

# Exact mass GC-MS analysis: protocol, database, advantages and application to plant metabolic profiling

Cyril Abadie<sup>1</sup>, Julie Lalande<sup>1</sup>, and Guillaume Tcherkez<sup>1</sup>

<sup>1</sup>Institut de Recherche en Horticulture et Semences

June 27, 2022

## Abstract

Plant metabolomics has been used widely in plant physiology, in particular to analyse metabolic responses to environmental parameters. Derivatization (via trimethylsilylation-methoximation) followed by GC-MS metabolic profiling is a major technique to quantify low molecular weight, common metabolites of primary carbon, sulphur and nitrogen metabolism. There are now excellent opportunities for new generation analyses, using high resolution, exact mass GC-MS spectrometers that are progressively becoming relatively cheap. However, exact mass GC-MS analyses for routine metabolic profiling are not common, there is no dedicated available database. Also, exact mass GC-MS is usually dedicated to structural resolution of targeted secondary metabolites. Here, we present a curated database for exact mass metabolic profiling (made of 336 analytes, 1,064 characteristic exact mass fragments) focused on molecules of primary metabolism. We show advantages of exact mass analyses, in particular to resolve isotopic patterns, localise S-containing metabolites, and avoid identification errors when analytes have common nominal mass peaks in their spectrum. We provide a practical example using leaves of different *Arabidopsis* ecotypes and show how exact mass GC-MS analysis can be applied to plant samples and identify metabolic profiles.

## Introduction

Plant metabolome analysis is now of crucial importance to describe responses to environmental conditions and thus understand associated metabolic mechanisms. Standardised metabolomics protocols have been proposed to characterise crop metabolism (Zheng, Johnson, Mandal, & Wishart, 2021). The term “metabolomics” is used to refer to techniques exploited to investigate “small” biological molecules (metabolites), i.e., extractable molecules with limited molecular weight, usually less than 500 atomic mass units (a.m.u.) (Roessner & Bowne, 2009). Two mainstream mass spectrometry techniques can be used: gas chromatography coupled to mass spectrometry (GC-MS) and liquid chromatography coupled to mass spectrometry (LC-MS) (Allwood et al., 2011; Perez de Souza, Alseekh, Naake, & Fernie, 2019). The output of metabolomics is a dataset of metabolic features (m/z metabolite peaks with retention time in LC-MS; analytes resulting from metabolite derivatisation in GC-MS) that can be used for statistics and detect metabolic changes between samples.

GC-MS metabolomics (also sometimes referred to as metabolic profiling) has been used extensively in plants under many conditions, species, or genetic backgrounds: a simple search with a literature database with query keywords ‘plant’, ‘metabolomics’ and ‘gc-ms’ returns 43,700 entries. When restricted to *Arabidopsis*, it returns 13,500 entries. It shows the massive utilisation of GC-MS for plant physiology and molecular biology. In particular, this technique is very useful to have, in a single sample analysis, a relative quantitation of most important metabolites of plant primary metabolism, such as amino acids, small soluble sugars, polyamines, or organic acids. It has thus been used to describe the response of C and N primary metabolism to major environmental cues, for example, herbivores (Jansen et al., 2009), CO<sub>2</sub> mole fraction (Högy, Keck, Niehaus, Franzaring, & Fangmeier, 2010; Misra & Chen, 2015), drought (Bowne et al., 2012; Sanchez, Schwabe, Erban, Udvardi, & Kopka, 2012), nutrient conditions (Cui, Abadie, Carroll, Lamade, & Tcherkez, 2019; Cui, Davanture, Lamade, Zivy, & Tcherkez, 2021; Cui, Davanture, Zivy, Lamade, & Tcherkez, 2019), or abiotic

stress combinations (Ghatak, Chaturvedi, & Weckwerth, 2018; Nakabayashi & Saito, 2015; Shulaev, Cortes, Miller, & Mittler, 2008).

To date, the vast majority of GC-MS analyses for metabolic profiling utilise nominal mass acquisition (i.e. at a.m.u. resolution). Accordingly, databases associated with GC-MS metabolomics such as the Golm Metabolomics Database (GMD) provide spectral data at a.m.u. resolution, and in a recent review of metabolomics resources, only nominal mass databases are discussed for GC-MS analyses (Vinaixa et al., 2016). In other words, to our knowledge, there is no directly accessible, high resolution (exact mass, i.e. at 0.0001 a.m.u. resolution or lower) and comprehensive GC-MS resource for plant metabolomics. This lack of curated, accessible and available resource for GC-MS analyses has three origins: (*i*) The availability of (affordable) exact mass GC-MS instruments is relatively recent, since the implementation of the orbitrap technology took place in the 2000s (Makarov, 2000; Makarov, Denisov, & Lange, 2009; Peterson, McAlister, Quarmby, Griep-Raming, & Coon, 2010) and the description of standard practices for high resolution GC-MS analyses has been proposed in 2021 only (Misra, 2021); (*ii*) Many ordinary applications of GC-MS metabolomics profiling do not require exact mass resolution since they are targeted on common, well-known compounds; and (*iii*) Whenever high resolution is required, LC-MS can be used. The use of high resolution in LC-MS may be important using full scan analyses, because there is a limited number of fragments (mainly parental ion and adducts) and therefore, identification essentially relies on both exact mass and isotopic pattern (De Vos et al., 2007; Kind & Fiehn, 2006). By contrast, in GC-MS analyses, the fragmentation pattern along with the retention index are used to identify analytes, with generally good accuracy. Several tools have been recently proposed to automatically annotate ions or fragments in mass spectra, in particular from LC-MS spectral data, for example in (Doerfler et al., 2014; Gaquerel, Kuhl, & Neumann, 2013; Matsuda et al., 2011; Qiu, Fine, Wherrett, Lei, & Sumner, 2016).

However, there are circumstances where high mass resolution may be desirable with GC-MS, since (*i*) several compounds with the same retention time could generate fragments with the same nominal mass and (*ii*) it could be useful to distinguish isotopic species (isotopologues) using their mass difference (for example, there is a mass excess of +1.003355 Da with  $^{13}\text{C}$  while it is +1.006277 Da with  $^2\text{H}$ ), and (*iii*) one may desire to perform untargeted GC-MS analyses with broad chemical coverage. Here, we describe an exact mass GC-MS method for high resolution routine plant metabolic profiling and provide the associated curated database, checked with authentic standards. This allows us to address aspects (*i*) and (*ii*) directly. We also provide the list of current compounds having similar nominal-mass fragments and similar retention time but can be distinguished easily with exact mass, avoiding quantification errors. We also take advantage of sulphur isotopes at natural abundance to allow the identification of S-containing fragments in datasets. Finally, we applied our protocol and the database using Arabidopsis leaves to show how it can be applied to real samples, allowing facile differentiation of genetic accessions.

