UC Davis
UC Davis Previously Published Works

Title
Metabolite Measurements

Permalink
https://escholarship.org/uc/item/4m44t1vm

ISBN
978-0-387-78744-2

Authors
Roessner, Ute
Beckles, Diane M

Publication Date
2009

DOI
10.1007/978-0-387-78745-9_3

Peer reviewed

Chapter 3
Metabolite Measurements

Ute Roessner and Diane M. Beckles

3.1 Introduction

Metabolites are the products of enzyme-catalyzed reactions that occur naturally within living cells. Metabolites are synthesized by the cell for the purpose of performing a useful, if not indispensable, function in the maintenance and survival of the cells by, for example, contributing to its infrastructure or energy requirements. To do so, they have to be recognized and acted upon by enzymes, which will change the properties of the metabolites by means of a chemical reaction. Therefore, the properties of metabolites and their functionality as they interact within their natural environment determine the chemistry of life. Thus, it can be argued that the metabolome in a biological system represents the final result of the expression of multiple genes in a cell. The analysis of metabolites has been an important part of any biological sciences. A large number of technologies have been developed for the analysis of metabolites in order to study metabolism in great detail. Today, the accumulation and combination of knowledge on analytical biochemistry from the last 50 years is commonly called metabolomics, and large investments are made to its application toward developments of new technologies with greater sensitivity, comprehensiveness, robustness, and higher throughput.

Oliver et al. introduced the term metabolomics in 1998 to describe the change of relative metabolite concentrations due to the alteration of gene expression [62]. Later, the definition of metabolomics describing the "comprehensive and quantitative analysis of all small molecules in a biological system" was introduced [26]. Today, metabolomics is commonly considered as the combination of analytics for metabolite determination with appropriate informatics for data extraction, mining, and interpretation.

U. Roessner (✉)

School of Botany, Australian Centre for Plant Functional Genomics, The University of Melbourne, 3010 Australia

e-mail: u.roessner@unimelb.edu.au

The range of chemical compounds synthesized by plants is astonishing. There may be more than 200,000 chemicals produced within the plant kingdom, representing a diverse array of structures, functional groups, and chemicals with different solubilities and reactivities [90]. This presents an enormous challenge for the researcher trying to assay these compounds in a massively parallel fashion. Thus far, only 50,000 compounds have been identified from a large number of species [20], highlighting the difficult nature of the task and the existence of thousands of undiscovered compound with different physical and chemical properties. Approximately 5,000–25,000 compounds may be produced in a single plant at a given developmental age and environmental state [61, 95, 99]. In contrast, about 500 are produced in bacteria, about 730 in yeast, and approximately 3,000 in human beings [96]. Apart from the sheer number and diversity of chemical structures, there are unique problems to cataloging the steady-state level of the plant metabolome. Reliable sampling and capture of the subset of compounds in a specific organ and subcellular location associated with the biological process of interest is, while not impossible, technically challenging and adds yet another layer of complexity to our ability to accurately record or view the metabolic state of the plant cell. The overall picture that emerges is that current extraction, analytical methods, and instrumentation are inadequate for truly assessing the plant metabolome, which may well remain a lofty vision to aspire toward in the future. These issues explain in part why the field of metabolomics lags behind genomics, transcriptomics, proteomics, and even glycomics. Identifying, reproducing, and curating the chemical species that make up the genome and transcriptome is comparatively facile, and even with the more difficult proteome and glycome, this goal is still feasible. This is easy to understand when the chemical structures of these compounds are considered. Nucleic acids, proteins, and glycoproteins have a common chemical structure(s) [12, 37, 40, 84]. Both DNA and RNA have a phosphodiester bond; all proteins have the peptide bond; glycoproteins have peptide bonds and glycoconjugates. In comparison, there is an astounding array of chemical groups that make up even the simplest of plant compounds. For example, several primary metabolites have inorganic and organic phosphates, amines, esters, hydroxides, carboxyl groups, and this does not even scratch the surface when one considers the exotic functional groups of secondary metabolites and that there are geometric and stereoisomers of many of these chemicals.

In the following, a short overview of the technologies and methodologies most commonly applied in metabolomics are described, with an emphasis on plant-based applications.

3.2 Technologies for Metabolite Analyses

The most commonly used platforms for the detection and measurement of metabolites involve their separation by gas chromatography (GC), liquid chromatography (LC), or capillary electrophoresis (CE) coupled with subsequent mass spectrometry
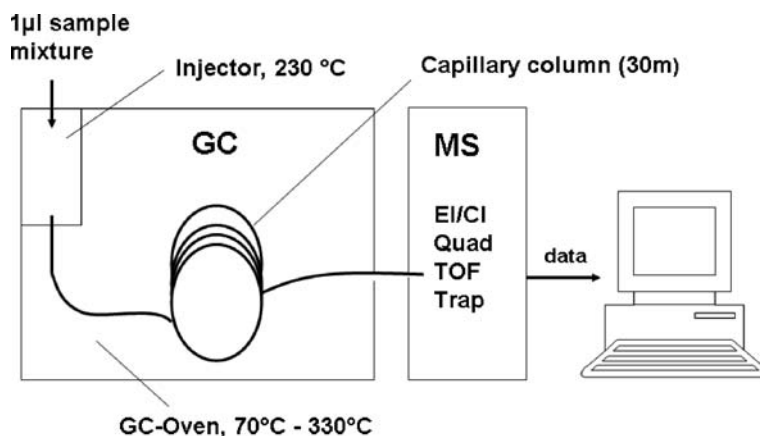
Fig. 3.1 Schematics of a typical GC-MS setup in metabolomics approaches
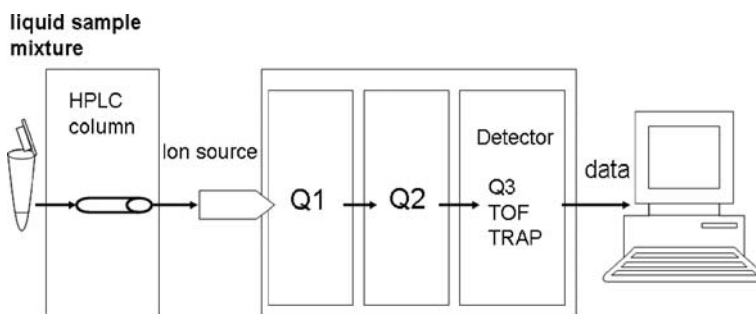

Fig. 3.2 Schematics of possible LC-tandem MS setups in metabolomics approaches

(MS) of the separated molecules (Figs 3.1 and 3.2). Compounds may also be measured directly without chromatographic separation. Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) and nuclear magnetic resonance spectroscopy (NMR) are two such examples (for review see [66]). The advantages and disadvantages of these technologies will be discussed in turn.

3.2.1 Mass Spectrometry

Different chemicals have different masses, and this fact is used in a mass spectrometer (MS) to determine which chemicals are present in a sample. The underlying principle of all MS is that the paths of gas phase ions in electric and magnetic fields are dependent on their mass-to-charge ratios which are then used by the mass analyzer to distinguish the ions from one another. The most important requirement of mass spectrometry is that compounds have to be vaporized and ionized (in an ion

source). Techniques for ionization have been key to determining what types of samples can be analyzed by mass spectrometry. Electron impact ionization (EI, Fig. 3.3) and chemical ionization (CI, Fig. 3.4) are mainly used for volatile compounds, e.g., in combination with gas chromatography (GC; see below, Fig. 3.1). In chemical ionization sources, the analyte is ionized by chemical ion–molecule reactions during collisions in the source.

Three ionization techniques often used with liquid and solid biological samples include electrospray ionization (ESI) (Fig. 3.5, also see Fig. 3.2), atmospheric pressure chemical or photon ionization (APCI/APPI), and matrix-assisted laser desorption/ionization (MALDI). Inductively coupled plasma sources are used primarily for metal analysis on a wide array of sample types.

Others include glow discharge, fast atom bombardment (FAB), thermospray, desorption/ionization on silicon (DIOS), secondary ion mass spectrometry (SIMS), and thermal ionization. The resulting ions are represented by their specific mass and charge, which means that ions from different chemical compounds will have different speed and directions within an electric or magnetic field. The ions are accelerated to a high speed and are thus separated in an electric and/or magnetic field. This process happens in the mass analyzer. There are many types of mass analyzers, using
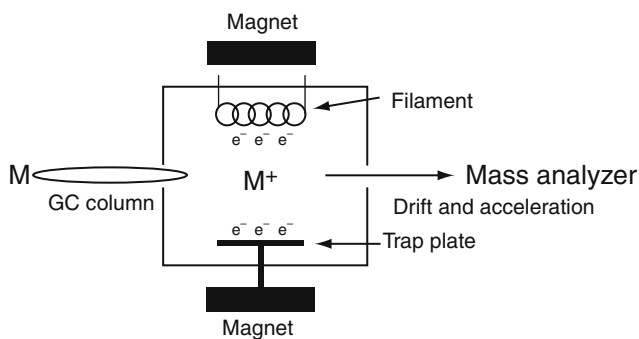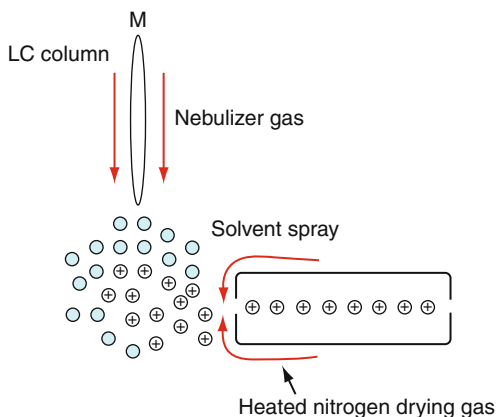

Fig. 3.3 Schematics of electron impact ionization. "M" is the molecule to be analyzed
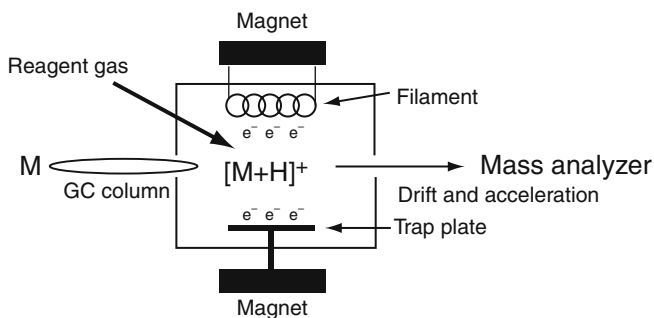
Fig. 3.4 Schematics of chemical ionization. "M" is the molecule to be analyzed
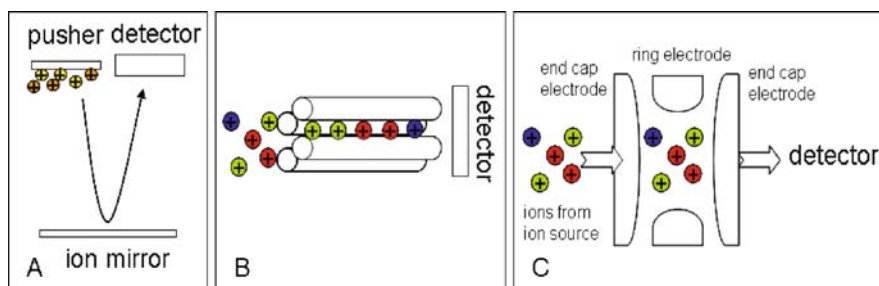
Fig. 3.5 Schematics of
electrospray ionization

M

LC column

Nebulizer gas

Solvent spray

Heated nitrogen drying gas

either static or dynamic fields and magnetic or electric fields, but all operate according to this same law. Most commonly used mass analyzers in biological applications include time-of-flight analyzer (TOF), ion trap analyzer (TRAP), quadrupole (Q), or Fourier transform ion cyclotron resonance MS (FT-ICR-MS).

