Metabolomics: Current analytical platforms and methodologies

Warwick B. Dunn, David. I. Ellis

During the previous decade, a new array of analytical methodologies and technologies were introduced related to the analysis of microbial, plant and animal metabolomes (complete collections of all low molecular weight compounds in a cell). The scientific field of metabolomics was born. In this review, we discuss advances in methodologies and technologies, and outline applications.

Keywords: Infrared spectroscopy; Mass spectrometry; Nuclear magnetic resonance spectroscopy; Metabolic fingerprinting; Metabolic profiling; Metabolite; Metabolome; Metabolomics; Metabonomics

Warwick B. Dunn*, David. I. Ellis
Bioanalytical Sciences Group, School of Chemistry, University of Manchester, Faraday Building, Sackville Street, P.O. Box 88, Manchester M60 1QD, UK

*Corresponding author.
Tel.: +44 161 200 4414;
Fax: +44 161 200 4556;
E-mail: Warwick.Dunn@manchester.ac.uk.

1. Introduction

The metabolomics experiment (sampling, sample preparation, instrumental analysis, data processing and data interpretation) provides unique challenges to fulfil the goal of improving the current status of biological information related to the metabolome and more generally functional genomics [1–5]. The metabolome has been defined as the qualitative and quantitative collection of all low molecular weight molecules (metabolites) present in a cell that are participants in general metabolic reactions and that are required for the maintenance, growth and normal function of a cell ([1], chapter 17, [6]). Generally, these include organic species (e.g., amino and fatty acids, carbohydrates, vitamins and lipids [see KEGG genome encyclopaedia for further examples http://www.genome.ad.jp/kegg/catalog/compounds.html), although inorganic and elemental species [7] can also be studied. The size of the metabolome varies greatly, depending on the organism studied; *Saccharomyces cerevisiae* contains approximately 600 metabolites [8], the plant kingdom has an estimated 200,000 primary and secondary metabolites [2] and the human metabolome can be expected to be even larger in size.

Metabolites constitute a diverse set of atomic arrangements when compared to the proteome (arrangement of 20 amino acids) and transcriptome (arrangement of four nucleotide bases bonded with sugar and phosphate backbone) and this provides wide variations in chemical (molecular weight, polarity, solubility) and physical (volatility) properties. The degree of diversity is indicated by the analysis of low molecular weight (MW), polar, volatile organic metabolites, such as ethanol or isoprene to the higher MW, polar (carbohydrates) and non-polar (terpenoids and lipids) metabolites. The metabolome also extends over an estimated 7–9 magnitudes of concentration (pmol–mmol). In genomics, pleiotropic effects cause seemingly unrelated biochemical pathways to be affected (e.g., by genetic alterations). Therefore, to be able to analyse all metabolites in a single analysis would be preferred (metabolomics). Technologically, this is not currently possible. Table 1 and Fig. 1 indicate the different strategies employed. Table 2 provides an overview of the advantages of studying the metabolome in comparison to the transcriptome or proteome.

Applications have developed from initial work in the 1980s employing soft ionisation mass spectrometry (SIMS) [9], gas chromatography-mass spectrometry (GC-MS) [10] and nuclear magnetic resonance (NMR) spectroscopy [11]. Genome sequencing has shown the relatively small volume of information we have attained relating to gene function (or genotype). For example, *S. cerevisiae* contains over 6000 genes of which only approximately 33% had been studied previously and whose function is known [12].

Metabolomics applications have expanded in-line with the complementary

0165-9936/$ - see front matter doi:10.1016/j.trac.2004.11.021

Table 1. Strategies for metabolomic analysis

Metabolomics	Non-biased identification and quantification of all metabolites in a biological system. Sample preparation must not exclude metabolites, and selectivity and sensitivity of the analytical technique must be high
Metabolite profiling	Identification and quantification of a selected number of pre-defined metabolites, generally related to a specific metabolic pathway(s). Sample preparation and instrumentation are employed so to isolate those compounds of interest from possible matrix effects prior to detection, normally with chromatographic separation prior to detection with MS. In the pharmaceutical industry, this is widely used to study drug candidates, drug metabolic products and the effects of therapeutic treatments (termed metabolite profiling [28])
Metabolic fingerprinting	High-throughput, rapid, global analysis of samples to provide sample classification. Quantification and metabolic identification are generally not employed. A screening tool to discriminate between samples of different biological status or origin. Sample preparation is simple and, as chromatographic separation is absent, rapid analysis times are small (normally 1 min or less)
Metabolite target analysis	Qualitative and quantitative analysis of one or a few metabolites related to a specific metabolic reaction. Extensive sample preparation and separation from other metabolites is required and this approach is especially employed when low limits of detection are required. Generally, chromatographic separation is used followed by sensitive MS or UV detection
Metabonomics	Evaluation of tissues and biological fluids for changes in endogenous metabolite levels that result from disease or therapeutic treatments

genomic techniques of proteomics and transcriptomics with the objective of gene-function determination in microbes [13], plants [14] and animals. Today, many different applications are also performed. These include:

- determination of metabolic biomarkers that change as an indicator of the presence of a disease or in response to drug-based intervention (for applications, see the sections below on Fourier transform infrared (FT-IR) spectroscopy, NMR and liquid chromatography-mass spectrometry (LC-MS); and,
- determination of the effect of biochemical or environmental stresses on plants or microbes, which include genetically modified (GM) plants [15],
- bacterial characterisations [16],
- human health assessments [17] and
- metabolic engineering [18].

Certainly analytical method and instrument validation is important to assess precision, accuracy and applicability of methods and this has been discussed previously ([1], chapter 11). Reviews on the challenges of data processing and data interpretation can be found elsewhere [1,3–5].

2. Sample preparation

In the metabolomics experiment, sampling provides a picture or snapshot of the metabolome at one point in time, although *in vivo* NMR measurements that have been reported recently do not require sampling before analysis [19]. The requirement of sampling and sample preparation that is not biased towards groups of metabolites provides challenges, which currently have not been fully resolved. The time and method of sampling can greatly influence the reproducibility of the analytical sample. Diurnal and dietary influences can have major effects on the composition of the metabolome [20,21], as can the section of a plant sampled [22]. Finally, the storage of samples is important, as the continued freeze/thawing of samples can be detrimental to stability and composition [23]. All these influence the precision, accuracy and reproducibility of results. However, generally, it is observed that biological variability is greater than analytical variability, even when controlled sampling and sample preparation are employed [24].

