Anal. Chem. **2000,** *72,* 3573−3580

Identification of Uncommon Plant Metabolites Based on Calculation of Elemental Compositions Using Gas Chromatography and Quadrupole Mass Spectrometry

Oliver Fiehn,* Joachim Kopka,† Richard N. Trethewey,† and Lothar Willmitzer

Max-Planck-Institute of Molecular Plant Physiology, 14424 Potsdam, Germany

Unknown compounds in polar fractions of *Arabidopsis thaliana* crude leaf extracts were identified on the basis of calculations of elemental compositions obtained from gas chromatography/low-resolution quadrupole mass spectrometric data. Plant metabolites were methoximated and silylated prior to analysis. All known peaks were used as internal references to construct polynomial recalibration curves of from raw mass spectrometric data. Mass accuracies of 0.005 ± 0.003 amu and isotope ratio errors of 0.5 ± 0.3% ($A + 1/A$), respectively, 0.3 ± 0.2% ($A + 2/A$), could be achieved. Both masses and isotope ratios were combined when the elemental compositions of unknown peaks were calculated. After calculation, compound identities were elucidated by searching metabolic databases, interpreting spectra, and, finally, by comparison with reference compounds. Sum formulas of more than 70 peaks were determined throughout single GC/MS chromatograms. Exact masses were confirmed by high-resolution mass spectrometric data. More than 15 uncommon plant metabolites were identified, some of which are novel in Arabidopsis, such as tartronate semialdehyde, citramalic acid, allothreonine, or glycolic amide.

Physiological changes during plant growth, senescence, and reactions to environmental stress or to changes due to plant transformation are usually investigated by analyzing only a limited number of target metabolites. For example, several thousand *Arabidopsis thaliana* mutants have been screened by thin-layer chromatography to search for a block in ferulate to 5-hydroxyferulate conversion.[1] However, unexpected results are often difficult to interpret if only selected metabolites are analyzed. A multitarget profiling analysis might therefore be a more valuable approach with which to study complex biological systems.[2] Such an analysis can be nicely done using the hyphenation of gas chromatography to mass spectrometry (GC/MS). GC/MS is a mature and highly reproducible technique for the analysis of thermostable compounds. It combines high chromatographic separation power with a universal detector to produce excellent sensitivity and selectivity. However, GC/MS-based metabolic screening approaches are rarely used in biological sciences. The use of GC/MS for multitarget profiling is mostly limited to human disease detection in clinical medicine.[3] The National Institute of Standards and Technology has recently released software for the automatic deconvolution of mass spectra of coeluting peaks. This software has been shown to be quite useful for the rapid identification of human metabolic disorders by pattern recognition using 68 trimethylsilylated compounds.[4] For plant analysis, however, much less has been done. In apricot fruit matrix, low standard deviations have been achieved using a simultaneous quantitation method of 14 common sugars, amino acids, and hydroxy acids.[5] In addition, optimal reaction conditions have been worked out for trimethylsilylation of 12 sugars and organic acids in plant leaf extracts.[6] However, to gain deeper insight into biological systems, analyses should not be restricted to a handful of compounds. More than 10 000 metabolites have been described in plants. In accordance with this complexity, several hundred peaks can be resolved in a GC/MS chromatogram of a single plant extract. However, the majority of these peaks cannot be identified using commercial libraries or references, especially since many metabolites are not commercially available. As Arabidopsis serves as a model organism in plant genomic research, each novel identified compound has the potential to lead to new biological information. We report here on the identification of unknown compounds in complex GC/MS chromatograms using an integrated approach of chemical derivatization, calculation of elemental compositions, database search, and proof of identity by reference compounds. This practical approach can be done using of widely available, inexpensive quadrupole mass spectrometers, even though the resolution and mass accuracy achieved with these spectrometers is worse compared to high-resolution mass spectrometers.[7,8]

* Corresponding author: (e-mail) fiehn@mpimp-golm.mpg.de; (telephone) ++49-331-5678-216; (fax) ++49-331-5678-250.

† Present address: Metanomics GmbH & Co KGaA, Tegeler Weg 33, 10589 Berlin, Germany.

(1) Chapple, C. C. S.; Voigt, T.; Ellis, B. E.; Somerville, C. R. *Plant Cell* **1992**, *4*, 1413−1424.

(2) Trethewey, R. N.; Krotzky, A. J.; Willmitzer, L. *Curr. Opin. Plant Biol.* **1999**, *2*, 83−85.

(3) Duez, P.; Kumps, A.; Mardens, Y. *Clin. Chem.* **1996**, *42*, 1609−1615.

(4) Halket, J. M.; Przyborowska, A.; Stein, S. E.; Mallard, W. G.; Down, S.; Chalmers, R. A. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 279−284.

(5) Katona, Z. F.; Sass, P.; Molnár-Perl, I. *J. Chromatogr., A* **1999**, *847*, 91−102.

(6) Adams, A. A.; Chen, Z. L.; Landman, P.; Colmer, T. D. *Anal. Biochem.* **1999**, *266*, 77−84.

