



Metabolite fingerprinting and profiling in plants using NMR

P. Krishnan^{1,2}, N. J. Kruger¹ and R. G. Ratcliffe^{1,*}

¹ Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

² Division of Biochemistry, Plant Physiology, and Environmental Sciences, Central Rice Research Institute, Cuttack (Orissa) 753 006, India

Received 20 May 2004; Accepted 31 August 2004

Abstract

Although less sensitive than mass spectrometry (MS), nuclear magnetic resonance (NMR) spectroscopy provides a powerful complementary technique for the identification and quantitative analysis of plant metabolites either *in vivo* or in tissue extracts. In one approach, metabolite fingerprinting, multivariate analysis of unassigned ¹H NMR spectra is used to compare the overall metabolic composition of wild-type, mutant, and transgenic plant material, and to assess the impact of stress conditions on the plant metabolome. Metabolite fingerprinting by NMR is a fast, convenient, and effective tool for discriminating between groups of related samples and it identifies the most important regions of the spectrum for further analysis. In a second approach, metabolite profiling, the ¹H NMR spectra of tissue extracts are assigned, a process that typically identifies 20–40 metabolites in an unfractionated extract. These profiles may also be used to compare groups of samples, and significant differences in metabolite concentrations provide the basis for hypotheses on the underlying causes for the observed segregation of the groups. Both approaches generate a metabolic phenotype for a plant, based on a system-wide but incomplete analysis of the plant metabolome. However, a review of the literature suggests that the emphasis so far has been on the accumulation of analytical data and sample classification, and that the potential of ¹H NMR spectroscopy as a tool for probing the operation of metabolic networks, or as a functional genomics tool for identifying gene function, is largely untapped.

Key words: ¹H NMR, mass spectrometry, metabolic phenotype, metabolite fingerprinting, metabolite profiling, metabolomics, nuclear magnetic resonance spectroscopy, plant metabolism.

Introduction

Metabolomics seeks to identify and quantify the complete set of metabolites in a cell or tissue type, and to do so as quickly as possible and without bias (Sumner *et al.*, 2003; Weckwerth, 2003). To achieve this objective, or at least to approach it, it is necessary to draw on a range of analytical methods. Mass spectrometry (MS) has established itself as the method of choice, but complementary information from other techniques, particularly nuclear magnetic resonance (NMR) spectroscopy is potentially useful in extending the coverage of the metabolome. In practice, most investigations of the plant metabolome tend to be based on either MS or NMR, with little attempt to exploit the synergy between them. In part this may reflect preferences arising from the analytical strengths of different research groups, as well as potential difficulties in reconciling and handling large datasets from two techniques based on unrelated physical principles, but it also reflects the usefulness of NMR and MS datasets in their own right. With the latter point in mind, it is useful to examine the practicality of using NMR as a source of data for plant metabolomic analysis and to review the emerging applications of NMR in metabolomics.

Until relatively recently the notion of developing experimental strategies for plant metabolomic analysis would have been greeted with some scepticism since there was little perceived need for such an unfocused approach. However, when the need arose for high-throughput, system-wide analyses of plant metabolism it was possible to harness existing MS and NMR methods that were already well-suited to generating the necessary data. These methods had been available for many years, and their scope for the analysis of complex mixtures was already well known, but these factors were not in themselves sufficient to trigger much interest in a metabolomic approach to plant analysis. It required a biological rationale to generate interest in

* To whom correspondence should be addressed. Fax: +44 (0)1865 275074. E-mail: george.ratcliffe@plants.ox.ac.uk

system-wide metabolic analysis and this continues to be based on two considerations.

First, it has become clear that the complexity of the plant metabolic network is such that it is not yet possible to construct predictive models of metabolic performance that allow rational metabolic engineering of plant genomes (Sweetlove *et al.*, 2003; Kruger and Ratcliffe, 2004). The fundamental problem is a shortage of quantitative information on the components of the metabolic network, and the interactions between them, with the result that many of the subtleties that determine the robustness of the network are not yet captured by the available modelling approaches. The progress that is being made in developing *in silico* genome-scale models of bacterial metabolism may eventually offer a way out of this impasse (Reed and Palsson, 2003), but an equally valid response is to increase the scale and scope of traditional metabolic analysis with the aim of defining more of the parameters that determine the properties of the network (Kruger and Ratcliffe, 2004).

Second, it can be argued that the metabolome is a fundamentally important biochemical manifestation of the genome, that, in effect, it defines a metabolic phenotype and that system-wide metabolic analysis could be a useful tool for functional genomics (Fiehn *et al.*, 2000; Fiehn, 2002). Strategies for identifying the function of unknown genes on the basis of metabolomic data have been proposed (Raamsdonk *et al.*, 2001; Allen *et al.*, 2003), and more generally there is considerable interest in using metabolic phenotypes as the basis for discriminating between plants of different genotypes, or between plants subjected to different treatments (Roessner *et al.*, 2001, 2002). Whether a phenotype based on the metabolic composition of a cell or tissue is the most appropriate choice for functional genomic applications, or whether it would be more revealing from a functional perspective to use the fluxes between metabolites as the basis for defining a metabolic phenotype (Kruger *et al.*, 2003) is a matter for debate (Ratcliffe and Shachar-Hill, 2005), but there is increasing evidence, for example from investigations of transgenic plants (Roessner-Tunali *et al.*, 2003), that metabolomic analysis is a useful phenotyping tool. Moreover, the value of a metabolic phenotype, however defined, is greatly increased by the possibility of correlating the data with the system-wide analysis of gene expression and protein content (Urbanczyk-Wochniak *et al.*, 2003).

This review assesses the contribution that NMR is making to system-wide metabolite analysis in plants, providing a description of the analytical capabilities of NMR, the suitability of the method for metabolomic analysis, and a survey of the applications of the technique. At the outset it should be emphasized that NMR, as with all the other techniques that have been recruited for system-wide metabolite analysis, is restricted by sensitivity considerations to an analysis of the subset of the metabolome that exceeds a concentration threshold. Thus, within the definitions

given elsewhere (Fiehn, 2002; Sumner *et al.*, 2003), NMR either generates a metabolite profile, in which the NMR signals are assigned to specific metabolites, or a metabolite fingerprint, in which the analysis is based on the distribution of intensity in the NMR spectrum rather than the assignment of the signals.

Metabolite detection by NMR

Any molecule containing one or more atoms with a non-zero magnetic moment is potentially detectable by NMR and, since the isotopes with non-zero magnetic moments include ^1H , ^{13}C , ^{14}N , ^{15}N , and ^{31}P , all biologically important molecules have at least one NMR signal. These signals are characterized by their frequency (chemical shift), intensity, fine structure, and magnetic relaxation properties, all of which reflect the precise environment of the detected nucleus. Thus NMR spectra often contain a wealth of information about the identity of the molecules in the sample, and it is on this basis that NMR can be used to identify and quantify metabolites in samples of biological origin.

