I.1 Gas Chromatography Mass Spectrometry

J. Kopka[1]

1 Introduction

GC-MS technology has been used for decades in studies which aim at the exact quantification of metabolite pool size and metabolite flux. Exact quantification has traditionally been focused on a single or small set of predefined target metabolites. Today GC-MS is one of the most widely applied technology platforms in modern metabolomic studies. Since early applications in unravelling the mode of action of herbicides (Sauter et al. 1988) it has experienced a renaissance (Fig. 1) in post-genomic, high-throughput fingerprinting and metabolite profiling of genetically modified (e. g. Roessner et al. 2001a,b, 2002; Fernie et al. 2004) or experimentally challenged plant samples (e. g. Cook et al. 2004; Kaplan et al. 2004; Urbanczyk-Wochniak and Fernie 2005). Metabolic phenotyping and analysis of respective phenocopies by metabolite profiling has become an integral part of plant functional genomics (Fiehn et al. 2000b; Roessner et al. 2002; Fernie et al. 2004). The essence of metabolite profiling, namely the non-biased screening of biological samples for changes of metabolite levels relative to control samples, has been thoroughly discussed earlier and is clearly distinguished from fingerprinting approaches and the concept of exact quantification (Fiehn et al. 2000b; Sumner et al. 2003; Birkemeyer et al. 2005).

GC-MS-based metabolome profiling analysis is on the verge of becoming a routine technology. This fact substantially contributes to the development of metabolomics as a fourth integral part of the Rosetta stone for functional genomics and molecular physiology (Trethewey et al. 1999; Fiehn et al. 2000b; Trethewey 2004). Nevertheless, GC-MS technology is already challenged again by new bottlenecks and demands for improved data sets which are optimised for the mathematical modelling tools currently developed in the fields of bioinformatics and biological systems analysis.

The challenges of modern, multi-parallel, GC-MS based metabolite analysis are manifold: (i) automation of sample preparation, wet chemistry and data processing after acquisition for increased throughput and reproducibility, (ii) extension of the analytical scope of metabolomics studies, for example by combined analysis of single samples using multiple analytical technology platforms, and combined analysis with the proteome and transcriptome

[1] Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany, e-mail: Kopka@mpimp-golm.mpg.de

Biotechnology in Agriculture and Forestry, Vol. 57
Plant Metabolomics (ed. by K. Saito, R.A. Dixon, and L. Willmitzer)

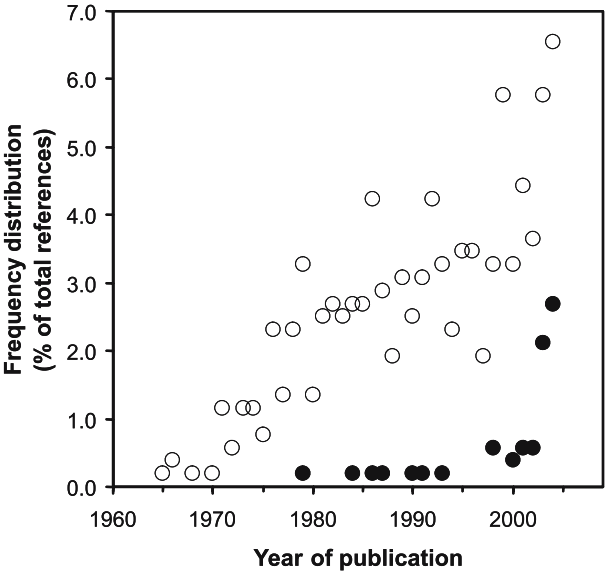


Fig. 1. Literature survey of publications which associate the concepts, "metabolite", "profiling", and "gas chromatography" performed on 1/2005. A total of ~500 citations without conference proceedings, abstracts and book chapters were found. The frequency of publications in all biological sciences (*open circles*) is compared to the contribution by plant metabolomics community (*closed circle*)

(Weckwerth et al. 2004b), (iii) profiling of trace compounds, or signalling molecules in the presence of bulk metabolites (Mueller et al. 2002; Birkemeyer et al. 2003; Schmelz et al. 2003, 2004), (iv) increasing accuracy in multi-parallel metabolite quantification (Birkemeyer et al. 2005), (v) combining profiling and flux analyses (Roessner-Tunali et al. 2004), (vi) establishment of quantitative repeatability, unambiguous nomenclature and comparability between analyses performed in different laboratories or using different analytical technology platforms (Schauer et al. 2005), and (vii) finally – perhaps the most important challenge of all metabolomic investigations – the identification of the unidentified majority of metabolic components from metabolite profiling experiments (Fiehn et al. 2000a; Schauer et al. 2005).

In agreement with the focus of this chapter the above challenges have predominantly analytic or technical motivation. The breakthrough of metabolomic investigations, however, will depend on the access to hitherto unavailable fundamental insights into metabolic and systems interactions. Increasingly integrative studies which consider the metabolome, proteome, transcriptome, and genome evolution of an organism have been initiated and are to be expected. Promising steps have been made – using GC-MS technology – towards network analysis (Fiehn 2003; Weckwerth et al. 2004a) and correlation studies between or within metabolome and transcriptome

constituents (Urbanczyk-Wochniak et al. 2003; Steinhauser et al. 2004; Kopka et al. 2005). A detailed discussion of these general aspects including GC-MS studies and beyond can be found in the applications section of this book.

