Metabolic profiles to define the genome: can we hear the phenotypes?

Julian L. Griffin

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK
(jlg40@mole.bio.cam.ac.uk)

There is an increased reliance on genetically modified organisms as a functional genomic tool to elucidate the role of genes and their protein products. Despite this, many models do not express the expected phenotype thought to be associated with the gene or protein. There is thus an increased need to further define the phenotype resultant from a genetic modification to understand how the transcriptional or proteomic network may conspire to alter the expected phenotype. This is best typified by the description of the silent phenotype in genetic manipulations of yeast. High-resolution proton nuclear magnetic resonance (^1H NMR) spectroscopy provides an ideal mechanism for the profiling of metabolites within biofluids, tissue extracts or, with recent advances, intact tissues. These metabolic datasets can be readily mined using a range of pattern recognition techniques, including hierarchical cluster analysis, principal components analysis, partial least squares and neural networks, with the combined approach being termed metabolomics. This review describes the application of NMR-based metabolomics or metabonomics to genetic and chemical interventions in a number of different species, demonstrating the versatility of such an approach, as well as suggesting how it may be integrated with other 'omic' technologies.

Keywords: metabolomics; metabonomics; magic-angle spinning NMR spectroscopy; phenotyping; Duchenne muscular dystrophy

1. INTRODUCTION

In the post-genomic era, as the number of species with sequenced genomes continues to rise, there has been greater awareness of functional genomic technologies (Evans 2000). This has also been driven by increased availability of technologies for global transcriptional and proteomic profiling, as well as the proliferation of techniques for the production of genetically modified organisms (Klose *et al.* 2002; Schoolnik 2002). However, many genes are not under transcriptional control, suggesting that an approach based solely on transcriptomics is inadequate, while the proteome of an organism remains an immense technical challenge. Furthermore, with the ever burgeoning number of modifications to be studied, there is also increasing awareness that further functional genomic technologies are required for the definition of an organism's phenotype.

A perfect example of the problem was described by Raamsdonk *et al.* (2001), investigating genetic modifications in yeast. The classical way to phenotype this organism is to study the rate of growth of a given strain on different substrates. However, in this study of several yeast mutants, many strains produced similar growth profiles despite the deletion of genes with very different metabolic functions. The term 'silent mutation' has been coined to describe such strains where a deletion produces no obvious effect in terms of growth of the organism. However, using proton nuclear magnetic resonance (^1H NMR) spectroscopy to study the metabolic changes induced in the different yeast strains, it was apparent that

metabolic perturbations within the yeast conspired to produce similar growth curves despite the different genetic deletions. Furthermore, these metabolic perturbations could be used to classify strains correctly, clustering mutants that arose from similar deletions together. This suggests such a process of defining a phenotype through the global changes induced in metabolism could be used to predict the function of genes deleted or upregulated in a given system.

Using techniques developed for monitoring organ toxicity in a non-invasive manner through biofluid NMR spectroscopy (Beckwith Hall *et al.* 1998; Holmes *et al.* 1998; Nicholson *et al.* 2002), it is similarly possible to derive a metabolic profile either from a biofluid, a tissue extract, the intact tissue or the whole organism itself following a genetic modification. As well as NMR spectroscopy, mass spectrometry and high-pressure liquid chromatography (HPLC) have been applied to detect low molecular weight metabolites in a range of animals and plants (Goodacre *et al.* 2000; Fiehn 2002), often in conjunction with pattern recognition algorithms to decipher the most significant changes that accompany a given modification, pathology or biological process.

This review provides an overview of the field, briefly examining how comprehensive analytical tools are being used to derive metabolic profiles for the use in functional genomics. Following this overview three specific applications of NMR-based analysis of mammalian metabolism will be considered to demonstrate how these approaches are implemented. In particular, tools for the direct analysis of intact tissues, such as magic-angle spinning (MAS), will

be considered and how these may help phenotype the ever increasing number of modified organisms.