10.1021/ac991142i CCC: $19.00
Published on Web 07/07/2000

MATERIALS AND METHODS

A total of 300 mg FW of *A. thaliana* C24 leaves were harvested, transferred into a 2-mL Eppendorf tube, immediately frozen in liquid nitrogen, and crushed in a Retsch -∼mill (Haan, Germany). A 1.4-mL aliquot of methanol and 50 μL of water were added, and the pH was checked to be between 5 and 6. Extraction was carried out at 70 °C for 15 min. The tube was centrifuged for 3 min at 14000*g*. The supernatant was decanted into a screw-top glass tube, and 1.4 mL of water and 0.75 mL of chloroform were added. The mixture was vortexed, and the tubes were centrifuged for 10 min at 4000 rpm. The methanol/water phase was dried in a SpeedVac concentrator overnight. For this study, the chloroform/methanol phase containing lipophilic compounds was discarded. Carbonyl moieties were protected either by methoximation or by ethoximation, using 50 μL of a 20 mg/mL solution of the corresponding alkoxyamine hydrochloride in pyridine at 30 °C for 90 min. Afterward, acidic protons were derivatized with 50 μL of *N*-methyl-*N*-*tert*-butyldimethylsilyltrifluoroacetamide (MTBSTFA) at 70 °C or *N*-methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA) at 37 °C, respectively, for 30 min.

One-microliter aliquots of these solutions were injected at a split ratio of 1:25 into a GC/MS system consisting of an AS 2000 autosampler, a GC 8000 gas chromatograph, and a Voyager quadrupole mass spectrometer including a dynode/phosphor/photomultiplier detector (all ThermoQuest, Manchester, U.K.), manufactured in 1998. Tuning was done according to the instrument manual using tris(perfluorobutyl)amine (CF43) as reference gas. Special attention was paid to the high mass resolution which was manually improved to gain resolution up to $R_{fwhm} = 2800$ at m/z 614. Installation requirements of ThermoQuest Voyager GC/MS instruments are at least $R_{fwhm} = 1500$. Mass spectra were recorded from m/z 50 to 600 at 0.5 s scan^{-1} for trimethylsilylated samples (TMS) and from m/z 50 to 800 at 0.7 s scan^{-1} for *tert*-butyldimethylsilylated samples (TBS). Accurate mass measurements were made using a Finnigan MAT magnetic sector field instrument (Finnigan, Bremen, Germany). Chromatography was performed using a 30 m × 250 μm SPB 50 column (Supelco, Bellefonte, PA) or a 30 m × 250 μm DB 5-MS column (J&W Scientific, Folsom, CA). Injection temperature was 230 °C, the interface was set to 250 °C, and the ion source was adjusted to 200 °C. Helium flow was 1 mL min^{-1}. After a 5-min solvent delay time at 70 °C, the oven temperature was increased at 5 °C min^{-1} to 310 °C, 1 min isocratic, cool-down to 70 °C, followed by an additional 5-min delay. Recording of mass spectra of the next sample would be hampered due to column bleeding, without preceding delay.

Ion trace integration was done using the MassLab FindTarget method for the characteristic M-57 fragments of all assigned peaks throughout the chromatograms of TBS samples. Raw masses of these fragments were determined by averaging all detected masses over the peaks widths for trace, minor, and major peaks, regularly including 12−15 scans. Raw masses of overloaded peaks with an ion trace intensity exceeding 1×10^7 units were only determined at the left or the right peak flank, including at least seven scans. Plant samples were derivatized and then used to confirm peak identities by comparison to reference compounds and use of the NIST, Wiley, and internally compiled spectra libraries. Retention time correction was done by internal reference compounds in order to minimize run-to-run errors. All chemicals were purchased at Sigma-Aldrich-Fluka (SAF, Deisenhofen, Germany). Stable isotope reference metabolites ([$^{13}C_{12}$]sucrose, [$^{13}C_6$]-glucose, glycerol-*d*8, ethanolamine-d_4, ethylene-d_6 glycol, aspartate-d_3, [$^{13}C_5$]glutamate, alanine-d_4, valine-d_8, leucine-*d*3 and benzoic-d_5 acid) were obtained from Campro Scientific (Emmerich, Germany).

RESULTS AND DISCUSSION

GC/MS with Trimethylsilylation. Plant metabolites such as sugars, amino acids, and hydroxy acids include many different chemical moieties, often present in the same molecule. As most of these compounds are not volatile, they have to be derivatized before GC analysis. There are limited possibilities for the choice of versatile reagents offering fast and complete derivatizations of nearly all the acidic protons present in an analyte. Most of them are silylating reagents. In preceding experiments, we found the trimethylsilylating reagent MSTFA to be superior to alternative chemicals such as *N*,*O*-bis(trimethylsilyl)-*N*-trifluoroacetamide (BSTFA) concerning the completeness of NH derivatization of amines and amino acids. Unwanted side reactions are also limited using MSTFA. Further, it is known that direct derivatization of reducing sugars such as fructose and glucose leads to a number of different peaks related to cyclic and open-chain structures of these hexoses that cannot be completely controlled by altering reaction conditions.[9] Cyclization of reducing sugars can be inhibited by oximation using hydroxylamine or alkoxyamine hydrochlorides.[10] After silylation, reducing sugars regularly result in a pair of chromatographically resolvable peaks due to their limited rotation along the C═N bond. For identification purposes, we found the use of alkoxyamines superior to hydroxylamines. When ethoximated samples are compared to methoximated plant extracts, the presence of carbonyl moieties in unknown peaks is clearly indicated by a chromatographic shift to longer retention times. Furthermore, an oximation step is valuable for metabolite profiling analyses because α-keto acids such as pyruvic acid are protected against decarboxylation.[11] For routine analyses, we found the stability of the stationary phase to be as important as chromatographic resolution. Use of a 5% phenyl/95% methyl-modified polysiloxane phase (DB5) gave narrower peaks and better resolution for critical peak groups compared to a 50/50 phase (SPB50), hence improving both the sensitivity and the selectivity of the analyses of leaf extracts. However, after 150 runs, column bleeding exceeded tolerable limits for the DB5 phase whereas the SPB50 column could be used for over 1000 runs.

