Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Biotechnology 124 (2006) 312–322
Journal of BIOTECHNOLOGY
www.elsevier.com/locate/jbiotec

Current challenges and developments in GC–MS based metabolite profiling technology

Joachim Kopka *

Max Planck Institute of Molecular Plant Physiology, Department Molecular Plant Physiology (Prof. L. Willmitzer), Am Muehlenberg 1, D-14476 Golm, Germany

Received 15 June 2005; received in revised form 20 October 2005; accepted 1 December 2005

Abstract

Gas chromatography–mass spectrometry (GC–MS) based metabolite profiling of biological samples is one of the key technologies for metabolite profiling and substantially contributes to our understanding of the metabolome. While the technology is in increasing use it is challenged with novel demands. Increasing the number of metabolite identifications within existing profiling platforms is prerequisite for a substantially improved scope of profiling studies. Clear, reproducible strategies for metabolite identification and exchange of identifications between laboratories will facilitate further developments, such as the extension of profiling technologies towards metabolic signals and other technically demanding trace compound analysis. Using GC–MS technology as an example the concept of mass spectral tags (MSTs) is presented. A mass spectral tag is defined by the chemometric properties, molecular mass to charge ratio, chromatographic retention index and an induced mass fragmentation pattern such as an electron impact mass spectrum (EI-MS) or secondary fragmentation (MS^2). These properties if properly documented will allow identification of hitherto non-identified MSTs by standard addition experiments of authenticated reference substances even years after first MST description. Strategies are discussed for MST identification and enhanced MST characterization utilizing experimental schemes such as in vivo stable isotope labelling of whole organisms and open access information distribution, for example the GMG internet platform initiated in 2004 (GMD, http://www.csbdb.mpimp-golm.mpg.de/gmd.html).

Keywords: Gas chromatography–mass spectrometry (GC–MS); Mass spectral library; Metabolite profiling; Mass spectral tag (MST); Metabolomics; Retention time index

1. Introduction

Since the crystallization of the metabolomics (Oliver et al., 1998; Tweeddale et al., 1998) or metabonomics (Nicholson et al., 1999) concept gas chromatography hyphenated to mass spectrometry has developed into a widely spread basic and general metabolomics technique. From the early proposal as a key technology for metabolite profiling (e.g. Jellum et al., 1975; Jellum, 1977, 1979) GC–MS is now applied as a routine technology for the screening of apparent

* Tel.: +49 331 5 67 8262; fax: +49 331 5 67 898262.
E-mail address: Kopka@mpimp-golm.mpg.de.

0168-1656/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2005.12.012

or up to now hidden metabolic phenotypes in functional genomic studies of plants (e.g. Trethewey et al., 1999; Fiehn et al., 2000; Fiehn, 2002; Roessner et al., 2001; Fernie et al., 2004; Roessner-Tunali et al., 2004; Trethewey, 2004) or microbes (e.g. Barsch et al., 2004; Stephanopoulos et al., 2004; Strelkov et al., 2004). Publications utilizing GC–MS are rapidly increasing since the year 2000 and laboratories which enter the metabolomics field add GC–MS to their suite of technology platforms. This decision is facilitated by low costs compared to CE–MS, LC–MS, or LC–NMR instrumentation, unsurpassed chromatographic reproducibility and resolution, highly repeatable mass spectral fragmentation upon electron impact ionization (EI) and few, if any, matrix effects. The downside of GC–MS technology appears to be the requirement for chemical derivatization prior to quantitative analysis. However, this requirement for the chemical modification of those compounds which are not volatile per se may – in the long run – be turned into the advantage of exploiting selective chemical enrichment and fractionation for the profiling of trace compounds in the presence of bulk metabolites (e.g. Mueller et al., 2002; Birkemeyer et al., 2003; Schmelz et al., 2003, 2004).

Metabolomic studies and respective key technologies, such as GC–MS, now have fully emerged in biological science and add to our capability to describe and functionally assess biological systems with increasing resolution at the levels of the genome, transcriptome, and proteome. The vision of a fully comprehensive metabolome analysis of relative changes in metabolite pool sizes and metabolic flux, recently termed fluxome (Fischer and Sauer, 2003; Sauer, 2004), may be called the fourth significant addition to the field of "-omics" technologies. Multi-parallel metabolite analyses contribute two highly important and novel aspects to functional genomic and molecular physiology investigations. (1) Metabolites are the same molecular entity irrespective of the organism which makes use of them. Thus, the role of metabolites and their interaction with other system levels can be investigated without the typical ambiguity arising from orthologous and paralogous sequences. The function of metabolites and the phylogenetic change in the use of metabolites are now both open to be analyzed. Common or differential use of metabolic signals may be found in comparative studies across species and phylum boundaries. (2) Other than the fields of genomics, transcriptomics, and proteomics, the science of metabolomics has a rich history in flux measurement and modelling. The dynamics of transcription or translation, such as transcript turnover and protein stability, are technically accessible to studies of specific transcriptional or translational regulation. But multi-parallel measurements in these fields are far from routine. While the field of transcript regulation is currently revolutionized by the recent discovery of the function of micro-RNAs and small interfering-RNAs, the modelling and our understanding of the regulation of metabolite accumulation appears at least in microbial organisms to be much more advanced.

Despite multiple efforts at establishing diverse and competing metabolite profiling techniques, a fully comprehensive metabolome analysis of small molecules is and will – perhaps for a long time – remain a vision to be approached (Sumner et al., 2003; Bino et al., 2004; Birkemeyer et al., 2005). Metabolites are highly chemically diverse as compared to proteins, RNA, and DNA. No single common analytical technology can currently be envisioned to cover all metabolite classes. Thus, the current feasible approach is a combination of minimally overlapping and within analytical limits non-biased profiling analyses dedicated to roughly uniform compound classes (e.g. Nikiforova et al., 2005).

