Chapter 11

Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Plant Metabolite Profiling and Metabolite Identification

J. William Allwood, David Parker, Manfred Beckmann, John Draper, and Royston Goodacre

Abstract

Mass spectrometry (MS) is usually the technique of choice for metabolomic studies where the volume of sample material is too limited for applications employing nuclear magnetic resonance (NMR) spectroscopy. With the advent of ultra-high accuracy mass spectrometers such as the Orbitrap (resolution ~ 10^5) and the Fourier Transform Ion Cyclotron Resonance (FT-ICR) analysers (resolution potentially in excess of 10^6) there is the opportunity to generate an accurate mass fingerprint (often referred to as a profile since the variables are considered as effectively discrete) of an infused sample extract. In such data representations mass "peaks" are detected in the raw data and the centroid mass intensity calculated. The resolving power and sensitivity of these ultra-high accuracy mass analysers is such that metabolite signals from molecules containing naturally abundant elemental isotopes (e.g. ^{13}C, ^{41}K, ^{15}N, ^{17}O, ^{34}S, and ^{37}Cl) are visible in the data. Such is the instruments precision that it allows for the calculation of highly accurate elemental compositions for the unknown signals, thus aiding greatly in the selection of potential metabolite candidates for the annotation of unknowns prior to their confirmation by comparisons to analytical standards. The application of FT-ICR-MS to plant metabolomics has thus far been limited to a few studies and clear step-by-step methodologies are as yet unavailable. This chapter presents a rigorous method for the extraction and FT-ICR-MS analysis of plant leaf tissues as well as downstream data processing.

Key words: FT-ICR-MS, DI, FI, ESI, CID, Plant metabolomics

Abbreviations

DI	Direct infusion
FI	Flow infusion
FT	Fourier transform
ICR	Ion cyclotron resonance
MS	Mass spectrometry

Nigel W. Hardy and Robert D. Hall (eds.), *Plant Metabolomics: Methods and Protocols*, Methods in Molecular Biology, vol. 860, DOI 10.1007/978-1-61779-594-7_11,

ESI Electrospray ionisation
LTQ Linear trap quadrupole
CID Collision-induced dissociation
QC Quality control
PCA Principal components analysis
LDA Linear discriminant analysis
RF Random forest

1. Introduction

Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) in the metabolomics field is currently regarded as an instrument of great potential due to its ultra-high mass accuracy and resolution which permits unequivocal mass assignment and the resolution of ion species currently not possible with alternative mass spectrometry (MS) analysers. The fact that a range of ionisation chemistries, including electrospray ionisation (ESI), atmospheric pressure chemical ionisation (APCI), atmospheric pressure photoionisation (APPI), matrix assisted laser desorption ionisation (MALDI), electron impact and chemical ionisation (EI and CI respectively), can be applied to FT-ICR-MS adds to its value further. In FT-ICR-MS metabolomics, mass resolutions up to several hundred thousand have routinely been achieved, although for analysis over narrow mass ranges, values of greater than one million have been reported (1). High mass accuracy is also a key FT-ICR-MS feature with average errors being less than one part per million (ppm), the limits of detection in MS mode are also comparable to those of alternative MS instrumentation. Studies of complex samples such as crude oils (2) report the resolution of more than 10,000 distinct chemical species in a single FT-ICR mass spectrum without prior chromatographic separation. To discuss all of the theory that is relevant to FT-ICR-MS and how it obtains such high performance is beyond the scope of this chapter. However, at this point we refer the reader to Barrow et al. (3) for an excellent review of the background of the technique.

In one of the first studies where FT-ICR-MS profiling was applied to plant metabolomics, Aharoni et al. (4) traced over 1,000 metabolites by direct infusion (DI) of aqueous methanol and acetone extracts of strawberry fruit collected at four different stages of fruit ripening. In comparison to DIMS analysis of complex plant extracts with conventional electrospray ionisation (ESI) and time-of-flight (ToF) instrumentation (5, 6), it is clear that DI-ESI-FT-ICR-MS is capable of resolving a much greater number and range of metabolites, although with the application of high performance or ultra-high performance liquid chromatography (HPLC and UHPLC respectively) this could potentially be enhanced further

since isobaric metabolites would also be resolved. However, with the low scan speeds which are commonly employed by FT-ICR-MS (typically 1 scan/s) it must be considered that the Thermo Hybrid LTQ Orbitrap instrument is more appropriate for the prior application of UHPLC (7).

Several notable FT-ICR-MS based plant studies in recent years have emerged from Japan. Hirai et al. (8) elegantly combined transcriptomics data obtained from Affymetrix *Arabidopsis thaliana* microarrays with FT-ICR-MS metabolomics to investigate the gene-metabolite networks controlling nitrogen and sulphur metabolism. FT-ICR-MS was also applied to investigate the metabolomics behind light/dark regulation in *A. thaliana* cell cultures (9). Here, the authors developed automated data processing software (Dr. DMASS: http://kanaya.naist.jp/DrDMASS/) and an accurate mass metabolite identification system (KNApSAcK: http://Kanaya.aist-nara.ac.jp/KNApSAcK/) (9, 10).

Metabolite putative identification can be performed against KNApSAcK, PubChem (http://pubchem.ncbi.nlm.nih.gov/) or Chemspider (http://www.chemspider.com/) databases as well as many others. However, whilst some of these databases contain information on large numbers of metabolites many of these are not of natural origin. Searching for likely annotation candidates based on accurate mass information in publicly accessible databases is in itself time-consuming as individual database coverage of natural chemistry varies and so a comprehensive search requires query of several if not all relevant databases. Unfortunately, with few exceptions, databases with appropriate metabolite mass information can contain much redundancy, resources for curation are often limited (consequently it is not uncommon to find mistakes relating to mass values, molecular formulae and structure), furthermore some of the entries relate to ionic states, often from interactions with salts, which whilst useful for understanding the metabolites role in metabolism is not so relevant for MS. A further consideration is that publically accessible databases for the most part only contain information regarding the uncharged metabolite, thus to search masses of interest all possible ionisation states must be calculated by the user (taking into account the mass of electrons) and ensuring that the elemental masses used are accurate.