## Material and methods

### Chemicals

Chemicals were from Sigma Aldrich (Merck; with reference numbers between parentheses): Mass Spectrometry Metabolite Library (MSMLS, IROA Technologies), 2-Oxoadipic acid (75447), 3-Iodo-L-Tyrosine (I8250), 5-Aminolevulinic acid (08339),  $\alpha$ -Tocopherol (T3251),  $\beta$ -Gentiobiose (G3000), Betain (W422312), Chlorogenic acid (C3878), Choline (C1879), Citraconic acid (C82604), D-2-Aminobutyric acid (116122), D-3-phosphoglyceric acid (P8877), D-Erythronic acid (75025), D-Erythronic acid  $\gamma$ -lactone (374385), D-Galacturonic acid (48280), D-Glucuronic acid (G5269), Diaminopimelate (33240), D-Melibiose (92413), D-Pinitol (441252), D-Quinic acid (138622), D-Threitol (377619), D-Xylonic acid  $\gamma$ -lactone (89339), Dulcitol (D0256), Galactinol (79544), Hydrocinnamic acid (135232), L- $\alpha$ -Glycerophosphocholine (G5291), Itaconic acid (I29204), L-carnitine (C0283), L-Cystathionine (C7505), Levoglucosan (06724), L-Fucose (F2252), L-Pyroglutamic acid (83160), Nicotinic acid (N4126), Norleucine (N6877), O-Acetyl-Serine (CDS020792), Phosphocholin (P0378), Sigmastrol (S2424), Sinapinic Acid (D7927), Mysteric acid (M3128), Tryptamine (193747), Urea (U5378), Xanthosine (CDS020790).

## Plant material

The height *Arabidopsis thaliana* accessions were obtained from INRA Versailles Genomic Resource Centre VNAT collection: An-1 (96AV, Belgium), Bl-1 (42 AV, Italy), Col-0 (186AV, Poland), Cvi-0 (166AV, Cape Verde), Ge-0 (101AV, Switzerland), Mt-0 (94AV, Libya), Oy-0 (224AV, Norway) and Shahdara (Tadjikistan). Plants were grown in a growth chamber under short photoperiod (8h light/16h dark), PPFD = 150  $\mu\text{mol m}^{-2} \text{s}^{-1}$ , temperature 22°C/18°C and relative humidity 60%. Plants were watered every 3 days and once a week with Plant-Prod (15-10-30, N-P-K, Fertil). Fifty-five (55) days after sowing, two fully developed leaves were sampled on each rosette and instantly quenched in liquid nitrogen. Frozen samples were then frozen-dried, dry material was grinded and 5 mg of dry powder used for metabolites extraction by adding 400  $\mu\text{L}$  of methanol/water (90/10) with adonitol (55  $\mu\text{M}$ ) as an internal standard.

## GC-MS analyses

A detailed, step-by-step protocol is provided as a Supplementary Material. This section summarizes how exact mass GC-MS analyses and data extraction with TraceFinder® were performed. GC-MS analyses were carried out using a GC-MS-Orbitrap Q Exactive (Thermo Scientific, Courtaboeuf, France). Fifty five (55)  $\mu\text{L}$  of each standard solution (50  $\mu\text{M}$ ), or 4  $\mu\text{L}$  from the plant leaf extract, were poured into a vial (with insert) and spin-dried at 39°C. Samples were derivatized (automatically with a preparative robot) with 20  $\mu\text{L}$  methoxyamine (20 mg  $\text{mL}^{-1}$  in pyridine; 90 min at 37°C) and 30  $\mu\text{L}$  N-methyl-N(trimethylsilyl)-trifluoroacetamide (MSTFA) for 30 min at 37°C. Prior to injection, 5  $\mu\text{L}$  of alkane mix (14 alkanes from C<sub>9</sub> to C<sub>36</sub>, 3  $\mu\text{g mL}^{-1}$ , Connecticut n-Hydrocarbon Mix, Supelco) were added in each sample to compute the retention index. Analyses were performed by injecting 1  $\mu\text{L}$  in splitless mode at 230°C (injector temperature) in a TG-5 SILMS column (30 m x 0.25 mm x 0.25  $\mu\text{m}$ ; Thermo Scientific) set in a Trace 1300 Series GC (Thermo Scientific). Helium was used as gas carrier with a constant flow of 1  $\text{mL min}^{-1}$ . After one minute at initial GC oven temperature (70°C), temperature was raised to 325°C at 15°C  $\text{min}^{-1}$  and finally kept at 325°C for 4 min. MS analyses were operated in positive polarity in full MS scan mode with the following source settings: mass scan range 50-750  $\text{m/z}$ , resolution 60,000, AGC target 1E6, MS transfer line 300°C and filament delay 4.12 min. Ionisation by electron impact (70 eV) was performed at 250°C ion source temperature. Amino acids were identified automatically using TraceFinder® (Thermo Scientific) using retention time, major characteristic fragment ( $\text{m/z}$  ion) and a confirmation fragment, with a maximum tolerance of 0.00007 Da via targeted screening.

## Results

### Overview of GC-MS analysis

As an example, the chromatogram associated with an *Arabidopsis* leaf sample derivatised with methoxyamine and N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) in pyridine is shown in Fig. 1, where the signal shown in panel (a) is total ion current (TIC). Abundant compounds form peaks that can be easily seen on the chromatogram, including some amino acids (e.g. proline), organic acids (fumarate) and sugars (e.g. sucrose), as trimethylsilylated ( $-\text{Si}(\text{CH}_3)_3$ , abbreviated TMS) derivatives. The TIC signal does not directly reflect the quantity (in moles) in the sample since analytes do not have all the same response coefficient in the mass spectrometer. The internal standard used for semi-quantitation (i.e. relative quantification) is adonitol (ribitol), the derivative of which (ribitol 5TMS) elutes at 10.68 min. As expected, there is some overlapping between analytes (i.e. several analytes have very close retention time and co-elute). This is visible, for example, at 15.05 (panel (b) of Fig. 1) where many  $\text{m/z}$  features related to sugars are visible together with 6-phosphogluconate specific  $\text{m/z}$  features (for example at 387.14254 Da). Interestingly, a relatively close mass (387.32912 Da) can be found at a similar retention time (15.25 min) from arachidic acid 1TMS, but such a mass difference is in practice too large (nearly 0.18 Da) to allow confusion with exact mass analysis.

### Mass resolution and isotopic pattern

Mass resolution across the  $\text{m/z}$  window (50-750 Da) was not found to be constant, varying from less than 0.0001 Da at low mass (< 100 Da) to 0.0017 Da at higher mass (> 350 Da), i.e. within [?]1 and 5 ppm. In

principle, this should allow to solve the natural isotopic pattern and distinguish  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^{33}\text{S}$  isotopologues ( $m+1$ ) and  $^{13}\text{C}_2$  and  $^{34}\text{S}$  isotopologue ( $m+2$ ). However, this depends on whether analytes of interest are silylated or not, because Si has two isotopes ( $^{29}\text{Si}$  and  $^{30}\text{Si}$ ), with mass excess values (with respect to the monoisotopic species) very close to that of  $^{33}\text{S}$  and  $^{34}\text{S}$  (sulphur-containing species are addressed further below). This is illustrated in Fig. 2, using a non-silylated compounds (naturally volatile and thus not requiring derivatisation), isopropyl-4-methylthiazole (IMT). In fact, using the example of the major fragment (molecule minus  $\text{CH}_3$ ) (Fig. 2a),  $^{15}\text{N}$ ,  $^{33}\text{S}$ ,  $^2\text{H}$  and  $^{13}\text{C}$  isotopologues are resolved (Fig. 2b,c). With methionine 2TMS (Fig. 2d), the presence of Si does not allow full resolution of the isotopic pattern. This is shown using the  $m+2$  isotopologues (Fig. 2e) where some isotopic species are found to be undistinguishable ( $^{34}\text{S}$ ,  $^{30}\text{Si}$ ). Despite this limitation, the expected isotopologue abundance (predicted from isotope abundance and elemental composition) is very close to observed abundance (Fig. 2f). A difference between observed and predicted abundance is visible for the deuterated isotopologue of IMT (red circle), likely due to the very small natural abundance of deuterium (0.015%) and thus the larger imprecision in quantitation. Also, it should be noted that  $^{13}\text{C}$  isotopologues are well-separated (resolved) from other isotopologues, allowing facile monitoring of  $^{13}\text{C}$  molecules during plant labelling experiments, for example.