Perhaps the easiest to understand is the time-of-flight (TOF) analyzer (Fig. 3.6A). It uses an electric field to accelerate the ions through the same potential, and then measures the time taken to reach the detector. If all the particles have the same charge, then their kinetic energies will be identical and their velocities will depend only on their masses. Lighter ions will reach the detector first and the heavier slower ions will be last.

Quadrupole mass analyzers use oscillating electrical fields to selectively stabilize or destabilize ions passing through a radio frequency (RF) quadrupole field (Fig. 3.6B). A quadrupole mass analyzer acts as a mass selective filter and is closely related to the quadrupole ion trap. An ion trap may use an electric or an electric and magnetic field to capture a cloud of ions in a region of a vacuum system or tube (Fig. 3.6C). The two most common types of ion traps are the Penning trap and the Paul trap. The Penning trap uses electric and magnetic fields while the

pusher detector

ion mirror

A

detector

B

end cap
electrode

ring electrode

end cap
electrode

detector

ions from
ion source

C

Fig. 3.6 Schematics of (**A**) time-of-flight, (**B**) quadrupole, and (**C**) ion trap mass analyzers (*see also Color Insert*)

Paul trap uses an electric field only. An ion trap MS may incorporate a Paul trap or the Orbitrap (which like the Paul trap, uses only electric field), and other types of mass spectrometers may also use a linear quadrupole ion trap as a selective mass filter.

The Orbitrap is the most recently introduced mass analyzer (commercially available since 2005, ThermoFisher). In the Orbitrap, ions are electrostatically trapped in an orbit around a central, spindle-shaped electrode. The electrode confines the ions so that they both orbit around the central electrode and oscillate back and forth along the central electrode's long axis. This oscillation generates an image current in the detector plates, which is recorded by the instrument. The frequencies of these image currents depend on the mass-to-charge ratios of the ions in the Orbitrap. Mass spectra are obtained by Fourier transformation of the recorded image currents. Similar to FT-ICR-MS (see below), Orbitraps have a high mass accuracy, high sensitivity, and a good dynamic range, but it is cheaper and less complex.

Fourier transform ion cyclotron resonance MS or FT-ICR-MS is regarded as the most complex type of mass analyzer. It determines the mass-to-charge ratio of ions based on the frequency of rotation of the ion inside a homogeneous magnetic field. The ions are first trapped in a magnetic field with electric trapping plates (Penning trap). The magnetic field causes the ions to adopt a circular motion perpendicular to the field. When an RF pulse is applied across the electric trapping plates, the ions are excited into a larger circular motion called the cyclotron frequency. The frequency of rotation is determined by the mass-to-charge ratio of each individual ion. Each ion will produce a current that is almost equivalent to the cyclotron frequency which is then recorded on a pair of detector plates. The resulting signal is called a free induction decay (FID), transient, or interferogram. The useful signal is extracted from this data by performing a Fourier transformation to give the mass spectrum. FT-ICR-MS is characterized by extremely high resolution in that masses can be determined with very high accuracy. Many applications of FT-ICR-MS use this mass accuracy to help determine the composition of molecules. This is possible due to the mass defect of the elements. Another place that FT-ICR-MS is useful is in dealing with complex mixtures since the resolution (narrow peak width) allows the signals of two ions of similar mass to charge ratio (m/z) to be detected as distinct ions. This high resolution is also useful in studying large macromolecules such as proteins with multiple charges, which can be produced by electrospray ionization. These large molecules contain a distribution of isotopes that produce a series of isotopic peaks. Because the isotopic peaks are close to each other on the m/z axis, due to the multiple charges, the high resolving power of the FT-ICR is extremely useful. FT-ICR-MS differs significantly from other MS techniques in that the ions are not detected by hitting a detector such as an electron multiplier but only by passing near detection plates. Additionally, the masses are not resolved in space or time as with other techniques but only in frequency. Thus, the different ions are not detected in different places as with sector instruments or at different times as with TOF instruments, but all ions are detected simultaneously over some given period of time.

Each analyzer type has its strengths and weaknesses. Many mass spectrometers use two or more mass analyzers for tandem mass spectrometry (MS/MS) (Fig. 3.7).
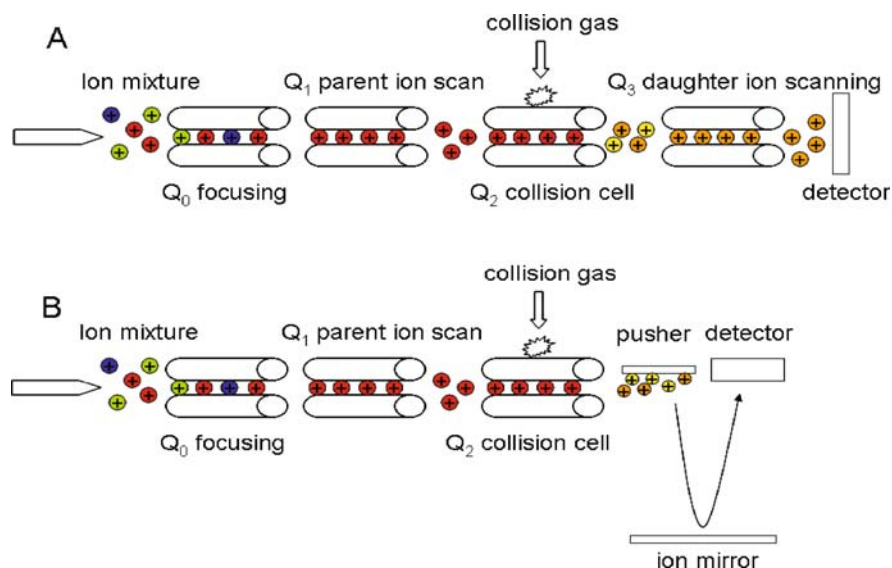
Fig. 3.7 Schemata of tandem MS/MS. (**A**) Triple quadrupole mass spectrometer (QqQ) and (**B**) quadrupole time-of-flight mass spectrometer (QqTOF) (*see* also Color Insert)

Tandem mass spectrometry involves multiple steps of mass selection or analysis, usually separated by some form of fragmentation. A tandem mass spectrometer is one capable of multiple rounds of mass spectrometry. For example, one mass analyzer can isolate one compound and determine its molecular weight. A second mass analyzer then stabilizes the molecular ion while it collides with a gas, causing them to fragment by collision-induced dissociation (CID). A third mass analyzer then catalogs the fragments produced from the original compound. Tandem MS can also be done in a single mass analyzer over time as in an ion trap. There are various methods for fragmenting molecules for tandem MS, including collision-induced dissociation (CID), electron capture dissociation (ECD), electron transfer dissociation (ETD), infrared multiphoton dissociation (IRMPD), and blackbody infrared radiative dissociation (BIRD). An important application, using tandem mass spectrometry, is in the structural elucidation of compounds.

The final element of the MS is the detector. The detector records the charge induced or current produced when an ion passes by or hits a surface. In a scanning instrument, the signal produced in the detector during the course of the scan vs. where the instrument is in the scan (at what m/q) will produce a mass spectrum, a record of ions as a function of m/q. Typically, some type of electron multiplier is used, though other detectors including Faraday cups and ion-to-photon detectors are also used. Because the number of ions leaving the mass analyzer at a particular instant is typically quite small, significant amplification is often necessary to get a signal. Microchannel Plate Detectors are commonly used in modern commercial instruments. In FT-ICR-MS and Orbitraps, the detector consists of a pair of metal

surfaces within the mass analyzer/ion trap region which the ions only pass near as they oscillate. No DC current is produced, only a weak AC image current is generated in a circuit between the electrodes.

In metabolomics, mass spectrometry is mostly used as a detector system following a specific separation procedure. Metabolite extracts from a biological source are characterized by great complexity, which has to be reduced before the compounds enter the mass spectrometer. Gas or liquid chromatography-based separation techniques allow the separation of complex mixtures with great precision. It involves passing a mixture dissolved in a "mobile phase" through a stationary phase, which separates the analyte to be measured from other molecules in the mixture and allows it to be isolated. Even very similar components, such as oligosaccharides that may only vary by a single sugar monomer or bond structure, can be separated with chromatography. In fact, chromatography can purify basically any soluble or volatile substance if the right adsorbent material, carrier fluid, and operating conditions are employed. Second, chromatography can be used to separate delicate compounds since the conditions under which it is performed are not typically severe. For these reasons, chromatography is quite well suited to a variety of uses in the field of biotechnology.

3.2.2 GC-MS

Gas chromatography linked to mass spectrometry is an effective and longstanding method for chemical analysis (Fig. 3.1). GC-MS presents several advantages; it is relatively easy to use, low in cost, gives reproducible results and excellent resolution of separated compounds. Hence GC-MS has been widely used in several separation applications since its first demonstrated use in 1964 for the quantitative and qualitative determination of polar/organic constituents of a sample [24].

The basis on which components are separated is on differential partitioning between a mobile gas phase and a solid stationary phase. Samples for GC-MS must first be converted from solid or liquid phase to a gas. This process called volatilization is accomplished by exposing the sample to high temperatures (up to 250°C). Once in the gas or mobile phase, the components are forced along a series of columns containing the solid or stationary phase. The volatilized compounds are partitioned between the two phases and the extent to which this occurs is determined by their chemical properties. Compounds that partition primarily in the mobile phase are eluted from the column faster than those with a greater affinity for the stationary matrix. The chemical behavior of each constituent, as it is eluted from the column, is recorded by its retention time.

Each separated compound eluting from the column must be subjected to ionization before entering the MS. Electron impact (EI) ionization produces electrons using a standardized filament voltage of 70 eV that effectively ionizes compounds (Fig. 3.3). These electrons are of high energy, and when they collide with separated compounds in the MS, they cause the compounds to fragment. The fragmentation

patterns are curated into mass spectral libraries for peak identification [51, 97]. Those produced from EI are highly reproducible which makes construction of spectral libraries using this technique consistent across experiments.