In conclusion the high diversity of chemical compounds, especially the specific biological use of stereo- and geometric isomers, as well as the demand for multiple analytical technologies currently poses three grand challenges to the science of metabolomics. (1) Metabolomic technologies allow multi-parallel analysis of hundreds of metabolites. However, the majority of covered metabolic components in metabolite profiles is still non-identified (Schauer et al., 2005a). Thus, the major challenge, even more so in other technologies than GC–MS, is the identification of the flood of hitherto non-identified metabolic components. (2) Well-known key metabolites and signalling compounds are still not accessible by routine multi-parallel profiling methods. Low metabolite concentration, unique chemical properties, such as chemical instability, and resulting laborious and time-consuming means of chemical analysis may be seen as the main obstacles which currently exclude these compounds from most comprehensive studies (Kopka et al., 2004). Therefore, the second challenge may be phrased as the demand for multi-parallel profiling analyses targeted to important

signalling compounds and crucial but not yet accessible metabolic intermediates. (3) All of the above tasks can only be achieved through a highly cooperative and interactive metabolomics community and thus the demand and challenge emerges to establish an efficient exchange of metabolite identifications (Kopka et al., 2005; Schauer et al., 2005a) and quantitative results.

In the following the GC–MS technology will be used to exemplify major aspects of these challenges. The subsequent discussion will be biased towards GC–MS technology, but many aspects will be put under general scrutiny. Thus, a contribution to the ongoing discussion and development of common concepts within the metabolomics community is intended.

2. Current developments and solutions

2.1. Compound identification: from mass fragment to metabolite

The first discovery of metabolic markers, such as the early increase of maltose during 1 h cold-response of *Arabidopsis thaliana* ecotype Col-0 (Kaplan et al., 2004) is initially made through observation of a statistically significant change in detector response. In mass spectral technologies this detector response is equivalent to an ion current, usually the ion current of a selected mass fragment, such as $m/z = 160$ (Fig. 1A), at a defined chromatographic retention time which can be standardized to units of retention time indices (RI). The relative change in response is routinely calculated as a response ratio of the baseline-corrected area or alternately the baseline-corrected height of the chromatographic peak divided by height or area of the same peak from biological reference samples, which are routinely included in metabolite profiling experiments (Fiehn et al., 2000; Roessner et al., 2001).

In the following all information is exemplified which is required for the full repeatability of an experiment in different laboratories using either the same technology, an alternate instrumental profiling platform, or a conventional targeted and fully quantitative analytical technology.

The first observation in MS-based profiling, here termed a mass fragment, has the properties, mass (or more precisely mass to charge ratio), chromatographic retention index and an abundance measured as a detector response. This mass fragment alone is hard if not impossible to unambiguously identify in different laboratories. Therefore, the mass fragment needs to be linked to a full GC–EI-MS mass spectrum. This mass spectrum, the result of an induced mass fragmentation of an initial molecular ion (M^+), is in the following called mass spectral tag (MST). Of any given MST, both MS and RI information are prerequisite for unambiguous compound identification in the context of multiple co-eluting compounds from typically complex biological extracts (Wagner et al., 2003). In LC–MS, CE–MS, or MALDI-TOF-MS the first observed mass fragment is usually M^+ or an adduct ion which can be described by an exact mass, and may be further characterized at MST level, for example by secondary fragmentation (MS^n) or post source decay (PSD).

A MST represents – in a strict sense – the chemical analyte not the native metabolite per se. In the case of the chosen example (Fig. 1), maltose is subject to methoxyamination and per-silylation. The resulting derivative structure is shown in Fig. 1D. The introduced C=N double bond creates two analytes, namely the *E*- and *Z*-isomers, which are resolved by GC–MS (Fig. 1A–C). The distinction between metabolite and the analyte(s) which represent(s) this metabolite is result of the GC–MS immanent requirement for chemical derivatization. Other MS-based technologies seemingly monitor the native metabolite. But these direct means of analysis may also have inherent and in cases multiple analytes which arise from the formation of adduct ions or from different ionization states.

In conclusion, metabolite identification programs of profiling platforms must document the full chain of evidence from a selected mass fragment, to MST "fingerprint" and analyte structure. Those metabolic components which are not yet identified should be documented by specific MST identifications, comprising a fragmentation pattern, RI, and, if possible, precise mass of the molecular ion, for future identification. MS technology platforms will differ in their respective capability to record such MST data. Most importantly a common understanding of the exact level of isomeric accuracy in metabolite determination is required for the exchange of identifications and tests of quantitative reproducibility.

The fact that metabolites occur in multiple biologically relevant or synthetic isomers is common place. For example, maltose of biological sources has