Using this derivatization scheme, GC/MS analyses of the polar fraction of Arabidopsis wild-type leaf extracts show a complex pattern of major, minor, and trace peaks. In total, more than 300 peaks could be detected. Using commercially available EI mass spectra libraries covering about 250 000 compounds, some 40 plant metabolites such as sucrose, citrate, malate, or common amino acids could be directly identified. An additional 30 compounds were identified by comparison of standard spectra libraries

(7) Hopfgartner, G.; Chernushevich, I. V.; Covey, T.; Plomley, J. B.; Bonner, R. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1305−1314.
(8) Huang, N.; Siegel, M. M.; Kruppa, G. H.; Laukien, F. H. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1166−1173.
(9) Asres, D. D.; Perrault, H. *Can. J. Chem.* **1997**, *75*, 1385−1392.
(10) Schweer, H. *J. Chromatogr.* **1982**, *236*, 355−360.
(11) Tam, Y. Y.; Normanly, J. *J. Chromatogr., A* **1998**, *800*, 101−108.

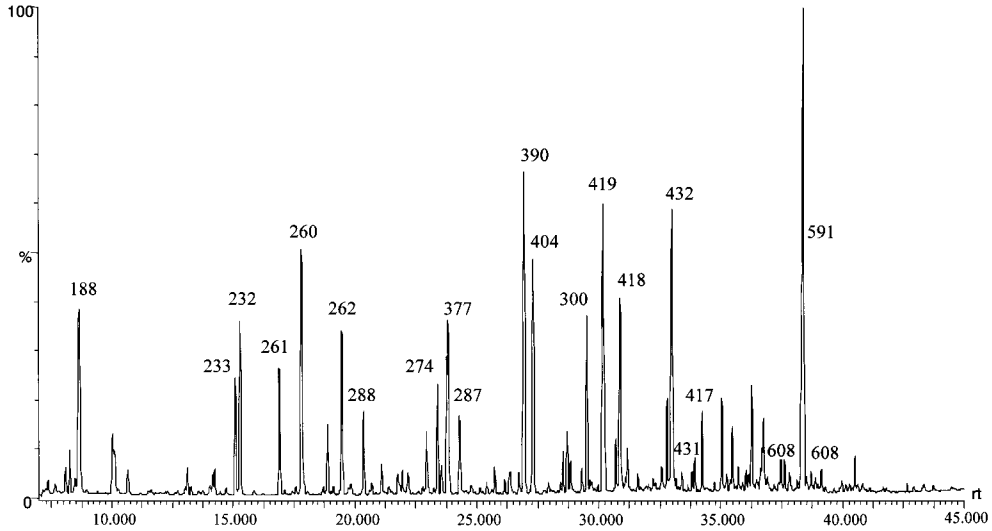

Figure 1. GC/MS chromatogram (TIC) of TBS derivatized Arabidopsis leaf extracts. Major peaks, M − 57 fragments: 188 ammonia, 233 carbonic acid, 232 ethanolamine, 261 lactate, 260 alanine, 262 carbonic acid (methoxim.), 288 valine, 274 γ-aminobutyrate, 377 glycerol, 287 fumarate, 390 serine, 404 threonine, 300 oxoproline, 419 malate, 418 aspartate, 432 glutamate, 431 lysine, 417 asparagine, 608 fructose, 591 citrate, and 608 glucose.

established in our group and the corresponding retention times. A major portion of the detected peaks, however, remains unknown. Many TMS spectra lack intense fragments in the high-m/z region, thereby hampering structural elucidation of unknown peaks. This is particularly true for physiologically important carbohydrates and sugar alcohols where molecular ions, M 15, or other known fragments are present at very low abundances[12] that undergo rapid further fragmentation. Even though it is difficult to follow TMS fragmentation pathways directly, unknown carbohydrates can still be classified by the relative abundances of their specific breakdown products.[13]

M − 57 Fragment Assignment of TBS Derivates. Since the 1990s, MTBSTFA is gaining increasing popularity as a silylating reagent for structural elucidation purposes[14] as well as isotope determination.[15] Using MTBSTFA for derivatizing Arabidopsis leaf extracts, acidic protons are replaced by TBS groups leading to complex chromatograms with more than 300 peaks (Figure 1). The stability of the *tert*-butyl radical (M = 57 amu) favors the cleavage of this moiety instead of the fragmentation of the total TBS group, hence giving rise to an intense M − 57 signal. In the first step of the identification procedure, this M − 57 assignment has to be done for both known and unknown peaks, assuming the most abundant signal at high m/z values to be the M − 57 ion. Occasionally, this assumption would be misleading. M − 57 assignment, thus, must be assured by the presence of the less intensive M − 15 ion, or, in a few cases, by the detection of the molecular ion. Only very few peaks exhibited M − 57 abundances below 5% of the base peak intensity. All of them were polyols such as glycerol, erythritol, or hexoses, the latter giving several peaks due to incomplete derivatization. But even for those peaks, M − 57 fragments could be unambiguously assigned. Aromatic compounds gave nearly exclusively M − 57 fragments of base peak intensity, a result that is concordant with the recent report of another research group.[16] Mass spectra of organic phosphates can easily be characterized by formation of abundant persilylated phosphate ions (m/z 383 for TBS derivates, m/z 299 for TMS derivates).