Strategies of sampling and sample preparation vary. Both invasive (blood, intra-cellular metabolites in plants and microbes) and non-invasive (urine, volatile components, metabolic footprint) sampling can be performed. Extra-cellular metabolites, such as metabolic footprint or urine, depict a picture over a period of metabolic activity and are normally stored at low temperatures to inhibit metabolic reactions. The extraction of intra-cellular metabolites provides a snapshot of the metabolome, can be time consuming, and is subject to certain difficulties when compared to other sampling strategies. Metabolic processes are rapid (reaction times less than 1 s), so rapid inhibition of enzymatic processes is required, generally by freeze clamping or freezing in liquid nitrogen after harvesting, and subsequent storage at −80°C. Freezing provides specific issues, such as loss of metabolites [25], the release of touch- or wound-induced metabolites or the non-reversible loss of compounds by absorption to cell walls. The application of acidic treatments using perchloric or nitric acid has been used but can result in a severe reduction in the number of metabolites detected and degradation compounds not stable at extreme pH. Polar/non-polar extractions are the most frequently applied method and are performed by physical/chemical disruption of the cells, removal of the cell pellet by centrifugation and distribution of metabolites to polar (methanol/water) and non-polar (chloroform) solvents. Hot alcoholic extractions are also performed. Metabolic footprinting has recently been used in microbial

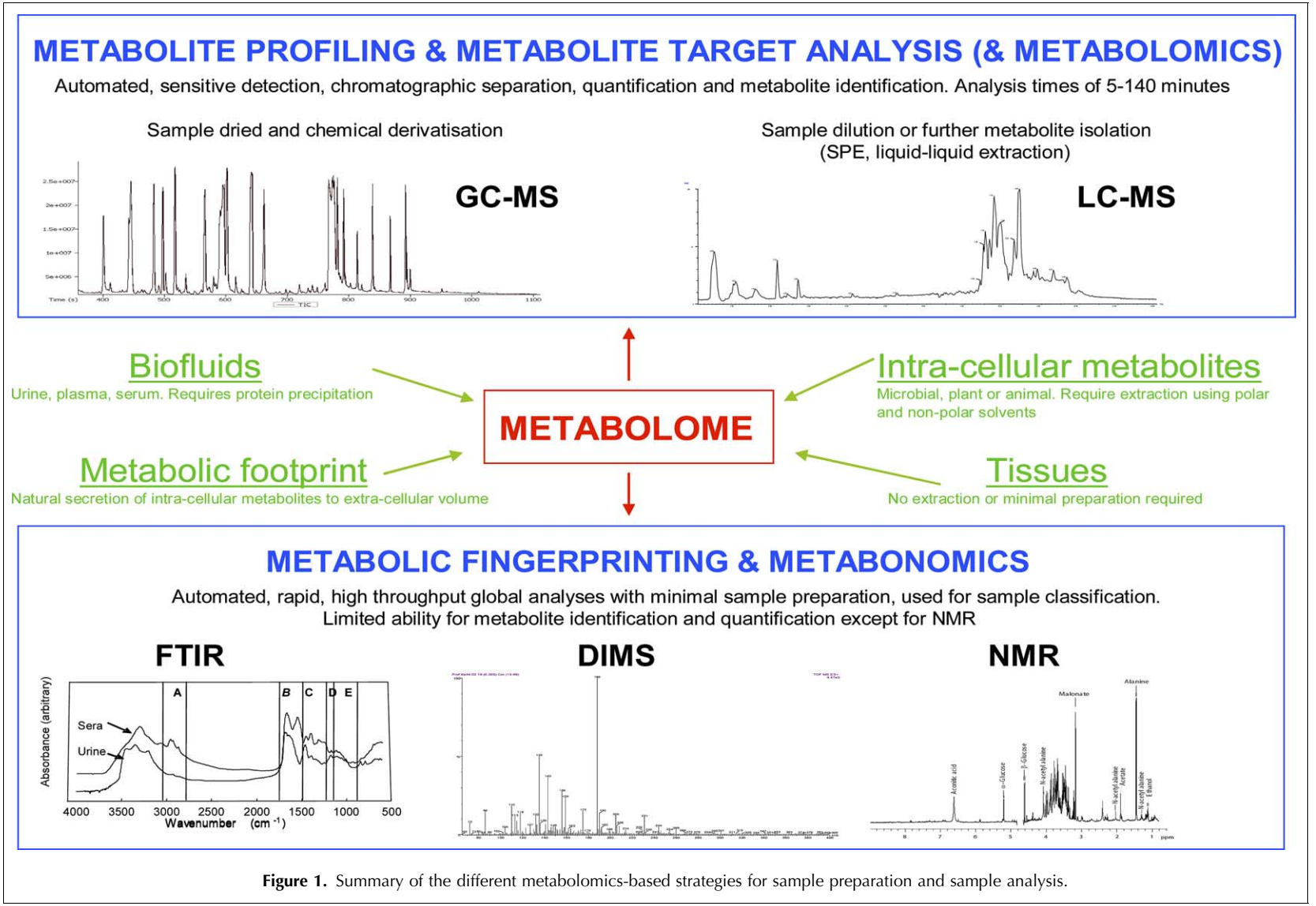

Figure 1. Summary of the different metabolomics-based strategies for sample preparation and sample analysis.

Table 2. Advantages of analysing the metabolome

- The number of metabolites should be lower than the number of genes and proteins in a cell. For example, *Saccharomyces cerevisiae* contains over 6000 genes but only 600 metabolites. Sample complexity is reduced
- The theory of Metabolic Control Analysis (MCA) [93] and experimental work have both shown that, although the concentration of enzyme and metabolic flux may not significantly change during a biochemical reaction, the concentration of metabolites can significantly change
- The metabolome is the downstream product of gene expression so reflects the functional level of the cell more appropriately and changes in the metabolome are expected to be amplified relative to proteome or transcriptome [94]
- Research is showing that metabolic fluxes are regulated not only by gene expression but also by environmental stresses (phenotype) and hence measurement of the downstream products (metabolites) is appropriate [77]
- It is estimated that metabolomics experiments cost two to three times less than proteomic and transcriptomic experiments (see [1], chapter 11)

metabolomics to exclude the extraction procedure and provide rapid, high-throughput sampling [26]. Here, metabolites naturally excreted from intra-cellular volumes to the extra-cellular supernatant are analysed. Sampling and collection of volatile compounds from plants has been discussed elsewhere [27], and a procedure for extraction and separation of metabolome, proteome and transcriptome has also been reported [28].

