Chapter 5

UNCOVERING THE PLANT METABOLOME: CURRENT AND FUTURE CHALLENGES

Ute Roessner
Australian Centre for Plant Functional Genomics, School of Botany, University of Melbourne, Victoria 3010, Australia

Abstract: Within the plant kingdom, it has been estimated that several hundred thousand different metabolic components may be produced, with abundances varying through of six orders of magnitude. The goal of metabolomics is a comprehensive and non-targeted analysis of metabolites in a biological system. Any valid metabolomic approach must be able to unbiasedly extract, separate, detect and accurately quantitate this enormous diversity of chemical compounds. These requirements dictate the challenges that are continually addressed in the field of plant metabolomics. To date, both gas- and liquid-based chromatography systems, in combination with various MS detection technologies, have been employed to analyse complex mixtures of extracted metabolites. In addition, nuclear magnetic resonance spectroscopy has been used to fingerprint plant systems, but will be not discussed in this context and has been reviewed elsewhere.

Although the technologies employed in metabolomic analyses are uncovering a huge amount of new knowledge in biology, a range of challenges are still to be faced. One bottleneck in metabolomic analysis is the identification of novel compounds. Additionally, in order to allow greatest spatial resolution, the sensitivity and selectivity of currently available technologies has to be increased. Multi-parallel and high-throughput analyses result in large data sets which need be evaluated, extracted, and interpreted. As a result, automated algorithms have to be developed. One of the major future challenges in the metabolomics field will be the integration of metabolic data with genomic and proteomic data sets. The ultimate goal is to comprehensively describe complex biological systems and as such, metabolomics has become an important player in systems biology. In the following text each of these challenges concurrently being connected with metabolomic analyses will be discussed.

Key Words: plant metabolome; chromatography; mass spectrometry; data analysis; data interpretation.

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 71–85.

1 INTRODUCTION

The development of tools to characterize genetic diversity in plant systems has made enormous progress over the last few years. Transgenic knockout populations, transposon insertions, and highly efficient ways to genotype single nucleotide polymorphisms within large populations have paved the way to a much broader base of diversity than imagined a few years ago. Furthermore, these developments have occurred in tandem with the elucidation of complete genomes and the rapid development of multiparallel technologies designed to access and describe the properties of biological systems (Celis et al., 2000). Most prominent amongst these new technologies has been the establishment of protocols for the determination of the expression levels of many thousands of genes in parallel (for review see Hardiman, 2004) and the detection, identification, and quantification of the protein complement (for review see Heazlewood and Millar, 2003). The logical progression from the large-scale analysis of transcripts to proteins is the determination of metabolite profiles in cells, tissues, and organisms. Importantly, the improvement of analytical instrumentations, such as mass spectrometry, has opened up the possibilities of determining and identifying a large number of metabolic compounds in parallel and in a high-throughput manner. The term metabolomics describes the comprehensive, non-targeted detection, and quantification of all compounds derived from a biological system.

2 ANALYTICAL TECHNOLOGIES FOR METABOLITE ANALYSES IN PLANT TISSUES

2.1 GC-MS

To date, both gas- and liquid-based chromatographic systems in combination with various mass spectrometry (MS) detection technologies, as well as nuclear magnetic resonance spectroscopy (NMR), have been employyed to analyse complex mixtures of extracted metabolites. Due to its overall robustness, gas chromatography coupled to electron impact ionization mass spectrometry (GC-EI-MS) has played a major role in high-throughput metabolite analyses (for review see Roessner et al., 2002). The use of GC allows separation of mixtures of compounds with high separation efficiency and sensitivity. In combination with MS, it also provides very accurate, sensitive and selective identification and quantification of separated compounds by their specific mass spectrum. Moreover, MS analysis further increases the resolution of the chromatography used as two co-eluting substances can be separated by their fragmentation pattern. Off-the-shelf

instruments are now able to rapidly and quantitatively detect up to 500 compounds simultaneously in crude plant extracts, depending on tissue and extraction procedure. In the past, GC-MS technology has been applied and optimized for simultaneous analyses of metabolites in many different plant species, such as *Arabidopsis thaliana* (Fiehn et al., 2000), *Solanum tuberosum* (Roessner et al., 2000), *Medicago truncatula* (Duran et al., 2003), *Lycopersicon esculentum* (Roessner-Tunali et al., 2003), *Saccharum officinarum* (S. Bosch, personal communication), *Lotus japonicus* (Colebatch et al., 2004), and *Cucubita maxima* (Fiehn, 2003).

