BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 60 Quark08 Quark08:Books:BLUK025-Plaxton:

3 Study of metabolic control in plants by metabolomics

Oliver Fiehn

3.1 Introduction

3.1.1 What is metabolomics?

The idea of 'metabolomics' has been coined and developed in the last decade to comprehensively study metabolism under genetic and environmental perturbations [1, 2]. However, the first papers involving metabolite profiling techniques were published well over 30 years ago, with the aim, at that time, of rapid medical diagnostics [3]. The underlying idea behind the use of metabolomics in plant biology today is to detect metabolic effects of genetic or environmental perturbation which may only distantly relate to known or presumed primary (enzymatic) alterations. Metabolomics, therefore, seeks to detect 'unexpected' events on a comprehensive scale, and it widely acknowledges the presence of novel metabolites with unknown chemical structure or biological function.

In this respect, it differs from classical control theory that has been applied more frequently to select well-known pathways or regulatory circuits with the objective to understand these pathways in a mathematical manner using well-defined models and assumptions. Usually, mathematical control models need to be supported by high level metabolite measurements such as flux data. Although some efforts have been reported to derive larger metabolic models from isotope calculations of protein hydrolysates, we are still far away from reaching the goal universal and global 'fluxome' [4] analysis, especially with regard to plant research. Metabolomics does not try to reach this goal. Its use in studying metabolism has so far been more of an observatory and confirmatory role. It aims less at directly deriving insights into the cellular organization of metabolism. Due to its power to detect broad classes of metabolites, including unknowns, metabolomics is best used for studying system properties (such as networks) and changes (control) of metabolite levels in disparate parts of metabolism.

In many studies involving genetic or environmental perturbations it appears that certain metabolic modules are in counter balance with others, such as sugars and amino acids (C/N balance), whereas other large modules such as lipid metabolism are less affected (or more tightly regulated) under these conditions. Hence, metabolomics may be best suited to identify which broader parts of metabolism are influenced in response to developmental, genetic or environmental changes. Data can then be used to generate novel hypotheses about potential cellular causes (changes in enzymatic or transport activities) that are responsible for such changes.

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 61 Quark08 Quark08:Books:BLUK025-Plaxton:

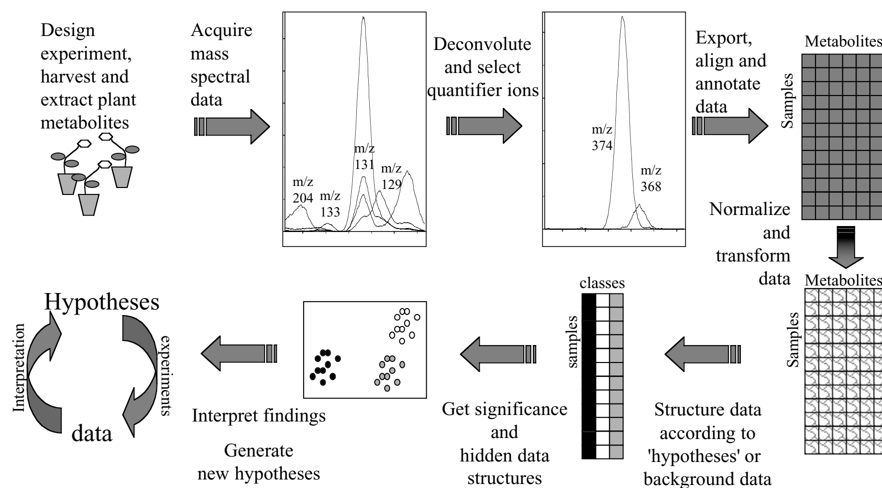

Figure 3.1 Flowchart of a plant metabolomic experiment.

Therefore, metabolomics will frequently generate more questions than answers, a concept that still needs to be embraced by classical hypothesis-driven research.

A general outline of this idea is depicted as Figure 3.1, which gives a flowchart of the way data are generated, annotated, transformed, structured and interpreted to gain novel hypotheses, before more experiments may verify these hypotheses. Starting from thoughts on system properties of metabolic networks, this chapter focuses on this flowchart, and specifically the problems associated with generating and annotating valid metabolite data. It adds a compilation of recent work in plant metabolomics to give an overview about the breadth and scope for which this technique is used in trying to understand plant metabolism.

3.1.2 Systemic properties in metabolic networks

Control and regulation are often used as synonyms, but this is actually not the case. As David Fell has pointed out in his famous book *Control of Metabolism* (1997 [5]), these two terms point to biochemical properties that are rather different in their respective meanings. Regulation is the ability of a complex system to maintain its basic properties (e.g. metabolite levels) independent of external factors that continuously try to push the system out of balance. A plant cell is exposed in short time intervals to many stochastic factors such as wind, light intensity differences, physical interferences or influx deviations of external transport metabolites. The system would become very unstable if each of these short-term pulses required immediate responses. There are a number of regulatory steps that inhibit metabolic overreactions, but instead introduce response lag times by using threshold systems, active transport steps, or reversibility of reactions. In total, these delay steps render the system to become 'robust' which is an important property to maintain

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 62 Quark08 Quark08:Books:BLUK025-Plaxton:

the system at a given steady state. Complementary to such robust regulation of steady state levels is the necessity to alter metabolite levels depending on certain stress conditions or developmental needs. The responsible general system property is called 'flexibility'. System flexibility is a prerequisite of the capability to 'control' or alter defined steady states without affecting other parts of the system, depending on external or internal stimuli. Any system needs capabilities to react in a fast and coordinated manner on immediate needs and threats, even if the triggering signals for such needs are of low abundance and transient in nature. Examples might be heat shock, wounding responses or herbivore attacks. The glucosinolate–myrosinase system commonly found in plants of the order Brassicales is one such example: myrosinases are thioglucosidases capable of hydrolyzing glucosinolates upon nonspecific generalist herbivore attack, which leads to a release of a suite of compounds with cytotoxic or feeding deterrent effects. Other examples of 'control' can be found in classic physiology. In physiological terms, cold acclimation (by increased values in carbohydrates) or leaf senescence (altered ratios of catabolism versus anabolism) are examples of 'control' or 'system flexibility', whereas the tendency to keep metabolic fluxes in a narrow range under a given set of environmental parameters (the steady state) is an example for metabolic 'regulation' or system 'robustness'.

3.2 Metabolomic methods

3.2.1 Historic perspective of plant metabolite analysis

How do plant systems manage to keep these two fundamental properties in balance? In principle, the global nature of metabolomic surveys should be directly suitable to answer this question. Metabolomics aims at quantification and identification of all metabolites of a given biological system under defined conditions. Metabolomic data may thus be used to assess network properties such as metabolite connectivities, or changes in metabolic ratios, individual metabolite levels and pathways. When stable isotope tracers are applied, even changes in fluxes or flux ratios can be assessed up to a certain extent [6]. However, it is still a methodological challenge to acquire comprehensive metabolic data, given the large differences in metabolite size, lipophilicity, volatility, charge state and other physicochemical properties.

Classically, analytical chemistry and plant physiology have focused on analyses of a limited number of select metabolite targets. The history of such target analyses tells us how introducing new instruments or methods has opened windows of research opportunities and how methodological advances have changed the view of plant metabolite functions and diversity.

Analyzing metabolite levels for plant physiological (or clinical) purposes has proceeded over the last 230 years, at least. At the end of the 18th century, Scheele and Vanquelin examined single primary plant metabolites such as citric, malic, oxalic, gallic and tartaric acid. Around the same time, microscopy was introduced

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 63 Quark08 Quark08:Books:BLUK025-Plaxton:

which enabled Markgraf to discover sucrose in sugar beet. In the 19th century, an array of analytical instruments was developed that fostered metabolite analysis, both for quantitative and qualitative purposes, including colorimeters, polarimeters, volumetric devices and photometers [7]. In that period, over 40 isolated plant metabolites were characterized by Berzelius. However, it was only during the last century that new techniques allowed the detection of the true richness of plant metabolomes, especially the so-called secondary metabolites.

The technological breakthrough did not come by a novel detection system but by better separation of metabolites. In 1906, the Russian botanist Mikhail Tswett invented chromatography by separating plant leaf extracts over powdered calcium carbonate and found chlorophyll pigments to be separated in several visible bonds [8]. The technique was eventually adopted and refined by Kuhn and Lederer in the 1930s who used it for carotenoid separations and purification of a large number of vitamins. This novel separation system boosted the number of compounds detected, particularly in combination with novel detectors which were based on a number of physical principles such as fluorescence emission, amperometry, light diffusion or light absorption in the visible and ultraviolet range. After the Second World War, the use of visible or UV absorbance, in particular, was widely adopted and allowed compound identification and comparison between laboratories based on spectral libraries [9].

3.2.2 Modern instrumentation in metabolite analysis

Today, virtually all these analytical methods are applied in physiological or medical research to separate, purify, detect and characterize compounds. However, it seems that for true metabolomics, only a select combination of methods is suitable. In general, a metabolomic method must be capable of detecting unambiguously, identifying and quantifying a large number of individual metabolites in a given sample. This requirement calls for high analytical resolution, universality (to detect metabolites irrespective of chemical substructures), selectivity (to acquire an analytical signal that is specific for a given metabolite), high dynamic range (to detect metabolites both at high and very low concentration), high precision (quantitative reproducibility), good accuracy (quantitative correctness) and high throughput (considering the need for statistically valid statements for a set of biological experiments). Metabolome number estimates range from about 500 (for prokaryotes) to many thousands of analytes (for vascular plants). There is no single method that can fulfil all the above mentioned requirements for so many analytes.