2. TERMINOLOGY

Several terms have been coined to describe the process of combining global analytical tools and pattern recognition analysis to define the metabolic phenotype of a tissue or organism. Oliver (1996, 2002) defines the metabolome as the complete set of metabolites and/or low molecular weight intermediates which is context dependent, varying according to the physiology, developmental or pathological state of the cell, tissue, organ or organism. This places the definition alongside the transcriptome and proteome, which represent the complete transcriptional profile and protein expression in a cell, tissue or organism during a given process. Furthermore, unlike genomics, all these 'omes' are context dependent, complicating any such attempts to define an organism's complete metabolome, as by definition this would also have to include all the influences of pathology, development, physiology and the environment.

However, Nicholson and co-workers have produced definitions for both metabolomics and metabonomics (Nicholson *et al.* 1999; Nicholson & Wilson 2003). They define metabolomics as 'the measurement of metabolite concentrations and fluxes and secretion in cells and tissues in which there is a direct connection between the genetic activity, protein activity and the metabolic activity itself' (Nicholson & Wilson 2003, p. 674). By contrast metabonomics is 'the quantitative measurement of the multivariate metabolic responses of multicellular systems to pathophysiological stimuli or genetic modification' (Nicholson & Wilson 2003, p. 674). These two definitions reflect the extra tier of complexity added by moving from unicellular to multicellular organisms.

Fiehn (2001, 2003) however, considers the definitions in part through the technologies they use. He suggests that the subject area needs three definitions for metabolomics, metabolic profiling and metabolic fingerprinting, stating that 'metabolomics has been defined ... as the analysis of the total complement of individual metabolites of a given biological sample ... in contrast to metabolic fingerprinting techniques which do not aim at analysing individual components' (Fiehn 2003, p. 876). Thus, metabolic profiling would involve techniques such as gas chromatography mass spectrometry (GC-MS) and liquid chromatography NMR (LC-NMR) where individual components are separated by one technique and identified by another, while NMR spectroscopy and direct infusion electrospray mass spectrometry (ESI-MS) would be metabolic fingerprinting tools.

In this manuscript I have attempted to use the terms in a context-dependent manner, using the terms that the researchers who actually carried out the work would use themselves. Personally, the arbitrary distinctions between these words often appear contradictory and unhelpful, especially in light of the inclusive nature of the fields of transcriptomics and proteomics. Furthermore, until a set of definitions under a global heading can be agreed upon the different areas of the field may not benefit from the work being carried out by fellow workers using a different set of definitions.

3. AN OVERVIEW OF METABOLOMICS

The concept of understanding disease processes through metabolic profiling is not a new idea, in much the same way that transcriptomics and proteomics through Northern and Western blotting pre-date the use of microarrays and proteomics using mass spectrometry and two-dimensional gel electrophoresis. For example, NMR spectroscopy has been used extensively to profile ^{31}P , ^1H and ^{13}C containing metabolites in a range of organisms including humans. Among one of the most impressive uses of ^1H NMR spectroscopy was the detection of *ca.* 50 metabolites in lymphocytes by Sze & Jardetzky (1994). For a comprehensive review of the use of NMR spectroscopy in metabolic profiling the reader is directed to Fan's (1996) review of the subject. Similarly, GC-MS has also been widely used as a metabolic profiling tool since its use to study steroids, amino acids and drug metabolites in the early 1970s (Devaux *et al.* 1971; Horning & Horning 1971).

However, one aspect of metabolomics that distinguishes it from previous studies of metabolism is the attempt to measure all global metabolites simultaneously. To capitalize on the wealth of information found in these datasets, multivariate pattern recognition techniques are also required so that normal metabolic variation can be separated from that induced by a disease process, a genetic manipulation or a drug intervention. The technology for these challenges has been developed from several different applications and these will be considered briefly in turn here.