This simple conclusion masks the versatility of the NMR technique as a tool for metabolite analysis in at least three ways. First, NMR is non-destructive, and spectra can be recorded from cell suspensions, tissues, and even whole plants, as well as from extracts and purified metabolites (Ratcliffe, 1994; Ratcliffe and Shachar-Hill, 2001). Second, NMR offers an array of detection schemes that can be tailored to the nature of the sample and the metabolic problem that is being addressed (Ratcliffe *et al.*, 2001). Thus analysing the metabolite composition of a tissue extract, determining the structure of a novel metabolite, demonstrating the existence of a particular metabolic pathway *in vivo*, and localizing the distribution of a metabolite in a tissue are all possible by NMR. However, the nature of the NMR measurements that are required for these tasks, particularly in relation to the hardware requirements, the detection scheme, and the sensitivity of the analysis are very different (Ratcliffe *et al.*, 2001). Third, the natural abundance of some of the biologically relevant magnetic isotopes is low and this allows these isotopes, particularly ^2H , ^{13}C , and ^{15}N , to be introduced into a metabolic system as labels prior to the NMR analysis. This has the effect of lowering the concentration threshold for the detection of these atoms, but more importantly it allows the exploration of metabolic pathways, leading to qualitative information on the links between labelled precursors and their products and quantitative information on metabolic fluxes (Bacher *et al.*, 1999; Roberts, 2000; Kruger *et al.*, 2003).

The high-throughput, system-wide objective of metabolomics puts a premium on sensitivity and ubiquity; the aim is to detect as many metabolites as possible in the shortest possible time. For NMR, this means restricting the detection scheme to the most sensitive magnetic nuclei and avoiding

the complications associated with *in vivo* detection by analysing tissue extracts. The most sensitive commonly occurring magnetic isotope is ^1H and this is inevitably the preferred nucleus for most metabolite fingerprinting and profiling applications of NMR (Fig. 1). Other possibilities include ^{19}F , which has a comparable sensitivity to ^1H , and which can be used to profile plant tissues after treatment with fluorinated agrochemicals (Aubert and Pallett, 2000; Bailey *et al.*, 2000a); ^{31}P , which can be used to profile the more abundant phosphate esters in tissue extracts; and ^{13}C , which can be used to profile amino acids, carbohydrates, lipids, and organic acids in extracts derived from labelled or unlabelled tissues. For ^1H NMR, the concentration threshold for routine detection of a metabolite in an extract using a modern high field spectrometer is probably $10\ \mu\text{M}$, corresponding to a quantity of $5\ \text{nmol}$ in the typical sample volume of $500\ \mu\text{l}$. In practice, the achievable sensitivity is strongly dependent on the field strength of the magnet, and on the design of the probehead that allows the signals to be detected. NMR spectrometers are available with field strengths up to 21 Tesla, corresponding to a ^1H NMR frequency of 900 MHz, but most metabolic analysis is done on the more commonly available instruments that operate in the range 300–600 MHz. Since increasing field strength also increases spectral resolution, reducing the number of overlapping signals in the spectrum, it is the spectrometers at the upper end of this frequency range that are the most effective for metabolite profiling by ^1H NMR.

Probehead design has a marked influence on the achievable sensitivity at a particular field strength, and there are two possibilities that extend the scope of ^1H NMR as a metabolite profiling technique. First, samples that are only available in very small quantities can be analysed in scaled-down probeheads ('microprobes'). Volumes as small as $50\ \mu\text{l}$ can be analysed and the quality of the spectra is better than can be obtained from the same quantity of material diluted into the larger volume required for a conventional probehead. Second, the increasing availability of cryogenic probeheads, in which the sensitivity is increased by cooling the detection system, offers the prospect of a substantial improvement in the detection of signals that are at the limit of detection in conventional probeheads. Cryogenic probeheads are mainly used to record spectra from macromolecules, but they are also suitable for metabolic analysis and the first results with these probeheads confirm that they can deliver substantial gains in sensitivity for both extracts

and *in vivo* applications (Exarchou *et al.*, 2003; Hinse *et al.*, 2003).

Apart from sensitivity, the other crucial feature of any metabolomic technique is that it should be able to generate

