RCM

Determination of elemental compositions by gas chromatography/time-of-flight mass spectrometry using chemical and electron ionization

Salvatore Abate[2], Yun Gyong Ahn[1,3], Tobias Kind[1], Tommaso R. I. Cataldi[2] and Oliver Fiehn[1]*

[1]UC Davis Genome Center, 451 Health Sci. Dr., Davis, CA 95616, USA
[2]Department of Chemistry, Università degli Studi della Basilicata, 85 N. Sauro, Potenza 85100, Italy
[3]Korea Basic Science Institute, 126-16 Anam-Dong Sungbuk-Ku, Seoul 136701, South Korea

Many metabolomic applications use gas chromatography/mass spectrometry (GC/MS) under standard 70 eV electron ionization (EI) parameters. However, the abundance of molecular ions is often extremely low, impeding the calculation of elemental compositions for the identification of unknown compounds. On changing the beam-steering voltage of the ion source, the relative abundances of molecular ions at 70 eV EI were increased up to ten-fold for alkanes, fatty acid methyl esters and trimethylsilylated metabolites, concomitant with 2-fold absolute increases in ion intensities. We have compared the abundance, mass accuracy and isotope ratio accuracy of molecular species in EI with those in chemical ionization (CI) with methane as reagent gas under high-mass tuning. Thirty-three peaks of a diverse set of trimethylsilylated metabolites were analyzed in triplicate, resulting in 342 ion species ($[M+H]^+$, $[M–CH_3]^+$ for CI and $[M]^{+\cdot}$, $[M–CH_3]^{+\cdot}$ for EI). On average, CI yielded 8-fold more intense molecular species than EI. Using internal recalibration, average mass errors of 1.8 ± 1.6 mm/z units and isotope ratio errors of $2.3 \pm 2.0\%$ (A+1/A ratio) and $1.7 \pm 1.8\%$ (A+2/A ratio) were obtained. When constraining lists of calculated elemental compositions by chemical and heuristic rules using the Seven Golden Rules algorithm and PubChem queries, the correct formula was retrieved as top hit in 60% of the cases and within the top-3 hits in 80% of the cases. Copyright © 2010 John Wiley & Sons, Ltd.

Gas chromatography coupled to mass spectrometry (GC/MS) is a robust technique that combines high sensitivity and specificity for suitable analyte classes and is one of the most frequently used tools for profiling primary metabolites. In most cases GC/MS experiments are performed in electron ionization (EI) mode with compound identification based on matching acquired spectra to mass spectral databases libraries.[1] The versatility of large libraries like the NIST08 mass spectral resource[2] lies in the fact that EI mass spectra are comparable over a wide range of different types of mass spectrometers from different vendors, although quadrupoles may be tuned to preferentially transmit high m/z ions which thus result in slightly different ion abundances from those in time-of-flight mass spectra. As outlined in Fiehn,[3] two major challenges remain to improve GC/MS methods for profiling metabolite: rapid annotation and identification of unknown peaks, and integration of biological background knowledge into data interpretation. There is, therefore, a need for new methods to speed up the structural elucidation of metabolites. In addition to mass spectral library searching and retention index matching, a number of steps can be taken to interpret the

mass spectrum, including accurate mass measurements by high-resolution mass spectrometry, study of isotope ratios, study of the neutral losses and tandem mass spectrometry (MS/MS).[4] When quadrupole GC/MS mass measurements of derivatized metabolites detected in complex mixtures were recalibrated and combined with isotope ratio constraints, it was shown that unknown low molecular weight compounds could be identified despite residual mass errors of up to 10 mm/z units that were verified by use of a magnetic sector instrument.[5] These efforts have been extended by adding chemical and heuristic rules to further constrain the hit list of possible elemental compositions calculated from accurate mass and accurate isotope ratio data.[6] However, derivatized metabolites in GC/MS often yield very low abundance or even absent molecular ions under EI conditions, instead producing highly abundant fragment ions at low m/z values. Such spectra are suitable for identification if the compounds are present in a MS library and if used in conjunction with retention index information. For spectra that are not represented in spectral libraries, first the authentic molecular mass needs to be determined. One way to obtain more abundant molecule-related ions is to operate the mass spectrometer under soft ionization conditions to keep molecules intact. Typical soft ionization methods available in GC/MS are:[7] (i) chemical ionization (CI), where analyte molecules are ionized by

Correspondence to: O. Fiehn, UC Davis Genome Center, 451 Health Sci. Dr., Davis, CA 95616, USA.
E-mail: ofiehn@ucdavis.edu

RCM

chemical ion-molecule reactions; (ii) field ionization (FI), where molecules are ionized by very high electrical fields in the close proximity of emitter needles; and (iii) photoionization (PI), where analyte molecules are ionized by absorption of two or more ultraviolet (UV) or one single vacuum ultraviolet (VUV) photon. On average, absolute ion yields in FI and CI are lower than in EI; however, molecular ions or molecular adduct ions have much higher relative intensity in FI and CI than in EI, rendering these the method of choice for the calculation of elemental formulae.

Another possible way to preserve the molecular ion in EI mode is to lower the ionization energy from 70 eV down to 30 eV or less. However, fragmentation of the molecular ions would still occur because average carbon–carbon bond energies are 348 kJ/mol or 3.6 eV. Moreover, at low ionization energies, the total ion current is several orders of magnitude lower and it cannot be used for practical GC/MS measurements, especially for low abundance compounds in metabolomic mixtures. Amirav et al.[8] developed a further kind of electron ionization called cold-EI that provides enhanced molecular ions. Currently, no commercial implementation of supersonic-molecular-beam GC/MS is available. If better mass resolving power is required, e.g. for complex samples, magnetic sector instruments have long been in use for this purpose.[9]

In addition, software-based approaches were developed for automated molecular weight estimation of EI mass spectra. Probability calculations for the determination of the molecular ion showed good performance[10,11] and were implemented into the freely available NIST MS Search program.[12] These algorithms rely on the presence of molecular ions within low base peak percent thresholds (1–5% bp). Such high-quality spectra require that ions from co-eluting compounds or noise ions are completely removed, for example using the AMDIS mass spectral deconvolution algorithms.[12] Without noise-free mass spectra, the determination of the molecular ion from chromatographically unresolved mixtures is not possible.