### *Differentiation of metabolites with exact mass*

Exact mass allows one to distinguish  $m/z$  features with the same nominal mass but with very slight differences (well below 0.1 Da), reflecting differences in elemental composition. This is particularly useful when retention time of analytes at the origin of similar  $m/z$  features are very close. We took advantage of GC-MS exact mass analyses of authentic standards used to construct the database (available as a Supplementary Files), to tackle this problem. Amongst most common plant metabolites, many analytes were found to have both similar retention time and identical nominal mass  $m/z$  signals, and they are listed in Table 1. This problem can potentially lead to identification mistakes with nominal mass analyses especially when mass spectra are rather similar. Two examples are further illustrated in Fig. 3. Ferulic acid 2TMS and tyramine 2TMS have the same retention time (12.09 min) and share the same nominal mass ion 338 a.m.u. However, there is a clear difference in exact mass (Fig 3a,b). Also, their spectra share little similarity (Fig. 3a). Malic acid 3TMS and threitol 4TMS have a retention time difference of less than 4 s and share a feature at a nominal mass of 189. In addition, their mass spectra are partly similar, in particular for features  $< 200$  a.m.u (Fig. 3c). However, the exact mass of features at 189 a.m.u. is readily distinguishable (189.11309 and 189.076741 Da) (Fig. 3d).

### *Arginine derivatives*

In terms of biochemical analysis, one of the most problematic metabolites of plant primary N metabolism is arginine, because it can be converted to several by-products upon derivatisation. It is particularly so during silylation, since the guanidium group of arginine may be cleaved to form ornithine, which in turn can lead to a cyclic form, ornithine lactam. In addition, one N atom of arginine can be lost and form citrulline. Taken as a whole, arginine derivatisation generates ornithine lactam 2TMS, ornithine 3TMS, arginine 3TMS and citrulline 3TMS (Fig. 4a). Arginine 3TMS is the minor derivative and has a retention time extremely close to citrulline 3TMS, and therefore it has often been overlooked. Here, arginine 3TMS is visible on the TIC, can also be identified with certainty with exact mass and thus cannot be confused with citrulline 3TMS. Although small, features typical of arginine 3TMS and citrulline 3TMS can be found at 187.108676 Da (guanidium group 2TMS) and 171.08278 Da (decarboxylated citrulline), respectively (Fig. 4b-e).

### *Identification of S-containing analytes*

The resolution of the isotopic pattern can be exploited to gain information on uncommon (or less common) elements contained by analytes. Nitrogen  $^{15}\text{N}$  is not sufficiently abundant (0.36% only) and can be overlapped by  $^{29}\text{Si}$  (4.7%) and  $^{33}\text{S}$  (0.76%) isotopes. Therefore, the use of the isotopic pattern to detect specific elements in analytes is only suitable for sulphur, provided (*i*) the signal coming from  $^{34}\text{S}$  is high enough ( $^{34}\text{S}$  abundance is 4.2%), and (*ii*) it can be somehow distinguished from  $^{30}\text{Si}$ . The mass excess of the  $^{34}\text{S}$  isotopologue is +1.995796 while that of  $^{30}\text{Si}$  is +1.996844, meaning a difference of 0.001048 Da (i.e., 1.048 resolution during

analysis is sufficient, mass excess at nearly exactly +1.995796 can be found. A similar analysis with high resolution LC-MS (without the issue of Si isotopes) has been undertaken previously (Nakabayashi & Saito, 2017). In Fig. 5, automatic searching of mass excess of +1.995796 is shown, with the difference (in of feature  $m/z$ ). Using a mixture of authentic standards that includes methionine and cysteine (Fig. 5a), only 9 (out of more than 2,000) features were found to fall in the mass excess +1.995796 $\pm$ 0.00005 (red shaded area). From these, manual checking allows confirmation of 3 of them, another one (labelled as **4** in Fig. 5a) being ambiguous: **1** and **2** are S-containing fragments of cysteine 2TMS (61.011195 and 131.035071 Da at 5.61 min), while **3** is a fragment of methionine 2TMS (176.092921 at 9.12 min). **4** appears at the retention time of methionine 2TMS but in the spectrum, the peak is too wide to ascertain that it is not another fragment with  $^{30}\text{Si}$  (i.e., other fragments without S with nearly the same exact mass). Using a plant sample (Arabidopsis leaf extract, Fig. 5b), about 10 features fall in the  $\pm 0.05$  two can be confirmed further: **1** is the same as in Fig. 5a (S-containing fragment of cysteine 2TMS) and **5** is  $\text{C}_3\text{H}_6\text{S}$  (74.018421 Da) and comes from methionine 2TMS, at 9.12 min. The appearance of this feature is further illustrated in Fig. 5c, where it is found to be very close to the isotopic pattern of another methionine 2TMS fragment ( $\text{C}_2\text{H}_6\text{OSi}$ ). Of course, looking at the mass excess (due to  $^{34}\text{S}$ ) with a high precision in Figs. 5a and 5b is only possible if resolution is high enough to have a better probability to have an acquired  $m/z$  datapoint exactly at, or very close to +1.995796. This is illustrated in Fig. 5d-f, where the observed signal (grey) is likely to show a mass signal coming from  $^{34}\text{S}$  at medium and high resolution. This can optimally be achieved for relatively small features ( $m/z < 150$  a.m.u.) for which resolution is larger.

### *Application to plant samples*

The compound database used here contains 1,064 target fragments that have all been confirmed with standards, representing 336 compounds with target identification exact mass features, and confirming exact mass features (see Supplementary File 1). Using Arabidopsis as a plant model, we found 234 compounds in leaves (266 analytes, some of them generating several derivatives). We assessed the usefulness and repeatability of exact mass analysis by conducting GC-MS analyses on several Arabidopsis ecotypes (BL1, AN1, Col0, Cvi0, Oy0, Ge0, Mt0 and Sha). Results are shown in Fig. 6 as a heatmap of analytes that are significantly different between ecotypes. Hierarchical clustering shows easy discrimination of samples (without any clustering error between ecotypes), suggesting good repeatability. Many metabolites were found to be different between ecotypes, and they are here subdivided into 7 groups (Fig. 6a). For example, Cvi0 was found to be enriched in various amino acid derivatives, while Shahdara was enriched in sugars. Repeatability (both technical and biological) was assessed by calculating standard errors, expressed in % of mean value, for each genotype. For the vast majority of metabolites, there was a very good reproducibility, with a median value of SE of 9.7% only across all genotypes (Fig. 6b). Some metabolites were associated with a high variability, larger than 40%. They were all metabolites present in very low amount, and/or having very poor ionisation, and/or encountering degradation during derivatisation, such that their relative signal was extremely small,  $10^{-5}$  or less (Fig. 6c).

