Talanta 103 (2013) 267–275

Contents lists available at SciVerse ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Targeted mass spectral ratio analysis: A new tool for gas chromatography—mass spectrometry

Benjamin Kehimkar, Jamin C. Hoggard, Jeremy S. Nadeau, Robert E. Synovec *

Department of Chemistry, Box 351700, University of Washington Seattle, WA 98105-1700, USA

ARTICLE INFO

Article history:
Received 8 August 2012
Received in revised form
11 October 2012
Accepted 12 October 2012
Available online 23 October 2012

Keywords:
Targeted Analysis
Gas Chromatography
Mass Spectrometry
High Throughput
Automated

ABSTRACT

An algorithm, referred to as targeted mass spectral ratio analysis (TMSRA) is presented whereby the ratios between intensities as a function of mass channel (m/z) of a target analyte mass spectrum are used to automatically determine which m/z are sufficiently pure to quantify the analyte in a sample gas chromatogram. The standard perfluorotributylamine (PFTBA) was used to evaluate the reproducibility of the collected mass spectra, which aided in selecting a mass spectral threshold for TMSRA application to a subsequent case study. Results with PFTBA suggested that a threshold of all m/z at or above 1% of the highest recorded m/z intensity should be included for targeted analysis. For the case study, 1-heptene was selected as the target analyte and n-heptane was selected as the interfering compound. These two compounds were chosen since their mass spectra are very similar. Chromatographic data containing a pure peak for these analytes were extracted, and mathematically added at various temporal offsets to generate various degrees of chromatographic resolution, R_s, for the purpose of evaluating algorithm performance, and indeed, TMSRA successfully quantified 1-heptene. At the higher R_s studied ($0.6 \leq R_s \leq 1.5$) a deviation within $\pm$ 1% and a RSD generally below 1% were achieved for 1-heptene quantification. As the R_s decreased, the deviation and RSD both increased. At a $R_s = 0$, a deviation of ~9% and a RSD of~9% were achieved.

1. Introduction

Gas chromatography (GC) is a widely used analytical technique [1–7], for many applications including, but not limited to, the following areas: pesticides analysis [1,3], the petroleum industry [2], for toxicology [4], the food and beverage industry [1,5,6], and for metabolomic studies [1,5,7]. As with any chemical analysis technique, data analysis plays a necessary role. It is critical to rapidly and confidently analyze the data. For many applications, the analyst is interested in specific compounds, referred to as targeted analysis. In contrast, if the analyst is searching for initially unknown analytes; this is often referred to as untargeted analysis. This report is aimed at improving GC data analysis for targeted analysis.

The algorithm presented herein is designed for processing data acquired from a GC coupled to mass spectrometry (MS). The algorithm is also designed for situations where high throughput is needed. The samples may be complex (the total number of compounds may far exceed that of targeted analytes) and chromatographically and mass spectrally interfering compounds are likely. Taking these factors into account, it is expected that the chromatographic resolution, R_s, between compounds could be compromised for targeted analytes.

There are many methods one may use to quantify targeted analytes [3,4,7,8–10]. Some MS quantification methods use the selective ion monitoring (SIM) mode. By preselecting the mass channels (m/z) to monitor, and using the intensities to quantify the targeted analyte(s), there are benefits of greater sensitivity, greater signal-to-noise ratio (S/N), and lower limit-of-detection (LOD) when compared to scan mode. However, the SIM mode is less comprehensive than the scan mode, since the m/z range collected is generally small, thus it may be difficult to confidently identify all compounds of interest [3]. Also, the m/z used in the SIM mode may be interfered by signals from non-targeted compounds, which may change from run-to-run. While it may be possible to detect interfering compounds in many situations, this cannot be corrected without either a loss of confidence in results or additional experiments required to provide confidence.

There is also software used to deconvolute (mathematically resolve) and quantify analytes such as AMDIS [3,4,8], or ChromaTOF [10]. Though both AMDIS and ChromaTOF can be used for targeted analysis, they are primarily non-targeted analysis methods, and can be time consuming for data computation and interpretation steps. Alternatively, one could choose chemometric tools to deconvolute and quantify targeted analytes. Parallel

Abbreviation: TMSRA, Targeted Mass Spectral Ratio Analysis

* Corresponding author. Tel.: +1 206 685 2328; fax: +1 206 685 8665.
E-mail address: synovec@chem.washington.edu (R.E. Synovec).

0039-9140/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2012.10.043

factor analysis (PARAFAC) [11] and trilinear decomposition (TLD) [12] are two examples; both however, are multi-way methods, so to work on GC–MS data they require simultaneously analyzing several temporally aligned GC–MS runs, and this process can be time consuming. Another chemometric tool to consider is the generalized rank annihilation method (GRAM) [13]. GRAM is used for targeted analysis and can be applied to a single GC–MS run when data from a standard exists (i.e., requiring another GC–MS run with the standard, typically spiked into the original sample). A major drawback is that GRAM may fail if the targeted analyte is severely overlapped with another compound, or if the analyte peak in the standard (spiked sample) is misaligned with the analyte peak in the original sample, or if there is a difference between the shape of the peaks of the standard and the sample. Alternatively, multivariate curve resolution-alternating least squares (MCR-ALS) [14,15] could be used, though ambiguities can pose problems in "un-mixing" the targeted analyte. A daunting drawback for applying chemometrics is that in most instances the algorithms require their respective input parameters to be selected on a case by case basis, so having them run automatically is generally not possible [16]. Of course, one may achieve a better separation for a particular analyte by adjusting instrumental parameters; this approach requires running the sample more times, may compromise the separation of other analytes, and may not always be possible, especially when time and/or resources are limited. Hence, often modifying the GC separation conditions may not be an appealing solution.

In contrast, the data analysis method presented herein is designed to rapidly analyze GC–MS data, one run at a time, even with significant chromatographic peak overlap. The method and algorithm presented is referred to as targeted mass spectral ratio analysis (TMSRA), and it runs automatically with the following inputs and specifications: (1) a standard or representative sample chromatogram for alignment, (2) the processed spectra of the targeted analytes (see Theory Section), (3) the respective retention times and the m/z range scanned for each targeted analyte, and (4) code to automate the initialization process of the algorithm and storing the results. The basis of TMSRA is the following: as long as there are a few m/z of a given targeted analyte that are objectively determined to be sufficiently pure (not interfered significantly so bias is minimal), quantification is possible even at low R_s. TMSRA should be compatible with various ionization techniques including soft-ionization such as chemical ionization (CI-MS), provided the tuning of the instrument yields reproducible mass spectra. The determination of sufficient mass spectral purity is done by evaluating the ratios between intensities of m/z in a spectrum and comparing the results relative to those from a library spectrum to identify all m/z sufficiently pure to ensure accurate quantification. Caution must be taken to avoid overloading the signal due to a high concentration, thereby altering the observed fragmentation pattern with respect to the library. For low signal peaks of interest, with low S/N, it is recommended to apply noise reduction techniques such as averaging multiple replicates. While quantifying targeted analytes using selective m/z is not new, the approach we utilize to evaluate m/z is novel by requiring "internal consistency" with the list of sufficiently pure m/z that are found.