Once potential molecular ions have been identified, accurate mass and isotope data can be used to calculate elemental compositions. Calculations should be more successful if highly accurate data are used, which are more likely to be achieved from high-abundance ions than from low-abundance molecular species. Here we present a new approach to generate high-intensity molecular ions under standard 70 eV ionization conditions using a gas chromatography/time-of-flight mass spectrometry (GC/TOF MS) by tuning the instrument in a way to preferentially detect high m/z ions. If the molecular ion is present in a normal EI mass spectrum, it is possible to tune the mass spectrometer in such a way that the abundance of the high mass ions is increased. This enhances the possibility for the identification of molecular formulae of unknown compounds because a high signal intensity yields better mass accuracy (<10 ppm) and accurate isotopic distributions (with a residual mean error of less than 5%). We have used the high-mass-tune EI method in comparison with CI analyses on a mixture of 31 metabolites to test its suitability for retrieving correct elemental compositions from unknown trimethylsilylated metabolites.

EXPERIMENTAL

Two-step derivatization of reference compound mixtures

First, methoximation is performed to inhibit ring formation of reducing sugars by protecting aldehyde and ketone groups. A solution of 40 mg/mL O-methylhydroxylamine hydrochloride, (CAS: [593-56-6]; Formula CH5NO.HCl; Cat. No. 22,690-4 (98%); Sigma-Aldrich, St. Louis, MO, USA) in pyridine (99.99%) was prepared, and 10 μL of this solution was added to dried mixtures of 28 reference compounds (quality control mix, QC) as published in detail.[13] This mixture comprised 1 μg of each compound, namely amino acids (alanine, asparagine, aspartic acid, N-acetylaspartic acid, 4-hydroxyproline, serine, valine, lysine, methionine, glutamic acid), hydroxyl, keto acids and dicarboxylic acids (α-ketoglutaric acid, pyruvic acid, succinic acid, citric acid, shikimic acid), fatty acids (arachidic acid, stearic acid), carbohydrates (glucose, glucose-6-phosphate), aromatic acids (chlorogenic acid, salicylic acid, nicotinic acid, adenosine), aliphatic and aromatic amines (creatinine, putrescine, serotonine) and isoprenoids (α-tocopherol, cholesterol). The methoximation reaction was allowed to proceed for 90 min at 30°C. Subsequently, 90 μL of the trimethylsilylating reagents, N-methyl-N-trifluoroacetamide/1% trimethylchlorosilane (1 mL bottles; Pierce, Rockford, IL, USA), was added and shaken at 37°C for 30 min to exchange acidic protons in order to increase compound volatilities. A mixture of internal retention index (RI) markers was prepared using fatty acid methyl esters (FAME markers) of C8, C9, C10, C12, C14, C16, C18, C20, C22, C24, C26, C28 and C30 linear chain length, dissolved in chloroform at a concentration of 0.8 mg/mL (C8–C16) and 0.4 mg/mL (C18–C30). A volume of 2 μL of this RI mixture was added to the samples. The solution was subsequently transferred to 2 mL glass screw-top amber vials for use in the autosampler.

GC/TOF MS data acquisition

A Waters Micromass GCT Premier (Milford, MA, USA) orthogonal time-of-flight mass spectrometer coupled to a Agilent 6890N gas chromatograph with the Agilent 7683 autosampler (Santa Clara, CA, USA) was used for all the mass spectral experiments. The system was controlled by MassLynx 4.1 software (Waters Micromass). The GCT was equipped with a DB-5MS capillary column (30 m × 0.320 mm i.d. × 0.25 μm; J&W Scientific, Folsom, CA, USA). Ultrahigh purity Airgas BIP 300 helium (99.999%; Airgas Northern California, Sacramento, CA, USA) was used as a carrier gas in constant flow mode (1 mL/min). The GC temperature program was set at an initial temperature 50°C with a hold time of 1 min, with a following temperature ramp of 20°C/min and a final temperature of 325°C and final hold time of 5.25 min for a total run time of 20 min. The injector was set to a temperature of 250°C and was used in split mode with 10:1 ratio. The injection volume was 1 μL. The transfer line temperature was set to 280°C (GC re-entrant temperature). The mass spectrometer was operated in positive EI mode or in positive CI mode at 7000 resolving power (FWHM m/z 658). All EI spectra were collected using an electron energy of

70 eV, trap current of 250 µA, emission current of 600 µA, filament current of 4.5 A, and source temperature of 250°C. All CI data were acquired at 200°C source temperature and an inner source electron energy of 40 eV. Methane was used as the reagent gas at half of the maximum gas flow. The tuning and data acquisition parameters for both EI and CI operation are given below.

The instrument was tuned in two different ways by controlling the outer source assembly:

– A *Normal-mass-tune* file was obtained by generating mass spectra of perfluorotributylamine (PFTBA) with a base peak of m/z69. The voltages on the ion repeller (2.0 arbitrary units (AU)), beam steering (−1.4 AU), focus 1 (59.0 AU), focus 2 (13.4 AU) and focus 3 (46.2 AU) were adjusted to the indicated values. The exact numerical values for the tune voltages depend on the status of the actual ion source and are given here as example of a successful tune file.

– A *High-mass-tune* file was attained by critically evaluating the PFTBA spectra and adjusting the voltages of the ion repeller (2.0 arbitrary unit (AU)), beam steering (−0.8 AU), focus 1 (59.0 AU), focus 2 (13.4 AU) and focus 3 (48.2 AU) so as to maximize, as absolute signal intensity, the two ions at m/z 264 and 502 in the spectrum of PFTBA. In the Waters GCT Premier instrument, the transmission of high mass ions and lower mass ions into the pushout region of the TOF analyzer is reduced by restriction slits in the ion source. The primary beam exiting the inner source is partially mass dispersed by a magnetic field. The field deflects low mass ions more than high mass ions. Consequently, the beam does not travel along the centerline of the instrument and, instead, high masses are detected at a higher relative ratio and at higher absolute intensity than with a non-focused beam (see Supplementary Fig. 1, Supporting Information). Exact numerical values for tune voltages depend on the status of the actual ion source and are given here as example of a successful tune file. Tune parameters were accepted when the mean residual mass values were <10 ppm.