## **Discussion**

### *Exact mass GC-MS database for routine analysis*

The present curated database contains 336 compounds, 234 of them being identified and quantified in Arabidopsis leaves. This might appear relatively small compared with the total estimated number of small metabolites (several thousands) in plants. However, this compares well with most targeted routine GC-MS analyses for metabolic profiling, which yield a list of about 80-100 metabolites in the vast majority of cases (for example, there are 162 metabolites in (Cui, Davanture, et al., 2019), and 178 in (Cui et al., 2021) found in leaves). Of course, the ability of instruments and softwares to extract a proper dataset from raw data using the database depends on the quality of analyses. In effect, despite the considerable dynamic range of modern instruments (here, 6 orders of magnitude in peak height), precise quantification can only be carried out when analytes are not too concentrated (inadequate peak shape do not allow peak extraction by softwares like Tracefinder(r)) (Kaufmann & Walker, 2017). This can be challenging when some metabolites are present in high amounts (e.g. sucrose or proline) while others are present in trace amounts or generate a weak signal

(e.g. salicylamide) (Fig. 6). It should be noted that data extraction from raw data can also be processed via untargeted peak searching, providing a much more powerful way to appreciate the diversity of molecules present in extracts (Perez de Souza et al., 2019). However, this has two drawbacks: (i) processing time is very long (at least 20 times slower with Tracefinder(r)), and (ii) many peaks would appear as unidentified, with only the  $m/z$  value and retention time (and thus post-hoc identification is required using exact mass and potentially, co-occurring fragments). Therefore, for routine analyses, it is probably more convenient to rely on targeted analyses with the database we propose here.

### *Advantages of exact mass analysis*

Nominal mass GC-MS analyses rely on two main criteria: retention index, and mass spectrum (Garcia & Barbas, 2011). A well-recognised difficulty with mass spectrum analysis and comparisons in biological samples is co-elution of analytes, leading to mixed mass spectra. To address this problem, deconvolution algorithms can be used (e.g. in ChromaTOF(r), or the deconvolution plugin in TraceFinder(r)) (Du & Zeisel, 2013; H. Lu, Liang, Dunn, Shen, & Kell, 2008) or alternatively, data extraction can focus on individual target  $m/z$ , e.g., in MetabolomeExpress (Carroll, Badger, & Harvey Millar, 2010). However, there remains a risk that co-elution leads to erroneous identification and quantitation when common  $m/z$  nominal mass features are present. Here, we have covered this aspect by searching systematically co-eluting nominal masses  $m/z$  (Table 1, Fig. 3). Rather common plant metabolites are concerned by this problem (e.g., malate, serine, etc.). Exact mass offers an efficient way to avoid it since exact mass values of targeted  $m/z$  features are clearly different. In addition, the isotopic pattern reflects elemental composition (Fig. 2) and can be used as a post-hoc verification. Erroneous identification is therefore much less probable. Here, using Arabidopsis samples, there was only one case of misidentification: iodo-tyrosine 3TMS has been identified by TraceFinder(r) at 16.12 min (Fig. 6c). This molecule is very unlikely since enzymes of iodo-tyrosine metabolism (thyroperoxidase and iodo-tyrosine deiodinase) are absent in plants (Phatarpkehar, Buss, & Rokita, 2014; Taurog, 1999). Our database available in Supplementary material has been curated accordingly. Coincidentally, several unusual sugar-containing molecules with similar retention time can generate fragments ( $m/z$  features) with the same exact mass. For example, rutinose 6TMS 1Meox elutes just before sucrose 8TMS (16.17 min here) and can form a target fragment at 218.1033 like iodo-tyrosine (Fig. SX). The observed signal is very low ( $10^{-6}$  that of sucrose 8TMS), suggesting it is a minor compound. Further work is needed (for example using  $MS^2$  on a purified fraction where this compound is more concentrated and thus generates well-visible  $MS^2$  fragments) to determine what compound generates a signal similar to iodo-tyrosine in Arabidopsis samples.

Another advantage of exact mass analyses is the resolution of isotopologues. This is of particular interest to identify  $^{13}C$ -molecules upon double-labelling (typically  $^{13}C$ - $^{15}N$ , or  $^{13}C$ - $^{33}S$ ) since the signal associated with  $^{13}C$  mass excess (+1.003355 Da) is readily visible. This allows direct quantitation of  $^{13}C$  isotopologues unlike in nominal mass analyses where all  $m+1$  isotopologues are under the same peak. In a recent study, the systematic, non-targeted analysis of  $^{13}C$  isotopologues (+1.003355 Da, +2.006710 Da, +3.010065 Da, etc.) has been used with high resolution LC-MS to look at variations in leaf metabolites with  $CO_2$  and  $O_2$  gaseous conditions (Abadie, Lalande, Limami, & Tcherkez, 2021). For  $^{34}S$ -isotopologues, there is an interference with  $^{30}Si$  (silicium being carried by trimethylsilyl groups). Despite this limitation, it is worth noting that whenever mass resolution is sufficiently high (in particular for low  $m/z$  values), it is possible to identify S-containing fragments (Fig. 5). Resolution is less of an issue upon labelling since  $^{34}S$  would prevail over  $^{30}Si$  and thus the peak apex would be closer to the mass excess of  $^{34}S$  (+1.995796 Da).

### *Application to plant samples*

Here, we took advantage of Arabidopsis samples to show how exact mass GC-MS analysis can be performed routinely on plant samples, with a rather good outcome, i.e. identification and quantification of more than 200 analytes. Another attempt to carry out high-resolution GC-MS for primary metabolites and pesticides analysis in Arabidopsis can be found in, e.g., (Peterson et al., 2010). In terms of methodology and instrumentation, high resolution exact mass GC-MS is similar to nominal mass GC-MS analysis in that it requires standardised derivatisation (optimally with a robotic facility) ensuring that all samples are treated

evenly (including the same time lag between the end of derivatisation and injection), and quality controls to assess reproducibility and quantitativity (for a specific discussion about automation of derivatisation for GC-MS, see (Zarate et al., 2016)). Of course, a specific feature of exact mass analysis is that it requires checking mass accuracy (W. Lu et al., 2017), and here this was performed on a day-to-day basis with calibration gas FC43 containing perfluorotributylamine PFTBA (from ThermoFisher Scientific). Also, it is extremely useful to have specific samples to appreciate mass accuracy, for example with a compound that does not require derivatisation like IMT (Fig. 2a-b). This allows one to check not only mass accuracy but also precision of mass excess values associated with all relevant isotopes (C, N, S isotopes; Fig. 2c).

It should be noted that exact mass GC-MS, however, does not solve the issue of multiple derivatives for some specific metabolites. In particular, amino acids can form by-products upon silylation. So is the case of glutamic acid and glutamine, that can both yield pyroglutamate (for a recent discussion on this issue, see (Miyagawa & Bamba, 2019)). It is also the case of arginine, which is known to yield several products (Molnar-Perl & Katona, 2000), and here it formed four products (Fig. 4). High resolution analysis nevertheless provides a method to have a better picture of derivatives and here, we identified 4 main products of arginine derivatisation: ornithine lactam 2TMS, ornithine 3TMS, arginine 3TMS and citrulline 3TMS. Since these different compounds do not have the same response coefficient in the mass spectrometer (i.e. distinct observed signal response curve to concentration), arginine cannot be quantified very precisely by this method.