2 GC-MS Profiling Technology in a Nutshell

Metabolite profiling with GC-MS involves six general steps:

1. *Extraction* of metabolites from the biological sample, which should be as comprehensive as possible, and at the same time avoid degradation or modification of metabolites (e. g. Kopka et al. 2004).
2. *Derivatisation* of metabolites making them amenable to gas chromatography. Metabolites which are not volatile per se require chemical modification prior to GC analysis.
3. *Separation* by GC. High resolution GC can also be highly reproducible as it involves automated sample injection robotics, highly standardised conditions of gas-flow, temperature programming, and standardised capillary column material.
4. *Ionisation* of compounds as they are eluted from the GC. Electron impact (EI) ionisation is most widely used, as it is the technology which is least susceptible to suppression effects and produces reproducible fragmentation patterns.
5. Time resolved *detection* of molecular and fragment ions. Mass separation and detection can be achieved with different mass-detection devices, including sector field detectors, quadrupole detectors (QUAD), ion trap technology, and time-of-flight detectors (TOF). The choice of detectors depends on the targeted analytical niche. GC-MS systems with QUAD detection are most widely spread for routine analysis. Ion trap technology allows MS×MS (two-dimensional MS) analysis for structural elucidation and targeted quantification of trace compounds (e. g. Mueller et al. 2002). TOF detection can either be tuned to fast scanning rates (van Deursen et al. 2000) or to high mass precision comparable to sector field systems. Fast scanning GC-TOF-MS enables the, today, most advanced technology in the GC-MS field, namely two dimensional GC×GC-TOF-MS (two-dimensional GC-TOF-MS) (Ryan et al. 2004; Sinha et al. 2004a–c).
6. *Acquisition and evaluation* of GC-MS data files. All GC-MS system manufacturers provide software which is tuned for targeted, quantitative metabolite analysis. The targeted approach involves unequivocal identification of predefined metabolites by expected chromatographic retention times and mass-spectral fragmentation patterns and quantitative calibration by authentic standard concentrations. Recent software developments support the non-targeted analysis of GC-MS patterns, and the full evaluation of all resolved compounds. This feature of GC-MS allows discovery of novel hitherto

unknown metabolites. As we are far from knowing all possible metabolites of a given organism, non-biased, truly comprehensive data evaluation is the most essential requirement of metabolite profiling.

2.1 Chemical Derivatisation and Chromatography

The principles of fast metabolic sample inactivation and nondestructive extraction are common to all metabolome analyses. In contrast to all other technologies GC-MS is inherently restricted to volatile and temperature-stable compounds. The scope of GC-MS for metabolite analysis is limited by the typical temperature range of commercial capillary columns, for example up to 320–350 °C. The lower temperature range is determined by ambient temperature, but cold trapping devices and isothermal GC allow analysis of low molecular weight gases and highly volatile metabolites. GC received a considerable extension of applications through the development of a highly versatile tool box of derivatisation reagents, which chemically transform non-volatile metabolites into volatile analytes for GC-MS analysis (e. g. Knapp 1979; Blau and Halket 1993; Toyo'oka 1999). To date, GC-MS profiling of metabolites in plants has largely been confined to compounds, recovered in the methanol-water phase after methanol-water/chloroform extraction of tissues (Fiehn et al. 2000a; Roessner et al. 2000; Duran et al. 2003; Barsch et al. 2004; Gullberg et al. 2004; Strelkov et al. 2004; Broeckling et al. 2005). Although not all hydrophilic compounds can be volatilised by derivatisation, the following classes of compounds are detected routinely: amino-, organic-, and aromatic-acids, amines, sugars up to trisaccharides, alcohols and polyols, and some mono-phosphorylated metabolites.

The current limitations of metabolite preparation and derivatisation strategy, namely methoxyamination with subsequent direct trimethylsilylation of predominantly polar metabolites, call for extension. Application of other technology platforms is an obvious route and will be discussed in the following chapters. Here a short appraisal of the potential of chemical derivatisation is attempted. Four main types of reaction schemes will be discussed.

1. *Alkoxyamination* by reagents, such as methoxyamine CH_3-O-NH_2, stabilises carbonyl moieties in native metabolite structures, but forms E- and Z-isomers of the $-N=C<$ double-bond substituents. Keto-enol tautomerism is suppressed, as is the decarboxylation of unstable β-carbonyl-carboxylic acids. In addition, the formation of acetal- or ketal-structures in aqueous solution is inhibited. These equilibrium reactions generate multiple intramolecular and water adducts, for example the typical α- and β-conformers of reducing sugars. Ether- and ester-conjugates are mostly stable when exposed to methoxyamine reagent and maintain conformation. So far other alkoxy-reagents – for example hydroxylamine, ethyloxyamine, or benzyloxyamine – have not been exploited for systematic discovery of metabolites with carbonyl moieties:

2. *Silylation* reagents classify into those which introduce either a trimethylsilyl (TMS) moiety, $-Si(CH_3)_3$, or a dimethyl-(*tert*-butyl)-silyl (TBS) moiety, $-Si(CH_3)_2-C(CH_3)_3$. TMS reagents have been well investigated and are known to have the widest derivatisation spectrum (Little 1999; Halket et al. 2005). TMS has the potential to substitute all exchangeable, "acidic" protons of a metabolite. Steric hindrance of TMS substitution is rare but common with the bulkier TBS reagent. The benefit of the TBS reagent is higher tolerance for the presence of water and clear mass spectral fragmentation. However, vicinal diols, which typically occur in sugars, are only partially derivatised.
3. *Alkylation* reactions, mostly methylation, are widely used to derivatise carboxylic acids and alcohols. The enormous reactivity of available reagents – some allow for flash derivatisation during hot GC injection – leads to transalkylation of ester-bonds und consequently breaks down complex metabolites, such as glycero- and phospholipids. Alkylation of sugars leads to derivatives which are more volatile than the TMS derivatives and therefore allow analysis of higher sugar oligomers.
4. *Acylation* reactions, mostly acetylation or trifluoro-acetylation, are less reactive than transalkylation. Reagents usually form stable ester and amide bonds and break down only activated metabolic intermediates, e. g. thioesters.

In conclusion further developments of alternate GC-MS profiling techniques need to employ more selective combinations of metabolite fractionation and derivatisation schemes. Solid phase extraction can be explored to partition and concentrate metabolites amenable to alternate subsequent derivatisation. On the other hand, vapour phase extraction (VPE) for the separation and concentration of volatile derivatisation products prior to GC injection may prove promising (Schmelz et al. 2003, 2004). VPE has the potential to be a robust technique and was shown to operate with a range of commonly used reagents.

2.2 Mass Detection and Quantitative Calibration Techniques

One of the major criticisms and pitfalls of metabolome analyses is best explained by so-called matrix effects. This well-known effect describes unexpected losses or increased recovery of metabolites in complex extracts compared to pure authentic preparations. Matrix effects on one hand are caused by the presence of compounds which either specifically inhibit extraction or chemical analysis of metabolites. Positive matrix effects can stabilise otherwise labile compounds in the presence of suitable chemicals. Typical examples are suppression effects of soft ionization techniques, for example electrospray ionization (ESI) or matrix assisted laser desorption ionization (MALDI). Electron-impact ionization (EI) typically used in GC-MS profiling is not susceptible to suppression. Instead GC injection is the crucial step which may

cause discriminations, especially in view of the complex and rather crude extracts which are typically injected.