For the first study, the targeted analyte was perfluorotributylamine (PFTPA), commonly used to calibrate mass spectrometers. The objective of the first study was to evaluate the mass spectrometer reproducibility and aid in selecting a mass spectral threshold for the subsequent case study. The threshold is applied as a means of removing m/z channels from the analysis that either contained only noise or an irreproducible peak due to a low S/N. For the subsequent case study the targeted analyte was 1-heptene, and the interferent was n-heptane. They were selected due to their mass spectra being very similar with signal at many m/z in common. Chromatographic data regions containing a pure GC–MS peak of these analytes were extracted (for 5 replicates), and mathematically added at various temporal offsets to generate various degrees of peak overlap to facilitate evaluation of the TMSRA algorithm as the R_s between the two compounds was controlled, spanning a R_s range of 0 to 1.5. The quantitative deviation (and bias) and precision for 1-heptene was studied as a function of R_s.

2. Theory

When chromatographic resolution, R_s, is compromised between a targeted analyte, T, and an interfering unknown compound, U, the challenge is to determine which m/z of the target analyte are sufficiently pure for quantification. For this discussion, resolution is defined as [17]

$$R_s = ({}^{T}t_R - {}^{U}t_R)/\left(({}^{T}W_b + {}^{U}W_b)/2\right) \quad (1)$$

where t_R and W_b are the retention time and peak width at the base, respectively. W_b is defined as four times the standard deviation of the peak using a Gaussian peak model for simplicity.

For targeted analysis, the ratios between intensities of paired m/z channels serve as a guide to identify sufficiently pure i.e. uninterfered, m/z (or mass channels), meaning the m/z in a sample are compared to those in a library spectrum, L, which is a designated standard of a given target analyte. For subsequent clarity during visualization purposes, the m/z of L are sorted (indexed) according to intensity in descending order, although computationally, the sorting makes no difference in the results of the algorithm. Taking the respective indexed m/z for L, each intensity is divided by other intensities as shown in Eq. (2) to produce the ratio matrix, ${}^{L}R_{i,j}$. Where ${}^{L}I_i$ is the signal intensity of indexed m/z i in L, and ${}^{L}I_j$ is the signal intensity of indexed m/z j in L; the elements of the ratio matrix are given by,

$${}^{L}R_{i,j} = {}^{L}I_i / {}^{L}I_j \quad (2)$$

Alternatively, one may see this procedure as treating the intensities of indexed m/z values as a vector and taking the outer product between the vector and its respective (element by element) inverse. Calculating ${}^{L}R_{i,\,j}$ (for all i and j) generates a matrix of ratios between all m/z as shown in Table 1A. ${}^{L}R_{i,\,j}$ values on the diagonal are equal to 1, (because there, $i=j$, thus ${}^{L}I_i = {}^{L}I_j$). In the lower left triangle of the matrix, ${}^{L}R_{i,\,j}$ is less than a value of 1 (since there, ${}^{L}I_i < {}^{L}I_j$). The upper right portion of the matrix is simply the inverse of the lower left triangle (${}^{L}I_i > {}^{L}I_j$). To remove redundancy in ${}^{L}R_{i,\,j}$, it is useful to exclusively use the lower left triangle.

Likewise, the m/z index of L is used to sort the m/z of the given targeted analyte in the sample spectrum, S, resulting in the sorted

Table 1
(A) A resulted matrix when calculating $R_{i,j}$ for all i and j in the spectrum. (B) The resulted matrix for calculating $X_{i,j}$ for all i and j values.

(A) $R_{i,\,j}$ Intensity ratios for two m/z values			
$i=1$	$R_{(1,1)}=1$	$R_{(1,2)}>1$	$R_{(1,3)}\gg 1$
$i=2$	$R_{(2,1)}<1$	$R_{(2,2)}=1$	$R_{(2,3)}>1$
$i=3$	$R_{(3,1)}\ll 1$	$R_{(3,2)}<1$	$R_{(3,3)}=1$
	$j=1$	$j=2$	$j=3$
(B) $X_{i,\,j}$ Relative mass purity			
$i=1$	$X_{(1,1)}=0$	$X_{(1,2)}\approx \mp 0$	$X_{(1,3)}\approx \mp 0$
$i=2$	$X_{(2,1)}\approx \pm 0$	$X_{(2,2)}=0$	$X_{(2,3)}\approx \pm 0$
$i=3$	$X_{(3,1)}\approx \pm 0$	$X_{(3,2)}\approx \mp 0$	$X_{(3,3)}=0$
	$j=1$	$j=2$	$j=3$

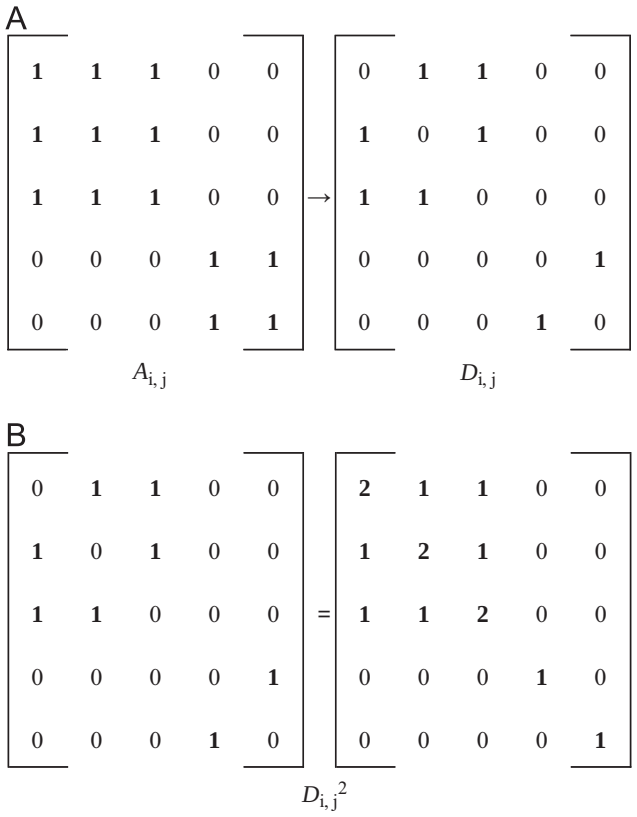

Fig. 1. (A) An example of the diagonal being set to a value of 0 (removed) in $A_{i,j}$ resulting in $D_{i,j}$. (B) The second matrix power of $D_{i,j}$ results in $D^2_{i,j}$. In the off diagonal, nonzero elements of $D^2_{i,j}$ indicates sets of at least three m/z that are candidates for quantification.

m/z of S corresponding with the sorted m/z of L. Similar to Eq. (2), the ratio matrix ${}^{S}R_{i,j}$ is calculated as follows,

$${}^{S}R_{i,j} = {}^{S}I_i / {}^{S}I_j \tag{3}$$

where ${}^{S}I_i$ is the signal intensity of indexed m/z i in S, and ${}^{S}I_j$ is the signal intensity of indexed m/z j in S. Next, the deviation between ${}^{L}R_{i,j}$ and ${}^{S}R_{i,j}$ relative to ${}^{L}R_{i,j}$ is calculated by Eq. (4), and the results are the elements of matrix $X_{i,j}$ (see Table 1B),

$$X_{i,j} = ({}^{S}R_{i,j} - {}^{L}R_{i,j}) / {}^{L}R_{i,j} \tag{4}$$

where $X_{i,j}$ is the *indexed mass purity matrix* and the division is done element by element. Pure m/z in S, relative to each other, have corresponding $X_{i,j}$ values close to 0, affected primarily by noise. If a m/z in S is being interfered, the corresponding $|X_{i,j}|$ absolute values should be greater since the intensity relative to other m/z would not match those in L. Looking at Table 1B, the elements on the diagonal have a value equal to 0 (because ${}^{S}R_{i,j} = {}^{L}R_{i,j} = 1$). Another noticeable pattern is that when the values of i and j in $X_{i,j}$ are transposed, then the sign changes, as confirmed by Eq. (4). Again, to remove redundancy, only the lower left triangle is used.