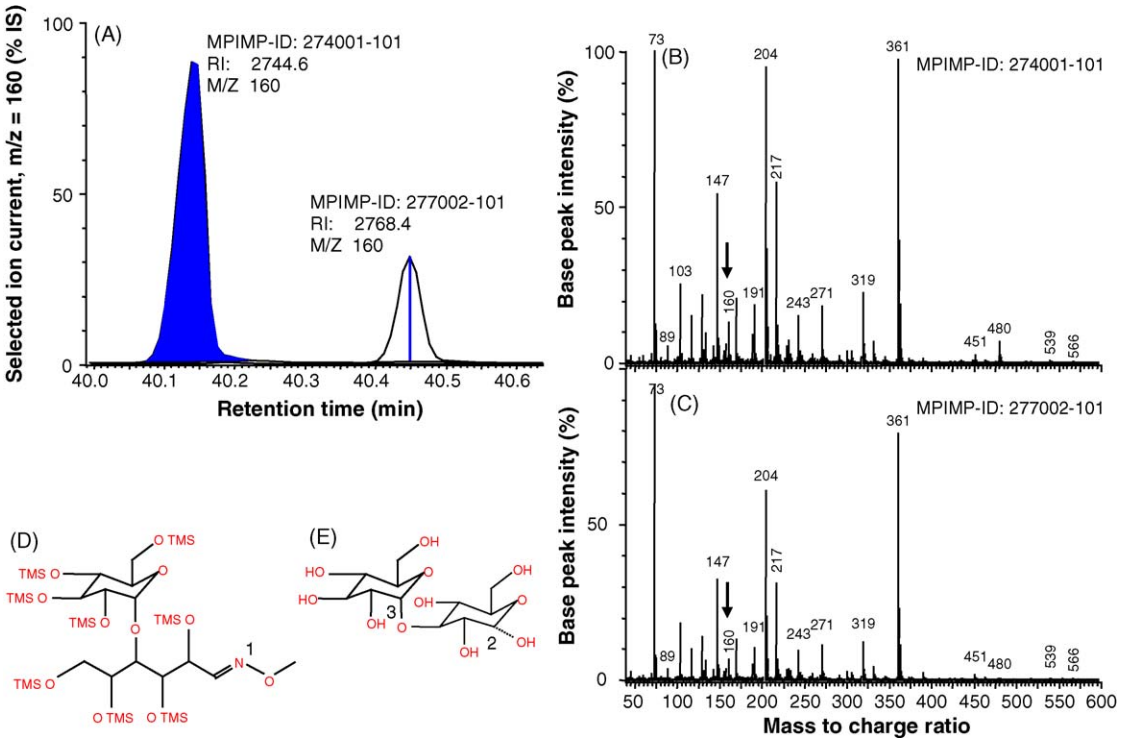

Fig. 1. An unbroken chain of evidence is required for reproducible metabolite identification in complex GC–MS chromatograms. (A) Metabolite profiling relies on the quantification of selected mass fragments with specific chromatographic behaviour using base line corrected peak area or height. (B and C) Each mass fragment, arrows indicate the selected *m/z* 160, belongs to a mass spectral tag (MSTs), i.e. a fingerprint of induced mass fragmentation at a defined retention index (RI) indicative of the underlying chemical compound. (D) These compounds, so-called analytes, are chemical derivatives which represent metabolites in GC–MS analyses. Chemical derivatization may produce more that one product. In this example the two *E*- and *Z*-geometrical isomers (1) are formed by methoxyamination. Chromatography and derivatization determine the level of isomeric accuracy at which a metabolite can be monitored. (E) In routine GC–MS profiles D- and L-stereoisomers cannot be distinguished because of lacking chromatographic separation, reducing-end anomer information of sugars (2) is lost, but resolution of glycosidic anomers (3) is usually maintained.

D-configuration. Routine profiling GC–MS, however, like most LC and CE methods does not allow stereoselective detection. In addition, reducing sugars form α- and β-anomers (Fig. 1E) and may exist in equilibrium with an open-chain form. All these variants are stably transformed into the same analyte after chemical modification for GC–MS or cannot be analyzed without disturbing the native equilibrium. Other anomeric structures, such as glycosidic bonds are fixed and are maintained throughout chemical derivatization and analysis. Thus, cellobiose and maltose can be distinguished by GC–EI-MS.

In order to perform a valid comparison of different profiling methods, each method must clearly define and document the precise level of isomeric accuracy and metabolite separation, which is achieved. For example, if the maltose measurement by GC–MS profiling is compared to LC-based profiling or to an enzyme assay, the chromatographic separation of different disaccharides must be previously characterized or – in the case of enzyme assays – possible side reactions and alternate substrate specificities must be communicated.

2.2. *Selective compound quantification: the benefit of stable mass fragmentation in GC–EI-MS*

In GC–EI-MS the initial M^+ ion is highly fragmented and usually of rather low abundance. This unavoidable property of GC–EI-MS strongly reduces the potential sensitivity and increases the complexity of