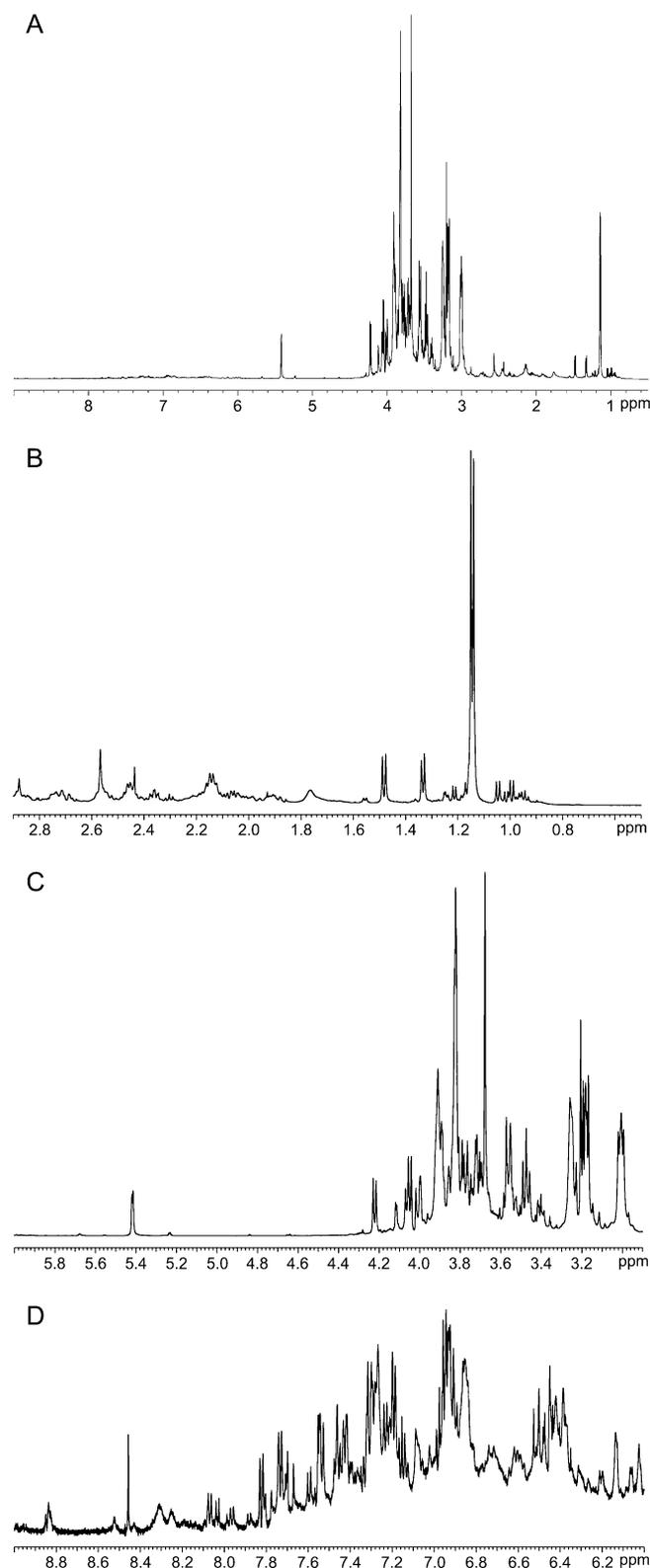


Fig. 1. ^1H NMR spectrum of a chloroform-methanol extract of 10 d rice (*Oryza sativa*) leaves recorded in D_2O , showing (A) the full spectrum and (B–D) expansions of the three main spectral regions. One-dimensional ^1H NMR spectra of tissue extracts contain a multitude of overlapping signals, with multiple signals from each detected metabolite in almost all cases. Pattern recognition techniques can be used to compare sets of spectra for fingerprinting purposes, while detailed analysis of the spectra can lead to the identification of 20–40 metabolites in typical extracts. Unpublished observations of P Krishnan, NJ Kruger and RG Ratcliffe.

identifiable signals from as many metabolites as possible. Since most molecules of biological interest contain hydrogen, ^1H NMR is again the obvious choice for system-wide metabolite analysis by NMR (Fig. 1). However, ^1H NMR suffers from the significant disadvantage that the dispersion of the signals in the NMR spectrum is rather small, resulting in extensive overlap in the signals in most regions of the spectrum. This problem is much less acute in ^{13}C , ^{15}N , ^{19}F , and ^{31}P NMR spectra, but each of these alternatives is less attractive for metabolite profiling than ^1H NMR. Thus while carbon, and to a lesser extent nitrogen, are ubiquitous, the sensitivity of NMR detection is greatly reduced by the low natural abundance of ^{13}C and ^{15}N (1.1% and 0.37%, respectively). By contrast, ^{19}F and ^{31}P are naturally abundant, and the sensitivity with which ^{19}F can be detected is comparable to ^1H , but these nuclei can only report on a very restricted subset of compounds. In the light of these considerations, it is unsurprising that ^1H NMR is the most commonly used NMR technique for metabolite profiling, but this raises the question of how the problems arising from the limited spectral dispersion can be circumvented, or at least minimized.

One option is to combine the NMR analysis with an in-line chromatographic separation technique, so that the extract is effectively fractionated before recording a sequence of NMR spectra (Lindon and Nicholson, 1997; Exarchou *et al.*, 2003). This approach, which is analogous in its effect to the coupling of gas chromatography to MS, largely eliminates the overlap problem in the ^1H NMR spectrum, and indeed it is often combined with a parallel analysis by MS to create a powerful strategy for the analysis of pre-selected groups of metabolites. The extra information that can be teased out of the ^1H NMR spectrum in this way more than compensates for the extra time required for the analysis.

The second option is to take advantage of NMR detection schemes that increase spectral resolution by distributing the signals along two frequency axes (Fig. 2; Fan, 1996; Ratcliffe *et al.*, 2001). Manipulating the NMR signals to produce a two-dimensional spectrum takes longer than a simple one-dimensional experiment, and it requires a more complicated, although still routine, detection scheme. However, the increased information content of the spectrum, in particular, the increase in the number of detected signals that are diagnostic for a specific molecule, makes this an attractive option for the analysis of tissue extracts. These experiments exploit the interactions between the NMR-detectable isotopes in a molecule, and they result either in homonuclear correlation, where the two frequency axes of the spectrum correspond to the same nucleus, usually ^1H , or in heteronuclear correlation, where one frequency axis corresponds to ^1H and the other corresponds to ^{13}C , ^{15}N , or occasionally, ^{31}P . Homonuclear correlation experiments are particularly useful in metabolite profiling, allowing linked subsets of peaks in the conventional one-dimensional

^1H NMR spectrum to be identified and assigned to particular compounds. Heteronuclear correlation experiments are also useful when the extract is derived from a tissue that has been labelled with ^{13}C or ^{15}N , and both types of two-dimensional NMR experiment have been used successfully for the analysis of plant tissue extracts (Fan, 1996).

Comparison of NMR and MS as metabolomic techniques

The next question to consider is the extent to which NMR measures up to the metabolomic ideal of a high-throughput, system-wide analytical technique and, in particular, to consider the advantages and disadvantages of NMR relative to the more commonly used MS approach. Sensitivity is perhaps the most important requirement for metabolomics, since high sensitivity favours the rapid analysis of a greater fraction of the metabolome. Here ^1H NMR, with a detection threshold of perhaps 5 nmol, is several orders of magnitude less sensitive than MS, which has a detection threshold of 10^{-12} mol (Sumner *et al.*, 2003). Moreover, with the exception of ^{19}F , the comparison becomes even less favourable for the NMR detection of most other nuclei. This difference in sensitivity translates into a more complete coverage of the metabolome with MS: 326 polar and lipophilic metabolites were detected, and 164 identified, in an analysis of *Arabidopsis thaliana* leaves based on a single phase separation (Fiehn *et al.*, 2000); while more than 150 polar metabolites were detected, and 77 identified, in an extract of potato tuber tissue (Roessner *et al.*, 2000); and over 70 metabolites were identified in a recent analysis of hexokinase overexpression in tomato plants (Roessner-Tunali *et al.*, 2003). These figures comfortably exceed the 20 (Fan *et al.*, 1988; Sobolev *et al.*, 2003), 30 (Le Gall *et al.*, 2004), or 40 (Le Gall *et al.*, 2003) metabolites that have typically been identified in metabolite profiling studies of plant samples by ^1H NMR; and while the disparity with MS may diminish as the use of in-line chromatographic separation steps in NMR increases, the greater sensitivity of MS will ensure that MS will retain the advantage of being able to analyse smaller samples.