For this reason, the technology for true metabolomic approaches has only been enabled by breakthroughs in two areas, namely computer power and instrument engineering. This duo has fostered applications of the two dominant detection systems, nuclear magnetic resonance (NMR) and mass spectrometry (MS). An advantage of NMR is the linearity of quantitative responses on increasing metabolite concentrations, almost irrespective of the chemical compound class. Large signals in NMR can directly be interpreted as high level concentrations, whereas in MS,

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 64 Quark08 Quark08:Books:BLUK025-Plaxton:

quantitative responses strongly rely on the ionization potential of each metabolite. Therefore, quantitation in MS is limited to relative abundances of a given metabolite between samples, or requires calibration curves if absolute comparisons of different metabolites are needed. Nevertheless, MS is generally more favored than NMR due to four reasons: (i) MS is advantageous with respect to the capability to resolve complex mixtures of compounds, (ii) for most compounds, MS is far more sensitive than NMR, (iii) due to the need for high end magnets, NMR instruments are usually far more expensive than most MS instruments and (iv) coupling of NMR to separation techniques such as liquid chromatography is far from straightforward. This leads to the main use of NMR as a tool for 'metabolite fingerprinting' in which complex, unresolved spectra are compared from tens to hundreds of samples. This allows the classification of differences in the global control of metabolite levels according to the underlying experimental design. Correspondingly, even experienced research consortia rarely identify more than 30 individual metabolites per NMR spectra of plant extracts [10, 11], whereas the use of chromatography-coupled MS leads to routine identifications of up to 150 metabolites per sample [16].

3.2.3 Sample preparation for metabolomics

There is no optimal way to prepare comprehensive plant extracts. Some compounds such as ADP/ATP have such high turnover rates that anything beyond freeze clamping may just be too slow to efficiently stop any of the residual postharvest enzyme activity. Other compounds such as plant hormones may have low turnover rates but are of such low abundance (in whole organs) that large plant biomasses need to be prepared. These two cases mark the opposite and contradictory ends of the range of extraction prerequisites. Another conflicting constraint is the ranges of lipophilicity versus hydrophilicity. For example, leaf waxes need very nonpolar solvents (such as hexane), whereas sugars can only be extracted with solvents of high polarity (such as water). The solubilization power of a given solvent mixture may be altered by additional modifications such as application of heat, microwaves, pressure or ultrasonication. However, the general problem remains of the contradiction between metabolome-wide comprehensiveness and quantitative completeness (recovery) of the extraction method. *Arabidopsis thaliana* leaves have been shown to serve as an example of how to systematically maximize comprehensiveness and reproducibility in sample preparation procedures [12]. However, this study was restricted to primary metabolites that are detectable by gas chromatography-based methods.

In general, therefore, there are only two possible solutions for unbiased and comprehensive metabolomics. Either extractions are performed with sequential steps of solvent polarity (which may cause miscibility and precipitation problems when combining the resultant extracts) or compromises are accepted. The general notion is that it is unavoidable to take compromises in metabolomics. Even if these are accepted, it is mandatory to test and report the limits of the chosen method of sample preparation. For example, many protocols suggest the use of methanol during extraction. However, even slight amounts of methanol at ambient temperature

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 65 Quark08 Quark08:Books:BLUK025-Plaxton:

and physiological pH cause chlorophyll to decompose by autoxidation, demetallation and methylation [13, 14]. Consequently, if methanol extraction is performed, a range of porphyrins and other chlorophyll allomerization products are unavoidable and are detected by LC/MS. Other metabolites can potentially get altered in analogous reactions under comparably mild conditions. Apart from autoxidation, there are a number of other factors contributing to the formation of artifacts or loss of compounds. For example, thawing of biomass must be carefully avoided as long as proteins are not fully precipitated (for complete enzyme inactivation). Some enzymes such as hydrolases or phosphatases are still active even in methanolic solutions at ambient temperatures.

The large losses in compounds that have been observed have been compared in a direct comparison of two published plant extraction methods: the 70ºC hot $MeOH:H_2O$ (4:1 v/v) protocol [15] and the −15ºC cold $CHCl_3:MeOH:H_2O$ (2:5:2, v/v/v) strategy [16]; see Figure 3.2. Glucose-6-phosphate and other compounds with high turnover rates were barely detectable using the hot enzyme inactivation/extraction method, whereas recovery was high using the cold protein precipitation method. A likely reason for this striking difference is that frozen plant material might not reach 70ºC in the first seconds after addition of the methanolic mixture, and needs time to heat up. This time frame needed for heating the extraction slurry of ground-frozen Arabidopsis leaves and methanolic solvent may then last long enough to reduce the already low abundance levels of hexose phosphates

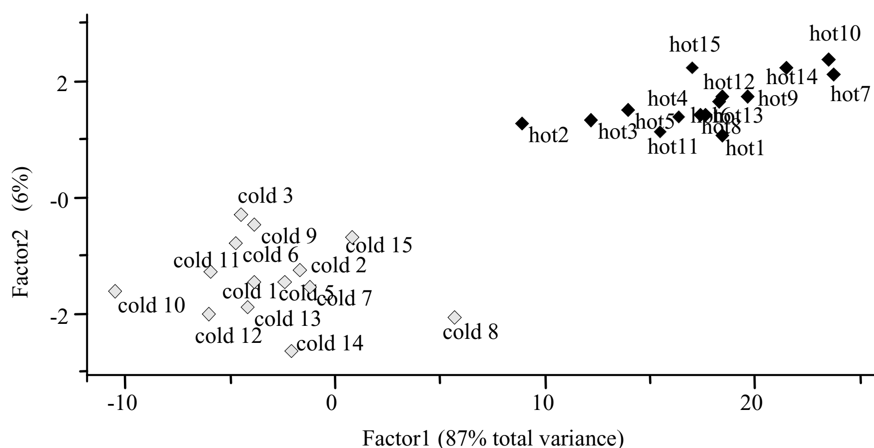

Figure 3.2 Principle component analysis of a direct comparison of the +70ºC (hot) and the −15ºC cold methanolic extraction method (cold) (unpublished results). Each method was applied 15 times on identical homogenized, deep frozen Arabidopsis leaf material. For each replicate, 20 mg FW material was taken. Data acquisition was performed by GC-TOF mass spectrometry. Principle component vector 1 separated the two methods, explaining 87% of the total variance of the data set. Investigation of the vector loading scores and *t*-test analysis resulted in identification of phosphorylated intermediates as most important discriminatory metabolites with clearly less being recovered from the +70ºC (hot) extracts.

and other phosphorylated intermediates to below the detection limit. Conversely, during cold extraction, enzymes are kept inactive at all times and in addition, the simultaneous presence of chloroform ensures immediate protein precipitation. Some protocols favor lyophilization instead of using fresh (frozen) plant material. However, even this procedure bears risks as many ligands directly interact with proteins or are tightly attached to cell walls or membranes. The degree of this interaction may be even higher in lyophilized material, which may cause losses in extraction of subsequent certain caged metabolites. Unfortunately, no comprehensive study has yet been published that compares these protocols.

Other physicochemical factors pose even larger risks in reducing metabolite recovery. Catecholamines, which are genuine metabolites in potato leaves, decompose when exposed to light for longer than 15–30 min [17]. Even more severe is the effect of oxygen that may be dissolved in the extraction solutions. Minute amounts of O_2 will suffice to oxidize cysteine, glutathione, tocopherol or ascorbate. Correspondingly, all extraction solvents and storage containers must be carefully degassed with argon or other noble gases.

Another important step that potentially leads to metabolite losses is solvent volume reduction. Concentration of solvents to complete dryness will inevitably cause losses of volatile and semivolatile components (such as terpenes). In addition, other compounds, such as complex lipids, may face reduced recoveries. For example, lipids may precipitate on surfaces during sample preparation or fractionation, with limited potential to get resolubilized (depending on the actual solvents and conditions). To conclude, therefore, published reports about the number of 'detected metabolites' in metabolomics need to be taken with caution, particularly if reports do not specify (a) the precise extraction method, sample preparation conditions and comparisons to method blank controls and (b) the number and names of unambiguously identified compounds and how confidence in metabolite annotation was achieved.

3.2.4 Metabolome coverage

3.2.4.1 The quest for combining sensitivity and selectivity

There are many challenges and open questions in plant metabolome analysis. For example, the simple question as to how large a plant metabolome is for a given species in a set of typical environments is still unanswered. There is growing evidence that the size of metabolomes cannot simply be computed using reconstructed pathways from genome annotations. On the one hand, many genes (and enzymes) as yet lack clear functional annotation, but on the other hand, enzymes may be far less specific than classically anticipated [18]. For example, deletion of a single amino acid residue may already lead to different enzyme substrate specificity and lead to new products [19]. Therefore, gene annotations may easily be misled by the sole reliance on gene sequence homology or the degree of amino acid identity.