Further sample preparation depends on the sample and metabolomics strategy employed. In many applications, no further isolation of metabolites from the sample matrix is performed, and samples are diluted and analysed directly or analysed after chemical derivatisation. However, some isolation of metabolites can be undertaken, especially with biofluids, where an initial preparation stage is protein precipitation with organic solvents. Further isolation from the sample matrix can be used including solid phase extraction (SPE) or liquid–liquid extraction (LLE). These can especially be observed in the analysis of pharmaceuticals and related metabolites [28] or for metabolite target analysis.

3. Mass spectrometry

MS (including combined chromatography-MS) is the most widely applied technology in metabolomics, as it provides a blend of rapid, sensitive and selective qualitative and quantitative analyses with the ability to identify metabolites. Mass spectrometers operate by ion formation, separation of ions according to their mass-to-charge (m/z) ratio and detection of separated ions.

3.1. GC-MS

GC-MS is a combined system where volatile and thermally stable compounds are first separated by GC and then eluting compounds are detected traditionally by electron-impact mass spectrometers. In metabolomics, GC-MS has been described as the gold standard ([1], chapter 11), although it is biased against non-volatile, high-MW metabolites. Volatile, low-MW metabolites can be sampled and subsequently analysed directly, including breath [29] and plant volatiles [30]. However, the majority of metabolites analysed require chemical derivatisation at room or elevated temperatures to provide volatility and thermal stability prior to analysis. Due to the range of chemical functionality of metabolites, two-stage derivatisation is employed [24]. First, carbonyl functional groups are converted to oximes with *O*-alkylhydroxylamine solutions, followed by formation of trimethylsilyl (TMS) esters with silylating reagents (typically *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide), to replace exchangeable protons with TMS groups. Oxime formation is required to eliminate undesirable slow and reversible silylation reactions with carbonyl groups, whose products can be thermally labile. Some metabolites contain a number of exchangeable protons and hence a range of derivatisation products are formed (e.g., amino acids and carbohydrates will form multiple derivatisation products, whereas organic acids often react to create only one detected product). Sample stability is a concern. The presence of water can result in the breakdown of TMS esters (esterification being a reversible reaction), although extensive sample drying and the presence of excess silylating reagent can limit this process. However, extensive sample drying can result in the loss of volatile metabolites (e.g., lactic acid). An automated system employing a 'derivatisation-when-required' approach is desirable to ensure maximum sample stability. However, if on-line automated derivatisation is not available, ensuring derivatised samples are not stored at room temperatures for long periods is essential. Sample stability depends on sample and environmental conditions.

During analysis, small aliquots of derivatised samples are analysed (typical injection volumes of 1 μl or less) by split and splitless techniques on GC columns of differing polarity (DB-5 or DB-50 or similar stationary-phase capillary columns are most commonly used). These provide both high chromatographic resolution of compounds and high sensitivity (typical limits of detection being pmol or nmol concentrations). Chromatograms are complex, containing hundreds of metabolite peaks and are further complicated by multiple derivatisation products. There needs to be either long run times (greater than 60 min) [24] or a combination of fast acquisition rate time-of-flight instruments coupled with deconvolution software (AMDIS [http://chemdata.nist.

gov/mass-spc/amdis/] or instrument-specific software [e.g., LECO ChromaTOF]). Deconvolution software [31] acts by using pure mass spectra to define chromatographic peaks (including overlapping peaks) and allows reductions in run time (from 60 min to less than 15 min) as full chromatographic separation is not required, and, in many metabolomics analyses, is difficult or impossible to achieve. Deconvolution can detect co-eluting peaks with peak apexes separated by less than 1 s, and also detect low-abundance peaks co-eluting in the presence of metabolites at much higher concentrations. Chromatographic alignment of raw data can be performed before deconvolution when comparing large data sets created on different days (e.g., MSFACTS [32]), although data analysis without deconvolution has also been reported [33].

Quantification is provided by either external calibration or response ratio (peak area of metabolite/peak area of internal standard). External calibration is labour intensive and not all metabolites are commercially available to enable preparation of standard solutions. Metabolome coverage is largely determined by the volatility of sample components, either non-derivatised or derivatised. Metabolite identification is provided by matching the retention time or retention index and mass spectrum of the sample peak with those of a pure compound previously analysed on the same or different instrument under identical instrumental conditions [34]. However, many metabolites are not available commercially, so creating user-defined mass spectral libraries is limited. Mass spectral searches against commercially available databases (e.g., NIST/EPA/NIH, www.nist.gov/srd/nist1.htm) can be successful. Unfortunately, commercially available mass spectral databases do not contain all metabolites that would be expected from studying metabolic reaction networks. Within the field, efforts are being made to create metabolomics-specific mass spectral libraries. As a final strategy, and because the electron impact mass spectrometer provides standard fragmentation of molecular ions during ionisation, structural identification can be performed *via* interpretation of fragment ions and fragmentation patterns.

Applications of GC-MS are wide ranging in metabolomics, especially as it is a mature analytical technique. The first applications of GC-MS in metabolomics can be traced to urinary screening procedures to indicate the presence of diseases related to organic acidemias [10]. More recently, applications in plant metabolomics are becoming widespread mainly through the initial research of workers in Germany [14,24]. Plants including Arabidopsis [35], potatoes [24] and tomatoes [36] have all been studied to measure effects of genetic or environmental modifications and stressors, either by analysis of intra-cellular metabolites or volatile metabolites, including scent [37]. Other applications, including microbial and clinical metabolomics, are now employing GC-MS to analyse samples such as biofluids or breath [38]. Certainly the application of multi-dimensional separations (e.g., GC × GC-TOF-MS) in the future will be advantageous because of greater specificity and sensitivity [39].

3.2. LC-MS

Another combined system, LC-MS, provides metabolite separation by LC followed by electrospray ionisation (ESI) or, less typically, atmospheric pressure chemical ionisation (APCI) [40]. This technique differs from GC-MS in distinct ways (lower analysis temperatures, and sample volatility not required) and this simplifies sample preparation. In the majority of non-pharmaceutical applications (e.g., microbial, plant, mammalian biomarker discovery), samples are prepared after intracellular extraction and/or protein precipitation by dilution in an appropriate solvent. Within pharmaceutical applications, further sample preparation can be employed, including SPE or LLE [28,40]. Sample derivatisation is generally not required, although it can be beneficial to improve chromatographic resolution and sensitivity [41] or to provide ionisable groups on metabolites otherwise undetectable by ESI-MS. Electrospray instrumentation operates in positive and negative ion modes (either as separate experiments or by polarity switching during analyses) and only detects those metabolites that can be ionised by addition or removal of a proton or by addition of another ionic species. Metabolites are generally detected in one but not both ion modes, so wider metabolome coverage can be obtained by analysis in both modes. Quantification is performed by external calibration or response ratio in pharmaceutical applications and peak areas are currently employed in animal, disease, plant and microbial work. Metabolite identification is more time intensive. ESI does not result in fragmentation of molecular ions as observed in electron impact mass spectrometers, so it does not allow direct metabolite identification by comparison of ESI mass spectra, as ESI mass spectral libraries are not commonly available (as is the case for GC-MS). However, with the use of accurate mass measurements and/or tandem MS (MS/MS) to provide collisional induced dissociation (CID) and related mass spectra (MS/MS), metabolite identification can be performed [42]. Recent research into the reproducibility of mass spectral libraries from CID have been undertaken [43], and show that mass spectra produced on different instrument types and different manufacturers' instruments are similar. Further studies with a wider range of electrospray sources and manufacturers' instrumentation are required, as variations in mass spectra collected after in-source fragmentation can be observed with different electrospray sources.