So far, only exemplary – albeit time demanding – thorough tests for unexpected matrix effects have been performed with selections of chemically

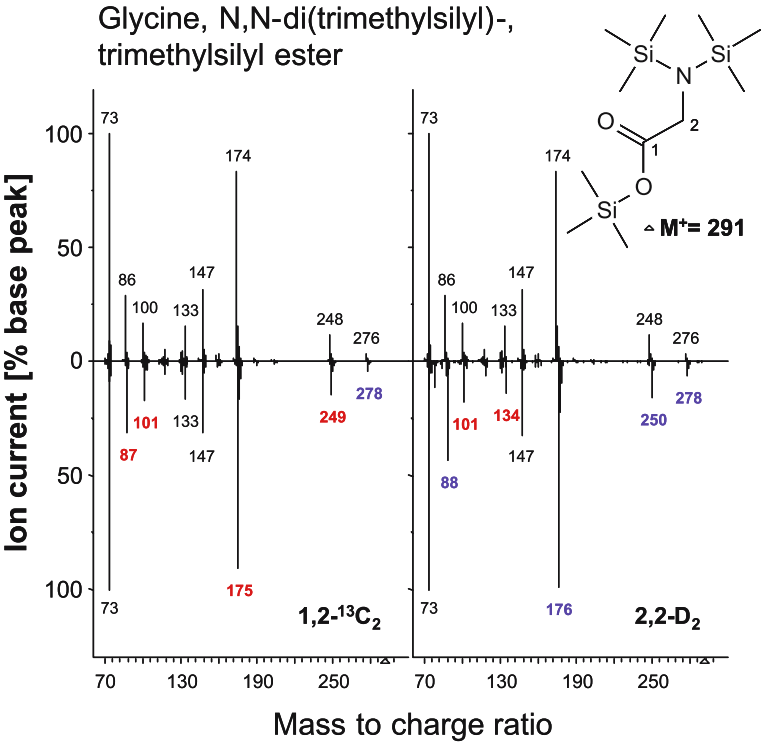
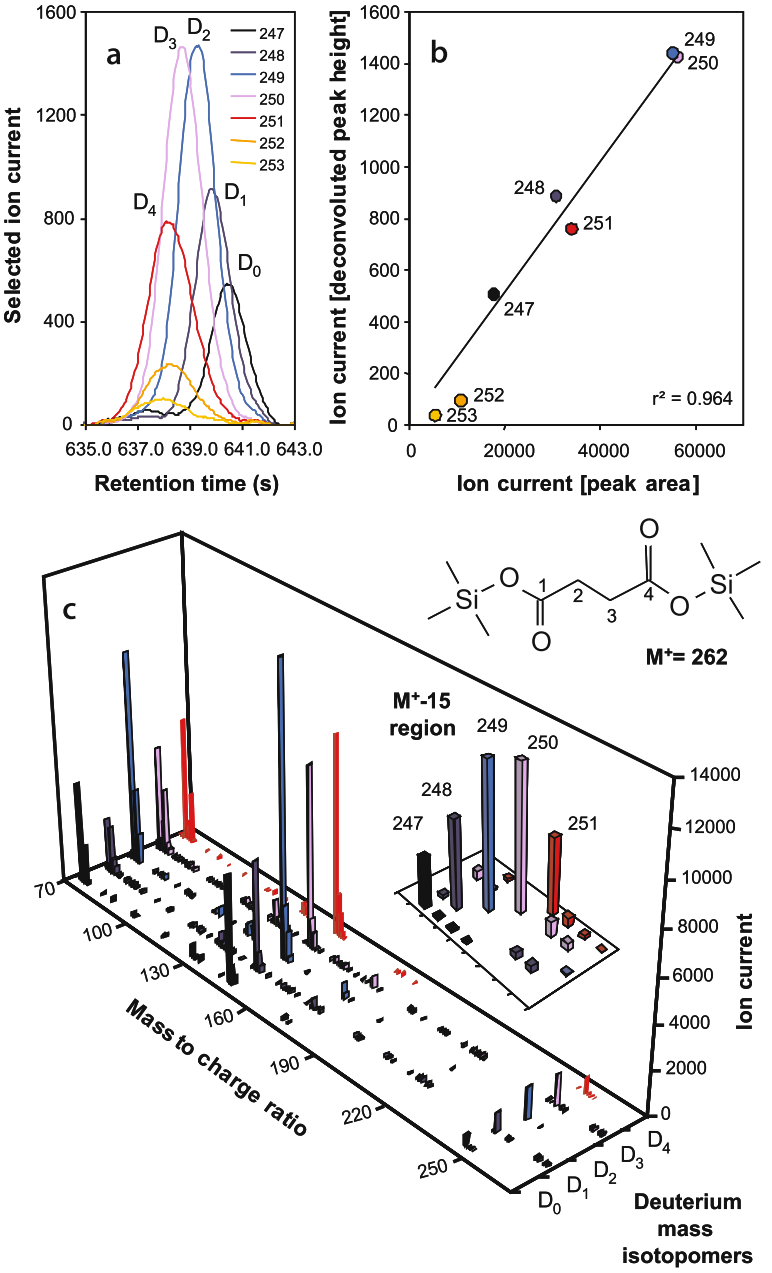


Fig. 2. Mass spectra of deuterated and ^{13}C labeled MSTs help structural elucidation and recovery analysis of metabolites. Labeled and non-labeled MSTs of Glycine *N*,*N*-di(trimethylsilyl)-, trimethylsilyl ester are shown. *Oryza sativa* L. cv. Nipponbare was labelled in vivo using deuterated water or $^{13}CO_2$. MSTs representing the fully labeled mass isotopomers demonstrate presence of two carbon atoms (*left panel*) and two non-exchangeable hydrogen atoms (*right panel*). Mass fragments which exhibited a mass shift of 1 amu (*red*) or 2 amu (*blue*) are indicated