Chemical ionization (CI)-MS in contrast to EI uses gases, such as methane or ammonia, to provide the collision energy for fragmentation (Fig. 3.4). It is often described as a "softer" form of ionization, especially when compared to EI. The applications for CI, however, are limited. The fragmentation patterns produced by this method are less reproducible than for EI because it is difficult to control temperature and pressure of the ion source conditions. However, even with this drawback, CI-MS is an advantage for some types of applications, such as identifying compounds based on both the mass of the parent ion and its isotopic pattern. For example, most metabolic flux experiments use stable isotope labeling to monitor the fate of the target compound when metabolized in the fed tissues [25, 77]. Here, CI-MS is a better option for quantifying isotopic label in individual compounds produced from the labeled precursor, because the ligand–label complex likely remains intact [69].
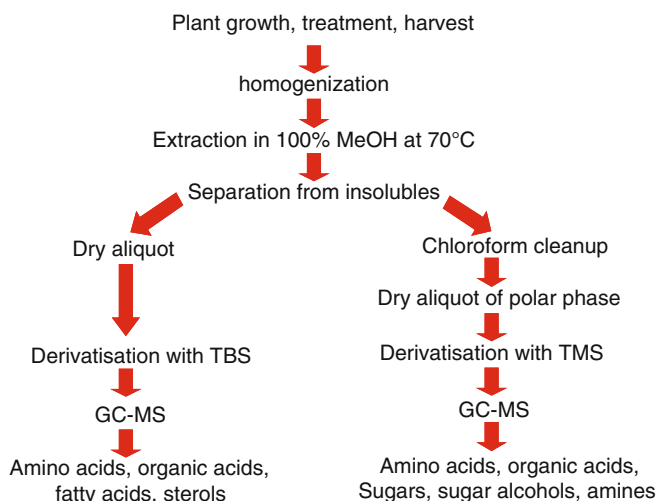
After ionization, the fragmented compounds are analyzed by mass detection by which quantification of ions can be achieved. Mass detection may be performed using quadrupole, ion-trap technology, or TOF detectors. The low-resolution quadrupole-type instruments are most commonly used, although fast scanning time-of-flight (TOF) MS detectors are becoming more popular.

The components eluted from the GC can be identified with a high degree of accuracy by comparing mass spectral and specific retention time indices of the eluted compounds to that of a reference database [51, 98]. Still, the identity of each compound should be verified by co-elution of authentic standards when available, and recovery assays should be performed to assess accuracy of measurements. Accomplishing these seemingly minor objectives is not as straightforward as they should be, especially when analyzing plant extracts. Purified standards of most plant-derived compounds are not commercially available. There are more than 350,000 curated chemical structures in GC-EI-MS libraries (e.g., NIST: http://www.nist.gov/srd/nist1a.htm and MSRI: http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/msri/gmd˙msri.html); yet a significant fraction of the detected compounds, in a typical GC-MS chromatogram, are not yet chemically identifiable. This is due, in part, to our ignorance of the diversity of chemicals manufactured in plants; it may also be due to the artifacts produced during extraction, GC ionization, and analysis. Recovery assays, although useful, add a significant time and cost factor to the analysis; however, it should be used in instances when assaying novel, low abundant, or unstable compounds.

For the researcher wishing to perform a basic characterization of polar compounds in plant extracts, GC-MS presents several advantages; it is reasonably inexpensive and easy to perform. Capital outlay is substantially lower when compared with other techniques such as LC-MS or NMR, and bench-top GC-MS instrumentation is readily available. From a quality control standpoint, GC-MS is exceptional in that there are only minimal matrix effects, i.e., the ionization efficiency of an analyte is not confounded by the presence of co-eluting substances [43]. Finally, there

are a growing number of easily adaptable methods available to survey the catalog of low molecular weight compounds in different plant systems (for review see [40, 96]).

There are some limitations to the application of GC-MS for metabolite profiling of any complex mixture, especially plant extracts. GC-MS is only suitable for analyzing compounds that can be volatilized either at high temperatures or by chemical modification. The result is that only small compounds ($\sim$1kD) can be effectively evaluated and thermolabile compounds are excluded. Further, samples must be derivatized, i.e., chemically modified to make them adaptable for GC-MS analysis. Derivatization simultaneously increases the volatility, thermostability, and detection limits of low-abundant compounds, and it is often achieved by silylation, alkylation, and acylation of extracts. Trimethylsilylation (TMS) is the most common method used in GC-MS-based metabolomics because it can derivatize a broad spectrum of compounds simultaneously, including sugars, amines, alcohols, amino, and organic acids [49], thus minimizing the probability that a chemical bias is introduced due to selective modification of one class of compounds. Derivatization is achieved when acidic protons in the target compound are effectively exchanged with TMS via a nucleophilic attack. It is imperative to use rigorous analytical procedures and standardized conditions when using TMS derivatives, as the derivatized compounds, especially amino acids, are sensitive to water and oxygen and are thermodynamically unstable [5]. An alternative derivatization method based on tert-butyldimethylsilylation (TBS) increases the chemical stability of amino acids and a range of other compounds thus widening the coverage of analytes potentially detected and measured by GC-MS [44]. A typical approach of GC-MS analysis using both types of derivatization is presented in Fig. 3.8.

Plant growth, treatment, harvest

↓

homogenization

↓

Extraction in 100% MeOH at 70°C

↓

Separation from insolubles

↙   ↘

Dry aliquot   Chloroform cleanup

↓   ↓

                       Dry aliquot of polar phase

↓   ↓

Derivatisation with TBS   Derivatisation with TMS

↓   ↓

GC-MS   GC-MS

↓   ↓

Amino acids, organic acids,   Amino acids, organic acids,
fatty acids, sterols   Sugars, sugar alcohols, amines

Fig. 3.8 Simplified workflow for GC-MS analysis of metabolites using TMS and TBS derivatization

3.2.3 LC-MS

LC-MS is gaining wider adoption as new technologies have allowed the high pressure liquid chromatography (HPLC) module to be interfaced with the low pressure (vacuum) of the MS (Fig. 3.2). Normal HPLC separation is accomplished by coupling LC with UV/VIS or diode-array detectors; however, the use of mass spectrometry enhances the specificity and selectivity of the system, thus improving the likelihood that determining the structural properties of compounds in a complex matrix can be identified. A requisite step in LC-MS is to convert the eluting compounds from a solute to a gas phase ion. This involves solute vaporization with concomitant ionization at the interface with the MS. Electrospray ionization (ESI, Fig. 3.5) [101], and atmospheric pressure ionization (API) [13] are good options for solute transformation. However, if the sample has a high concentration of salts or other ionizable compounds, these methods produce matrix effects such as ion suppression and ion enhancement [43], requiring additional downstream validation of the results [8].

LC-MS however presents several advantages over GC-MS. Samples do not have to be volatilized, which maintains the compounds in their native state before separation. Classes of compounds incompatible for analysis by GC such as those of higher molecular mass, greater polarity, and lower thermostability can be detected. Processing time may also be reduced in LC-MS because direct infusion methods can be used for analysis. The downside of direct infusion is that matrix effects are often a problem because of the larger number of compounds present in crude plant samples that can simultaneously enter the MS. Therefore, in most applications, LC separation prior to MS is advisable in order to reduce the complexity of ions to be scanned, even though this is less expedient.

A wide selection of column matrices is now available that support LC separations based on ion exchange, reversed phase, and hydrophobic interaction chromatography. Further, LC protocols that optimize the elution of constituents in complex compound mixtures have been developed. Very recently, a new sophisticated LC technology for nanoelectrospray application has been introduced. Nanospray LC-MS provides much higher sensitivity than normal flow electrospray LC-MS, however it uses several small capillary tubing connections which lead to frequent clogging and/or leaking at the column and spray needle. This feature has made nanospray-based LC-MS applications challenging. Agilent Technologies, Inc. has developed an HPLC-Chip interface for mass spectrometry (Fig. 3.9). All components required for LC and spray into the mass spec are integrated directly onto a reusable biocompatible polymer chip. This supports the delivery of solvent and sample, high pressure switching of flows, automated chip loading, and spray performance into the MS. An illustrating video of how the chip technology works can be retrieved from Agilent Technologies, Inc. website (www.agilent.com). Chips are being manufactured with different column packings, and users are even able to provide their own packing for custom-built chips. The HPLC-Chip has been already extensively applied in proteomics and peptidomics applications showing very high reproducibility, long shelf-life, and ease of use. Currently, its potential
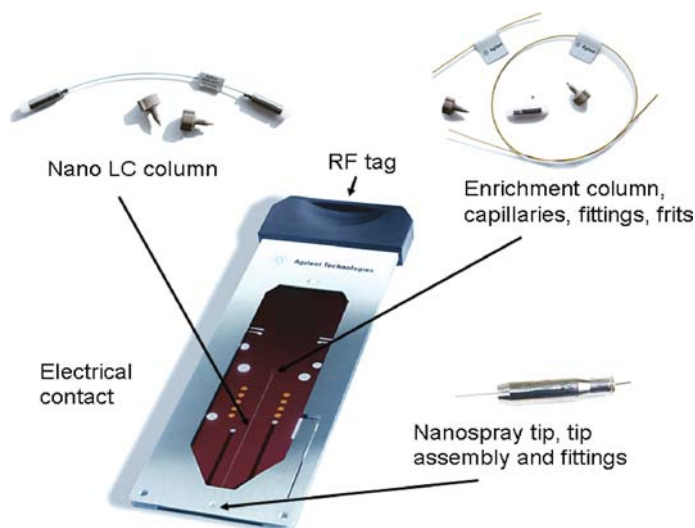
Fig. 3.9 HPLC-Chip and its integrated components related to the components of conventional LC (picture provided in courtesy of Agilent Technologies, Inc.) (*see* also Color Insert)

and applicability is being explored in the metabolomics field. LC, used in tandem with MS, produces a spectrum of separated compounds which can be detected with great selectivity. The fragmentation patterns produced provide information on the chemical structure of the compounds, and the system is also able to detect low-abundance metabolites. LC-MS has found utility in the separation and analysis of both primary and secondary metabolites in plant extracts [42, 89]. Mass detection may be dramatically increased to 5,000 signals from a single plant extract if Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) is used with LC [2].

The major obstacle that hinders the widespread use of LC-MS for many biological applications is that it is difficult to establish robust mass spectral libraries for peak identification [58]. The type of mass spectra produced by LC-MS is largely dictated by the instrument used (i.e., QqQ, QqTOF, Ion Trap, etc.), and the reference LC-MS spectral libraries constructed are of limited use because they are instrument specific. Each research group has to develop its own "in-house" LC-MS reference library infrastructure, which is often difficult and beyond the capabilities of the average plant biology lab.