HPLC column chemistry and dimensions used will affect the chromatographic resolution and sensitivity.

Analytical columns (i.d. = 4.6–2.0 mm) do not provide the chromatographic resolving power to separate these complex samples and 10-minute run-times are used followed by chemometric tools to mine the non-chromatographically resolved data and classify differences between samples [42]. Currently, the availability of deconvolution software (that is commonplace in GC-MS) is limited. Due to ionisation suppression (otherwise termed matrix effects) [44], the inability to provide full quantification of metabolites eluting in the presence of other metabolites is not possible. The presence of multiple metabolites (or matrix components, such as ionic salts) influences the ionisation and the transfer of metabolites from liquid to gas phase. Microbial, plant and some clinical metabolomics applications require only classification of biologically different samples with small run-times to allow high-throughput analyses without full quantification. Current research looking into the application of capillary columns promises better resolution, sensitivity and full quantification possibilities, although currently at the expense of time [45]. Alternatively, the application of ultra high-pressure chromatographic systems (e.g., the Waters Acquity UPLC™ system) can enhance chromatographic resolution. Generally, hundreds of peaks are detected, whichever strategy is employed. The most common column chemistries used today are reversed-phase C_{18} or C_8 columns. However, for polar metabolites injected onto these columns, the retention of these metabolites is minimal and they generally elute in or near the solvent front, thus reducing the volume of interpretative data. To overcome this, other column chemistries are required, including HILIC [46], and other weak ion-exchange chemistries (e.g., the Waters Atlantis metabonomics column). ESI is not amenable to high concentrations of salt, acid or base, so LC mobile phase compositions are limited.

Applications of LC-MS metabolomics are mainly focussed on clinical applications [40], along with NMR as a complementary tool. The technique has been employed in the discovery of biomarkers for a number of diseases or chronic disruptions in metabolism in both rat and human studies, including cyclosporine-A [42]. All these applications are screening techniques with some metabolite identification being confirmed with later targeted analysis or by exact mass and/or MS/MS. In fact, LC-NMR-MS techniques are currently being explored to aid in metabolite identification (NMR) and sensitive analyses (MS) [47]. Plant systems have been studied by LC-MS systems [45,48]. Other detection systems can be employed with LC, although much less frequently than MS (including, e.g., UV photodiode array detectors [49]).

3.3. Direct-injection mass spectrometry (DIMS)

DIMS advertises itself as a high-throughput screening tool (hundreds of samples per day with a 1-min analysis time applied generally). Here, crude sample extracts are injected or infused in to an electrospray mass spectrometer resulting in one mass spectrum per sample, which is representative of the composition of the sample. Metabolome coverage, as for LC-MS, depends on the ability of the metabolite to be ionised. The mass spectrum or mass list (*m/z vs.* response) is used for sample classification. Specific studies have shown the presence of ionisation suppression and its effect on sensitivity (micromolar limits of detection), which means this is a screening and not a quantitative tool. Metabolite identification can be tentatively performed using high-resolution instruments and accurate mass determination. Time-of-flight (TOF) instruments can provide mass resolutions greater than 4000 at mass 200, which allows the resolution and the detection of metabolites of the same nominal mass but different monoisotopic mass [e.g., the amino acids glutamine (MW 146.0689Da) and lysine (MW 146.1052Da)]. Combined with 5-ppm mass accuracy, this can tentatively determine the molecular formula. However, the absence of chromatographic separation can allow overlap of peaks with differences in monoisotopic masses (e.g., less than 0.025 Da), and hence a reduction in the mass accuracy of accurate mass determinations. The application of Fourier transform ion cyclotron resonance (FTICR)-MS overcomes this problem ([1], chapter 8, [50]). FTICR provides both high resolving power (100,000+) and high mass accuracy (<1 ppm external calibration, <0.1 ppm internal calibration). All possible metabolite peaks can be resolved and their molecular formula calculated very accurately. FTICR has a large future in the field of metabolomics, as it allows both high-throughput analyses and lower detection limits than those of other MS techniques, although generally with smaller dynamic ranges. Structural isomers have the same monoisotopic mass, and therefore cannot be detected separately so chromatographic separation would be required.

Applications of DIMS are mainly concentrated in the microbial [51] and plant arenas. Applications in yeast functional genomics have shown that discrimination of wild type and mutant knockout strains can be performed in metabolic footprinting of *S. cerevisiae* mutants [52]. Other applications include metabolic fingerprinting of microbial secondary metabolites [53]. Generally, quadrupole or TOF instruments are employed, although FTICR instruments have also been used [50]. In clinical applications, DIMS with MS/MS has been used in disease diagnosis [54].

3.4. Other MS-based techniques

To a lesser extent, other MS-based techniques have been employed in analysing the metabolome. Capillary electrophoresis–mass spectrometry (CE–MS) has significant potential [55]. In the last few years, both high-resolution chromatographic separation and sensitive detection

methods have shown the ability to detect up to 1600 metabolites in positive and negative ion modes.

Another area of growing interest is the application of matrix-assisted laser desorption ionisation (MALDI), laser desorption ionisation (LDI) or direct ionisation on silicon (DIOS) to provide ionisation of metabolite solutions spotted directly on a target plate, so allowing minimal sample preparation and high-throughput analysis.