▶ **Fig. 3a–c.** Mass spectral deconvolution of deuterated mass isotopomers. Succinic acid di(trimethylsilyl) ester was partially labelled in vivo by exposing *Oryza sativa* L. cv. Nipponbare to deuterated water. Metabolite profiles were performed on a Pegasus II GC-TOF-MS system (LECO, St. Joseph, MI, USA) with 20 scans s^{-1}. Mass spectra were deconvoluted using ChromaTOF software version 1.00, with baseline offset just above noise, smoothing and peak width set to 10 and 2 scans, respectively: **a** selected ion traces of non-deuterated (D_0, m/z = 247) and deuterated (D_{1-4}, m/z = 248 – 251) $M^+ - 15$ mass fragments. Mass fragments at 252 and 253 amu are carbon mass isotopomers of D_4; **b** peak area compared to deconvoluted peak height. Peak area integration does not allow differentiation of contributions by carbon mass isotopomers; **c** deconvoluted mass spectra of D_{0-4}. *Inset* shows partial deconvolution of D_{0-4} carbon mass isotopomers and missing carbon mass isotopomers of D_{1-3}

diverse, representative metabolites (e. g. Roessner-Tunali et al. 2003; Gullberg et al. 2004). Therefore technologies are required to improve quantitative standardisation for the comparison of increasingly diverse biological samples and experimental conditions.



For this purpose, full saturating ^{13}C in vivo labelling was developed using yeast which is one of the most important organisms in systems biology (e. g. Stephanopoulos et al. 2004). Metabolites of yeast were demonstrated to be fully labelled when provided with an exclusive carbon source, such as U-^{13}C-glucose (Mashego et al. 2004; Birkemeyer et al. 2005). Refer to Birkemeyer et al. (2005) for detailed discussion of potential applications for ^{13}C-labelled metabolomes. Similar approaches are possible in plants (Figs. 2 and 3).

In short, standardised in vivo labelled extracts of yeast or other microorganisms can substitute the rather small number of chemically synthesised mass isotopomers used in earlier studies (Fiehn et al. 2000a; Gullberg et al. 2004). Typically a standardised labelled reference sample is combined in equal amounts with non-labelled experimentally challenged samples. The advantages of this approach are (i) the presence of a mass isotopomer for all identified but also all hitherto non-identified metabolites, (ii) the concentration of each mass isotopomer is inherently adjusted to the endogenous metabolite concentration, (iii) metabolic components can easily be distinguished from laboratory contaminations, and (iv) recovery of all metabolic components can be determined with the appropriate mass isotopomer.

Thus metabolite profiling will achieve the same level of transcriptome and proteome experiments, which utilize differential fluorescent probes or differential isotope coded tagging, respectively. In conclusion, comprehensive in vivo isotope labelling will help to establish quantitative between laboratory comparability of GC-MS based metabolome experiments. More importantly, we expect metabolome experiments with full mass isotopomer standardisation to be also independent of the mass spectrometric platform, e. g. CE-MS, LC-MS, or possibly even MALDI-TOF-MS.

3 Short Excursion into Nomenclature and Definitions

Concise and unambiguous description of GC-MS metabolite profiling results requires clear definitions. The definitions suggested within this section are biased towards the specifics of GC-MS technology but may also be applied to other technology platforms. This section is intended as a contribution to the ongoing process of unifying data formats and concepts within the field of plant metabolomics (e. g. Fiehn 2002; Bino et al. 2004; Jenkins et al. 2004).

3.1 Metabolite and Analyte

Routine GC-MS profiling analysis (Fiehn et al. 2000b; Roessner et al. 2000) has an upper size exclusion limit which is roughly equivalent to a persilylated trisaccharide derivative (MW:1296), hexatriacontane (MW:506), or hentriacontanoic acid trimethylsilylester (MW:523). Even though it may appear tempting, metabolite and analyte are best not defined by molecular weight.

A *metabolite* may be described as a compound which is internalised, chemically converted or secreted by an organism, but is not synthesised by DNA replication, transcription, or translation. Post-processing events of DNA, RNA and proteins, such as DNA methylation, RNA splicing, sequence specific protease cleavage or post-translational modification are not attributed to the metabolome. The origin of a metabolite is not exclusively dependent on the biosynthetic capacity of an organism or delimited by the genomic inventory. Metabolites may readily be exchanged between organisms, for example in plant microbe interactions, and – like drugs or pesticides – can today be of anthropogenic/xenobiotic origin.

In contrast to LC- or CE-MS, GC-MS analysis requires clear distinction between metabolite and analyte, because – depending on choice of chemical derivatisation – metabolites may be chemically transformed before quantification. The term *analyte* may be used to address the chemical structure and compound which is submitted to GC-MS and finally detected and quantified. An analyte can be identical with the metabolite, if the metabolite is not chemically derivatised. Single metabolites may have more than one analyte, if the chosen derivatisation reaction generates more than one derivative, for example methoxyamination (see above). In these cases *preferred* and *alternate* analytes exist for quantification. Analytes of one metabolite may differ in abundance, i. e. a *major* and one, even multiple, *minor* analytes may exist. Standardisation by stable mass isotopomers corrects the quantification errors which may arise from unforeseen matrix effects on analyte ratios during chemical derivatisation of GC injection.

Different metabolites may be chemically transformed into the same analyte structure. In addition a single analyte may arise from inadequate chromatographic separation of isomers. For example, the biochemically distinct stereoisomeric structures of DL-amino acids are only separated by specialised chiroselective chromatography. These analytes have *composite* properties in contrast to absolutely *specific* analytes.

These concepts are not unique to GC-MS technology. Analyte sensitivity, accuracy, and potentially composite analyte properties need to be thoroughly considered in MS-MS applications, non-chiroselective capillary electrophoresis or liquid chromatography, and in cases of adduct-formation or multiply charged ions.

3.2 Mass Spectral Tag (MST) and Mass Fragment

GC-MS metabolite profiles resolve hundreds of analytes, which represent metabolites, but also internal standard substances and laboratory contaminations. Typical GC-MS profiles may contain approximately 100 identified analytes of metabolites. The chemical structure of the majority of GC-MS analytes, however, is still unknown. Each new biological object or experimental condition still gives rise to new, hitherto unidentified, chemical components.