Quadrupole MS Mass Correction. To calculate the elemental compositions of these M − 57 ions, either their masses or their isotopic peak patterns have to be determined very precisely. Unfortunately, quadrupole mass spectrometers offer neither accurate mass nor exact isotopic measurements. Although our external mass calibrations using CF43 regularly resulted in calibration curves with standard deviations of about 0.020−0.030 amu, the actual mass errors were regularly much higher. For the example of citramalic acid, it can be seen that the scan-to-scan deviation (0.010 amu) was much lower than the total mass error, which was calculated to 0.185 amu (Figure 2). The standard error of the mean was as small as 0.003 amu, since it decreases with the square root of the number of scans. Similar results were obtained for other peaks. Therefore, we looked for ways by which to decrease the total mass errors. We found that the extent of total mass errors does not only depend on the actual instrument status, the tuning parameters, and the scanning speed but also on the m/z value under investigation. Furthermore, we found mass error shifts from chromatogram to chromatogram. It can be concluded that exact mass measurements cannot be done with quadrupole instruments and external mass calibration. Recently, however, accurate mass measurements have been reported using electrospray ionization-quadrupole mass spectrometry by adding internal references.[17,18] Neither hardware nor software of routine GC/quadrupole MS instruments offers an easy way of adding internal reference compounds for scan-to-scan mass calibration. We tried to use the observed masses of m/z 73 (TMS) and 147

(12) DeJongh, D. C.; Radford, T.; Hribar, J. D.; Hanessian, S.; Bieber, M.; Dawson, G.; Sweeley, C. C. *J. Am. Chem. Soc.* **1969**, *91*, 1728−140.
(13) Bleton, J.; Mejanelle, P.; Sansoulet, J.; Goursaud, S.; Tchapla, A. *J. Chromatogr., A* **1996**, *720*, 27−49.
(14) Baker, A.; Dodd, C. D.; Parsons, R. *Plant Cell Environ.* **1997**, *19*, 1249−1260.
(15) Godber, I. M.; Parsons, R. *Plant Cell Environ.* **1998**, *21*, 1089−1099.
(16) Heberer, T.; Stan, H. J. *Anal. Chim. Acta* **1997**, *341*, 21−34.
(17) Kostiainen, R.; Tuominen, J.; Luukkanen, L.; Taskinen, J.; Green, B. N. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 283−285.
(18) Tyler, A. N.; Clayton, E.; Green, B. N. *Anal. Chem.* **1996**, *68*, 3561−3569.

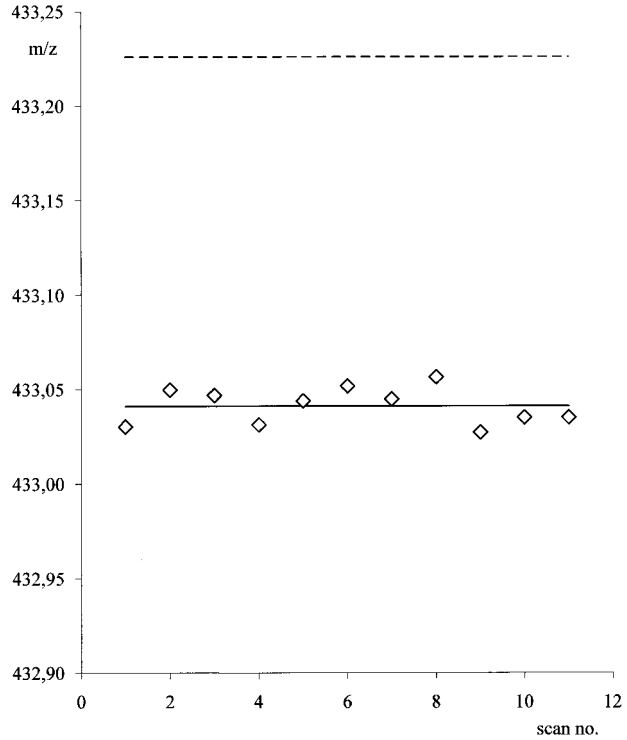

Figure 2. M − 57 raw mass accuracy over 11 scans of an unknown peak in an Arabidopsis GC/MS chromatogram. This metabolite was later identified as tris(-*tert*-butyldimethylsilyl)citramalic acid. Dotted line, calculated mass. Solid line, average of determined masses.

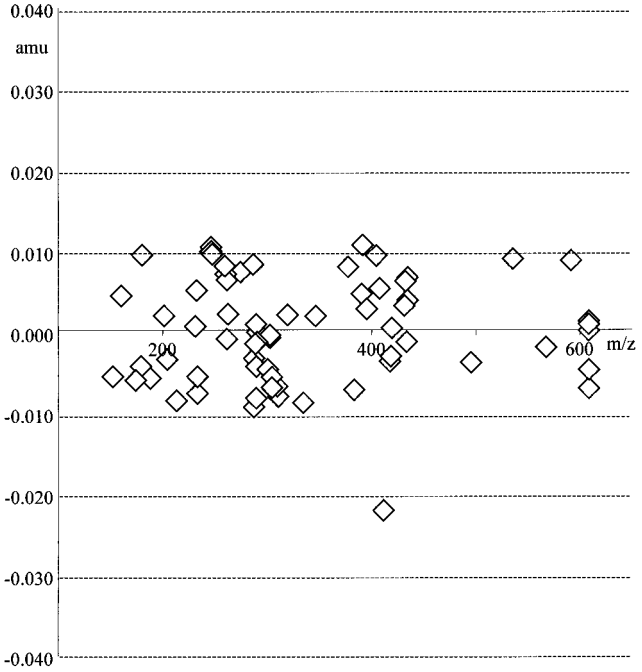

Figure 3. M − 57 mass accuracy of all known Arabidopsis metabolites detected in a single GC/MS chromatogram after raw mass correction using a cubic recalibration curve.