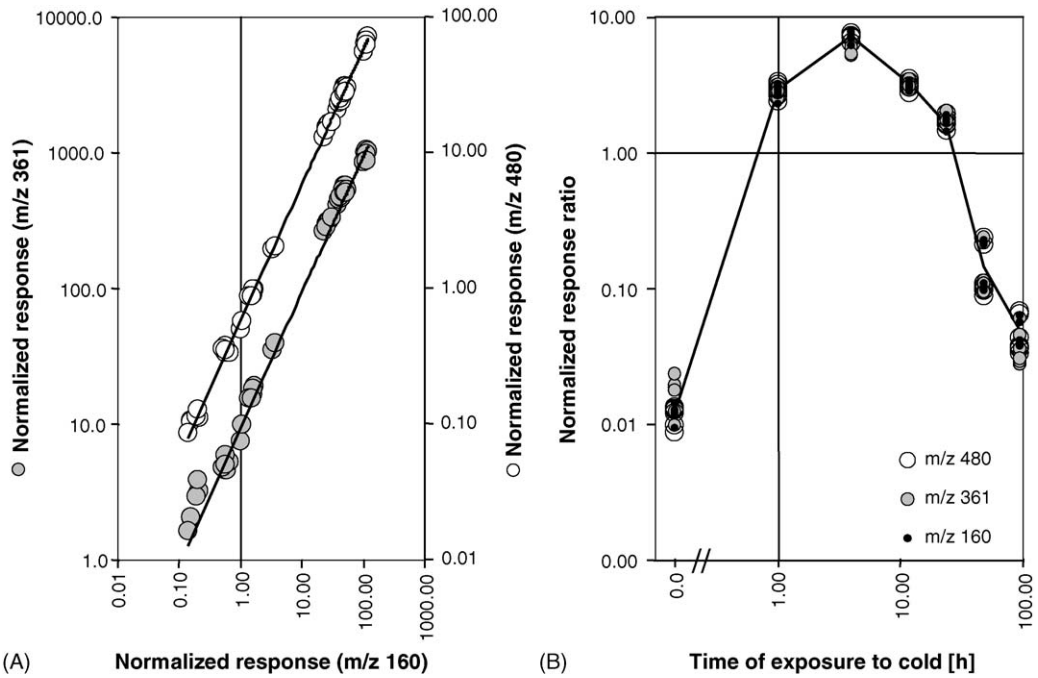

Fig. 2. Multiple mass fragments of the same MST allow independent verification of quantitative changes. (A) Mass fragments which belong to the same MST, for example *m/z* 160, 361, and 480 of MPIMP-ID 274001-101, exhibit a concentration-independent fragmentation pattern. Deviations from the resulting linear correlation of fragment abundance occur beyond upper and lower detection limits and can be used to eliminate non-selective mass fragments of multiple co-eluting compounds. (B) Response ratios which are based on multiple mass fragments independently generate the same profile of relative quantitative changes. In this example GC–MS detector responses were normalized to the amount of sample and to the response of an internal standard, such as ribitol. Normalized response ratios were calculated using the average normalized response of all samples in this experiment as quotient denominator. Data were taken from the compound validation process of Kaplan et al. (2004).

resulting chromatograms. In addition the strong fragmentation increases the risk of occurrence of identical mass fragments which may arise from common substructures of co-eluting analytes. Indeed, in GC–MS profiles of complex samples each selected mass fragment is best tested for its unambiguous selectivity. The task is highly similar to establishing gene identity in transcript profiling experiments through the use of multiple oligonucleotides or gene specific primer pairs. In GC–EI-MS the strong fragmentation can be exploited to verify fragment specificity and selectivity. All mass fragments which constitute a MST can be independently used to quantify the relative change in the pool size of the respective compound (Fig. 2). The specific mass fragments of a MST exhibit strictly linear changes in abundance. The linear range or reproducible fragmentation may extend across more than four orders of magnitude (Fig. 2A). Non-selective mass fragments and those exhibiting high noise can be excluded by deviation from the axiom of linearity. Thus, the GC–MS profiling inherent fragmentation allows independent validation of relative metabolite changes by use of multiple alternate mass fragments (Fig. 2B).

2.3. Linking metabolite profiles to public databases

Reconciliation of metabolite profiles with metabolite definitions made by public databases such as KEGG (Kanehisa et al., 2004), BRENDA (Schomburg et al., 2004), MetaCyc (Krieger et al., 2004), the PubChem project (http://pubchem.ncbi.nlm.nih.gov/), or the chemical abstracts service (CAS, http://www.cas.org/) is certainly advised to improve our common understanding of metabolites, metabolic processes,

regulation, and pathways. However, all presently available databases have non-reconciled, and in part redundant or ambiguous metabolite definitions. Especially metabolite isomers are distinguished at different levels of precision. Thus, a Babel of competing compound ontologies is created.

The current solution to this dilemma, instead of inventing new metabolite ontologies, is the precise evaluation of the capability of each metabolite profiling platform for isomer separation. The analytically correct level of isomeric accuracy of each metabolite can then be defined. Subsequently, all appropriate indices from the different public databases may be used for a detailed metabolite characterization and finally the valid linkage of profiling data to respective public database information. For example, the two methoxyaminated and per-silylated maltose analytes in GC–MS profiles, namely MST 274001-101 and 277002-101 (Fig. 1A–C), may be characterized and defined to represent the sum of at least the following, in part redundant database indices: KEGG|C00208 (maltose), KEGG|C00897 and KEGG|C01971 (α- and β-maltose), CAS|69-79-4 (maltose), and PubChem|3508 (maltose). This information can be modelled in relational database designs as a typical many to many relationship, such as the framework suggested by Jenkins et al. (2004).

2.4. Two strategies for MST identification

Strategies for a highly detailed identification of hitherto non-identified mass spectral tags will substantially increase the value of metabolite profiles and fully eliminate chemical artefacts from profiling analyses. Among the currently available collection of about 1000 non-redundant GC–EI-MSTs, which are characterized by fragmentation pattern and RI only 33.1% are chemically identified. The collection is still very small as compared to commercial mass spectral libraries, which may encompass as many as 163.198 non-redundant mass spectra (NIST05, National Institute of Standards and Technology, Gaithersburg, MD, USA, http://www.nist.gov/srd/mslist.htm), more than 338.332 partly redundant MS entries (Wiley Publishers, USA, http://eu.wiley.com/WileyCDA/Section/id-3047.html) or even 495.000 non-redundant mass spectra but encompassing the previously mentioned libraries (Palisade Mass Spectrometry, Ithaca, NY, USA, http://www.palisade-ms.com/). These libraries contain only a small fraction of metabolically relevant compounds, lack yet non-identified MSTs and most importantly – except for NIST05 – do not provide information on chromatographic retention behaviour. Chromatographic retention, however, is essential for unambiguous identification of MSTs from metabolite profiles. Mass spectral match alone – while suitable for mass spectral classification – is insufficient for identification (Wagner et al., 2003). The NIST05 mass spectral library now provides RI information but does not integrate RI information into an automated matching procedure. Therefore, combined MS and RI matching may now be seen as the next step of chemometric development.

Two strategies currently prevail in metabolite identification. The first approach, which we might want to call “top-down”, starts with a biological phenomenon or a yet non-identified but validated metabolic marker. The identification process continues with the tedious enrichment of biologically active fractions, and, if finally successful, reaches the goal of compound purification, chemical characterization and ultimate proof of structure by chemical synthesis. This process certainly does not fit to the ultimate demand to simultaneously identify hundreds or in the next wave of GC × GC–MS instrumentation possibly thousands of metabolites (Sinha et al., 2004a,b,c; Kell et al., 2005).