In many of these detailed characterizations, it was shown that a one-dimensional GC separation approach does not resolve all compounds in high-complex extracts of plants. Recently a new approach has been taken, in which a second dimension of GC is applied to further separate the mixtures. GC × GC-TOF-MS has been already successfully applied to highly resolve volatile compounds of roasted coffee beans (Ryan et al., 2004). In the future, this technology will allow a more complete definition of the chemical composition of plants.

Despite the many advantages that GC-EI-MS has in metabolomics applications, there are also limitations of this technology. One of these is that GC can only be used for low molecular weight (<1000 Da) compounds, which are either volatile at relatively low temperatures, or which can be chemically transformed into volatile derivates. Thus, for a comprehensive analysis of a greater range of plant metabolites, complementary techniques have to be established (Kopka et al., 2004).

2.2 LC-MS

One complementary approach to GC-EI-MS in metabolite analyses is the application of liquid chromatography coupled to electrospray ionization mass spectrometry (LC-ESI-MS). The main advantages of LC-ESI-MS are twofold. Firstly, compounds do not have to be chemically altered prior to analysis and secondly, highly polar, thermo-unstable and high-molecular weight compounds, such as oligosaccharides or lipids, are able to be separated and quantified. LC in combination with an ultraviolet or visible light (UV/VIS) or diode-array detection (DAD) has been applied for many years in plant metabolite analyses. An enormous range of different columns and elution procedures exist for the separation and detection of many different classes of compounds. When coupled to MS, these provide further selectivity, unbiased detection, and most importantly, information about the structure of detected compounds. This multidimensional approach has been successfully applied for the analysis of a wide range of primary and secondary metabolites in plant tissues (Tolsitkov and Fiehn, 2002; Huhmann and Sumner, 2002). Recently, the use of a monolithic column enabled the separation of several hundred chromatographic peaks derived from extracts

of *Arabidopsis* (Tolstikov et al., 2003). Another research group has reported the detection of 1,400 components (based on mass-to-charge ratios) by direct injection of *Arabidopsis* extracts into a quadrupole time-of-flight (QTOF) hybrid mass spectrometer (von Roepenack-Lahaye et al., 2004). The resolution and selectivity of mass detection can be dramatically increased to up to 5,000 signals from a single plant extract by application of Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) as shown by Aharoni et al. (2002). In the future, this technique will play an increasing role in metabolic fingerprinting approaches where large mutant collections are screened for metabolic alterations following random mutations.

2.3 Increasing sensitivity

An additional challenge in metabolite analyses is the development of technologies for the isolation and detection of metabolites from very small samples sizes in order to increase spatial resolution in single cell or tissue-specific investigations. These techniques have to be designed to combine high sensitivity with selectivity. First remarkable reports have been given on the determination of the distribution of IAA in *Arabidopsis* plants (Muller et al., 2002) or even the distribution of ATP in *Vicia faba* embryos (Borisjuk et al., 2003). Future research has now to face multiparallel analyses of metabolites on a cell and organ level. One attractive technology to increase sensitivity is capillary electrophoretic separation techniques in combination with laser-induced fluorescence (CE-LIF) or mass spectrometric detection (CE-MS), which has been already proven to give promising results. For example, CE-LIF allowed the separation and quantification of a large range of amino acids and sugars in approximately 50 pL of phloem sap or in five pooled mesophyll cells of *Cucurbita maxima* (Arlt et al., 2001; S. Brandt, 2004, personal communication). By using CE-MS, more than 80 main metabolites belonging to glycolysis, photorespiration or the oxidative pentose phosphate pathway could be analysed in rice leaf extracts (Sato et al., 2004). It has to be especially noted that in this study, the ability to analyse many unstable substances in parallel, which only occur in low concentrations in *planta,* such as fructose-1,6-bisphosphate or ribulose-1,5-bisphosphate, was presented.

In the past decade many new technologies have been established which are currently used in novel biological information discovery in plant physiology and functional genomics. In summary, if the working definition for metabolomics means the analysis of all metabolites in a biological system, it requires a platform of complementary analytical technologies for comprehensive selectivity and sensitivity.

