Pharmaceutical Research, Vol. 23, No. 6, June 2006 (© 2006)
DOI: 10.1007/s11095-006-0025-z

Expert Review

Metabonomics Techniques and Applications to Pharmaceutical Research & Development

John C. Lindon,[1,2] Elaine Holmes,[1] and Jeremy K. Nicholson[1]

Received December 21, 2005; accepted January 13, 2006

Abstract. In this review, the background to the approach known as metabonomics is provided, giving a brief historical perspective and summarizing the analytical and statistical techniques used. Some of the major applications of metabonomics relevant to pharmaceutical Research & Development are then reviewed including the study of various influences on metabolism, such as diet, lifestyle, and other environmental factors. The applications of metabonomics in drug safety studies are explained with special reference to the aims and achievements of the Consortium for Metabonomic Toxicology. Next, the role that metabonomics might have in disease diagnosis and therapy monitoring is provided with some examples, and the concept of pharmacometabonomics as a way of predicting an individual's response to treatment is highlighted. Some discussion is given on the strengths and weaknesses, opportunities of, and threats to metabonomics.

KEY WORDS: biomarkers; diagnostics; drug safety; metabonomics; spectroscopy.

METABONOMICS BACKGROUND AND ITS ROLE IN PHARMACEUTICAL RESEARCH & DEVELOPMENT

There has been a revolution in the techniques and approaches used in molecular biology over the past decade or so, and following the decoding of the human and other genomes, studies have largely switched to simultaneous determination of gene expression changes between subjects, or following drug treatment or other intervention, mainly carried out using microarray technology (1). This type of study has been given the name of transcriptomics. Later, an equivalent impetus to map out all protein expression changes in a cell or tissue has evolved and has been termed proteomics. Nowadays, there are nearly 200 different named "omics," many of which terms will not survive because of their very specialist application, and indeed, many of these terms are not necessary because they serve only to describe a methodology that already has a perfectly valid name.

Prior to the development of the various omics approaches, the simultaneous analysis of the plethora of metabolites seen in biological fluids had been carried out largely using nuclear magnetic resonance (NMR) spectroscopy (2), and when these complex data sets were first interpreted using multivariate statistics (3,4), the concept of metabonomics was born. Thus, metabonomics encompasses the comprehensive and simultaneous systematic profiling of metabolite levels and their systematic and temporal changes through such effects on diet, lifestyle, environment, genetics, and pharmaceuticals, both beneficial and adverse, in whole organisms. This is achieved by the study of biofluids and tissues with the data being interpreted using chemometrics techniques (5,6). A parallel approach mainly from plant science and from the study of *in vitro* systems has led to the term metabolomics also being coined (7), and the methods and approaches used in the two disciplines are now highly convergent.

This multivariate approach holds out the promise for the pharmaceutical communities of a means by which disease and drug effect endpoints can be obtained. One of the problems with transcriptomics, in particular, is the difficulty in some cases of relating observed gene expression fold changes to such conventional disease and pharmaceutically relevant endpoints. This does not apply so much to proteomics, but this technology is still slow and labor-intensive, and techno-

[1] Biological Chemistry, Biomedical Sciences Division, Imperial College London, Sir Alexander Fleming Building, South Kensington, London SW7 2AZ, UK.
[2] To whom correspondence should be addressed. (e-mail: j.lindon@imperial.ac.uk)

ABBREVIATIONS: CE, capillary electrophoresis; CLOUDS, classification of unknowns by density superposition; COMET, Consortium for Metabonomic Toxicology; COSY, correlation spectroscopy; CPMG, Carr-Purcell-Meiboom-Gill (an NMR pulse sequence); CSF, cerebrospinal fluid; DA, discriminant analysis; FT, Fourier transform; GC, gas chromatography; HPLC, high-performance liquid chromatography; IR, infrared; LC, liquid chromatography; LIMS, laboratory information management system; MAS, magic angle spinning; MS, mass spectrometry; NMR, nuclear magnetic resonance; OSC, orthogonal signal correction; PCA, principal component analysis; PLS, partial least squares; QC, quality control; SHY, statistical heterospectroscopy; SMRS, Standard Metabolic Reporting Structures; STOCSY, statistical total correlation spectroscopy; TOCSY, total correlation spectroscopy; TOF, time of flight; TSP, trimethylsilylpropionic acid sodium salt; UPLC, ultra-performance liquid chromatography.

logical advances are required before it can be made high throughput.

Nevertheless, there remains a strong imperative to be able to integrate information at the transcriptomic, proteomic, and metabonomic levels despite these different levels of biological control showing very different timescales of change. This is because some time courses can be very rapid, such as gene switching, some require much longer timescales, e.g., protein synthesis, or, in the case of metabolic changes, can encompass enormous ranges of timescales. Counterintuitively, biochemical changes do not always occur in the order, transcriptomic, proteomic, metabolic, because, for example, pharmacological or toxicological effects at the metabolic level can induce subsequent adaptation effects at the proteomic or transcriptomic levels. One important potential role for high throughput and highly automated metabonomics methods, therefore, could be to direct the timing of more expensive or labor-intensive proteomic and transcriptomic analyses to maximize the probability of observing meaningful and relevant biochemical changes using those techniques.

In addition, overlaid with this temporal complexity is the fact that environmental and lifestyle effects have a large effect at all levels of molecular biology. Gene and protein expression effects and metabolite levels can be altered by such factors, and this variation has to be incorporated into any analysis as part of intersample and interindividual variation. Even healthy animals and man can be considered as "super organisms," with an internal ecosystem of diverse symbiotic gut microflora that have metabolic processes that interact with the host and for which, in many cases, the genome is not known. The complexity of mammalian biological system and the diverse features that need to be measured to allow omics data to be fully interpreted have been reviewed recently (8), and it has been argued that novel approaches will continue to be required to measure and model metabolic processes in various compartments from such global systems with different interacting cell types, and with various genomes, connected by cometabolic processes (9).

In this review, the main technologies used in metabonomics are summarized, brief details of the types of samples used are given, and the current pharmaceutical applications of metabonomics are described. Some prospects for the future are then discussed.

METABONOMICS SAMPLES

Metabonomics studies of pharmaceutical relevance generally use biofluids or cell or tissue extracts. These are often easy to obtain and, for mammalian biofluids, can provide an integrated view of the whole systems biology. Urine and plasma are obtained essentially noninvasively and hence can be used more easily for disease diagnosis and, in a clinical trial setting, for monitoring drug therapy. However, there is a wide range of fluids that have been studied, including seminal fluids, amniotic fluid, cerebrospinal fluid (CSF), synovial fluid, digestive fluids, blister and cyst fluids, lung aspirates, and dialysis fluids. In addition, a number of metabonomics studies have used analysis of tissue biopsy samples and their lipid and aqueous extracts, such as from vascular tissue in studies of atherosclerosis (10). The approach can also be used to characterize *in vitro* cell systems such as Caco-2 cells, commonly used for cell uptake studies (11), and other model cell systems such as yeast (12), tumor cells (13), and tissue spheroids, which can be used as model systems for liver or tumor investigations, for example (14).

METABONOMICS ANALYTICAL TECHNOLOGIES

The main analytical techniques that are employed for metabonomic studies are based on NMR spectroscopy and mass spectrometry (MS). The latter technique requires a preseparation of the metabolic components using either gas chromatography (GC) after chemical derivatization or liquid chromatography (LC), with the newer method of ultra-performance liquid chromatography (UPLC) being used increasingly. The use of capillary electrophoresis (CE) coupled to MS has also shown some promise. Other more specialized techniques such as Fourier transform infrared (FTIR) spectroscopy and arrayed electrochemical detection have been used in some cases (15,16). The main limitation of the use of FTIR is the low level of detailed molecular identification that can be achieved, and indeed, in the case quoted above, MS was also employed for metabolite identification.

Similarly, although an array of coulometric detectors following high-performance liquid chromatography (HPLC) separation does not identify compounds directly, the combination of retention time and redox properties can serve as a basis for database searching of libraries of standard compounds. The separation output can also be directed to a mass spectrometer for additional identification experiments (16).

All metabonomics studies result in complex multivariate data sets that require visualization software and chemometric and bioinformatic methods for interpretation. The aim of these procedures is to produce biochemically based fingerprints that are of diagnostic or other classification value. A second stage, crucial in such studies, is to identify the substances causing the diagnosis or classification, and these become the combination of biomarkers that define the biological or clinical context.

There have been a number of reviews of metabonomics recently that describe the various techniques used and which also summarize the main areas of application (17,18).

Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a nondestructive technique, widely used in chemistry, that provides detailed information on molecular structure, both for pure compounds and in complex mixtures (19). NMR spectroscopic methods can also be used to probe metabolite molecular dynamics and mobility as well as substance concentrations through the interpretation of NMR spin relaxation times and by the determination of molecular diffusion coefficients (20).

Automatic sample preparation is possible for NMR spectroscopy involving buffering and addition of D_2O as a magnetic field lock signal for the spectrometer, and standard NMR spectra typically take only a few minutes to acquire using robotic flow-injection methods. For large-scale studies, bar-coded vials containing the biofluid can be used, and the contents of these can be transferred and prepared for analysis using robotic liquid handling technology into 96-well plates

under laboratory information management system (LIMS) control. Currently, using such approaches, well over 100 samples per day can be measured on one spectrometer, each taking a total data acquisition time of only around 5 min. Alternatively, for more precious samples or for those of limited volume, conventional 5-mm or capillary NMR tubes are usually used, either individually or using a commercial sample tube changer and automatic data acquisition.