(TMS-O-DMS) as internal references. These ions are formed by rearrangement processes and appear in nearly every scan in complex plant chromatograms after derivatization with MTBSTFA. However, resulting mass accuracy was not satisfying. In a more successful approach, we used all known M − 57 fragments throughout a chromatogram for internal mass calibration. Linear and polynomial regression curves were constructed by plotting the determined versus the exact masses, followed by a calculation of the residual mass errors. Generally, the lowest average mass error could be found using third-order regression curves with no further improvements using higher order calibration curves. This recalibration of quadrupole mass accuracy resulted in an average error of only 0.005 ± 0.003 amu with nearly no peak exceeding an error of 0.010 amu (Figure 3). It can be seen that mass accuracy is independent from m/z values, and thus, errors may be given in amu instead of ppm units. The only outlier was the determination of a cystine TBS trace peak that was not completely chromatographically resolved from a major alanine TBS peak. The same cystine mass error was also found using a magnetic sector field instrument, hence indicating the fact that mass measurements may also be dependent on signal-to-noise ratios. Due to quadrupole mass resolution limits (R_{fwhm} > 1500 at m/z 614), exact masses could not be determined when M − 57 fragments were less than 10-fold abundant compared to noise ions or when fragments of coeluting compounds had the same nominal mass. However, this was only the case in less than 5% of the peaks that were examined.

Quadrupole MS Isotope Determination. Even when M − 57 fragments were used for internal mass calibration, however, the derived mass accuracy was still not good enough to directly calculate elemental compositions of unknown peaks. The number of possible structures had to be further minimized by calculating elemental compositions obtained from isotopic peak patterns. Theoretical isotopic patterns for a given formula can be found on the Internet[19] or may be calculated by exact equations[20] or by approximations. After plotting the calculated isotopic ratios versus the determined ratios for known analytes, several conclusions could be drawn (Figure 4). First, the diisotopic ratios ($A + 2/A$) form clusters, thus indicating the dominance of the TBS $^{30}Si/^{28}Si$ ratio over the diisotopic ratio of the analyte skeletons. Therefore, the number of TBS groups can easily be determined for each peak by integration of its $A + 2/A$ area ratios for the M − 57 ions. Second, regression curves must be constructed to minimize the errors of isotopic ratio determination. Monoisotopic ratios usually had to be corrected by linear calibration curves whereas the average diisotopic error could not be minimized with regression curves. Average isotope ratio errors of 0.5 ± 0.3% ($A + 1/A$), respectively, 0.3 ± 0.2% ($A + 2/A$) could be achieved. However, several outliers can be seen in the isotopic plots and a somewhat greater error has to be taken into account for the calculation of unknown peaks. Generally, isotopic ratios could be estimated most reliably when M − 57 fragments were at least 10-fold higher than noise peaks or fragments of coeluting peaks. Clearly, possible elemental compositions can only be estimated with a high degree of uncertainty by determining isotopic ratios alone. Errors were on the order of one to two N atoms and about one O atom. Therefore, for studies of stable isotope distribution in plant science, combustion/isotope ratio mass spectrometers are used instead of quadrupole instruments.[21,22]

Calculation of Elemental Compositions and Database Searches. As examples, the complete elemental composition

(19) http://www.sis.com.

(20) McLafferty, F. W.; Turecek, F. *Interpretation of Mass Spectra*, 4th ed.; University Science Books: Mill Valley, CA, 1993; Chapter 2.

Table 1. Iterative Computing of Elemental Compositions Obtained from Quadrupole Mass Spectrometric Data[a]

starting information	derived information		elemental composition						error			result	
	TBS	MW	C	H	N	O	S	P	$A+1/A$	$A+2/A$	mass	comments	database entries
example 1	2	139	0	0	0	0			−7.6	−2.3		molecule has 6−7 C	
M − 57 = 310.137 amu			5	5	3	5			−0.8	−0.4	0.004	alternative sum formula	3
$A+1/A$ ratio = 26.5%			*6*	*5*	*1*	*3*			−0.4	−0.1	−0.008	*6-hydroxynicotinic acid*	*12*
$A+2/A$ ratio = 10.5%			5	5	3	0	1		−0.1	3.6	−0.014	$A+2$ too high	
no methoxime formation			7	9	1	2			0.7	0.0	0.029	mass too high	
example 2	1	144	0	0	0	0			−9.6	−0.9		molecule has 8−9 C	
M − 57 = 201.129 amu			*8*	*16*	*0*	*2*			*−0.4*	*0.5*	*0.002*	*2-ethylhexanoic acid*	*63*
$A+1/A$ ratio = 16.9%			5	12	4	1			−2.4	0.0	−0.012	$A+1$ too low	
$A+2/A$ ratio = 4.4%			7	16	2	1			−0.8	0.2	0.013	alternative sum formula	9
no methoxime formation			8	16	0	0	1		0.3	4.5	−0.015	$A+2$ too high	
example 3	3	148	0	0	0	0			−5.2	−2.6		molecule has 4−5 C	
M − 57 = 433.226 amu			*5*	*8*	*0*	*5*			*0.6*	*0.3*	*0.000*	citramalic acid	13
$A+1/A$ ratio = 36.3%			3	8	4	1	1		0.5	3.6	0.004	$A+2$ too high	
$A+2/A$ ratio = 17.0%			4	8	2	4			0.2	−0.1	0.011	alternative sum formula	6
no methoxime formation	2	262	13	18	4	0	1		−0.2	−0.1	0.001	sulfur replacing TBS	0

[a] Three peaks were chosen as examples, later identified by comparison with external references as uncommon metabolites (in italic).