4. FT-IR

FT-IR spectroscopy is an established, constantly developing analytical technique that enables the rapid, non-destructive, reagentless and high-throughput analysis of a diverse range of sample types. The principle of FT-IR lies in the fact that, when a sample is interrogated with light (or electromagnetic (EM) radiation), chemical bonds at specific wavelengths absorb this light and vibrate in one of a number of ways, such as stretching or bending vibrations. These absorptions/vibrations can then be correlated to single bonds or functional groups of a molecule for the identification of unknown compounds. Based on (bio)chemical information content and spectral richness (i.e., numbers of clearly defined peaks), the major region of interest is in the mid-IR, which is usually defined as 4000–600 cm^{-1}. In relation to biological applications, the mid-IR can be further broken down into what are termed spectral windows of interest, where strong absorption bands are able to be directly related to specific compounds. For example, stretching vibrations from CH_3 and CH_2 groups associated with fatty acids can be found within the spectral region 3050–2800 cm^{-1}. Due to its holistic nature, FT-IR spectroscopy is a valuable metabolic fingerprinting/footprinting tool owing to its ability to analyse carbohydrates, amino acids, lipids and fatty acids as well as proteins and polysaccharides simultaneously ([1], chapter 7).

FT-IR is also a highly versatile technique within the laboratory as minimum sample preparation is required and little in terms of background training is needed. Samples can be simply spotted (typically 0.5–20 μl) onto a variety of plates (i.e. Si or ZnSe) for high-throughput analyses (spectra collected within seconds). However, FT-IR does have some drawbacks when compared to the many other methods discussed. IR absorption of water, for example, is very intense but this can be overcome with dehydration of samples (which is a prerequisite for samples prepared on the plates mentioned above), the water signal can be subtracted electronically, or attenuated total reflectance (ATR) can be used. With ATR, liquid samples such as sera or plasma can be spotted directly onto the surface of a trapezoidal crystal (i.e. ZnSe or Ge) with a high-refractive index and spectra collected with little or no sample preparation. Portable FT-IR instruments are also available, allowing the analyst much scope in terms of spectral collection in a variety of environments. Whilst sensitivity and selectivity are not as high as the other methods mentioned in this article, the rapidity of spectral collection and the fact that FT-IR readily lends itself to high-throughput analyses cannot be overstressed, as 1000s of spectra can be collected within 24 h.

To date, the vast majority of metabolomic studies undertaken using vibrational spectroscopy have been with FT-IR spectroscopy. However, some work has been carried out using Raman and, in terms of metabolomics, this is an emerging technology with significant potential for monitoring metabolites [56].

The metabolic fingerprinting potential of near IR (NIR) spectroscopy should also be recognised, and studies undertaken using NIR include measurement of lactate in human blood [57] as well as the investigation of metabolites in faeces [58].

This being said, within vibrational spectroscopy, FT-IR has been applied to metabolomic questions of interest far more than either Raman (which measures predominantly non-polar bonds, as opposed to the polar bonds that FT-IR measures) or NIR. In the case of NIR, it is predominantly overtones and combination vibrations that are measured, whilst FT-IR spectra, which as already mentioned are collected in the mid-IR, are much more information-rich in terms of chemical content, as it is the fundamental vibrations that are being measured.

Applications are diverse in the fields of microbiology, plant and clinical sciences. In microbiology, rapid, accurate and reproducible identification of bacteria has been demonstrated to the sub-species level [59]. Other work applying FT-IR to microbial questions has included differentiation of clinically relevant bacteria [60], rapid detection of food spoilage [61] and metabolic footprinting of tryptophan-metabolism mutants [62].

However, within the last few years, others within the biosciences have realised that FT-IR has much potential as a metabolic fingerprinting technique for the rapid detection and diagnosis of disease or dysfunction. A number of studies have been undertaken using FT-IR to analyse tissues, cells and biofluids, with the aim being to develop a rapid methodology for the detection and diagnosis of disease or dysfunction (termed infrared pathology [63]). Attempts at cancer diagnostics have included: analysis of cell maturation in cervical tissue [64]; IR microscopic studies of colorectal adenocarcinoma sections [65]; and, detailed analyses of spectral variations within the cell cycle [66].

FT-IR has also been used to acquire metabolic fingerprints from a variety of biofluids, and the majority of metabolomic strategies for biomarker discovery are devoted to biofluid analysis [1]. These have included studies of follicular fluids to provide a biomarker for

oocyte quality [67] and synovial fluid to aid in the diagnosis of arthritic disorders [68]. In terms of analysis of serum, IR spectra could be used to discriminate between diabetes type 1, diabetes type 2 and healthy donors [69], or to identify scrapie from serum, which could lead to a rapid diagnostic test for scrapie and related disorders, such as bovine spongiform encephalopathy (BSE) and variant Creutzfeldt–Jakob disease (vCJD) [70].

Metabolic profiling of athletes in terms of doping and overtraining has also been investigated using a variety of body fluids, demonstrating that FT-IR could be applied to routine clinical analyses [71]. Other studies of interest include the multi-component assay of metabolites in urine [72] and investigation into metabolic disorders associated with urine [73]. For in-depth reviews into the use of biomedical spectroscopy, the reader is directed to [74,75].

Very recently, FT-IR has also been introduced as a metabolic fingerprinting technique within the plant sciences. These studies have included investigations into plant–plant interference, whereby global metabolite changes associated with abiotic/biotic perturbation and/or interactions were demonstrated [76]. Other applications within this field have also included metabolic fingerprinting of salt-stressed tomatoes, where functional groups of importance related to tomato salinity were identified [77], and studies of the dynamics of silage fermentations to identify the underlying biochemical differences and responses to various silage inoculants [78].

5. NMR spectroscopy

In metabolomics (also termed metabonomics for NMR-based clinical applications), NMR spectroscopy provides a rapid, non-destructive, high-throughput method that requires minimal sample preparation [79,80]. NMR spectroscopy functions by the application of strong magnetic fields and radio frequency (RF) pulses to the nuclei of atoms. For atoms with either an odd atomic number (e.g., ^{1}H) or odd mass number (e.g., ^{13}C), the presence of a magnetic field will cause the nucleus to possess spin, termed nuclear spin. Absorption of RF energy will then allow the nuclei to be promoted from low-energy to high-energy spin states, and the subsequent emission of radiation during the relaxation process is detected.

The majority of applications employ ^{1}H (proton) NMR for clinical studies and, as the majority of known metabolites contain hydrogen atoms, the system is non-biased to particular metabolites, unlike other techniques discussed in this article (i.e. all metabolites at concentrations greater than instrument limit of detection are detected). Initially, sensitivity depends on the natural abundance of the atom studied (^{1}H, ^{31}P, ^{19}F 100%; ^{13}C 1.10%; ^{15}N 0.37%) though improvements in sensitivity can be obtained by longer analysis times, application of higher magnetic fields and the use of cryogenic probes [81]. MS does provide significant improvements in sensitivity in comparison to NMR [42].