In fact, a simple calculation suggests that both NMR and MS should be capable of detecting signals from the whole metabolome, provided the extraction procedure is scaled correctly. Given that the Michaelis constants of most enzymes are in the range 1 μM to 10 mM, it is likely that 1 μM is the lower limit for the concentration of most intracellular metabolites. Assuming that a metabolite with this concentration is restricted to just 10% of the tissue volume, the tissue content would be 0.1 nmol g^{-1} fresh weight. Thus the extraction of 50 g of tissue should permit the detection of the whole metabolome by ^1H NMR, and just 10 mg of tissue should be sufficient for MS. If it is further assumed that a typical cell might contain 5000

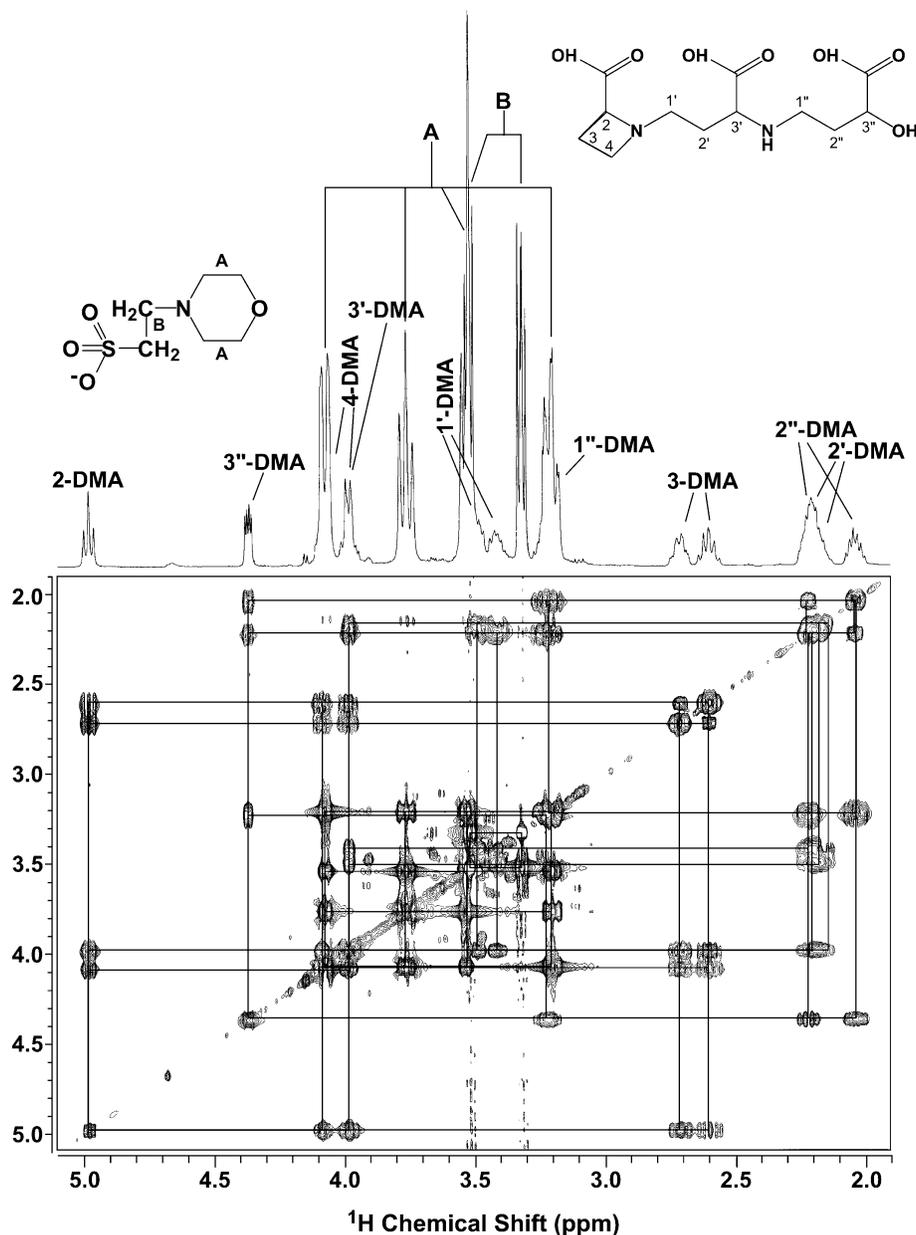


Fig. 2. Homonuclear correlation between ¹H NMR signals in the two-dimensional TOCSY spectrum of a wheat (*Triticum aestivum*) exudate. The exudate was collected under conditions that enhanced the production of 2'-deoxymugineic acid (DMA; structure top right). The normal one-dimensional ¹H NMR spectrum is shown in the projection at the top of the figure, and the TOCSY spectrum distributes the intensity in two dimensions in a way that reflects the structure of the detected metabolites. In this example the TOCSY spectrum links the hydrogen atoms in three isolated groups, revealing the expected correlations between H2, H3, and H4, between H1', H2', and H3', and between H1'', H2'', and H3''. The spectrum also contains signals from MES buffer (structure top left) apparently as a result of uptake and re-secretion. Adapted from Fan *et al.* (2001) with the permission of Elsevier Science.

metabolites, then it appears that MS identifies considerably less than 5% of the metabolome, even though the whole metabolome is potentially detectable. This argument ignores any bias against particular classes of compound arising from the extraction method, but it serves to emphasize that sensitivity is not the only issue for metabolomic techniques, and that the difficulty of detecting minor components in the presence of much larger signals may be the most serious obstacle to a complete analysis.

The actual size of the metabolome for a plant cell is unknown, and the estimate of 5000 metabolites may be conservative (Trethewey, 2004), so the number of metabolites identified by MS and NMR may be an even smaller fraction of the total than the calculation suggests. However, in considering the likely size of the metabolome, it may be useful to consider the solvation and osmotic implications of large numbers of metabolites present at concentrations that permit significant binding to typical enzymes. 10 000

soluble metabolites present at 1 μM would be equivalent to a total concentration of only 10 mM, which would have negligible implications for solvating power and water potential; whereas 10 000 metabolites at 100 μM , corresponding to the midpoint of the Michaelis constant range, would generate an implausibly high concentration of 1 M. This consideration suggests that a large fraction of the metabolome may be present at very low concentrations in pools that turn over slowly because of weak interactions with protein binding sites. The significance of an abundance of very low concentration metabolites may be difficult to assess (Nicholson and Wilson, 2003) and indeed many of them may merely reflect the haphazard way in which material flows through a metabolic network in which each metabolic conversion is the result of a chance encounter with a binding site that cannot guarantee an absolute outcome. However, in order to investigate the implications of this analysis, it becomes even more important to develop routine procedures for measuring the minor components of the metabolome.

Overlapping signals, and the dynamic range problem associated with a potential concentration range of some five orders of magnitude, limit the scope of both MS and NMR for metabolomic analysis. For example, minor signals in MS can be difficult to identify in the isotopomeric noise surrounding the mass ions of major components, and minor components can be obscured by the wings of major components as they leave the GC column. However, the precision and range of the mass measurements is an advantage in unravelling the mass spectrum, as is the routine use of GC to fractionate the extract. Similarly, the restricted chemical shift range in a one-dimensional ^1H NMR spectrum is a major hindrance to metabolite identification at even the highest magnetic fields, but the problem is alleviated by using two-dimensional NMR techniques or by fractionating the sample with LC.

There are several other factors that influence the suitability of MS and NMR as metabolomic techniques. Both methods require tissue extraction, but many classes of compound need derivatization before GC-MS, so preparing MS samples is more time-consuming and more likely to generate an unrepresentative sample. NMR also has a slight advantage for quantitative analysis, since the high stability of modern spectrometers makes this task straightforward, by contrast with MS where frequent calibration and variable retention times can complicate quantitative analysis. Both techniques usually generate multiple signals, which is an advantage for metabolite identification and a disadvantage in terms of spectral complexity. However, in MS, some of this multiplicity comes from the fragmentation of the mass ion, complicating quantitative analysis; whereas in NMR, multiple signals arise directly from the same molecule and thus provide a cross-check on metabolite quantitation.