Work carried out at Aberystwyth University has lead to the development of FT-ICR-MS data processing methods as well as two further new resources for the analysis of ESI-MS based data. First, a data analysis package FIEmspro (http://users.aber.ac.uk/jhd) written in the R environment and requiring a moderate knowledge of R command-line usage (11) can be used for all data analysis (i.e. visualisation, normalisation, feature ranking) and thus can be used to narrow down the number of signals of interest for further interpretation. Second, a database constructed using metabolite "structures" harvested from publicly accessible databases and

converted into a common format to generate a comprehensive archive known as MZedDB (http://maltese.dbs.aber.ac.uk:8888/hrmet/search/addsearch0.php) (12). This database is based on an archive in a common format of all metabolite "structures" derived from several widely used and publicly accessible databases. Using "rules" derived from structural information and physical properties (such as number of H-bond acceptors/donors, number of OH/COOH/NH_2 groups, number of acidic H or basic O^- in molecule) MZedDB generates a list of potential adducts and neutral loss fragments that are likely to be observed for each structure and calculates on the fly the accurate mass of every potential ionisation product which provides targets for searches based on accurate mass. Starting with a list of m/z signals MZedDB supports a range of manual or semi-automated (via R environment) annotation strategies based on either m/z mass or predicted elemental composition at a range of mass resolutions (12). The database can also be used to generate elemental compositions for m/z signals using the seven golden rules (13) (see Table 1), this process can be done manually on the Web interface or in batch mode via R (code available at http://maltese.dbs.aber.ac.uk:8888/hrmet/supp/rhrmet.html) (12).

A range of feature ranking methods can be used to identify m/z signals that significantly alter ($p \leq 0.001$) between sample classes (11, 14). Multivariate analyses such as principal components analysis (PCA) (15) or linear discriminant analysis (LDA) can be used in an exploratory manner to assess the shape of the model as well as to determine errors such as mislabelled or out-lying samples (16, 17). Univariate methods such as N-Way ANOVA and non-parametric Kruskal–Wallis tests, as well as Random Forest (RF) decision trees (11, 14, 18, 19), and evolutionary computation (20, 21) are used to identify significant differential metabolite variables which are of the most interest with regard to the biological system/question under study. This can be followed by correlation analyses such as the Pearson Coefficient to indicate related signals from the sample matrix (adducts, isotopes, neutral losses) (22). The selected metabolites can then undergo a simple calculation of the accurate mass differences between individual pairs of correlated signals indicating their likely relationships. This then allows any annotation suggestions to focus on, potentially, the correct ionisation products (12, 22). For example, a correlated signal with a mass difference of 1.0033 indicates the difference between ^{12}C and ^{13}C containing signals.

As several overlapping solutions predicting the presence of different metabolites can be possible, the most likely ions putatively identifying a specific metabolite can be confirmed by MS^n experiments (1, 3, 14, 23) to confirm if the tentative (database matched) identification is correct by comparison of the fragmentation patterns of the sample analytes and those of pure analytical

Table 1
The seven golden heuristic and chemical rules for the selection of accurate and correct elemental compositions

Rule	Description
Rule 1: Restriction of elements	Natural compounds contain restricted numbers of each element, thus by dividing the mass range by the element mass allows sensible ranges of atoms for that specific element to be predicted, i.e. C has a mass of 12 Da, data were collected over the mass range 1–1,000 Da 1,000/12 = 83, 83 C atoms is the maximum expected for a mass of 1,000 Da Heuristic filtering based upon information for the numbers of atom present for each element within compounds that are found commonly in the PubChem, Wiley, NIST02, and DNP databases, was used to reduce the predicted numbers of atoms further. Based upon database information, maximum element ratios can also be applied to heuristic filtering, e.g. for 47 C the maximum H is 150.
Rule 2: Lewis and Senior check	The LEWIS rule in simple terms demands that a molecule consisting of simple elements (C, H, N, O, especially) share electrons so that the *s*, *p*-valence shells are filled completely, i.e. the "octet" rule. However, the LEWIS rule excludes all nitroso compounds and so is combined with the, SENIOR rule that requires three essential conditions for the existence of an elemental composition: (a) The sum of valences or total number of odd number atoms is equal (b) The sum of valences is even to or greater than double the maximum valence (c) The sum of valences is even to or greater than double the total atom number – 1
Rule 3: Isotopic pattern filter	Natural compounds comprise monoisotopic and isotopic masses according to the natural average abundance of stable isotope abundances for each element. For MS instruments with low relative errors of 2–5% RSD and assuming high-quality data with a good signal-to-noise ratio and accurate detection of the M + 1 and M + 2 isotope ions, inclusion of the calculation for isotope ratio abundance permits the removal of the majority of incorrectly assigned elemental compositions. Of all seven golden rules this is the most important for the removal of incorrect elemental compositions.
Rule 4: H/C element ratio check	By including element ratio constraints to the heuristic filtering (especially for H/C), the calculated elemental compositions are further restricted to the most probable candidates. For most natural molecules, the H/C ratio is rarely greater than 3 or less than 0.125 and by applying a filtering range of 0.2 > 3.1 the majority of drug and natural compounds can be filtered for. However, in extreme cases such as fluorines, when the experimenter expects to find such compounds, the range needs to be extended for them to be fully accounted for.
Rule 5: Heteroatom ratio check	Many formulas, alkanes for one example, comprise no heteroatom. Cases of high ratios of heteroatom to carbon number are extremely rare, thus a simple exclusion of very high heteroatom ratio elemental compositions helps to further remove unlikely candidates.

(continued)

Table 1 (continued)

Rule	Description
Rule 6: Element probability check	Based upon the NIST02, Wiley, and DNP database searches and element combinations of N, O, P and N, O, S, with C and H, a high number of entries are found which have high element ratios. From this information specific thresholds for the numbers of atoms for each element can be accordingly defined.
Rule 7: TMS check	TMS derivatisation is commonly performed in GC-MS analyses in order to enhance volatility and permit the detection of otherwise undetectable compounds. To calculate elemental compositions of neutral masses, the replacement of acidic H^+ with TMS groups must be accounted for in order to calculate the non-derivatised molecules mass. The number of TMS groups is easily deduced via the calculation of isotopic abundances. The TMS check also mandates that for each Silicon there has to be three methyl groups.