The outer source parameters were tuned at every source switch. Every week the tune conditions were checked with PFTBA. The two different tune files need different calibrations. Therefore, two calibrations were performed using PFTBA, after the tuning of the outer source of the instrument. The PFTBA was injected through the outer source reference reservoir that was set at the temperature of 50°C. The reference re-entrance temperature was 125°C. The ion counts/s were kept below 1500 for the ions of PFTBA during the calibration. Calibration data was acquired with a scan time of 0.9 s and an inter scan delay of 0.1 s in continuum mode. The calibration curve was generated with a maximum of 27 points in the range of m/z 44–615. The curve was fitted with a second-order polynomial such that the standard deviation of the residuals was lower than 0.0008 u.

During GC analysis the acquisition time per spectrum in centroid mode was set to 0.15 s with an interscan delay of 0.03 s (about 5.5 spectra acquisitions per second) over a mass range of 55–700 u. For the duration of the GC run accurate mass measurements were obtained by external calibration and single point lock-mass correction at m/z 284.9943 for EI using 2,4,6-tris(trifluoromethyl)-1,3,5-triazine (TTT) injected from the

outer source as the internal reference and at m/z 286.0027 for CI analysis. Data were acquired in dynamic range enhancement (DRE) mode to allow the accurate mass analysis of very abundant GC/MS peaks that would otherwise saturate the detector and become unusable for accurate mass analysis. In DRE, the instrument acquires alternate high and reduced intensity scans when peaks saturate to provide data with low enough intensity to be still applicable to the internal lock mass correction and tune files.

RESULTS AND DISCUSSION

Tuning the instrument to increase the intensity of ions at higher m/z values

In the process of the chemical identification of unknown compounds it is important to obtain overall high signal intensities for molecular ions (or defined adducts or fragments of molecular ions) and therefore optimal signal-to-noise ratios for each peak. Higher signal intensities yield better ion statistics, thus improving accurate mass and isotopic abundance measurements, and which subsequently lead to higher confidence in determining elemental compositions. Molecular ions of trimethylsilylated metabolites are mostly found in the mass range between 300–800 u (Supplementary Table 2, see Supporting Information). To obtain more abundant molecular ions in this mass region, we developed a method to tune the TOF mass spectrometer in a way that preferentially increases the signal intensity of high mass ions in a given mass spectrum. Conversely, the method discriminated against ion intensities at lower m/z ranges. The instrument was tuned with PFTBA by setting the value of four different voltages of the outer source, beam steering and focus 1, 2 and 3, to maximize the absolute intensity of the two PFTBA fragment ions at m/z 264 and 502. Hence, for the instrument used here it is possible to steer the outer source ion beam to detect preferentially ions in higher or lower mass ranges. The mass spectrum of PFTBA obtained with this high-mass-tune file showed that the relative abundance of high mass ions was much higher than in the PFTBA spectrum in the NIST mass spectral database. Most importantly, the absolute intensities (counts/s) for the higher mass ions were greatly increased under high-mass-tune compared with the normal-mass-tune file for the total mass range. Mass spectra of 2,4,6-tris(trifluoromethyl)-1,3,5-triazine (TTT) (the ion at m/z 284.9943 used as the lock mass) obtained with the high-mass-tune showed the molecular ion (M$^{+\cdot}$ m/z 284.9943) as the base peak (100%) which is more than two-fold its intensity in the TTT mass spectrum in the NIST05 library (37%). The high-mass-tune file was also able to greatly enhance the intensity of the molecular ion in alkanes, fatty acid methyl esters (FAMEs), and compounds containing nitrogen and oxygen and derivatized with trimethylsilyl group (TMS). Samples of a mixture of normal alkanes (C_nH_{2n+2} with n = 12, 14, 16, 18, 20, 24, 28, 32, 36) and fatty acid methyl esters ($C_nH_{2n+1}COOCH_3$ with n = 10, 11, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30) were acquired in EI positive mode under the high-mass-tune file and the normal-mass-tune file. In Table 1 the relative and absolute abundances of the molecular ions for those compounds with the two tune file settings are reported. The ratios of the relative abundance of the molecular ions obtained with the high-mass-tune were 4 to 20 times higher

Table 1. Molecular ion abundances for alkanes and fatty acid methyl esters (FAMEs) increase in the case of the high-mass-tune approach

n-Alkanes	rt[a] (min)	Abundance M[+.] relative to base peak (%)			Absolute abundance M+ (cps)[b]		
		High-mass-tune	Normal-mass-tune	Ratio	High-mass-tune	Normal-mass-tune	Ratio
$C_{36}H_{74}$	18.37	1.0	0.1	10.0	13	—[c]	>[c]
$C_{32}H_{66}$	15.95	6.1	0.3	20.3	88	19	4.6
$C_{28}H_{58}$	14.41	13.2	1.2	11.0	217	93	2.3
$C_{24}H_{50}$	12.87	22.4	1.9	11.8	931	487	1.9
$C_{20}H_{42}$	11.07	37.5	3.7	10.1	1470	900	1.6
$C_{18}H_{38}$	10.03	40.0	4.6	8.7	4620	3090	1.5
$C_{16}H_{34}$	8.90	49.0	6.4	7.7	11800	7790	1.5
$C_{14}H_{30}$	7.63	59.3	8.2	7.2	15100	12000	1.3
$C_{12}H_{26}$	6.21	62.5	11.4	5.5	12500	13100	0.95
FAMEs							
$C_{31}H_{62}O_2$	16.62	100	26.1	3.8	1613	501	3.2
$C_{29}H_{58}O_2$	15.66	100	26.1	3.8	1720	538	3.2
$C_{27}H_{54}O_2$	14.88	100	27.4	3.6	1850	654	2.8
$C_{25}H_{50}O_2$	14.16	100	24.4	4.1	1810	712	2.5
$C_{23}H_{46}O_2$	13.39	100	19.0	5.3	2250	996	2.3
$C_{21}H_{42}O_2$	12.56	100	15.8	6.3	2240	1030	2.2
$C_{19}H_{38}O_2$	11.66	100	11.5	8.7	2070	998	2.1
$C_{17}H_{34}O_2$	10.68	90.7	9.8	9.3	4140	2580	1.6
$C_{15}H_{30}O_2$	9.61	54.0	8.3	6.5	2870	2060	1.4
$C_{13}H_{26}O_2$	8.42	31.8	4.7	6.8	1520	1270	1.2
$C_{11}H_{22}O_2$	7.10	12.2	2.2	5.5	574	623	0.9
$C_{10}H_{20}O_2$	6.37	7.3	1.5	4.9	332	363	0.9

[a] Retention time.