Overall the comparison between MS and NMR shows a clear sensitivity advantage for MS and a better developed

protocol for separating the contributions from the different components of complex mixtures. However, the ease with which NMR spectra can be recorded and quantified, plus the fact that even MS seems to be capable of identifying only a small fraction of the metabolome, indicates that metabolite analysis by NMR adds value to metabolomics through complementarity. Currently, MS is closer to the metabolomic ideal than NMR, but the fingerprinting and profiling investigations described in the following sections emphasize the important role that NMR can play in the system-wide analysis of plant metabolism.

Metabolite fingerprinting of plant tissues using NMR

Fingerprinting ignores the assignment problem presented by the multitude of signals in a high resolution ^1H NMR spectrum and, instead, uses multivariate analysis to compare sets of spectra and hence the samples from which the spectra were derived. The signals in these spectra have their origin in the metabolites present in biological samples used for the NMR analysis, but the identity of the metabolites is secondary to the task of establishing whether the spectra from a set of samples are similar or different. This pattern recognition approach to the analysis of NMR spectra was developed in the biomedical field (El-Dereby, 1997; Lindon *et al.*, 2001) and, more recently, the approach has been adopted for the analysis of extracts and materials of plant origin (Defernez and Colquhoun, 2003).

In outline, metabolite fingerprinting involves sorting datasets into categories so that conclusions can be drawn about the classification of individual samples. Typically, the starting point is a principal components analysis of the digitized spectrum and this in itself may be sufficient to divide the sample set into a number of categories. Subsequently, it may be informative to investigate the variables that are most important in discriminating between the samples and this leads back to the NMR signals and the metabolites that they represent (Fig. 3). Thus the approach has the great merit of avoiding the often time-consuming process of signal assignment before it is necessary, and when that point is reached, attention is focused on those parts of the spectrum that are most relevant to the question being addressed. The approach also has the advantage that it is rapid, and that it is largely unbiased in detecting the metabolites that happen to be present in the sample, making NMR fingerprinting an attractive analytical technique for defining metabolic phenotypes.

Some of the early applications of NMR fingerprinting in plant analysis focused on food products, particularly fruit juices, with the aim of discriminating between juices derived from different varieties of the same fruit. For example, in a study of apple juices, it was shown that the ^1H NMR spectra could be used to distinguish between three types of apple with a success rate of up to 100% under favourable

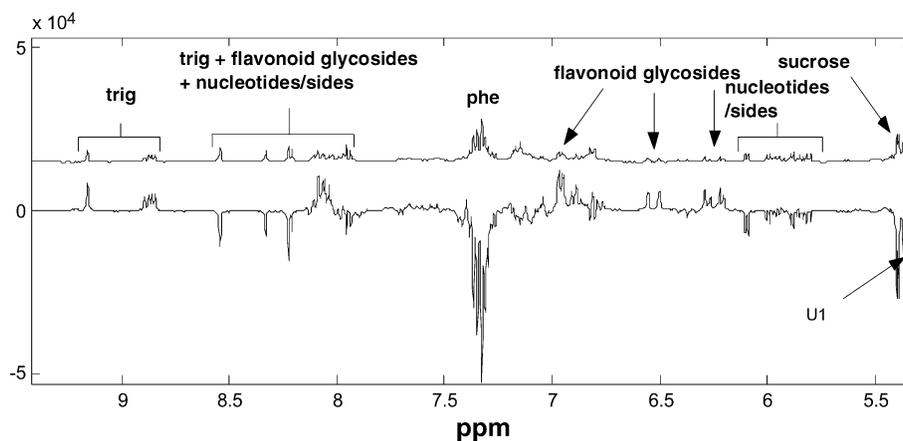


Fig. 3. Comparison between a region of the one-dimensional ^1H NMR spectrum of an extract obtained from a transgenic tomato (*Lycopersicon esculentum*) fruit and the loading for a partial least squares (PLS) score derived from a multivariate analysis of the spectra of control and transgenic tomatoes at three stages of development. This particular PLS score discriminated between control and transgenic tomatoes, and the positive and negative loadings indicate the signals that tended to be higher in the transgenic and control fruits, respectively. Adapted from Le Gall *et al.* (2003) with the permission of the American Chemical Society.

conditions (Belton *et al.*, 1998). This study was largely based on principal components analysis and inspection of the loadings identified sucrose and malate as the major, but not exclusive, basis for the classification. A similar approach was recently taken in a comparison of commercial feverfew preparations (Bailey *et al.*, 2002), and this study demonstrated the ease with which NMR, as a non-selective analytical technique, can pick out anomalous or unusual samples, as well as the power of multivariate data analysis for discriminating between sets of similar spectra.

In fact, the NMR fingerprinting principle illustrated by these investigations can be extended to many other types of comparison, and substantial investigations have been reported in several areas, including the impact of environmental factors, for example, fluctuating growth conditions (Lommen *et al.*, 1998), exposure to cadmium (Bailey *et al.*, 2003) and herbicides (Ott *et al.*, 2003), as well as comparisons between ecotypes of *Arabidopsis thaliana* (Ward *et al.*, 2003) and between wild-type and transgenic genotypes of tomato (Noteborn *et al.*, 2000; Le Gall *et al.*, 2003) and pea (Charlton *et al.*, 2004). Fingerprinting provided the basis for an informative classification of the samples in each of these applications and it is possible to draw several general conclusions about the approach from this work.

First, NMR fingerprinting has revealed substantial variability in the metabolic composition of plants grown under nominally the same conditions. For example, in an NMR analysis of tomato fruit it was found that up to 30% of the metabolites that contributed to the fingerprint varied significantly (Noteborn *et al.*, 2000) and, in a study of the mode of action of herbicides using 5–10-d-old maize seedlings, several hours under altered conditions following a growth cabinet malfunction was sufficient to produce detectable effects on the fingerprint (Ott *et al.*, 2003). Similar observations have been made with MS (Fiehn *et al.*,

2000; Sumner *et al.*, 2003) and it is clear that growing conditions need to be tightly controlled if the sensitivity and precision of the analysis is not to be swamped by factors that are extraneous to the investigation. Moreover, even with tightly controlled growth conditions, it may be necessary to pool samples to reduce biological variation (Sumner *et al.*, 2003).