The NMR spectrum (specifically the chemical shift) depends on the effect of shielding by electrons orbiting the nucleus. The chemical shift for ^{1}H NMR is determined as the difference (in ppm) between the resonance frequency of the observed proton and that of a reference proton present in a reference compound (for ^{1}H NMR experiments, tetramethylsilane in solution, set at 0 ppm). The measured chemical shifts vary: 0–10 ppm for ^{1}H; and, 0–250 ppm for ^{13}C. The signal intensity depends on the number of identical nuclei, and the presence of complex samples does not interfere with measured intensity as ionisation suppression does with electrospray ionisation. This allows quantification to be performed.

NMR spectroscopy is a high-throughput fingerprinting technique. Crude samples are mixed with a reference-compound solution (e.g., tetramethylsilane dissolved in D_2O for ^{1}H NMR), added to an NMR probe (generally less than 2 ml), inserted into the instrument and analysed. Normally, the application of 96-well plates and flow-injection probes allows the analysis of hundreds of samples per day [79]. NMR probes are generally high µl volume based and that adds constraints on sample volume required. However, the introduction of 1 mm µl probes has enabled 2 µl volumes to be analysed, hence allowing smaller volume invasive sampling of study subjects, which is important for small-animal-based studies [82].

The spectra are complex, containing thousands of signals relating to metabolites. For data processing, the spectrum is generally split into buckets of chemical shifts with widths of 0.02–0.04 ppm. All signals in this bucket are summed. The chemical shifts can be assigned to specific metabolites and pure metabolite can be added for further clarification. However, the spectrum pattern is generally used in classification of samples, similar to that employed for FT-IR and DIMS. The robustness of NMR technology in metabolomics has been discussed previously [83].

The technique is used extensively in clinical and pharmaceutical applications for the analysis of biofluids or tissues, especially with ^{1}H NMR. Here, studies are based on cells responding to stress, including disease or therapeutic interventions by adjustment of their intra- and extra-cellular environments to ensure homeostasis (constant intra-cellular environment). Normally, this involves toxin removal from the body *via* metabolite changes in biofluids, such as urine and blood. A range of pulse sequences can enhance spectral information and minimise effects from large excesses of components [79,80]. For example, biofluids contain large excesses

of water that cause peak broadening and spectra containing minimal metabolic information. However, the application of water-suppression pulse sequences can eliminate this effect. Also, the application of spin-echo (small MW components) and diffusion-edited (macromolecular) sequences can enhance detection of different component types. Finally, pulse sequences can be used to separate 1-D spectroscopy into further dimensions [e.g., correlation spectroscopy (COSY) or diffusion ordered spectroscopy (DOSY)]. Applications include the measurement of toxic insults in rat and mouse urine, including paracetomol [84], and determination of biomarkers in coronary heart disease of humans [85]. The investigation of UV-absorbing vitamin C in the eye by metabolomics mapping using ^{19}F NMR after addition of 6-deoxy-6-fluoro-ascorbic acid has also been reported [86].

The study of intact cells or tissues is becoming more popular with the application of magic angle spinning (MAS), where the sample is rapidly spun (5 kHz) at an angle of 54.7° to overcome the difficulties of sample heterogeneity and short relaxation times that produces resonance broadening and reduces spectral information. This has been applied to the analysis of whole cells of the microalga *Thalassiosira pseudonana* [87].

NMR spectroscopy has been employed in other fields for the analysis of plant-cell extracts, such as Arabidopsis [88] and tobacco [89], to analyse cold stresses on worms [90], to determine disease biomarkers of environmentally stressed red abalone [91] and to determine the mode of action of biochemicals [92].

A ground-breaking use of NMR was to analyse *S. cerevisiae* mutants, where it was discovered that, although enzyme flux was small, the changes in metabolite concentrations were large and that the use of Functional Analysis by Co-responses in Yeast (FANCY) could be applied to correlate the unknown function genes with the funtion of known genes [13].

Acknowledgements

The authors kindly thank Hilary Major for LC-MS analyses, Nigel Bailey for NMR spectra and Marie Brown for proof reading the manuscript.