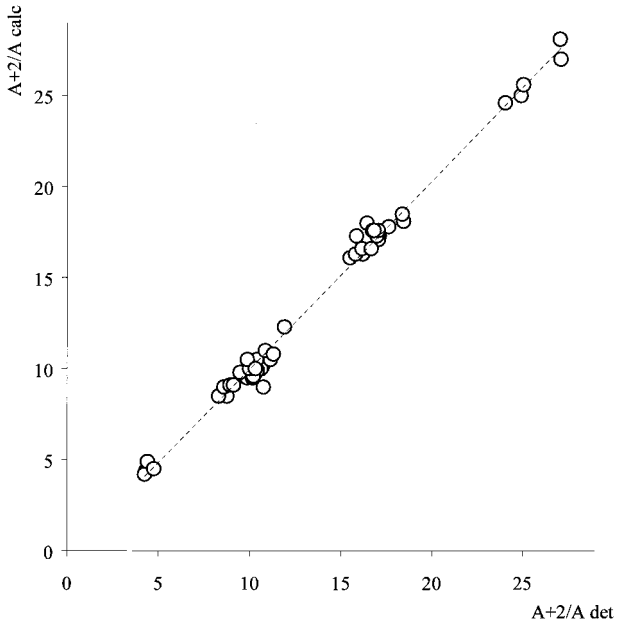

Figure 4. Calculated versus determined $A+2/A$ ratios of the M − 57 fragments of known metabolites, detected in the polar fraction of an Arabidopsis leaf extract.

calculation procedure is given that leads to the identification of three uncommon plant metabolites (Table 1). First, the exact masses for the M − 57 peaks were calculated from the detected raw masses using the cubic recalibration curve established for this chromatogram, as described above. Second, the isotopic ratios were determined and the $A+1/A$ ratios were corrected using the corresponding recalibration curve. Third, the methoximated sample was compared to an ethoximated chromatogram in order to determine whether the unknown compound included carbonyl moieties. Next, the number of TBS groups was derived from the $A+2/A$ ratio and the molecular mass of each underivatized molecule was calculated. Using this information, the most likely elemental composition can be computed iteratively for each compound. The subtraction of the $A+1/A$ ratio calculated for the TBS groups from the monoisotopic ratio for the total M − 57 fragment gives the monoisotopic ratio of the underivatized molecule. For each example, the number of carbon atoms can be approximately determined. The number of potential sum formulas is further limited by several facts. Due to the nitrogen rule, even M − 57 fragments can only be explained by an uneven number of nitrogen atoms, and uneven M − 57 ions account for molecules including an even number of nitrogen atoms. Furthermore, each TBS group has to be bound to one N, O, or S atom, the presence of P in biochemical samples requires at least three O atoms, and the total number of H atoms must not exceed 2 times the number of C atoms plus two plus the number of N atoms ($H \leq C_{2n+2} + N$). Thus, many theoretical elemental compositions are chemically impossible or biochemically highly unlikely. However, these examples also show the limits of our approach. The mass differences between the correct and the alternative elemental compositions were in each case ~0.010 amu. As given above, the same number was the limit of our quadrupole mass accuracy. Calculations could have been limited to only one possible elemental composition, if the maximum mass error was <0.005 amu. For complex samples, this number is still only achievable by high-resolution mass spectrometers. The examples further demonstrate that the absence of sulfur is easily indicated as large $A+2/A$ ratio errors occur when S is included. Under the assumption that the number of TBS groups was misinterpreted by the $^{34}S/^{32}S$ ratio of 4.4%, a high number of carbon atoms were required to match the $A+1/A$ ratio. In no case were database entries found for these elemental compositions. Therefore, one to two possible elemental compositions can be derived from the quadrupole mass spectrometric data of each unknown peak. The most likely formula is the one with a mass error below 0.010 amu. The next step in the identification process is to enter the most likely sum formula into chemical and metabolic database search forms. For $C_6H_5NO_3$, for example, 12 isomers can be found by search in chemical databases.[23,24] Possible molecular structures

(21) Gleixner, G.; Scrimgeour, C.; Schmidt, H.-L.; Viola, R. *Planta* **1998**, *207*, 241−245.
(22) Liang, Y.; White, W. S.; Yao, L.; Serfass, R. E. *J. Chromatogr., A* **1998**, *800*, 51−58.

Table 2. Tentatively Identified *A. Thaliana* Metabolites by GC/MS Accurate Mass Measurements, Derivatization Characteristics, Calculation of Elemental Compositions, Chemical and Metabolic Database Search, and Comparison with Reference Compounds