Second, stringent control of sample preparation, NMR data collection and spectrum processing are all essential if the benefits of the fingerprinting approach are to be maximized (Lommen *et al.*, 1998; Defernez and Colquhoun, 2003; Ott *et al.*, 2003; Ward *et al.*, 2003). Operationally the key objectives are (i) to maximize the reproducibility of NMR data collection and (ii) to minimize misalignment of the NMR signals before embarking on the statistical analysis (Defernez and Colquhoun, 2003). Ideally, two replicate samples would generate two identical NMR fingerprints, but, in practice, there will be discrepancies in lineshape and chemical shift, both of which will hinder the classification of the spectra as identical. Changes in lineshape can be minimized by using exactly the same sample volume and by optimizing the magnetic field homogeneity before data acquisition. This should be sufficient to generate reproducible lineshapes, but if not then the lineshapes can be manipulated during processing, for example, by varying the linebroadening parameter to ensure that the processed linewidth for a particular signal is the same in every spectrum (Lommen *et al.*, 1998). Changes in chemical shift can also be problematic and the negative impact of such changes on a principal components analysis has been demonstrated (Defernez and Colquhoun, 2003). Careful sample preparation, with precise control of the pH, and precise temperature control during the acquisition of the spectrum should ensure that most signals occur at exactly the same chemical shift. However, differences in ionic strength and specific

interactions between components in the sample may lead to discrepancies in peak position and these can only be corrected by aligning regions of the spectra before the statistical analysis (Defernez and Colquhoun, 2003).

Third, several of these investigations have provided the strongest possible argument for NMR fingerprinting by generating interesting biological insights into the systems that have been analysed. For example, investigations of transgenic plant material have highlighted the potential value of NMR fingerprinting as a tool for investigating the substantial equivalence of conventional and engineered crop plants (Noteborn *et al.*, 2000; Le Gall *et al.*, 2003; Charlton *et al.*, 2004). Thus in one study of transgenic tomatoes only minimal compositional differences were found between isogenic transgenic and wild-type lines after allowing for the substantial variation caused by external factors (Noteborn *et al.*, 2000); while in a second study of a line that had been modified to increase the flavonol content, statistically significant changes in metabolites, other than those that had been deliberately increased, were found to be minor and within the variation that would be observed in a field-grown crop (Le Gall *et al.*, 2003). Interestingly, in a recent study of the differences in composition between wild-type and transgenic pea lines, the wild-type material showed greater variation, and a detailed analysis of this effect led to the conclusion that it was caused by the transformation process selecting for a subset of individuals rather than by the transgene itself (Charlton *et al.*, 2004).

Biological insights have also been forthcoming from fingerprinting analyses of abiotic stress. For example, in a study of cadmium toxicity (Bailey *et al.*, 2003), principal components analysis of the ^1H NMR spectra showed clear discrimination between control plants and plants exposed to cadmium, and the loadings plot for the first principal component was dominated by changes in a handful of signals that could be readily assigned to particular metabolites. Clearly the next step will be to use this information to test hypotheses about the metabolic response to cadmium but this was beyond the scope of the fingerprinting study. Finally, in an impressive investigation of herbicide action, NMR fingerprints of maize seedlings were used as the input for a neural network analysis (Ott *et al.*, 2003). This led to a robust method for discriminating between 19 modes of action and it was argued that it was sufficiently reliable to be a useful tool in the discovery of bioactive compounds with novel modes of action.

Metabolite profiling of plant tissues using NMR

Although multivariate analysis of ^1H NMR spectra from samples of plant origin is a relatively recent development, NMR spectroscopy has been used as an analytical tool to identify the metabolites in such samples for many years (Fan, 1996). In fact, using multivariate analysis as a tool to identify the interesting features of a spectrum prior to

embarking on an assignment exercise provides an efficient way to explore the plant metabolome, and it is likely to become a standard procedure in investigations that proceed beyond the discrimination step in a fingerprinting exercise (Le Gall *et al.*, 2003, 2004). However, profiling applications of NMR in plant tissues have usually focused immediately on the identification of particular metabolites, and so the techniques that have been developed to extract useful analytical information from the spectra of complex mixtures have been developed from a purely analytical perspective (Fan, 1996).

^1H NMR spectroscopy is the most commonly used form of NMR for metabolite identification, and in some early profiling experiments on plant tissues this technique was applied *in vivo* (Fan *et al.*, 1986a, b, 1988). This allowed the metabolic response of the tissues to anoxia to be monitored directly, and it was also possible to deduce information about the subcellular distribution of a limited number of metabolites. However, the poorer resolution of the *in vivo* spectra made the assignment problem more difficult, both by increasing overlap and by masking the characteristic splitting of spin-coupled resonances, and it was concluded that a more detailed profile of the metabolic composition of the tissue could only be obtained by working with tissue extracts. The narrower lines in the extract spectra made it easier to implement two dimensional NMR techniques, allowing assignments to be made on the basis of correlations between signals in different regions of the spectrum. This increased the number of metabolites that could be identified in the extract, and it also increased the number of assignments that could be made retrospectively *in vivo*. Thus the conclusion from this early work was that quantitative information on around 20 metabolites could be easily obtained from unfractionated extracts of plant tissues and that the key to maximizing this number was to use two-dimensional spectroscopy (Fan, 1996).

Analysing the metabolic composition of a tissue with ^1H NMR spectroscopy has several advantages: only a crude extract, and therefore minimal sample preparation, is required; a wide range of compounds can be analysed, providing definitive structural information with no restrictions relating to volatility, polarity or the presence of specific chromophores; the method is non-destructive, permitting subsequent analysis by other methods; and the method can be applied with little prior knowledge of the composition of the sample (Fan, 1996). Despite these advantages, and in contrast to the biomedical field, there have been relatively few applications of the ^1H NMR method to samples of plant origin, with the main interest being in the analysis of root exudates and fruit juices. For example, there have been two substantial investigations of unfractionated root exudates, both of which resulted in the identification of numerous organic acids, amino acids, and derivatives of mugineic acid (Fig. 2; Fan *et al.*, 1997, 2001). One- and two-dimensional NMR analyses, including both homonuclear (^1H - ^1H) and

heteronuclear (^1H - ^{13}C) correlations, were complemented by GC-MS, to identify the presence of components present at low concentration, and by high resolution MS spectrometry, for accurate molecular mass measurements. This combination of techniques showed clear differences between the exudates of different species and genotypes, and it was also possible to monitor changes in the composition of the exudates brought about by iron deficiency (Fan *et al.*, 1997, 2001).

Fruit juices have been the other main target for compositional analysis of unfractionated samples from plant material and a recent study of tomato juice provides a representative example (Sobolev *et al.*, 2003). A comprehensive analysis of the spectra, again using a full range of two-dimensional NMR techniques, led to the assignment of more than 90 signals from 19 metabolites, and the spectra were then used for a comparison between two tomato cultivars. In a similar study of transgenic tomato lines, NMR analysis of extracts prepared from freeze-dried fruits led to the identification of more than 40 compounds and this formed the basis for a quantitative analysis of the unintended metabolic consequences of the transformation (Le Gall *et al.*, 2003).

It is notable that in almost all these profiling investigations the emphasis has been on determining composition and relating it to the classification of the samples into different groups, rather than on testing hypotheses about the control and regulation of metabolic processes. Thus, in contrast to metabolite profiling by GC-MS, where one of the objectives is to use the data to further understanding of intermediary metabolism and its manipulation (Roessner-Tunali *et al.*, 2003), the data from ^1H NMR analyses have been used almost exclusively for analytical and screening purposes.