Kind and Fiehn (13) developed an algorithm based upon seven heuristic and chemical rule-based filters for the accurate selection of the correct elemental formula from the hundreds that may be generated for any one given accurate mass. For liquid chromatography (LC) data, adducts must first be identified and removed, thus giving a list of neutral ions alone. Likewise, for gas chromatography (GC) data, products of derivatisation must be identified and the original neutral ion calculated. Elemental compositions are then generated for the accurate masses of each neutral ion. The algorithm performs at its best providing that the elemental compositions are based upon high resolution and mass accuracy data from instruments such as FT-ICR-MS and Thermo hybrid LTQ Orbitrap system (i.e. within 3 ppm mass accuracy and resolution of 100,000 >) for molecules which are purely resolved with either liquid chromatography, gas chromatography, or capillary electrophoresis. The seven golden rules are explained in the following table, when applied to the elemental compositions generated for 6,000 database entries, the seven golden rule algorithm selected the correct elemental composition as the top hit with an 80–99% probability rate. Adapted from ref. 13
Abbreviations: *DNP* Dictionary of Natural Products, *NIST02* National Institute of Standards and Technology 2002 MS library, *TMS* Trimethylsilyl

standards. Since FT-ICR-MS is a "trapping" instrument, multiple stages of analysis (MS^n) can be undertaken, such as fragmentation of an ion selected from a mixture, followed by further fragmentation of the product daughter fragment ions (as required for unambiguous confirmation of metabolite identification along with accurate mass measurement of the parent ion, and comparison to an analytical standard). The most commonly employed form of MS^n uses collision-induced dissociation (CID), although alternative methods that could potentially be applied to metabolite analysis include infrared multiphoton dissociation (IRMPD) and electron capture dissociation (ECD), although the latter is more commonly employed for the analysis of positively charged peptides (1, 3) since the target analyte must be doubly or higher charged. The method presented within this chapter will employ CID (24) as this is the most commonly used.

2. Materials

2.1. Harvest of Plant Material

1. Clean stainless steel scissors (sharp), forceps, and spatulas of appropriate size for sample material (see Note 1).
2. Liquid nitrogen, a 1–2 L Dilvac (Day-Impex, Colchester, Essex, UK) and long-arm forceps to retrieve tubes from the liquid nitrogen (see Note 2).
3. Pre-labelled (alcohol resistant marker pen) high-quality 2-mL polypropylene microcentrifuge tubes and/or 15- or 50-mL polypropylene falcon tubes (see Notes 3 and 4) (Greiner Bio One, Stonehouse, Gloucestershire, UK) depending upon volume of sample material.
4. Stainless steel 5-mm ball bearings (Retsch, Hunslet, Leeds, UK) cleaned in methanol and air-dried three times, placed in pre-labelled 2-mL microcentrifuge tubes (see Notes 3 and 4) (Greiner Bio One, UK).
5. Denver Instrument Balance—Summit SI-234 (Denver, Colorado, USA), or similar.
6. Appropriate freezer boxes suitable for long-term –80°C storage of samples.

2.2. Extraction for the Capture of Polar Metabolites and Chloroform Purification of Non-polar Metabolites

1. Ice and insulated ice box (see Note 5).
2. Liquid nitrogen, a 1–2 L Dilvac (Day-Impex, UK) and long-arm forceps (see Note 2).
3. Retsch MM200 ball mill and two 5 or 10 position microcentrifuge tube adapters (Retsch, UK).
4. Eppendorf Concentrator 5301 at 30°C and setting 1 (Eppendorf UK Ltd., Histon Cambridge, UK).
5. Pre-labelled (alcohol resistant marker pen) high-quality 2-mL polypropylene microcentrifuge tubes (Greiner Bio One, UK), two sets should be prepared for storage of the final extracts and one set for preparation of the extracts (see Note 3).
6. High-quality methanol (trace analysis grade), water (ultrapure), and chloroform (HPLC grade or better) (Mallinckrodt-J.T. Baker, Leadenhall Street, London, UK).
7. Prepare a mixture of 100 mL chloroform–250 mL methanol–100 mL water using a solvent washed (see Note 1) measuring cylinder and storage bottle fitted with a PTFE lined lid. Prepare and store at –20°C for 24 h minimum prior to extraction (see Note 6).
8. High-quality P1000 and P200 polypropylene pipette tips (Greiner Bio One, UK) (see Note 3).
9. Appropriate freezer boxes suitable for long-term, –80°C storage of samples.

2.3. Preparation of Metabolite Standards

1. High-quality methanol or isopropanol (trace analysis grade) and water (ultra-pure) (Mallinckrodt-J.T. Baker, UK).
2. Prepare 70% Aqueous methanol and 50% Aqueous isopropanol (see Note 1).
3. Substance P (Thermo Corp., DE) for calibration (see Note 7).
4. Acetaminophen ($[M+H]^+$ 152.0712, $[M-H]^-$ 150.0555) (Sigma-Aldrich Ltd., Gillingham, Dorset, UK).
5. Caffeine ($[M+H]^+$ 195.0882, $[M-H]^-$ 193.0726) (Sigma-Aldrich Ltd., UK).
6. Sulfaguanidine ($[M+H]^+$ 215.0603, $[M-H]^-$ 213.0446) (Sigma-Aldrich Ltd., UK).
7. Sulfamethoxine ($[M+H]^+$ 311.0814, $[M-H]^-$ 309.0658) (Sigma-Aldrich Ltd., UK).
8. Valine-Tyrosine-Valine ($[M+H]^+$ 380.2125, $[M-H]^-$ 378.2029) (Sigma-Aldrich Ltd., UK).
9. Terfenadine ($[M+H]^+$ 472.3216, $[M-H]^-$ 470.3059) (Sigma-Aldrich Ltd., UK).
10. Reserpine ($[M+H]^+$ 609.2812, $[M-H]^-$ 607.2656) (Sigma-Aldrich Ltd., UK) (see Note 7).
11. Erythromycine ($[M+H]^+$ 734.4691, $[M-H]^-$ 732.4534) (Sigma-Aldrich Ltd., UK).

2.4. Preparation of Samples and Addition of Internal Standards

1. 70% Aqueous methanol and 50% Aqueous isopropanol prepared from trace analysis grade solvents and ultra-pure water (Mallinckrodt-J.T. Baker, UK) (see Note 1).
2. Minisart RC4 single-use syringe filter non-sterile, regenerated cellulose membrane, polypropylene housing, pore size 0.20 μm (Sartorius, Goettingen, DE) (see Note 3).
3. Disposable polypropylene 1 mL syringes (Becton Dickinson, Oxford, UK) (see Note 3).
4. High quality P1000 and P200 polypropylene pipette tips (Greiner Bio One, UK) (see Note 3).

2.5. Instrumentation and Analysis

1. Standard multi-well plate for nanospray ESI or borosilicate glass mass spectrometry vials appropriate for the autosampler being employed for standard ESI.
2. TriVersa™-NanoMate chip technology (Advion Biosystems, NY, USA) coupled to an LTQ-FT™ ICR mass spectrometer (Thermo Corp, DE) (see Fig. 1).
3. Thermo Xcalibur version 2.0 (Thermo Corp, DE).

2.6. Data Processing and Statistical Analysis

1. MatLab R2008a (The Mathworks Inc., Natick, MA, USA).
2. R environment using the FIEMSpro metabolomics data analysis package (11, 23, 25) Web accessible (http://users.aber.ac.uk/jhd).