3 IDENTIFICATION OF UNKNOWN COMPOUNDS

Non-targeted metabolite detection in plant tissue results in a large number of chromatographic peaks and mass spectra, which cannot be identified easily with respect to the chemical nature of the compound. It has been shown in many metabolomic approaches that, for example, up to 70% of all peaks in a typical GC-MS chromatogram of a plant extract still remain unidentified. The interpretation of mass spectra following GC-EI-MS analysis is very difficult for two reasons. Firstly, derivatization dramatically alters the chemical structure of the compounds. Secondly, the use of electron impact (EI) to ionize the compounds is a very harsh method that leads to complex fragmentation patterns. As a result, two strategies are used to identify the chemical nature of as many peaks as possible. Firstly, the spectra of all resolved peaks are compared to commercially available EI mass spectrum libraries such as NIST (http://www.nist.gov/: National Institute of Standards and Technology, Gaithersburg, USA). However, although these libraries contain over 350,000 entries, the majority of these are non-biological compounds. In a second approach, commercial standard compounds, that are assumed to be present at detectable levels within plant tissues, are analysed. A reference library containing both the retention time of these compounds (as determined under the same conditions) and the corresponding mass spectrum can be created (Wagner et al., 2003). Identification by retention time is verified by co-chromatography of each standard substance with substances obtained in the plant extract. A major problem of this approach is that most plant compounds are not commercially available, especially the enormous number of secondary metabolites. Recently the publication of the first "biological" public domain GC-MS mass spectra library (MSRI; http://csbdb.mpimp-golm.mpg.de/gmd.html) was described (Kopka et al., 2005; Schauer et al., 2005). This library contains a large number of identified and unknown, but repeatedly observed EI-mass spectra of many different plant species and organs. A feature of this library is its compatibility with the NIST software and GC-MS evaluation software packages, such as automated mass spectral deconvolution and identification system (AMDIS) (see below). Further references to this mass spectral and retention time index library and its applications may be found in Chapter 4X by C. Birkemeyer and J. Kopka.

For LC-MS signal identification the situation is much more difficult. Mass spectra generated by LC-MS are typically instrument dependent and therefore, standard reference LC-MS spectral libraries are of limited use. The minimum information acceptable for the identification of novel organic compounds or metabolites has been traditionally defined by the scientific

literature criteria and often includes elemental analysis, NMR and MS spectral data for the isolated compound. One method for preliminary identification of unknowns appears to be the use of multidimensional instrumental techniques (based on combinations of GC-MS, LC-MS, MS/MS, or MS/NMR), which enable both comparative profiling and structural elucidation. For example, LC-QTOF-MS/MS (liquid chromatographic quadrupole tandem time-of-flight mass spectroscopy) has the potential to provide accurate mass and product-ion information of chromatographically separated metabolites. Experimental mass data can then be used for the calculation of an elemental composition and be compared with available mass information in, e.g., the NIST or KEGG database for possible structure suggestions. Further stepwise fragmentation by tandem MS (MS^n) leads to product-ion information, which can be used to determine/confirm structure. Although this gives much information about the potential structure of the compound, the final confirmation of the identity of the compound has to be done by either analysis of an authentic standards substance or by analysis of the purified sample using NMR.

The method of choice for unambiguous peak identification is NMR, which offers high chemical selectivity. In combination with LC and MS (LC-MS-NMR), it represents the ultimate technology for peak identification and structure elucidation (Wolfender et al., 2003) although the in-line version of this combination to date is still highly limited by the low sensitivity of the NMR instrument.

4 AUTOMATION OF DATA EVALUATION

Once an analytical platform is established a large number of samples can be analysed very quickly. This makes it an impractical and tedious task to manually extract information of each single chromatogram. One challenge of multitargeted compound analysis is the development of automated chromatogram evaluation. Many software packages delivered with the GC- or LC-MS system (Xcalibur, ThermoElectron, Austin, USA or HP Chemstation, Agilent, Palo Alto, USA) are able to use either self-created or commercial mass spectra libraries for peak detection, identification, and integration. The limitation of these software packages are that they search and integrate only targets, which the researcher has to know and enter into the search lists. This situation as been improved recently with the development of novel software packages for untargeted chromatogram evaluation based on mass spectral deconvolution. Deconvolution means the separation of corresponding fragments to one mass spectrum and thus for a single compound. This can be either achieved in an automated fashion by the software packages provided