While one of the advantages of NMR as a metabolite profiling technique is that informative spectra can be recorded from crude extracts, it is also possible to use in-line separation techniques to simplify the analysis. This approach has had a major impact on the use of NMR for metabolic analysis in the biomedical field (Lindon *et al.*, 1996), and the use of LC-NMR is now well established as a technique for phytochemical analysis (Bringmann *et al.*, 1998; Exarchou *et al.*, 2003). LC-NMR eliminates the need to purify a compound before analysis, and the NMR spectra can either be recorded in continuous flow or stopped flow mode, the latter involving so-called peak parking in which fractions from the column are diverted to a storage loop while a spectrum of sufficient quality is recorded from the fraction that is already in the NMR magnet. A further elaboration of this scheme is to split the output from the HPLC column into two pathways, one destined for NMR analysis and the other for MS. This powerful combination has been used to investigate xenobiotic metabolism in plants (Bailey *et al.*, 2000a, b) and to identify natural products in plant extracts (Exarchou *et al.*, 2003). In the

latter study, sensitivity was further increased by using a solid phase extraction (SPE) step between the HPLC column and the NMR spectrometer, and a cryogenic probe to detect the NMR signals. The SPE cartridge serves to concentrate the sample and it also avoids the use of large volumes of expensive deuterated solvents for the LC separation, since it is only necessary to use a deuterated solvent, which is required for the ^1H NMR analysis, to elute the sample from the cartridge.

In principle, these so-called hyphenated NMR techniques have the potential to increase the number of metabolites that can be identified in ^1H NMR fingerprints of crude extracts. However, techniques such as LC-UV-SPE-NMR-MS (Exarchou *et al.*, 2003) are mainly used in the realm of natural products research and xenobiotic metabolism, and they have yet to be combined with the fingerprinting approach described in the preceding section. It seems likely that it will only be possible to obtain a true measure of the fraction of the plant metabolome that can be analysed by NMR when fractionation of tissue extracts becomes a routine tool for the assignment of the spectra used in fingerprinting analysis.

Concluding remarks

Although compromised to some extent by its sensitivity, ^1H NMR spectroscopy is an effective technique for both metabolite fingerprinting and metabolite profiling applications in samples of plant origin. These analyses have the potential to complement high-throughput system-wide analyses by MS, and the application of coupled techniques that allow parallel MS and NMR analyses on the same sample would seem to be an ideal way to increase the fraction of the metabolome that can be revealed by routine analysis. Curiously, the two factors that have driven the recent growth in metabolomics, the need for more comprehensive analyses to underpin our understanding of the metabolic network and the possibility of using metabolic phenotypes based on composition to uncover gene function, have yet to be greatly advanced by either the NMR or MS approaches to system-wide metabolite analysis. In this regard it is notable that the potentially powerful strategy of using metabolome data to reveal the phenotype of silent mutations exploited metabolite fingerprints based on ^1H NMR analysis (Raamsdonk *et al.*, 2001). The original demonstration of this method was in yeast and it is disappointing that it has apparently not yet been applied successfully to plants.

Acknowledgements

NJK and RGR thank the UK Biotechnology and Biological Sciences Research Council for financial support. PK thanks the Indian Department of Science and Technology for the award of a BOY-SCAST Fellowship.