name	elemental compn	M − 57 mass		derivatization		comments	pathways
		quadrupole	sector field	meox	TBS		
glycolic amide	$C_2H_5NO_2$	246.124	246.134	0	2	RT, MS nearly identical to *N*-hydroxyacetamide	
erythronolactone isomer	$C_4H_6O_4$	289.128	289.128	0	2	RT, MS nearly identical to (−)-D-erythronolactone	acsorbate/aldarate metabolism
tartronate semialdehyde	$C_3H_4O_4$	304.146	304.139	1	2	RT, MS similar to *β*-hydroxypyruvate	glyoxylate/dicarboxylate e, ascorbate/aldarate metabolism
5-hydroxyferulic acid	$C_{10}H_{10}O_5$	495.245	495.245	0	3	RT, MS similar to sinapic acid	flavanoid, stilbene, and ignin biosynthesis
2-ketothreonic acid	$C_4H_6O_5$	334.159	334.153	1	2		acsorbate/aldarate metabolism
methyleneglutarate *or* methylitaconate	$C_6H_8O_4$	315.138	315.143	0	2		C5 branched dibasic acid metabolism
>30 possible database entries	$C_6H_{10}O_4$	317.150	317.155	0	2	RT, MS similar to $C_6H_8O_4$	
2-dehydro-3-deoxy-rhamnoate *or* -fuconate	$C_6H_{10}O_5$	447.239	447.244	0	>3	incomplete derivatization	fructose/mannose metabolism

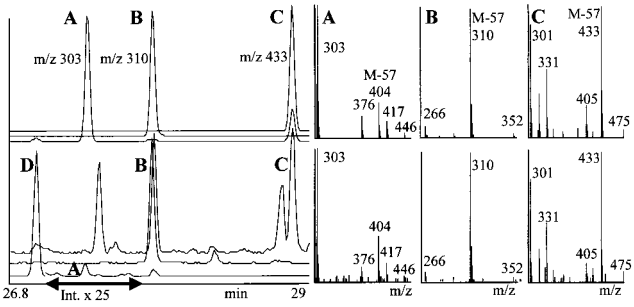

Figure 5. Extracted ion chromatograms and mass spectra of reference compounds (above) in comparison to leaf extracts (below). Identification by external reference of allothreonine (A), 6-hydroxynicotinic acid (B), and citramalic acid (C) as TBS derivatives. From 27.1 to 27.8 min, the EIC *m*/*z* 303 of the plant extract is increased 25-fold to visualize the allothreonine separation from its isomer threonine (D).

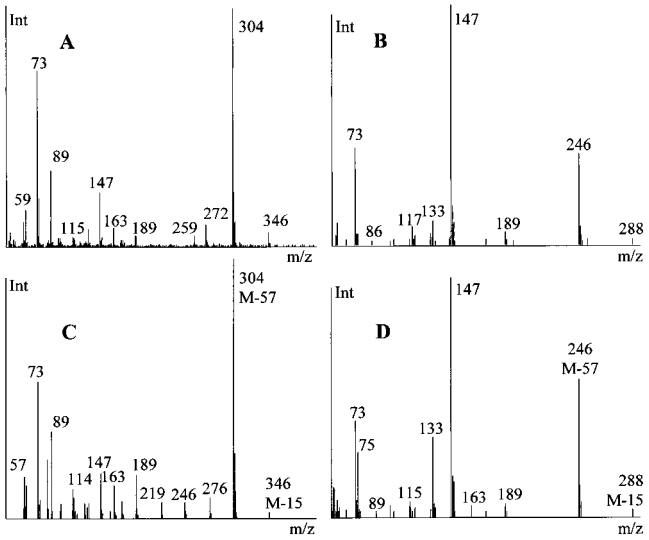

Figure 6. Mass spectra comparison of the tentatively identified Arabidopsis metabolites tartronate semialdehyde (A) and glycolic amide (B) with reference compounds *β*-hydroxypyruvate (C) and *N*-hydroxyacetamide (D), all as methoximated TBS derivatives.

have to include at least as many acidic protons as TBS groups (determined by the isotopic data), taking into account that primary amines cannot be derivatized twice by MTBSTFA and that derivatization might be incomplete for polyols. The highest ranking is given to compounds that are known to occur in standard metabolic pathway databases[25,26] that today include over 5600 metabolite entries. Artificial compounds found only in chemical databases were ranked lower. After ranking, the most likely candidates were purchased when commercially available, and retention times and mass spectra were compared to plant extract chromatograms to confirm the identity. For comparison of chromatographic elution profiles and mass spectra, the addition of internal stable isotope compounds is generally regarded to be superior to comparison with external reference compounds. We have proven the identity of 11 compounds by addition of stable isotopes as TMS derivatives (not shown) and for the proof of identity of all others by external references. For example, in Figure 5 the identity of three uncommon plant metabolites is proven by external references following TBS derivatization. For each compound, elution profiles were exactly matching, and after purification, mass spectra were nearly identical at high *m*/*z*, despite the coelution of further unknown compounds. However, quite a lot of known metabolites could not be purchased from any manufacturer. For these compounds, identifications can only be done tentatively. Mass accuracies were confirmed using gas chromatography coupled to a magnetic sector field instrument. In all cases, the most likely elemental compositions calculated from quadrupole mass spectrometric data were verified. For each calculated elemental composition, all entries in chemical databases were tested concerning the preceding information about the presence of carbonyl moieties and acidic protons. For several sum formulas, only one compound was left after all the commercially available compounds were compared to the unknown peak (Table 2) and eliminated. Further structural elucidation can be achieved by interpretation of the fragmentation pattern. For example, a loss of 32 amu was observed for the M − 57 fragment of tartronate semialdehyde, indicating cleavage of a methoxy group (Figure 6 A). The corresponding ethoximated peak showed the expected mass shift of 14 amu for the M − 57 fragment, followed by a loss of 46 amu that we believe indicates the cleavage of an ethoxy group (spectrum not shown). For this elemental composition, the

(23) http://www.chemfinder.com.
(24) http://www.chemweb.com.
(25) http://wit.mcs.anl.gov/WIT2.
(26) http://www.genome.adj.jp.