A typical ^{1}H NMR spectrum of urine contains thousands of sharp lines from predominantly low molecular weight metabolites. The large interfering NMR signal arising from water in all biofluids is easily eliminated by use of appropriate standard NMR solvent suppression methods, either by secondary RF irradiation at the water peak chemical shift or by use of a specialized NMR pulse sequence that does not excite the water resonance. The position of each spectral band (known as its chemical shift and measured in frequency terms, in parts per million, from that of an added standard reference substance) gives information on molecular group identity and its molecular environment. The reference compound used in aqueous media is usually the sodium salt of 3-trimethylsilylpropionic acid (TSP) with the methylene groups deuterated to avoid giving rise to peaks in the ^{1}H NMR spectrum. The multiplicity of the splitting pattern on each NMR band and the magnitudes of the splittings (caused by nuclear spin–spin interactions mediated through the electrons of the chemical bonds and known as J coupling) provide knowledge about nearby protons, their through-bond connectivities, the relative orientation of nearby C–H bonds, and hence also molecular conformations. The band areas relate directly to the number of protons giving rise to the peak and hence to the relative concentrations of the substances in the sample. Absolute concentrations can be obtained if the sample contains an added internal standard of known concentration, or if a standard addition of the analyte of interest is added to the sample, or if the concentration of a substance is known by independent means (e.g., glucose in plasma can be quantified by a conventional biochemical assay).

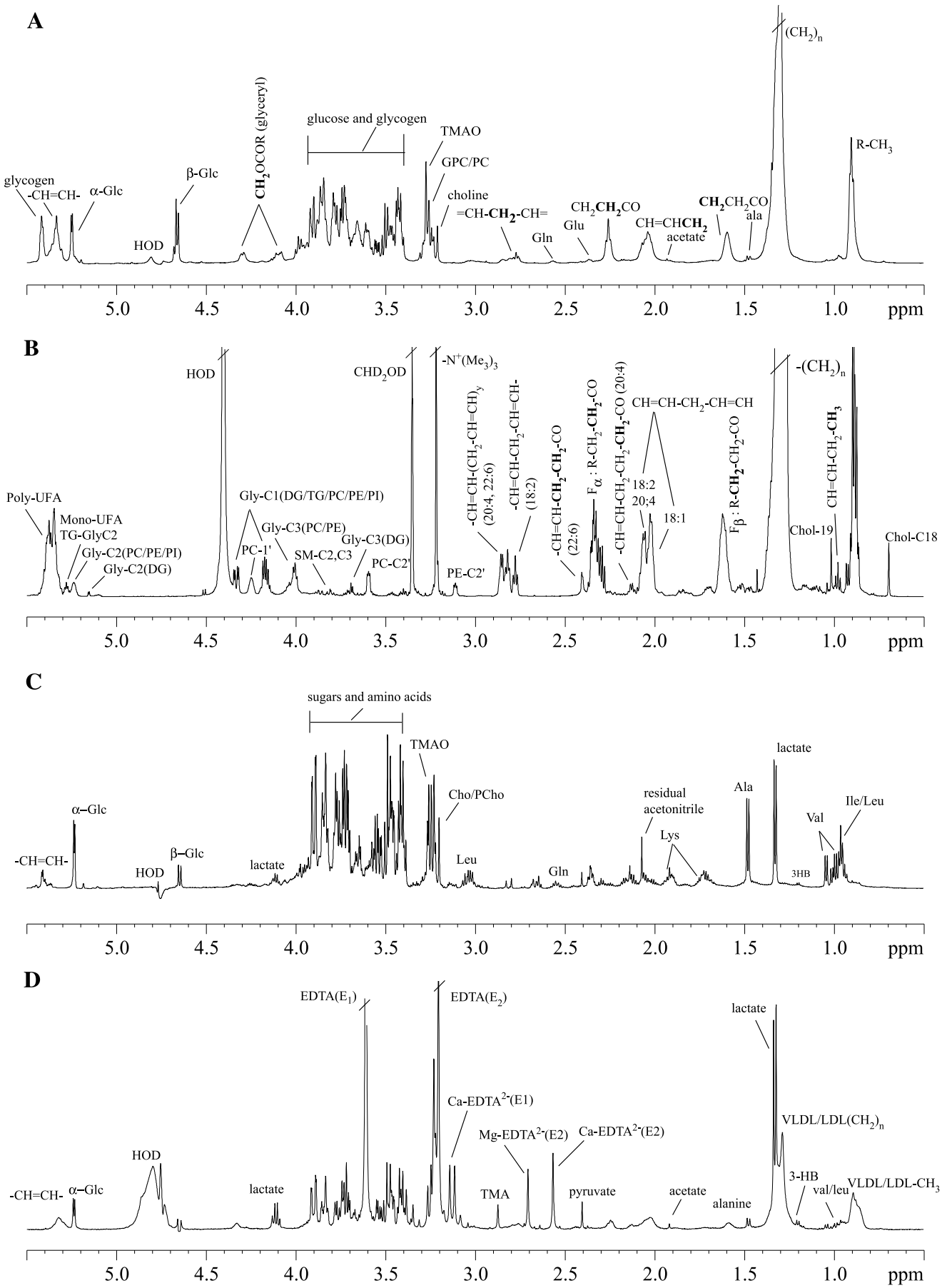
Blood plasma and serum contain both low and high molecular weight components, and these give a wide range of signal linewidths. Broad bands from protein and lipoprotein signals contribute strongly to the ^{1}H NMR spectra, with sharp peaks from small molecules superimposed on them (21). Standard NMR pulse sequences, where the observed peak intensities are edited on the basis of molecular diffusion coefficients or on NMR relaxation times [such as the Carr-Purcell-Meiboom-Gill (CPMG) spin-echo sequence], can be used to select only the contributions from macromolecules or, alternatively, to select only the signals from the small molecule metabolites, respectively (19). It is also possible to use these approaches to investigate molecular mobility and flexibility and to study intermolecular interactions such as the reversible binding between small molecules and proteins (20). Some typical ^{1}H NMR spectra are given in Fig. 1, showing the varied profiles from mouse liver tissue, lipid and aqueous extracts of liver tissue, and blood plasma.

Identification of biomarkers can involve the application of a range of techniques including two-dimensional (2-D) NMR experiments (19). Although all of the armory of the usual analytical physical chemistry can be used, including MS, ^{1}H NMR spectra of urine and other biofluids, even though they are very complex, allow many resonances to be assigned directly based on their chemical shifts, signal multiplicities, and by adding authentic material, and further information can be obtained by using the spectral editing techniques described above.

2-D NMR spectroscopy can be useful for increasing signal dispersion and for elucidating the connectivities between signals, thereby enhancing the information content and helping to identify biochemical substances. These include the ^{1}H–^{1}H 2-D J-resolved experiment, which attenuates the peaks from macromolecules and yields information on the multiplicity and coupling patterns of resonances, a good aid to molecule identification. The appropriate projection of such a spectrum onto the chemical shift axis yields a fingerprint of peaks from only the most highly mobile small molecules, with the added benefit that all of the spin-coupling peak multiplicities have been removed. Other 2-D experiments known as correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) provide ^{1}H–^{1}H spin–spin coupling connectivities, giving information on which hydrogens in a molecule are close in chemical bond terms. Use of other types of nuclei, such as naturally abundant ^{13}C or ^{15}N, or where present ^{31}P, can be important to help assign NMR peaks, and here, such heteronuclear correlation NMR experiments are achievable. These benefit from the use of so-called inverse detection, where the lower sensitivity or less abundant nucleus NMR spectrum (such as ^{13}C) is detected indirectly using the more sensitive/abundant nucleus (^{1}H) by making use of spin–spin interactions such as the one-bond ^{13}C–^{1}H spin–spin coupling between the nuclei to effect the connection. These yield both ^{1}H and ^{13}C NMR chemical shifts of CH, CH_2, and CH_3 groups, useful again for identification purposes. There is also a sequence that allows correlation of protons to quaternary carbons based on long-range ^{13}C–^{1}H spin–spin coupling between the nuclei.

A very useful recent advance in NMR technology has been the development of cryogenic probes where the detector coil and preamplifier (but not the samples) are cooled to around 20 K. This has provided an improvement in spectral signal-to-noise ratios of up to a factor of 5 by reducing the thermal noise in the electronics of the spectrometer. Conversely, because the NMR signal-to-noise ratio is proportional to the square root of the number of coadded scans, shorter data acquisition times by up to a factor of 25 become possible for the same amount of sample. NMR spectroscopy of biofluids detecting the much less sensitive ^{13}C nuclei, which also only have a natural abundance (1.1%), also becomes possible because of the increase in signal-to-noise ratio (22). This technology also makes the use of tissue-specific microdialysis samples more feasible (23).

Within the last few years, the development of a technique called high-resolution ^{1}H magic angle spinning (MAS) NMR spectroscopy has made feasible the acquisition of high-resolution NMR data on small pieces of intact tissues with no pretreatment (24–26). Rapid spinning of the sample (typically at ~4–6 kHz) at an angle of 54.7° relative to the applied magnetic field serves to reduce the loss of information caused by line-broadening effects seen in nonliquid samples such as tissues. These broadenings are caused by sample heterogeneity and residual anisotropic NMR parameters that are normally averaged out in free solution where


A
glycogen
-CH=CH-
α-Glc
HOD
β-Glc
CH2OCOR (glyceryl)
glucose and glycogen
TMAO
GPC/PC
choline
=CH-CH2-CH=
Gln
Glu
CH2CH2CO
CH=CHCH2
acetate
CH2CH2CO
ala
(CH2)n
R-CH3
5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 ppm
B
HOD
CHD2OD
-N+(Me3)3
-CH=CH-(CH2-CH=CH)y (20:4, 22:6)
-CH=CH-CH2-CH=CH- (18:2)
-CH=CH-CH2-CH2-CO (22:6)
Fα : R-CH2-CH2-CO
-CH=CH-CH2-CH2-CO (20:4)
CH=CH-CH2-CH=CH
18:2
20;4
18:1
Fβ : R-CH2-CH2-CO
-(CH2)n
CH=CH-CH2-CH3
Poly-UFA
Mono-UFA
TG-GlyC2
Gly-C2(PC/PE/PI)
Gly-C2(DG)
Gly-C1(DG/TG/PC/PE/PI)
Gly-C3(PC/PE)
PC-1'
SM-C2,C3
Gly-C3(DG)
PC-C2'
PE-C2'
Chol-19
Chol-C18
5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 ppm
C
sugars and amino acids
TMAO
Cho/PCho
lactate
Ala
α–Glc
residual acetonitrile
Ile/Leu
Val
Lys
-CH=CH-
HOD
β–Glc
lactate
Leu
Gln
3HB
5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 ppm
D
EDTA(E1)
EDTA(E2)
lactate
Ca-EDTA2-(E1)
VLDL/LDL(CH2)n
Mg-EDTA2-(E2)
Ca-EDTA2-(E2)
HOD
3-HB
-CH=CH-
α–Glc
lactate
TMA
pyruvate
acetate
alanine
val/leu
VLDL/LDL-CH3
5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 ppm

molecules can tumble isotropically and rapidly. MAS NMR spectroscopy has straightforward, but manual, sample preparation. NMR spectroscopy on a tissue sample in a MAS experiment is the same as solution-state NMR, and all common pulse techniques can be employed to study metabolic changes and to perform molecular structure elucidation and molecular dynamics studies.