[b] The counts-per-spectrum (cps) refers to the combination of all the mass spectra acquired during the elution of a chromatographic peak, noise-subtracted by an equal number of spectra adjacent to the peak.

[c] The signal was relatively low in intensity and the molecular ion M[+.] in the mass spectrum was not detected.

than those gained with the normal-mass-tune. Figure 1 shows the profile of the two mass spectra for tetracosane ($C_{24}H_{50}$) obtained under the two tune settings; the mass spectrum acquired with the high-mass-tune clearly shows the higher molecular ion signal with 24% of relative abundance compared
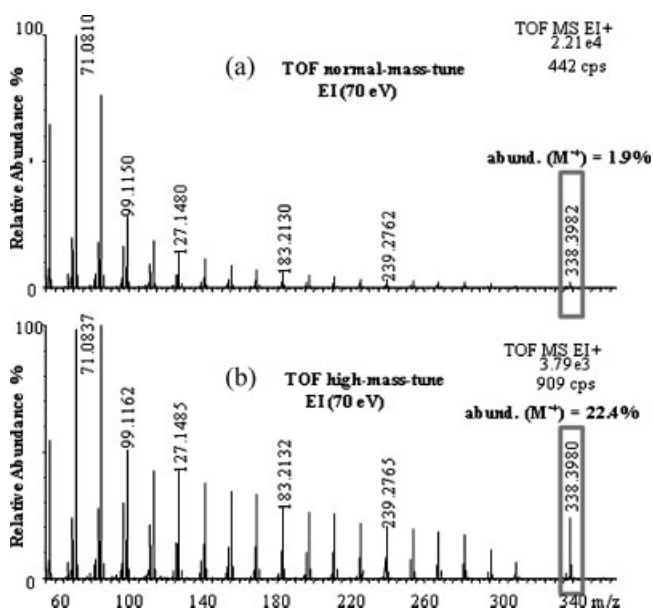

Figure 1. EI mass spectra of tetracosane (MW 338.39123, $C_{24}H_{50}$) obtained under 70 eV electron energy using (a) the normal EI tune and (b) the high-mass-tune file.

with the mass spectrum obtained with the normal tune file (only 2% of molecular ion). The absolute intensities of ions > m/z 200 were 2 to 5 times higher under high-mass-tune than for the same ions obtained with the normal tune file, especially for FAMEs and triemthylsilylated metabolites. For alkanes and FAMEs, the average increase of absolute intensities was two-fold. Compounds with low abundance molecular ion intensities below 1030 cps were significantly more increased ($p < 0.02$) than the molecular ions at above 1030 cps with an average increase of 2.3-fold compared with an average increase of a mere 34% for the higher abundance group. For one compound ($C_{36}H_{74}$), the molecular ion was not even detectable without high-mass-tune. Therefore, the high-mass-tune appears to be specifically beneficial for compounds that have very low abundance molecular ions.

The method to obtain high intensity molecular ions also was utilized with compounds containing oxygen and nitrogen and derivatized (methoximation and silylation). Figure 2 shows the mass spectrum of N,N,N',N'-tetrakis(trimethylsilyl)-1,4-buta-nediamine (putrescine-4TMS) obtained with the high-mass-tune compared with the normal-mass-tune spectrum contained in the FiehnLib GC-TOF mass spectral library.[14] It becomes apparent that even for trimethylsilylated compounds, molecular ions become more intense with a relative intensity of 15% when high-mass-tune is used.

Spectra matching under high-mass-tune parameters

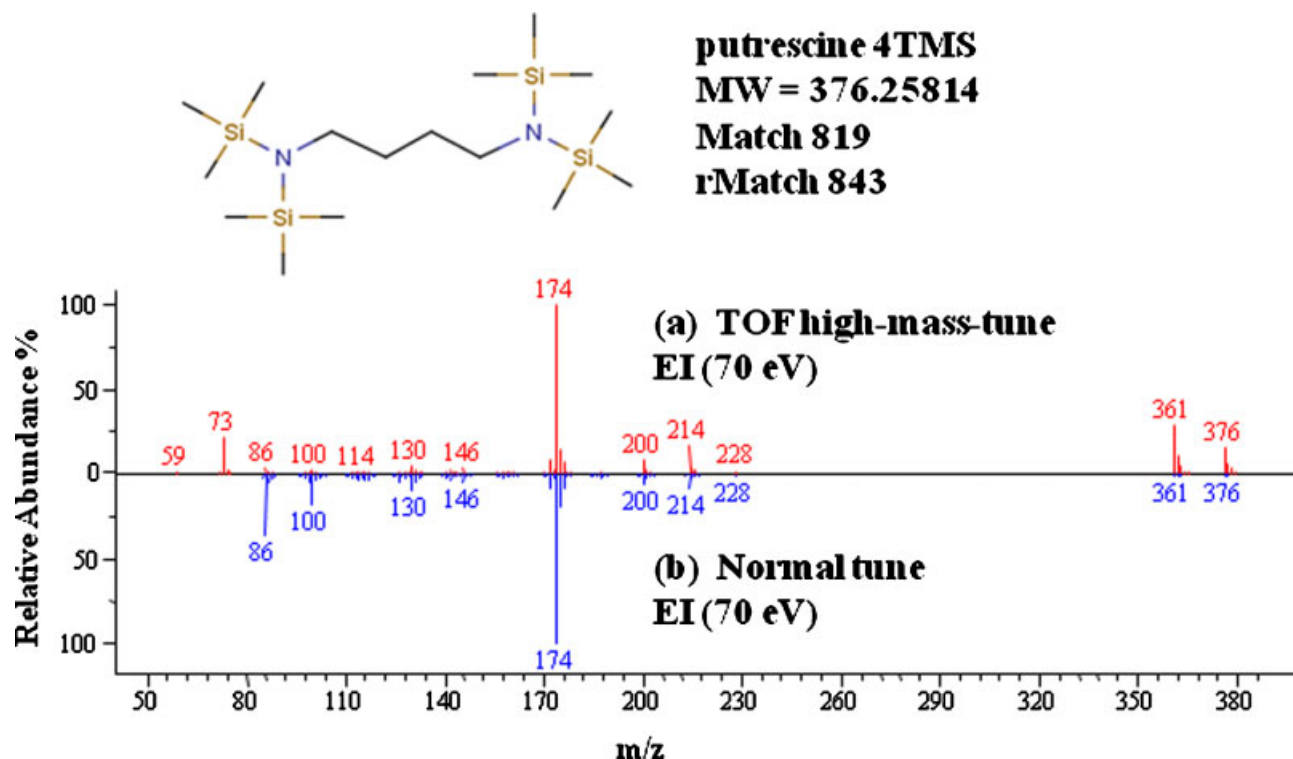As the abundances of several ions under high-mass-tune conditions were higher than under normal-mass-tune, we