4. LISTENING TO SILENT PHENOTYPES IN YEAST

Yeast (*Saccharomyces cerevisiae*) is the most widely used organism for modelling eukaryotic functional genomics as a whole, being the first eukaryote to be fully sequenced (Goffeau *et al.* 1996). Furthermore, mutant strains are now available for the *ca.* 6000 potential protein-encoding genes of yeast (Winzeler *et al.* 1999), suggesting that the entire genome could be phenotyped on a gene-by-gene basis. However, an approach based solely on growth-rate competition analysis is inadequate because a network of genes can conspire to mask the effect of the deletion of one gene. To phenotype such 'silent' strains an approach dubbed functional analysis by co-responses in yeast (FANCY) has been developed (Oliver 1996; Raamsdonk *et al.* 2001). Using ^1H NMR spectroscopy (Raamsdonk *et al.* 2001) or ESI-MS (Castrillo *et al.* 2003) to study the metabolic changes induced in the different yeast strains, it was apparent that these profiles could be used to classify strains, clustering mutants that arose from similar deletions together. Figure 1 demonstrates the metabolic information obtainable using NMR spectroscopy which can be used in such approaches. This suggests that such a process of defining a phenotype through the global changes induced in metabolism may be used to predict the function of genes deleted or upregulated in a given system through comparative metabolomics. Indeed, Raamsdonk and colleagues were able to show that mutants involving the separate deletion of two genes encoding the same enzyme, 6-phosphofructo-2-kinase, produced the same metabolic phenotype, and that deletions

Figure 1. High-resolution MAS ^1H NMR spectrum of intact wild-type yeast cells acquired at 600 MHz and a spinning rate of 5000 Hz. Thirty-five metabolites were identified through chemical shift and coupling patterns with reference to metabolites found in mammalian cells. 1, Lipid (CH_3CH_2-); 2, short-chain fatty acids; 3, leucine; 4, isoleucine; 5, valine; 6, lipid (CH_2CH_2); 7, ethanol; 8, alanine; 9, lipid ($\text{CH}_2\text{CH}_2\text{CO}$); 10, leucine; 11, lysine; 12, acetate; 13, glutamate; 14, glutamine; 15, lipid ($\text{CH}_2\text{CH}=\text{CH}$); 16, lipid (COCH_2); 17, lipid ($=\text{CH}_2\text{CH}_2\text{CH}=\text{CH}$); 18, aspartate; 19, citrate; 20, creatine; 21, tyrosine, leucine and cysteine; 22, choline; 23, taurine; 24, glucose; 25, *myo*-inositol; 26, aliphatic amino acids; 27, lipids ($\text{CH}=\text{CH}$); 28, orotic acid; 29, adenosine; 30, tyrosine; 31, tryptophan; 32, phenylalanine; 33, uridine triphosphate; 34, adenosine monophosphate; 35, adenosine triphosphate.

involved in similar pathways, such as oxidative phosphorylation, also clustered together.

Potentially one of the most lucrative areas of metabolomics is in the field of metabolic engineering of microbial populations. Buchholz *et al.* (2002) have described a process using substrate pulses in microbial fermentation and following the perturbations using HPLC, LC-MS and enzymatic assays to define the metabolic dynamics. In turn this may be used to identify key sites for modification to upregulate the production of pharmaceutically important metabolites.

The analysis does not have to be confined to cell extracts. Allen *et al.* (2003) have shown using ESI-MS that metabolites excreted into the culture media can be used to distinguish mutant strains of yeast, dubbing the approach 'metabolic footprinting' in contrast to metabolic fingerprinting. Furthermore, by examining the media directly a high throughput is possible of the order of *ca.* 2 min per sample, with the advantage that this merely samples the media, allowing dynamic studies. With the availability of *ca.* 6000 mutant strains of yeast, this area is set to expand significantly, providing functional genomics with the tentative promise of the first comprehensive metabolome for an organism.