3.2.4 CE-MS

Capillary electrophoresis (CE), either coupled to MS or to laser-induced fluorescence (LIF) detection, is a highly efficient and sensitive method for both targeted and unbiased profiling in plant extracts. It has not been widely adopted in

high-throughput metabolomic approaches, but has proven useful in particular applications because of its high sensitivity, especially compared to LC-MS. Derivatization is not necessary; solvent consumption is less; separation of low-molecular weight compounds with minimal pre-treatment is easily achieved; small sample sizes can be used; separation runs are faster than with LC-MS [65]; and unlike LC-MS, the separated species do not need to be ionized because they are already charged [80]. However, there are some complications to the use of CE for plant metabolomics. Only charged compounds or those that can be charged by changing the pH of the solution can be analyzed, and because there are limits to the volume of sample that can be injected onto the capillary, this can make detecting some compounds difficult. Still, some promising results have been already achieved. CE coupled with UV analysis has supported targeted profiling of some plant compound classes, e.g., organic acids [98], flavonoids [23] and amino acids [100]. When coupled to MS, more than 80 primary metabolites belonging to glycolytic, photorespiratory, and oxidative pentose phosphate pathways can be analyzed in rice leaf extracts [75]. However, the resolution power of CE coupled to LIF holds even greater promise for analyzing micro-amount of biological fluids. It has the sensitivity to separate and quantify a large number of amino acids and sugars in only ~50 picoliters of phloem sap, or the pooled sap of five leaf mesophyll cells from *Cucurbita maxima* [3]. Once the technology becomes more routinely used, it is likely to become very important for the development of cell-type-specific metabolite analyses.

3.2.5 NMR

NMR is potentially one of the most potent but underused methods available for plant metabolomics. NMR is a non-destructive, non-targeted fingerprinting technique that can detect a multitude of different metabolite classes irrespective of size, charge, volatility, or stability [21]. It is a powerful tool for comparative high-throughput profiling of plant extract, for determining metabolite structure and for elucidating metabolic fluxes. Although NMR use in the plant field lags that in others, e.g., medical research, there is renewed interest and increasing applications of this technology to addressing basic biological questions related to plant processes (67).

NMR spectroscopy uses the magnetic properties of atoms that make up the chemical structure of compounds. A strong magnetic field is combined with radio frequency pulses to produce high-energy spin states in nuclei with odd atomic or mass numbers (e.g., ^{1}H or ^{13}C). The radiation emitted when these nuclei return to the lower energy spin state is detected and used to eventually construct the chemical structure of the analyte [21]. One of the most attractive features of NMR is that metabolites can be measured non-destructively; permitting in vivo measurement of metabolites in intact tissues [54]. Most importantly, NMR provides high-resolution structural information about the metabolites for unambiguous identification. When combined with stable-isotopic labeling, NMR becomes very informative. Real-time

in vivo flux of compounds can be monitored and even resolved between subcellular compartments [54, 66, 73]. For example, the location of ^{31}P in different plastids can be determined based on the chemical shift caused by the different pH of the subcellular compartments [73].

There are a number of factors that need to be considered when NMR is applied to plant metabolomics. NMR is in general less sensitive and less discriminating than other established separation techniques [21]. The signal-to-noise ratio and resolution of in vivo samples are much lower when compared to those in vitro extracts; metabolites below 5 nmol will not be detected by NMR [52], and although volumes of around 2μl may be used [38], NMR generally requires larger sample input. In addition, NMR involves a considerably greater capital investment in instrumentation, and when the number of compounds actually detected is considered vs. cost, it is not always the best option. NMR is best suited for specialized applications involving in vivo real-time imaging of flux, structural elucidation of compounds, and for metabolite fingerprinting. There are emerging technologies that may supersede NMR in some applications because of their resolving power. Once such example is imaging matrix assisted laser desorption ionization mass spectrometry (I-MALDI) coupled with TOF-MS. This was used to detect and measure a range of metabolites even at micromolar concentrations in different subcellular fractions of the plant cell [14].

3.2.6 Other Novel, Highly Valuable Approaches for Metabolite Analysis

As mentioned, the described technologies are currently the most commonly applied analytical instrumentations for metabolite analyses. A range of well-established and routinely applicable methodologies are available for the analysis of a number of different compound classes, ranging from primary metabolites, such as sugars, amino, organic or fatty acids to highly complex secondary metabolites, such as alkaloids and flavonoids [40]. Although claimed as being high-throughput methods, they often are limited by time per sample due to time-consuming chromatographic separation upfront MS analysis. A recently introduced separation technique that shows great promise is based on ion mobility [94]. Ion mobility spectrometry coupled to mass spectrometry (IMS-MS) is a technique where ions are first separated by drift time through a neutral gas which was given an electrical potential gradient before being introduced into a mass spectrometer. The IMS technique separates and detects ions that have been sorted according to how fast they travel through an electrical field in a tube. Small ions travel very fast, and they reach the detector first, with successively larger ions following along. Unlike the mass spectrometry technique, which relies on very low pressures to keep the ions from colliding with each other, IMS operates at normal atmospheric pressure, and the ions collide with each other repeatedly. The ionized gas moves through an electrical field inside a drift tube. Smaller ions collide less frequently than large ions because they present a smaller target and are

harder to hit. Thus, they move through the tube relatively unimpeded and reach the detector first. The largest ions take several seconds longer to travel to the detector because they collide more frequently with other ions along the way. Because IMS only sorts molecules by size, and not by chemical properties or other identifying features, it is not a particularly good technique for making a positive identification of unknown compounds. However, the duty cycle of IMS is short relative to LC or GC separations and can thus be coupled to such techniques producing triply hyphenated techniques such as LC-IMS-MS. Perhaps IMS's greatest strength is the speed at which the separations occur – typically on the order of 10s of milliseconds. As a research tool, ion mobility has already shown great strides toward the analysis of biological materials, specifically in proteomics and metabolomics [16, 56, 85, 93].

A powerful way of assaying low-abundant metabolites has recently been re-introduced by Gibon and colleagues [33]. A large number of metabolites are present in very small quantities, especially metabolic intermediates and cofactors in plant cells. Also, in many cases, only small samples sizes can be harvested. To overcome these obstacles, Gibon et al. developed robotized microplate-based activity assays using enzymes involved in central carbon and nitrogen metabolism as tools to assay the metabolites involved in the reaction mechanism. The distinct advantage of this high-sensitivity assay is that limited sample size or metabolites in low abundance are not insurmountable problems for analysis [31, 32].

3.3 Data Analysis

Analyzing the large volumes of data produced by metabolite profiling technologies is still a major challenge facing researchers. The best approach for data analysis will depend on the aim and application of the specific experiment or process investigated. Although accurate detection of compounds is necessary regardless of research objectives, slightly different emphases on the downstream analyses will be required depending on if the primary aim is to identify hitherto unknown compounds, determine differences between samples, or evaluate broad changes in entire pathways. The first step is to ensure adequate quality control of the raw data, i.e., accurate peak identification, assignment, and quantification. Next, robust statistical validation of raw data is critical as it affects the interpretation of the data if comparisons are made between different systems. Finally, transformation and presentation of the data in a manner that allows efficient and maximal extraction of biological information from the system to be studied in an intuitive user-friendly interface are desirable.

Once validated, the data needs to be mined and presented in a manner to make hypothesis testing facile, to highlight pattern or relationships among variables, or to drive the generation of new biological questions. Several methods are currently used including cluster analysis, pathway mapping, and comparative overlays, as well as heatmaps. They each provide slightly different insight into multidimensional datasets and may even be complementary.

3.3.1 Data Extraction from Analytical Instrumentation

This is the most important, but often under-estimated process of data analysis in metabolomics applications both based on mass spectrometry or NMR. The process includes raw data format transformation (if necessary), chromatogram deconvolution, peak detection, peak alignments, baseline corrections, noise reduction, peak assignment (identification), peak quantification, and data validation. Most importantly, the procedure has to be applicable for each batch measurement with the desired or necessary level of correctness. A vast number of software packages are available, both commercially and as free open-public packages. The description of those software packages is beyond the scope of this chapter. Table 3.1 lists a few of the most common software packages available for mass spectral data analysis. Although much progress has been made in developing software application, manual inspection and validation of raw data are still essential and represent a limiting step in a process that could be automated. Therefore, constant evolution and improvements of sophisticated programs capable of evaluating raw data are critical. This would have the dual advantage of increased confidence that the effects of false positives are minimized and that operator time on data analysis is reduced. However, the usefulness of such programs will be determined largely by how it meets the needs of each individual application. Each metabolomics user should invest the time and effort to test some of those packages to decide best suitability for their analytical method as well as comfort in using the package.

3.3.2 Statistics

Data analysis is the process of transforming data with the aim of extracting useful information from which to draw conclusions and develop new working hypotheses. Depending on the type of data and the question asked, this includes the application

Table 3.1 Some examples of software available for mass spectral data analysis

Name	URL
MSFACTs	http://www.noble.org/PlantBio/MS/MSFACTs/MSFACTs.html [22]
MET-IDEA	http://www.noble.org/plantbio/ms/MET-IDEA/index.html [11]
HiRes	http://hatch.cpmc.columbia.edu/highresmrs.html [103]
MZmine	http://mzmine.sourceforge.net/ [48]
XCMS	http://metlin.scripps.edu/download/ [79]
SpectConnect	http://spectconnect.mit.edu [83]
AMDIS	http://chemdata.nist.gov/mass-spc/amdis/ [81]
AnalyzerPro	www.spectralworks.com
metAlign	www.metalign.nl
SIEVE	http://www.thermo.com/
GeneSpringMS	http://www.agilent.com/

of statistical methods, selecting or discarding certain subsets based on specific criteria, or other mathematical methods.

In applying statistics to a scientific problem, one begins with a biological process or sets of organisms to be studied. This might be a population of transgenic plants compared to untransformed wild type, plants treated with a specific stress elicitor compared to unstressed plants, or plants growing in one environment vs. another. It may instead be a comparison observed at various times; data collected about this kind of "population" constitute what is called a time series. For practical reasons, a chosen subset of the population – a sample – is studied, rather than compiling data about an entire population. Once data are collected about the sample, they can be subjected to statistical analysis which serves two related purposes: description and inference. Descriptive statistics can be used to summarize the data, either numerically or graphically, to describe the sample. Basic examples of numerical descriptors include the mean and standard deviation which can be presented as tables or various kinds of graphs and charts. Inferential statistics is used to model patterns in the data, accounting for randomness and drawing inferences about the larger population. These inferences are then presenting a form of hypothesis testing, estimation of numerical characteristics, correlation (description of associations), or regression (modeling of relationships).

If a sample is representative of the population, then inferences and conclusions made from the sample can be extended to the population as a whole. A major problem lies in determining the extent to which the chosen sample is representative. Statistics offers methods to estimate and correct for randomness in the sample and in the data collection procedure, as well as methods for designing robust experiments in the first place. The fundamental mathematical concept employed in understanding such randomness is probability.