Basically, GC-MS analysis provides semi-quantitative information on plant samples. A good method to have information on effective quantity (i.e., absolute quantity) of analytes is GC-C-IRMS (GC coupled to combustion and isotope ratio mass spectrometry). In effect, the same sample (same derivatisation) can be injected and the GC-C-IRMS converts quantitatively each peak to CO<sub>2</sub> and N<sub>2</sub> and thus can provide direct information on the amount of carbon (via mass 44 monitoring) or nitrogen (via mass 28 monitoring) in each peak of interest. This method has been used recently (using alfalfa seed protein extracts) and shown to provide satisfactory estimates of absolute amino acid contents (Domergue, Lalande, Abadie, & Tcherkez, 2022). More classical method can also be used, such as external calibration curves, or deuterated standards (internal references). The use of deuterated standards has three drawbacks: First, it complicates mass spectra, with some probability to coincide with m/z features of interest. An example with plant samples is cysteine 3TMS where the target peak (C<sub>8</sub>H<sub>22</sub>NSSi<sub>2</sub>) at 220.1009 Da might coincide with a deuterated fragment ([<sup>2</sup>H]-C<sub>7</sub>H<sub>21</sub>NOSi<sub>3</sub>, 220.0995 Da). Second, having all deuterated standards is tedious and expensive. Third, deuteration can cause an isotope effect in either chromatography or ionisation efficiency so that the GC-MS signal differs between deuterated and the protiated forms (Alzweiri, Khanfar, & Al-Hiari, 2015; Caban & Stepnowski, 2020; Matucha, Jockisch, Verner, & Anders, 1991; Ripszam, Grabic, & Haglund, 2013).

### *Perspectives*

Our results show that high resolution exact mass GC-MS can be used in plants, as an efficient way to carry out metabolic profiling. This technique shows several advantages over classical, nominal mass GC-MS, such as identification less prone to errors thanks to exact mass m/z features. The database we propose here to perform targeted analyses covers more than 200 metabolites of different chemical families. This database is evolutive, in that it will be enriched over the coming years and updated versions kept accessible via the journal website. We recognise that plants contain a lot of sugars, with many isomers. Some of them have identical retention times, and thus cannot be resolved with exact mass since m/z features have the same elemental composition. To solve this problem, specific devices like ion mobility may be required (Morrison & Clowers, 2018; Mu, Schulz, & Ferro, 2018; Przybylski & Bonnet, 2021). In terms of isotopic analysis via exact mass GC-MS, Si isotopes represent a limitation to distinguish isotopologues. We showed that high mass precision allows reasonable access to <sup>34</sup>S isotopologues, but less so for <sup>15</sup>N or <sup>33</sup>S that are much less abundant. When isotopic enrichments are modest, GC-C-IRMS analysis is the best alternative, but this technology is presently not able to perform <sup>34</sup>S analysis. Therefore, exact mass LC-MS seems to be better suited to isotopic analysis of sulphur-containing compounds.

## Acknowledgements

Authors acknowledge the support of the *Region Pays de la Loire* and *Angers Loire Metropole* via the grant Isoseed, awarded to G.T. Authors also thank ThermoScientific for helping us setting up the robotic facility and modifying programs to fix the sample preparation sequence.

## Conflict of interest

The authors declare no conflict of interest.

## Supplementary material provided in this paper

**Protocol S1.** Protocol for exact-mass GC-MS analysis. (PDF)

**Supplementary Fig. S1 .** Alternative compounds generating same exact mass fragments as iodotyrosine.

**Supplementary File 1 .** Database is source format.

**Supplementary File 2.** List of compounds and fragments contained in the database, and compounds found in *Arabidopsis* samples (Excel format).

## References

- Abadie, C., Lalande, J., Limami, A. M., & Tcherkez, G. (2021). Non-targeted  $^{13}\text{C}$  metabolite analysis demonstrates broad re-orchestration of leaf metabolism when gas exchange conditions vary. *Plant, Cell & Environment*, *44* (2), 445-457.
- Allwood, J. W., De Vos, R. C. H., Moing, A., Deborde, C., Erban, A., Kopka, J., . . . Hall, R. D. (2011). Plant metabolomics and its potential for systems biology research: Background concepts, technology, and methodology. *Methods in enzymology*, *500*, 299-336.
- Alzweiri, M., Khanfar, M., & Al-Hiari, Y. (2015). Variations in GC-MS Response Between Analytes and Deuterated Analogs. *Chromatographia*, *78* (3), 251-258.
- Bowne, J. B., Erwin, T. A., Juttner, J., Schnurbusch, T., Langridge, P., Bacic, A., & Roessner, U. (2012). Drought Responses of Leaf Tissues from Wheat Cultivars of Differing Drought Tolerance at the Metabolite Level. *Molecular Plant*, *5* (2), 418-429.
- Caban, M., & Stepnowski, P. (2020). The application of isotopically labeled analogues for the determination of small organic compounds by GC/MS with selected ion monitoring. *Analytical Methods*, *12* (30), 3854-3864.
- Carroll, A. J., Badger, M. R., & Harvey Millar, A. (2010). The MetabolomeExpress Project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets. *BMC bioinformatics*, *11* (1), 1-13.
- Cui, J., Abadie, C., Carroll, A., Lamade, E., & Tcherkez, G. (2019). Responses to K deficiency and waterlogging interact via respiratory and nitrogen metabolism. *Plant, Cell & Environment*, *42* (2), 647-658.
- Cui, J., Davanture, M., Lamade, E., Zivy, M., & Tcherkez, G. (2021). Plant low-K responses are partly due to Ca prevalence and the low-K biomarker putrescine does not protect from Ca side effects but acts as a metabolic regulator. *Plant, Cell & Environment*, *44* (5), 1565-1579.
- Cui, J., Davanture, M., Zivy, M., Lamade, E., & Tcherkez, G. (2019). Metabolic responses to potassium availability and waterlogging reshape respiration and carbon use efficiency in oil palm. *New Phytologist*, *223* (1), 310-322.
- De Vos, R. C. H., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., & Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature protocols*, *2* (4), 778-791.