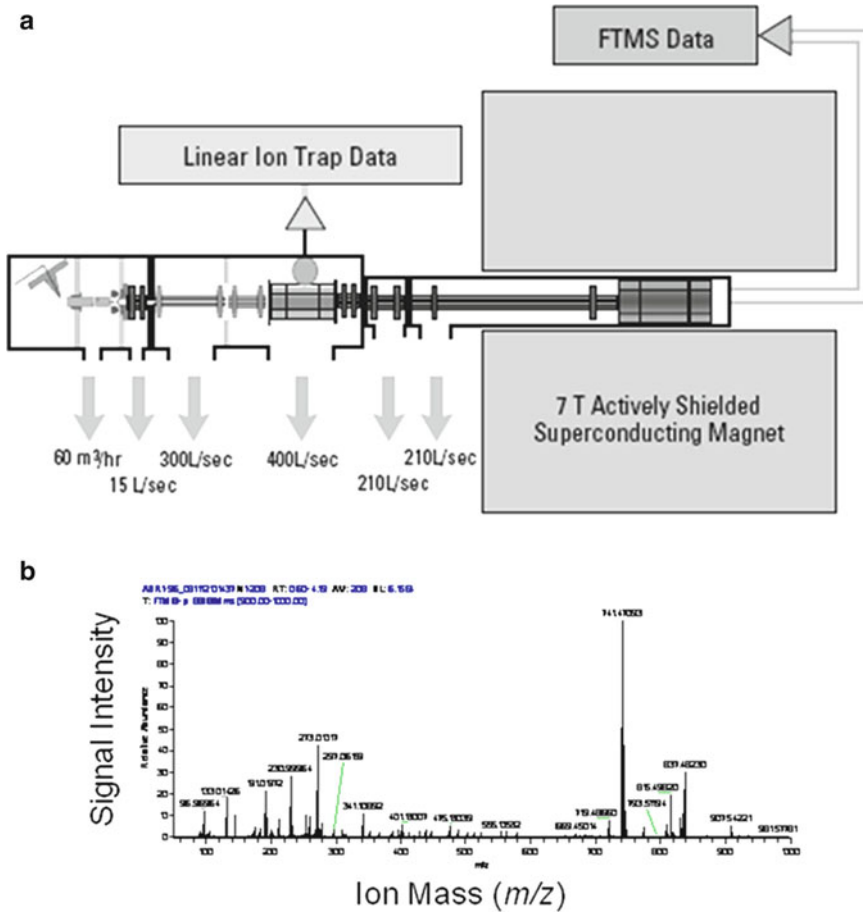

Fig. 1. FT-ICR-MS schematic and example FT-ICR-MS profile. (**a**) Diagram of the Thermo LTQ-FT-MS system (Reproduced with thanks to Thermo Fisher Scientific). (**b**) An example Nano-infusion FT-ICR-MS fingerprint of a polar extract taken from *Brachypodium distachyon* leaf tissue. The sample preparation and mass spectral acquisition was performed as presented in the methods within this chapter.

3. Determining the mathematical relationships between m/z and automated database searches are performed in R using the code described at http://maltese.dbs.aber.ac.uk:8888/hrmet/supp/rhrmet.html (12).

3. Methods

3.1. Harvest of Plant Material

1. Plant material should be rapidly excised using clean sharp scissors whilst maintaining that there are no soil particles coating the material and that contact is not made between the plant material and laboratory gloves (see Note 8).

2. Rapidly transfer the material (100 mg ±2 mg) with clean forceps into 2-mL microcentrifuge tubes each containing a single 5-mm stainless steel ball bearing (cleaned three times in methanol). Alternatively, for larger sample material directly grind in liquid nitrogen with a pestle and pre-cooled mortar and weigh the still-frozen powder (100 mg ±2 mg) into 2-mL microcentrifuge tubes.
3. Once weighed, the samples should again be plunged into liquid nitrogen prior to −80°C storage until extraction.

3.2. Extraction for the Capture of Polar Metabolites and Chloroform Purification of Non-polar Metabolites

The following extraction procedure was originally devised by Fiehn et al. (26) and updated by Lisec et al. (27). It was designed for GC-MS analyses and has been successfully applied to each of the META-PHOR target species of melon, broccoli, and rice (28, 29) but in our experience is equally applicable to direct infusion mass spectrometry with ESI for the analysis of polar (5) and non-polar metabolites (6) from the leaf material of *Arabidopsis thaliana* and *Brachypodium distachyon*. It is important to be well organised in advance of starting the procedure and to work quickly and precisely throughout using 1,000 and 200 μL pipettes (see Note 8). It must be taken into consideration that analysis of a single sample provides only a single metabolic snapshot without further information on biological variation or analytical errors. To estimate such variations, sufficient biological replicates and sufficient technical replicates must be prepared and analysed. If excess material is available then excess samples should also be prepared to allow for optimisation of reconstitution solvents and their final volume, as well as instrument conditions and to assess analytical and technical errors.

1. Samples should be removed from −80°C storage and flash frozen in liquid N_2, non-ground samples are homogenised using a Retsch MM200 ball mill set on a frequency of 30 Hz for 1 min, and placed on ice.
2. To each sample 1 mL of −20°C extraction solvent, chloroform–methanol–water (1:2.5:1), is added and the sample placed back on ice.
3. The samples are then mixed on a vortex and vigorously shaken in a cold room at 3°C for 15 min and returned back onto ice.
4. The samples are then centrifuged at 3°C and 14,500 × *g* for 3 min with a microcentrifuge, after which the supernatants are decanted to clean labelled 15-mL falcon tubes and kept on ice.
5. Repeat steps 2–4 on the same sample pellet, thus extracting each sample twice.
6. To 2 mL of the clean combined sample supernatants, 1 mL of ultra-pure water is added and the samples are then mixed with a vortex and centrifuged at 3°C and 14,500 × *g* for 3 min with a desktop centrifuge to aid solvent phase separation.

7. The polar phase is recovered (carefully avoiding the interphase) as 250–500 μL aliquots (depending upon sample concentration) into clean labelled 2-mL microcentrifuge tubes, approximately 200 μL of non-polar phase can also be recovered to a clean labelled 2-mL microcentrifuge tube.
8. The polar and non-polar samples are then dried via speed vacuum concentration in an Eppendorf Concentrator 5301, on setting 1, for 8 h and stored at −80°C prior to analysis. Alternatively, if the samples are for immediate analysis the extract can be directly injected into the mass spectrometry system (see Note 9).