Mass Spectrometry

Mass spectrometry (MS) has also been widely used in metabolic fingerprinting and metabolite identification. Although most studies to date have been on plant extracts and model cell system extracts, its application to mammalian studies is increasing. In general, a prior separation of the complex mixture sample using chromatography is required. MS is inherently considerably more sensitive than NMR spectroscopy, but it is necessary generally to employ different separation techniques (e.g., different LC column packings) for different classes of substances. MS is also a major technique for molecular identification purposes, especially using tandem MS (MS-MS) methods for fragment ion studies. Analyte quantitation by MS in complex mixtures of highly variable composition can be impaired by variable ionization and ion suppression effects. For plant metabolic studies, most investigations have used chemical derivatization to insure volatility and analytical reproducibility, followed by GC-MS analysis. Some approaches using MS rely on more targeted studies, for example, by detailed analysis of lipids (27).

For metabonomics applications on biofluids such as urine, an HPLC chromatogram is generated with MS detection, usually using electrospray ionization, and both positive and negative ion chromatograms can be measured. At each sampling point in the chromatogram, there is a full mass spectrum; thus, the data are three dimensional in nature, i.e., retention time, mass, and intensity. Given this very high resolution, it is possible to cut out any mass peaks from interfering substances such as drug metabolites without affecting the integrity of the data set.

The recently introduced UPLC is a combination of a 1.7-μm reversed-phase packing material and a chromatographic system, operating at around 12,000 psi. This has enabled better chromatographic peak resolution and increased speed and sensitivity to be obtained for complex mixture separation. UPLC provides around a 10-fold increase in speed and a 3- to 5-fold increase in sensitivity compared with a conventional stationary phase. Because of the much improved chromatographic resolution of UPLC, the problem of ion suppression from coeluting peaks is greatly reduced. UPLC-MS has already been used for metabolic profiling of urines from males and females of two groups of phenotypically normal mouse strains and a nude mouse strain (28). A comparison of MS-detected HPLC and UPLC chromatograms from a mouse urine sample is shown in Fig. 2.

Recently, CE coupled to MS has also been explored as a suitable technology for metabonomics studies (29). Metabolites are first separated by CE based on their charge and size and then selectively detected using MS monitoring. This method has been used to measure 352 metabolic standards and then employed for the analysis of 1692 metabolites from *Bacillus subtilis* extracts, revealing changes in metabolite levels during the bacterial growth.

For biomarker identification, it is also possible to separate out substances of interest on a larger scale from a complex biofluid sample using techniques such as solid-phase extraction or HPLC. For metabolite identification, directly coupled chromatography-NMR spectroscopy methods can also be used. The most general of these "hyphenated" approaches is HPLC-NMR-MS (30), in which the eluting HPLC peak is split, with parallel analysis by directly coupled NMR and MS techniques. This can be operated in on-flow, stopped-flow, and loop-storage modes and thus can provide the full array of NMR- and MS-based molecular identification tools. These include 2-D NMR spectroscopy as well as MS-MS for identification of fragment ions and Fourier transform-mass spectrometry (FT-MS) or time of flight-mass spectrometry (TOF-MS) for accurate mass measurement and, hence, derivation of molecular empirical formulae.

In summary NMR and MS approaches are highly complementary, and use of both is often necessary for full molecular characterization. MS can be more sensitive with lower detection limits provided the substance of interest can be ionized, but NMR spectroscopy is particularly useful for distinguishing isomers, for obtaining molecular conformation information, and for studies of molecular dynamics and compartmentation.

Chemometrics Methods

An NMR spectrum of a biofluid sample can be thought of as an object with a multidimensional set of metabolic coordinates, the values of which are the spectral intensities at each data point, and the spectrum is therefore a point in a multidimensional metabolic hyperspace. The initial objective in metabonomics is to classify a spectrum based on identification of its inherent patterns of peaks and, secondly, to identify those spectral features responsible for the classification. The approach can also be used for reducing the dimensionality of complex data sets, for example, by 2-D or 3-D mapping procedures, to enable easy visualization of any clustering or similarity of the various samples. Alternatively, in what are known as supervised methods, multiparametric data sets can be modeled so that the class of separate samples (a validation set) can be predicted based on a series of mathematical models derived from the original data or training set (31).

One of the simplest techniques that has been used extensively in metabonomics is principal components analysis

Fig. 1. (A) ^{1}H magic angle spinning (MAS) nuclear magnetic resonance (NMR) Carr-Purcell-Meiboom-Gill (CPMG) spectrum (600 MHz) of intact control liver tissue, (B) ^{1}H NMR (600 MHz) spectrum of a control lipid soluble liver tissue extract, (C) solvent presaturation ^{1}H NMR spectrum (600 MHz) of a control aqueous soluble liver tissue extract, and (D) ^{1}H NMR CPMG spectrum (500 MHz) of control blood plasma. 3HB, 3-D-hydroxybutyrate; Cho, choline; Chol, cholesterol; Glu, glucose; GPC, glycerophosphorylcholine; Gly, glycerol; LDL, low density lipoprotein; PCho, phosphocholine; TMAO, trimethylamine-*N*-oxide; VLDL, very low density lipoprotein. Reproduced with permission from Coen *et al.* (65).

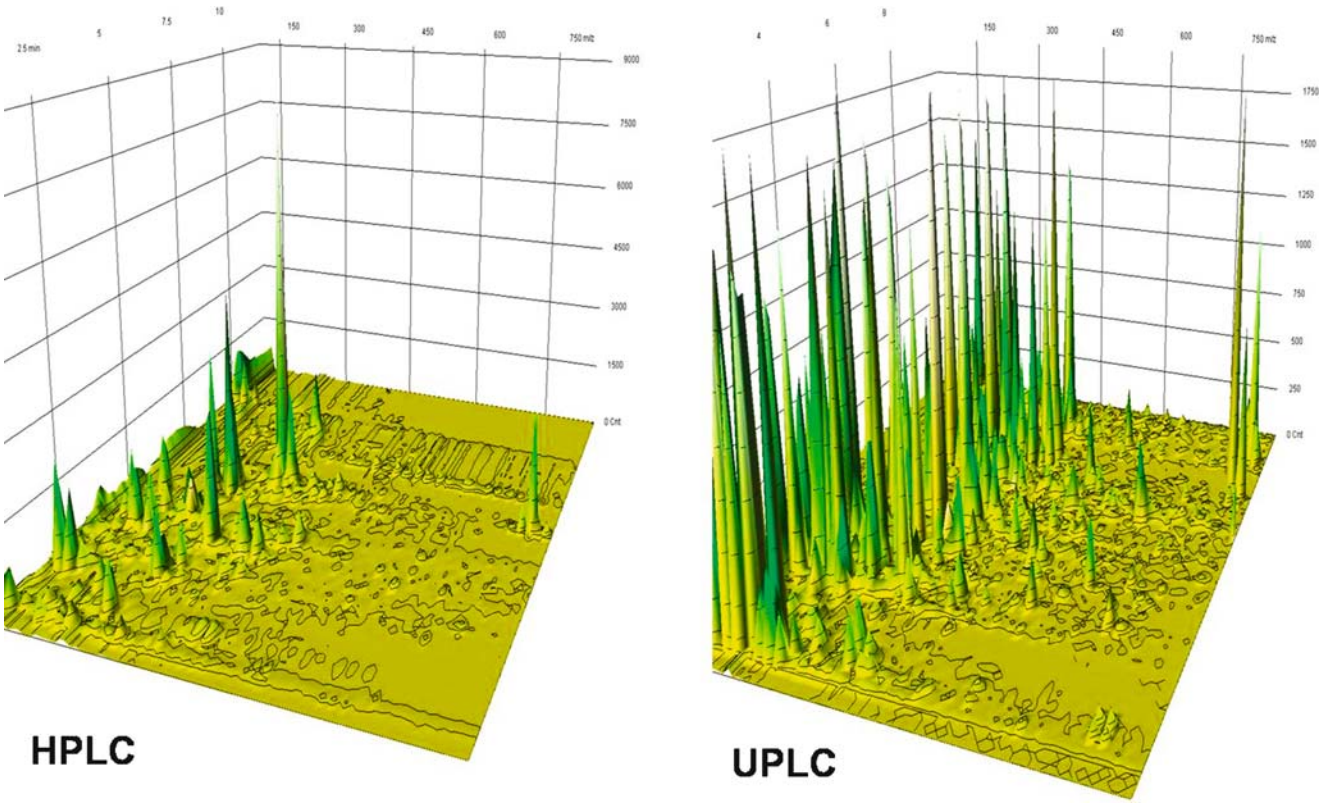


Fig. 2. Three-dimensional plots of retention time, m/z, and intensity from control white male mouse urine using (left) high-performance liquid chromatography-mass spectrometry (HPLC-MS) with a 2.1 cm × 100 mm Waters Symmetry 3.5 μm C18 column, eluted with 0–95% linear gradient of water with 0.1% formic acid/acetonitrile with 0.1% formic acid over 10 min at a flow rate of 0.6 mL/min and (right) ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) with 2.1 cm × 100 mm Waters ACQUITY 1.7 μm C18 column, eluted with the same solvents at a flow rate of 0.5 mL/min. In both cases, the column eluent was monitored by electrospray ionization orthogonal acceleration-time of flight-mass spectrometry from 50 to 850 m/z in positive ion mode. Reproduced with permission from Wilson *et al.* (28).