Figure 2. EI mass spectra of *N,N,N′,N′*-tetrakis(trimethylsilyl)-1,4-butanediamine (putrescine-4TMS) (MW 376.25814) obtained with 70 eV electron energy using (a) the high-mass-tune file and (b) the spectrum obtained from MS NIST search. The MS NIST search gives the correct result with a match of 819 and a reverse match of 843.

tested how mass spectral similarity matching was affected. For the example of putrescine-4TMS, the NIST MS search resulted in correct identification with a match value of 819/ 1000 and a reverse match of 843/1000 which are above the thresholds commonly used for metabolomic identifications.[15] Table 2 shows the five best hits obtained for the putrescine-4TMS with the NIST MS search operating in hybrid mode on the FiehnLib library[14] and the NIST05[2] library. For high-mass-tune spectra we therefore recommend use of the NIST hybrid mode search which applies both the logic of normal screen with four different prescreen algorithms plus the logic of neutral loss searching to compensate for mass spectrum intensity differences between library spectra and acquired high-mass-tune spectra. Final match factors were calculated without using m/z weighting.

Mass accuracy and isotope accuracy under high-mass-tune parameters

We have tested the effects of the high-mass-tune parameters on mass and isotope accuracy for the abundant molecular ions of FAMEs, and subsequently on low to very low abundance molecular ions and fragment ions of trimethylsilylated metabolites. For determining accurate mass errors,

the mass of the electron ($m_e = 0.00054858026$ u) was considered in the calculations as its omission would introduce errors of 1–5 ppm at mass ranges below m/z 500.[16,17] For FAMEs, very high abundances of relative molecular ion intensities were obtained (Table 1) which resulted in approximately 5 ppm mass errors, 1–2% isotope accuracy errors (Table 3). As previously suggested, isotopic abundances should always be included in calculating elemental formulae.[6] For all FAMEs, correct elemental compositions were calculated using the data acquired under high-mass-tune conditions and applying the constraints given in the Seven Golden Rules.[6]

We have subsequently tested whether derivatized primary metabolites could also be successfully investigated under high-mass-tune parameters. Typically, metabolomics by GC/MS applies methoximation to inhibit anomerization of carbohydrates with subsequent trimethylsilylation as a mild and universal reaction to increase the volatility of otherwise non-volatile hydrophilic compounds. However, the trimethylsilyl-group (TMS) easily fragments under EI conditions, yielding low abundance molecular ions and a range of low m/z fragmentation and rearrangement products, e.g. caused by initial neutral loss of TMS-OH (M–90). Sometimes,

Table 2. Results of mass spectral library search for the putrescine-4TMS peak under high-mass-tune parameters

Hit	Library	Match	RMatch	Name
1	NIST05	819	843	1,4-Butanediamine, *N,N,N′,N′*-tetrakis(trimethylsilyl)- C16H44N2Si4 CAS: 39772-63-9; ID: 111012
2	FiehnLib Leco	813	824	Putrescine_RI 588298
3	NIST05	697	780	Cadaverine, *N,N,N′,N′*-tetrakis(trimethylsilyl) C17H46N2Si4 CAS: 65898-76-2; ID: 110996

Table 3. Mass accuracy and isotopic abundance accuracy are slightly better in high-mass-tune due to increased ion signal. Data represent combined and baseline-subtracted spectra across chromatographic peaks

FAMEs	$M^{+\cdot}$ calc.	Error (ppm)	Error A+1/A (%)[a]	Error A+2/A (%) [a]
$C_{31}H_{62}O_2$	466.4790	8.6	−4.6	−1.4
$C_{29}H_{58}O_2$	438.4449	2.7	−1.4	−1.2
$C_{27}H_{54}O_2$	410.4133	2.2	−2.5	−1.4
$C_{25}H_{50}O_2$	382.3806	−0.5	0.1	−0.6
$C_{23}H_{46}O_2$	354.3534	10.2	−2.3	0.1
$C_{21}H_{42}O_2$	326.3214	8.9	1.3	−1.2
$C_{19}H_{38}O_2$	298.2843	−9.7	−0.7	−0.6
$C_{17}H_{34}O_2$	270.2558	−0.4	−0.8	0.3
$C_{15}H_{30}O_2$	242.2260	5.8	0.4	0
$C_{13}H_{26}O_2$	214.1921	−5.6	−0.5	−0.3
$C_{11}H_{22}O_2$	186.1611	−4.8	1.4	−0.2
$C_{10}H_{20}O_2$	172.1452	−6.4	−0.6	n.d.
avg error[b]		5.5 ± 3.5	1.4 ± 1.3	0.7 ± 0.5

[a] Difference between the experimental and the theoretical value of the relative abundance of isotopic peaks A+1/A and A+2/A.
[b] Arithmetic mean with standard deviation.

molecular ions are not detectable but only methyl losses [M–CH$_3$]$^{+\cdot}$. We have acquired accurate mass and isotope data for [M]$^{+\cdot}$ and [M-CH$_3$]$^{+\cdot}$ ions for a mixture of 28 common metabolites (33 peaks) in triplicate injections, ranging from sugars and sugar phosphates to sterols, aromatics, fatty acids, amino and hydroxyl acids (Supplementary Table S2, see Supporting Information). Despite the high-mass-tune, for almost half of all peaks, the [M]$^{+\cdot}$ ion intensities were too low to be used to determine accurate data, leaving 144 ion species for further investigations. On average, [M–CH$_3$]$^{+\cdot}$ ions were about 3-fold more intense in EI than [M]$^{+\cdot}$ ions; however, the ratios varied greatly depending on the metabolite structure. The results are summarized in Table 4 and Fig. 3 (upper panel). Raw accurate mass data were internally re-calibrated[5] to correct for mass shifts as the acquired masses exceeded the range of the PFTBA calibration masses. In general, accurate masses were determined reliably over all intensity values, with an average of 1.6 m/z units mass error

for the analyzed [M]$^{+\cdot}$ and [M–CH$_3$]$^{+\cdot}$ ions. However, at very low abundance for the monoisotope ions (at <1000 counts per averaged and background-subtracted spectrum, cps), isotope ratios were less reliably determined than at higher abundance (preferably at >11 000 cps).