3.3. Preparation of Metabolite Standards

For tuning the FT-ICR-MS across a suitable mass range for the analysis of polar phase plant extracts, a cocktail of analytical standards containing a final concentration of 100 μM of each standard (all of a minimum 99% purity) should be prepared. Standards should be weighed precisely on an accurate balance, when possible standards should be dissolved and diluted in 70% [v/v] aqueous methanol or 50% [v/v] aqueous isopropanol; on occasion standards may first require a pure non aqueous solvent to dissolve completely prior to dilution with aqueous solvents. Just prior to FT-ICR-MS tuning, further dilute the cocktail 1:10 with 70% [v/v] aqueous methanol or 50% [v/v] aqueous isopropanol (depending on the initial dissolvent).

The range of standards used should be appropriate to the mass range of metabolites present within the sample. The standards should also be of relevance to the plant biology of interest, i.e. if you wish to study glucosinolates then also use relevant glucosinolate standards within the calibration cocktails. This is of importance since, due to ESI suppression effects, pure compounds or compounds present in simple mixtures may respond differently to ESI than when present in a complex matrix such as a plant extract.

3.4. Preparation of Samples and Addition of Internal Standards

1. In our experience, lyophilised polar and non-polar samples are best reconstituted in 200 μL methanol (trace analysis grade)–water (ultra-pure) (70:30, [v/v]) for ESI applications.
2. Prior to analysis, reconstituted samples are sonicated for 15 min and either centrifuged at 0°C for 4 min at 14,000 × *g* (12) or may be filtered using Minisart RC4 syringe filters.
3. Prepare also an extraction blank in a clean 2-mL microcentrifuge tube which is also subjected to the above centrifugation or filtration steps. This sample permits the removal of mass signals which originate from plasticides within the pipette tips, microcentrifuge tubes, syringe, and filters (see Note 3).
4. The samples are then randomised and directly transferred into borosilicate glass mass spectrometry vials (200 μL) or multi-well plates (20 μL) (see Note 10) suitable for the auto-sampler

being employed. The remaining sample is stored in a liquid state at –80°C, and for long-term storage the vials are topped off with argon or nitrogen.

5. Prior to analysis, prepare also a representative sample pool containing an equal volume of every biological sample (~200 µL total volume) to serve as a quality control (QC). Aliquot 90 µL of QC into a clean 2-mL microcentrifuge tube and add 10 µL of the 100 µM calibration cocktail, this will provide an assessment of how the sample matrix effects the analytical standards FT-ICR-MS detection when compared to the cocktail of analytical standards in solution. The QC also provides a data quality check for the true experimental samples. The QC sample should be included after every tenth biological sample within the analytical run sequence.

3.5. Instrument Set up, Tuning and Calibration for FT-ICR-MS Sample Profiling and MS^n

For reasons of clarity, the described protocol focuses on the use of a single instrument, the Thermo-Finnigan LTQ fitted with a 7-Telsa FT-ICR mass analyser (Thermo-Finnigan, DE; Fig. 1), for sample profiling. If required, multiple MS/MS (MS^n) experiments are possible to follow up secondary ionisations of either the most abundant or predefined mass ions. Generally speaking, increases in mass resolution are concomitant with a proportional increase in data dimensionality, which in turn effects experimental design with regards to the numbers of replicates required to achieve statistical robustness (11, 30). A workflow from FT-ICR-MS analysis through to data processing, statistics, and metabolite assignments is available for reference (see Fig. 2).

1. Before starting the analytical run sequence, ensure that the LTQ instrument is fully operational according to the manufacturer's recommended instrumental conditions and performance (see Note 11). Also using a single representative sample (the QC being ideal), you must check that its concentration is optimal for FI-ESI-FT-ICR-MS analysis.
2. Place the extracted samples as described above into the autosampler (see Note 12). The tray holder is maintained at 5°C (31). An equivalent method for standard flow ESI is described by Beckmann et al. (25).
3. Typical nanospray conditions comprise 200 nL/min flow rate, 0.5 psi back pressure, and +1.6 kV (positive ion data) or –1.6 kV (negative ion data) electrospray voltage, controlled by Chipsoft software (Advion Biosystems, USA). Prior to starting a run sequence of polar plant extracts maintain that the nanospray is stable for at least 3 min. FT-ICR-MS parameters include an automatic control gain setting of 1×10^5 and a mass resolution of 100,000 (defined at m/z 400). Data is recorded for 5 min per replicate infusion using the Xcalibur software (Thermo Corp., DE) (12, 25).

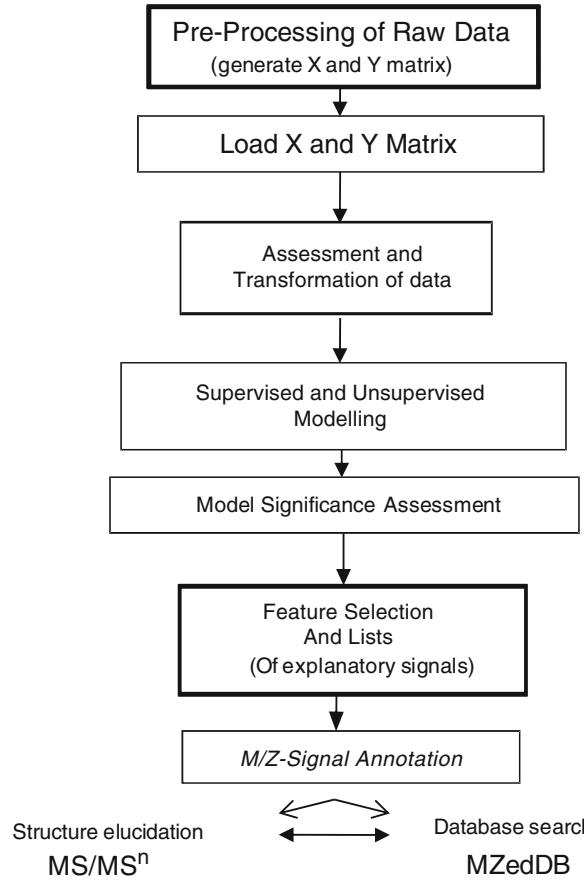

Fig. 2. Overall workflow for metabolic profiling using FT-ICR-MS. Overview of the major components of data analysis starting with raw-data conversion and first-pass data analysis, followed by data mining and finally annotation and database searches. Adapted from ref. 11.