Thus, the “bottom-up” approach of establishing chemical libraries of commercially available metabolites and the brute-force mapping of these compounds to metabolite profiling platforms currently represents the first feasible path of fast metabolite identification. Once established chemical metabolite libraries can be applied to characterize the scope of any metabolomics technology. The distribution of pure reference substances may prove to be a highly effective way of establishing comparability between laboratories using different technology platforms. For those laboratories which utilize common technologies, such as GC–MS, electronic platforms which facilitate the exchange of metabolite identifications have already been established (Kopka et al., 2005; Schauer et al., 2005a). Similar approaches appear to be feasible for some LC–MS–MS technologies as well (Halket et al., 2005). One of the most important but not unexpected findings in establishing these libraries was the fact that the temporal order of MSTs in GC–MS profiles was

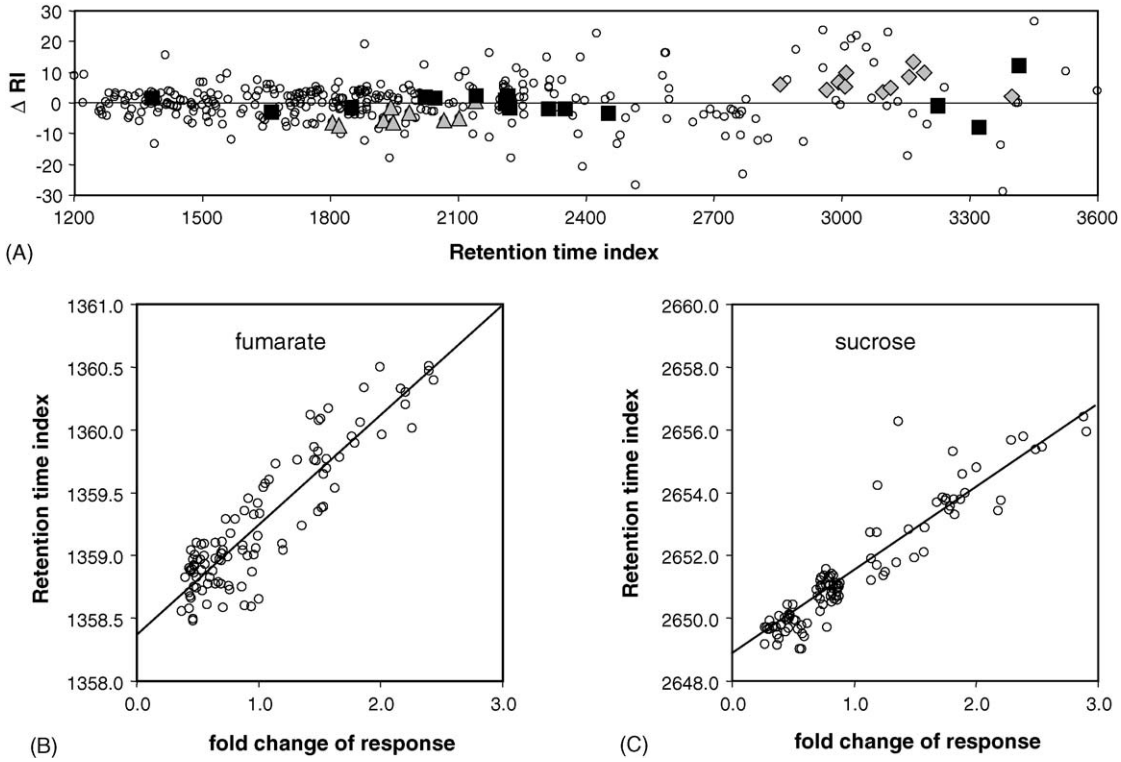

Fig. 3. Analysis of the composite residual error in predicting GC–MS retention time indices (RIs). (A) The accuracy of RI prediction from a quadrupole GC–MS system (Kaplan et al., 2004) to a GC–TOF-MS system (Wagner et al., 2003) exhibits an average error ±4.81 RI units. The difference of a local regression result and authenticated compound RIs (ΔRI) is plotted. Fatty acid trimethylsilyl esters (closed squares), phenylpropanoids (grey triangles), and phenylpropanoyl-quinic acid conjugates (grey diamonds) are indicated. (B and C) Changes in metabolite abundance induce a specific linear shift of compound RI. The slope appears to be a compound property. Plots exemplifying the effect of the change in response ratios on RI were taken from the compound validation process of Kaplan et al. (2004).

highly repeatable and roughly independent of the choice of GC–MS system or slope of temperature ramp, given the same type of capillary column was used. Nevertheless prediction of RI for unequivocal identification of metabolite isomers, such as hexoses or disaccharides, still requires further refinement. Fig. 3A shows the residual error of a RI prediction by local regression compared to the authenticated compound RI. In this case the standard deviation of prediction accuracy was 5.98 RI units. Two MSRI libraries available through GMD, http://csbdb.mpimp-golm.mpg.de/gmd.html, were used for this analysis (Kopka et al., 2005). A quadrupole GC–MS RI system with an approximately 60 min temperature ramp (e.g. Kaplan et al., 2004) was compared to 37 min temperature ramp using the same column type but a time of flight GC–TOF-MS system (e.g. Wagner et al., 2003). The residual error in RI prediction is demonstrated to be composite. On one hand, chemical compound classes exhibit common trends of over or under estimation, e.g. phenylpropanoids or phenylpropanoyl conjugates compared to the almost perfectly predicted class of fatty acid trimethylsilyl esters (Fig. 3A). A second strong influence is brought about by differences in analyte concentration. Changes in metabolite levels induce a substance specific and concentration dependent increase in RI as is demonstrated here for fumaric acid and sucrose (Fig. 3B–C). In contrast to these examples RIs of other compounds may hardly respond to changes in concentration (data not shown).