References

[1] G.G. Harrigan, R. Goodacre (Editors), Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis, Kluwer Academic Publishers, London, UK, 2003.
[2] O. Fiehn, Plant Mol. Biol. 48 (2002) 155.
[3] R. Goodacre, S. Vaidyanathan, W.B. Dunn, G.G. Harrigan, D.B. Kell, Trends Biotechnol. 22 (2004) 245.
[4] D.B. Kell, Curr. Opin. Microbiol. 7 (2004) 296.
[5] L.W. Sumner, P. Mendes, R.A. Dixon, Phytochemistry 62 (2003) 817.
[6] S.G. Oliver, M.K. Winson, D.B. Kell, F. Baganz, Trends Biotechnol. 16 (1998) 373.
[7] B. Lahner, J.M. Gong, M. Mahmoudian, E.L. Smith, K.B. Abid, E.E. Rogers, M.L. Guerinot, J.F. Harper, J.M. Ward, L. McIntyre, J.I. Schroeder, D.E. Salt, Nat. Biotechnol. 21 (2003) 1215.
[8] J. Forster, I. Famili, P. Fu, B.O. Palsson, J. Nielsen, Genome Res. 13 (2003) 244.
[9] J. Vandergreef, Trends Anal. Chem. 5 (1986) 241.
[10] K. Tanaka, D.G. Hine, J. Chromatogr. A 239 (1982) 301.
[11] J.K. Nicholson, M.J. Buckingham, P.J. Sadler, Biochem. J. 211 (1983) 605.
[12] A. Goffeau, B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston, E.J. Louis, H.W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S.G. Oliver, Science (Washington, DC) 274 (1996) 546.
[13] L.M. Raamsdonk, B. Teusink, D. Broadhurst, N.S. Zhang, A. Hayes, M.C. Walsh, J.A. Berden, K.M. Brindle, D.B. Kell, J.J. Rowland, H.V. Westerhoff, K. van Dam, S.G. Oliver, Nat. Biotechnol. 19 (2001) 45.
[14] U. Roessner, L. Willmitzer, A.R. Fernie, Plant Cell Rep. 21 (2002) 189.
[15] G. Le Gall, I.J. Colquhoun, A.L. Davis, G.J. Collins, M.E. Verhoeyen, J. Agric. Food Chem. 51 (2003) 2447.
[16] S. Vaidyanathan, D.B. Kell, R. Goodacre, J. Am. Soc. Mass Spectrom. 13 (2002) 118.
[17] S.M. Watkins, J.B. German, Curr. Opin. Biotechnol. 13 (2002) 512.
[18] B. Teusink, H.V. Westerhoff, Eur. J. Biochem. 267 (2000) 1889.
[19] I. Tkac, P.G. Henry, P. Andersen, C.D. Keene, W.C. Low, R. Gruetter, Magn. Reson. Med. 52 (2004) 478.
[20] E.M. Lenz, J. Bright, I.D. Wilson, S.R. Morgan, A.F.P. Nash, J. Pharm. Biomed. Anal. 33 (2003) 1103.
[21] K.S. Solanky, N.J.C. Bailey, B.M. Beckwith-Hall, A. Davis, S. Bingham, E. Holmes, J.K. Nicholson, A. Cassidy, Anal. Biochem. 323 (2003) 197.
[22] F.W. Ma, L.L. Cheng, Plant Sci. 165 (2003) 819.
[23] O. Al Jowder, E.K. Kemsley, R.H. Wilson, Food Chem. 59 (1997) 195.
[24] U. Roessner, C. Wagner, J. Kopka, R.N. Trethewey, L. Willmitzer, Plant J. 23 (2000) 131.
[25] C. Wittmann, J.O. Kromer, P. Kiefer, T. Binz, E. Heinzle, Anal. Biochem. 327 (2004) 135.
[26] A. Tava, L. Pecetti, M. Povolo, G. Contarini, Phytochem. Anal. 11 (2000) 148.
[27] W. Weckwerth, K. Wenzel, O. Fiehn, Proteomics 4 (2004) 78.
[28] D.T. Rossi, M.W. Sinz, Mass Spectrometry in Drug Discovery, Marcel Dekker, New York, USA, 2002.
[29] C.H. Deng, X.M. Zhang, N. Li, J. Chromatogr. B 808 (2004) 269.
[30] R.M.M. Perera, P.J. Marriott, I.E. Galbally, Analyst (Cambridge, UK) 127 (2002) 1601.
[31] J.M. Halket, A. Przyborowska, S.E. Stein, W.G. Mallard, S. Down, R.A. Chalmers, Rapid Commun. Mass Spectrom. 13 (1999) 279.
[32] A.L. Duran, J. Yang, L.J. Wang, L.W. Sumner, Bioinformatics 19 (2003) 2283.
[33] P. Jonsson, J. Gullberg, A. Nordstrom, M. Kusano, M. Kowalczyk, M. Sjostrom, T. Moritz, Anal. Chem. 76 (2004) 1738.
[34] C. Wagner, M. Sefkow, J. Kopka, Phytochemistry 62 (2003) 887.
[35] O. Fiehn, J. Kopka, P. Dormann, T. Altmann, R.N. Trethewey, L. Willmitzer, Nat. Biotechnol. 18 (2000) 1157.
[36] U. Roessner-Tunali, B. Hegemann, A. Lytovchenko, F. Carrari, C. Bruedigam, D. Granot, A.R. Fernie, Plant Physiol. 133 (2003) 84.
[37] J.C. Verdonk, C.H.R. de Vos, H.A. Verhoeven, M.A. Haring, A.J. van Tunen, R.C. Schuurink, Phytochemistry 62 (2003) 997.
[38] C. Grote, J. Pawliszyn, Anal. Chem. 69 (1997) 587.
[39] A. Guttman, M. Varoglu, J. Khandurina, Drug Discovery Today 9 (2004) 136.