4. To alleviate the loss of low mass ions when analysing wide mass ranges, ions are transferred from the linear ion trap to the ICR detector for full scans by segmenting the total m/z spectral range into an optimal number of smaller windows. This helps to minimise the loss of low mass ions due to time-of-flight effects. Mass resolution is fixed at 100,000 (defined for an ion at m/z 400) throughout. Automatic gain control (AGC) is set to correspond to the number of charges transferred from the front-stage ion trap to the ICR detector cell.
5. The data acquisition method set to acquire data in positive or negative ionisation mode is as follows: run time 5 min; one segment was used; number of scan events 17 (0.25 min per event), scan rate "normal" (1 scan/s), scan type "full," and data type centroid. The SIM window scan events are set as follows: scan event 1: positive polarity, mass range from m/z 50 to m/z

120; scan event 2: positive polarity, mass range from m/z 100 to m/z 200; scan event 3: positive polarity, mass range from m/z 180 to m/z 280; and so on until the mass range 50–1,400 m/z is covered (see Table 2). The number of events can be customised to meet the objective m/z range of the study. Each scan event is 0.25 min with the first scan event longer to incorporate a 0.75 min delay to allow the system time to normalise. The scan event acquisition time can be increased to allow acquisition of more scans per SIM window. For negative mode the same method is used only changing the polarity. Prior to any statistical analysis the data is log transformed to reduce the chance of high-intensity peaks dominating in the multivariate data analyses.

Table 2
FT-ICR-MS SIM window data acquisition method for polar plant leaf extracts

SIM window segment	Duration (min)	Start scan *(m/z)*	End scan *(m/z)*	Acquisition time (min)
1	1	50	120	1
2	0.25	100	200	1.25
3	0.25	180	280	1.5
4	0.25	260	360	1.75
5	0.25	340	440	2
6	0.25	420	520	2.25
7	0.25	500	600	2.5
8	0.25	580	680	2.75
9	0.25	660	760	3
10	0.25	740	840	3.25
11	0.25	820	920	3.5
12	0.25	900	1,000	3.75
13	0.25	980	1,080	4
14	0.25	1,060	1,160	4.25
15	0.25	1,140	1,240	4.5
16	0.25	1,220	1,320	4.75
17	0.25	1,300	1,400	5

In order for FT-ICR-MS to maintain high mass accuracy across a large mass range, especially with regard to metabolites of low m/z, SIM window methodologies are employed. The table presents clearly the recommended SIM window methodology for the analysis of polar extracts of plant leaf material

6. Run blank samples comprising extraction solvents, calibration cocktail, and QC samples (mix of all plant samples), interspersed at random into the run sequence, to monitor instrument performance and detect system peaks. The mass spectral response of the analytical standards within the calibration cocktail should also be compared to their response within the complex QC sample matrix to check for variation in the reported molecular weight (analyte *m/z* value) as well as to monitor ESI suppression effects and differential ionisation efficiencies.
7. Accurate mass measurements are performed in the FT-ICR-MS by the collection of 30 mass spectra and averaging the masses acquired over these scans. An initial scan window of 70 Da (50–120 *m/z*) is acquired and followed by scan windows of 100 Da with a 20-Da overlap (front and back) between windows across the mass range 100–1,400 *m/z* (see Table 2).
8. MS*n* data can be recorded throughout the profiling analysis for either the most abundant or predefined *m/z* targets, or alternatively, target *m/z's* selected by multivariate analysis of the profiling data can be analysed at a later stage. MS*n* is achieved via first isolating the target *m/z* and applying CID within the LTQ; collect 30 mass spectral scans and sum the data for each target analyte (25).
9. For XY-matrix generation, subsequent data mining and MZedDB searches, infusion data acquired in profile mode are obtained as processed mass spectra with associated peak lists (Xcalibur, Thermo Corp., DE) and exported as exact mass text files (see Note 13). Accurate mass alignment of all mass spectra, "peak-picking", integration, and centroiding of mass signals is performed in Matlab (11, 12, 25). Another data acquisition and XY-matrix generation strategy providing maximum *m/z* accuracy uses the stitching of transient files (i.e. scans recorded in the time domain), customised mass calibration using known *m/z*-ions for each SIM window and XY matrix processing in custom-written MATLAB software (31–33).

3.6. Data Analysis

1. Data within each biological XY matrix class are aligned and any peaks not represented in 70% of class replicates should be removed from the matrix.
2. Carry out all statistical tests in the R environment using the FIEMSpro metabolomics data analysis package (11) which is Web accessible (http://users.aber.ac.uk/jhd).
3. Perform explanatory feature selection using RF decision trees (11, 34, 35).
4. Perform signal correlation analysis by the Pearson correlation method on the explanatory *m/z* obtained by the feature selection methods such as RF, ANOVA, and non-parametric Kruskal–Wallis (11, 22).

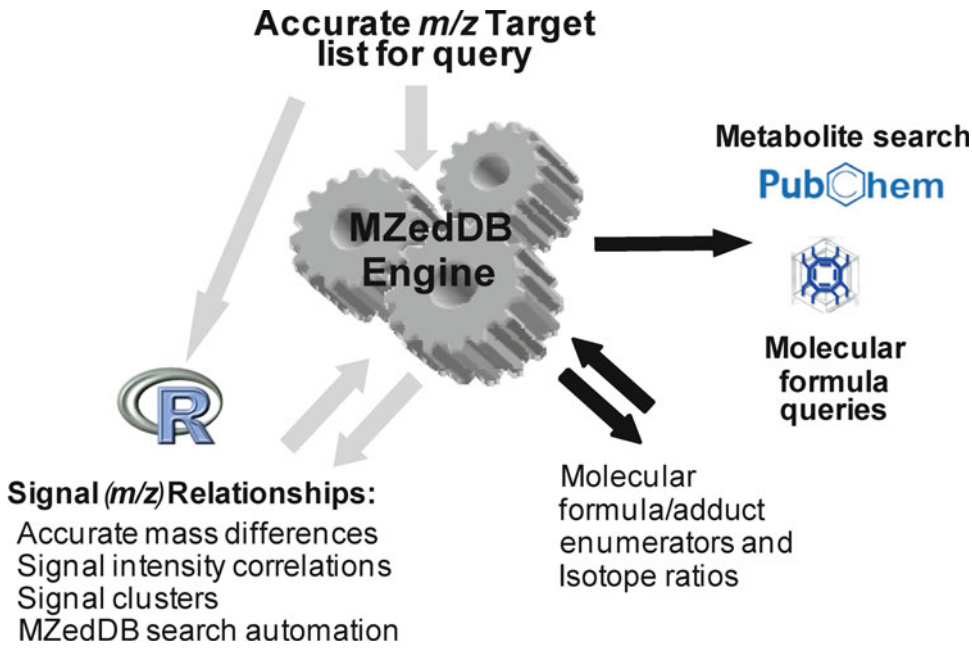

Fig. 3. MZedDB Web-resource workflow. MZedDB architecture for accurate *m/z* searches. *Grey arrows* represent MZedDB's general functionalities and *black arrows* indicate some common query pathways. Adapted from ref. 12.