(PCA). This technique expresses most of the variance within a data set using a smaller number of factors or principal components. Each PC is a linear combination of the original data parameters whereby each successive PC explains the maximum amount of variance possible not accounted for by the previous PCs. Each PC is orthogonal and therefore independent of the other PCs; thus, the variation in the spectral set is usually described by many fewer PCs compared to the number of original data point values because the less important PCs simply describe the noise variation in the spectra. Conversion of the data matrix to PCs results in two matrices known as scores and loadings. Scores, the linear combinations of the original variables, are the coordinates for the samples in the established model and may be regarded as the new variables. In a scores plot, each point represents a single sample spectrum. The PC loadings define the way in which the old variables are linearly combined to form the new variables and indicate those variables carrying the greatest weight in transforming the position of the original samples from the data matrix into their new position in the scores matrix. In the loadings plot, each point represents a different spectral intensity. Thus, the cause of any spectral clustering observed in a PC scores plot is interpreted by examination of the loadings that cause any cluster separation. In addition, there are many other visualization (or unsupervised) methods, such as nonlinear mapping and hierarchical cluster analysis.

One widely used supervised method (i.e., using a training set of data with known outcomes) is partial least squares (PLS) (32). This is a method which relates a data matrix containing independent variables from samples, such as spectral intensity values (an **X** matrix), to a matrix containing dependent variables (e.g., measurements of response, such as toxicity scores) for those samples (a **Y** matrix). PLS can also be used to examine the influence of time on a data set, which is particularly useful for biofluid NMR data collected from samples taken over a time course of the progression of a pathological effect. PLS can also be combined with discriminant analysis (DA) to establish the optimal position to place a discriminant surface which best separates classes. It is possible to use such supervised models to provide classification probabilities and quantitative response factors for a wide range of sample types, but given the strong possibility of chance correlations when the number of descriptors is large, it is important to build and test such chemometric models using independent training data and validation data sets. Extensions of this approach allow the evaluation of those descriptors that are completely independent (orthogonal) to the **Y** matrix of endpoint data. This orthogonal signal correction (OSC) can be used to remove irrelevant and confusing parameters and has been integrated into the PLS algorithm for optimum use (33).

Apart from the methods described above that use linear combinations of parameters for dimension reduction or

classification, other methods exist that are not limited in this way. For example, neural networks comprise a widely used nonlinear approach for modeling data. A training set of data is used to develop algorithms, which "learn" the structure of the data and can cope with complex functions. The basic software network consists of three or more layers, including an input level of neurons (spectral descriptors or other variables), one or more hidden layers of neurons which adjust the weighting functions for each variable, and an output layer which designates the class of the object or sample. Recently, probabilistic neural networks, which represent an extension to the approach, have shown promise for metabonomics applications in toxicity (34). Other approaches that are currently being tested include genetic algorithms, machine learning, and Bayesian modeling (35).

New Approaches to Biomarker Identification Using Chemometrics

Recently, a new method for identifying multiple NMR peaks from the same molecule in a complex mixture, hence providing a new approach to molecular identification, has been introduced. This is based on the concept of statistical total correlation spectroscopy (STOCSY) (36). This takes advantage of the multicolinearity of the intensity variables in a set of spectra (e.g., ^{1}H NMR spectra) to generate a pseudo-2-D NMR spectrum that displays the correlation among the intensities of the various peaks across the whole sample. This method is not limited to the usual connectivities that are deducible from the more standard 2-D NMR spectroscopic methods, such as TOCSY. Added information is available by examining lower correlation coefficients or even negative correlations because this leads to connection between two or more molecules involved in the same biochemical pathway. In an extension of the method, the combination of STOCSY with supervised chemometrics methods offers a new framework for analysis of metabonomic data. In a first step, a supervised multivariate DA can be used to extract the parts of NMR spectra related to discrimination between two sample classes. This information is then combined with the STOCSY results to help identify the molecules responsible for the metabolic variation. To illustrate the applicability of the method, it has been applied to ^{1}H NMR spectra of urine from a metabonomic study of a model of insulin resistance based on the administration of a carbohydrate diet to three different mice strains, in which a series of metabolites of biological impor-

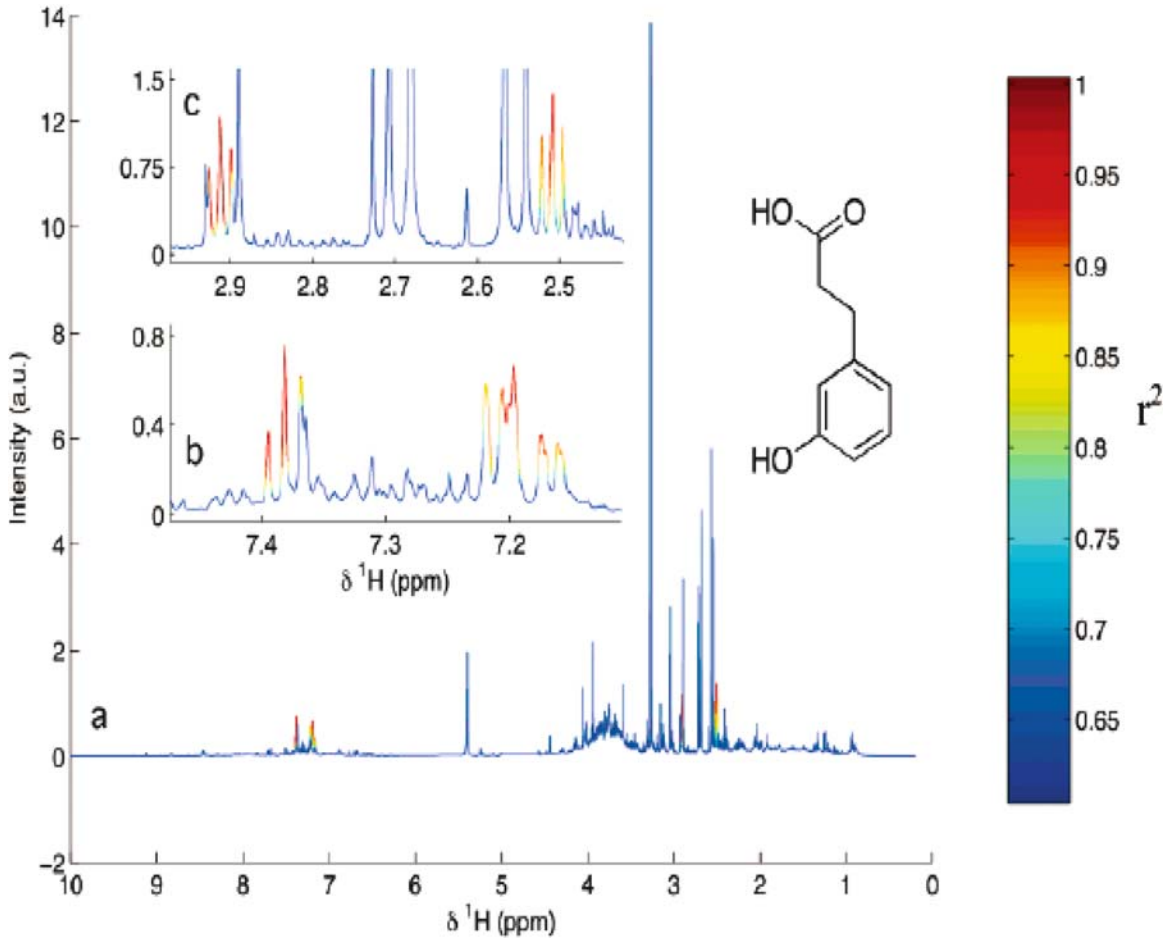


Fig. 3. One-dimensional statistical total correlation spectroscopy (STOCSY) analysis to identify peaks correlated to that at the chemical shift, δ 2.51. The degree of correlation across the spectrum has been color-coded and projected on the spectrum. (a) Full spectrum, (b) partial spectrum between δ 7.1 and 7.5, and (c) partial spectrum between δ 2.4 and 3.0. The STOCSY procedure enabled the assignment of this metabolite as 3-hydroxyphenylpropionic acid. Reproduced with permission from Cloarec *et al.* (36).

tance could be conclusively assigned and identified by use of the STOCSY approach (36). This is illustrated in Fig. 3, where the approach has been used to identify the metabolite 3-hydroxyphenylpropionic acid.

The approach is not limited to NMR spectra alone and has been extended to other forms of data. It has recently been applied to coanalysis of both NMR and mass spectra from a metabonomic toxicity study (37). This allowed better assignment of biomarkers of the toxin effect by using the correlated but complementary information available from the NMR and mass spectra taken on a whole sample cohort.

APPLICATIONS OF METABONOMICS RELEVANT TO PHARMACEUTICAL RESEARCH & DEVELOPMENT

Phenotypic and Physiological Effects

To relate therapeutic or toxic effects to normality or to understand the biochemical alterations caused by disease, it is necessary to have a good comprehension of what constitutes a normal biochemical profile. A number of studies have used metabonomics in this type of application to identify metabolic differences, in experimental animals such as mice and rats, caused by a range of inherent and external factors (38). Such differences may help explain differential toxicity of drugs between strains and interanimal variation within a study. Many effects can be distinguished using NMR-based metabonomics, including male/female differences, age-related changes, estrus cycle effects in females, diet, diurnal effects, and interspecies differences and similarities (38). Similarly some preliminary results have been obtained using the UPLC-MS method on normal and obese Zucker rats and on black, white, and nude mice (39). Considerable effort is being spent trying to elucidate the complex interactions between diet, health, and therapy (40).

Metabonomics has also been used for the phenotyping of mutant and transgenic animals and the investigation of the consequences of transgenesis such as the transfection process used to introduce a new gene (41). The development of a genetically engineered animal is often made using such transfection procedures, and it is important to differentiate unintended consequences of this process from the intended result. Metabonomic approaches can give insight into the metabolic similarities or differences between mutant or transgenic animals and the human disease processes that they are intended to simulate. This leads to a better evaluation of their appropriateness for use as disease models and for drug efficacy studies.