with the GC-MS instrument (Pegasus, Leco, St. Jospehs, USA) or separate software can be applied, such as AMDIS (http://chemdata.nist.gov/mass-spc/amdis/; National Institute of Standards and Technology, Gaithersburg, USA). Recently other helpful commercial and free software packages have become available. Examples include MSFacts for GC-MS (Duran et al., 2003) or MetAlign for GC- and LC-MS (www.metalign.nl), which automatically import, reformat, align, correct the baseline and export large chromatographic data sets to allow more rapid visualization and interrogation of metabolomic data. To date, these software packages are indispensable for unambiguous data extraction. Very recently a novel software package named AnalyzerPro (www.spectralworks.com; Runcorn, Cheshire, UK) has been made available which meets the high requirements of an automatic GC-MS and also LC-MS^n chromatogram evaluation. In addition to signal deconvolution, mass spectra library matching and quantification, the implementation of retention time indices (RI) for improved signal identification are beneficial. Retention times of eluted substances following chromatographic separation do change dramatically over time. Retention time indices include for their calculation a range of added time references (e.g., long-chain alkanes) and therefore provide a better prediction of the absolute retention time of the analytes. In addition, retention time indices are very stable both within and between systems, allowing valid system to system comparisons, provided that injection, separation and ionization parameters are kept similar (Schauer et al., 2005).

5 DATA INTERPRETATION AND VISUALIZATION

As mentioned above, high-throughput analysis of a collection of samples results in large data sets, which have to be interpreted in a biological context. To date, statistical tools for pattern-recognition, such as hierarchical clustering (HCA) or principle component analysis (PCA), are routinely used for ease of comparison, and visualization of similarities and differences between data sets by definition of clusters (Fiehn et al., 2000; Roessner et al., 2001a, 2001b). Another approach is to detect dependencies and connections between metabolites and more recently, between genes, proteins, and metabolites by using pair-wise analysis of linear correlations (Urbanczyk-Wochniak et al., 2003; Steuer et al., 2003). Interestingly, when significant correlations are connected, the construction of regulatory networks becomes possible. The comparison of network connectivity between different genotypes allows not only the identification of novel pathways, it also represents a way of uncovering “silent” mutations, which do not show any obvious phenotype in any of the parameters under analysis (Weckwerth et al., 2004).

6 COMBINATION OF STEADY-STATE METABOLOMICS WITH METABOLIC FLUX ANALYSIS

The measurement of steady-state levels of metabolites, as described in this review, gives new insights into metabolic networks at a given time. But the real behaviour of plant metabolism can be only understood by determination of the dynamics of metabolism. The basis of metabolic flux analysis (MFA) is a combination of stable isotope labelling under steady-state conditions and NMR or MS-based detection systems to follow the distribution of label. This technique has been applied in detail in micro organism research but will play an increasingly important role in plant research (for review see Schwender et al., 2004). The application of a multiparallel detection method such as GC-or LC-MS allows determination of isotope label in very many metabolites in one experiment and therefore gives the opportunity to calculate metabolic fluxes of many different pathways simultaneously (Schwender et al., 2003; Roessner-Tunali et al., 2004). The power of this method becomes striking when it is incorporated with steady-state metabolite level determinations. This has been demonstrated by Foerster et al. (2002), showing *in silico* pathway analysis using stoichiometric models in yeast, which were constructed from knowledge of biochemical reaction networks in the cells. By further implementation of available genomic, biochemical, and physiological information, these authors reported the reconstruction of a genome-scale metabolic network from *S. cerevisiae* (Famili et al., 2003), which produced computed predictions for phenotypes following *in silico* mutation and therefore allowed gene function identification. In conclusion, a metabolomics approach in combination with stable isotope metabolic flux analysis will provide important insights in plant functional genomics studies. Another obvious use of this information will be in more rational approaches in metabolic engineering of novel, valuable biotech-crops (Sweetlove et al., 2003).