5. METABONOMICS IN DRUG TOXICOLOGY AND DISEASE PROCESSES

High-resolution ^1H NMR spectroscopy of biofluids has been used extensively to follow the metabolic modifications of xenobiotics such as pharmaceuticals as well as the metabolic changes they induce in mammals (Nicholson *et al.* 1989; Holmes *et al.* 1994; Beckwith Hall *et al.* 1998). The use of urine for the derivation of metabolic profiles makes the technique minimally invasive, and it is hence ideal as a screening tool, especially in drug toxicity trials being carried out in animal models. Indeed, if the metabolic perturbations are indicative of particular organ toxicity it not only becomes possible to classify a drug as potentially toxic but also the approach determines the site and mode of activity (Holmes *et al.* 1998; Nicholson *et al.* 1989, 2002; Griffin *et al.* 2001b). This has led

to many drug companies' developing metabonomic centres for drug safety assessment.

These approaches are also applicable to understanding toxicity at the tissue level. Garrod *et al.* (2000) explored the potential of NMR-based metabonomics of linking histopathology and urinary biomarkers to metabolic changes in the tissue by investigating 2-bromoethanamine (BEA) toxicity, a known renal papillary toxin. They examined changes in two regions of the kidney and the liver using MAS NMR spectroscopy, a technique capable of producing high-resolution NMR spectra in intact tissues directly. From this approach it was found that the drug induced mitochondrial dysfunction and inhibited fatty acyl-CoA dehydrogenases, with ^1H MAS NMR spectroscopy detecting a transient rise in glutaric acid in all three tissue types. Both the renal cortex and papilla demonstrated evidence of changed osmolarity, with decreases in known osmolytes detected in both tissues. Furthermore, toxicity was also detected in the liver where BEA caused increases in lipid triglycerides, lysine and leucine.

Waters *et al.* (2001, 2002) took this analysis one stage further when investigating alpha-naphthylisothiocyanate toxicity in the livers of rats, by comparing urinary and blood plasma biomarkers of toxicity directly with metabolite changes in liver tissue. ^1H MAS NMR spectra of intact liver clearly showed increases in hepatic liver triglycerides, accompanied by decreases in glucose and glycogen. This perturbation in lipid and carbohydrate handling was correlated with increased plasma ketone bodies and a decrease in tricarboxylic acid cycle intermediates in urine. Such a holistic approach clearly demonstrates the correlation between the biofluid markers detected and hepatic tissue toxicity, and hence could be used to confirm specific organ toxicity during the drug development process.

To maximize the information obtainable from multivariate datasets a high-throughput technology is desirable so that the data matrices produced can fully define both the variation associated with a disorder as well as the innate variation associated with the biological system. This is particularly true for the application of functional genomic strategies in safety assessment for drug toxicity studies where target compounds are typically dosed at

pharmacological concentrations, well below estimated LD₅₀ (lethal dose for 50% of the population) concentrations, and hence will have extreme effects in a minority of animals. In this respect NMR-based metabonomics is particularly attractive, being relatively cheap on a per sample basis. With improvements both in automated shimming of samples and flow probe technology, sample throughput for metabolite-rich fluids such as urine and blood plasma is as high as 300 samples per day, with no significant costs or time associated with sample preparation.

Using such an approach the consortium for metabonomic toxicology consisting of Imperial College London, UK, Bristol-Myers Squibb, Eli Lilly and Company, Hoffman-LaRoche, NovoNordisk, Pfizer Incorporated and the Pharmacia Corporation, are currently investigating *ca.* 150 model liver and kidney toxins through NMR-based analysis of urinary metabolites over a 3 year period (Lindon *et al.* 2003). To achieve this it has first been necessary to determine how reproducible such a database would be in terms of both the collection of samples and the resultant NMR analysis. Examining hydrazine toxicity in the rat, Lindon *et al.* (2003) found that the variation in NMR-based metabonomics as a result of conducting a study across seven different laboratories was minor in comparison to the variation associated with the toxic lesion. It is hoped that such an approach will allow the generation of expert systems where liver and kidney toxicity can be predicted for model drug compounds, with the databases being easily transferable between laboratories.