[40] R. Bakhtiar, L. Ramos, F.L.S. Tse, J. Liq. Chromatogr. Relat. Technol. 25 (2002) 507.
[41] W.J. Leavens, S.J. Lane, R.M. Carr, A.M. Lockie, I. Waterhouse, Rapid Commun. Mass Spectrom. 16 (2002) 433.
[42] E.M. Lenz, J. Bright, R. Knight, I.D. Wilson, H. Major, J. Pharm. Biomed. Anal. 35 (2004) 599.
[43] A.W.T. Bristow, K.S. Webb, A.T. Lubben, J. Halket, Rapid Commun. Mass Spectrom. 18 (2004) 1447.
[44] R. King, R. Bonfiglio, C. Fernandez-Metzler, C. Miller-Stein, T. Olah, J. Am. Soc. Mass Spectrom. 11 (2000) 942.
[45] E. von Roepenack-Lahaye, T. Degenkolb, M. Zerjeski, M. Franz, U. Roth, L. Wessjohann, J. Schmidt, D. Scheel, S. Clemens, Plant Physiol. 134 (2004) 548.
[46] V.V. Tolstikov, O. Fiehn, Anal. Biochem. 301 (2002) 298.
[47] O. Corcoran, M. Spraul, Drug Discovery Today 8 (2003) 624.
[48] D.V. Huhman, L.W. Sumner, Phytochemistry 59 (2002) 347.
[49] P.D. Fraser, M.E.S. Pinto, D.E. Holloway, P.M. Bramley, Plant J. 24 (2000) 551.
[50] S.C. Brown, G. Kruppa, J.-L. Dasseux, Mass Spectrom. Rev. 24 (2005) 223.
[51] J.I. Castrillo, A. Hayes, S. Mohammed, S.J. Gaskell, S.G. Oliver, Phytochemistry 62 (2003) 929.
[52] J. Allen, H.M. Davey, D. Broadhurst, J.K. Heald, J.J. Rowland, S.G. Oliver, D.B. Kell, Nat. Biotechnol. 21 (2003) 692.
[53] J. Smedsgaard, J.C. Frisvad, J. Microbiol. Methods 25 (1996) 5.
[54] M.S. Rashed, J. Chromatogr. B 758 (2001) 27.
[55] T. Soga, Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita, T. Nishioka, J. Proteome Res. 2 (2003) 488.
[56] U. Utzinger, D.L. Heintzelman, A. Mahadevan-Jansen, A. Malpica, M. Follen, R. Richards-Kortum, Appl. Spectrosc. 55 (2001) 955.
[57] D. Lafrance, L.C. Lands, D.H. Burns, Talanta 60 (2003) 635.
[58] T. Nakamura, T. Takeuchi, A. Terada, Y. Tando, T. Suda, Int. J. Pancreatol. 23 (1998) 137.
[59] D. Naumann, D. Helm, H. Labischinski, Nature (London) 351 (1991) 81.
[60] R. Goodacre, E.M. Timmins, P.J. Rooney, J.J. Rowland, D.B. Kell, FEMS Microbiol. Lett. 140 (1996) 233.
[61] D.I. Ellis, D. Broadhurst, D.B. Kell, J.J. Rowland, R. Goodacre, Appl. Environ. Microbiol. 68 (2002) 2822.
[62] N.N. Kaderbhai, D.I. Broadhurst, D.I. Ellis, R. Goodacre, D.B. Kell, Comp. Funct. Genom. 4 (2003) 376.
[63] M. Diem, S. Boydston-White, L. Chiriboga, Appl. Spectrosc. 53 (1999) 148A.
[64] L. Chiriboga, P. Xie, H. Yee, V. Vigorita, D. Zarou, D. Zakim, M. Diem, Biospectroscopy 4 (1998) 47.
[65] P. Lasch, W. Haensch, E.N. Lewis, L.H. Kidder, D. Naumann, Appl. Spectrosc. 56 (2002) 1.
[66] M. Diem, L. Chiriboga, P. Lasch, A. Pacifico, Biopolymers 67 (2002) 349.
[67] N. Thomas, R. Goodacre, E.M. Timmins, M. Gaudoin, R. Fleming, Hum. Reprod. 15 (2000) 1667.
[68] H.H. Eysel, M. Jackson, A. Nikulin, R.L. Somorjai, G.T.D. Thomson, H.H. Mantsch, Biospectroscopy 3 (1997) 161.
[69] W. Petrich, B. Dolenko, J. Fruh, M. Ganz, H. Greger, S. Jacob, F. Keller, A.E. Nikulin, M. Otto, O. Quarder, R.L. Somorjai, A. Staib, G. Warner, H. Wielinger, Appl. Opt. 39 (2000) 3372.
[70] J. Schmitt, M. Beekes, A. Brauer, T. Udelhoven, P. Lasch, D. Naumann, Anal. Chem. 74 (2002) 3865.
[71] C. Petibois, G. Cazorla, G. Deleris, Appl. Spectrosc. 56 (2002) 10.
[72] H.M. Heise, G. Voigt, P. Lampen, L. Kupper, S. Rudloff, G. Werner, Appl. Spectrosc. 55 (2001) 434.
[73] A. Markus, D.W. Swinkels, B.S. Jakobs, R.A. Wevers, J.M.F. Trijbels, H.L. Willems, Anal. Chim. Acta. 429 (2001) 287.
[74] D. Naumann, Appl. Spectrosc. Rev. 36 (2001) 239.
[75] W. Petrich, Appl. Spectrosc. Rev. 36 (2001) 181.
[76] E. Gidman, R. Goodacre, B. Emmett, A.R. Smith, D. Gwynn-Jones, Phytochemistry 63 (2003) 705.
[77] H.E. Johnson, D. Broadhurst, R. Goodacre, A.R. Smith, Phytochemistry 62 (2003) 919.
[78] H.E. Johnson, D. Broadhurst, D.B. Kell, M.K. Theodorou, R.J. Merry, G.W. Griffith, Appl. Environ. Microbiol. 70 (2004) 1583.
[79] J.C. Lindon, E. Holmes, J.K. Nicholson, Anal. Chem. 75 (2003) 384A.
[80] N.V. Reo, Drug Chem. Toxicol. 25 (2002) 375.
[81] H.C. Keun, O. Beckonert, J.L. Griffin, C. Richter, D. Moskau, J.C. Lindon, J.K. Nicholson, Anal. Chem. 74 (2002) 4588.
[82] J.L. Griffin, A.W. Nicholls, H.C. Keun, R.J. Mortishire-Smith, J.K. Nicholson, T. Kuehn, Analyst (Cambridge, UK) 127 (2002) 582.
[83] M. Defernez, I.J. Colquhoun, Phytochemistry 62 (2003) 1009.
[84] M. Coen, S.U. Ruepp, J.C. Lindon, J.K. Nicholson, F. Pognan, E.M. Lenz, I.D. Wilson, J. Pharm. Biomed. Anal. 35 (2004) 93.
[85] J.T. Brindle, H. Antti, E. Holmes, G. Tranter, J.K. Nicholson, H.W.L. Bethell, S. Clarke, P.M. Schofield, E. McKilligin, D.E. Mosedale, D.J. Grainger, Nat. Med. 8 (2002) 1439.
[86] M. Satake, B. Dmochowska, Y. Nishikawa, J. Madaj, J. Xue, Z.W. Guo, D.V. Reddy, P.L. Rinaldi, V.M. Monnier, Invest. Ophthalmol. Vis. Sci. 44 (2003) 2047.
[87] M.S. Chauton, T.R. Storseth, G. Johnsen, J. Appl. Phycol. 15 (2003) 533.
[88] J.L. Ward, C. Harris, J. Lewis, M.H. Beale, Phytochemistry 62 (2003) 949.
[89] H.K. Choi, Y.H. Choi, M. Verberne, A.W.M. Lefeber, C. Erkelens, R. Verpoorte, Phytochemistry 65 (2004) 857.
[90] J.G. Bundy, D.J. Spurgeon, C. Svendsen, P.K. Hankard, D. Osborn, J.C. Lindon, J.K. Nicholson, FEBS Lett. 521 (2002) 115.
[91] M.R. Viant, E.S. Rosenblum, R.S. Tjeerdema, Toxicol. Sci. 72 (2003) 1165.
[92] N. Aranibar, B.K. Singh, G.W. Stockton, K.H. Ott, Biochem. Biophys. Res. Commun. 286 (2001) 150.
[93] D.B. Kell, P. Mendes, in: A. Cornish-Bowden, M.L. Cardenas (Editors), Technological and Medical Implications of Metabolic Control Analysis, Kluwer Academic Publishers, London, UK, 1999, p. 3.
[94] E. Urbanczyk-Wochniak, A. Luedemann, J. Kopka, J. Selbig, U. Roessner-Tunali, L. Willmitzer, A.R. Fernie, EMBO Rep. 4 (2003) 989.

Warwick Dunn obtained an analytical chemistry-based B.Sc. (Hons) degree and a MS-based Ph.D. degree from The University of Hull, UK. Subsequently, he has employed organic and isotope-ratio MS along with gas and liquid chromatographic techniques in solving biological and chemical problems in academia and industrial environments. He currently works as Research Associate in the Bioanalytical Sciences Group at Manchester University studying the metabolomes of microbes, plants and animals.

David Ellis obtained an environmental science B.Sc. (Hons) degree and a Ph.D. in microbiology/FT-IR from The University of Wales, Aberystwyth. He now works as Experimental Officer/Laboratory Manager in the Bioanalytical Sciences Group at the University of Manchester on a variety of metabolomics-based projects, which will continue in the Manchester Interdisciplinary Biocentre (MIB), which is to be completed soon.