The alternative possibility of using instrumental analytics to tackle metabolome size is equally difficult. An important reason for this is that both NMR and MS still lack the resolution, sensitivity or universality necessary to detect and identify all

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 67 Quark08 Quark08:Books:BLUK025-Plaxton:

components of a given sample. Developments for improved resolution may remedy this problem, for example, by exploiting more than one physicochemical property for separation prior to detection. An obvious possibility is to utilize gas phase ion mobility [20] in addition to classical chromatography. However, this technique has not yet been applied to plant samples. Other choices may include combining different chromatographic techniques such as coupling liquid chromatography to capillary zone electrophoresis (LC × CE; lipophilicity versus charge) or using gas chromatography with different column polarities (GC × GC, volatility versus lipophilicity) [21]. Even classical high pressure liquid chromatography (HPLC) (LC in short) has undergone dramatic improvements in performance, from the 4.6 mm i.d. columns that were used in the 1980s, to the 0.2 mm i.d. capillary columns used in the 1990s, to today's monolithic columns [27] (Figure 3.3) or ultrahigh pressure HPLC that leads to increased chromatographic resolution with over 100 000 theoretical plates.

Other advances in instrumentation to try to improve universality includes ionization for MS, which is achieved by introducing coupled interfaces (ESI × APCI,

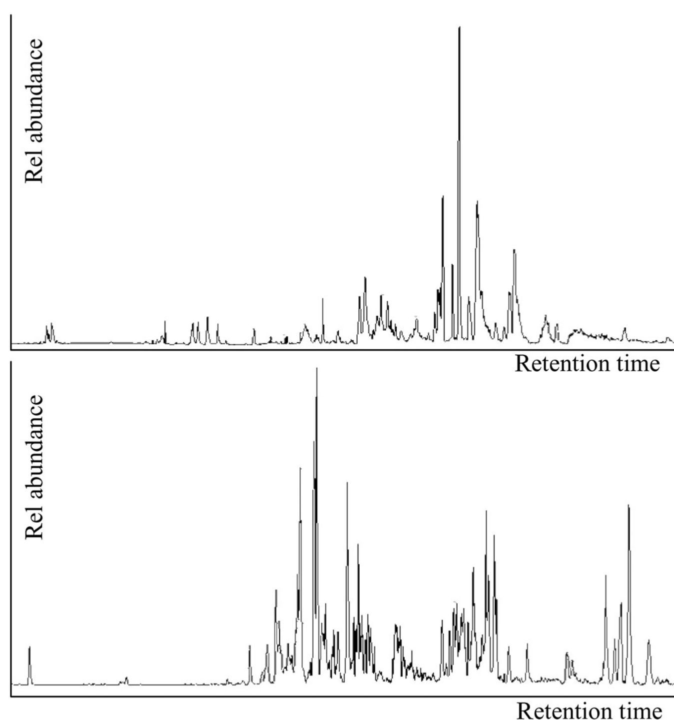

Figure 3.3 Chromatographic resolution in LC/MS. Upper panel: injection of 100 μl Arabidopsis thaliana leaf extract (i-propanol:water 2:1) onto a classical 4.6 mm C18 reverse phase column (20 000 theoretical plates/m, 1 ml/min flow rate.) Lower panel: injection of 0.2 μl of the same Arabidopsis extract onto a monolithic capillary C18 column (100 000 theoretical plates/m, 0.01 ml/min flow rate).

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 68 Quark08 Quark08:Books:BLUK025-Plaxton:

electrospray and chemical ionization) or characteristics of different types of mass spectrometers (linear ion trap, capable of tandem mass spectrometry 'in time' and 'in space'). However, it seems obvious that improvements in resolution for a single method will not cause a quantum leap in metabolome coverage. To date, uses of COSY and TOCSY two-dimensional NMR methods [22] have not succeeded in significantly increasing the number of identified compounds in metabolomic surveys. Approaches to use high-end Fourier-transform ion cyclotron mass spectrometers (FT-MS) are equally limited in the potential to target full metabolome surveys [59]. Despite this ultimate mass spectrometric resolution of $R > 500\,000$, isomeric compounds (such as glucose and fructose) cannot be distinguished due to their identical elemental composition, and even mass accuracy in FT-MS of <1 ppm does not allow unambiguous calculation of elemental compositions above approximately 400–500 Da without additional information [23, 24]. In addition, each metabolite subjected to electrospray/mass spectrometry usually gives rise to 3–5 further signals apart from isotopic ions. This is typically caused by adduct formation or in-source fragmentation ions. Therefore, the number of detected ions (mass signals or *m/z* values) must not be mistaken with the number of detected metabolites. Consequently, publications that report hundreds to thousands of mass signals based on MS-based metabolite fingerprinting [25] or LC/MS profiles [26, 27] do not give a reasonable estimate of metabolome size or coverage. Many of these ions certainly will account for novel and unknown compounds; however, others may simply be catabolic products which have been produced during postharvest biological processes or chemical by-products that occurred during sample preparation (or simply chemical artifacts due to impurities in solvents and plastic ware).

3.2.4.2 Cellular and subcellular metabolomics

A further reason why the complement of all metabolites present for a given plant, say Arabidopsis or rice, is not known is the lack of spatial resolution. It is a truism that a plant consists of many organs, and that each organ may include many tissue types and each tissue type may comprise various cell types. All published reports so far support the notion that different tissue types comprise varying metabolomes. Different biological roles of individual cell types support the further expectation of detecting striking differences on the low-level spatial resolution, e.g. between trichome and epidermis cells, or between parenchyma and bundle sheath cells. Lastly, intracellular organization of metabolism is also highly structured into compartments, each of which serves specific functions that lead to large metabolic differences. One example is a report of metabolite profiling of isolated chloroplasts and subfractions including the envelope, the stroma and the thylakoids in a study on the activity of three 13-lipoxygenases under stress conditions [28]. In this study, barley leaves treated with methyl jasmonate resulted in a remarkable increase of linolenic acid, free 9- and 13-hydroperoxy linolenic acid and the corresponding hydroxy- and aldehyde derivatives. The subcellular fractionation confirmed that these compounds were preferentially accumulated in the envelope and the stroma, therefore directly

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 69 Quark08 Quark08:Books:BLUK025-Plaxton:

linking the localization of the products and substrates with the corresponding lipoxygenases.

Apart from lack of spatial resolution, the number and concentrations of metabolites are controlled in response to plant development and environmental stimuli. On the other hand, the very nature of this flexible and unsteady metabolome state may serve as a valuable source of information of the physiological condition and the underlying regulatory network if carefully designed physiological (and genetic) plant experiments are carried out. Unfortunately, given the challenges of spatial and temporal resolution, today's analytical methods still seem to be inadequate with respect to acquiring the full complement of metabolites at the required sensitivity and for multiple biological snapshots.

3.2.4.3 Compound identification

In order to come closer to high metabolomic coverage, there are basically two approaches. One is to work toward a comprehensive analysis of all known and previously described metabolites by combining and unifying extraction, fractionation and improved detection methods. The metabolome coverage would increase in a stepwise manner by building blocks of certain compound classes until large overview analyses have been achieved. The advantage of this approach is that exact quantifications would enable comparison of results across experiments and databases. A clear disadvantage is that most known metabolites are simply not commercially available. As an example, from the general metabolite list given in the LIGAND database in KEGG [29], only about 1000 compounds can be purchased. Furthermore, this approach would always be biased by past knowledge and disregard the potential importance and impact of novel metabolites. The alternative approach is to embark on comprehensive structural elucidation by *de novo* analysis of (NMR and MS) spectra, called dereplication [30]. In the past 10–15 years, some progress has been made in computational methods and the establishment of metabolite spectra databases. However, many (secondary) metabolites are structurally so complicated that current methods are incapable of performing an automatic structural dereplication. Instead, detailed interpretations by experienced natural product chemists are needed. On-line acquisition of UV, NMR and MS spectra after separation of compounds by liquid chromatography enables rapid gathering of the necessary structural information which potentially leads to a partial or a complete online *de novo* structure determination of natural products [31]. However, there is always a lesser level of confidence in metabolite annotations if these are based on spectral interpretation rather than comparison with authentic standards. It is hardly imaginable that dereplication reaches the level of acceptance like gene sequencing or peptide mapping. Consequently, an increase in metabolome coverage can only be achieved by running both approaches simultaneously, i.e. building blocks of metabolite profiles of known standards, and *de novo* compound identification. It is mandatory to specify in metabolomic reports how exact a given metabolite annotation is. For example, whether or not the applied method is able to distinguish isobaric and/or isomeric compounds, or if even higher confidence levels are reached

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 70 Quark08 Quark08:Books:BLUK025-Plaxton:

that allow distinguishing chiral compounds such as D/L-isomers, allomers, enantiomers or diastereomers.

In addition, the choice of the data acquisition method is dictated by the biological question, as unbiased metabolomics is inadequate for many research projects. If, for example, biological hypotheses are narrowed to selected pathways or a small number of metabolic elementary modes, there is no need to use metabolomics. Instead, 'metabolite profiling' methods (that look for a limited set of pre-defined analytes) or classical target analyses can be applied. Conversely, target methods are unsuitable to answer questions about systemic control of network properties for which metabolomics is appropriate. In comparison to metabolomics, the focus on a selection of 'known' metabolites by metabolite profiling or target analysis disables an unbiased search for novel phytochemicals which may bear important physiological relevance (for example, as signaling molecules).