As the number of indexed m/z values increases, so does the complexity of ${}^{S}R_{i,j}$, ${}^{L}R_{i,j}$ and $X_{i,j}$. To help interpret the results embedded in plots of $X_{i,j}$, *connectivity (or adjacency) matrices* $C_{i,j}$ are derived from the lower left triangle of $X_{i,j}$. Detailed explanation on connectivity matrices can be found elsewhere [18–20], and interested readers are encouraged to further read about connectivity matrices as desired. Briefly, as used herein, connectivity matrices describe the relationship (i.e., the connection) between the elements of two vectors (here, indexed m/z) as a binary map. The matrix $C_{i,j}$ is constructed by setting a tolerance value for the lower left triangle of $|X_{i,j}|$, the most deviation the analyst is willing to accept from ${}^{S}R_{i,j}$ (analyte in a sample) with respect to ${}^{L}R_{i,j}$, (the target analyte standard), with the result being a logical matrix containing a value of 1 where $X_{i,j}$ elements are within the desired tolerance, and a value of 0 where they are not. A lax tolerance will have many elements with a value of 1 in $C_{i,j}$, while a strict tolerance will have few. The aforementioned tolerance is a value between 0 and 1, and can alternatively be seen as a percent deviation from the standard; for example a value of 0.1 can be seen as a tolerance of up to 10% deviation from ${}^{L}R_{i,j}$.

The $C_{i,j}$ matrix is transposed $C^{T}_{i,j}$ and added to itself $C_{i,j} + C^{T}_{i,j}$, producing an 'uncondensed' connectivity matrix $A_{i,j}$. In interpreting the matrix $A_{i,j}$, each element can be thought of as indicating the purity of pairs of indexed m/z, or alternatively, each row i, having a length equal to the number of indexed m/z being analyzed, indicating the purity of the m/z indexed by i with respect to those m/z indexed by j. Each element whose value is 1, meets the tolerance criteria with respect to the ratio of indexed m/z i and j, meaning the relative intensities of the m/z are consistent (to within tolerance) with those found in L.

One often finds patterns or rows in $A_{i,j}$ that appear identical to each other; this can be interpreted as a confirmation (from a different indexed m/z as a function of i) that the ratios of the particular set of indexed m/z indicated in $A_{i,j}$ are consistent. The diagonal is made equal to elements with a value of 0, because these elements will always give a value of 1 for tolerances greater or equal to a value of 0, (since $X_{i,j} = 0$ when $i = j$). This also clears rows that only have a value of 1 on the diagonal as they can be interpreted as "the only indexed m/z as a function of j that are consistent with indexed m/z as a function of i is when $i = j$." The resulting matrix, after setting the diagonal of $A_{i,j}$ to a value of 0, was denoted as $D_{i,j}$. However, slight differences between sets of m/z in $D_{i,j}$ indicated as pure complicate the logic to determine which set of m/z are consistent. To proceed, the $D_{i,j}$ matrix is taken to the second matrix power, $D^2_{i,j}$. The resulting effect is to soften the strict logical tolerance criteria of the connectivity matrix (as explained below) in order to achieve the goal of including at least three indexed m/z in each valid set. A minimum of three indexed m/z are needed; this is because it is not possible to recover an original set of m/z having fewer then three members found in $A_{i,j}$ after setting the diagonal to a value of 0 (as seen in Fig. 1A) and then squaring the resulting matrix, $D_{i,j}$ to $D^2_{i,j}$ (seen in Fig. 1B). The matrix multiplication of $D_{i,j}$ to itself annexes or connects the indexed m/z that are consistent with at least one indexed m/z within the set (known as including or connecting through the nearest neighbor) hence moderately increasing the size of the indexed m/z set and softening the connectivity matrix criteria (as seen in Fig. 1B). All nonzero elements in $D^2_{i,j}$ are set to a value of 1 and plotted as a matrix of dots, each dot represents $X_{i,j}$ values that meet the set tolerance (including through the nearest neighbor). All sets (rows) with three or more elements equal to a value of 1 are extracted (excluding duplicate sets); the result is a collection of various sets of m/z that could (in principle) be used to quantify the target analyte T.

Before explaining the decision process to determine which set of m/z to use for quantification, the process of how the peak height is calculated needs to be explained. Referring back to L, the sum of intensities of a set of m/z in L can be thought of as $\Sigma^{L}I_{set}$, or a fraction, q, of the summed intensities of all m/z in L,

$$q = \sum{}^{L}I_{set} / \sum{}^{L}I_{total} \tag{5}$$

The peak height of analyte T, in S (the sample) is calculated by taking the sum of intensities of a set of m/z multiplied by its respective reciprocal of q (calculated from L), and is given by,

$$\sum^{S} I_{\text{total(calc)}} = \sum^{S} I_{\text{set}}/q \tag{6}$$

The calculated TIC equivalent peak height $\Sigma^{S} I_{\text{total (calc)}}$ for the sample is a close approximation of a pure TIC peak height of S; this procedure facilitates comparison across different sets of m/z. Implementing this procedure for every set of m/z extracted from $D_{i,j}^2$, gives a list of possible quantitative answers for the TIC equivalent peak height of T. At this point one may ask "which peak height is the best answer?" Logically, if T is present, and if there are several m/z of T sufficiently unaffected by interfering compounds U, then any interference from U would only add to the signal of the T's peak height. Therefore, the most correct set of m/z to use for quantifying T is the set of m/z producing the lowest TIC equivalent peak height (i.e., which minimizes bias). Indeed, using a SIM mode approach, while fast, does not provide quantitative information to minimize bias, such as done in the method we employ.

Finally, for the purposes of this report to study quantitative accuracy and precision as a function of R_s, per the experimental design for the case study with 1-heptene and n-heptane, the calculated peak heights for each run are divided by their corresponding pure peak heights (same sample replicate prior to undergoing simulated peak overlap). Thus, the concentration ratio, F, of the analyte in the sample relative to the analyte in the library standard, is determined using the ratio of TIC equivalent peak heights, following from Eqs. (5) and (6),

$$F = \sum^{S} I_{\text{total(calc)}} / \sum^{S(pure)} I_{\text{set}} \tag{7}$$

thereby 'normalizing' the quantitative result for determining concentration. For the case study, ideally, $F=1$ should be experimentally obtained.