7 DEVELOPMENT OF DATABASES FOR METABOLOMICS-DERIVED DATA

In the past it has been noted by several scientists, that the large data sets generated by post-genomics technologies have to be transmitted, stored safely and be made available in convenient and accessible formats (Goodarce et al., 2004). The implementation of relational databases for data storage requires well-designed data standards. The DNA microarray community has agreed on the development of a minimum information about a microarray experiment

(MIAME, Brazma et al., 2001) and its structure has been widely accepted. Similar initiatives are underway for the proteomics community (PEDRo, Taylor et al., 2003). Whilst metabolic databases such as the KEGG system (Goto et al., 2002) or MetaCyc (Krieger et al., 2004) provide detailed information about metabolic pathways and enzymes of a variety of organisms, the development of a data standard equivalent to MIAME and PEDRo describing metabolomics data in their experimental context has been proposed only very recently (MIAMET, Bino et al., 2004; ArMet, Jenkins et al., 2004). On the other hand it will be not only important to store metabolic profiling data but to also integrate these data with metabolic pathway information which will be the future source of knowledge discovery. Recently, a database has been developed, which assembles information about different *Arabidopsis* metabolic pathways (AraCyc) and provides diagrams showing metabolites and genes encoding the enzymes in each pathway (Mueller et al., 2003). For a holistic integration of numerous multiparallel genomic, proteomic, metabolomic, and metabolic flux analysis data sets with metabolic pathway information the "Pathway Tools Omics Viewer" (http://www.arabidopsis.org:1555/expression.html) has been enabled, which in an easy and powerful manner paints experimental data onto the biochemical pathway map. Another example for such "mapping" tool is MapMan (Thimm et al., 2004), which allows users to visualize comparative metabolic and also transcriptional profiling data sets on existing metabolic templates and design their own templates. For a holistic integration of numeric multiparallel genomic, proteomic and metabolomic data sets a data managing system for editing and visualization of biological pathways was developed, which on a publicly available domain will be very important for data-mining in the functional genomics field (MetNetDB, Syrkin Wurtele et al., 2003; PaVESy, Luedemann et al., 2004). These software tools henceforth will become important to map novel findings onto metabolic pathways and fully understand the function of each gene, encoded protein and metabolite.

8 FROM TECHNOLOGY TO BIOLOGY

Once a robust metabolite analysis platform has been established and reliable data can be produced, the range of plant research applications is enormous. These can vary from answering simple biological questions, i.e., what are the metabolic differences between two cultivars, to investigations regarding complex metabolic networks. For example, a metabolomics approach can be used to determine the influence of transgenic and environmental manipulations on the metabolite profile as demonstrated by a detailed characterization of the metabolic complement of a number of transgenic potato tubers altered in their starch biosynthetic pathway and wild

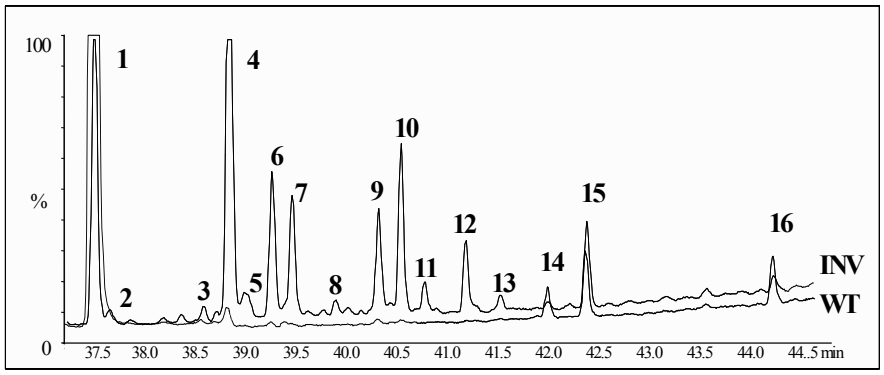

Figure 5-1. Comparison of a specific region of a GC-MS chromatogram of wild type potato tuber (WT, lower line) compared to tubers expressing a yeast invertase in the cytosol (INV, upper line). 1: sucrose; 3: maltose TMS; 4: maltose MEOX1; 5: trehalose TMS; 6: maltose MEOX2; 7: maltitol TMS; 12: isomaltose MEOX1; 13: isomaltose MEOX2, 2, 8, 9, 10,11, 14, 15 and 16 are not identified, mass spectra suggest they are sugars or sugar derivatives.

type tubers incubated in different sugars using GC-MS (Roessner et al., 2001a, 2001b). Due to this non-targeted approach, many unintended differences of transgenic tubers compared to wild type were detected (Roessner et al., 2001a; Figure 5-1). This study showed that using a metabolomic approach, it is possible to easily phenotype genetically and environmentally diverse plant systems.