The importance of the symbiotic relationship between mammals and their gut microfloral populations has been recognized (8) and highlighted by a study in which axenic (germ free) rats were allowed to acclimatize in normal laboratory conditions, and their urine biochemical profiles were monitored for 21 days using ^{1}H NMR spectroscopy (42). An interesting example of the phenotypic differences caused by variations in gut microflora has been highlighted by the study of the same strain of rat from the same supplier but housed in separate colonies at the supplier (43). It was commented that the effect on drug metabolism and drug safety assessment of having different microfloral populations, in what would otherwise seem to be a homogenous population, is still unknown. Furthermore, the situation can be complicated by infections or pathological agents, and the combined influence of gut microflora and parasitic infections on urinary metabolite profiles has also been elucidated (44).

Preclinical Drug Candidate Safety Assessment

The selection of robust candidate drugs for development based on minimization of the occurrence of drug adverse effects is one of the most important aims of pharmaceutical Resarch & Development, and the pharmaceutical industry is now embracing metabonomics for evaluating the adverse effects of candidate drugs.

Metabonomics can be used for definition of the metabolic hyperspace occupied by normal animals and the consequential rapid classification of a biofluid sample as normal or abnormal. If the sample is regarded as abnormal, then classification of the target organ or region of toxicity, the biochemical mechanism of that toxin, the identification of combination biomarkers of toxic effect, and evaluation of the time course of the effect, e.g., the onset, evolution, and regression of toxicity, can all be determined. There have been many studies using ^{1}H NMR spectroscopy of biofluids to characterize drug toxicity going back to the 1980s (2), and the role of metabonomics in particular, and magnetic resonance in general, in toxicological evaluation of drugs has been comprehensively reviewed recently (45). However, the combined use of NMR spectroscopy and HPLC-MS is beginning to be used for toxicity studies, and this has been exemplified by a study on the nephrotoxin gentamycin (46).

The usefulness of metabonomics for the evaluation of xenobiotic toxicity effects has recently been comprehensively explored by the successful Consortium for Metabonomic Toxicology (COMET). This was formed between five pharmaceutical companies and Imperial College London, UK (47), with the aim of developing methodologies for the acquisition and evaluation of metabonomic data generated using ^{1}H NMR spectroscopy of urine and blood serum from rats and mice for preclinical toxicological screening of candidate drugs. A flowchart for the project operation is shown in Fig. 4 (top).

A feasibility study was carried out at the start of the project, using the same detailed protocol and using the same model toxin, over seven sites in the companies and their appointed contract research organizations. This was used to evaluate the levels of analytical and biological variations that could both arise using metabonomics on a multisite basis. The intersite NMR analytical reproducibility revealed the high degree of robustness expected for this technique when the same samples were analyzed both at Imperial College and at various company sites. This gave a multivariate coefficient of regression between paired samples of only about 1.6% (48). Additionally, the biological variability was evaluated by a detailed comparison of the ability of the companies to provide consistent urine and serum samples for an in-life study of the same toxin, with all samples measured at the Imperial College. There was a high degree of consistency between samples from the various companies, and dose-related effects could be distinguished from intersite variation.

To achieve the project goals, new methodologies for analyzing and classifying the complex data sets were

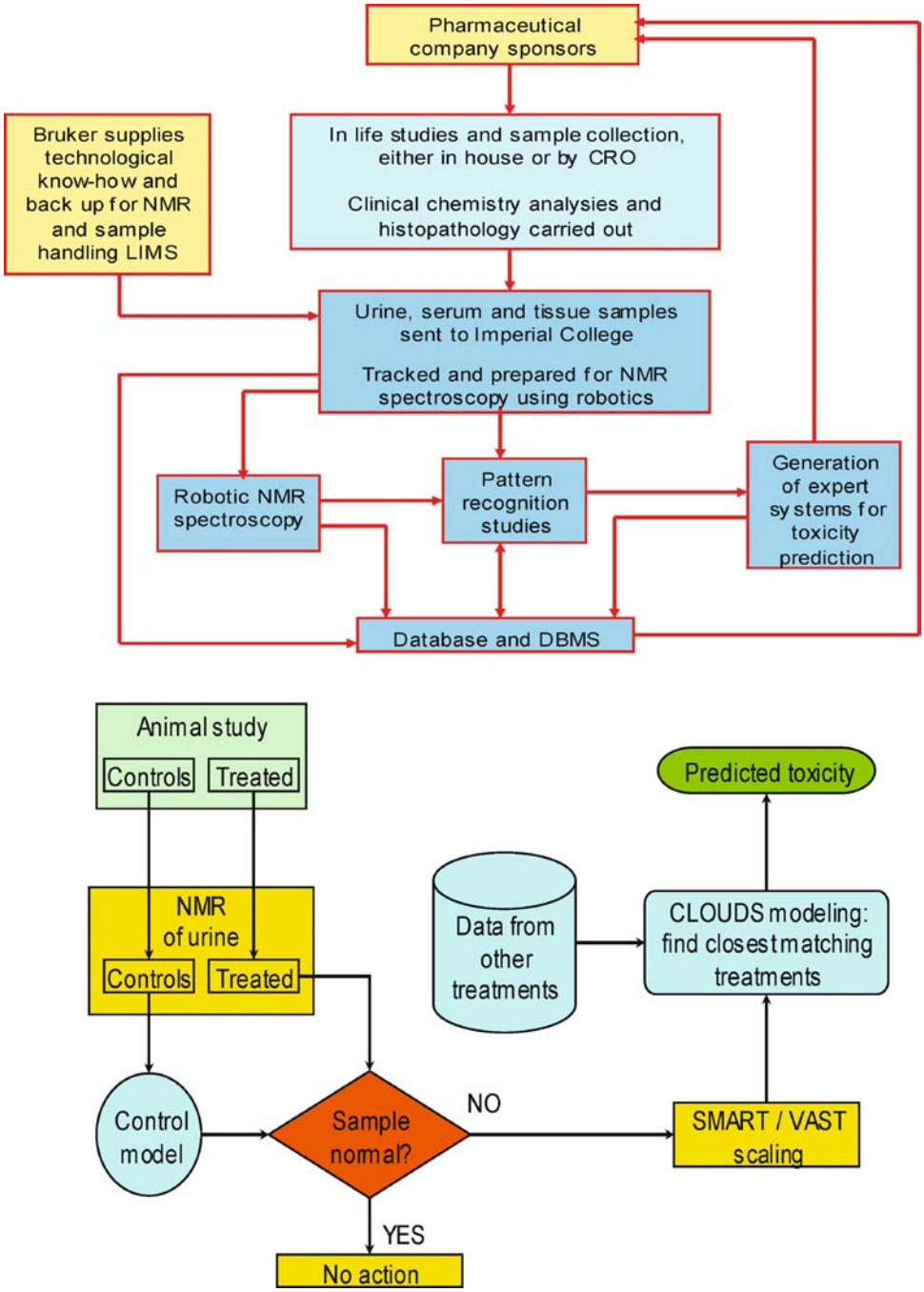


Fig. 4. Top: operational flowchart for the Consortium for Metabonomic Toxicology (COMET) project. CRO, Contract Research Organisation; LIMS, Laboratory Information Management System; DBMS, Database Management System. Bottom: component activities in the COMET expert system operation.

developed. For example, because the predictive expert system that was developed takes into account the metabolic trajectory over time, a new way of comparing and scaling these multivariate trajectories was developed (called SMART) (49). Additionally, a novel classification method for identifying the class of toxicity based on all of the NMR data for a given study has been generated. This has been termed classification of unknowns by density superposition (CLOUDS) and is a novel nonneural implementation of a classification technique developed from probabilistic neural networks (50). The flowchart for the diagnostic expert system is shown in Fig. 4 (bottom).

This consortium showed that it is possible to construct predictive and informative models of toxicity using NMR-based metabonomic data, delineating the whole time course of toxicity. The successful outcome is evidenced by the generated

databases of spectral and conventional results for a wide range of model toxins (147 in total) that served as the basis for computer-based expert systems for toxicity prediction. The project goals of the generation of comprehensive metabonomic databases (now around 35,000 NMR spectra) and successful and robust multivariate statistical models (expert systems) for prediction of toxicity, initially for liver and kidney toxicity in the rat and mouse, have now been achieved, and the predictive systems and databases have been transferred to the sponsoring companies (51). In addition, interesting species differences (rat and mouse) in the toxicity of one compound have been published (52).

Disease Diagnosis and Therapeutic Efficacy

Many examples exist in the literature on the use of NMR-based metabolic profiling to aid human disease diagnosis, such as the use of plasma to study diabetes, CSF for investigating Alzheimer's disease, synovial fluid for osteoarthritis, seminal fluid for male infertility, and urine in the investigation of drug overdose, renal transplantation, and various renal diseases. A promising use of NMR spectroscopy of urine and plasma, as evidenced by the number of publications on the subject, is in the diagnosis of inborn errors of metabolism in children (53). Most of the earlier studies using NMR spectroscopy have been reviewed previously (54).

More recently, CSF sample analysis using NMR spectroscopy has been used to distinguish control subjects from those with meningitis, and the various types of infection (bacterial, viral, and fungal) could also be differentiated (55). In another study, CSF analysis was used to investigate aneurismal subarachnoid hemorrhage, and it was shown that metabolic profiles derived using NMR spectroscopy correlated with vasospasm and clinical outcome (56).

Tissues themselves can be studied by metabonomics through the MAS technique, and published examples include prostate cancer (57), renal cell carcinoma (58), breast cancer (59), and various brain tumors (26,60). A number of mouse models of cardiac disease, including Duchene muscular dystrophy, cardiac arrhythmia, and cardiac hypertrophy, have been investigated using cardiac tissue MAS NMR spectroscopy (61). It was shown that although the mouse strain was a major component of the mouse phenotype, it was possible to discover underlying profiles characteristic of each abnormality.

One area of disease where progress is being made using NMR-based metabonomics studies of biofluids is cancer. This if highlighted by a publication on the diagnosis of epithelial ovarian cancer based on analysis of serum (62).

Metabonomics using NMR spectroscopy has been used to develop a method for diagnosis of coronary artery disease noninvasively through analysis of a blood serum sample (63). Based on angiography, patients were classified into two groups: those with normal coronary arteries and those with triple coronary vessel disease. Around 80% of the NMR spectra were used as a training set to provide a two-class model after data filtering techniques had been applied, and the samples from the two classes were easily distinguished. The remaining 20% of the samples were used as a test set, and their class was then predicted based on the derived model, with a sensitivity of 92% and a specificity of 93%. It was also possible to diagnose the severity of the disease that was present by employing serum samples from patients with stenosis of one, two, or three of the coronary arteries. Although this is a simplistic indicator of disease severity, separation of the three sample classes was evident, although none of the wide range of conventional clinical risk factors that had been measured was significantly different between the classes.