3. Experimental

3.1. Instrumentation

The instrument used was an Agilent 6890 GC equipped with a 5973A Mass Spectrometer with unit mass resolution and a 7683B auto injector (Agilent Technologies, Palo Alto, CA, USA). Samples were run through a 30 m J&W 122-503E DB-5 capillary column, with 250 μm i.d., and 0.5 μm film thickness (Agilent Technologies, Palo Alto, CA, USA). The inlet temperature was 250 °C. The temperature program utilized an initial temperature of 35 °C for 1.15 min, then ramped at a rate of 5 °C/min to 60 °C, then ramped at 35 °C/min to 300 °C with this final temperature maintained for 1.5 min. The flow rate was held at 1.4 ml/min with an initial head pressure of 2.48 psig (17.1 kPa above ambient pressure) using H_2 as carrier gas. The mass spectrometer parameters were as follows: The electron energy (filament voltage) parameter was set to 70 eV, and the m/z scan range was 50 to 220 for the PFTBA study and 40 to 140 for the case study with 1-heptene and n-heptane. The 'sample' parameter (the number of times the signal abundance of each m/z is recorded then averaged) was set to 4, and the data acquisition rate was set to 20 Hz. For the case study, 0.5 μl of each sample was injected via the auto injector using split-less injection mode.

3.2. The PFTBA study

Since the proposed computational method relies heavily on the ratios between the intensities of m/z, the reproducibility of the spectrum and the mass spectral threshold needed investigation. Perfluorotributylamine (PFTBA) (Agilent Technologies, Palo Alto, CA, USA) was used to evaluate these issues since it is the tuning compound for the quadrupole MS. Since perfluoroalkanes are commonly used in tuning MS instruments [21] it is assumed that for the specific vial of PFTBA the spectrum is fairly reproducible, and therefore could be used to gauge the run-to-run variation of m/z over a wide range of intensities. The intensities of the aforementioned m/z in the range 50 to 220 were measured for each spectrum. Twelve PFTBA spectra were collected over three weeks, in order to assess longer term reproducibility issues. The spectra were then manually imported into Matlab2009b (Mathworks, Natick MA, USA) where the pool of spectra were analyzed using principal component analysis (PCA) software from PLS_Toolbox Version 6.2, (Eigenvector Research, Inc., Wenatchee WA, USA), and the scores of each PFTBA spectrum were used to calculate the Mahalanobis distance to determine the spectrum best suited, i.e., that with the smallest average Mahalanobis distance, for use as the library [26–28]. The remaining spectra were then analyzed with respect to the library using a subsection of the TMSRA algorithm; the subsection entails from the start of the original algorithm to the calculation of the index mass purity plot (Eq. (4)). The reason for omitting the quantification portion of the algorithm is because the focus of this particular study was to qualitatively analyze the spectra and to validate that the run-to-run variance was relatively small.

3.3. Sample preparation

The test compounds reported herein for the case study are 1-heptene and n-heptane, (Sigma-Aldrich, St. Louis MO, USA) were used to make mixtures (by mass between 450 to 750 mg). A 100 to 150 mg aliquot from the stock of standard mixtures was diluted in 5000 mg of acetone, resulting in seven standards of various mixtures of compounds of known concentration between 2500 to 3700 ppm. In the case study, 1-heptene was designated as the target while n-heptane was used as the interferent. Other case studies with other compounds were performed, with results not presented for brevity, since the 1-heptane: n-heptane case was the most challenging and interesting case. Samples were run in sets of 6 replicates and the regions containing the target analytes were extracted. One 1-heptene replicate was arbitrarily chosen as the reference and its mass spectrum was designated as the library spectrum, the remaining five peaks were used for simulated peak overlap with the interfering n-heptane peak.

3.4. Data processing (case studies)

Data analysis was performed on an Athlon II X2 (AMD, 3.1 GHz) with 4 GB of RAM. Prior to analysis, GC–MS data files were imported into Matlab2009b (Mathworks, Natick MA, USA) using an in-house developed Matlab function. GC–MS data were subjected to m/z channel 'de-skewing' (details found elsewhere [22]) and the chromatograms were aligned using in-house software [2,23–25]. For simulated chromatographic overlap, regions of the chromatograms containing well resolved analyte peaks (1-heptene and n-heptane) were extracted. For each replicate, the two chromatographic regions, one containing the targeted 1-heptene peak, with the other containing the n-heptane peak serving as the interferent, underwent baseline correction, and then were summed along the time axis to achieve peak overlap at a desired R_s as calculated in Eq. (1). The R_s range studied was 0 to 1.0 in increments of 0.1, and included the R_s values 1.25 and 1.5. The 'simulated peak overlap data' underwent data analysis processing following the algorithmic steps described

in the Theory section. All the necessary information (the chromatographic regions used to extract the library spectrum (L), m/z range, expected t_R) was input into the algorithm, which was allowed to process without further user intervention. The spectrum of S used for analysis was collected (via summation) across $\pm\sigma$ of the expected ${}^{T}t_R$ of the 1-heptene peak from which L was derived. The mass spectral threshold used was 1%. This threshold was chosen because of the results of the PFTBA experiment (as will be described later herein), which indicated the relative intensities of m/z with an intensity greater than 1% (of the most intense m/z) are sufficiently reproducible. Since there was no prior knowledge of what tolerance would be best to use in generating $C_{i,j}$ (and the eventual $D^2_{i,j}$) with respect to the calculation of the peak height for a given T, a broad range of tolerances were investigated, ranging from 0.025 to 0.25, in increments of 0.025 (corresponding to deviation of 2.5% to 25%) for each R_s investigated. For brevity and clarity, only tolerances ranging from 0.05 to 0.2 are presented. Results (calculated peak heights and corresponding sets of m/z) obtained using TMSRA, for each R_s and at each tolerance value were plotted in Excel 2007 (Microsoft, Redmond WA, USA). After analysis, for each run, calculated peak heights were divided by their corresponding pure peak heights (same run prior to undergoing simulated peak overlap with an interfering peak), thereby 'normalizing' the quantitative result that should ideally have a value of 1 based upon Eq. (7). This process was performed to minimize the error due to instrumentation and help study the method presented in terms of both precision and accuracy.

4. Results and discussion

The purpose of the PFTBA study was to assess the precision of the TMSRA algorithm and to determine suitable algorithmic tolerance levels. The library spectrum and a representative sample spectrum are shown in Fig. 2. The index key for PFTBA, indexed m/z (ordered by intensity) is shown in Table 2. PFTBA exhibited appreciable signal at 14 m/z values, and for example, m/z=219 is the most intense so it has an index position of 1 (m/z=69 is normally the most intense, but the tune parameters were set to emphasize higher, more selective m/z). Fig. 2 shows the spectra of the library and the sample and how they directly compare. In this case there are a few m/z in the targeted analyte PFTBA in the sample (Fig. 2B) that are lower in intensity when compared to the library (Fig. 2A); the variation is subtle but noticeable. Though it helps to look at the spectra directly it is important to go back to the original premise of this report and focus on the intensity ratios of the m/z as they offer a more sensitive means of comparing spectra.