Another useful application of metabolomics is in the field of functional genomics, which aims to identify of gene functions using high-throughput phenotyping technologies, as for example in investigations of responsible genes and their products in plant adaptations to different abiotic stresses. Often the role of certain metabolites in stress response could be assigned, as for example proline plays a major role in salt stress adjustments in rice (Garcia et al., 1997). The detailed characterization of metabolic adaptations to low and high temperatures in *Arabidopsis thaliana* has already demonstrated the power of this approach (Kaplan et al., 2004; Cook et al., 2004). Interestingly, it could be shown that low temperatures have more profound effects than heat, and novel findings of metabolic adaptation to temperature stress were identified (Kaplan et al., 2004). Another important report on using metabolomics as a tool in investigating metabolic responses of *Medicago truncatula* cell cultures to biotic and abiotic elicitors has revealed both elicitor-specific responses of metabolite levels as well as more generic responses in which similar metabolites responded independently of the type of stress (Broeckling et al., 2004). Nutrient deficiencies and

toxicities represent another example of common stress situations, e.g., it has been already demonstrated that the availability of inorganic nitrogen can reprogram carbohydrate metabolism (Stitt et al., 2002). This has been recently verified in more detail by a metabolomic investigation of the effects on tomato leaf metabolism grown in saturated, replete, and deficient nitrogen supplement conditions (Urbanczyk-Wochniak et al., 2005), showing the impact of nitrogen levels in the growth solutions on a wide range of metabolites. Similar striking effects on metabolite levels have been found when barley plants were grown in conditions where other inorganic nutrients are unavailable, e.g., phosphate or zinc (Roessner-Tunali, unpublished results). In the future, this approach will lead to the determination of the role of both metabolites and genes in stress tolerance and thus provide new ideas for genetic engineering of novel stress-resistant crops.

The next step of interpretation of metabolomic data sets can be achieved when they are integrated with other "omics" data such as transcriptomic or proteomic data. First attempts to face this challenge have been presented by Urbanczyk-Wochniak and co-workers who combined data obtained from microarray analysis and metabolite profiling of the same sample (Urbanczyk-Wochniak et al., 2003). A co-response analysis of both data sets has resulted in a large number of significant correlations between mRNA transcripts and metabolites. Some of these could be explained easily with existing biochemical knowledge but most were found to be novel, and thus highlighted the power of these integrated approaches for gene and metabolite function identifications. A similar investigation simultaneously analysed transcripts and metabolite levels in *Lotus japonicus* nodules to study symbiotic nitrogen fixation in detail (Colebatch et al., 2004). This report has shown clear interrelationships between transcript and metabolite responses dependent on a physiological event.

Last but not least, it has to be noted that a detailed characterization of the metabolome of a biological organism plays an integral role in a systems biology approach (Weckwerth, 2003). The aim of the emerging area of systems biology is to investigate the dynamics of all genetic, regulatory and metabolic processes in a cell and to understand the complexity of cellular networks (Kitano, 2002). Further this will give the opportunity to investigate the behaviour of biological systems with respect to the environment.

ACKNOWLEDGEMENTS

I would like to thank Dr. J. Patterson and Dr. Siria Natera for proof-reading the manuscript. My special thanks to Dr. Joachim Kopka for helpful discussions and to the Australian Centre for Plant Functional Genomics for funding.

REFERENCES

Aharoni, A., Ric de Vos, C.H., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R., and Goodenowe, D.B., 2002, Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry, *OMICS* **6**:217–234.

Arlt, K., Brandt, S., and Kehr, J., 2001, Amino acid analysis in five pooled single plant cell samples using capillary electrophoresis coupled to laser-induced fluorescence detection, *J. Chrom. A* **926**:319–325.

Bino, R.J., Hall, R.H., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B., Mendes, P., Roessner-Tunali, U., Beale, M., Trethewey, R.N., Lange, B.M., Syrkin Wurtele, E., and Sumner, L., 2004, Opinion: Potential of Metabolomics as a Functional Genomics Tool, *Trends Plant Sci.* **9**:418–425.