3.2.5 Quality control

Once a certain protocol for plant metabolomics has been developed, and is established for a research project, it is a prerequisite to monitor the quality of metabolite identification and quantification over the whole project period, preferentially over years. This is essentially the difference between method development and method validation. Developing a protocol basically means a proof-of-principle that a certain analytical objective can be fulfilled. However, a method is only validated if these objectives are strictly defined, and if the exact parameters and conditions are given as to how to achieve and monitor these results, e.g. by quality control charts. In metabolomics, this is relevant to both quantification and compound identification. For example, for each method the relative standard deviations need to be given, and it also needs to be specified how these limits are controlled and monitored on a routine (daily) basis. In addition to validation by relative error ranges, some compounds may be quantified in absolute concentrations using reference compounds and comparison to control methods to give quantification accuracies. Use of absolute instead of relative quantification refers to the question under study. For example, for some biological studies, determination of nanomolar concentrations may be essential for calculating turnover rates or crop nutritional quality. For other studies, such as functional genomics, assessment of metabolic control by *x*-fold values may be sufficient.

Once the validation criteria are clearly laid out, metabolomic protocols become 'standard operating procedures' (SOPs). SOPs are commonly seen for industrial processes including agro-biotechnology companies, but for academic laboratories, combining larger genomic programs with exploratory metabolomics, such SOPs are becoming more and more essential. In any case, high quality reports on metabolomics or metabolite profiling should always include detail as to how many of the claimed metabolite signals were detectable in all the samples of a given biological experiment, what level of variance was found for these variables within the experiment and how method blanks were used to ensure that detected peaks were not artificially formed

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 71 Quark08 Quark08:Books:BLUK025-Plaxton:

during the process. Without such specifications, the reported results and sometimes even the biological interpretations may become questionable.

3.3 Metabolomic databases

Publishing metabolite data is straightforward in classical target analysis. The experimental sections in peer-reviewed scientific journals usually refer to established and widely accepted methods, and data can be presented as average values or even as individual results for each sample, when appropriate. For metabolomics, publishing data is not so straightforward. Metabolomic results are usually data-rich, but poor in information. If only *x*-fold changes of metabolites are published with respect to the controls, then the data may contain only limited information for comparison with other experiments or conditions. Instead, metabolite levels must be deposited in a publicly accessible way that allows reusing the data under different aspects by giving the SOPs of sample preparation, data acquisition, data processing and the corresponding results. In 2004, a variety of reports have highlighted the importance of providing such information, among them being a general architecture for metabolomic databases ArMet [32] and considerations about the minimal information of a metabolomic experiment, MIAMet [33]. These considerations have only partially materialized in publicly available plant metabolomic databases [34]. For a range of compounds, agro-biotechnology companies have published validated metabolite data of crop nutritional value [35]. However, for fundamental research, no equivalent is known that is as comprehensive and validated. The basic reason for this lack is that there are very many aspects and parameters that need to be associated with 'metabolite levels' in order to turn these into informative and interpretable patterns that are useful for external researchers. Biochemical properties and cellular relationships can be mapped onto software platforms that can be interrogated in order to enhance the interpretability of data [36], but the very details of biological experiments and data acquisition are hard to capture in a standardized way. Some progress can be reviewed in a public forum that originated from a biomedical, pharmaceutical and toxicological background, *Standard Metabolic Reporting Structures* (SMRS) led by the Imperial College, London, UK [37]. Some of the associated problems are common for all databases that are reporting metabolite levels. The metabolites need to be named by unique identifiers in a consistent and traceable way to allow data exchange between different databases, for example, for system biology applications. Astonishingly enough, there are no such (publicly accessible) repositories of unique metabolite names.

The most comprehensive repository may be CAS, the Chemical Abstracts Service of the American Chemical Society, that includes information about millions of compounds, among them being biogenic metabolites. However, this service comes with high charges and it does not contain links to genomic databases. Furthermore, some compound identifiers in CAS have been changed or erased over time, and some components have multiple entries. Despite widespread popularity among (plant) biologists, databases such as Goto's LIGAND repository [29] cannot serve

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 72 Quark08 Quark08:Books:BLUK025-Plaxton:

as the authoritative resource for metabolite identifiers because many compounds, especially lipids, are not well covered. Efforts have been launched recently to compile comprehensive metabolite lists such as MetaCyc [38] and INCHI [39], and it is therefore very likely that the problem of consistent metabolite annotations will soon be solved.

In addition, the underlying plant biology experiments need to be described in detail to allow reuse of the metabolite results. This is a serious problem that seems to be very hard to tackle in an appropriate way. For publishing experiments in a peer-reviewed plant journal, it is expected to explain experimental details in both the 'materials and methods' section and the flow text. However, publication of data in database repositories cannot follow the same path. Any unstructured flow text description of biological study designs is insufficient. The concept of publicly available databases is that results can be queried and downloaded for comparative studies. This concept requires, therefore, a logical and consistent structure of entering information about the underlying experimental design and details. So far, unfortunately, there is no consensus in the plant community on vocabularies and items that are mandatory for describing a given experiment. One reason is that experimental designs are at the heart of a study and are therefore very different and hard to describe, and to capture in a fixed database structure. The other reason is that, so far, the biological community has relied on the peer-review system to ensure that sufficient information is given to enable reuse of data or the repeat of a study. However, there is usually no peer-review system associated with database entries. Metabolomics research groups may learn from Web-based entry forms that have been developed for describing transcript microarray data using a study annotator [40] that supports quantitative data with a structured ontology on the relationships and properties of various study designs and experimental details.

For general acceptance of metabolomic databases, a consensus needs to be sought how to name and structure plant biological experiments with respect to terms, structural hierarchies, ontologies and controlled vocabularies. Related efforts have resulted in compulsory repositories such as for naming species (in NCBI [41], for Arabidopsis germplasm in *t*he *A*rabidopsis *i*nformation *r*esource TAIR [42] or for naming plant organs in Plantontology.org [43]. In the meantime, metabolomic databases need to describe, but not prescribe, which experimental details need to be given.

3.4 Pathways, clusters and networks: applications of plant metabolomics

Recent studies in plant biology involving the four different types of metabolite analysis (target analysis, metabolite profiling, metabolomics, metabolite fingerprinting) may be classified into five broad research areas, some of which are overlapping: confirming the effects of bioengineering plant metabolic enzymes (Section 3.4.1), studying plant biochemistry including the connectivity of pathways (Section 3.4.2), observing and cataloguing physiological effects during developmental or

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 73 Quark08 Quark08:Books:BLUK025-Plaxton:

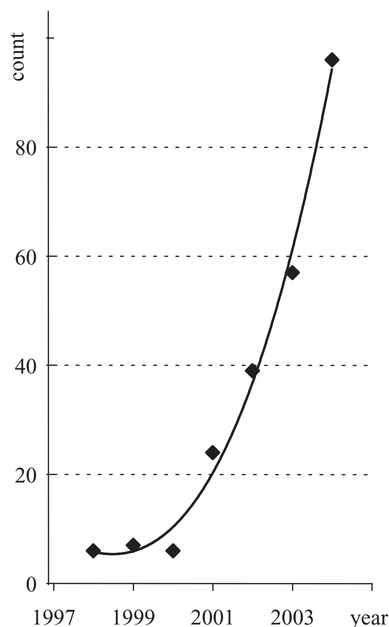

Figure 3.4 Count of hits in the literature database of the Institute for Scientific Information (ISI). Search words relating to metabolomics were restricted to the plant specific literature. 71% of the counts were original reports, and 20% were classified as review articles. 77 authors were found with three or more articles.

environmental transitions (Section 3.4.3), designing and utilizing better analytical methods (Section 3.4.4) and applying profiling methods in food science (Section 3.4.5). A literature survey using metabolomics/metabolite profiling terminology in the ISI database in May 2005 resulted in over 3000 hits, and 296 hits were still found when the search was restricted to plant-related search terms (Figure 3.4). Consequently, only select publications are presented here to serve as an overview of how these techniques may be applied to plant metabolism research.

3.4.1 Bioengineering of metabolism

In bioengineering, metabolite analysis is typically restricted to only a few target compounds or select pathways are profiled to conclude the effectiveness of a certain treatment. An example here is the selection of plant lines with high secondary metabolite levels, such as ginsenosides, for which 993 EMS mutant lines were tested by LC/MS and LC/UV [44]. If metabolite profiling or target analysis is used, hundreds of lines involving thousands of analyses can easily be scrutinized to guide the bioengineering efforts because both sample purification and instrumental methods can easily be optimized for this task. Sometimes more directed efforts in bioengineering are put forth by overexpressing novel genes into plant systems, such as for formation of ketocarotenoids in tomato and tobacco [45]. Metabolite profiling can then demonstrate the effectiveness of this transformation, and in addition, the level of formation of side products or substrates and catabolites of the primary products (such as hydroxylated intermediates). An example of how

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 74 Quark08 Quark08:Books:BLUK025-Plaxton:

metabolite profiling may unravel unexpected effects of plant bioengineering is the transformation of stilbene sythase genes from grape to tomato plants [46]. This transgenic overexpression resulted in accumulation of trans-resveratrol and trans-resveratrol-glucopyranoside, but also of soluble antioxidants such as ascorbate and glutathione. In contrast, membrane-bound antioxidants such as tocopherol and lycopene were not affected.