Because in non-biased analysis of GC-MS profiles identified and unidentified components are equally important, we created the term *mass spectral tag* (MST), i. e. a mass spectrum which is characterised by a specific chromatographic retention and by repeated occurrence in a single or multiple types of biological samples (Colebatch et al. 2004; Desbrosses et al. 2005). MSTs represent analytes. MSTs can be identified, in other words, unequivocally linked to a chemical structure. The use of MSTs allows uncoupling of metabolite profiling experiments from the time consuming process of chemical identification. MSTs can be used to track analytes in different experiments or laboratories (Schauer et al. 2005). Thus MST identification can be performed even years after the first discovery.

MSTs of GC-EI-MS profiles are composed of multiple characteristic *mass fragments* in constant relative abundances. In most cases residual, non-fragmented molecular ions are rare or even absent. In consequence GC-MS allows selection of multiple mass fragments which all represent the same MST and exhibit the same quantitative changes. Typically one quantifying mass fragment (QM) and a set of specific, supporting qualifying mass fragments are selected in GC-MS analysis (Halket et al. 2005). The criteria for the proper choice of QMs are equal to the choice of a preferred analyte. QMs need to be selective, i. e. not composite, in the context of the complexity of co-eluting MSTs. Therefore, the best QM is the most abundant among the available selective mass fragments.

3.3 Response and Relative Quantification of Metabolite Pools

GC-MS metabolite profiling studies monitor relative changes in metabolite pool sizes and but also allow insight into flux, i. e. the dynamic turnover of metabolite pools or metabolite substructures (e. g. Fischer and Sauer 2003; Sauer 2004; Roessner-Tunali et al. 2004). Flux experiments are easily distinguished from above mentioned saturating in vivo labelling experiments. Flux experiments monitor the initial kinetics of labelling and thus stable isotopes are only partially incorporated into metabolite pools. In contrast, saturating in vivo labelling reaches the endpoint of a completely stable isotope labelled metabolome.

MSTs are quantified by ion currents of QMs which are recorded after analyte ionization, fragmentation and mass separation. Ion currents in GC-MS are monitored either by peak area or peak height. Both measurements need to be baseline corrected for electronic and chemical noise. The resulting corrected values are defined to be what we call *responses*, i. e. X_{QM} of fragment QM (Colebatch et al. 2004; Desbrosses et al. 2005). The fragment response is routinely normalised to the amount of the sample, for example fresh or dry weight. In addition each response is corrected for recovery effects, which may occur at any step of the analytical process between metabolic inactivation of the sample and final recording of ion currents. Different levels of recovery correction exist: (i) correction by extract and sample volume, (ii) correction

by addition of a constant amount of a representative internal standard compound (IS), and (iii) normalization by chemically identical, but stable-isotope labelled mass isotopomers of each metabolite. The *normalised response* (N_{QM}) is, consequently, $N_{QM} = X_{QM} \times X_{IS}^{-1} \times \text{sample weight}^{-1}$, where X_{IS} ideally represents a mass isotopomer response of QM. In a further step, the normalised response of a fragment, N_{QM}, is divided by the average relative response of QM as determined in a set of reference samples, $\text{avg}N_{QM(ref)}$. The resulting quotient, $R_i = N_{QM} \times \text{avg}N_{QM(ref)}^{-1}$, is called *response ratio* R_i. R_i describes the x-fold changes in metabolite pools sizes relative to the reference samples. Typical reference samples are taken at the start of a time series experiment or are mock-treated biological controls.

In GC-MS profiling analyses the standard deviation of normalised responses is dependent on the chemical nature of metabolite and analyte. Average relative standard deviations (RSD) of 10% (Weckwerth et al. 2004b) or 13.8% (5.5–33.4%; Gullberg et al. 2004) were reported for replicate GC-MS analyses. These analyses included extraction as well as derivatisation and were performed using representative analytes. Use of isotope labelled standardisation was reported to reduce RSD further to approximately 6.9–9.7% residual experimental variance (Gullberg et al. 2004).

4 Present Challenges of GC-MS Profiling

4.1 Standardisation of GC-MS Systems

GC-MS profiles, with the exception of GC×GC-TOF-MS data, are in essence three-dimensional and comprise a chromatographic time-resolved axis, a second coordinate axis which represents the mass to charge ratio (m/z, z = 1 in GC-MS with only rare exceptions), and an intensity axis which monitors the ion current (IC) and thus the abundance of molecules or mass fragments. A substantial breakthrough for GC-MS analyses was the early establishment of generally accepted calibration substances and procedures, so-called tuning routines, which allowed comparison of mass spectra from GC-MS systems of virtually all manufacturers and from different hyphenated mass detection technologies. In addition the widely used electron-impact ionisation technique (EI) ensured stable fragmentation ratios, which are in first approximation independent of analyte concentration. However, comparability was only achieved by restriction to 1 amu precision.

The chromatography axis is less standardised, not least because of multiple types of available capillary GC-columns which have different chromatographic properties and thus serve different separation problems. In addition slight changes in temperature program, pressure and flow settings of both carrier gas and injection technique, as well as slight production differences of column manufacturers cause minor but perceptible changes in retention

time. Retention time indices (RI), based on homologous series of internal reference substances, such as *n*-alkanes, have been introduced early to aid GC analyses (Kovàts 1958). Use of an *n*-alkane RI system in GC-MS metabolite profiling substantially improves the reproducibility of the chromatography axis. The currently achievable accuracy of RI prediction was recently investigated in three different profiling laboratories which use the same type of capillary column but different GC-MS systems (Schauer et al. 2005). In this investigation the possibility of predicting RIs of more than 100 identified analytes was tested. Mathematical regression resulted in an average accuracy of ±5.4 RI units.

The IC intensity axis in GC-MS is standardised for high vs low mass discrimination. The GC-MS tuning includes processes which ensure constant ratios of high vs low mass intensities. However, mass spectra which are recorded by either QUAD-MS or fast scanning GC-TOF-MS detection may differ in this respect. Fast scanning GC-TOF-MS systems (e. g. Pegasus II MS system, LECO, St. Joseph, MI, USA) have increased sensitivity of small mass fragments and reduced sensitivity in the high mass range.