In conclusion, because of the limited RI predictability each laboratory still has to perform standard addition experiments especially for the identification of closely eluting metabolite isomers. For an improved

prediction of RI behaviour from MS-RI libraries at least compound concentration and substance class need to be considered.

2.5. The role of in vivo stable isotope labelling in MST characterization

For the characterization of those mass spectral tags which cannot be identified "bottom-up", in vivo stable isotope labelling will be an essential tool (Birkemeyer et al., 2005). Accurate mass spectral interpretation substantially reduces the number of potential candidates for any given non-identified MST. Isotope labelling does not directly yield the structural information. However, fully labelled mass isotopomers of metabolites yield unambiguous information on elemental composition of the respective molecular ion or induced mass spectral fragments and inherently proof metabolic origin (Fig. 4). The example in Fig. 4 shows ^{13}C, ^{15}N, and deuterated mass isotopomers of

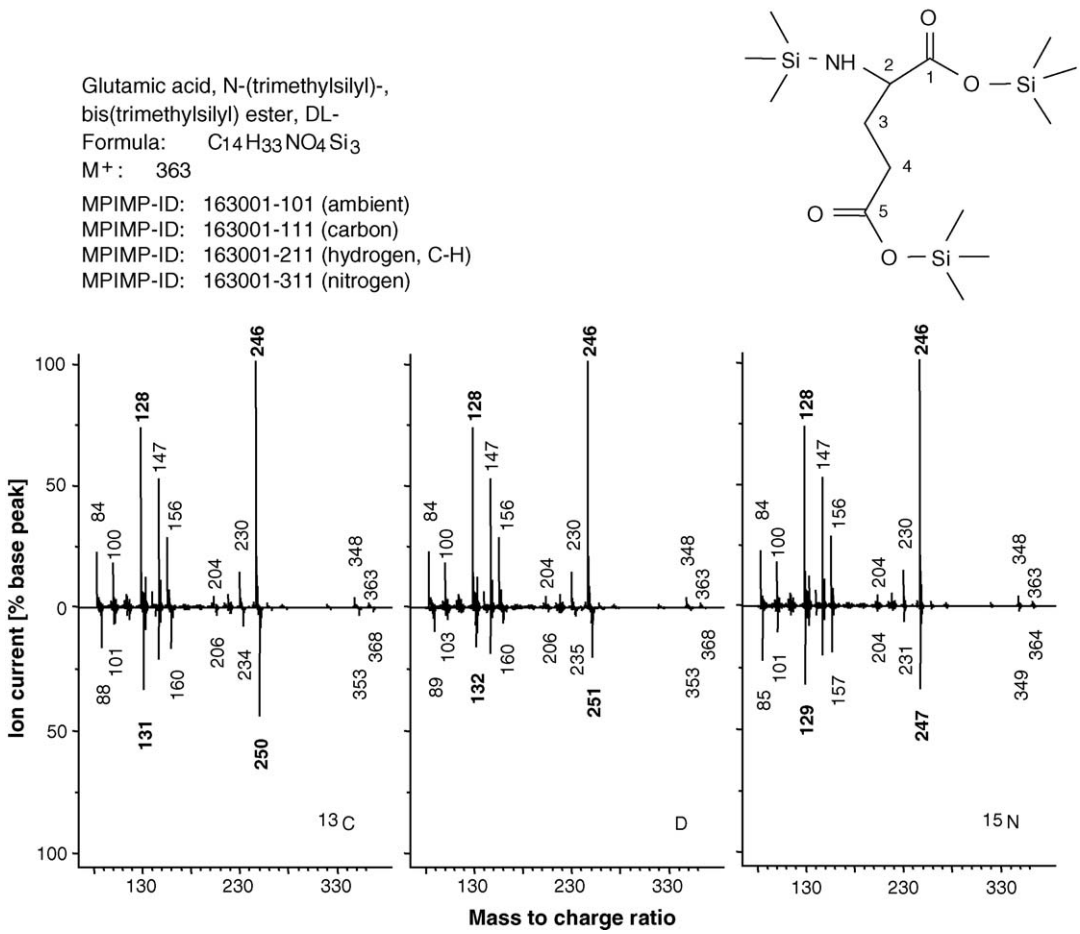

Fig. 4. Mass spectra of ambient, ^{13}C-, D- and ^{15}N-labelled MSTs help structural elucidation of non-identified metabolites. This example shows labelled MSTs of glutamic acid, *N*-(trimethylsilyl)-, bis(trimethylsilyl) ester in head to tail view with the ambient mass isotopomer above (MPIMP-ID, 163001-101, 111, 211, and 311). *Oryza sativa* L. cv. Nipponbare was labelled in vivo using deuterated water. Note that D_2O is lethal to plants. Only partial labelling was possible, and labelling efficiency differed between compounds. Mass spectral deconvolution of a GC–TOF-MS profile, however, allowed extraction of a fully deuterated MST from a partially labelled sample. *Saccharomyces cerevisiae* was fully labelled with U-^{13}C-glucose. Ninety-eight percent ^{15}N-labelled *Spirulina spec.* was obtained from Spectra Stable Isotopes Inc. (Columbia, MD, USA). Note that the non-silylated N–H hydrogen is exchangeable upon aqueous extraction. Major GC–EI-MS fragments are highlighted in bold format.