One practical advantage metabolomics and metabonomics has over other 'omic' technologies such as transcriptomics and proteomics is that metabolism is readily transferable from one species to the next. Thus, ¹H MAS NMR spectroscopy provides a quick and convenient technique to investigate potential toxicity in tissues from any species. Using this approach metabolism and the responses to inorganic metals could be compared in the bank vole (*Clethrionomys glareolus*), wood mouse (*Apodemus sylvaticus*) and the white-toothed shrew (*Crocidura suaveolens*) with that of a widely used strain of laboratory rat (Sprague Dawley) using ¹H NMR spectroscopy-based metabolomics (Griffin *et al.* 2000a,b). Furthermore, this approach is potentially very sensitive to toxic insults. Using ¹H MAS NMR spectroscopy to examine cadmium toxicity in the bank vole, biochemical changes in lipid and glutamate metabolism that preceded classical nephrotoxicity were detected (Griffin *et al.* 2000b). These changes occurred after chronic dosing, and at a renal cadmium concentration (8.4 µg g⁻¹ dry weight) that was nearly two orders of magnitude below the World Health Organization critical organ concentration (200 µg g⁻¹ wet weight; World Health Organization 1992). Similar approaches have also been applied to invertebrates, monitoring environmental contamination using worms as reporter species (Bundy *et al.* 2002). To follow such investigations using transcriptomics and proteomics would be vastly complicated by the lack of a sequenced genome for the animal and species-related differences in protein structures. Given the versatility of the approach it is understandable that pharmaceutical companies are increasingly using these techniques to follow drug

safety and development issues (Robertson *et al.* 2000; Smith 2001; Shockcor & Holmes 2002).

Gavaghan *et al.* (2000) have also shown that strains of mouse have their own metabolic phenotype or 'metabotype' using pattern recognition to investigate NMR spectra of urine. This has immediate impact on the use of particular mouse strains to follow drug metabolism, suggesting that certain strains may be prone to certain types of toxicity. Furthermore, given the extensive use of genetically modified mice it is becoming important to appreciate the metabolic profiles of the strain background so as to fully characterize the changes induced by a gene deletion or modification.

Metabonomics has also found a role in the diagnosis of patients, being used to identify inborn errors of metabolism (Moolenaar *et al.* 2003) and the severity of coronary heart disease (Brindle *et al.* 2002). In this latter study the metabolic profile, largely influenced by saturated fats of blood sera and plasma, were correlated with the degree of coronary artery disease, suggesting that these metabolic profiles might have a diagnostic use. If such systems could be applied to the clinical situation, significant financial savings could be made over invasive angiography, currently the gold standard for diagnosis.

To date, NMR-based metabolic profiling has centred on ¹H NMR spectroscopy. However, given the relatively small chemical shift range of the nucleus, there is usually significant overlap between metabolites in conventional one-dimensional spectroscopy. An alternative is to examine metabolites through ¹³C NMR spectroscopy, where resonances are spread over a *ca.* 200 p.p.m. chemical shift range. To compensate for the lower sensitivity of the ¹³C nucleus and a natural abundance of only 1.1%, superconducting NMR probe technology ('cryoprobes') can be used, significantly reducing NMR acquisition times and allowing natural abundance detection of metabolites. This approach relies on cooling the NMR radiofrequency detector and preamplifier to *ca.* 20 K or less (Styles *et al.* 1984). As thermal noise is reduced by a factor equivalent to *ca.* temperature^{1/2}, the thermal noise is reduced approximately fourfold, giving a *ca.* 16-fold reduction in acquisition time for the same signal to noise using a conventional probe. Using this approach Keun *et al.* (2002) readily detected hepatic toxicity using ¹³C NMR spectroscopy of urine detecting metabolites via natural abundance labelling.