- Doerfler, H., Sun, X., Wang, L., Engelmeier, D., Lyon, D., & Weckwerth, W. (2014). mzGroupAnalyzer-predicting pathways and novel chemical structures from untargeted high-throughput metabolomics data. *Plos One*, 9 (5), e96188.
- Domergue, J.-B., Lalande, J., Abadie, C., & Tcherkez, G. (2022). Compound-Specific  $^{14}\text{N}/^{15}\text{N}$  Analysis of Amino Acid Trimethylsilylated Derivatives from Plant Seed Proteins. *International Journal of Molecular Sciences*, 23 (9), Article no. 4893.
- Du, X., & Zeisel, S. H. (2013). Spectral deconvolution for gas chromatography mass spectrometry-based metabolomics: current status and future perspectives. *Computational and structural biotechnology journal*, 4 (5), e201301013.
- Gaquerel, E., Kuhl, C., & Neumann, S. (2013). Computational annotation of plant metabolomics profiles via a novel network-assisted approach. *Metabolomics*, 9 (4), 904-918.
- Garcia, A., & Barbas, C. (2011). Gas chromatography-mass spectrometry (GC-MS)-based metabolomics. In *Metabolic Profiling* (pp. 191-204): Springer.
- Ghatak, A., Chaturvedi, P., & Weckwerth, W. (2018). Metabolomics in plant stress physiology. *Plant Genetics and Molecular Biology*, 187-236.
- Hogy, P., Keck, M., Niehaus, K., Franzaring, J., & Fangmeier, A. (2010). Effects of atmospheric  $\text{CO}_2$  enrichment on biomass, yield and low molecular weight metabolites in wheat grain. *Journal of Cereal Science*, 52 (2), 215-220.
- Jansen, J. J., Allwood, J. W., Marsden-Edwards, E., van der Putten, W. H., Goodacre, R., & van Dam, N. M. (2009). Metabolomic analysis of the interaction between plants and herbivores. *Metabolomics*, 5 (1), 150-161.
- Kaufmann, A., & Walker, S. (2017). Comparison of linear intrascan and interscan dynamic ranges of Orbitrap and ion-mobility time-of-flight mass spectrometers. *Rapid Communications in Mass Spectrometry*, 31 (22), 1915-1926.
- Kind, T., & Fiehn, O. (2006). Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC bioinformatics*, 7 (1), 1-10.
- Lu, H., Liang, Y., Dunn, W. B., Shen, H., & Kell, D. B. (2008). Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS. *TrAC Trends in Analytical Chemistry*, 27 (3), 215-227.
- Lu, W., Su, X., Klein, M. S., Lewis, I. A., Fiehn, O., & Rabinowitz, J. D. (2017). Metabolite measurement: pitfalls to avoid and practices to follow. *Annual Review of Biochemistry*, 86 , 277.
- Makarov, A. (2000). Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical Chemistry*, 72 (6), 1156-1162.
- Makarov, A., Denisov, E., & Lange, O. (2009). Performance evaluation of a high-field Orbitrap mass analyzer. *Journal of the American Society for Mass Spectrometry*, 20 (8), 1391-1396.
- Matsuda, F., Nakabayashi, R., Sawada, Y., Suzuki, M., Hirai, M. Y., Kanaya, S., & Saito, K. (2011). Mass spectra-based framework for automated structural elucidation of metabolome data to explore phytochemical diversity. *Frontiers in plant science*, 2 , Article no. 40.
- Matucha, M., Jockisch, W., Verner, P., & Anders, G. (1991). Isotope effect in gas—liquid chromatography of labelled compounds. *Journal of Chromatography A*, 588 (1), 251-258.
- Misra, B. B. (2021). Advances in high resolution GC-MS technology: a focus on the application of GC-Orbitrap-MS in metabolomics and exposomics for FAIR practices. *Analytical Methods*, 13 (20), 2265-2282.

- Misra, B. B., & Chen, S. (2015). Advances in understanding CO<sub>2</sub> responsive plant metabolomes in the era of climate change. *Metabolomics*, 11 (6), 1478-1491.
- Miyagawa, H., & Bamba, T. (2019). Comparison of sequential derivatization with concurrent methods for GC/MS-based metabolomics. *Journal of bioscience and bioengineering*, 127 (2), 160-168.
- Molnar-Perl, I., & Katona, Z. F. (2000). GC-MS of amino acids as their trimethylsilyl/t-butyldimethylsilyl Derivatives: In model solutions III. *Chromatographia*, 51 (1), S228-S236.
- Morrison, K. A., & Clowers, B. H. (2018). Contemporary glycomic approaches using ion mobility-mass spectrometry. *Current Opinion in Chemical Biology*, 42 , 119-129.
- Mu, Y., Schulz, B. L., & Ferro, V. (2018). Applications of Ion Mobility-Mass Spectrometry in Carbohydrate Chemistry and Glycobiology. *Molecules (Basel, Switzerland)*, 23 (10), Article no. 2557.
- Nakabayashi, R., & Saito, K. (2015). Integrated metabolomics for abiotic stress responses in plants. *Current opinion in plant biology*, 24 , 10-16.
- Nakabayashi, R., & Saito, K. (2017). Ultrahigh resolution metabolomics for S-containing metabolites. *Current Opinion in Biotechnology*, 43 , 8-16.
- Perez de Souza, L., Alseekh, S., Naake, T., & Fernie, A. (2019). Mass Spectrometry-Based Untargeted Plant Metabolomics. *Current Protocols in Plant Biology*, 4 (4), e20100.
- Peterson, A. C., McAlister, G. C., Quarmby, S. T., Griep-Raming, J., & Coon, J. J. (2010). Development and characterization of a GC-enabled QLT-Orbitrap for high-resolution and high-mass accuracy GC/MS. *Analytical Chemistry*, 82 (20), 8618-8628.
- Phatarphekar, A., Buss, J. M., & Rokita, S. E. (2014). Iodotyrosine deiodinase: a unique flavoprotein present in organisms of diverse phyla. *Molecular BioSystems*, 10 (1), 86-92.
- Przybylski, C., & Bonnet, V. (2021). Discrimination of isomeric trisaccharides and their relative quantification in honeys using trapped ion mobility spectrometry. *Food Chemistry*, 341 , Article no. 128182.
- Qiu, F., Fine, D. D., Wherrett, D. J., Lei, Z., & Sumner, L. W. (2016). PlantMAT: A Metabolomics Tool for Predicting the Specialized Metabolic Potential of a System and for Large-Scale Metabolite Identifications. *Analytical Chemistry*, 88 (23), 11373-11383.
- Ripszam, M., Grabic, R., & Haglund, P. (2013). Elimination of interferences caused by simultaneous use of deuterated and carbon-13 standards in GC-MS analysis of polycyclic aromatic hydrocarbons (PAHs) in extracts from passive sampling devices. *Analytical Methods*, 5 (12), 2925-2928.
- Roessner, U., & Bowne, J. (2009). What is metabolomics all about? *Biotechniques*, 46 (5), 363-365.
- Sanchez, D. H., Schwabe, F., Erban, A., Udvardi, M. K., & Kopka, J. (2012). Comparative metabolomics of drought acclimation in model and forage legumes. *Plant, Cell & Environment*, 35 (1), 136-149.
- Shulaev, V., Cortes, D., Miller, G., & Mittler, R. (2008). Metabolomics for plant stress response. *Physiologia plantarum*, 132 (2), 199-208.
- Taurog, A. (1999). Molecular evolution of thyroid peroxidase. *Biochimie*, 81 (5), 557-562.
- Vinaixa, M., Schymanski, E. L., Neumann, S., Navarro, M., Salek, R. M., & Yanes, O. (2016). Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects. *TrAC Trends in Analytical Chemistry*, 78 , 23-35.
- Zarate, E., Boyle, V., Rupprecht, U., Green, S., Villas-Boas, S. G., Baker, P., & Pinu, F. R. (2016). Fully automated trimethylsilyl (TMS) derivatisation protocol for metabolite profiling by GC-MS. *Metabolites*, 7 (1), 1.

Zheng, J., Johnson, M., Mandal, R., & Wishart, D. S. (2021). A Comprehensive Targeted Metabolomics Assay for Crop Plant Sample Analysis. *Metabolites*, 11 (5), Article no. 303.

**Table 1. Common analytes that share both similar nominal mass signals and retention time but can be distinguished using exact mass .** For each example, the elemental composition of the fragment and its exact mass is shown. When the observed mass of interest belongs to the isotopic pattern, it is indicated with the isotope in front ( $^{13}\text{C}$ ). See also Fig. 4 for a detailed analysis of two examples, with fragment chemical structure. This table shows couple of analytes with a difference in retention time of less than 0.4 min. (\*) Note that m/z ions at [?]226.092 Da can also come from a fragment of allantoin 2TMS ( $\text{C}_4\text{H}_{18}\text{N}_4\text{O}_3\text{Si}_2$ , 226.09119 Da). All analytes and fragments in this table have been checked using authentic standards.