Another important field of research is geared toward more complex traits such as protein content or optimized carbon partitioning and fluxes. Metabolite profiling of transgenic bean plants which expressed a *Corynebacterium glutamicum* phosphoenolpyruvate carboxylase (PEPC) in a seed-specific manner [47] showed that metabolic fluxes shifted from sugars and starch into organic acids and free amino acids. This ultimately led to a gain of 20% more protein per gram seed dry weight and an increase of total seed dry weight of more than 20%. This report also shows that there is some overlap between bioengineering (complex) traits, efforts toward the biochemical relationships and the use of metabolite profiling to verify these. Consequently, a major use of metabolomics is found in plant biochemistry.

3.4.2 Plant biochemistry

3.4.2.1 Pathway analysis

Interestingly, potentially conflicting data on the biological role of PEPC were reported for Arabidopsis lines with up to 75% reduced PEPC activity [48]. In this report, it was found by ^{1}H-NMR fingerprinting that levels of various primary metabolites were indeed affected by reduced PEPC activity but without having an impact on overall plant growth. The authors concluded that this finding supported the idea that PEPC had little impact on anplerotic carbon fluxes and was just the opposite to what was found with *C. glutamicum* PEPC in transgenic beans [47]. These reports may serve to illustrate that even major conclusions may differ depending on the experimental setup (e.g. underexpression of an endogenous enzyme-coding gene vs overexpression of a transgene) and, obviously, on the plant species that is being investigated.

A particularly nice example for a successful plant biochemical study involving metabolomics is a report on malate synthase involved in Arabidopsis seedling growth [49]. In this report, it was shown that this glyoxylate shunt enzyme is partially dispensible for lipid utilization and gluconeogenisis, in contrast to bacteria. This study may also serve as an example that each enzyme needs to be studied individually to determine its functions and biological roles, *in vivo,* and that homology searches and comparisons alone carry high risks of misleading conclusions. Another excellent study compared the effects of inducible activation of cytosolic yeast invertase with constitutive gene overexpression in growing potato tubers [50]. By this approach, the primary effects of increased sucrose mobilization caused by overexpression could be distinguished from pleiotropic effects such as a switch from starch synthesis to respiration. This switch was only observed to occur in constitutive gene overexpression during plant development, but not in short-term

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 75 Quark08 Quark08:Books:BLUK025-Plaxton:

inducible overexpression. Such experiments are extremely important to better understand the control of metabolism *in vivo,* and to distinguish causes from mere observations of end-point effects.

Other studies have focused on secondary plant biochemistry [51]. In this research field, an elegant study reported the metabolic and transcriptional characterization of phenylpropanoid biosynthesis in Arabidopsis. Metabolomic approaches could show that despite absence of phenotypic alterations, specific functions of the Arabidopsis phenylalanine lyases PAL1 and PAL2 were elaborated [52]. The authors showed that PAL1 is the primary factor involved in the formation of phenylpropanoids, as well as more complex and less understood alterations in other metabolic modules.

Further reports apply plant metabolomics by integrating classical methods such as isotope labeling or by extending data mining toward network analysis. Isotope labeling with ^{13}C tracers is particularly useful for ^{13}C-NMR analysis. Consequently, this technique has been used for plant biochemistry. A study on rice coleoptiles under anaerobic conditions revealed that glutamine and malate pools were generated from multiple turns of the TCA cycle and that there was a high contribution of the glyoxylate shunt toward malate formation under these conditions [53]. Primary metabolism was also the focus of a study investigating potato plants with underexpression of sucrose synthase II that was found to be primarily localized in vascular tissues [54]. Largely different effects of underexpression of this enzyme isoform were found for source and sink tissues, with major effects on control of sugar alcohol metabolism in leaves and of control of amino acid metabolism in tubers. Besides classical and multivariate statistics, these effects were revealed by directly ranking differences in metabolic network connectivities.

Comparative correlation analysis has also been used to study control of primary metabolism responding to elicitation by methyl jasmonate, yeast elicitor or UV light [55]. In this study, glycine, serine and threonine pathways were found to be perturbed and induction of threonine aldolase activity was suggested from these data.

3.4.2.2 Flux measurements

The result of metabolomic analyses is a series of measurements of metabolite levels, that is, snapshots of metabolism. However, this is just one way to study the control of metabolism. Metabolic snapshot data are usually not sufficient to directly derive enzyme activities and hierarchical structures of pathways and, even more importantly, the dynamics of carbon partitioning between the different organs or the different metabolic cycles. A range of possible scenarios may explain a finding of altered metabolite levels: anabolic reactions might be faster, the catabolic fate might be different or transport activities may have been changed. Conversely, even if metabolite levels remain unchanged in an experiment, the underlying enzyme activities may still be different. If, for example, both anabolic and catabolic reactions are altered in the same way and intensity, the steady state levels of a given metabolite should not change (although fluxes are increased), Therefore, metabolite snapshot data should be complemented by flux data.

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 76 Quark08 Quark08:Books:BLUK025-Plaxton:

Theoretically, it should be possible to derive fluxes from snapshot measurements if we had the ability to measure true concentrations of all substrates and products at faster time intervals, assuming we would know the total network structure. However, all these constraints are not fulfilled at the current state of metabolomic practice, as outlined above. Even if snapshots were taken in a time series, and even if all substrates and products of a 'pathway' were covered, today we are still unable to detect the flow of products back into the pathway (either by reversible reactions or via other routes through the metabolic network) or find potential new side fluxes out of (or into) the pathway by unforeseen additional enzymatic activities. Here, the use of labeled compounds is most appropriate, either by radioactive labeling or by stable isotope tracers (e.g. isotopomer analysis). These techniques, however, are also restricted in use by the need to feed in labeled substrates which (a) may not be taken up quickly enough and (b) are then quickly diluted through the metabolic network. Therefore, only short distances or small parts of the total metabolic network can be imputed that have reasonably high metabolic turnover rates or that lead to and from strong carbon sinks such as starch. A global view on all metabolic fluxes (a 'fluxome' [4]) is still out of reach by current techniques, even if fluxes are inferred from other metabolic sinks such as proteins. For vascular plants, a highly suitable way to analyze fluxes is to use NMR-based techniques [56]. Using current methods, a combination of metabolomic snapshot data at high number of biological replicates (to get the breadth of metabolic networks at high statistical significance levels) and flux measurements (on select and important pathways [57]) therefore seems to be the most practical solution to reach a more complete picture of metabolic control and regulation.

3.4.3 Physiological studies

Physiological adaptation to environmental stress has been the focus of several studies on Arabidopsis plants. By comparing Ws-2 and Cvi-1 ecotypes to Arabidopsis lines overexpressing CBF transcription factor genes, it has been shown that CBF overexpression configured the metabolome of Arabidopsis in a way that resembles cold acclimation treatments [58], proving the major contribution of this gene family in the cold response pathway. Two other studies compared plants under sulfur stress. By using both mass-spectrometry-based metabolite fingerprints and gene transcription levels [59], it was confirmed that an immediate response on sulfur starvation is a decrease in glucosinolate biosynthesis. Another report combined data from GC/MS and LC/MS measurements upon sulfur deprivation [60] and found that the metabolic system was rebalancing not only sulfur metabolism but also partitioning of carbon and nitrogen in a time dependent manner. This occurred mainly by accumulation along the O-acetylserine-serine-glycine pathway which led to storage of nitrogen in glutamine and allantoin pools. Apart from nitrogen shuffling, further effects were found in lipid catabolism, purine metabolism and enhanced photorespiration.

Other reports have focused on metabolic effects upon environmental perturbations. For example, responses to heat stress, drought stress and a combination of

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 77 Quark08 Quark08:Books:BLUK025-Plaxton:

heat and drought stress were compared in Arabidopsis by metabolite profiling [61]. It was suggested that sucrose and other sugars replace proline as the major osmoprotectant under combined drought and heat stress, whereas proline levels were controlled in response to drought stress alone. Another important area of very active research has been the study of metabolic responses to pests and pathogens. Interestingly, an infestation of tomato plants with spider mites *(Tetranychus urticae)* caused a delay of 4 days between increased levels of terpene biosynthetic transcripts and the emission of volatile terpenoids [62]. This work sheds light on the time-decoupling of control layers in metabolic responses, which also questions the validity of approaches using direct transcript–metabolite correlations for generating hypothesis on the functional annotation of metabolic genes [63].