4.2 Deconvolution and Alignment of Mass Spectral Tags

The principal challenge in GC-MS profiling analysis is the automated unravelling of the multiple partially co-eluting MSTs which comprises a GC-MS chromatogram. One of the fundamental advances in GC-MS technology has been the development of algorithms and software for the so-called deconvolution of mass spectra from GC-MS chromatograms (Halket et al. 1999; Stein 1999; Shao et al. 2004; AnalyzerPro, http://www.spectralworks.com), specific software for the deconvolution of fast scanning GC-TOF-MS data files, e. g. ChromaTOF software used by Vreuls et al. (1999), Veriotti and Sacks (2001) or Jonsson et al. (2005), and ongoing developments for automated processing of GC×GC-TOF-MS chromatograms (Ryan et al. 2004; Sinha et al. 2004a,b). The inherent steps of deconvolution are (i) mass resolved baseline subtraction of electronic and chemical noise, (ii) assignment of retention time and/or retention time indices (RI) to chromatographic peak apices and respective MSTs, and (iii) accurate separation of MSTs from closely co-eluting analytes, the most challenging and advanced but still error-prone part (Fig. 3).

Even though automated mass spectral deconvolution has fundamentally facilitated GC-MS analyses of complex mixtures, accuracy and limitations of respective software have so far not been systematically compared and assessed. Typical errors of mass spectral deconvolution are (i) accidental generation of MSTs due to noise fluctuations, (ii) deconvolution of multiple MSTs from a single component, (iii) incomplete MSTs which lack one or multiple mass fragments (Fig. 3c), and (iv) chimeric MSTs, i. e. composite mass spectra of co-eluting compounds. The co-elution problem of complex mixtures has been fundamentally improved by introduction of fast scanning GC-TOF-MS and is

today technically best solved by GC×GC-TOF-MS, using a set of two capillary GC columns with alternate phase-polarity (Sinha et al. 2004c).

Reliable alignment of identical MSTs in sets of consecutive GC-MS chromatograms is required for rapid, repeatable and automated comparative high-throughput analysis of large samples sets. So far, software solutions and novel algorithm developments for the alignment of complex mixtures depend on close to constant chromatographic retention within series of consecutive GC-MS chromatograms (Duran et al. 2003; Jonsson et al. 2004; metAlign, http://www.metalign.nl). Indeed, consistent run-to-run retention times are considered to be crucial to the application of chemometrics on complex mixtures, especially in the field of two-dimensional separations (Sinha et al. 2004c).

In conclusion, automated mass spectral deconvolution of GC-MS profiles appears to be in principal solved by both GC×GC technology and deconvolution algorithm, but the optimum solution still has to be found (Halket et al. 2005). In contrast prediction of chromatographic shifts in complex mixtures with highly dynamic range of concentrations has not been satisfyingly solved. As there is currently no solution – other than recalibration with pure standard substances – addressing the problem of RI shifts will be crucial for future GC-MS based metabolite profiling and identification of MSTs.

4.3 Identification of Mass Spectral Tags

Identification of MSTs requires chromatographic separation as well as mass spectrometric information (Wagner et al. 2003), mainly because plants like microorganisms contain a multitude of isomeric metabolites (e. g. Barsch et al. 2004; Stephanopoulos et al. 2004; Strelkov et al. 2004). These isomers give rise to MSTs, which can be chromatographically resolved but have almost identical mass spectra. Today, GC-MS appears to have found a generally accepted standard for mass spectral comparison. The NIST mass spectral search and comparison software (Ausloos et al. 1999; Stein 1999) has been integrated into the customised operating software of most GC-MS manufacturers. The GC-MS technology is in this respect more advanced than LC-MS (Halket et al. 2005). However, mass spectral search and comparison software, which harbours information on chromatographic retention in what we suggested to call MSRI libraries (Wagner et al. 2003; Kopka et al. 2005) and the automated utilization of this information for probability based matching, would be highly desirable. The new version NIST05 (National Institute of Standards and Technology, Gaithersburg, MD, USA) of a mass spectral search and comparison software now makes RI information available but currently does not utilize RI for automated matching. The result of a hitherto manual inventory of *Oryza sativa* L. cv. Nipponbare leave profiles is shown in Fig. 4.

Two different approaches exist for the identification of unknown MSTs from GC-MS profiles: (i) the “bottom up” approach in which metabolites of interest to a particular researcher are analysed by the purchase of authentic standard

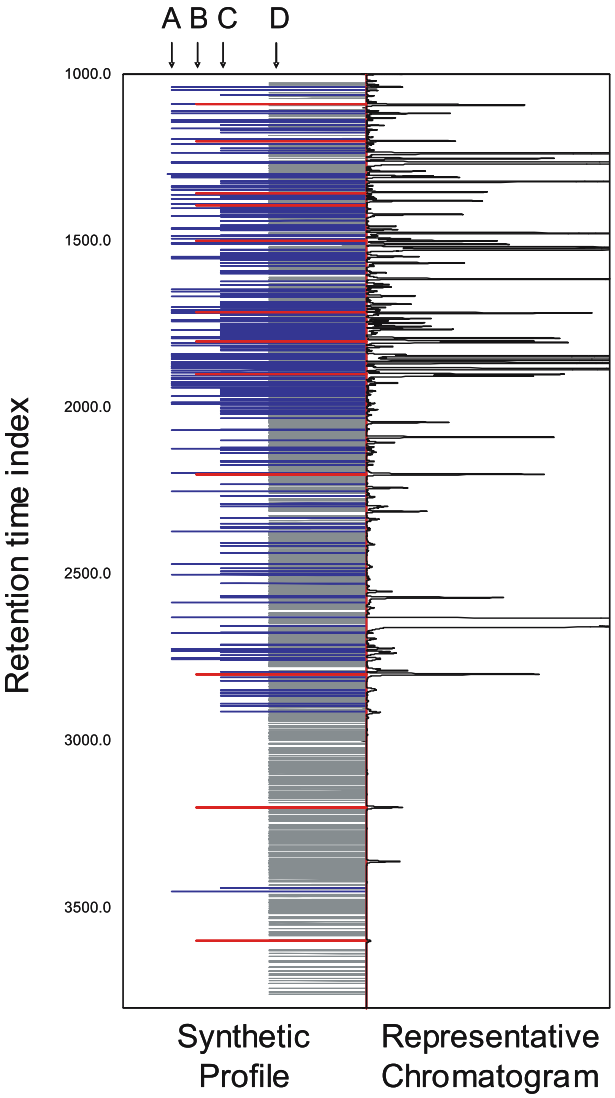


Fig. 4. Synthetic and representative GC-MS profiles of *Oryza sativa* L. cv. Nipponbare leaves: A – 132 identified MSTs representing 109 known metabolites; B – 12 added internal standard substances; C – 148 unidentified MSTs which match previous MSRI library entries; D – all previously observed MSTs present in the MSRI library at GMD (http://csbdb.mpimp-golm.mpg.de/gmd.html)