We have therefore explored the use of chemical ionization (CI) to increase overall ion intensities, improve ion statistics and, specifically, to improve data acquisition for isotope ratios. For many metabolites, either the protonated molecule or the methyl-loss fragment [M–CH$_3$]$^{+\cdot}$ were detected as the base peak. For example, the [M–CH$_3$]$^{+\cdot}$ fragment ion of putrescine-4TMS was detected as its base peak ion (bp), with [M+H]$^+$ at 11% bp and m/z 174 at 3% bp (conversely, m/z 174 is 100% bp under EI conditions, see Fig. 2). Consequently, CI accurate mass data were acquired under dynamic range exclusion according to the instrument software's magnification factors of each mass region in order to avoid detector saturation of the base peaks. Identification of the molecular species is facilitated by detection of additional methane reagent gas adducts (M+29 and M+41) with abundances of 5–50% bp. Under these CI conditions, [M+H]$^{+\cdot}$ and [M–CH$_3$]$^{+\cdot}$ ions were detected for all peaks, yielding 198 trimethylsilylated metabolite ions to be investigated (Table 4). Average ion intensities were 7.5-fold higher than under EI conditions and, consequently, more than half of the detected metabolites had ion intensities of >11 000 cps. Mean mass errors of 1.9 m/z units were observed. After internal re-calibration of all masses to correct for high m/z mass shifts in the same manner as for EI data, the error distribution did not show a significant trend over the m/z range (Fig. 3, lower panel) or the intensity range (Table 4). While we found most peaks to yield better than 3% isotope accuracy error (Fig. 3, lower panel), the number of outliers were higher for mid- and low-intensity monoisotopic peak abundances at less than 11 000 cps. Due to the dynamic range exclusion, even for high intensity levels good isotope accuracies were achieved (Table 4). Overall, the high-mass-tune approach was shown to work for a wide range of compounds including TMS derivatives of metabolites as well as alkanes and FAMEs even under EI. Generally, CI was more convenient than EI for determining the elemental compositions of TMS metabolites,

Table 4. Comparison of chemical and electron ionization. Average ± one standard deviation mass and isotope accuracy data for trimethylsilylated metabolites under high-mass-tune for all detected molecular species ions. Detailed data are given as Supplementary Table S2 (see Suppoting Information)

|	CI	EI	Combined
Total number of molecular species	198 [M+H]$^+$ or [M–CH$_3$]$^+$	144 [M]$^{+\cdot}$ or [M–CH$_3$]$^{+\cdot}$	342 [M+H]$^+$, [M]$^{+\cdot}$ or [M–CH$_3$]$^+$
>11 000 cps	136 ions	24 ions	160 ions
mass error	1.9 ± 1.6 m/z units	1.6 ± 1.3 m/z units	1.8 ± 1.6 m/z units
A+1/A ratio error	1.8 ± 1.4%	1.7 ± 1.2%	1.8 ± 1.4%
A+2/A ratio error	1.1 ± 0.9%	1.5 ± 1.2%	1.2 ± 1.0%
1000 to 11,000 cps	57 ions	85 ions	142 ions
mass error	2.1 ± 2.0 m/z units	1.4 ± 1.4 m/z units	1.7 ± 1.6 m/z units
A+1/A ratio error	3.2 ± 2.8%	2.6 ± 2.0%	2.8 ± 2.4%
A+2/A ratio error	2.6 ± 2.9%	1.9 ± 1.7%	2.2 ± 2.2%
<1000 cps	5 ions	35 ions	40 ions
mass error	1.4 ± 1.2 m/z units	2.5 ± 2.8 m/z units	2.4 ± 2.7 m/z units
A+1/A ratio error	3.9 ± 0.4%	4.7 ± 4.6%	4.6 ± 4.3%
A+2/A ratio error	4.3 ± 4.1%	5.2 ± 4.9%	5.1 ± 4.8%
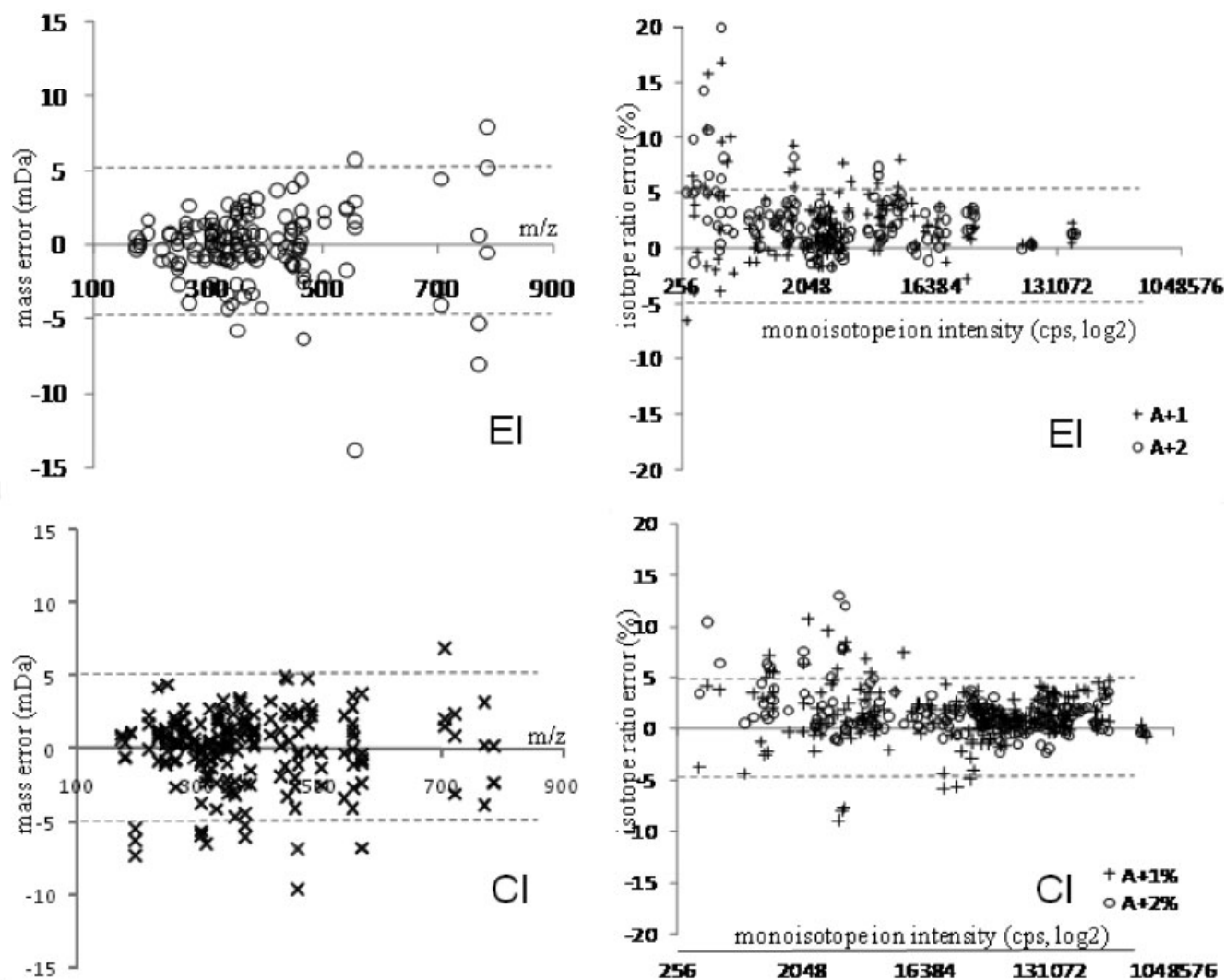
Figure 3. Mass and isotope ratio accuracy for $[M]^{+\cdot}$ and $[M-CH_3]^+$ ions of trimethylsilylated metabolites. Ion intensities are given in log2 scale for visual clarity. Isotope errors are given as crosses for A+1 isotopes and open squares for A+2 isotopes. Thresholds indicate ±5 m*m/z* units for mass errors and ±5% errors for isotope ratio accuracy. Upper panel: Electron ionization data for mass (left) and isotope ratio errors (right), n = 144. Lower panel: Chemical ionization data for mass (left) and isotope ratio errors (right), n = 198.