RT window (min)	Nominal mass (a.m.u.)	Analyte	Fragment and exact mass (Da)	Analyte	Fragment and exact mass (Da)
4.40-4.42	<b>160</b>	Pentanoic acid 1TMS	$^{13}\text{C}-\text{C}_8\text{H}_{19}\text{OSi}$ 160.12380	Glyoxylic acid 1TMS	$\text{C}_5\text{H}_{10}\text{NO}_3\text{Si}$ 160.04299
8.00-8.32	<b>174</b>	<i>O</i> -acetylserine 2TMS	$\text{C}_7\text{H}_{16}\text{NO}_2\text{Si}$ 174.09503	$\beta$ -alanine 3TMS	$\text{C}_7\text{H}_{20}\text{NSi}_2$ 174.11343
8.80-8.86	<b>189</b>	Malic acid 3TMS	$\text{C}_8\text{H}_{21}\text{OSi}_2$ 189.11309	Threitol 4TMS	$\text{C}_7\text{H}_{17}\text{O}_2\text{Si}_2$ 189.07671
5.35-5.55	<b>190</b>	2-hydroxybutyric acid 2TMS	$\text{C}_7\text{H}_{18}\text{O}_2\text{Si}_2$ 190.08453	Alanine 2TMS	$\text{C}_7\text{H}_{20}\text{NOSi}_2$ 190.10834
13.17-13.20	<b>191</b>	Caffeic acid 2TMS	$\text{C}_{10}\text{H}_{11}\text{O}_2\text{Si}$ 191.05283	Myoinositol 6TMS	$\text{C}_7\text{H}_{19}\text{O}_2\text{Si}_2$ 191.09236
9.85-10.0	<b>192</b>	Anthranilic acid 2TMS	$\text{C}_{10}\text{H}_{14}\text{NOSi}$ 192.08446	Phenylalanine 2TMS	$\text{C}_{11}\text{H}_{18}\text{Nsi}$ 192.12085
12.21-12.24	<b>203</b>	Glucosamine 6TMS	$\text{C}_8\text{H}_{21}\text{NOSi}_2$ 203.11617	Tryptophan 4TMS	$^{13}\text{C}-\text{C}_{12}\text{H}_{16}\text{NSi}$ 203.10855
7.70-7.92	<b>204</b>	Serine 3TMS	$\text{C}_8\text{H}_{22}\text{NOSi}_2$ 204.12399	<i>N</i> -formylglycine 2TMS	$\text{C}_7\text{H}_{18}\text{NO}_2\text{Si}_2$ 204.08760
9.4-9.64	<b>205</b>	Cinnamic acid 1TMS	$\text{C}_{11}\text{H}_{13}\text{O}_2\text{Si}$ 205.06848	Glutathione 4TMS	$\text{C}_{10}\text{H}_{13}\text{N}_2\text{OSi}$ 205.07971
6.91-7.13	<b>218</b>	Hydroxypyruvic acid 2TMS	$\text{C}_9\text{H}_{22}\text{O}_2\text{Si}_2$ 218.11583	Isoleucine 2TMS	$\text{C}_8\text{H}_{20}\text{NO}_2\text{Si}_2$ 218.10326
7.94-8.00	<b>218</b>	Threonine 3TMS	$\text{C}_9\text{H}_{24}\text{NOSi}_2$ 218.13964	<i>O</i> -acetylserine 2TMS	$\text{C}_8\text{H}_{20}\text{NO}_2\text{Si}_2$ 218.10326
9.37-9.39	<b>220</b>	Threonic acid 4TMS	$\text{C}_8\text{H}_{20}\text{O}_3\text{Si}_2$ 220.09510	Cysteine 3TMS	$\text{C}_8\text{H}_{22}\text{NSSi}_2$ 220.10115
11.82-11.92	<b>226</b>	Methionine sulfoximine 2TMS* 2MeOx	$\text{C}_5\text{H}_{18}\text{N}_4\text{O}_2\text{SSi}$ 226.09197	Histamine 3TMS	$\text{C}_{10}\text{H}_{22}\text{N}_2\text{Si}_2$ 226.13215
9.13-9.16	<b>230</b>	Pyroglutamic acid 2TMS	$\text{C}_9\text{H}_{20}\text{NO}_2\text{Si}_2$ 230.10326	Hydroxyproline 3TMS	$\text{C}_{10}\text{H}_{24}\text{NOSi}_2$ 230.13964
8.34-8.70	<b>247</b>	Erythrulactone 2TMS	$\text{C}_9\text{H}_{19}\text{O}_4\text{Si}_2$ 247.08219	Citramalic acid 3TMS	$\text{C}_{10}\text{H}_{23}\text{O}_3\text{Si}_2$ 247.11857
10.31-10.42	<b>261</b>	<i>N</i> -acetylmethionine 2TMS	$\text{C}_{10}\text{H}_{23}\text{NO}_3\text{Si}_2$ 261.12165	$^{13}\text{C}$ -methylglutamic acid 3TMS	$\text{C}_{11}\text{H}_{26}\text{NO}_2\text{Si}_2$ 261.15356
11.78-12.09	<b>264</b>	Adenine 2TMS	$\text{C}_{10}\text{H}_{18}\text{N}_5\text{Si}_2$ 264.11007	Tyramine 3TMS	$\text{C}_{13}\text{H}_{22}\text{NOSi}_2$ 264.12399

RT window (min)	Nominal mass (a.m.u.)	Analyte	Fragment and exact mass (Da)	Analyte	Fragment and exact mass (Da)
12.66-12.80	<b>267</b>	Urocanic acid 2TMS	C <sub>11</sub> H <sub>19</sub> N <sub>2</sub> O <sub>2</sub> Si <sub>2</sub> 267.09851	Octopamine 4TMS	C <sub>13</sub> H <sub>23</sub> O <sub>2</sub> Si <sub>2</sub> 267.12366
11.93-12.22	<b>280</b>	Pyridoxine 3TMS	C <sub>13</sub> H <sub>22</sub> NO <sub>2</sub> Si <sub>2</sub> 280.11891	Tyrosine 3TMS	C <sub>14</sub> H <sub>26</sub> NOSi <sub>2</sub> 280.15529
12.09-12.26	<b>308</b>	Ferulic acid 2TMS	C <sub>14</sub> H <sub>20</sub> O <sub>4</sub> Si <sub>2</sub> 308.09001	Coumaric acid 2TMS	C <sub>15</sub> H <sub>24</sub> O <sub>3</sub> Si <sub>2</sub> 308.12640
11.94-12.09	<b>308</b>	4-hydroxyphenyl lactic acid 3TMS	C <sub>15</sub> H <sub>24</sub> O <sub>3</sub> Si <sub>2</sub> 308.12640	Ferulic acid 2TMS	C <sub>14</sub> H <sub>20</sub> O <sub>4</sub> Si <sub>2</sub> 308.09001
12.13-12.38	<b>319</b>	Mannitol 6TMS	C <sub>13</sub> H <sub>31</sub> O <sub>3</sub> Si <sub>3</sub> 319.15810	3-indoleacetic acid 2TMS	C <sub>16</sub> H <sub>25</sub> NO <sub>2</sub> Si <sub>2</sub> 319.14238
12.09 (same RT)	<b>338</b>	Ferulic acid 2TMS	C <sub>16</sub> H <sub>26</sub> O <sub>4</sub> Si <sub>2</sub> 338.13696	Tyramine 3TMS	C <sub>16</sub> H <sub>32</sub> NOSi <sub>3</sub> 338.17917

### Hosted file

image1.emf available at <https://authorea.com/users/491574/articles/574533-exact-mass-gc-ms-analysis-protocol-database-advantages-and-application-to-plant-metabolic-profiling>

**Fig. 1. Overview of a GC-MS chromatogram from an Arabidopsis leaf extract** (here, ecotype Cvi-0). (a) total ion current chromatogram, showing annotation of typical peaks (in red). The observed mass spectrum at 15.05 min is shown in (b), with characteristic ions of 6-phosphogluconate 3TMS annotated in red. Note the contribution of ion  $m/z = 387.14254$  Da (fragment C<sub>13</sub>H<sub>32</sub>O<sub>7</sub>PSi<sub>2</sub><sup>+</sup>). Note that a close, but easily distinguishable mass (387.32912 Da, fragment C<sub>22</sub>H<sub>47</sub>O<sub>3</sub>Si<sup>+</sup>) can be observed at 15.25 min, from arachidic acid 1TMS; nominal mass analysis would have generated the same apparent mass at 387 (or 388) Da.