substances, which are subsequently mapped by standard addition experiments onto established standardised GC-MS systems, and (ii) the "top down" approach whereby structural elucidation is performed on a hitherto unknown, but important target MST. The work of "top down" structural identification

is highly time-demanding and involves preparative enrichment, purification, spectroscopic, mass spectral and NMR analyses of the preparation and finally chemical synthesis of the predicted structure. As a consequence the "bottom up" approach prevails in most laboratories and "top down" identification is currently restricted to potentially novel signalling compounds or marker substances of specific biological samples and experimental conditions.

In order to avoid unnecessary "top down" investigations reliable identification by prior standard addition experiments is essential. MSRI library collections of mass spectra (Kopka et al. 2005), which comprise frequently observed identified and non-identified MSTs, appear to represent the most effective means to pool the identification efforts currently performed in many laboratories around the world (Schauer et al. 2005). Identified and yet unidentified, MSTs can efficiently be shared by public resources such as GMD@CSBDB (http://csbdb.mpimp-golm.mpg.de/gmd.html). In addition mass spectral identifications and chromatographic sequence of analyte elution can be transferred between laboratories. "Bottom up" identifications performed in parallel may be used for inter-laboratory confirmation of identifications and reduce the risk of unnecessary structural elucidation projects.

Currently the MSRI libraries available from GMD@CSBDB include in total more than 2000 evaluated mass spectral data sets obtained using GC-QUAD- and GC-TOF-MS technology platforms with 1089 non-redundant and 360 identified MSTs. Future efforts at GMD aim to refine mass spectral quality and annotation, and will add stable isotope labelled variants of MSTs (e. g. Fig. 2) for improved mass spectral interpretation of unidentified MSTs. The number of identified analytes and metabolites will continuously be extended and annotations updated.

Acknowledgements. I would like to thank A.R. Fernie, A. Erban, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany for critically reading and discussing my manuscript. My thanks extend to Prof. Dr. Le Tran Binh, Institute of Biotechnology (IBT), Hanoi, Vietnam, for sharing his expertise in rice cultivation. This work was supported by the Max-Planck society, and the Bundesministerium für Bildung und Forschung (BMBF), grant PTJ-BIO/0312854.

References

Ausloos P, Clifton CL, Lias SG, Mikaya AI, Stein SE, Tchekhovskoi DV, Sparkman OD, Zaikin V, Zhu D (1999) The critical evaluation of a comprehensive mass spectral library. J Am Soc Mass Spectrom 10:287–299

Barsch A, Patschkowski T, Niehaus K (2004) Comprehensive metabolite profiling of *Sinorhizobium meliloti* using gas chromatography-mass spectrometry. Funct Integrat Genomics 4:219–230

Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW (2004) Potential of metabolomics as a functional genomics tool. Trends Plant Sci 9:418–425

Birkemeyer C, Kolasa A, Kopka J (2003) Comprehensive chemical derivatization for gas chromatography-mass spectrometry-based multi-targeted profiling of the major phytohormones. J Chromatogr A 993:89–102

Birkemeyer C, Luedemann A, Wagner C, Erban A, Kopka J (2005) Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling. Trends Biotechnol 23:28–33

Blau K, Halket JM (1993) Handbook of derivatives for chromatography, 2nd edn. Wiley, New York

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

Colebatch G, Desbrosses G, Ott T, Krusell L, 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

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

Desbrosses G, Kopka J, Udvardi MK (2005) Legume metabolomics: development of GC-MS resources for functional genomics of plant-microbe interactions. Plant Physiol 137:1302–1318

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

Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. Nat Rev Mol Cell Biol 5:763–769

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

Fiehn O (2003) Metabolic networks of *Cucurbita maxima* phloem. Phytochem 62:875–86

Fiehn O, Kopka J, Trethewey RN, Willmitzer L (2000a) Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. Anal Chem 72:3573–3580

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

Fischer E, Sauer U (2003) Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. Eur J Biochem 270:880–891

Gullberg J, Jonsson P, Nordström A, Sjöström M, Moritz T (2004) Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry. Anal Biochem 331:283–295

Halket JM, Przyborowska A, Stein SE, Mallard WG, Down S, Chalmers RA (1999) Deconvolution gas chromatography mass spectrometry of urinary organic acids – potential for pattern recognition and automated identification of metabolic disorders. Rapid Commun Mass Spectrom 13:279–284

Halket JM, Waterman D, Przyborowska AM, Patel RKP, Fraser PD, Bramley PM (2005) Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J Exp Bot 56:219–243

Jenkins H, Hardy N, Beckmann M, Draper J, Smith AR, Taylor J, Fiehn O, Goodacre R, Bino RJ, Hall RD, Kopka J, Lane GA, Lange BM, Liu JR, Mendes P, Nikolau BJ, Oliver SG, Paton NW, Rhee S, Roessner-Tunali U, Saito K, Smedsgaard J, Sumner LW, Wang T, Walsh S, Wurtele ES, Kell DB (2004) A proposed framework for the description of plant metabolomics experiments and their results. Nature Biotechnol 22:1601–1606

Jonsson P, Gullberg J, Nordström A, Kusano M, Kowalczyk M, Sjöström M, Moritz T (2004) A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. Anal Chem 76:1738–1745

Jonsson P, Johansson AI, Gullberg J, Trygg J, A J, Grung B, Marklund S, Sjostrom M, Antti H, Moritz T (2005) High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. Anal Chem 77: 5635–5642