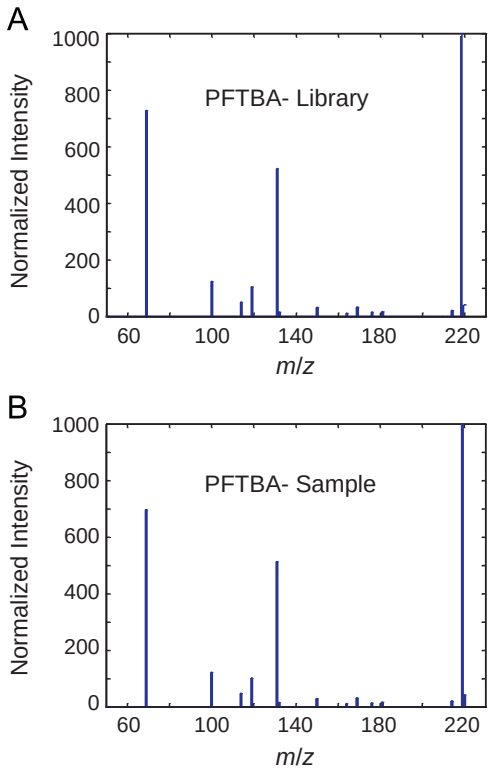

Fig. 2. (A) The spectrum of PFTBA (library), m/z range of 50–220. (B) The spectrum of a PFTBA run (sample), m/z range of 50–220.

Table 2
The index key for PFTBA containing m/z of intensity above the minimum threshold.

Index	m/z
1	219
2	69
3	131
4	100
5	119
6	114
7	220
8	169
9	150
10	214
11	132
12	176
13	181
14	164

Results for comparing the spectra of the sample to the library are shown in Fig. 3. In Fig. 3A, the indexed mass ratio matrix ${}^{L}R_{i,j}$ of the library spectrum is presented. In Fig. 3B, the first column of ${}^{L}R_{i,j}$ (j=1) is shown and using Eq. (2), the mass ratio matrix becomes easier to interpret: for the first column, the indexed m/z are divided by the first indexed m/z. Fig. 3C is analogous to Fig. 3A, only the sample spectrum is analyzed and ${}^{S}R_{i,j}$ is obtained, using Eq. (3). Likewise, Fig. 3D presents the first column of ${}^{S}R_{i,j}$ in the same manner as described in Fig. 3B (only Eq. (3) is used instead). Fig. 3E is the indexed mass ratio matrices or 'Mass Purity Plots' obtained by applying Eq. (4); the differences in the intensity ratios between the library and the sample are subtle, which is expected for the PFTBA replicates. Another reason to use the intensity ratios instead of the intensities directly is that the ratios also serve as a means of normalization across the various m/z channels. In Fig. 3E, the 'Indexed Mass Purity Plot', one also observes that most m/z have $X_{i,j}$ values that fall below 0.05, (corresponding to 5% deviation) from the library. Fig. 3F presents the first column of $X_{i,j}$; only three m/z deviate beyond the arbitrary 0.05 tolerance value. More importantly there are no indexed mass channels with a large m/z ratio variance, a result that would indicate the existence of a bias in the m/z intensities.

Finally, statistical results for all 11 replicates of PFTBA showing the variance with respect to the library are shown in Fig. 4. Fig. 4A is the result of averaging $X_{i,j}$, i.e., the results after using Eqs. (2, 3 and 4) for the aforementioned 11 replicates with respect to the library standard spectrum. Fig. 4B complements Fig. 4A by showing the RSD for $X_{i,j}$. One may notice that the ratios between m/z are generally reproducible. The more intense m/z (top 9) typically have a RSD below 5%, and the remaining m/z have RSD values typically below 10%. These