5. Employ hierarchical cluster analysis based on the correlation coefficient to identify the set of clusters, which satisfy some setting, for example, signal correlation coefficient larger than 0.75.
6. Determine the mathematical relationships between m/z in R (see Fig. 3) using the code described at http://maltese.dbs.aber.ac.uk:8888/hrmet/supp/askMZedDBworkflow.r (12). This code searches for operator predetermined mass differences between measured accurate masses at an adjustable sensitivity (see Note 14).

4. Notes

1. Glassware such as bottles for the storage of extraction solvents and measuring cylinders for their preparation, as well as the scissors, forceps, and spatulas used to prepare and weigh samples and standards must be very clean. In our experience, washing and repeat washing clean glassware and metal ware with polar solvents such as methanol, ethanol, propan-2-ol, and acetonitrile, as well as non-polar solvents such as chloroform, prior to rinsing several times with HPLC grade water, oven drying, and capping with kitchen foil, helps to prevent sample contamination.

2. Liquid nitrogen requires careful handling. Please refer to your organisation's guidelines on safety for its use.
3. It is best to use high-quality polypropylene plastic ware (microcentrifuge and falcon tubes as well as pipette tips, syringes, and syringe filters) from a reputable supplier (e.g. Eppendorf, Greiner or Sarstedt), since this helps to reduce the range of plasticides that are introduced to the sample extracts. Alternatively, disposable borosilicate glass tubes may be used for extraction and borosilicate glass MS vials for storage and concentration of extracts. The mass signals of plasticides frequently mask the signals from metabolites of interest. As recommended in the methods an extraction blank should be prepared identically to the plant material samples, this sample can be used to account for mass signals introduced through sample preparation.
4. Do not label the lids of the 2-mL microcentrifuge tubes used for milling plant material since the lids can crack and transfer ink into the sample. If a lid does crack, provided that no plastic enters the sample material, remove the lid with scissors and replace with a lid removed from a clean microcentrifuge tube.
5. Perspex ice boxes are the best, but polystyrene can be used.
6. This solution can be stored at –20°C and used for up to 1 month after preparation.
7. Calibration is undertaken following procedures set out in the Thermo FT-MS handbook. However, please note that Reserpine and substance P do not calibrate into a low enough mass range for metabolite applications. To calibrate into a suitable low mass range, researchers should select appropriate low molecular weight standards, for a list of recommended compounds for online calibration of ESI-based instrumentation refer to Subheading 2.3. It must also be considered that deuterated internal standards could be added to samples or alternatively known metabolites within the samples can be used as lock-mass for off-line calibration (32, 33).
8. It is important to work on a sample-by-sample basis as rapidly and precisely as possible. When harvesting plant material and undertaking sample extraction, it is best to be well organised and to work quickly but precisely. The scissors, forceps, and spatulas used in the sample harvest procedure must be rinsed in HPLC grade water and dried between the collections of each sample. It is easy to underestimate the importance of this, but technical variance is frequently seen as being greater than the analytical variance of instruments such as FT-ICR-MS.
9. Polar samples may be too dilute to be amenable to the detection of minor metabolites of low concentration and may therefore require some form of concentration prior to injection.
10. For reduced contamination from multi-well plates buy pre-washed plates.

11. Re-calibrate the system if not performed within the previous 4 days before analysis.
12. The described method utilises an Advion Nanomate chip-based direct infusion nanospray ionisation system to introduce the sample. Chip nozzles block very easily, so ensure that samples are filtered or spun down, and are free of precipitates.
13. Check sample file size as an indicator that the sample ran correctly (i.e. spray current was stable throughout the run), e.g. if three replicate injections have file sizes of around 800 Mb and one has a file size of around 700 Mb, then something is not right with the fourth sample. Additionally, check occurrence of different levels of total ion count (TIC) for all SIM windows in the Xcalibur chromatogram view: a failed analysis shows TIC of near zero in combination with missing *m/z* signals in spectrum view especially at longer infusion times and should be removed for XY-matrix generation.
14. In theory, any mass difference can be searched for providing the operator knows the exact expected mass difference between the measured masses (see Fig. 3). This process is important in indicating possible isotope signals present in the matrix for which a prediction would not be wanted, and as an indication of ionisation products within the matrix.

Acknowledgements

JWA and RG would like to acknowledge the EU Frame work VI initiative for research funding and support as part of the plant metabolomics project META-PHOR (FOOD-CT-2006-036220). RG is also grateful to the UK BBSRC for financial support of the MCISB (Manchester Centre for Integrative Systems Biology). DP, JD, and MB would like to acknowledge research support received from Aberystwyth University and UK BBSRC grant BB/D006953/1; MB is further supported by a Research Councils UK Fellowship.

References

1. Brown, S.C., Kruppa, G., Dasseux, J.-L. (2005) Metabolomics applications of FT-ICR mass spectrometry. *Mass Spec. Rev.* **24**, 223–231.
2. Hughey, C.A., Rodgers, R.P., Marshall, A.G. (2002) Resolution of 11,000 compositionally distinct components in a single electrospray ionization Fourier transform ion cyclotron resonance mass spectrum of crude oil. *Anal. Chem.* **74**, 4145–4149.
3. Barrow, M.P., Burkitt, W.I., Derrick, P.J. (2005) Principles of Fourier transform ion cyclotron mass spectrometry and its application in structural biology. *The Analyst* **130**, 18–28.
4. Aharoni, A., De Vos, C.H.R., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R., Goodenowe, D.B. (2002) Nontargeted Metabolome Analysis by Use of Fourier Transform Ion Cyclotron Mass Spectrometry. *Omics* **6**, 217–234.
5. Parker, D., Beckmann, M., Enot, D.P., Overy, D.P., Caracuel Rios, Z., Gilbert, M., Talbot, N., Draper, D. (2008) Rice blast infection of

Brachypodium distachyon as a model system to study dynamic host pathogen interactions. *Naure. Prot.* **3**, 435–445.