The use of any statistical method is valid only when the system or population under consideration satisfies the basic mathematical assumptions of the method. Inappropriate use can produce subtle but serious errors in description and interpretation – subtle in that even experienced professionals sometimes make such errors, and serious is that they may affect data interpretation and decisions. Even when statistics is correctly applied, the results can be difficult to interpret for a non-expert – for example, the statistical significance of a trend in the data, which measures the extent to which the trend could be caused by random variation in the sample and may not agree with one's intuitive sense of its significance. Therefore, a set of basic statistical skills (and skepticism) is needed in the process of metabolomics data analysis and interpretation to deal with the vast amount of information metabolomics is providing.

The most common statistical methods applied for metabolomics data are Student's *t* test and analysis of variance (ANOVA). A *t* test is any statistical hypothesis test in which the test statistic has a Student's *t* distribution if the null hypothesis is true. A requirement is that the means of two normally distributed populations are equal. Given two datasets, which are characterized by their means, standard deviations, and a number of data points, the *t* test is used to determine whether the means are distinct, provided that the underlying distributions can be assumed

to be normal. Two versions of the *t* test exist; the two samples are either independent or dependent of each other. When samples are dependent of each other, or paired, it means that each member of one sample has a unique relationship with a particular member of the other sample. Once a *t* value is determined, a *p* value can be determined using lookup tables or integral calculus (http://www.danielsoper.com/statcalc/calc08.aspx). A threshold chosen for statistical significance (usually below 0.05 Student's *t* distribution for *P* or with 95% confidence level) indicates that the two sample groups differ from each other.

ANOVA is a collection of statistical models, in which the observed variance is partitioned into components due to different variables. There are several types of ANOVA depending on the number of comparing pairs (treatment, genotype, time) under analysis. One-way ANOVA is used to test for differences among three or more independent groups. Factorial or two-way ANOVA is used when the effects of two or more treatment variables are under investigation. The most commonly used two-way ANOVA is the "two-by-two" design, where there are two independent variables and each variable has two levels or distinct values – for example, a control and a mutant plant at two developmental stages.

Three problems occur when both *t* test and ANOVA are applied to metabolomics data. First, both methods assume that the data under analysis are normal distributed which may or may not be true for metabolite data. Secondly, inadequate sample sizes for the large number of variables (metabolites, signals) measured that are required to prove a metabolite are discriminant, and thirdly, multiple hypothesis testing using only univariate statistical tests across all metabolites in parallel [9]. It is worth to note here that the same issues are of concern in other "omics" or high-throughput biological sciences, where a high number of variables are determined in a comparably small sample collection. The issues will increase the so-called false discovery rate (FDR), leading to the identification of apparently significant metabolites/biomarkers which are in fact incorrect. The false discovery rate (FDR) can be described as the expected percent of false prediction from the whole number of prediction and increases with the number of tests performed on a given dataset.

The first problem is difficult to tackle, in theory for each metabolite measured the distribution has to be determined and the appropriate statistical analysis chosen. This is not always feasible and commonly it has been agreed to treat metabolomics data as not normally distributed datasets. Therefore, before applying either *t* test or ANOVA data should be transformed (e.g., by either $\log_2$ or $\log_{10}$ transformation) to achieve close to normal distribution. The second problem is only solvable if the number of biological replications per sample would be increased dramatically, which again is not feasible for time, cost, and other logistical factors. Therefore, additional more stringent statistical methods have to be applied in order to deal with the inadequate sample sizes. The third issue of multiple hypothesis testing using univariate methods can be in fact only excluded by applying multivariate data analysis methods, which will be described below. Broadhurst and Kell [9] summarized those issues arising in metabolomics experiments very clearly and provided a framework of tackling the issues in an easy way to implement. This method is based on the Bonferroni correction [1] in which a new, more stringent *p* value is calculated by

dividing the originally chosen p value threshold by the number of variables under t test or ANOVA test ($p\prime = p/n$). For example, if 100 metabolites are determined and the p value threshold was set to 0.05, the new $p\prime$ value threshold would be 0.0005, leading to a substantially smaller number of statistically significant metabolites (with now p value below 0.0005) giving much higher confidence for biological interpretation. The Bonferroni correction is often criticized as being too stringent. Metabolites that fail to meet the new p threshold may still be involved in the biological process studied. Therefore, the validity of using this parameter should be examined on a case-by-case basis.

3.3.3 Data Mining, Classification, and Visualization

Data mining is the extraction of potentially useful information from large datasets by multidimensional analysis. Dataset are sorted, analyzed, and sifted through to extract non-intuitive, but potentially important information. The broad aims are to discover new patterns between the variables examined, to test existing scientific models, to refine the theoretical understanding of a system to the point where it is fact based, and to enhance the ability to predict behavior or trends. Data mining is distinct from data analysis in that it is the platform itself and the dynamic interaction between the user and that platform that eventually provides the user with insight into the data not readily obvious by a precursory examination.

Common data mining methods in metabolomics applications include clustering, such as hierarchical clustering (HCA), multivariate data analysis, principle or independent component analysis, and co-response or correlation analysis. It is outside the scope of this chapter to describe the methods in detail and the ways in which they are applied in metabolomics applications; therefore, the reader is referred to Chapter 6 of this book. In addition, a number of summarizing reviews are available for more detailed understanding (e.g., [96] and [45] and references therein).

A number of web-based tools for the visualization and interpretation of omic data are becoming rapidly available. Ideally, these tools should allow the user to easily extract meaningful biological data from large datasets in a highly interactive, user-driven manner. With metabolomics, the ultimate goal is to understand the regulation of biochemical pathways. This perhaps can be best achieved by integrating gene, protein, and metabolite profiling data and examining the interrelationships of each level of the system. Here we give a brief overview of some of the more popular resources available for analyzing omic data generated from plant species. Each has advantages and disadvantages, and their usefulness will depend on the specific needs of the individual researcher.

The software module of the Pathway Tools database called the "Omics Viewer" (Pathway Tools Omics Viewer – PTOV) is one of the most extensive available. Gene and protein expression data, metabolite levels, and metabolic flux analysis data can be integrated onto a single metabolic map or interface [63]. The full utility of this tool is limited to species which have near-complete genomic sequence, i.e.,

Arabidopsis (dicot; AraCyc) (http://www.arabidopsis.org:1555/expression.html), rice (monocot; RiceCyc) (http://pathway.gramene.org/expression.html), and *Medicago* (legumes; MedicCyc) [92]. The output shows an overview of grouped metabolic pathways where measured metabolites, transcript, etc. can be color-coordinated based on their relative amount in the two samples compared. It is anticipated that as resources grow for other model species, they will be interfaced with PTOV. For example, preliminary data from the *Solanaceace* genomic projects are already being integrated at RiceCyc. MetNet (*Met*abolic *Net*work Exchange) is another plant-specific tool for functional genomics, which permits the user to visualize and overlay pathway components from comparative metabolite and transcriptomic studies. What distinguishes MetNet from similar programs is that it incorporates regulatory pathway information and contains applications for network modeling. In addition, MetNet contains rigorous statistical capabilities and it dovetails with AraCyc at PTOV. One disadvantage is that it is developed solely for *Arabidopsis*. MapMan [87] is a popular interface that is being increasingly used because it is a standalone, is relatively easy to navigate, and can be easily adapted to other species, e.g., legumes [34] and tomato [91]. It, however, depicts pathway data that are categorized by function, and only two experiments can be compared on the same map. Other publicly available tools include kaPPA-View and VANTED [46]. kaPPA-View was developed using *Arabidopsis* genomic data; it boasts an extensive secondary metabolic pathway depiction (added from other plant species), and gene expression data from multiple enzyme isoforms can be shown on the pathway. The user also has the option of showing transcript and metabolite data side by side, which is not possible with any other programs. VANTED (*V*isualization and *A*nalysis of *N*etworks with related *E*xperimental *D*ata) has been more recently introduced and allows the user to show detailed information for individual metabolites, which is advancement over most other tools. It is also possible to compare more than two experiments on the same map, but the trade-off is that this limits the number of pathway steps shown [46].
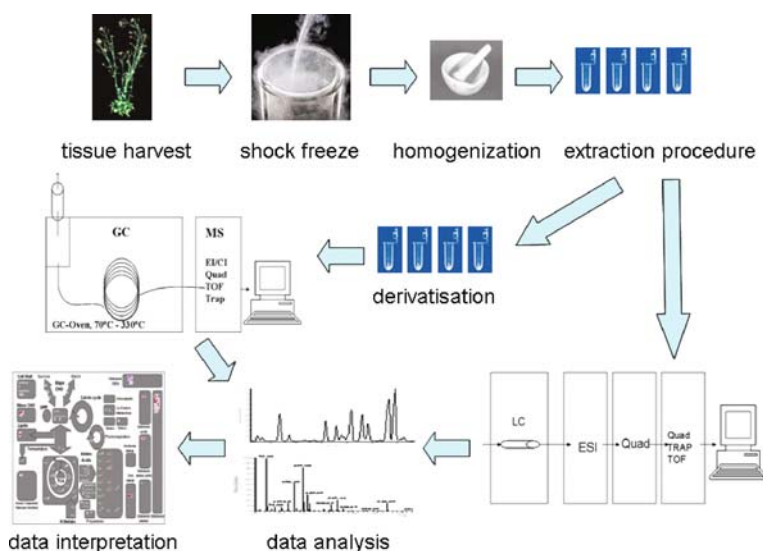
3.4 Metabolomics Approaches

After some confusion over the correct nomenclature to accurately describe various metabolite analytical approaches, Fiehn has provided clear definitions for each [27] which are now commonly accepted and used in the metabolomics community.

Target analysis describes the determination and quantification of a small set of known metabolites using one particular analytical technique, such as HPLC with UV detection. This type of approach has been done for many decades, before even the word metabolomics was invented. A number of different columns and elution protocols for many different compounds or compound classes have been developed and successfully applied for metabolic studies. Methods exist for small sugar analysis [15], amino acids [41], nucleotides [57, 68], flavonoids [59], volatiles [6], or alkaloids [30].

The second type of approach is called metabolite profiling, which includes the analysis of a larger set of both, identified and unknown metabolites, in a more unbiased manner, for instance using GC-MS. A number of exciting applications of GC-MS in plant metabolomics have been already described. To date, GC-MS is considered as the "workhorse" of separation and profiling technologies because the methodologies are well established and routinely applicable. Off the shelf, comparably affordable instrumentation is available from a range of vendors. Existing methods allow the detection and quantification of a large number of small metabolites, such as amino, organic and fatty acids, sugars, sugar alcohols, amines, or sterols. The development of a GC-MS mass spectral library (MSRI) in the public domain [51] is a great help to the researcher planning to use GC-MS technology. This library contains not only the electron impact originated mass spectra of metabolites but also their corresponding retention time indices which together, provides a highly valuable tool for peak identification. In the past decade, the GC-MS technology has been successfully applied for comparative metabolomics approaches in a range of different plant species, for example in *Arabidopsis thaliana* [28], *Solanum tuberosum* [71], *Medicago truncatula* [10], *Lotus japonicus* [17] or *Hordeum vulgare* [72], *Oryza sativa* [86], and *Solanum esculentum* [68].