INTEGRATION OF -OMICS RESULTS

The value of obtaining multiple NMR spectroscopic (or indeed other types of analysis) data sets from various biofluid samples and tissues of the same animals collected at different time points has been demonstrated. This procedure has been termed integrated metabonomics (6) and can be used to describe the changes in metabolism in different body compartments affected by exposure to, for example, toxic drugs (64,65). Such timed profiles in multiple compartments are themselves characteristic of particular types and mechanisms of pathology and can be used to give a more complete description of the biochemical consequences than can be obtained from one fluid or tissue alone.

Integration of metabonomics data with that from other multivariate techniques in molecular biology such as from gene array experiments or proteomics is also feasible. Thus, it has also been possible to integrate data from transcriptomics and metabonomics to find, after acetaminophen administration to mice, common metabolic pathways implicated by both gene expression changes and changes in metabolism (66). It has also been found that evaluation of both transcriptomic and metabolic changes following administration of the toxin bromobenzene provides a more sensitive approach for detecting the effects of the toxin (67). In a similar fashion, changes in gene expression detected in microarray experiments can lead to the identification of changed enzyme activity, and this can also be achieved by analysis of metabolic perturbations (68).

The integration of multiple complex data sets over a cohort of samples is a major challenge, but if achieved successfully, this will resent a novel approach to the identification of combination biomarkers of a disease or pharmaceutical effect. An initial approach that has coanalyzed NMR and mass spectra (statistical heterospectroscopy, SHY) from a toxin effect study has recently been published (37).

FUTURE PROSPECTS

It has become accepted that the main pharmaceutical areas where metabonomics is impacting include validation of animal models of disease, including genetically modified animals, preclinical evaluation candidate drugs in safety studies, assessment of safety in humans in clinical trials and after product launch, quantitation, or ranking of the beneficial effects of pharmaceuticals, improved understanding of the causes of highly sporadic idiosyncratic toxicity of marketed drugs, and patient stratification for clinical trials and drug treatment (pharmacometabonomics).

In addition, in terms of disease studies, metabonomics is playing a role in improved, differential diagnosis and prognosis of human diseases, particularly for chronic and degenerative diseases, and for diseases caused by genetic effects. A better understanding of large-scale human population differ-

ences through epidemiological studies is also being achieved.

Other applications include studies on nutrition, sports, medicine, and lifestyle, including the effects of diet, exercise, and stress and evaluation of the effects of interactions between drugs and between drugs and diet.

One of the long-term goals of using pharmacogenomic approaches is to understand the genetic makeup of different individuals (their genetic polymorphisms) and their varying abilities to handle pharmaceuticals, both for their beneficial effects and for identifying adverse effects. If personalized healthcare is to become a reality, an individual's drug treatments must be tailored so as to achieve maximal efficacy and avoid adverse drug reactions. Very recently, an alternative approach to understanding intersubject variability in response to drug treatment using a combination of multivariate metabolic profiling and chemometrics to predict the metabolism and toxicity of a dosed substance, based solely on the analysis and modeling of a predose metabolic profile, has been developed (69). Unlike pharmacogenomics, this approach, which has been termed "pharmacometabonomics," is sensitive to both the genetic and modifying environmental influences that determine the basal metabolic fingerprint of an individual because these will also influence the outcome of a chemical intervention. This new approach has been illustrated with studies of the toxicity and metabolism of compounds with very different modes of action, allyl alcohol, galactosamine, and acetaminophen (paracetamol), administered to rats.

A major initiative has been underway to investigate the reporting needs and to consider recommendations for standardizing reporting arrangements for metabonomics studies, and to this end, a Standard Metabolic Reporting Structures (SMRS) group has been formed (http://www.smrsgroup.org). This has produced a draft policy document that covers all of those aspects of a metabolic study that are recommended for recording, from the origin of a biological sample, the analysis of material from that sample, and chemometric and statistical approaches to retrieve information from the sample data, and a summary publication has appeared (70). The various levels and consequent detail for reporting needs, including journal submissions, public databases, and regulatory submissions, have also been addressed. In parallel, a scheme called ArMet for capturing data and metadata from metabolic studies has been proposed and developed (71). This has been followed up with a workshop and discussion meeting sponsored by the US National Institutes of Health, from which firm plans are being developed to define standards in a number of areas relevant to metabonomics, including characterization of

Strengths	Weaknesses
Robust and stable analytical platforms	Multiple analytical platforms
Excellent analytical/biological reproducibility	Analytical sensitivity
No pre-selection of analytes	Analytical dynamic range
Minimally invasive	Complexity of data sets
Exploratory studies possible	Over-fitting of data possible
Real biological end points	No current standardisation of methods
Whole system integration	Regulatory body training needed
Multi-genomic	High capital cost
Cross-species biomarkers	
Low cost per sample/analyte	
Opportunities	**Threats**
Use of marker species in environmental studies	Scepticism of non-hypothesis led studies
Much experience from mammalian system studies (e.g. pathways)	Conservatism (one significant analyte per test approach)
Potential of multi-omics integration	Lack of well trained scientists
Benefits of a central laboratory approach	Disadvantages of a central laboratory approach
Web-based diagnostics	

Fig. 5. Strengths, weaknesses, opportunities, and threats (SWOT) analysis of metabonomics.

sample-related metadata, technical standards and related data, metadata and QC matters for the analytical instrumentation, data transfer methodologies and schema for implementation of such activities, and development of standard vocabularies to enable transparent exchange of data (72).

NMR- and MS-based metabonomics are now recognized as independent and widely used techniques for evaluating the toxicity of drug candidate compounds, and it has been adopted by a number of pharmaceutical companies into their drug development protocols. For drug safety studies, it is possible to identify the target organ of toxicity, derive the biochemical mechanism of the toxicity, and determine the combination of biochemical biomarkers for the onset, progression, and regression of the lesion. Additionally, the technique has been shown to be able to provide a metabolic fingerprint of an organism (metabotyping) as an adjunct to functional genomics and hence has applications in design of drug clinical trials and for evaluation of genetically modified animals as disease models.

Using metabonomics, it has proved possible to derive new biochemically based assays for disease diagnosis and to identify combination biomarkers for disease, which can then be used to monitor the efficacy of drugs in clinical trials. Thus, based on differences observed in metabonomic databases from control animals and from animal models of disease, diagnostic methods and biomarker combinations might be derivable in a preclinical setting. Similarly, the use of databases to derive predictive expert systems for human disease diagnosis and the effects of therapy require compilations from both normal human populations and patients before, during, and after therapy.

In summary, it is clear that metabonomics will have an impact in pharmaceutical Research & Development, and Fig. 5 summarizes a strengths, weaknesses, opportunities, and threats (SWOT) analysis of the discipline. The analytical procedures used are stable and robust and have a high degree of reproducibility, and although advances will obviously be made in the future, current data will always be readable and interpretable. In contrast to other -omics, metabonomics enjoys a good level of biological reproducibility, and the cost per sample and per analyte is low. It has the advantage of not having to preselect analytes, and through use of biofluids, it is minimally invasive, with hypothesis generation studies being easily possible. Metabolic biomarkers are closely identifiable with real biological endpoints and provide a global systems interpretation of biological effects, including the interactions between multiple genomes such as humans and their gut microflora. One major potential strength of metabonomics is the possibility that metabolic biomarkers will be more easily used across species than transcriptomic or proteomic biomarkers, and this should be important for pharmaceutical studies. On the other hand, metabonomics suffers from the use of multiple analytical technologies, there are questions of the sensitivity and dynamic range of the technologies used, and the data sets are complex. Using chemometrics, it is possible to overinterpret the data, but this is easily avoided by correct statistical rigor. At present, the groups using metabonomics are moving toward defining standards for data and operations, and a good start has been made, but there remains a need for the regulatory agencies to be trained in the data interpretation and for more well trained practitioners.

There is an inherent conservatism that would like to be able to use a single biomarker or analyte for each diagnostic test. However, the reality of the complexity of disease and drug effects means that biomarker combinations will be more usual, and thus, there will be many opportunities for metabonomics that are as yet underexplored, such as its use in environmental toxicity studies, in directing the timing of transcriptomic and proteomic experiments, and for deriving theranostic biomarkers. It will surely be an integral part of any multiomics study where all the data sets are combined to derive an optimum set of biomarkers.

The ultimate goal of systems biology must be the integration of data acquired from living organisms at the genomic, protein, and metabolite levels. In this respect, transcriptomics, proteomics, and metabonomics will all play an important role. Through the combination of these and related approaches will come an improved understanding of an organism's total biology and, with this, better understanding of the causes and progression of human diseases and, given the 21st century goal of personalized healthcare, the improved design and development of new and better targeted pharmaceuticals.