N-(trimethylsilyl)-glutamic acid di(trimethylsilyl) ester, the major analyte of glutamate in GC–MS metabolite profiling. Availability of multiply labelled mass spectral tags of the same compound aids interpretation of mass spectra irrespectively of the organism used for labelling the metabolome. For example, the fragment $[M\text{-}117]^+ = 246$ of this test case originates from inductive cleavage of either of the terminal carboxyl moieties and contains one N ($246 \rightarrow 247$ amu), five C–H ($246 \rightarrow 251$ amu), and four C, either C_{2-5} or C_{1-4} ($246 \rightarrow 250$ amu). Fragment $m/z = 128$ is formed after cleavage of one, -117 amu, and elimination, -118 amu, of the second carboxyl group, as is indicated by a shift of 1, 4, and 3 amu upon ^{15}N, D, and ^{13}C labelling, respectively. In the case of non-identified MSTs a detailed interpretation, similar to known compounds, is impossible, however, the elemental composition of mass fragments and, if present, M^+ can be deduced.

2.6. The role of in vivo stable isotope labelling in MST quantification

A fully isotope labelled metabolome may be used as a quantitative multiplex chemical standard in profiling analyses (Mashego et al., 2004; Birkemeyer et al., 2005; Wu et al., 2005). Using this approach each identified and non-identified compound is provided with the perfect internal standard, namely a chemically identical mass isotopomer, for relative quantification. However, in routine GC–MS profiling the high costs of multiplex stable isotope internal standardization appears to be hardly justified. Indeed, the gain of quantitative precision in GC–MS based metabolite profiling was an improvement of the mean error from 8.2% (6.9–9.7%) to 13.8% (5.5–33.4%) for metabolites with and without specific isotope-labelled internal standards (Gullberg et al., 2004). A similar gain in quantitative precision was reported for LC–MS–MS (Wu et al., 2005). The substantial advantage of a systematic use of mass isotopomers in metabolite profiling studies will be (1) besides better quantitative precision an extended linear range of quantification, especially at low concentrations (Wu et al., 2005) and (2) improved standardization of those mass spectral technology platforms which exhibit strong matrix effects. Most importantly, (3) mass isotopomer standardization will allow access to highly refined multi-step enrichment schemes which remove bulk metabolites such as glucose from microbial cultures or sucrose from plant extracts. Removal of these bulk metabolites will facilitate enrichment of trace compounds and development of metabolite profiling protocols for these hitherto not accessible compounds. Thus, mass isotopomer standardization by in vivo labelled metabolomes will ultimately extend the metabolite profiling concept to known and potentially novel metabolic signals.

Finally complex metabolite mixtures which are standardized by stable isotope labelled metabolomes may be regarded as the best experimental design to ultimately establish reproducibility of metabolite profiles across technology platforms and between laboratories. All technology or handling dependent variances of recovery, matrix suppression effects of mass spectral ionization, and the enhancement of metabolite measurements through chemical stabilization can ultimately be controlled by use of differentially labelled mass isotopomers.

Isotope dilution mass spectrometry will enhance metabolomics technologies to the same or even a better level of quantitative precision as is currently accepted for semi-quantitative proteome or transcriptome profiles, which rely on differential fluorophor labelling, such as difference gel electrophoresis (DIGE) of proteins (e.g. Unlu et al., 1997; Van den Bergh and Arckens, 2004), glass chip technology for transript analyses (e.g. Lockhart et al., 1996; Lockhart and Winzeler, 2000) or isotope coded tagging (ICAT) methods for protein profiling (e.g. Gygi et al., 1999; Aebersold and Mann, 2003).

Acknowledgements

I would like to thank A.R. Fernie and A. Erban, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany, and Prof. R. Trengove, Department of Biotechnology, Murdoch University, Murdoch, WA, Australia, for critically discussing this manuscript or specific aspects of the presented concept. My thanks extend to Prof. Dr. Lothar Willmitzer for his long-standing support. This work was supported by the Max-Planck society, and the Bundesministerium für Bildung und Forschung (BMBF), grant PTJ-BIO/0312854.

References

Aebersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. Nature 422, 198–207.

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

Bino, R.J., Hall, R.D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B.J., Mendes, P., Roessner-Tunali, U., Beale, M.H., Trethewey, R.N., Lange, B.M., Wurtele, E.S., Sumner, L.W., 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.

Fernie, A.R., Trethewey, R.N., Krotzky, A.J., Willmitzer, L., 2004. Metabolite profiling: from diagnostics to systems biology. Nat. Rev. Mol. Cell Biol. 5, 763–769.

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N., Willmitzer, L., 2000. Metabolite profiling for plant functional genomics. Nat. Biotechnol. 18, 1157–1161.

Fiehn, O., 2002. Metabolomics—the link between genotypes and phenotypes. Plant Mol. Biol. 48, 155–171.

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.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R., 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 17, 994–999.

Halket, J.M., Waterman, D., Przyborowska, A.M., Patel, R.K.P., Fraser, P.D., Bramley, P.M., 2005. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J. Exp. Bot. 56, 219–243.

Jellum, E., Helland, P., Eldjarn, L., Markwardt, U., Marhofer, J., 1975. Development of a computer-assisted search for anomalous compounds (CASAC). J. Chromatogr. 112, 573–580.

Jellum, E., 1977. Profiling of human-body fluids in healthy and diseased states using gas-chromatography and mass-spectrometry, with special reference to organic-acids. J. Chromatrogr. 143, 427–462.

Jellum, E., 1979. Application of mass-spectrometry and metabolite profiling to the study of human-diseases. Philos. Trans. R. Soc. Lond. Ser. A-Math. Phys. Eng. Sci. 293, 13–19.

Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A.R., Taylor, J., Fiehn, O., Goodacre, R., Bino, R.J., Hall, R.D., Kopka, J., Lane, G.A., Lange, B.M., Liu, J.R., Mendes, P., Nikolau, B.J., Oliver, S.G., Paton, N.W., Rhee, S., Roessner-Tunali, U., Saito, K., Smedsgaard, J., Sumner, L.W., Wang, T., Walsh, S., Wurtele, E.S., Kell, D.B., 2004. A proposed framework for the description of plant metabolomics experiments and their results. Nat. Biotechnol. 22, 1601–1606.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., Hattori, M., 2004. The KEGG resource for deciphering the genome. Nucleic Acid Res. 32, D277–D280.