Metabolomics itself would represent the determination and quantification of as many metabolites as possible, again both identified and unidentified, using complementary analytical methodologies to ensure maximal comprehensiveness, e.g., LC-MS/MS, GC-MS, and CE-MS. A schematic approach for metabolomics starting from sample harvest to data analysis is presented in Fig. 3.10.

Fig. 3.10 Schematic workflow of a metabolomics approach from tissue harvest to data interpretation using complementary analytical instrumentation for greater comprehensiveness of metabolite detection and quantification (*see* also Color Insert)

The study of Tohge et al. [88] illustrates the utility of using multiple technologies platforms to investigate gene functions in rice. The aim of the study was to determine the effects of over-expression of the *PAP1* gene encoding an MYB transcription factor on *Arabidopsis* plants. This T-DNA activation-tagged line is known to produce higher levels of anthocyanins compared to untransformed plants. In order to unravel the mode of action of this transcription factor, the transgenic plants were analyzed by targeted profiling confirming the higher levels of flavonoids using liquid chromatography with photodiode array detection followed by mass spectral analysis (LC-PAD-MS). In addition, amino acid levels were determined by HPLC and anions and sugars by CE-MS, and a non-targeted analysis was performed using FT-ICR-MS. One outcome of the study was that the sample origin (plant organ) and the growth conditions have greater influences on the metabolite composition than the transgenic event. When different analytical platforms are employed for comprehensive metabolite analysis, attention must be paid to sample harvest (sufficient tissue for the different methods) and data integration. Most importantly, when data are integrated for data mining and interpretation, each type of analysis should be performed using the same sample, or more ideally, the same homogenate. The aliquots of this homogenate may then undergo different extraction procedures to prepare the sample to be amendable for the chosen analytical techniques. If possible, a single extraction method for all measurements would be preferable to reduce influences of variability of extraction onto the data generation.

The last more general approach is called metabolic fingerprinting, in which a metabolic "signature" of the sample is generated, for instance, using direct infusion ESI-MS (DIMS) or FT-ICR-MS. This approach is often used for high-throughput screening of large mutant collection or mapping populations. In the first instance, no attempt is made to assign detection signals to a particular metabolite but once a significantly discriminating signal is obtained, identification becomes crucial for interpretation. A recent example is provided by Oikawa and colleagues who described a strategic approach of using FT-ICR-MS for high-throughput metabolomics in plants [60]. The authors suggest a scheme for plant sample analysis starting from the data generation in a reproducible manner using FT-ICR-MS without any further chromatographic separation of the complex extracts. Obtained mass fingerprints (*m/z* values with respective ion intensity) were then mass-error corrected and directly submitted to multivariate analysis using a newly developed software tool (DMASS). Marker metabolites were identified by searching the open-source metabolite relationship database KNApSAcK, and putative identifications were confirmed by structural analysis using MS/MS mode of the FT-ICR-MS. The described scheme will be extremely helpful for high-throughput metabolic phenotyping studies not only in plants but also in other biological systems.

3.5 Application Examples in Plant Sciences

Even with the challenges and limitation of truly producing a comprehensive metabolite profile of plant extracts, metabolomics has already been successfully applied to many fields in plant science. This discipline is rapidly becoming useful to

address fundamental and longstanding questions in biochemistry and physiology, which in turn has, and will continue to broaden our understanding of plant biology. Metabolomics has found applications for the comprehensive phenotyping of genetic varieties or genetically modified plants (GMOs), to determine gene function, to monitor plant behavior, responses to biotic and abiotic stress, and to make determinations of substantial equivalence. Because metabolites are the end products of gene expression, metabolic profiling also has the potential to bridge the gap between genotype and phenotype [36] and thus provides a more comprehensive and integrated view of how cells function in multicellular organisms. In addition, metabolic profiling has the potential to uncover new or dramatic changes in specific metabolites that can point to new hitherto undiscovered regulatory mechanisms or to a prediction of gene or protein function. Academically intriguing questions, testable hypotheses, and potentially new biotechnological targets can be generated and identified from analysis and data mining of metabolic metadata sets. Here we summarize a selection of metabolomics applications in plant research. A vast amount of primary research literature and excellent books and reviews demonstrating and summarizing the fast-growing field of plant metabolomics are available, and the reader is referred to [40, 74, 96] or [78]. In this chapter, we present only a small selection of examples for potential applications of metabolomics.

Metabolomics can be used for comprehensive phenotyping of genetic varieties or genetically modified plants (GMOs). This was capably demonstrated in pioneering work by Fiehn et al. [29] in which GC-MS-based metabolomics was applied to compare the metabolite profiles of two *Arabidopsis* ecotypes (parentals) and two respective, well-characterized mutants produced from each of the two parental lines. This study showed that the differences in the metabolomes of the ecotypes were far greater than those observed in the respective mutants. Roessner et al. [71] used GC-MS-based metabolite profiling to investigate the influence of transgenic modifications of the sucrose degradation and starch synthetic pathway in potato tubers. Large differences in the metabolite profile as a result of the genetic modification were detected. However, when the wildtype tubers were incubated under different in vitro culture conditions, the metabolite profiles of wildtype tubers could be drastically altered, such that even greater modulations in metabolite levels were introduced than that in the transgenics [71]. These two works established the discriminatory power of metabolomics to resolve and molecularly separate organisms based on genotype and culture conditions. Thereafter, metabolomics has been used to test substantial equivalence of GMO crops to examine potential "unintended" effects on the chemical constitution of the cell caused by transgenic manipulation [4, 16, 53, 104]. Although limited in scope and thus requiring caution in interpretation, the results thus far tend to support the view that traditional breeding, different ecotypes/cultivars, or growth in different environments may produce inherently more genetic variation in plants than that from some transgenic manipulations.

Metabolic profiling has also helped to redirect our knowledge of plant primary metabolism. For example, Roessner-Tunali et al. established that the de novo synthesis of amino acids occurs in non-photosynthetic organs (potato tuber) and was independent on import from leaves which was a departure from classical textbook description of this pathway [70]. Another long-held view was that import of sugars

into tomato fruit was symplastic early in development and then switched to an apoplastic route during ripening [64]. This was challenged after analysis of metabolite profiles in transgenic tomatoes with reduced expression of a sucrose importer, as the results implied that sugar import was apoplast [39]. These two examples show the power of the metabolomics approach in putting new knowledge within reach by taking an unbiased approach to studying metabolism.

Metabolomics is set to become important as a high-throughput method to screen and accurately quantify the phenotypes of large genetic mapping populations. An important tool for identifying novel genetic variation and new genes determining plant performance and fitness is quantitative trait locus (QTL) mapping. Almost by necessity, traditional QTL analysis was done on easily scorable phenotypes such as fruit color, yield, or stress tolerance. However, the availability of novel technologies for high-throughput and simultaneous analysis of transcripts or metabolites has proved to be efficient for the rapid and efficient dissection of multiple traits at the molecular level, offering unprecedented access to QTLs. Large recombinant populations can therefore be characterized with greater precision for desired features, e.g., carotenoids, vitamins, acid, and/or sugar content in fruits, thus connecting DNA content to the measurable change in phenotype.

Novel QTLs that control the level of a single or a network of metabolites can be rapidly identified by direct comparison of the metabolite profile of progeny with the parents from whom they were derived. For example, Schauer et al. [76] used GC-MS technology to profile fruit from a well-characterized mapping population made from a cross of the cultivated tomato species (*Solanum lycopersicum*) and a wild, non-ripening tomato species (*Solanum pennellii*). Genomic regions from *S. lycopersicon*, which were populated with markers, were introgressed into homologous regions of *S. pennellii* thus allowing mapping of the genes responsible for contrasting traits, e.g., yield, fruit color, fruit sugars between the parentals [76]. A large number of single metabolite QTLs were identified; for example, four lines with overlapping genomic regions correlated with an increase in malate, suggesting that the QTL may map to potential metabolic enzyme(s) or regulatory gene(s) that controls or regulates malate levels. Many QTLs that affected entire pathways and/or metabolic networks in tomato fruit were also discovered, and due to careful measurement of physiological parameters, the study also showed that events in source or photosynthetic tissues has a large control over traits in fruit, even though there is large degree of spatio-temporal separation of biological activities of the two organs. This work will undoubtedly provide tomato breeders with an arsenal of new gene targets that could drive future tomato genetic improvement strategies. Further, it has heightened our awareness of the inseparable and indivisible nature of the biochemical and physiological behavior of crop plants in determining quality traits.

In another example, Keurentjes et al. [50] used a non-targeted LC-qTOF-MS method to produce metabolic fingerprints of 14 *Arabidopsis thaliana* accessions which could serve as parentals in a subsequent mapping experiment. The goal of this study is to support QTL detection by comparing the metabolite profiles of these fingerprinted parental accessions and their resulting progeny in a recombinant inbred line (RIL) population. Over 2,000 individual mass peaks were detected by

LC-qTOF-MS, and careful examination of the metabolomes of the two most divergent accessions allowed QTLs to be assigned to about 75% of all mass peaks [50]. If each mass peak can be unequivocally identified and chemically determined, this application has the power to chart new metabolic pathways and simultaneously elucidate their underlying genetic control which is potentially exciting. The two examples we offer here may well be pioneering work for future QTL identification and mapping; regardless, they demonstrate the potential that metabolomics offers to this field dominated by DNA sequencing technologies. As innovations in metabolomics support faster, easier, and more robust measurements, it will be possible to broadly apply metabolomics to study genetic segregation and to identify novel genes that contribute to biotechnologically important phenotypic traits. It can be envisioned that this new "reverse-QTL" approach will identify pathways based on metabolic profiles that underlie a trait of interest thus enabling a new paradigm in QTL screening and mapping.

Metabolomics represents a powerful approach of monitoring response, adaptation, and tolerance mechanisms of plants to challenging environments and is therefore ripe for use as a diagnostic and investigative tool. Plants are often able to survive and protect themselves against an onslaught of environmental stresses such as extreme temperature, aridity, and salinity by modulating intracellular solutes. This is usually achieved by increased synthesis and accumulation of compatible solutes or osmoprotectant such as polyols (including glycerol and sorbitol), amino acids (especially proline), quaternary compound (glycine betaine), or tertiary sulfonium compounds (dimethylsulfoniopropionate) [7, 67, 82, 102]. Integrating metabolomics with physiology is proving to be yet another valuable resource to dissect plant molecular response to abiotic stress and has generated new and exciting results in this emerging field. Kaplan et al. [47] and Cook et al. [19] detailed the plant metabolic adaptations after perturbation of the systems by exposure to variations in temperature. Low temperatures had greater repercussions for metabolite levels than did high temperatures, and several previously unknown adaptive mechanistic responses to cold stress were revealed, including changes in cellular amino acids, intermediates of both the TCA cycle and carbohydrate metabolism [18, 47].