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

Knapp DR (1979) Handbook of analytical derivatization reactions. Wiley, New York
Kopka J, Fernie AF, Weckwerth W, Gibon Y, Stitt M (2004) Metabolite profiling in Plant Biology: Platforms and Destinations. Genome Biol 5(6):109–117
Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmüller E, Dörmann P, Gibon Y, Stitt M, Willmitzer L, Fernie AR, Steinhauser D (2005) GMD@CSBDB: The Golm Metabolome Database. Bioinformatics 21:1635–1638
Kovàts ES (1958) Gas-chromatographische Charakterisierung organischer Verbindungen: Teil 1. Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. Helv Chim Acta 41:1915–1932
Little JL (1999) Artifacts in trimethylsilyl derivatization reactions and ways to avoid them. J Chromatorgr A 844:1–22
Mashego MR, Wu L, van Dam JC, Ras C, Vinke JL, van Winden WA, van Gulik WM, Heijnen JJ (2004) MIRACLE: mass isotopomer ratio analysis of U-^{13}C-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites. Biotech Bioeng 85:620–628
Mueller A, Duechting P, Weiler EW (2002) A multiplex GC-MS/MS technique for the sensitive and quantitative single-run analysis of acidic phytohormones and related compounds, and its application to *Arabidopsis thaliana*. Planta 216:44–56
Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. Plant J 23:131–142
Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie AR (2001a) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell 13:11–29
Roessner U, Willmitzer L, Fernie AR (2001b) High-resolution metabolic phenotyping of genetically and environmentally diverse plant systems – identification of phenocopies. Plant Physiol 127:749–764
Roessner U, Willmitzer L, Fernie AR (2002) Metabolic profiling and biochemical phenotyping of plant systems. Plant Cell Rep 21:189–196
Roessner-Tunali U, Hegemann B, Lytovchenko A, Carrari F, Bruedigam C, Granot D, Fernie AR (2003) Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development. Plant Physiol 133:84–99
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
Ryan D, Shellie R, Tranchida P, Casilli A, Mondello L, Marriott P (2004) Analysis of roasted coffee bean volatiles by using comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry. J Chromatogr A 1054:57–65
Sauer U (2004) High-throughput phenomics: experimental methods for mapping fluxomes. Curr Opin Biotechnol 15:58–63
Sauter H, Lauer M, Fritsch H (1988) Metabolite profiling of plants – a new diagnostic technique. Abstr Pap Am Chem Soc 195:129
Schauer N, Steinhauser D, Strelkov S, Schomburg D, Allison G, Moritz T, Lundgren K, Roessner-Tunali U, Forbes MG, Willmitzer L, Fernie AR, Kopka J (2005) GC-MS libraries for the rapid identification of metabolites in complex biological samples. FEBS Lett 579:1332–1337
Schmelz EA, Engelberth J, Alborn HT, O'Donnell P, Sammons M, Toshima H, Tumlinson JH (2003) Simultaneous analysis of phytohormones, phytotoxins, and volatile organic compounds in plants. Proc Natl Acad Sci USA 100:10552–10557
Schmelz EA, Engelberth J, Tumlinson JH, Block A, Alborn HT (2004) The use of vapor phase extraction in metabolic profiling of phytohormones and other metabolites. Plant J 39:790–808
Shao XG, Wang GQ, Wang SF, Su QD (2004) Extraction of mass spectra and chromatographic profiles from overlapping GC/MS signal with background. Anal Chem 76:5143–5148

Sinha AE, Fraga CG, Prazen BJ, Synovec RE (2004a) Trilinear chemometric analysis of twodimensional comprehensive gas chromatography-time-of-flight mass spectrometry data. J Chromatogr A 1027:269–277

Sinha AE, Hope JL, Prazen BJ, Nilsson EJ, Jack RM, Synovec RE (2004b) Algorithm for locating analytes of interest based on mass spectral similarity in GC×GC-TOF-MS data: analysis of metabolites in human infant urine. J Chromatogr. A 1058:209–215

Sinha AE, Prazen BJ, Synovec RE (2004c) Trends in chemometric analysis of comprehensive two-dimensional separations. Anal Bioanal Chem 378:1948–1951

Stein SE (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. J Am Soc Mass Spectrom 10:770–781

Steinhauser D, Usadel B, Luedemann A, Thimm O, Kopka J (2004) CSB.DB: a comprehensive systems-biology database. Bioinformatics 20:3647–3651

Stephanopoulos G, Alper H, Moxley J (2004) Exploiting biological complexity for strain improvement through systems biology. Nat Biotechnol 22:1261–1267

Strelkov S, von Elstermann M, Schomburg D (2004) Comprehensive analysis of metabolites in *Corynebacterium glutamicum* by gas chromatography/mass spectrometry. Biol Chem 385:853–861

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

Toyo'oka T (1999) Modern derivatization methods for separation science. Wiley, New York

Trethewey RN (2004) Metabolite profiling as an aid to metabolic engineering in plants. Curr Opin Plant Biol 7:196–201

Trethewey RN, Krotzky AJ, Willmitzer L (1999) Metabolic profiling: a Rosetta stone for genomics? Curr Opin Plant Biol 2:83–85

Urbanczyk-Wochniak E, Fernie AR (2005) Metabolic profiling reveals altered nitrogen nutrient regimes have diverse effects on the metabolism of hydroponically-grown tomato (*Solanum lycopersicum*) plants. J Exp Bot 56:309–321

Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner-Tunali U, Willmitzer L, Fernie AR (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. EMBO Reports 4:989–993

Van Deursen MM, Beens J, Janssen HG, Leclercq PA, Cramers CA (2000) Evaluation of time-of-flight mass spectrometric detection for fast gas chromatography. J Chromatogr A 878:205–213

Veriotti T, Sacks R (2001) High-speed GC and GC/time-of-flight MS of lemon and lime oil samples. Anal Chem 73:4395–4402

Vreuls RJJ, Dallüge J, Brinkman UAT (1999) Gas chromatography – time-of-flight mass spectrometry for sensitive determination of organic microcontaminants. J Microcolumn Sep 11:663–675

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. Phytochem 62:887–900

Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004a) Differential metabolic networks unravel the effects of silent plant phenotypes. Proc Natl Acad Sci USA 18:7809–7814

Weckwerth W, Wenzel K, Fiehn O (2004b) Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. Proteomics 4:78–83