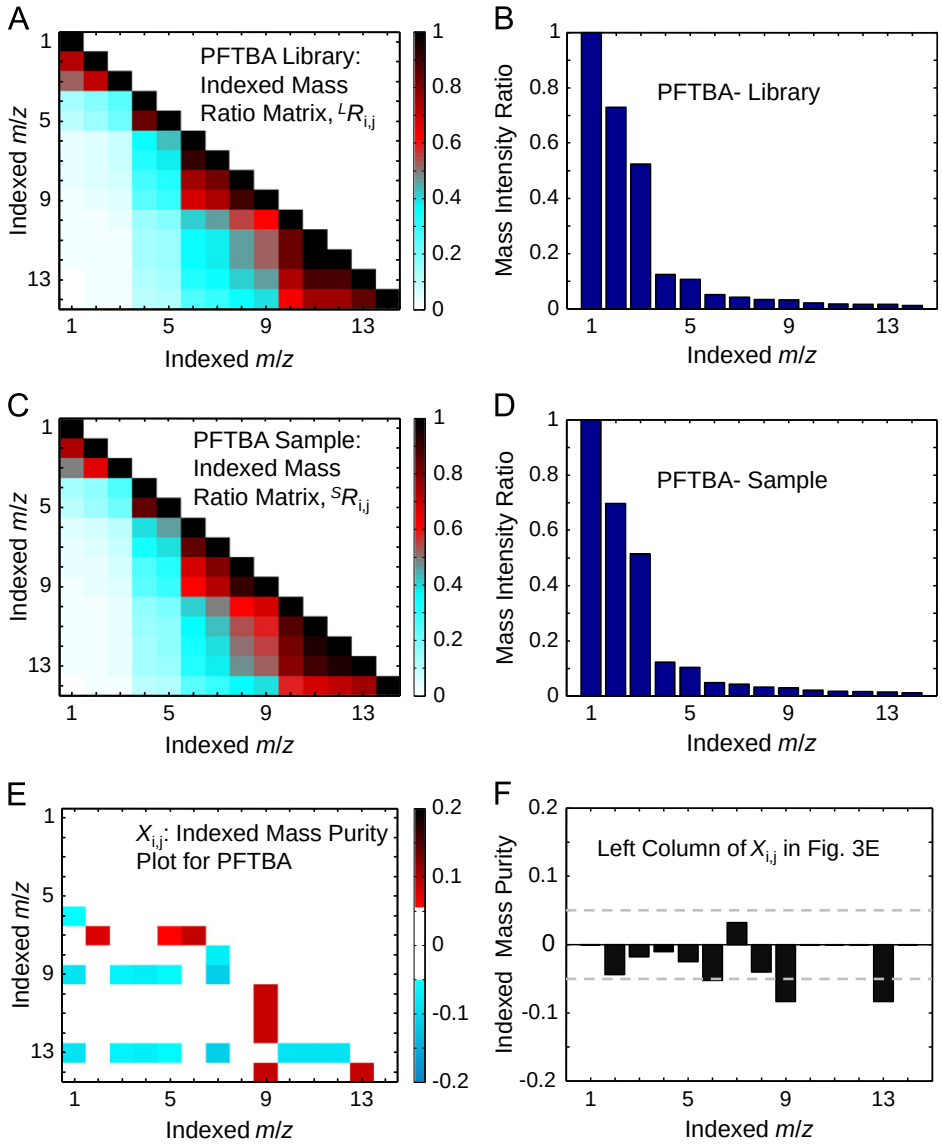

Fig. 3. (A) Indexed Mass Ratio Matrix, $^{L}R_{i,j}$, for the PFTBA indexed library spectrum, obtained by taking the outer product of the spectrum with respect to its inverse as calculated in Eq. (2). (B) Indexed Mass Ratios of the first column of $^{L}R_{i,j}$ seen in Fig. 3A. (C) Indexed Mass Ratio Matrix, $^{S}R_{i,j}$, for the PFTBA indexed sample spectrum, obtained by taking the outer product of the spectrum with respect to its inverse as calculated in Eq. (3). (D) Indexed Mass Ratios of the first column in $^{S}R_{i,j}$ seen in Fig. 3C. (E) Indexed Mass Purity Plot results from calculating the difference between $^{L}R_{i,j}$ and $^{S}R_{i,j}$, as shown in Eq. (4). (F) Indexed Mass Purity Ratios of the first column in $X_{i,j}$ seen in Fig. 3E. The dashed lines mark the $\pm$5% deviation.

results suggest that based upon using PFBTA for assessing tolerances, 1% of the maximum intensity appears to be a suitable threshold for analyzing mass spectra. This threshold was applied in the subsequent case study demonstration in the next section of this report. It is important to note that the performance of the TMSRA algorithm depends heavily on the accuracy of the library spectrum as a representation of a given target analyte.

For the case study we computationally generated a full range of chromatographic overlap (from a R_s of 0 to 1.5, using Eq. (1)) of the target analyte, 1-heptene, with the interferent, n-heptane. Their respective spectra are shown in Fig. 5A. One can see that this is a relatively challenging case study since the two compounds exhibit significant signal intensity at many of the same m/z. Since there are few m/z that are selective for 1-heptene, its quantification would be a difficult feat at lower R_s, especially if the analyst does not know *a priori* which m/z may or may not be sufficiently selective. The chromatographic peak heights of the two compounds are similar, as seen in Fig. 5B. The sorted mass index information is provided in Table 3. The R_s values between 1-heptene and n-heptane that were analyzed were 1.5, 1.25 and from 1.0 to 0 at 0.1 increments. As stated in the Theory section, there was no *a priori* knowledge of which tolerance values were best suited to apply on the lower left triangle of $|X_{i,j}|$ for constructing the connectivity matrices ($C_{i,j}$ and eventually $D^2_{i,j}$, shown in Fig. 1B) to use in the m/z selection process for the eventual quantification of the target analyte peak. Moreover it was conceivable that the mass spectral run-to-run variation would result in different runs having different optimal tolerance values. Therefore, a range of different tolerance values were studied for each R_s; the range of tolerance values were from 0.025 to 0.25 at 0.025 increments (corresponding to the maximum deviation and/or bias) the sample spectrum is allowed to deviate from the library in order to be considered as 'pure'. An example of data encountered in intermediate steps in the algorithm is shown in Fig. 6 at R_s=0.3. The core of the process is using

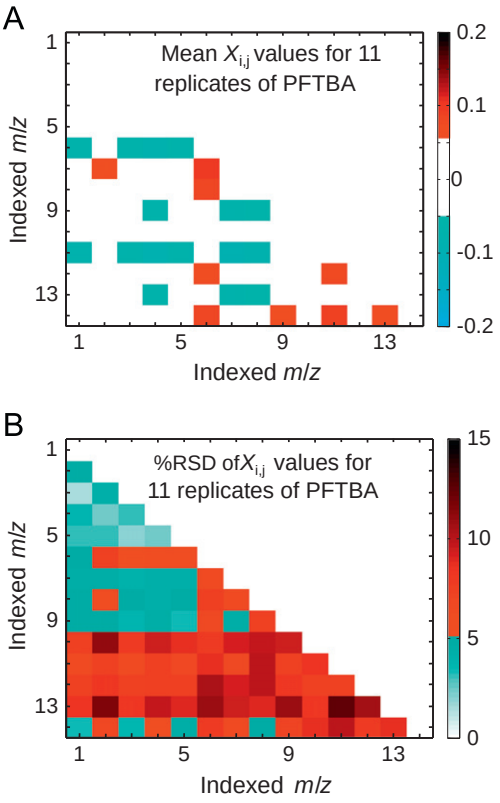

Fig. 4. (A) Statistical results (average of PFTBA pairing of 11 samples with the single library standard) for the Indexed Mass Purity plots of PFTBA. (B) Statistical results (%RSD) for Indexed Mass Purity Plots of PFTBA.

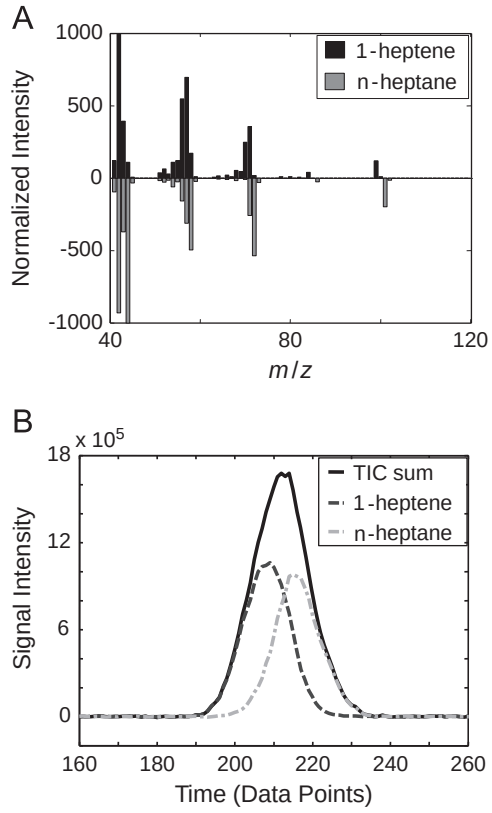

Fig. 5. (A) Case study with the mass spectrum of 1-heptene compared to n-heptane. (B) Total ion current (TIC) signal illustrates the overlap resulting from overlaying and summing two GC–MS chromatographic portions, 1-heptene (on the left) at R_s=0.3 from n-heptane.

Table 3
The index key for sorted mass channels of 1-heptene.