especially for generating accurate isotope ratio data that are a very important constraint for calculating chemical formulae of unknown compounds.

Retrieving correct elemental compositions using accurate mass, accurate isotope GC/TOF data

The overall average mass accuracy of all combined 344 molecular species, acquired in either EI or CE mode, was 1.8 ± 1.8 m*m/z* units (4.9 ± 4.8 ppm), and overall isotope ratio accuracies were obtained as $2.6 \pm 2.5\%$ for the A+1/A ratio and $2.1 \pm 2.6\%$ for the A+2/A ratio. However, mass accuracy and especially isotope ratio accuracies were worse for the 40 metabolite peaks with very low abundance of less than 1000 cps for the monoisotopic molecular species (Table 4 and Fig. 3). When excluding these ions, the mass accuracy was found to be 1.8 ± 1.6 m*m/z* units with an isotope ratio accuracy of $2.3 \pm 2.0\%$ (A+1/A ratio) and $1.7 \pm 1.8\%$ (A+2/

A), only slightly worse than for the top-160 abundant ions above 11 000 cps (Table 4). How well would such data be suitable to help identifying unknown compounds by retrieving the correct elemental formulae?

Many software versions from mass spectrometry vendors do not apply chemical and heuristic constraints to the elemental formulae that are calculated from accurate mass data. Such omission of chemical rules and discrimination against unlikely formulae leads to long hit lists of potential formula candidates that need to be manually investigated for potential correct (and likely) elemental compositions. Hits are most often scored only by distance of experimental to calculated accurate mass or by using a fit-function, taking the isotope abundance into account or the accurate masses of isotopes. We have previously shown that accurate masses alone are not capable of retrieving correct elemental formulae due to the large search space of chemically possible solutions.[18] We have subsequently devised an algorithm

termed the 'Seven Golden Rules' that constrains and scores all chemically possible formulae by the likelihood that certain ratios of atoms occur, in addition to using the isotope ratios as further important rule,[6] and we have published this algorithm for free downloads.[19] Using these rules, hit lists can be further constrained by actual existence in small molecule databases such as ChemSpider or PubChem that currently comprise more than 27 million compounds. The freely available Seven Golden Rules software[19] uses the largest publicly available database, the Chemical Structure Lookup Service (CSLS),[20] for formula and structure lookups. This free service, provided by the Computer-Aided Drug Design (CADD) Group of the Laboratory of Medicinal Chemistry (LMC), National Cancer Institute (NCI), currently covers 74 million indexed structures from over 100 databases (46 million unique structures).

The number of hits and the likelihood that the correct formula is retrieved as the top-hit or at least among the top-3 hits of the candidate list is directly impacted by the search query criteria: if a very large experimental error is assumed (e.g. 40 ppm mass and 20% isotope accuracy), a very high number of hits will result, with an increased likelihood that some of those hits would be closer to the experimental data than the correct formula. Conversely, when the search query is limited to the average experimental error (here: around 5 ppm mass accuracy and 3% isotope accuracy), many

formulae are excluded as the experimental data are outside the accuracy boundaries.