REFERENCES

1. P. Baldi and G. W. Hatfield. *DNA Microarrays and Gene Expression*, Cambridge University Press, Cambridge, UK. ISBN 0521800226, 2002.
2. J. K. Nicholson and I. D. Wilson. High-resolution proton magnetic resonance spectroscopy of biological fluids. *Prog. Nucl. Magn. Reson. Spectrosc.* **21**:449–501 (1989).
3. K. P. R. Gartland, S. M. Sanins, J. K. Nicholson, B. C. Sweatman, C. R. Beddell, and J. C. Lindon. Pattern recognition analysis of high resolution ^{1}H NMR spectra of urine. A nonlinear mapping approach to the classification of toxicological data. *NMR Biomed.* **3**:166–172 (1990).
4. K. P. R. Gartland, C. R. Beddell, J. C. Lindon, and J. K. Nicholson. The application of pattern recognition methods to the analysis and classification of toxicological data derived from proton NMR spectroscopy of urine. *Mol. Pharmacol.* **39**:629–642 (1991).
5. J. K. Nicholson, J. C. Lindon E. Holmes. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29**:1181–1189 (1999).
6. J. K. Nicholson, J. Connelly, J. C. Lindon, and E. Holmes. Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* **1**:153–162 (2002).
7. O. Fiehn. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**:155–171 (2002).
8. J. K. Nicholson and I. D. Wilson. Understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat. Rev. Drug Discov.* **2**:668–676 (2003).
9. J. K. Nicholson, E. Holmes, J. C. Lindon, and I. D. Wilson. The challenges of modeling mammalian biocomplexity. *Nat. Biotechnol.* **22**:1268–1274 (2004).
10. M. Mayr, Y. L. Chung, U. Mayr, X. K. Yin, L. Ly, H. Troy, S. Fredericks, Y. H. Yu, J. R. Griffiths, and Q. B. Xu. Proteomic and metabolomic analyses of atherosclerotic vessels from apolipoprotein E-deficient mice reveal alterations in inflammation, oxidative stress, and energy metabolism. *Arterioscler. Thromb. Vasc. Biol.* **25**:2135–2142 (2005).
11. R. J. A. N. Lamers, E. C. H. H. Wessels, J. J. M. van der Sandt, K. Venema, G. Schaafsma, J. van der Greef, and H. J. J. van Nesselrooij. A pilot study to investigate effects of inulin on Caco-2 cells through *in vitro* metabolic fingerprinting. *J. Nutr.* **133**:3080–3084 (2003).

12. S. G. Villas-Boas, J. Hojer-Pedersen, M. Akesson, J. Smedsgaard, and J. Nielsen. Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* **22**:1155–1169 (2005).
13. J. E. Ippolito, J. Xu, S. J. Jain, K. Moulder, S. Mennerick, J. R. Crowley, R. R. Townsend, and J. I. Gordon. An integrated functional genomics and metabolomics approach for defining poor prognosis in human neuroendocrine cancers. *Proc. Natl. Acad. Sci. USA* **102**:9901–9906 (2005).
14. M. E. Bollard, J. S. Xu, W. Purcell, J. L. Griffin, C. Quirk, E. Holmes, and J. K. Nicholson. Metabolic profiling of the effects of D-galactosamine in liver spheroids using ^{1}H NMR and MAS NMR spectroscopy. *Chem. Res. Toxicol.* **15**:1351–1359 (2002).
15. N. N. Kaderbhai, D. I. Broadhurst, D. I. Ellis, and D. B. Kell. Functional genomics via metabolic footprinting: monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT-IR and direct injection electrospray mass spectrometry. *Compar. Funct. Genom.* **4**:376–391 (2003).
16. P. H. Gamache, D. F. Meyer, M. C. Granger, and I. N. Acworth. Metabolomic applications of electrochemistry/mass spectrometry. *J. Am. Soc. Mass Spectrom.* **15**:1717–1726 (2004).
17. J. C. Lindon, J. K. Nicholson, E. Holmes, and J. R. Everett. Metabonomics: metabolic processes studied by NMR spectroscopy of biofluids. *Concepts Magn. Reson.* **12**:289–320 (2000).
18. J. C. Lindon, E. Holmes, M. E. Bollard, E. G. Stanley, and J. K. Nicholson. Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers* **9**:1–31 (2004).
19. T. D. W. Claridge. *High-Resolution NMR Techniques in Organic Chemistry*, Elsevier, Oxford, UK. ISBN 0080427995, 1999.
20. M. Liu, J. K. Nicholson, and J. C. Lindon. High resolution diffusion and relaxation edited one- and two-dimensional ^{1}H NMR spectroscopy of biological fluids. *Anal. Chem.* **68**: 3370–3376 (1996).
21. J. K. Nicholson, P. J. D. Foxall, M. Spraul, R. D. Farrant, and J. C. Lindon. 750 MHz ^{1}H and ^{1}H–^{13}C NMR spectroscopy of human blood plasma. *Anal. Chem.* **67**:793–811 (1995).
22. H. C. Keun, O. Beckonert, J. L. Griffin, C. Richter, D. Moskau, J. C. Lindon, and J. K. Nicholson. Cryogenic probe ^{13}C NMR spectroscopy of urine for metabonomic studies. *Anal. Chem.* **74**:4588–4593 (2002).
23. K. E. Price, S. S. Vandaveer, C. E. Lunte, and C. K. Larive. Tissue targeted metabonomics: metabolic profiling by microdialysis sampling and microcoil NMR. *J. Pharm. Biomed. Anal.* **38**:904–909 (2005).
24. A. Tomlins, P. J. D. Foxall, J C. Lindon, J. M. Lynch, M. Spraul, J. R. Everett, and J. K. Nicholson. High resolution magic angle spinning ^{1}H nuclear magnetic resonance analysis of intact prostatic hyperplastic and tumour tissues. *Anal. Commun.* **35**: 113–115 (1998).
25. S. L. Garrod, E. Humpfer, M. Spraul, S. C. Connor, S. Polley, J. Connelly, J. C. Lindon, J. K. Nicholson, and E. Holmes. High-resolution magic angle spinning ^{1}H NMR spectroscopic studies on intact rat renal cortex and medulla. *Magn. Reson. Med.* **41**:1108–1118 (1999).
26. L. L. Cheng, I. W. Chang, D. N. Louis, and R. G. Gonzalez. Correlation of high-resolution magic angle spinning proton magnetic resonance spectroscopy with histopathology of intact human brain tumor specimens. *Cancer Res.* **58**:1825–1832 (1998).
27. M. Morris and S. M. Watkins. Focused metabolomic profiling in the drug development process: advances from lipid profiling. *Curr. Opin. Chem. Biol.* **9**:407–412 (2005).
28. I. D. Wilson, J. K. Nicholson, J. Castro-Perez, J. H. Granger, K. A. Johnson, B. W. Smith, and R. S. Plumb. High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *J. Proteome Res.* **4**:591–598 (2005).
29. T. Soga, Y. Ohashi, Y. Ueno, H. Nasaoka, M. Tomita, and T. Nishioka. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J. Proteome Res.* **2**:488–494 (2003).
30. J. C. Lindon, J. K. Nicholson, and I. D. Wilson. Directly-coupled HPLC-NMR and HPLC-NMR-MS in pharmaceutical research and development. *J. Chromatogr., B, Biomed. Sci. Appl.* **748**: 233–258 (2000).
31. J. C. Lindon, E. Holmes, and J. K. Nicholson. Pattern recognition methods and applications in biomedical magnetic resonance. *Prog. Nucl. Magn. Reson. Spectrosc.* **39**:1–40 (2001).
32. H. Wold. Partial least squares. In S. Kotz and N. L. Johnson (eds.), *Encyclopedia of Statistical Sciences,* Vol. 6, Wiley, New York, 1985, pp. 581–591.
33. J. Trygg and S. Wold. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **16**:119–128 (2002).
34. E. Holmes, J. K. Nicholson, and G. Tranter. Metabonomic characterization of genetic variations in toxicological and metabolic responses using probabilistic neural networks. *Chem. Res. Toxicol.* **14**:182–191 (2001).
35. D. B. Kell. Metabolomics and machine learning: explanatory analysis of complex metabolome data using genetic programming to produce simple, robust rules. *Mol. Biol. Rep.* **29**:237–241 (2002).
36. O. Cloarec, M. E. Dumas, A. Craig, R. H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J. C. Lindon, E. Holmes, and J. Nicholson. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic ^{1}H NMR data sets. *Anal. Chem.* **77**:1282–1289 (2005).
37. D. J. Crockford, E. Holmes, J. C. Lindon, R. S. Plumb, S. Zirah, S. J. Bruce, P. Rainville, C. L. Stumpf, and J. K. Nicholson. Statistical heterospectroscopy, (SHY), an approach to the integrated analysis of NMR and UPLC-MS data sets: application in metabonomic toxicology studies. *Anal. Chem.* **78**(2):363–371 (2006).
38. M. E. Bollard, E. G. Stanley, J. C. Lindon, J. K. Nicholson, and E. Holmes. NMR-based metabonomics approaches for evaluating physiological influences on biofluid composition. *NMR Biomed.* **18**:143–162 (2005).
39. R. S. Plumb, J. H. Granger, C. L. Stumpf, K. A. Johnson, B. W. Smith, S. Gaulitz, I. D. Wilson, and J. Castro-Perez. A rapid screening approach to metabonomics using UPLC and oa-TOF mass spectrometry: application to age, gender and diurnal variation in normal/Zucker obese rats and black, white and nude mice. *Analyst* **130**:844–849 (2005).
40. J. B. German, M. A. Roberts, and S. M. Watkins. Genomics and metabolomics as markers for the interaction of diet and health: lessons from lipids. *J. Nutr.* **133**:2078S–2083S (2003).
41. K. K. Lehtimaki, P. K. Valonen, J. L. Griffin, T. H. Vaisanen, O. H. J. Grohn, M. I. Kettunen, J. Vepsalainen, S. Yla-Herttuala, J. Nicholson, and R. A. Kauppinen. Metabolite changes in BT4C rat gliomas undergoing ganciclovir-thymidine kinase gene therapy-induced programmed cell death as studied by H-1 NMR spectroscopy *in vivo*, *ex vivo*, and *in vitro*. *J. Biol. Chem.* **278**: 45915–45923 (2003).
42. A. W. Nicholls, R. J. Mortishire-Smith, and J. K. Nicholson. NMR spectroscopic based metabonomic studies of urinary metabolite variation in acclimatizing germ-free rats. *Chem. Res. Toxicol.* **16**:1395–1404 (2003).
43. L. C. Robosky, D. F. Wells, L. A. Egnash, M. L. Manning, M. D. Reily, and D. G. Robertson. Metabonomic identification of two distinct phenotypes in Sprague-Dawley (Crl:CD(SD)) rats. *Toxicol. Sci.* **87**:277–284 (2005).
44. Y. Wang, E. Holmes, J. K. Nicholson, O. Cloarec, J. Chollet, M. Tanner, B. H. Singer, and J. Utzinger. Metabonomic investigations in mice infected with *Schistosoma mansoni*: an approach for biomarker identification. *Proc. Natl. Acad. Sci. USA* **101**:12676–12681 (2004).
45. J. C. Lindon, E. Holmes, and J. K. Nicholson. Toxicological applications of magnetic resonance. *Prog. Nucl. Magn. Reson. Spectrosc.* **45**:109–143 (2004).
46. E. M. Lenz, J. Bright, R. Knight, F. R. Westwood, D. Davies, H. Major, and I. D. Wilson. Metabonomics with 1H-NMR spectroscopy and liquid chromatography–mass spectrometry applied to the investigation of metabolic changes caused by gentamycin-induced nephrotoxicity in the rat. *Biomarkers* **10**:173–187 (2005).
47. J. C. Lindon, J. K. Nicholson, E. Holmes, H. Antti, M. E. Bollard, H. Keun, O. Beckonert, T. M. Ebbels, M. D. Reily, D. Robertson,

G. J. Stevens, P. Luke, A. P. Breau, G. H. Cantor, R. H. Bible, U. Niederhauser, H. Senn, G. Schlotterbeck, U. G. Sidelmann, S. M. Laursen, A. Tymiak, B. D. Car, L. Lehman-McKeeman, J. M. Colet, A. Loukaci, and C. Thomas. Contemporary issues in toxicology: the role of metabonomics in toxicology and its evaluation by the COMET project. *Toxicol. Appl. Pharmacol.* **187**: 137–146 (2003).