Index	m/z	Index	m/z
1	41	12	43
2	56	13	51
3	55	14	67
4	42	15	68
5	70	16	83
6	69	17	50
7	57	18	53
8	40	19	65
9	54	20	71
10	98	21	63
11	53	22	66

Eqs. (2) and (3) to generate the intensity ratio matrices ${}^{L}R_{i,j}$, ${}^{S}R_{i,j}$, (such as Fig. 6A, B, respectively), and analyzing the differences between them using Eq. (4), resulting in $X_{i,j}$ (shown in Fig. 6C). Setting the right tolerance values for $C_{i,j}$ is a balancing act; too strict and some good m/z will be thrown out, too lax and contributions from an interfering compound will add to the sum of intensity, $\Sigma^{S}I_{\text{set}}$. Fortunately, the algorithmic computation time is short, so many tolerance values can be rapidly evaluated. An example of a connectivity matrix, $D^{2}_{i,j}$, resulting from a good tolerance selection is shown in Fig. 6D. Each row pertains to the indexed m/z. A dot represents an m/z that is consistent with respect to the referenced indexed m/z (row number) while a blank space represents otherwise. The black dots in the sixth row represent the m/z set that TMSRA has found as the least biased solution at the given tolerance value. In Fig. 6D, those indexed m/z are 2, 3, 6, 10, 15, 16, and 22, which correspond to actual m/z 56, 55, 69, 98, 68, 83, and 66, respectively. The final results (TIC equivalent peak height as a function of tolerance values) can be compared either by the user, or ultimately by an automated tolerance value selection process.

Statistical results of quantifying 1-heptene using Eq. (5) can be found in Fig. 7A and B. The sets of m/z used for quantification were recorded and the average size of sets of mass channels is shown in Fig. 7C. One can observe in Fig. 7A that for all tolerances at R_s between 0.6 to 1.5, quantitative results are in good agreement, with the concentration ratios, F, having RSD values typically below 2% (most being below 1%) and deviating generally within $\pm$ 1% from the expected value of 1. At low R_s (0 to 0.5, see Fig. 7B) as the R_s decreases, the deviation in F from a value of 1 increases (for the best tolerance values, which are typically 0.1 and 0.125, the deviation in F is $\pm$ 1% to 9%) and the RSD increases again, (for the best tolerances, the RSDs is between 2% to 9%). Looking at Fig. 7C one can see that as the R_s decreases, the average size of the set of m/z used for quantification decreases. Also as the tolerance values decrease, the average size of the set of m/z shrinks. Note that if the size of the best set of m/z becomes too small (below three m/z) due to a decrease in tolerance or R_s, the next set of m/z (regardless of its size) that produces the smallest $\Sigma^{S}I_{\text{total (calc)}}$ (and therefore the smallest F) is chosen for quantification.

Regarding the tolerance value for the connectivity matrix for this case study (and others not shown for brevity), upon comparing results for individual replicates, no single tolerance value repeatedly performed best in every replicate sample, though in the case study presented the tolerance value of 0.1 generally fared best, it did not produce the best results in every replicate run and at every R_s value (sometimes the tolerance value 0.125 performed better). For practical use, an algorithm that converges on the most correct tolerance value on a run by run basis should be considered. Finally, the computation

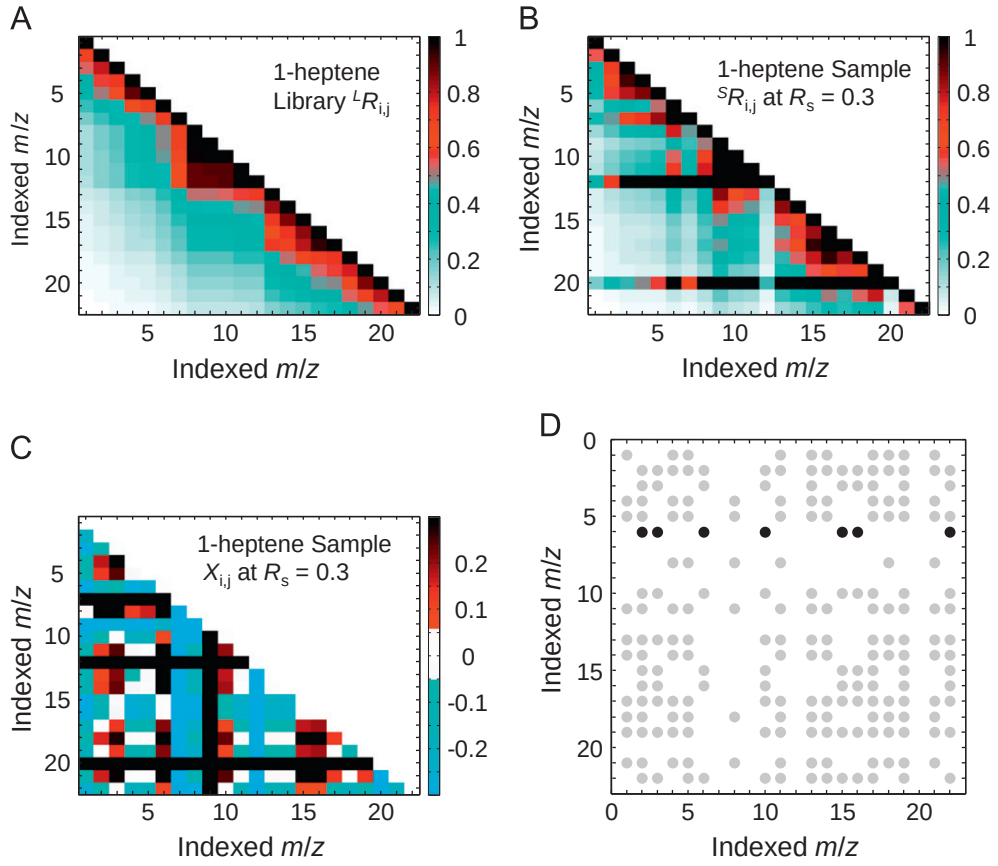

Fig. 6. (A) Indexed Mass Ratio Matrix $^{L}R_{i,j}$ for the case study (data in Fig. 5). Obtained using the 1-heptene indexed library spectrum and using Eq. (2). (B) Indexed Mass Ratio Matrix $^{S}R_{i,j}$ of the case study (at R_s=0.3). Obtained by using the sample spectrum indexed according to $^{L}R_{i,j}$, and using Eq. (3). (C) Indexed Mass Purity Plot ($X_{i,j}$) result from calculating the difference between Fig. 6A and B, using Eq. (4). (D) $D^2_{i,j}$ connectivity matrix with tolerance at 0.10 (10%), derived from Indexed Mass Purity Plot (Fig. 6C) for the condition that R_s=0.3.