In order to show these dependencies, we have queried the 342 accurate mass and isotope data by search queries ranging from 5 ppm/3% isotope accuracy to 40 ppm/20% isotope accuracy (Fig. 4). In addition, we have used a further data set for which we have averaged the data for each molecular species using the three independent triplicate analyses, i.e. removing some of the experimental error by averaging data from three different chromatograms (yielding 114 molecular species). As demonstrated in Fig. 4, optimal results are found if search criteria are used that include about one standard deviation around the average technical error, i.e. 10 ppm mass accuracy and 5% isotope accuracy. Using these constraints and searching elemental compositions present in the PubChem database, more than half of all queries yielded the correct elemental composition as the top-hit, with an additional 30% of the queries found in the top-3 hits of the lists of possible elemental formulae. This result was further improved by averaging data from three different chromatograms, yielding up to 87% of the queries as correct hits within the top-3 hits of PubChem formulae. It is important to understand that each of such elemental compositions may point to multiple structural isomers, as given in Supplementary Table S3 (see Supporting Information) for the formulae used here. However, assuming that really novel elemental compositions would need to be considered, i.e. without a database constraint, only around 20% of the queries would find the correct elemental composition as the top-hit with an additional 25% of the correct formulae to be found within the top-3 hits. This result shows that compound identifications could greatly benefit if mass and isotope accuracies were improved, but it also shows that it is important to constrain hit lists by chemical (and biochemical) databases.
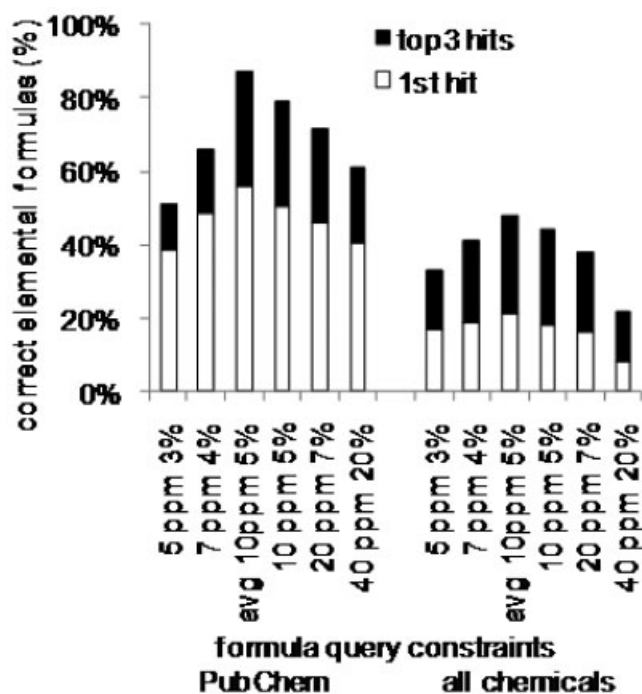
CONCLUSIONS

A high-mass-tune method is presented which allows the determination of molecular ions and calculation of elemental compositions with a higher level of confidence by steering the electron beam of the outer source to preferentially increase relative high m/z ion intensities up to 20-fold, even under the standard and universal EI mode at 70 eV energy. Although the average increase in absolute intensities was only 2-fold under high-mass-tune, this method is particularly important for compounds that have very low absolute intensities under normal tune parameters. This approach facilitates the calculation of the elemental compositions of differentially regulated unknown metabolites in metabolomic screens. We have shown that the correct elemental formula can be regularly found within the top-3 hits of query lists for trimethylsilylated metabolites. Chemical ionization with methane as the reagent gas has proven to be very versatile for TMS metabolites, giving abundant molecular species that could readily be assigned correct molecular formulae when constrained to PubChem query lists.

Figure 4. Percentage retrieved of the correct elemental composition using accurate mass and isotope information for 342 molecular species data using the 'Seven Golden Rules' algorithm and PubChem constraints in comparison with all chemically allowed formulae. Query boundaries are given as mass accuracy (ppm) and isotope ratio accuracy (%). Average data for triplicate measurements were applied to only the best query option ('avg' 10 ppm 5%).

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article.

RCM

Acknowledgements

REFERENCES

1. Halket JM, Waterman D, Przyborowska AM, Patel RKP, Fraser PD, Bramley PM. *J. Exp. Bot.* 2005; **56**: 219.
2. NIST. *NIST Scientific and Technical Databases*, 2008. Available: http://www.nist.gov/srd/nist1.htm.
3. Fiehn O. *Trac-Trends Anal. Chem.* 2008; **27**: 261.
4. McLafferty FW, Turecek F. *Interpretation of Mass Spectra.* University Science Books, 1993.
5. Fiehn O, Kopka J, Trethewey RN, Willmitzer L. *Anal. Chem.* 2000; **72**: 3573.
6. Kind T, Fiehn O. *BMC Bioinformatics* 2007; **8**: 115.
7. Zimmermann R, Welthagen W, Gröger T. *J. Chromatogr. A* 2008; **1184**: 296.
8. Amirav A, Gordin A, Poliak M, Fialkov AB. *J. Mass Spectrom.* 2008; **43**: 141.
9. Bonin MA, Ashley DL, Cardinali FL, McCraw JM, Patterson DG. *J. Am. Soc. Mass Spectrom.* 1992; **3**: 831.
10. Scott DR. *Anal. Chim. Acta* 1994; **285**: 209.
11. Scott DR. *Chemometrics and Intelligent Laboratory Systems* 1995; **27**: 129.
12. Stein SE. *J. Am. Soc. Mass Spectrom.* 1999; **10**: 770.
13. Fiehn O, Wohlgemuth G, Scholz M, Kind T, Lee DY, Lu Y, Moon S, Nikolau B. *Plant J.* 2008; **53**: 691.
14. Kind T, Wohlgemuth G, Lee DY, Lu Y, Palazoglu M, Shahbaz S, Fiehn O. *Anal. Chem.* 2009; **81**: 10038.
15. Wagner C, Sefkow M, Kopka J. *Phytochemistry* 2003; **62**: 887.
16. Scheppele SE, Grindstaff QG, Grigsby RD, McDonald SR, Hwang CS. *Int. J. Mass Spectrom. Ion Phys.* 1983; **49**: 179.
17. Ferrer I, Thurman EM. *Rapid Commun. Mass Spectrom.* 2007; **21**: 2538.
18. Kind T, Fiehn O. *BMC Bioinformatics* 2006; **7**: 234.
19. Kind T, Fiehn O. *Seven Golden Rules* 2007. Available: http://fiehnlab.ucdavis.edu/projects/Seven_Golden_Rules/.
20. Nicklaus MC, Sitzmann M, Filippov IV, Ihlenfeldt W-D. *Chemical Structure Lookup Service 2008.* Available: http://cactus.nci.nih.gov/cgi-bin/lookup/search.