In order to distinguish cause and effects, most reports involve studying the time dependence of metabolic responses instead of mere data associations or multivariate clustering. A good example of studying the hormonal effects on control of metabolism was an investigation of a 48-day time course upon elicitation of Arabidopsis roots with salicylic acid, jasmonic acid, chitosan and two fungal cell wall elicitors [64]. Upon treatment, 289 secondary metabolite peaks were profiled by LC/MS of which 10 peaks were confirmed by NMR structural elucidation to be compounds exhibiting antimicrobial activity at concentrations detected in the root exudates. Investigating metabolic relationships has been the focus of a physiological study on sink–source transitions in developing aspen leaves [65]. Besides confirming anticipated changes in sugar and amino acid metabolism, the study also revealed that control of nitrogen storage (determined by altered asparagine concentrations) was sequestered by changes in malate concentration and transaminase activity in this developmental time course.

3.4.4 Plant metabolomic methods

Novel insights into the control of metabolism go hand in hand with improved methods. For this reason, various research groups validated the usefulness of novel methods or improved protocols by using model plant experiments. Among these methods is the use of vapor phase extraction for phytohormone analysis upon pathogen infection of Arabidopsis plants [66], fractionation methods applied to rice plants to reduce metabolite complexity prior to instrumental analysis [67], or application of capillary electrophoresis coupled to mass spectrometry for rice leaf analysis [68], a method that was primarily used previously for microbial metabolomics. Volatile compounds responding to infection with pathogens are best monitored by gas chromatography/mass spectrometry, for example using head space sampling [69]. For example, when applied to onion bulbs that were inoculated with *Erwinia carotovora ssp. carotovora*, *Fusarium oxysporum* or *Botrytis allii* [70], 259 compounds were detectable in the headspace of which 25 were found to be specific to one or more pathogens. For phytohormones, an elegant and simple novel extraction procedure has been proposed [71]: plant tissues were extracted in mixtures of n-propanol and dichloromethane followed by methylation using trimethylsilylcliazomethane. The reaction products were heated, collected on a

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 78 Quark08 Quark08:Books:BLUK025-Plaxton:

polymeric absorbent and analyzed by GC/NCI-MS. This approach has been applied to Arabidopsis, tobacco and tomato plants, enabling to quantify signaling crosstalk interactions at the level of synthesis and accumulation of phytohormones.

Most methods used for studying metabolic control involve invasive techniques. A new idea has recently been proposed to enable detecting metabolites *in vivo* and in real time at subcellular resolution by using protein-based nanosensors based on FRET fluorescence-based microscopy. The prototypes of these sensors have been shown to work in yeast and in mammalian cell cultures [72, 73]. An extension to multiple metabolites and detection in plant cells would offer extreme versatility and direct insights into metabolic control, e.g. following time courses after glucose pulses or inhibition of specific enzymes. Analyzing time-related metabolomic data has also been the focus of a study of growth related gradients in poplar trees by magic angle spinning NMR analysis [74]. When investigating trees with underexpression of the PttMYB76 gene involved in the phenylpropanoid and lignification pathways, growth-related metabolic gradients were detected in the plant internode direction. Factors affecting NMR spectra have been investigated for potato and tomato samples [75]. This study emphasizes that, as with any method, great care must be taken to control method parameters in order to allow robust assessments of metabolite levels over hundreds of samples.

Methods based on gas chromatography/mass spectrometry have evolved in two directions. Two methods have been published that avoid detailing individual metabolites, but rather compare full spectra sections in order to align and compare hundreds of chromatograms, followed by multivariate analysis and retrospective investigation of differences related to the plant experimental designs [76, 77]. However, apart from looking at differences in the control of metabolism, it is equally interesting to note which metabolites are tightly regulated at a defined steady state level. That is, which metabolites are not altered between experiments. Hence, it seems a favorable option to quantify levels for each individual compound that is detectable in the metabolomic experiment. For GC/MS, LC/MS or CE/MS approaches, this involves individual peak detection with subsequent mass spectral deconvolution of overlapping peaks (Figure 3.5) and peak alignment by retention indices, in order to be able to compare data between experiments, laboratories or metabolomic databases. Examples for this strategy is the compilation of known and unknown metabolites from *Lotus japonicus* nodules, roots, leaves and flowers [78] or the investigation of metabolites from Arabidopsis leaves [16].

3.4.5 Food science

The last typical field of application of metabolomics in control of metabolism is food science. Food quality is easily and rapidly deteriorated by a number of different pests, and therefore it must be tightly monitored to prevent major losses. One typical report on such efforts is the use of GC/MS to profile volatile metabolites in apple to diagnose fungal infections [79]. In this report, four different fungi were applied to infect apples and the responses to these infections were classified by 20 significant biomarkers out of a total of 498 detected different volatile metabolite peaks. Such methods enable a

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 79 Quark08 Quark08:Books:BLUK025-Plaxton:

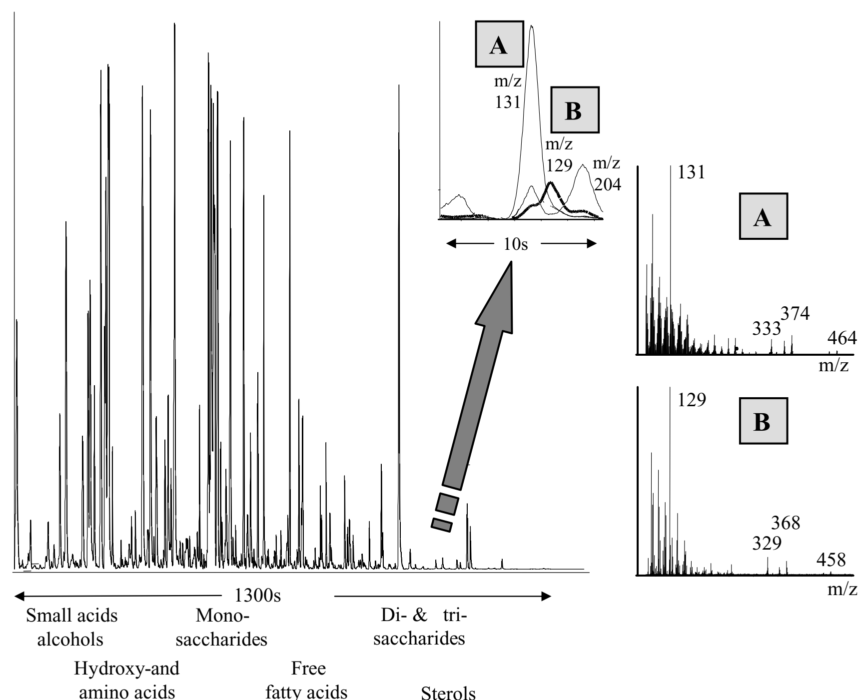

Figure 3.5 Deconvolution of overlapping peak analysis of an Arabidopsis eaf extract by GC-TOF. Left panel: primary plant metabolites are separated within 1300 s. Due to the complexity of metabolomic extracts, co-elution of compounds is inevitable. Right panels: Mass spectra of d6-cholesterol (A) and cholesterol (B) are deconvoluted by automatic algorithms (ChromaTOF 3.0), enabling unambiguous metabolite annotations.

rapid survey during food storage by observing metabolic effects for disease diagnostics rather than trying to understand the biochemical or physiological control mechanisms. A more recent but important branch of research tries to improve nutritional quality or metabolic traits in foods by genetic breeding using the analysis of quantitative trait loci (QTL). Metabolic effects were assessed as a result of the introgression of a 9 cM region of the wild tomato species, *Lycopersicon pennelli*, into a cultivated tomato line (*Lycopersicon esculentum* IL9-2-5) [80]. Metabolite contents in ripe fruits were found to have increased sucrose and glucose levels that were due to altered kinetic properties of a fruit apoplastic invertase. A few other metabolic perturbations were found, including aspartate and alanine biosynthesis.

In food quality control, concerns have arisen that genetic modifications may result in potentially harmful or undesirable metabolite alterations. In order to study the substantial equivalence of genetically modified (GM) potato tubers to classical cultivars, 40 GM lines, modified in primary carbon metabolism, glycoprotein processing or polyamine and ethylene metabolism, were analyzed by NMR and LC-UV [81]. Differences in average metabolite levels were less than threefold, which was

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 80 Quark08 Quark08:Books:BLUK025-Plaxton:

found to be negligible when compared with the natural variability within each cultivated tuber population.

3.5 Outlook

Most of the published work in plant metabolite analysis is, so far, either classical hypothesis-driven target analysis or multitarget metabolite profiling, that is restricted to usually below 100 identified compounds. Although considerable hypothesis-driven research can be undertaken using these methods, the prospects of using truly unbiased metabolomics are alluring. Two major bottlenecks need to be tackled. The first is that too many metabolic peaks remain unidentified. This raises concerns that many of these may not genuinely reflect control of metabolic states but rather arise from insignificant enzymatic side reactions, or, even worse, are indeed chemical artifacts produced during sample preparation. In principle, this argument is hard to rebut, especially as long as there are no consistent metabolomic databases and no major efforts for rapid identification *de novo* of unknown compounds. Various metabolomic databases are or will be made public in the near future. However, it is doubtful how many known metabolites these databases will include. The second major drawback is a gap in the interpretation of metabolic snapshot data. Very often, the general finding of metabolomic studies is that a large number of compounds have been altered in response to a given experimental treatment. Such observations remain mere physiological descriptions if other levels of information are not integrated to result in a comprehensive picture of plant biology (for example, spatial and temporal resolution of such metabolic snapshot data). A further way to improve the interpretation of snapshot data is to refer these to a 'plant physiology and plant biochemistry knowledge database' that may be inferred from both theoretical considerations (such as metabolic control analysis) and text mining approaches. The ultimate aim of all these efforts must remain an understanding of events and effects in plant physiology, which can be tested by constructing data models and predict metabolic alterations under (genotype *x* environment) conditions that were not tested before [82]. This level of understanding of plant metabolism is still out of reach, but with modern methods in dissecting plants on the molecular and cellular levels it is not impossible anymore!