Borisjuk, L., Rolletschek, H., Walenta, S., Panitz, R., Wobus, U., and Weber, H., 2003, Energy status and its control on embryogenesis of legumes: ATP distribution within Vicia faba embryos is developmentally regulated and correlated with photosynthetic capacity, *Plant J.* **36**:318–329.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M., 2001, Minimum information about a microarray experiment (MIAME)-toward standards for microarray data, *Nat. Genet.* **29**:365–371.

Broeckling, C.D., Huhman, D.V., Farag, M.A., Smith, J.T., May, G.D., Mendes, P., Dixon, R.A., and Sumner, L.W., 2005, Metabolic profiling of *Medicago truncatula* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism, *J. Exp. Bot.* **56**: 323–336.

Celis, J.E., Kruhoffer, M., Gromova, I., Frederiksen, C., Ostergaard, M., Thykjaer, T., Gromov, P., Yu, J., Palsdottir, H., Magnusson, N., and Ornoft, T.F., 2000, Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics, *FEBS Lett.* **480**:2–16.

Colebatch, G., Desbrosses, G., Ott, T., Krusell, L., Montanari, O., Kloska, S., Kopka, J., and Udvardi, M.K., 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., and Thomashow, M.F., 2004, A prominent role for the CBF cold response pathway in configuring the low-temperature metabolomie of *Arabidopsis*, *Proc. Natl. Acad. Sci. USA* **101**:15243–15248.

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

Famili, I., Foerster, J., Nielsen, J., and Palsson, B.O., 2003, *Saccharomyces cerevisiae* phenotypes can be predicted by using contraint-based analysis of a genome-scale reconstructed metabolic network, *Proc. Natl. Acad. Sci. USA* **100**:13134–13139.

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

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

Foerster, J., Gombert, A.K., and Nielsen, J., 2002, A functional genomics approach using metabolomics and in silico pathway analysis, *Biotechnol. Bioengineering* **79**:703–712.

Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G., and Kell, D.B., 2004, Metabolomics by numbers: acquiring and understanding global metabolite data, *Trends Biotechnol.* **22**:245–252.

Goto, S., Okuno, Y., Hattori, M., Nishioka, T., and Kanehisa, M., 2002, LIGANS: database of chemical compounds and reactions in biological pathways, *Nucleic Acid Res.* **30**:402–404.

Garcia, A.B., Engler, J., Iyer, S., Gerats, T., Van Montagu, M., and Caplan, A.B., 1997, Effects of Osmoprotectants upon NaCl Stress in Rice, *Plant Physiol.* **115**:159–169.

Hardiman, G., 2004, Microarray platforms – comparisons and contrasts, *Pharmacogenomics* **5**:487–502.

Heazlewood, J.L. and Millar, A.H., 2003, Integrated plant proteomics – putting the green genomes to work, *Funct. Plant Biol.* **30**:471–482.

Huhman, D.V. and Sumner, L.W., 2002, Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer, *Phytochemistry* **59**:347–360.

Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A.R., Taylor, J., Fiehn, O., Goodacre, R., Bino, R.J., Hall, R., 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., Syrkin Wurtele, E., and Kell, D.B., 2004, A proposed framework for the description of plant metabolomics experiments and their results, *Nat Biotechnol* **22**:1601–1606.

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

Kitano, H., 2002, Systems Biology: A Brief Overview, *Science* **295**:1662–1664.

Kopka, J., Fernie, A.R., Weckwerth, W., Gibon, Y., and 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., and Steinhauser, D., 2005, GMD@CSB.DB: 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., and Karp, P.D., 2004, MetaCyc: a multiorganism database of metabolic pathways and enzymes, *Nucleic Acid Res* **32**(Database issue):D438–442.

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

Mueller, L.A., Zhang, P., and Rhee, S.Y., 2003, AraCyc: a biochemical pathway database for Arabidopsis, *Plant Physiol.* **132**:453–460.

Muller, A., Duchting, P., and 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.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N., and 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., and Fernie, A.R., 2001a, Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems, *Plant Cell* **13**:11–29.