Table 3. Positively Identified Metabolites in GC/MS Chromatograms of Polar Fractions of *A. thaliana* Leaf Extracts Following Comparison with Reference Compounds[a]

	amino acids		*phosphates*		*acids and hydroxy acids*
1	4-aminobutyric acid	37	fructose 1,6-bisphosphate	64	ascorbic acid
2	alanine	38	fructose 6-phosphate	65	benzoic acid
3	β-alanine	39	glucose 6-phosphate	66	citramalic acid
4	allothreonine	40	glycerol 2-phosphate	67	citric acid
5	asparagine	41	glycerol 3-phosphate	68	2-ethylhexanoic acid
6	aspartic acid	42	phosphate	69	fumaric acid
7	arginine (carbodiimide)	43	phosphohomoserine	70	galacturonic acid
8	cysteine			71	gluconic acid
9	cystine		*sugar alcohols*	72	gluconic acid-1,5-lactone
10	glutamic acid	44	erythritol	73	glyceric acid
11	glutamine	45	ethyleneglycol	74	glyoxylic acid
12	glycine	46	galactinol	75	4-hydroxybenzoic acid
13	histidine	47	glycerol	76	6-hydroxynicotinic acid
14	homoserine	48	inositol	77	isocitric acid
15	hydroxyproline	49	mannitol	78	lactic acid
16	isoleucine	50	propyleneglycol	79	maleic acid
17	leucine			80	malic acid
18	lysine		*sugars*	81	nicotinic acid
19	methionine	51	1,6-anhydroglucose	82	oxalic acid
20	ornithine	52	erythrose	83	oxamide
21	phenylalanine	53	fructose	84	2-oxoglutamic acid
22	proline	54	fucose	85	pyruvic acid
23	pyroglutamic acid	55	galactose	86	quinic acid
24	serine	56	glucose	87	shikimic acid
25	threonine	57	glyceraldehyde	88	sinapic acid
26	tryptophan	58	maltose	89	succinic acid
27	tyrosine	59	mannose	90	threonic acid
28	valine	60	raffinose	91	uric acid
		61	rhamnose		
	N-compounds	62	sucrose		*artifacts*
29	putrescine	63	trehalose	92	ammonia
30	spermidine			93	carbonic acid
31	spermine			94	hydroxylamine
32	ethanolamine				
33	1,3-diaminopropane				
34	2-hydroxypyridine				
35	4-hydroxypyridine				
36	urea				

[a] Identifications based on TMS and TBS derivatized samples are combined.

only alternative isomer to tartronate semialdehyde is β-hydroxypyruvate, which was excluded by comparison of mass spectra (Figure 6 C) and retention times. Additional cleavages can be deduced, but as for TMS derivatives, the fragmentation of TBS compounds cannot easily be predicted, and therefore, possible structures of fragment ions remain arbitrary without use of isotopically labeled compounds and high-resolution mass spectrometers. A few compounds are not found even in standard metabolic pathways, such as glycolic amide (2-hydroxyacetamide, Figure 6B). Spectra comparison with its isomer *N*-hydroxyacetamide shows clear differences (Figure 6D) such as the higher abundance of m/z 75, 89, 115, 133, and 163 for *N*-hydroxyacetamide. By comparison with blank samples, no glycolic amide was detected. No indications of artificial cleavages of larger metabolites giving rise to smaller compounds could be found except for carbodiimide, which is known to be a major breakdown product of arginine during the silylation procedure. However, artificial compound formation can only be definitely excluded by comparison of wild-type plant tissues to mutant or transgenic plant tissues with specifically blocked biochemical pathways.

In total, 102 compounds have been identified in the polar fraction of *A. thaliana* leaf extracts (Table 3). A total of 94 of these were positively confirmed by mass spectra and by comparing retention times to reference compounds. Eight compounds have been tentatively identified by interpretation of mass spectra, comparison with related reference compounds, and search in metabolic pathways. The TBS derivatization has been especially useful for unexpected compounds and metabolites that are not commercially available. About 30 peaks were first identified as TBS derivatives by calculation of elemental compositions and database searching before identifying the corresponding TMS analogues. To our knowledge, 15 of these compounds have not been previously identified in Arabidopsis extracts, and some of them have never been identified in plants. A disadvantage of this method is that complete derivatization seems to be limited to compounds bearing from one to four acidic protons. It is therefore more difficult to identify sugars, amino sugars, and larger polyols this way. After unknown metabolites have been successfully identified, their corresponding TMS spectra are recorded and integrated into routine profiling methods.

CONCLUSIONS

It is demonstrated that the elemental compositions of unknown small metabolites can be calculated successfully using GC/MS data from routine quadrupole instruments after TBS derivatization.

Mass accuracies with errors below 0.010 amu are achievable when raw masses are recalibrated using fragments of known compounds throughout a given chromatogram. Determination of isotopic ratios further limits the range of possible sum formulas. By combining both mass and isotopic ratio determination, unknown peaks are conveniently characterized. Several uncommon polar metabolites have been identified in Arabidopsis leaf extracts. In current work, these compounds are included in routine profiling studies in order to gain a deeper understanding of metabolite regulation and gene function.

ACKNOWLEDGMENT

We gratefully acknowledge the use of the magnetic sector field instrument at the German Federal Institute of Testing Materials (BAM, Berlin) and the help of Dr. Andreas Lehmann. This work has been supported by generous financing of the Max-Planck-Society by federal research funds.

Received for review October 4, 1999. Accepted May 17, 2000.

AC991142l