References

1. O. Fiehn (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology*, **48**, 155–171.
2. S. Oliver (1997) Yeast as a navigational aid in genome analysis. *Microbiology* **143**, 1483–1487.
3. E.C. Horning and M.G. Horning (1971) Human metabolic profiles obtained by GC and GC/MS. *Journal of Chromatographic Science* **9**, 129–140.
4. N. Zamboni and U. Sauer (2004) Model-independent fluxome profiling from H-2 and C-13 experiments for metabolic variant discrimination. *Genome Biology* **5**, Art. No. R99 (http:// genomebiology.com/).

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 81 Quark08 Quark08:Books:BLUK025-Plaxton:

5. D. Fell (1997) *Understanding the Control of Metabolism.* Portland Press, London.
6. J. Kikuchi, K. Shinozaki and T. Hirayama (2004) Stable isotope labeling of Arabidopsis thaliana for an NMR-based metabolomics approach. *Plant and Cell Physiology* **45**, 1099–1104.
7. J. Buettner (2005) Impacts of laboratory methodology on medical thinking in the 19th century. *Clinical Chemistry and Laboratory Medicine* **38**, 57–63.
8. M. Tswett (1906) Physical chemical studies on chlorophyll adsorptions. *Berichte der Deutschen botanischen Gesellschaft* **24**, 316–323.
9. R.A. Friedel and M. Orchin (1951) *Ultraviolet Spectra of Aromatic Compounds.* John Wiley & Sons, New York.
10. P. Krishnan, N.J. Kruger and R.G. Ratcliffe (2005) Metabolite fingerprinting and profiling in plants using NMR. *Journal of Experimental Botany* **56**, 255–265.
11. N.J.C. Bailey, M. Oven, E. Holmes, M.H. Zenk and J.K. Nicholson (2004) An NMR-based metabolomic approach to the analysis of the effects of xenobiotics on endogenous metabolite levels in plants. *Spectroscopy – An International Journal* **18**, 279–287 Sp. Iss. SI.
12. J. Gullberg, P. Jonsson, A. Nordstrom, M. Sjostrom and T. Moritz (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. *Analytical Biochemistry* **331**, 283–295.
13. R.G. Brereton, A. Rahmani, Y.Z. Liang and O.M. Kvalheim (1994) Investigation of the allomerization reaction of Chlorophyll-alpha. Use of diode-array HPLC, mass spectrometry and chemometric factor analysis for the detection of early products. *Photochemistry and Photobiology* **59**, 99–110.
14. K. Hyvarinen, J. Helaja, P. Kuronen, I. Kilpelainen and P.H. Hynninen (1995) H-1 and C-13 NMR-spectra of the methanolic allomerization products of 13(2)(R)-Chlorophyll-alpha. *Magnetic Resonance in Chemistry* **33**, 646–656.
15. O. Fiehn, J. Kopka, P. Doermann, T. Altmann, R. Trethewey and L. Willmitzer (2000) Metabolic profiling for plant functional genomics. *Nature Biotechnology* **18**, 1157–1161.
16. W. Weckwerth, K. Wenzel and O. Fiehn (2004) 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.
17. J. Szopa, G. Wilczynski, O. Fiehn, A.Wenczel and L. Willmitzer (2001) Identification and quantification of catecholamines in potato plants (*Solarium tuberosum*) by GC-MS. *Phytochemistry* **58**, 315–320.
18. W. Schwab (2003) Metabolome diversity: too few genes, too many metabolites?. *Phytochemistry* **62**, 837–849.
19. C. Lindermayr, J. Fliegmann and J. Ebel (2003) Deletion of a single amino acid residue from different 4-coumarate: CoA ligases from soybean results in the generation of new substrate specificities. *Journal of Biological Chemistry* **278**, 2781–2786.
20. J.M. Clark, K.A. Daum and J.H. Kalivas (2003) Demonstrated potential of ion mobility spectrometry for detection of adulterated perfumes and plant speciation. *Analytical Letters* **36**, 215–244.
21. J.L. Hope, B.J. Prazen, E.J. Nilsson, M.E. Lidstrom and R.E. Synovec (2005) Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection: analysis of amino acid and organic acid trimethylsilyl derivatives, with application to the analysis of metabolites in rye grass samples. *Talanta* **65**, 380–388.
22. H.R. Tang, Y.L. Wang, J.K. Nicholson and J.C. Lindon (2004) Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma. *Analytical Biochemistry* **325**, 260–272.
23. C.A. Hughey, R.P. Rodgers and A.G. Marshall (2002) Resolution of 11000 compositionally distinct components in a single electrospray ionization Fourier transform ion cyclotron resonance mass spectrum of crude oil. *Analytical Chemistry* **74**, 4145–4149.
24. A.C. Stenson, A.G. Marshall and W.T. Cooper (2003) Exact masses and chemical formulas of individual Suwannee River fulvic acids from ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectra. *Analytical Chemistry* **75**, 1275–1284.

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 82 Quark08 Quark08:Books:BLUK025-Plaxton:

25. S.J. Murch, H.P. Rupasinghe, D. Goodenowe and P.K. Saxena (2004) A metabolomic analysis of medicinal diversity in Huang-qin (Scutellaria baicalensis Georgi) genotypes: discovery of novel compounds. *Plant Cell Reports* **23**, 419–425.
26. E. von Roepenack-Lahaye, T. Degenkolb, M. Zerjeski, *et al.* (2003) Profiling of Arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. *Plant Physiology* **134**, 548–559.
27. V.V. Tolstikov, A. Lommen, K. Nakanishi, N. Tanaka and O. Fiehn (2003) Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. *Analytical Chemistry* **75**, 6737–6740.
28. A. Bachmann, B. Hause, H. Maucher *et al.* (2002) Jasmonate-induced lipid peroxidation in barley leaves initiated by distinct 13-LOX forms of chloroplasts. *Biological Chemistry* **383**, 1645–1657.
29. S. Goto, T. Nishioka and M. Kanehisa (2000) LIGAND: chemical database of enzyme reactions. *Nucleic Acids Research* **28**, 380–382.
30. C. Steinbeck (2001) The automation of natural product structure elucidation. *Current Opinion in Drug Discovery & Development* **14**, 338–342.
31. J.L. Wolfender, K. Ndjoko and K. Hostettmann (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. *Journal of Chromatography A* **1000**, 437–455.
32. H. Jenkins, N. Hardy, M. Beckmann *et al.* (2004) A proposed framework for the description of plant metabolomics experiments and their results. *Nature Biotechnology* **22**, 1601–1605.
33. R.J. Bino, R.D. Hall, O. Fiehn *et al*. (2004) Potential of metabolomics as a functional genomics tool. *Trends in Plant Science* **9**, 418–425.
34. J. Kopka, N. Schauer, S. Krueger *et al.* (2005) GMD@CSB.DB: the Golm metabolome database. *Bioinformatics* **21**, 1635–1638.
35. W.P. Ridley, R.D. Shillito, I. Coats *et al.* (2004) Development of the international life sciences institute crop composition database. *Journal of Food Composition and Analysis* **17**, 423–438.
36. E.S. Wurtele, J. Li, L.X. Diao *et al.* (2003) MetNet: software to build and model the biogenetic lattice of Arabidopsis. *Comparative and Functional Genomics* **4**, 239–245.
37. J. Lindon (chair). Author group from organisations: Imperial College (London), Nestlé, Aventis, Novo Nordisk, Bristo-Myers Squibb, TNO, Umetrics, University of Wales (Aberystwyth), GlaxoSmithKline, Unilever, Pfizer, European Bioinformatics Institute, The Medical Research Council, Bruker, Roche, Eli Lilly (2005) The standard metabolic reporting structure – an open standard for reporting metabolic data. http://www.smrsgroup.org/, March 09, 2005.
38. W.A. Warr (ed.) (2004) IUPAC Project Meetings: Extensible Markup Language (XML) Data dictionaries and chemical identifier. http://www.warr.com/inchi.pdf NIST, Gaithersburg, MA, USA, June 2004.
39. P. Murray-Rust, H. Rzepa and S.E. Stein (2005) The INChI as an LSID for molecules in lifescience. http://lists.w3.org/Archives/Public/public-swls-ws/2004Sep/att-0026/inchi.html, March 9, 2005.
40. E. Manduchi, G.R. Grant, H. He *et al.* (2004) RAD and the RAD study-annotator: an approach to collection, organization and exchange of all relevant information for high-throughput gene expression studies. *Bioinformatics* **20**, 452–459.
41. D.L. Wheeler, T. Barrett, D.A. Benson *et al.* (2005) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **33**, D39–D45 (Special Issue).
42. S.Y. Rhee, W. Beavis, T.Z. Berardini *et al*. (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Research* **31**, 224–228.
43. Plant Ontology Consortium (POC) (2002) The plant ontology™ consortium and plant ontologies. *Comparative and Functional Genomics* **3**(2), 137–142.
44. S.S. Woo, J.S. Song, J.Y. Lee *et al.* (2004) Selection of high ginsenoside producing ginseng hairy root lines using targeted metabolic analysis. *Phytochemistry* **65**, 2751–2761.
45. L. Ralley, E.M.A. Enfissi, N. Misawa, W. Schuch, P.M. Bramley and P.D. Fraser (2004) Metabolic engineering of ketocarotenoid formation in higher plants. *Plant Journal* **39**, 477–486.