Roessner, U., Willmitzer, L., and Fernie, A. R., 2001b, High-resolution metabolic phenotyping of genetically and environmentally diverse plant systems – identification of phenocopies, *Plant Physiol.* **127**:749–764.

Roessner, U., Willmitzer, L., and Fernie, A.R., 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., and Fernie, A.R., 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., and 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.

Ryan, D., Shellie, R., Tranchida, P., Casilli, A., Mondello, L., and Marriott, P., 2004, Analysis of roasted coffee bean volatiles by using comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, *J. Chrom. A* **1054**:57–65.

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

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

Schwender, J., Ohlrogge, J.B., and Shachar-Hill, Y., 2003, A flux model of glycolysis and the oxidative pentosephosphate pathway in developing *Brassica napus* embryos, *J. Biol. Chem.* **278**:29442–29453.

Schwender, J., Ohlrogge, J., and Shachar-Hill, Y., 2004, Understanding flux in plant metabolic networks, *Curr. Opin. Plant Biol.* **7**:309–317.

Stitt, M., Muller, C., Matt, P., Gibon, Y., Carillo, P., Morcuende, R., Scheible, W.R., and Krapp, A., 2002, Steps toward an integrated view of nitrogen metabolism, *J. Exp. Bot.* **53**:959–570.

Steuer, R., Kurths, J., Fiehn, O., and Weckwerth, W., 2003, Observing and interpreting correlations in metabolomic networks, *Bioinformatics* **19**:1019–1026.

Sweetlove, L.J., Last, R.L., and Fernie, A.R., 2003, Predictive metabolic engineering: A goal for systems biology, *Plant Physiol.* **132**:420–425.

Syrkin Wurtele, E., Li, J., Diao, L., Zhang, H., Foster, C.M., Fatland, B., Dickerson, J., Brown, A., Cox, Z., Cook, D., Lee, E-K. and Hofmann, H., 2003, MetNet: software to build and model the biogenetic lattice of *Arabidopsis*, *Comp. Funct. Genom.* **4**:239–245.

Taylor, C.F., Paton, N.W., Garwood, K.L., Kirby, P.D., Stead, D.A., Yin, Z., Deutsch, E.W., Selway, L., Walker, J., Riba-Garcia, I., Mohammed, S., Deery, M.J., Howard, J.A., Dunkley, T., Aebersold, R., Kell, D.B., Lilley, K.S., Roepstorff, P., Yates, J.R. 3rd, Brass, A., Brown, A.J., Cash, P., Gaskell, S.J., Hubbard, S.J. and Oliver, S.G., 2003, A systematic approach to modeling, capturing, and disseminating proteomics experimental data, *Nat. Biotechnol.* **21**:247–254.

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y., and Stitt, M., 2004, MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes, *Plant J.* **37**:914–939.

Tolstikov, V.V., and Fiehn, O., 2002, Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion mass trap spectrometry, *Anal. Biochem.* **301**:298–307.

Tolstikov, V.V., Lommen, A., Nakanishi, K., Tanaka, N., and Fiehn, O., 2003, Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics, *Anal. Chem.* **75**:6737–6740.

Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L., and Fernie, A.R., 2003, Parallel analysis of transcript and metabolic profiles: a new approach in systems biology, *EMBO Rep.* **4**:989–992.

Urbanczyk-Wochniak, E. and Fernie, A.R., 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.

von Roepenack-Lahaye, E., Degenkolb, T., Zerjeski, M., Franz, M., Roth, U., Wessjohann, L., Schmidt, J., Scheel, D., and Clemens, S., 2004, Profiling of arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry, *Plant Physiol.* **134**:548–559.

Wagner, C., Sefkow, M., and 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.

Weckwerth, W., 2003, Metabolomics in systems biology, *Annu Rev. Plant Biol.* **54**:669–689.

Weckwerth, W., Loureiro, M.E., Wenzel, K., and Fiehn, O., 2004, Differential metabolic networks unravel the effects of silent plant phenotypes, *Proc. Natl. Acad. Sci. USA* **18**:7809–7814.

Wolfender, J.L., Ndjoko, K., and Hostettmann, K., 2003, Liquid chromatography with ultraviolet absorbance-mass spectrometric detection and with nuclear magnetic resonance spectroscopy: a powerful combination for the on-line structural investigation of plant metabolites, *J. Chromatogr. A* **1000**:437–455.