Kaplan, F., Kopka, J., Haskell, D.W., Zhao, W., Schiller, K.C., Gatzke, N., Sung, D.Y., Guy, C.L., 2004. Exploring the temperature-stress metabolome of *Arabidopsis*. Plant Physiol. 136, 4159–4168.

Kell, D.B., Brown, M., Davey, H.M., Dunn, W.B., Spasic, I., Oliver, S.G., 2005. Metabolic footprinting and systems biology: the medium is the message. Nat. Rev. Microbiol., doi:10.1038/nrmicro1177 (published online 10 June 2005).

Kopka, J., Fernie, A.F., Weckwerth, W., Gibon, Y., Stitt, M., 2004. Metabolite profiling in plant biology: platforms and destinations. Genome Biol. 5, 109–117.

Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., Dörmann, P., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A.R., Steinhauser, D., 2005. GMD@CSBDB: The Golm Metabolome Database. Bioinformatics 21, 1635–1638.

Krieger, C.J., Zhang, P., Mueller, L.A., Wang, A., Paley, S., Arnaud, M., Pick, J., Rhee, S.Y., Karp, P.D., 2004. MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acid Res. 32, D438–D442.

Lockhart, D.J., Dong, H.L., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C.W., Kobayashi, M., Horton, H., Brown, E.L., 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. 14, 1675–1680.

Lockhart, D.J., Winzeler, E.A., 2000. Genomics, gene expression and DNA arrays. Nature 405, 827–836.

Mashego, M.R., Wu, L., van Dam, J.C., Ras, C., Vinke, J.L., van Winden, W.A., van Gulik, W.M., Heijnen, J.J., 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, E.W., 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.

Nicholson, J.K., Lindon, J.C., Holmes, E., 1999. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29, 1181–1189.

Nikiforova, V.J., Kopka, J., Tolstikov, V., Fiehn, O., Hopkins, L., Hawkesford, M.J., Hesse, H., Hoefgen, R., 2005. Systems rebalancing of metabolism in response to sulfur deprivation, as revealed by metabolome analysis of *Arabidopsis* plants. Plant Physiol. 138, 304–318.

Oliver, S.G., Winson, M.K., Kell, D.B., Baganz, F., 1998. Systematic functional analysis of the yeast genome. Trends Biotechnol. 16, 373–378.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., Fernie, A.R., 2001. Metabolic profiling allows

comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell 13, 11–29.

Roessner-Tunali, U., Lui, J., Leisse, A., Balbo, I., Perez-Melis, A., Willmitzer, L., Fernie, A.R., 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.

Sauer, U., 2004. High-throughput phenomics: experimental methods for mapping fluxomes. Curr. Opin. Biotechnol. 15, 58–63.

Schauer, N., Steinhauser, D., Strelkov, S., Schomburg, D., Allison, G., Moritz, T., Lundgren, K., Roessner-Tunali, U., Forbes, M.G., Willmitzer, L., Fernie, A.R., Kopka, J., 2005a. GC–MS libraries for the rapid identification of metabolites in complex biological samples. FEBS Lett. 579, 1332–1337.

Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., Schomburg, D., 2004. BRENDA, the enzyme database: updates and major new developments. Nucleic Acid Res. 32, D431–D433.

Schmelz, E.A., Engelberth, J., Alborn, H.T., O'Donnell, P., Sammons, M., Toshima, H., Tumlinson, J.H., 2003. Simultaneous analysis of phytohormones, phytotoxins, and volatile organic compounds in plants. Proc. Natl. Acad. Sci. 100, 10552–10557.

Schmelz, E.A., Engelberth, J., Tumlinson, J.H., Block, A., Alborn, H.T., 2004. The use of vapor phase extraction in metabolic profiling of phytohormones and other metabolites. Plant J. 39, 790–808.

Sinha, A.E., Fraga, C.G., Prazen, B.J., Synovec, R.E., 2004a. Trilinear chemometric analysis of two dimensional comprehensive gas chromatography–time-of-flight mass spectrometry data. J. Chromatogr. A 1027, 269–277.

Sinha, A.E., Hope, J.L., Prazen, B.J., Nilsson, E.J., Jack, R.M., Synovec, R.E., 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, A.E., Prazen, B.J., Synovec, R.E., 2004c. Trends in chemometric analysis of comprehensive two-dimensional separations. Anal. Bioanal. Chem. 378, 1948–1951.

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, L.W., Mendes, P., Dixon, R.A., 2003. Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 62, 817–836.

Trethewey, R.N., 2004. Metabolite profiling as an aid to metabolic engineering in plants. Curr. Opin. Plant Biol. 7, 196–201.

Trethewey, R.N., Krotzky, A.J., Willmitzer, L., 1999. Metabolic profiling: a Rosetta stone for genomics? Curr. Opin. Plant Biol. 2, 83–85.

Tweeddale, H., Notley-McRobb, L., Ferenci, T., 1998. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis. J. Bacteriol. 180, 5109–5116.

Unlu, M., Morgan, M.E., Minden, J.S., 1997. Difference gel electrophoresis: a single method for detecting changes in protein extracts. Electrophoresis 18, 2071–2077.

Van den Bergh, G., Arckens, L., 2004. Fluorescent two-dimensional difference gel electrophoresis unveils the potential of gel based proteomics. Curr. Opin. Biotechnol. 15, 38–43.

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.

Wu, L., Mashego, M.R., van Dam, J.C., Proell, A.M., Vinke, J.L., Ras, C., van Winden, W.A., van Gulik, W.M., Heijnen, J.J., 2005. Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly ^{13}C-labeled cell extracts as internal standards. Anal. Biochem. 336, 164–171.