### Hosted file

image2.emf available at <https://authorea.com/users/491574/articles/574533-exact-mass-gc-ms-analysis-protocol-database-advantages-and-application-to-plant-metabolic-profiling>

**Fig. 2. Isotopic pattern in two N- and S-containing molecules, isopropyl-4-methylthiazole (IMT) and methionine 2TMS**. (a) overall mass spectrum of IMT, showing the prevalence of the demethylated fragment at 126.03770 Da. (b) magnification of the mass spectrum showing the isotopic pattern of the major fragment of IMT and (c) detailed view of  $m + 1$  isotopologues showing the resolution of <sup>15</sup>N, <sup>33</sup>S, <sup>13</sup>C and <sup>2</sup>H isotopologues. (d) overall spectrum of methionine 2TMS with two major fragments (128.08955 Da and 176.09290 Da) and (e) magnification showing  $m + 2$  isotopologues. (f) relationship between expected and observed relative abundance of isotopologues in IMT using two fragments (blue, green) and methionine 2TMS (pink), plotted using a log scale. Note the underestimation of deuterium (<sup>2</sup>H) content (red circle). The black line stands for the 1:1 line. In (a) and (d), fragmentations are shown with chemical formulas to illustrate how the fragments of interest are formed. Note IMT does not require derivatisation while methionine is silylated (two TMS groups). In (e), note that there is a common peak for <sup>34</sup>S (+1.995796 Da) and <sup>30</sup>Si (+1.996844 Da) and also <sup>33</sup>S (+0.999387 Da) and <sup>29</sup>Si (+0.999568 Da).

### Hosted file

image3.emf available at <https://authorea.com/users/491574/articles/574533-exact-mass-gc-ms-analysis-protocol-database-advantages-and-application-to-plant-metabolic-profiling>

**Fig. 3. Illustration of some common nominal mass  $m/z$  ions between analytes with the same retention time:** ferulic acid 2TMS and tyramine 2TMS, both at  $c$  . 12.09 min (a,b) and malic acid 3TMS

and threitol 4TMS, both at  $c \approx 8.85$  min (c,d). Comparisons of overall spectra (a,c) with nearly common masses distinguishable with exact mass (insets, in red), and origin of fragments (b,d). In (c), ions "other" refer to  $C_7H_{18}O_2Si_2^+$  (190.083983 Da) and its protonated (+1.007825 Da) species (191.091808 Da). Note that tyramine yields a much higher signal (order of magnitude  $10^9$ ) than ferulic acid ( $10^7$ ) despite the use of identical concentration in samples ( $500 \text{ ng mL}^{-1}$ ).  $m+1$  and  $m+2$  refer to +1 a.m.u. (mostly  $^{13}C$ , and  $^{29}Si$ ) isotopologues and +2 a.m.u. (mostly  $^{13}C_2$ ,  $^{30}Si$ , and  $^{13}C$ - $^{29}Si$ ) isotopologues, respectively.

### Hosted file

image4.emf available at <https://authorea.com/users/491574/articles/574533-exact-mass-gc-ms-analysis-protocol-database-advantages-and-application-to-plant-metabolic-profiling>

**Fig. 4. Mass spectrum of arginine derivatives .** (a) chromatogram showing the four products of arginine derivatisation: ornithine lactam 2TMS, ornithine 3TMS, arginine 3TMS and citrulline 3TMS. The mass spectrum of arginine 3TMS and citrulline 3TMS are shown in panels (b) and (c). Typical fragments are illustrated in (d) and (e), respectively. Note that arginine 3TMS and citrulline 3TMS have very close retention times (11.37 and 11.41 min) and thus might appear merged in the same peak under low resolute conditions.

### Hosted file

image5.emf available at <https://authorea.com/users/491574/articles/574533-exact-mass-gc-ms-analysis-protocol-database-advantages-and-application-to-plant-metabolic-profiling>

**Fig. 5. Identification of S-containing fragments using exact mass data :** appearance of  $m+2$  isotopologues at +1.995796 with a precision of 0.05 cysteine (a) and an extract from Arabidopsis leaves (b). Ions that effectively contain S are circled in red. Other ions do not contain S despite their mass difference value (w.r.t. the isotopologue at +1.995796) close to zero (see main text for calculations). In (a) and (b), red frames illustrate the chemical structure of identified S-containing fragments. (c) detailed observed isotopic pattern ( $m+2$  isotopologues) of the fragment ion with a monoisotopic mass of 74.018 Da, showing the  $^{34}S$  isotopologue (observed at +1.99549) of structure **2** in (b), as well as  $^{30}Si$  and  $^{29}S+^{13}C$  isotopologues of  $C_2H_6OSi$  (74.018791 Da), the monoisotopic form of which also contributes to the observed signal at [?] $74.018$  Da. The signal at +2.00738 likely corresponds to another fragment rather than the  $^{13}C_2$  isotopologue considering the relatively high signal. Other signals (na) are unrelated to the ion of interest. (d-f) simulation of the peak of  $m+2$   $^{34}S$  and  $^{30}Si$  isotopologues using the arbitrary example of a monoisotopic mass at 156.04700 Da containing one Si and one S, and three levels of mass resolution. The observed signal (sum) is in grey, while the separate contribution of  $^{34}S$  and  $^{30}Si$  is in blue and orange, respectively. The sampled (i.e. measured by the mass spec) mass that is closest to  $^{34}S$  is circled in red. It shows that there is no sampled mass at exactly +1.995796 ( $^{34}S$  mass excess w.r.t. monoisotopic) except at high resolution, and that  $^{34}S$  can be easily confused with  $^{30}Si$  due to their very small mass difference (0.001048 Da).

### Hosted file

image6.emf available at <https://authorea.com/users/491574/articles/574533-exact-mass-gc-ms-analysis-protocol-database-advantages-and-application-to-plant-metabolic-profiling>

**Fig. 6. Overview of exact GC-MS metabolic profiling of Arabidopsis genotypes .** (a) heatmap with hierarchical clustering showing groups of metabolites that are significantly different ( $P < 0.0002$ , one-way ANOVA) between genotypes. Numbers refers to groups named on top left. (b) Standard error between 6 biological replicates, in % of average value, in each genotype. The red dotted line stands for the median value (9.7%). The dotted frame comprises metabolites that are also framed and named in panel (c). Data are shown as whisker plots with median, 25% and 75% quartiles, and interquartile range. (c) Relationship between average standard error (in % of overall mean) and relative signal (relative to sucrose set at 1) obtained on the GC-MS instrument (quantifying target mass). Metabolites that appear highly variable are framed (and are the same as metabolites framed in panel (b)). Note that metabolite 14 is labelled "3-iodotyrosine" because it is identified with high certainty by Tracefinder® using the database although it is

not expected in plants (see main text).