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 83 Quark08 Quark08:Books:BLUK025-Plaxton:

46. G. Giovinazzo, L. D'Amico, A. Paradiso, R. Bollini, F. Sparvoli and L. DeGara (2005) Antioxidant metabolite profiles in tomato fruit constitutively expressing the grapevine stilbene synthase gene. *Plant Biotechnology Journal* **3**, 57–69.
47. H. Rolletschek, L. Borisjuk, R. Radchuk *et al.* (2004) Seed-specific expression of a bacterial phosphoenolpyruvate carboxylase in *Vicia narbonensis* increases protein content and improves carbon economy. *Plant Biotechnology Journal* **2**, 211–219.
48. A. Moing, M. Maucourt, C. Renaud *et al.* (2004) Quantitative metabolic profiling by 1-dimensional H-1-NMR analyses: application to plant genetics and functional genomics. *Functional Plant Biology* **31**, 889–902.
49. J.E. Cornah, V. Germain, J.L. Ward, M.H. Beale and S.M. Smith (2004) Lipid utilization, gluconeogenesis, and seedling growth in Arabidopsis mutants lacking the glyoxylate cycle enzyme malate synthase. *Journal of Biological Chemistry* **279**, 42916–42923.
50. B.H. Junker, R. Wuttke, A. Tiessen *et al.* (2004) Temporally regulated expression of a yeast invertase in potato tubers allows dissection of the complex metabolic phenotype obtained following its constitutive expression. *Plant Molecular Biology* **56**, 91–110.
51. K. Morino, F. Matsuda, H. Miyazawa, A. Sukegawa, H. Miyagawa and K. Wakasa (2005) Metabolic profiling of tryptophan-overproducing rice calli that express a feedback-insensitive alpha subunit of anthranilate synthase. *Plant and Cell Physiology* **46**, 514–521.
52. A. Rohde, K. Morreel, J. Ralph *et al.* (2004) Molecular phenotyping of the pal1 and pal2 mutants of Arabidopsis thaliana reveals far-reaching consequences on and carbohydrate metabolism. *Plant Cell* **16**, 2749–2771.
53. T.W.M. Fan, A.N. Lane and R.A. Higashi (2003) *In vivo* and *in vitro* metabolomic analysis of anaerobic rice coleoptiles revealed unexpected pathways. *Russian Journal of Plant Physiology* **50**, 787–793.
54. W. Weckwerth, M.E. Loureiro, K. Wenzel and O. Fiehn (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7809–7814.
55. C.D. Broeckling, D.V. Huhman, M.A. Farag *et al.* (2005) Metabolic profiling of Medicago truncatula cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. *Journal of Experimental Botany* **56**, 323–336.
56. R.G. Ratcliffe and Y. Shachar-Hill (2005) Revealing metabolic phenotypes in plants: inputs from NMR analysis. *Biology Reviews* **80**, 27–43.
57. A.R. Fernie, P. Geigenberger and M. Stitt (2005) Flux an important, but neglected, component of functional glenomics. *Current Opinion in Plant Biology* **8**, 174–182.
58. D. Cook, S. Fowler, O. Fiehn and M.F. Tomashow (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15243–15248.
59. M.Y. Hirai, M. Yano, D.B. Goodenowe *et al.* (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10205–10210.
60. V.J. Nikiforova, J. Kopka, V. Tolstikov *et al.* (2005) Systems rebalancing of metabolism in response to sulfur deprivation, as revealed by metabolome analysis of Arabidopsis plants. *Plant Physiology* **138, 304–318**.
61. L. Rizhsky, H.J. Liang, J. Shuman, V. Shulaev, S. Davletova and R. Mittler (2004) When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. *Plant Physiology* **134**, 1683–1696.
62. M.R. Kant, K. Ament, M.W. Sabelis, M.A. Haring and R.C. Schuurink (2004) Differential timing of spider mite-induced direct and indirect defenses in tomato plants. *Plant Physiology* **135**, 483–495.
63. E. Urbanczyk-Wochniak, A. Luedemann, J. Kopka *et al.* (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Reports* **4**, 989–993.
64. T.S. Walker, H.P. Bais, K.M. Halligan, F.R. Stermitz and J.M. Vivanco (2003) Metabolic profiling of root exudates of Arabidopsis thaliana. *Journal of Agricultural and Food Chemistry* **51**, 2548–2554.

BLUK025-C03[60-84].qxd 10/20/05 5:16 AM Page 84 Quark08 Quark08:Books:BLUK025-Plaxton:

65. M.L. Jeong, H.Y. Jiang, H.S. Chen, C.J. Tsai and S.A. Harding (2004) Metabolic profiling of the sink-to-source transition in developing leaves of quaking aspen. *Plant Physiology* **136**, 3364–3375.
66. E.A. Schmelz, J. Engelberth, J.H. Tumlinson, A. Block and H.T. Alborn (2003) The use of vapor phase extraction in metabolic profiling of phytohormones and other metabolites. *Plant Journal* **39**, 790–808.
67. T. Frenzel, A. Miller and K.H. Engel (2002) Metabolite profiling – a fractionation method for analysis of major and minor compounds in rice grains. *Cereal Chemistry* **79**, 215–221.
68. S. Sato, T. Soga, T. Nishioka and M. Tomita (2004) Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. *Plant Journal* **40**, 151–163.
69. L.H. Lui, A. Vikram, Y. Abu-Nada, A.C. Kushalappa, G.S.V. Raghavan and K. Al-Mughrabi (2005) Volatile metabolic profiling for discrimination of potato tubers inoculated with dry and soft rot pathogens. *American Journal of Potato Research* **82**, 1–8.
70. B. Prithiviraj, A. Vikram, A.C. Kushalappa and V. Yaylayan (2004) Volatile metabolite profiling for the discrimination of onion bulbs infected by Erwinia carotovora ssp carotovora, Fusarium oxysporum and Botrytis allii. *European Journal of Plant Pathology* **110,** 371–377.
71. E.A. Schmelz, J. Engelberth, H.T. Alborn *et al.* (2003) Simultaneous analysis of phytohormones, phytotoxins, and volatile organic compounds in plants. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10552–10557.
72. S. Okumoto, K. Deuschle, M. Fehr *et al.* (2004) Genetically encoded sensors for ions and metabolites. *Soil Science and Plant Nutrition* **50**, 947–953.
73. M. Fehr, D.W. Ehrhardt, S. Lalonde and W.B. Frommer (2004) Minimally invasive dynamic imaging of ions and metabolites in living cells. *Current Opinion in Plant Biology* **7**, 345–351.
74. S. Wiklund, M. Karlsson, H. Antti *et al.* (2005) A new metabonomic strategy for analysing the growth process of the poplar tree. *Plant Biotechnology Journal* **3**, 353–362.
75. M. Defernez and I.J. Colquhoun (2003) Factors affecting the robustness of metabolite fingerprinting using H-1 NMR spectra. *Phytochemistry* **62**, 1009–1017.
76. P. Jonsson, J. Gullberg, A. Nordstrom *et al.* A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Analytical Chemistry* **204**, 1738–1745.
77. A.L. Duran, J. Yang, L.J. Wang and L.W. Sumner (2003) Metabolomics spectral formatting, alignment and conversion tools (MSFACTs). *Bioinformatics* **19**, 2283–2293.
78. G.G. Desbrosses, J. Kopka and M.K. Udvardi (2005) Lotus japonicus metabolic profiling. Development of gas chromatography-mass spectrometry resources for the study of plant-microbe interactions. *Plant Physiology* **137**, 1302–1318.
79. A. Vikram, B. Prithiviraj, H. Hamzehzarghani and A. Kushalappa (2004) Volatile metabolite profiling to discriminate diseases of McIntosh apple inoculated with fungal pathogens. *Journal of the Science of Food and Agriculture* **84**, 1333–1340.
80. C. Baxter, F. Carrari, A. Bauke *et al.* (2005) Fruit carbohydrate metabolism in an introgression line of tomato with increased fruit soluble solids. *Plant and Cell Physiology* **46**, 425–437.
81. M. Defernez, Y.M. Gunning, A.J. Parr, L.V.T. Shepherd, H.V. Davies and I.J. Colquhoun (2004) NMR and HPLC-UV profiling of potatoes with genetic modifications to metabolic pathways. *Journal of Agricultural and Food Chemistry* **52**, 6075–6085.
82. L.V. Lejay, D.E. Shasha, P.M. Palenchar *et al.* (2004) Adaptive combinatorial design to explore large experimental spaces: approach and validation. *Systems Biology* **2**, 1–7.