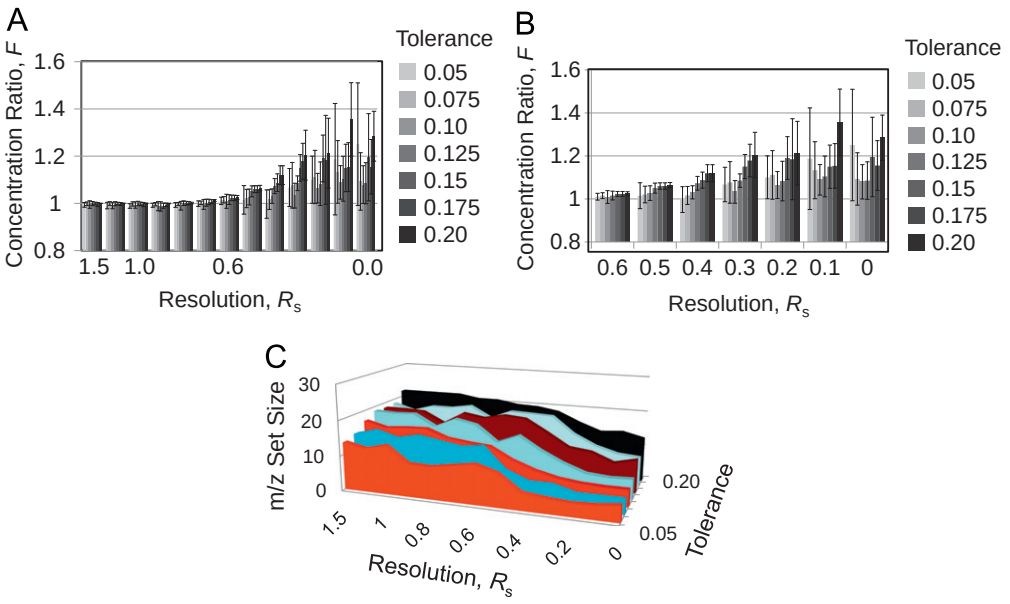

Fig. 7. (A) Quantitative results for the concentration ratio, *F*, for 1-heptene at various R_s, by selecting the best *m*/*z* set at various tolerance values, using Eqs. (5, 6, and 7). The best *m*/*z* set provides the minimum quantification value (lowest potential for bias). Tolerances 0.025, 0.225 and 0.25 were omitted, and only R_s values 1.5, 1.0, 0.6 and 0.0 were indicated on the *x*-axis for clarity. The full range of R_s values are 1.5, 1.25, and from 0.0 to 1.0 by increments of 0.1. (B) Quantitative results, *F*, for 1-heptene at various tolerances, at R_s ranging from 0 to 0.6. Tolerances 0.025, 0.225 and 0.25 were omitted for clarity. (C) Average size of set of *m*/*z* used to calculate the peak maximum shown in Fig. 7A, B as a function of R_s and tolerance for $C_{i,j}$. The tolerance values shown, in ascending order are 0.05 0.075, 0.10, 0.125, 0.15, 1.175, and 0.20.

time of the TMSRA algorithm was investigated. We measured the time required for the algorithm to analyze a portion of a given sample GC–MS chromatogram that contains a target analyte. For one hundred runs the average computation time was 0.027 s with a standard deviation of about 7.5 ms. Note that the computation time could potentially be shortened by further optimization of the code.

5. Conclusions

The TMSRA algorithm has the ability to quantify, using sufficiently selective m/z, a targeted analyte over a wide range of R_s relative to an interfering peak, with a fast computation time. At this stage the basic features of TMSRA have been presented with method development (PFTBA study) and method demonstration (case study with 1-heptene and n-heptane). For the targeted analyte 1-heptene in the case study, at the higher R_s studied ($0.6 \leq R_s \leq 1.5$) a deviation in the concentration ratio, F, within $\pm$ 1% and a RSD generally below 1% were achieved. As the R_s decreased, the deviation and RSD both increased. At a $R_s = 0$, a deviation of $\sim$9% and a RSD of $\sim$9% were achieved. For this challenging case study, as R_s goes to a value of 0, the deviation approached the tolerance level selected. While other approaches to use selective m/z with GC–MS have been reported, we utilized a novel approach requiring "internal consistency" with the list of sufficiently pure m/z that are found. This internal consistency should result in less deviation (more accurate) quantification for high speed analysis of targeted analytes with GC–MS. More rigorous testing and application of TMSRA will be the basis for future investigations.

References

[1] F.L. Dorman, J.J. Whiting, J.W. Cochran, J. Gardea-Torresdey, Anal. Chem. 82 (2010) 4775.
[2] J.S Nadeau, B.W. Wright, R.E. Synovec, Talanta 81 (2010) 120.
[3] W. Zhang, P. Wu, C. Li, Rapid Commun. Mass Spectrom. 20 (2006) 1563.
[4] M.R. Meyer, F.T. Peters, H.H. Maurer, Clin. Chem. 56 (4) (2010) 575.
[5] E.M. Humston, J.D. Knowles, A. McShea, R.E. Synovec, J. Chromatogr. A 1217 (2010) 1963.
[6] A. Genovese, R. Dimaggio, M.T. Lisanti, P. Piombino, L. Moio, Ann. di Chim. 95 (2005).
[7] E. Jellum, O. Stokke, L. Eldjarn, Anal. Chem. 45 (7) (1973) 1099.
[8] J.M. Halket, A. Przyborowska, S.E. Stein, W.G. Mallard, S. Down, R.A. Chalmers, Rapid Commun. Mass Spectrom. 13 (1999) 279.
[9] W.G. Pool, J.W. de Leeuw, B. van de Graaf, J. Mass Spectrom. 32 (1997) 438.
[10] Y. Koh, K.K. Pasikanti, C.W. Yap, E.C. Chan, J. Chromatogr. A 1217 (2010) 8308.
[11] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149.
[12] E. Sanchez, B.R. Kowalski, J. Chemom. 4 (1990) 29.
[13] C.G. Fraga, J. Chromatogr. A 1019 (2003) 31.
[14] M. Daszykowski, B. Walczak, Trends in Anal. Chem. 25 (11) (2006) 1081.
[15] M. Garrido, F.X. Rius, M.S. Larrechi, Anal. Bioanal. Chem. 390 (2008) 2059.
[16] J.C. Hoggard, R.E. Synovec, Anal. Chem. 80 (2008) 6677.
[17] J.C. Giddings, Unified Separation Science, John Wiley & Sons, USA, 1991.
[18] Y. Marrero-Ponce, E.R. Martínez-Albelo, G.M. Casañola-Martín, J.A. Castillo-Garit, Y. Echevería-Díaz, V.R. Zaldivar, J. Tygat, J.E. Rodriguez Borges, R. García-Domenech, F. Torrens, F. Pérez-Giménez, Mol. Divers 14 (2010) 731.
[19] M. Randic, J. Chem. Inf. Comput. Sci. 37 (1997) 672.
[20] E. Estrada, J. Chem. Inf. Comput. Sci. 35 (1995) 31.
[21] F.W. McLafferty, F. Tureček, in: Books (Ed.), Interpretation of Mass Spectra, University Science, 1993.
[22] W.G. Pool, J.W. de Leeuw, B. van de Graaf, J. Mass Spectrom. 31 (1996) 213.
[23] K.M. Pierce, J.L. Hope, K.J. Johnson, B.W. Wright, R.E. Synovec, J. Chromatogr. A 1096 (2005) 101.
[24] N.E. Watson, M.M. Van Wingerden, K.M. Pierce, B.W. Wright, R.E. Synovec, J. Chromatogr. A 1129 (2006) 111.
[25] K.M. Pierce, B.W. Wright, R.E. Synovec, J. Chromatogr. A 1141 (2007) 106.
[26] T.I. Dearing, J.S. Nadeau, B.G. Rohrback, L.S. Ramos, R.E. Synovec, Talanta 83 (2011) 738.
[27] M. Fransson, S. Folestad, Chemom. Intell. Lab. Syst. 84 (2006) 56.
[28] M. Daszykowski, B. Walczak, J. Chromatogr. A 1176 (2007) 1.