48. H. C. Keun, T. M. D. Ebbels, H. Antti, M. Bollard, O. Beckonert, G. Schlotterbeck, H. Senn, U. Niederhauser, E. Holmes, J. C. Lindon, and J. K. Nicholson. Analytical reproducibility in ^{1}H NMR-based metabonomic urinalysis. *Chem. Res. Toxicol.* **15**:1380–1386 (2002).
49. H. C. Keun, T. M. D. Ebbels, M. E. Bollard, O. Beckonert, H. Antti, E. Holmes, J. C. Lindon, and J. K. Nicholson. Geometric trajectory analysis of metabolic responses to toxicity can define treatment specific profiles. *Chem. Res. Toxicol.* **17**:579–587 (2004).
50. T. Ebbels, H. Keun, O. Beckonert, H. Antti, M. Bollard, E. Holmes, J. Lindon, and J. Nicholson. Toxicity classification from metabonomic data using a density superposition approach: 'CLOUDS'. *Anal. Chim. Acta* **490**:109–122 (2003).
51. J. C. Lindon, H. C. Keun, T. M. D. Ebbels, J. M. T. Pearce, E. Holmes, and J. K. Nicholson. The Consortium for Metabonomic Toxicology (COMET): aims, activities and achievements. *Pharmacogenomics* **6**:691–699 (2005).
52. M. E. Bollard, H. C. Keun, O. Beckonert, T. M. D. Ebbels, H. Antti, A. W. Nicholls, J. P. Shockcor, G. H. Cantor, G. Stevens, J. C. Lindon, E. Holmes, and J. K. Nicholson. Comparative metabonomics of differential hydrazine toxicity in the rat and mouse. *Toxicol. Appl. Pharmacol.* **204**:135–151 (2005).
53. S. H. Moolenaar, U. F. H. Engelke, and R. A. Wevers. Proton nuclear magnetic resonance spectroscopy of body fluids in the field of inborn errors of metabolism. *Ann. Clin. Biochem.* **40**:16–24 (2003).
54. J. C. Lindon, J. K. Nicholson, and J. R. Everett. NMR spectroscopy of biofluids. In G. A. Webb (ed.), *Annual Reports on NMR Spectroscopy,* Vol. 38, Academic Press, Oxford, UK, 1999, pp 1–88.
55. M. Coen, M. O'Sullivan, W. A. Bubb, P. W. Kuchel, and T. Sorrell. Proton nuclear magnetic resonance-based metabonomics for rapid diagnosis of meningitis and ventriculitis. *Clin. Infect. Dis.* **41**:1582–1590 (2005).
56. V. G. Dunne, S. Bhattachayya, M. Besser, C. Rae, and J. L. Griffin. Metabolites from cerebrospinal fluid in aneurysmal subarachnoid haemorrhage correlate with vasospasm and clinical outcome: a pattern-recognition H-1 NMR study. *NMR Biomed.* **18**:24–33 (2005).
57. M. G. Swanson, D. B. Vigneron, Z. L. Tabatabai, R. G. Males, L. Schmitt, P. R. Carroll, J. K. James, R. E. Hurd, and J. Kurhanewicz. Proton HR-MAS spectroscopy and quantitative pathologic analysis of MRI/3D-MRSI-targeted postsurgical prostate tissues. *Magn. Reson. Med.* **50**:944 – 954 (2003).
58. D. Moka, R. Vorreuther, H. Schicha, M. Spraul, E. Humpfer, M. Lipinski, P. J. D. Foxall, J. K. Nicholson, and J. C. Lindon. Biochemical classification of kidney carcinoma biopsy samples using magic-angle-spinning ^{1}H nuclear magnetic resonance spectroscopy. *J. Pharm. Biomed. Anal.* **17**:125–132 (1998).
59. B. Sitter, U. Sonnewald, M. Spraul, H. E. Fjosne, and I. S. Gribbestad. High-resolution magic angle spinning MRS of breast cancer tissue. *NMR Biomed.* **15**:327–337 (2002).
60. S. J. Barton, F. A. Howe, A. M. Tomlins, S. A. Cudlip, J. K. Nicholson, B. A. Bell, and J. R. Griffiths. Comparison of *in vivo* ^{1}H MRS of human brain tumors with ^{1}H HR-MAS spectroscopy of intact biopsy samples *in vitro*. *Magma* **8**:121–128 (1999).
61. G. L. A. H. Jones, E. Sang, C. Goddard, R. J. Mortishire-Smith, B. C. Sweatman, J. N. Haselden, K. Davies, A. A. Grace, K. Clarke, and J. L. Griffin. A functional analysis of mouse models of cardiac disease through metabolic profiling. *J. Biol. Chem.* **280**:7530–7539 (2005).
62. K. Odunsi, R. M. Wollman, C. B. Ambrosone, A. Hutson, S. E. McCann, J. Tammela, J. P. Geisler, G. Miller, T. Sellers, W. Cliby, F. Qian, B. Keitz, M. Intengan, S. Lele, and J. L. Alderfer. Detection of epithelial ovarian cancer using H-1-NMR-based metabonomics. *Int. J. Cancer* **113**:782–788 (2005).
63. 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, and D. J. Grainger. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using H-1 NMR-based metabonomics. *Nat. Med.* **8**:1439–1445 (2002).
64. N. J. Waters, E. Holmes, A. Williams, C. J. Waterfield, R. D. Farrant, and J. K. Nicholson. NMR and pattern recognition studies on the time-related metabolic effects of α-naphthylisothiocyanate on liver, urine, and plasma in the rat: an integrative metabonomic approach. *Chem. Res. Toxicol.* **14**:1401–1412 (2001).
65. M. Coen, E. M. Lenz, J. K. Nicholson, I. D. Wilson, F. Pognan, and J. C. Lindon. An integrated metabonomic investigation of acetaminophen toxicity in the mouse using NMR spectroscopy. *Chem. Res. Toxicol.* **16**:295–303 (2003).
66. M. Coen, S. U. Ruepp, J. C. Lindon, J. K. Nicholson, F. Pognan, E. M. Lenz, and I. D. Wilson. Integrated application of transcriptomics and metabonomics yields new insight into the toxicity due to paracetamol in the mouse. *J. Pharm. Biomed. Anal.* **35**:93–105 (2004).
67. W. H. M. Heijne, R. J. A. N. Lamers, P. J. van Bladeren, J. P. Groten, J. H. J. van Nesselrooij, and B. Van Ommen. Profiles of metabolites and gene expression in rats with chemically induced hepatic necrosis. *Toxicol. Pathol.* **33**:425–433 (2005).
68. J. L. Griffin, S. A. Bonney, C. Mann, A. M. Hebbachi, G. F. Gibbons, J. K. Nicholson, C. C. Shoulders, and J. Scott. An integrated reverse functional genomic and metabolic approach to understanding orotic acid-induced fatty liver. *Physiol. Genomics* **17**:140–149 (2004).
69. T. A. Clayton, J. C. Lindon, H. Antti, C. Charuel, G. Hanton, J.-P. Provost, J.-L. Le Net, D. Baker, R. J. Walley, J. R. Everett, and J. K. Nicholson. Pharmaco-metabonomic phenotyping and personalised drug treatment. *Nature* in press. (2006).
70. J. C. Lindon, J. K. Nicholson, E. Holmes, H. C. Keun, A. Craig, J. T. M. Pearce, S. J. Bruce, N. Hardy, S.-A. Sansone, H. Antti, P. Jonsson, C. Daykin, M. Navarange, R. D. Beger, E. R. Verheij, A. Amberg, D. Baunsgaard, G. H. Cantor, L. Lehman-McKeeman, M. Earll, S. Wold, E. Johansson, J. N. Haselden, K. Kramer, C. Thomas, J. Lindberg, I. Schuppe-Koistinen, I. D. Wilson, M. D. Reily, D. G. Robertson, H. Senn, A. Krotzky, S. Kochhar, J. Powell, F. van der Ouderaa, R. Plumb, H. Schaefer, and M. Spraul. Summary recommendations for standardization and reporting of metabolic analyses. *Nat. Biotechnol.* **23**:833–838 (2005).
71. H. Jenkins, N. Hardy, M. Beckmann, J. Draper, A. R. Smith, J. Taylor, O. Fiehn, R. Goodacre, R. J. Bino, R. Hall, J. Kopka, G. A. Lane, B. M. Lange, J. R. Liu, P. Mendes, B. J. Nikolau, S. G. Oliver, N. W. Paton, S. Rhee, U. Roessner-Tunali, K. Saito, J. Smedsgaard, L. W. Sumner, T. Wang, S. Walsh, E. S. Wurtele, and D. B. Kell. A proposed framework for the description of plant metabolomics experiments and their results. *Nat. Biotechnol.* **22**:1601–1606 (2004).
72. A. L. Castle, O. Fiehn, R. Kaddurah-Daouk, and J. C. Lindon. Metabolomics standards workshop and the development of international standards for reporting metabolomics experimental results. *Brief. Bioinform.* (2006), in press.