References

- Allen J, Davey HM, Broadhurst D, Heald JK, Rowland JJ, Oliver SG, Kell DB. 2003. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nature Biotechnology* **21**, 692–696.
- Aubert S, Pallett KE. 2000. Combined use of ^{13}C - and ^{19}F -NMR to analyse the mode of action and the metabolism of the herbicide isoxaflutole. *Plant Physiology and Biochemistry* **38**, 517–523.
- Bacher A, Rieder C, Eichinger D, Arigoni D, Fuchs G, Eisenreich W. 1999. Elucidation of novel biosynthetic pathways and metabolite flux patterns by retrobiosynthetic NMR analysis. *FEMS Microbiology Reviews* **22**, 567–598.
- Bailey NJC, Cooper P, Hadfield ST, Lenz EM, Lindon JC, Nicholson JK, Stanley PD, Wilson ID, Wright B, Taylor SD. 2000a. Application of directly coupled HPLC-NMR-MS/MS to the identification of metabolites of 5-trifluoromethylpyridone (2-hydroxy-5-trifluoromethylpyridine) in hydroponically grown plants. *Journal of Agricultural and Food Chemistry* **48**, 42–46.
- Bailey NJC, Oven M, Holmes E, Nicholson JK, Zenk MH. 2003. Metabolomic analysis of the consequences of cadmium exposure in *Silene cucubalus* cell cultures via ^1H NMR spectroscopy and chemometrics. *Phytochemistry* **62**, 851–858.
- Bailey NJC, Sampson J, Hylands PJ, Nicholson JK, Holmes E. 2002. Multi-component metabolic classification of commercial feverfew preparations via high-field ^1H -NMR spectroscopy and chemometrics. *Planta Medica* **68**, 734–738.
- Bailey NJC, Stanley PD, Hadfield ST, Lindon JC, Nicholson JK. 2000b. Mass spectrometrically detected directly coupled high performance liquid chromatography/nuclear magnetic resonance spectroscopy/mass spectrometry for the identification of xenobiotic metabolites in maize plants. *Rapid Communications in Mass Spectrometry* **14**, 679–684.
- Belton PS, Colquhoun IJ, Kemsley EK, Delgadillo I, Roma P, Dennis MJ, Sharman M, Holmes E, Nicholson JK, Spraul M. 1998. Application of chemometrics to the ^1H NMR spectra of apple juices: discrimination between apple varieties. *Food Chemistry* **61**, 207–213.
- Bringmann G, Günther C, Schlauer J, Rückert M. 1998. HPLC-NMR on-line coupling including the ROESY technique: direct characterization of naphthylisoquinoline alkaloids in crude plant extracts. *Analytical Chemistry* **70**, 2805–2811.
- Charlton A, Allnut T, Holmes S, Chisholm J, Bean S, Ellis N, Mullineaux P, Oehlschlager S. 2004. NMR profiling of transgenic peas. *Plant Biotechnology Journal* **2**, 27–35.
- Defernez M, Colquhoun IJ. 2003. Factors affecting the robustness of metabolite fingerprinting using ^1H NMR spectra. *Phytochemistry* **62**, 1009–1017.
- El-Dereby W. 1997. Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy: a review. *NMR in Biomedicine* **10**, 99–124.
- Exarchou V, Godejohann M, van Beek TA, Gerothanassis IP, Vervoort J. 2003. LC-UV-Solid-Phase-Extraction-NMR-MS combined with a cryogenic flow probe and its application to the identification of compounds present in Greek oregano. *Analytical Chemistry* **75**, 6288–6294.
- Fan TW-M. 1996. Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. *Progress in Nuclear Magnetic Resonance Spectroscopy* **28**, 161–219.
- Fan TW-M, Higashi RM, Lane AN. 1986a. Monitoring of hypoxic metabolism in superfused plant tissues by *in vivo* ^1H NMR. *Archives of Biochemistry and Biophysics* **251**, 674–687.
- Fan TW-M, Higashi RM, Lane AN. 1988. An *in vivo* ^1H and ^{31}P NMR investigation of the effect of nitrate on hypoxic metabolism in maize roots. *Archives of Biochemistry and Biophysics* **266**, 592–606.
- Fan TW-M, Higashi RM, Lane AN, Jardetzky O. 1986b. Combined use of ^1H -NMR and GC-MS for metabolite monitoring and *in vivo* ^1H -NMR assignments. *Biochimica et Biophysica Acta* **882**, 154–167.
- Fan TW-M, Lane AN, Pedler J, Crowley D, Higashi RM. 1997. Comprehensive analysis of organic ligands in whole root exudates using nuclear magnetic resonance and gas chromatography-mass spectrometry. *Analytical Biochemistry* **251**, 57–68.
- Fan TW-M, Lane AN, Shenker M, Bartley JP, Crowley D, Higashi RM. 2001. Comprehensive chemical profiling of graminaceous plant root exudates using high-resolution NMR and MS. *Phytochemistry* **57**, 209–221.
- Fiehn O. 2002. Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology* **48**, 155–171.
- Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L. 2000. Metabolite profiling for plant functional genomics. *Nature Biotechnology* **18**, 1157–1161.
- Hinse C, Richter C, Provenzani A, Stöckigt J. 2003. *In vivo* monitoring of alkaloid metabolism in hybrid plant cell cultures by 2D cryo-NMR without labelling. *Bioorganic and Medicinal Chemistry* **11**, 3913–3919.
- Kruger NJ, Ratcliffe RG. 2004. Metabolic organization in plants: a challenge for the metabolic engineer. In: Bohnert HJ, Nguyen HT, eds. *Advances in plant biochemistry and molecular biology*, Vol. 1. *Bioengineering and molecular biology of plant pathways*. Elsevier (in press).
- Kruger NJ, Ratcliffe RG, Roscher A. 2003. Quantitative approaches for analysing fluxes through plant metabolic networks using NMR and stable isotope labelling. *Phytochemistry Reviews* **2**, 17–30.
- Le Gall G, Colquhoun IJ, Davis AL, Collins GJ, Verhoeven ME. 2003. Metabolite profiling of tomato (*Lycopersicon esculentum*) using ^1H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *Journal of Agricultural and Food Chemistry* **51**, 2447–2456.
- Le Gall G, Colquhoun IJ, Defernez M. 2004. Metabolite profiling using ^1H NMR spectroscopy for quality assessment of green tea, *Camellia sinensis* (L.). *Journal of Agricultural and Food Chemistry* **52**, 692–700.
- Lindon JC, Holmes E, Nicholson JK. 2001. Pattern recognition methods and applications in biological magnetic resonance. *Progress in Nuclear Magnetic Resonance Spectroscopy* **39**, 1–40.
- Lindon JC, Nicholson JK. 1997. Recent advances in high-resolution NMR spectroscopic methods in bioanalytical chemistry. *Trends in Analytical Chemistry* **16**, 190–200.
- Lindon JC, Nicholson JK, Wilson ID. 1996. Direct coupling of chromatographic separations to NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy* **29**, 1–49.
- Lommen A, Weseman JM, Smith GO, Noteborn HPJM. 1998. On the detection of environmental effects on complex matrices combining off-line liquid chromatography and ^1H NMR. *Bio-degradation* **9**, 513–525.
- Nicholson JK, Wilson ID. 2003. Understanding ‘global’ systems biology: metabonomics and the continuum of metabolism. *Nature Reviews Drug Discovery* **2**, 668–676.
- Noteborn HPJM, Lommen A, van der Jagt RC, Weseman JM. 2000. Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *Journal of Biotechnology* **77**, 103–114.
- Ott K-H, Aranibar N, Singh B, Stockton GW. 2003. Metabonomics classifies pathways affected by bioactive compounds. Artificial neural network classification of NMR spectra of plant extracts. *Phytochemistry* **62**, 971–985.
- Raamsdonk LM, Teusink B, Broadhurst D, et al. 2001. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology* **19**, 45–50.

- Ratcliffe RG.** 1994. *In vivo* NMR studies of higher plants and algae. *Advances in Botanical Research* **20**, 43–123.
- Ratcliffe RG, Roscher A, Shachar-Hill Y.** 2001. Plant NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy* **39**, 267–300.
- Ratcliffe RG, Shachar-Hill Y.** 2001. Probing plant metabolism with NMR. *Annual Review of Plant Physiology and Plant Molecular Biology* **52**, 499–526.
- Ratcliffe RG, Shachar-Hill Y.** 2005. Revealing metabolic phenotypes in plants: inputs from NMR analysis. *Biological Reviews* **80**, (in press).
- Reed JL, Palsson BØ.** 2003. Thirteen years of building constraints-based *in silico* models of *Escherichia coli*. *Journal of Bacteriology* **185**, 2692–2699.
- Roberts JKM.** 2000. NMR adventures in the metabolic labyrinth within plants. *Trends in Plant Science* **5**, 30–34.
- Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie AR.** 2001. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *The Plant Cell* **13**, 11–29.
- Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L.** 2000. Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. *The Plant Journal* **23**, 131–142.
- Roessner U, Willmitzer L, Fernie AR.** 2002. Metabolic profiling and biochemical phenotyping of plant systems. *Plant Cell Reports* **21**, 189–196.
- Roessner-Tunali U, Hegemann B, Lytovchenko A, Carrari F, Bruedigam C, Granot D, Fernie AR.** 2003. Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development. *Plant Physiology* **133**, 84–99.
- Sobolev AP, Segre A, Lamanna R.** 2003. Proton high-field NMR study of tomato juice. *Magnetic Resonance in Chemistry* **41**, 237–245.
- Sumner LW, Mendes P, Dixon RA.** 2003. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* **62**, 817–836.
- Sweetlove LJ, Last RL, Fernie AR.** 2003. Predictive metabolic engineering: a goal for systems biology. *Plant Physiology* **132**, 420–425.
- Trethewey RN.** 2004. Metabolite profiling as an aid to metabolic engineering in plants. *Current Opinion in Plant Biology* **7**, 196–201.
- Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner-Tunali U, Willmitzer L, Fernie AR.** 2003. Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Reports* **4**, 989–993.
- Ward JL, Harris C, Lewis J, Beale MH.** 2003. Assessment of ^1H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. *Phytochemistry* **62**, 949–957.
- Weckwerth W.** 2003. Metabolomics in systems biology. *Annual Review of Plant Biology* **54**, 669–689.