To identify networks responsible for differential adaptation to salinity, Gong et al. [35] examined transcript and metabolite abundances of *Arabidopsis* and a closely related salt-tolerant species, *Thellungiella halophila*. They found that the metabolomes of both species were surprisingly similar except for two key differences – *Arabidopsis* had a greater flux of carbon to protein synthesis and *Thellungiella* appeared to have a pre-adaptation strategy to salinity stress that was not obvious or present in *Arabidopsis* [35]. Identifying novel metabolites or pathways that underlie plant adaptation to abiotic stress could lead to the production of crop species more tolerant to salt stress. It can be envisioned that in the near future, we will be able to compare the metabolite responses of different plant species to a range of different stresses, which will in turn allow the detection of metabolites affected by stress in all or most species (species independent) as well as species/genus-specific metabolite alterations (species specific).

3.6 Summary and Future Outlook

Plant metabolism is dazzlingly complex, and its study is one of the most fascinating and fast-exploding areas of biology. This complexity challenges any approach used to understand the detailed metabolic events occurring at the cellular or subcellular level. However, regardless of the difficulties or limitation of metabolomics, it is central to the development of the systems approach, which argues for and embraces the comprehensiveness, interrelationships, and interconnected nature of all levels of organism organization, i.e., gene, transcript, protein, metabolite, and physiology, in an attempt to complete our understanding of the system. Perhaps the greatest hurdle is the impossibility at the current time, of obtaining an accurate description of the total metabolic composition of a plant cell.

We have presented the current state of the art for identifying, measuring, and fingerprinting metabolites in plant cells. Although these analytical methods for assessing plant metabolites are becoming more efficient, repeatable, and even dynamic, the field of metabolomics is still very much in its infancy and we are just on the cusp of realizing its potential. Several areas of this new and burgeoning field need to be addressed; constant refinement and development of new enabling technologies and statistical and data mining tools to support data interpretation will broaden the realm of what is now possible in the plant metabolomic field. However, we stress that regardless of the metabolomic application used or the desired outcome, the need for quality control and rigorous statistical analysis cannot be compromised.

References

1. Abdi H (2007) Bonferroni and Sidak corrections for multiple comparisons. In: Salkind NJ (ed.) Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA.
2. Aharoni A, de Vos CHR, Verhoeven HA, Maliepaard CA, Kruppa G, Bino RJ, Goodenowe DB (2002) Nontargeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry. Omics 6:217–234.
3. Arlt K, Brandt S, Kehr J (2001) Amino acid analysis in five pooled single plant cell samples using capillary electrophoresis coupled to laser-induced fluorescence detection. J Chromatography A 926:319–325.
4. Baker JM, Hawkins ND, Ward JL, Lovegrove A, Napier JA, Shewry PR, Beale MH (2006) A metabolomic study of substantial equivalence of field-grown genetically modified wheat. Plant Biotechnol J 4:381–392.
5. Birkemeyer C, Kopka J (2007) Design of metabolite recovery by variations of the metabolite profiling protocol. In: Nikolau BJ (ed.) Proceedings of the 3rd Congress on Plant Metabolomics. Springer, Dordrecht, The Netherlands.
6. Block E, Naganathan S, Putman D, Zhao SH (1992) *Allium* chemistry – HPLC analysis of thiosulfinates from onion, garlic, wild garlic (Ramsoms), leek, scallion, shallot, elephant (Great-Headed) garlic, chive, and Chinese chive – Uniquely high allyl to methyl ratios in some garlic samples. J Agric Food Chem 40:2418–2430.
7. Bohnert HJ, Jensen RG (1996) Strategies for engineering water-stress tolerance in plants. Trends Biotechnol 14:89–97.

8. Böttcher C, van Roepenack-Lahaye E, Willscher E, Scheel D, Clemens S (2007) Evaluation of matrix effects in metabolite profiling based on capillary liquid chromatography electrospray ionization quadrupole time-of-flight mass spectrometry. Anal Chem 79: 1507–1513.

9. Broadhurst D, Kell DB (2006) Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2:171–196.

10. Broeckling CD, Huhman DV, Farag MA, Smith JT, May GD, Mendes P, Dixon RA, Sumner LW (2005) Metabolic profiling of *Medicago truncatula* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. J Exp Bot 56:323–336.

11. Broeckling CD, Reddy IR, Duran AL, Zhao X, Sumner LW (2006) MET-IDEA: Data extraction tool for mass spectrometry-based metabolomics. Anal Chem 78:4334–4341.

12. Brown M, Dunn WB, Ellis DI, Goodacre R, Handl J, Knowles JD, O'Hagan S, Spasic I, Kell DB (2004) A metabolome pipeline: from concept to data to knowledge. Metabolomics 1:39–51.

13. Bruins AP, Covey TR, Henion JD (1987) Ion spray interface for combined liquid chromatography/atmospheric pressure ionization mass spectrometry. Anal Chem 59:2642–2646.

14. Burrell MM, Earnshaw CJ, Clench MR (2006) Imaging matrix assisted laser desorption ionization mass spectrometry: a technique to map plant metabolites within tissues at high spatial resolution. J Exp Bot 58:757–763.

15. Cairns AJ (1992) A reconsideration of fructan biosynthesis in storage roots of *Asparagus-officinalis* L. New Phytol 120:463–473.

16. Catchpole GS, Beckmann M, Enot DP, Mondhe M, Zywicki B, Taylor J, Hardy N, Smith A, King RD, Kell DB, Fiehn O, Draper J (2005) Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. Proc Natl Acad Sci USA 102:11458–14462.

17. Clark JM, Daum KA, Kalivas JH (2003) Demonstrated potential of ion mobility spectrometry for detection of adulterated perfumes and plant speciation. Anal Lett 36:215–244.

18. Colebatch G, Desbrosses G, Ott T, Krusell L, Montanari O, Kloska S, Kopka J, Udvardi MK (2004) Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. Plant J 39:487–512.

19. Cook D, Fowler S, Fiehn O, Thomashow MF (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. Proc Natl Acad Sci USA 101:15243–15248.

20. De Luca V, St Pierre B (2000) The cell and developmental biology of alkaloid biosynthesis. Trends Plant Sci 361:168–173.

21. Dunn WB, Ellis DI (2005) Metabolomics: Current analytical platforms and methodologies. Trends Anal Chem 24:285–294.

22. Duran AL, Yang J, Wang L, Sumner LW (2003) Metabolomics spectral formatting, alignment and conversion tools (MSFACTs). Bioinformatics 19:2283–2293.

23. Edwards EL, Rodrigues JA, Feffeira J, Goodall DM, Rauter AP, Justino J, Thomas-Oates J (2006) Capillary electrophoresis-mass spectrometry characterisation of secondary metabolites from the antihyperglycaemic plant *Genista tenera*. Electrophoresis 27:2164–2170.

24. Eneroth P, Hellström K, Ryhage R (1964) Identification and quantification of neutral fecal steroids by gas-liquid chromatography and mass spectrometry: studies of human excretion during two dietary regimens. J Lipid Res 5:245–262.

25. Fernie AR, Geigenberger P, Stitt M (2005) Flux an important, but neglected, component of functional glenomics. Current Opin Plant Biol 8:174–182.

26. Fiehn O (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. Comp Funct Genomics 2:155–168.

27. Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol 48:155–171.

28. Fiehn O (2006) Metabolite Profiling in *Arabidopsis*. In: Salinas J, Sanchez-Serrano JJ (eds) Arabidopsis Protocols 2nd edition. Methods in Molecular Biology (323), Humana Press, Totowa NJ, pp. 439–447.

29. Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L (2000) Metabolite profiling for plant functional genomics. Nat Biotechnol 18:1157–1161.

30. Friedman M, Dao L (1992) Distribution of glycoalkaloids in potato plants and commercial potato products. J Agricult Food Chem 40:419–423.

31. Gibon Y, Blaesing OE, Hannemann J, Carillo P, Höhne M, Hendriks JHM, Palacios N, Cross J, Selbig J, Stitt M (2004) A robot-based platform to measure multiple enzyme activities in *Arabidopsis* using a set of cycling assays: Comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. Plant Cell 16: 3304–3325.

32. Gibon Y, Usadel B, Blaesing OE, Kamlage B, Hoehne M, Trethewey RN, Stitt M (2006) Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in *Arabidopsis* rosettes. Genome Biol 7:R76.1–R76.23.

33. Gibon Y, Vigeolas H, Tiessen A, Geigenberger P, Stitt M (2002) Sensitive and high throughput metabolite assays for inorganic pyrophosphate, ADPGlc, nucleotide phosphates, and glycolytic intermediates based on a novel enzymic cycling system. Plant J 30:221–235.

34. Goffard N, Weiller G (2006) Extending MapMan: application to legume genome arrays. Bioinformatics 22:2958–2959.

35. Gong Q, Li P, Ma S, Rupassara SI, Bohnert HJ (2005) Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. Plant J 44:826–839.

36. Goodacre R (2005) Metabolomics – the way forward. Metabolomics 1:1–2.

37. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol 22: 245–252.

38. Griffin JL, Nicholls AW, Keun HC, Mortishire-Smith RJ, Nicholson JK, Kuehn T (2002) Metabolic profiling of rodent biological fluids *via* ^{1}H NMR spectroscopy using a 1 mm microlitre probe. Analyst 127:582–584.

39. Hackel A, Schauer N, Carrari F, Fernie AR, Grimm B, Kuhn C (2006) Sucrose transporter LeSUT1 and LeSUT2 inhibition affects tomato fruit development in different ways. Plant J 45:180–192.

40. Hall RD (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytol 169:453–468.

41. Hodisan T, Culea M, Cimpoiu C, Cot A (1998) Separation, identification and quantitative determination of free amino acids from plant extracts. J Pharmaceutical Biomed Anal 18:319–323.

42. Huhman DV, Sumner LW (2002) Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. Phytochemistry 59:347–360.

43. Ikonomou MG, Blades AT, Kebarle P (1990) Investigations of the electrospray interface for liquid chromatography/mass spectrometry. Anal Chem 62:957–967.

44. Jacobs A, Lunde C, Bacic A, Tester M, Roessner U (2007) The impact of constitutive expression of a moss Na[+] transporter on the metabolomes of rice and barley. Metabolomics 3: 307–317.

45. Jewett MC, Hansen ME, Nielsen J (2007) *Saccharomyces cerevisiae* metabolomics: a driver for developing integrative analytical tools for discerning metabolic function. In: Jewett MC, Nielsen J (eds) Metabolomics. Springer, Heidelberg, Germany.

46. Junker BH, Klukas C, Schreiber F (2006) VANTED: A system for advanced data analysis and visualization in the context of biological networks. BMC Bioinformatics 7: 109.

47. Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL (2004) Exploring the temperature-stress metabolome of *Arabidopsis*. Plant Physiol 136: 4159–4168.

48. Katajamaa M, Miettinen J, Orešic M. (2006) MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics 22:634–636.

49. Katona ZF, Sass P, Molnár-Perl I (1999) Simultaneous determination of sugars, sugar alcohols, acids and amino acids in apricots by gas chromatography–mass spectrometry. J Chromat A 847:91–102.

50. Keurentjes JJB, Fu J, de Vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nat Gen 38:842–849.

51. Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmüller E, Dörmann P, Weckwerth W, Gibon Y, Stitt M, Willmitzer L, Fernie AR, Steinhauser D (2005) GMD@CSB.DB: the Golm metabolome database. Bioinformatics 21:1635–1638.

52. Krishnan P, Kruger NJ, Ratcliffe RG (2005) Metabolite fingerprinting and profiling in plants using NMR. J Exp Bot 56:255–265.

53. Kristensen C, Morant M, Olsen CE, Ekstrøm CT, Galbraith DW, Møller BL, Bak S (2006) Metabolic engineering of dhurrin in transgenic *Arabidopsis* plants with marginal inadvertent effects on the metabolome and transcriptome. Proc Natl Acad Sci USA 102: 1779–1784.

54. Lindon JC, Holmes E, Nicholson JK (2003) So what's the deal with metabonomics? Anal Chem 75:384A–391A.

55. Luedemann A, Weicht D, Selbig J, Kopka J (2004) PaVESy: pathway visualization and editing system. Bioinformatics 20:2841–2844.

56. Matz LM, Dion HM, Hill HH (2002) Evaluation of capillary liquid chromatography-electrospray ionization ion mobility spectrometry with mass spectrometry detection. J Chromat A 946:59–68.

57. Meyer R, Wagner KG (1985) Determination of nucleotide pools in plant-tissue by high-performance liquid-chromatography. Anal Biochem 148:269–276.

58. Moco S, Bino RJ, Vorst O, Verhoeven HA, de Groot J, van Beek TA, Vervoort J, de Vos CHR (2006) A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol 141:1205–1218.

59. Molnár-Perl I, Füzfai Z (2005) Chromatographic, capillary electrophoretic and capillary electrochromatographic techniques in the analysis of flavonoids. J Chromat A 1073: 201–227.

60. Oikawa A, Nakamura Y, Oqura T, Kimura A, Suzuki H, Sakurai N, Shinbo Y, Shibata D, Kanaya S, Ohta D (2006) Clarification of pathway-specific inhibition by Fourier transform ion cyclotron resonance/mass spectrometry-based metabolic phenotyping studies. Plant Physiol 142:398–413.

61. Oksman-Caldentey K-M, Inzé D (2004) Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. Trends Plant Sci 9:433–440.

62. Oliver SG, Winson MK, Kell DB, Baganz F (1998) Systematic functional analysis of the yeast genome. Trends Biotechnol 16:373–378.

63. Paley SM, Karp PD (2006) The pathway tools cellular overview diagram and Omics viewer. Nucleic Acid Res 34:3771–3778.

64. Patrick JW (1997) Phloem unloading: Sieve element unloading and post-sieve element transport. Annual Rev Plant Physiol Plant Mol Biol 48:191–222.

65. Ramautar R, Demirci A, de Jong GJ (2006) Capillary electrophoresis in metabolomics. Trends Anal Chem 25:455–466.

66. Ratcliffe RG, Shachar-Hill Y (2005) Revealing metabolic phenotypes in plants: inputs from NMR analysis. Biol Rev Cambridge Phil Soc 80:27–43.

67. Rathinasabapathi B (2000) Metabolic engineering for stress tolerance: Installing osmoprotectant synthesis pathways. Annals Bot 86:709–716.

68. Roessner-Tunali U, Hegemann B, Lytovchenko A, Carrari F, Bruedigam C, Granot D, Fernie AR (2003a) Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development. Plant Physiol 133:84–99.

69. Roessner-Tunali U, Lui J, Leisse A, Balbo I, Perez-Melis A, Willmitzer L, Fernie AR (2004) Flux analysis of organic and amino acid metabolism in potato tubers by gas chromatography-mass spectrometry following incubation in ^{13}C labelled isotopes. Plant J 39:668–679.

70. Roessner-Tunali U, Urbanczyk-Wochniak E, Czechowski T, Kolbe A, Willmitzer L, Fernie AR (2003b) De novo amino acid biosynthesis in potato tubers is regulated by sucrose levels. Plant Physiol 133:683–692.

71. Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie AR (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell 13:11–29.

72. Roessner U, Patterson JH, Forbes MG, Fincher GB, Langridge P, Bacic A (2006) An investigation of boron toxicity in barley using metabolomics. Plant Physiol 142:1087–1101.

73. Roscher A, Emsley L, Raymond P, Roby C (1998) Unidirectional steady state rates of central metabolism enzymes measured simultaneously in a living plant tissue. J Biol Chem 273:25053–25061.

74. Saito K, Dixon RA, Willmitzer L (2006) Plant Metabolomics, Berlin Heidelberg, Germany, Springer-Verlag.

75. Sato S, Soga T, Nishioka T, Tomita M (2004) Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. Plant J 40:151–163.

76. Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature Biotechnol 24:447–454.

77. Schwender J, Ohlrogge J, Shachar-Hill Y (2004) Understanding flux in plant metabolic networks. Current Opin Plant Biol 7:309–317.

78. Shulaev V, Cortes D, Miller G, Mittler R (2008) Metabolomics for plant stress response. Physiol. Plant 132:199–208.

79. Smith, CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78:779–787.

80. Soga T, Heiger DN (2000) Amino acid analysis by capillary electrophoresis electrospray ionization mass spectrometry. Anal Chem 72:1236–1241.

81. Stein SE (1999) An Integrated Method for Spectrum Extraction and compound identification from GC/MS data. J Amer Soc Mass Spectrom 10:770–781.

82. Stoop JMH, Williamson JD, Pharr DM (1996) Mannitol metabolism in plants: a method for coping with stress. Trends Plant Sci 1:139–144.

83. Styczynski MP, Moxley JF, Tong LV, Walther JL, Jensen KL, Stephanopoulos GN (2007) Systematic identification of conserved metabolites in GC/MS Data for metabolomics and biomarker discovery. Anal Chem 79:966–973.

84. Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 62:817–836.

85. Tang KQ, Li FM, Shvartsburg AA, Strittmatter EF, Smith RD (2005) Two-dimensional gas-phase separations coupled to mass spectrometry for analysis of complex mixtures. Anal Chem 77:6381–6388.

86. Tarpley L, Duran AL, Kebrom TH, Sumner LW (2005) Biomarker metabolites capturing the metabolite variance present in a rice plant developmental period. BMC Plant Biol 5:8.

87. Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37:914–939.

88. Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J-I, Awazuhara M, Inoue E, Taka-
    hashi H, Goodenowe DB, Kitayama M, Noji M, Yamazaki M, Saito K (2005) Functional
    genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants
    over-expressing an MYB transcription factor. Plant J 42:218–235.
89. Tolstikov VV, Fiehn O (2002) Analysis of highly polar compounds of plant origin: combina-
    tion of hydrophilic interaction chromatography and elctrospray ion mass trap spectrometry.
    Anal Biochem 301:298–307.
90. Trethewey RN (2004) Metabolite profiling as an aid to metabolic engineering in plants. Curr
    Opin Plant Biol 7:196–201.
91. Urbanczyk-Wochniak E, Usadel B, Thimm O, Nunes-Nesi A, Carrari F, Davy M, Bläsing O,
    Kowalczyk M, Weicht D, Polinceusz A, Meyer S, Stitt M, Fernie AR (2006) Conversion of
    MapMan to allow the analysis of transcript data from Solanaceous species: Effects of genetic
    and environmental alterations in energy metabolism in the leaf. Plant Mol Biol 60:773–792.
92. Urbanczyk-Wochniak E, Sumner LW (2007) MedicCyc: a biochemical pathway database
    for *Medicago truncatula*. Bioinformatics 23:1418–1423.
93. Valentine SJ, Kulchania M, Barnes CAS, Clemmer DE (2001) Multidimensional separa-
    tions of complex peptide mixtures: a combined high-performance liquid chromatography/ion
    mobility/time-of-flight mass spectrometry approach. Internat J Mass Spectrom 212:97–109.
94. Valentine SJ, Liu XY, Plasencia MD, Hilderbrand AE, Kurulugama RT, Koeniger SL, Clem-
    mer DE (2005) Developing liquid chromatography ion mobility mass spectometry tech-
    niques. Exp Rev Proteomics 2:553–565.
95. Verpoorte R (1998) Exploration of nature's chemodiversity: the role of secondary metabo-
    lites as leads in drug development. Drug Discovery Today 3:232–238.
96. Villas-Bôas SG, Roessner U, Hansen M, Smedsgaard J, Nielsen J (2007) Metabolome Anal-
    ysis: An Introduction, New Jersey, NJ, USA, John Wiley & Sons, Inc.
97. Wagner C, Sefkow M, Kopka J (2003) Construction and application of a mass spectral
    and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles.
    Phytochemistry 62:887–900.
98. Wang SY, Kuo YH, Chang HN, Kang PL, Tsay HS, Lin KF, Yang NS, Shyur LF (2002)
    Profiling and characterization antioxidant activities in *Anoectochilus formosanus* Hayata. J
    Agricult Food Chem 50:1859–1865.
99. Williams BJ, Cameron CJ, Workman R, Broeckling CD, Sumner LW, Smith JT (2007)
    Amino acid profiling in plant cell cultures: An inter-laboratory comparison of CE-MS and
    GC-MS. Electrophoresis 28:1371–1379.
100. Wurtele ES, Li J, Diao L, Zhang H, Foster CM, Fatland B, Dickerson J, Brown A, Cox Z,
     Cook D, Lee E-K, Hoffman H (2003) MetNet: software to build and model the biogenetic
     lattice of *Arabidopsis*. Comp Funct Genomics 4:239–245.
101. Yamashita M, Fenn JB (1984) Electrospray ion source. Another variation on the free-jet
     theme. J Phys Chem 88:4451–4459.
102. Yancey PH (2005) Organic osmolytes as compatible, metabolic and counteracting cytopro-
     tectants in high osmolarity and other stresses. J Exp Biol 208:2819–2830.
103. Zhao Q, Stoyanova R, Du S, Sajda P, Brown TR (2006) *HiRes* – a tool for comprehensive
     assessment and interpretation of metabolomic data. Bioinformatics 22:2562–2564.
104. Zywicki B, Catchpole G, Draper J, Fiehn O (2005) Comparison of rapid liquid
     chromatography-electrospray ionization-tandem mass spectrometry methods for determina-
     tion of glycoalkaloids in transgenic field-grown potatoes. Anal Biochem 336:178–186.