6. Allwood, J.W., Ellis, D.I., Heald, J.K., Goodacre, R., Mur, L.A.J. (2006) Metabolomic approaches reveal that phosphatidic and phosphatidyl glycerol phospholipids are major discriminatory non-polar metabolites in responses by *Brachypodium distachyon* to challenge by *Magnaporthe grisea*. *The Plant J* **46**, 351–368.
7. Koulman, A., Woffendin, G., Narayana, V.K., Welchman, H., Crone, C., Volmer, D.A. (2009) High-resolution extracted ion chromatography, a new tool for metabolomics and lipidomics using a second-generation orbitrap mass spectrometer. *Rapid Communications in Mass Spectr.* **23**, 1411 – 1418.
8. Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T., Saito, K. (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *PNAS USA* **101**, 10205–10210.
9. Nakamura, Y., Kimura, A., Saga, H., Oikawa, A., Shinbo, Y., Kai, K., Sakurai, N., Suzuki, H., Kitayama, M., Shibata, D., Kanaya, S., Ohta, D. (2007) Differential metabolomics unravelling light/dark regulation of metabolic activities in *Arabidopsis* cell cultures. *Planta* **227**, 57–66.
10. Ohta, D., Shibata, D., Kanaya, S. (2007) Metabolic profiling using Fourier-transform ion-cyclotron-resonance mass spectrometry. *Anal. Bioanal. Chem.* **389**, 1469–1475.
11. Enot, D.P., Lin, W., Beckmann, M., Parker, D., Overy, D.P., Draper, J. (2008) Preprocessing, classification modelling and feature selection using flow injection electrospray mass spectrometry metabolite fingerprint data. *Nature Prot.* **3**, 446–470.
12. Draper, J., Enot, D.P., Parker, D., Beckmann, M., Snowdon, S., Lin, W., Zubair, H. (2009) Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive m/z annotation tool utilising predicted ionisation behaviour 'rules'. *BMC Bioinformatics* **10**, 227.
13. Kind, T. and Fiehn, O. (2007) Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* **8**, 105.
14. Enot, D.P., Beckmann, M., Draper, J. (2007) Detecting a difference – assessing generalisability when modelling metabolome fingerprint data in longer term studies of genetically modified plants. *Metabolomics* **3**, 335–347.
15. Jolliffe (1986) *Principle Components Analysis.* Springer-Verlag, New York.
16. Goodacre, R. (2007) Metabolomics of a superorganism. *J. Nutrition* **137**, 259 S–266 S.
17. Goodacre, R. Vaidyanathan, S., Dunn, W.B., Harrigan, G.G., Kell, D.B. (2004) Metabolomics by numbers – acquiring and understanding global metabolite data. *Trends Biotech.* **22**, 245–252.
18. Enot, D.P., Beckmann, M., Overy, D., Draper, J. (2006) Predicting interpretability of metabolome models based on behavior, putative identity, and biological relevance of explanatory signals. *PNAS USA* **103,** 14865–14870.
19. Enot, D.P. and Draper, J. (2007) Statistical measures for validating plant genotype similarity assessments following multivariate analysis of metabolome fingerprint data. *Metabolomics* **3**, 349–355.
20. Goodacre, R., York, E.V., Heald, J.K., Scott, I.M. (2003) Chemometric discrimination of unfractionated plant extracts profiled by flow-injection electrospray mass spectrometry. *Phytochem.* **62**, 859–863.
21. Johnson, H.E., Broadhurst, D., Goodacre, R., Smith, A.R. (2003) Metabolic fingerprinting in salt-stressed tomatoes. *Phytochem.* **62**, 919–928.
22. Brown, M., Dunn, W.B., Dobson, P., Patel, Y., Winder, C.L., Francis-McIntyre, S., Begley, P., Carroll, K., Broadhurst, D., Tseng, A., Swainston, N., Spasic, I., Goodacre, R., Kell, D.B. (2009) Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *The Analyst* **134**, 1322–1332.
23. Overy, D.P., Enot, D.P., Tailliart, K., Jenkins, H., Parker, D., Beckmann, M., Draper, J. (2008) Explanatory signal interpretation and metabolite identification strategies for nominal mass FIE-MS metabolite fingerprints. *Nature Prot.* **3**, 471–485.
24. Laskin, J. and Futrell, J.H. (2005) Activation of large ions in FT-ICR mass spectrometry. *Mass Spec. Rev.* **24**, 135–167.
25. Beckmann, M., Parker, D., Enot, D.P., Duval, E., Draper, J. (2008) High-throughput metabolome fingerprinting using Flow Injection Electrospray Mass Spectrometry. *Nature Prot.* **3**, 486–504.
26. Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N., Willmitzer, L. (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **18**, 1157–1161.
27. Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., Fernie, A.R. (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nature Prot.* **1**, 387–396.
28. Biais, B. and Allwood, J.W., Deborde, C., Xu, Y., Maucourt, M., Beauvoit , B., Dunn, W.B.,

Jacob, D., Goodacre, R., Rolin, D., Moing, A. (2009) ^{1}H-NMR, GC-EI-TOF-MS, and data set correlation for fruit metabolomics, application to melon. *Anal. Chem.* **81**, 2884–2894.

29. Allwood, J.W. and Erban, A., de Koning, S., Dunn, W.B., Luedemann, A., Lommen, A., Kay, L., Löscher, R., Kopka, J., Goodacre, R. (2009) Inter-laboratory reproducibility of fast gas chromatography – electron impact – time of flight mass spectrometry (GC-EI-TOFMS) based plant metabolomics. *Metabolomics* **5**, 479–496.
30. Broadhurst, D.I. and Kell, D.B. (2006) Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2**, 171–196.
31. Taylor, N.S., Weber, R.J.M., Southam, A.D., Payne, T.G., Hrydziuszko, O., Arvanitis, T.N., Viant, M.R. (2009) A new approach to toxicity testing in *Daphnia magna*: application of high throughput FT-ICR mass spectrometry metabolomics. *Metabolomics* **5**, 44–58.
32. Southam, A.D., Payne, T.G., Cooper, H.J., Arvanitis, T.N., Viant, M.R. (2007) Dynamic Range and Mass Accuracy of Wide-Scan Direct Infusion Nanoelectrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Increased by the Spectral Stitching Method. *Anal. Chem.* **79**, 4595–4602.
33. Payne, T.G., Southam, A.D., Arvanitis, T.N., Viant, M.R. (2009) A Signal Filtering Method for Improved Quantification and Noise Discrimination in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Data. *JASMS* **20,** 1087–1095.
34. Beckmann, M., Enot, D.P., Overy, D.P., Draper, J. (2007) Representation, comparison, and interpretation of metabolome fingerprint data for total composition analysis and quality trait investigation in potato cultivars. *J. Ag. Food Chem.* **55**, 3444–3451.
35. Breitling, R., Pitt, A.R., Barrett, M.P. (2006) Precision mapping of the metabolome. *Trends Biotech.* **24**, 543–548.