6. METABOLIC PROFILING IN PLANT BIOCHEMISTRY

The challenge of metabolomics has also been taken up by plant biochemists. Here, the challenge is potentially even greater than that in mammals and yeast, as despite plant genomes typically containing 20 000–50 000 genes, currently 50 000 metabolites have been identified in the plant kingdom (De Luca & St Pierre 2000), with the number set to rise *ca.* 200 000 (Pichersky & Gang 2000; Fiehn 2001, 2002). Much of the current plant metabolomics is based upon metabolic profiling through mass spectrometry analysis of plant extracts. For example, Fiehn *et al.* (2000) have used GC-MS chromatograms to quantify 326 distinct compounds in *Arabidopsis thaliana* leaf extracts, assigning a chemical structure to half of these.

By applying principal components analysis (PCA) to the metabolic profiles of four genotypes of the plant (two homozygous ecotypes and a mutant of each ecotype) they were able to readily separate the individual strains. By further optimizing this process using time-of-flight mass spectrometry coupled with GC separation, Fiehn has since reported the detection and characterization of *ca.* 1000 metabolites (Hall *et al.* 2002).

An intriguing application of this approach and such comprehensive metabolic matrices is that the data can be used to probe for correlated aspects of metabolism, termed metabolic cliques (Kose *et al.* 2001). These correlations in turn represent the combined effect of both the result of direct enzyme action and indirect cellular control through transcription factors and other regulatory mechanisms. This promises to identify the metabolic pathways that are altered in a pathology directly and may be of immense benefit to anyone studying multifactorial diseases.

Other researchers have used similar approaches to investigate genetically modified plants. For example, transgenic potato plants overexpressing invertase in tubers have been studied using GC-MS (Roessner *et al.* 2000). This approach was able to demonstrate that reduced starch production in these plants was caused by an increase in glycolytic flux. Also, using liquid chromatography as the separation technique prior to mass spectroscopy, Liu *et al.* (2002) have followed the production of isoflavone phytoestrogens, potential chemopreventative agents for certain cancers and cardiovascular disease, in genetically engineered plants. Such metabolic profiling tools are not limited to mass spectrometry, with Bailey *et al.* (2000) demonstrating the analytical power of hyphenated LC-NMR-MS/MS in terms of studying plant metabolism.

One of the most impressive aspects of plant-based metabolomics is the scale of some of the analyses being conducted with throughputs as high as 7000 samples a day, generating gigabytes worth of data (Goodacre *et al.* 2000; Hall *et al.* 2002). Despite the huge technical challenges involved in plant metabolomics, there have been significant advances in terms of functional genomic approaches. In the legume *Medicago truncatula* Sumner has reported using metabolomics to identify and profile hundreds of proteins alongside comprehensive metabolic profiling in a variety of tissues using liquid chromatography ultraviolet mass spectrometry (LC UV MS), GC-MS and capillary electrophoresis-MS, including metabolite groups such as flavonoids, saponins and amino acids (Hall *et al.* 2002). Also it appears that methods that are specific to one class of compounds can still provide metabolic fingerprints that can distinguish different strains. Examining different inbred strains of maize, Nikolau of Iowa State University, Ames, could classify these plants using GC analysis of fatty alcohols, aldehydes, esters, alkanes, ketones and fatty acids (Hall *et al.* 2002). For further information on the rapidly expanding subject area the reader is referred to www.metabolomics.nl.

7. OTHER TECHNIQUES

The field of metabolomics is not limited to those using NMR spectroscopy or mass spectrometry. Spectroscopic

methods based on Fourier transform mass spectrometry (FT-IR) or ultraviolet spectroscopy can also be used, with Oliver *et al.* (1998) using the former to differentiate the respiratory mutants from wild-type strains. While providing only limited metabolic information the approach has the advantage that it is easily scalable, allowing high throughput and making it useful as first screen tool.

Tweeddale *et al.* (1998) have used thin-layer chromatography to follow the metabolic fate of ^{14}C -glucose in *Escherichia coli* under different culture conditions. While the approach is simple it has a great deal of potential for being used with other less global techniques, and has the advantage of being relatively cheap to develop.

Arrays are also in development for fingerprinting metabolism in *E. coli* (Bochner *et al.* 2001). A 96-well plate assay has been developed relying on nicotinamide adenine dinucleotide reduction of tetrazolium dye, producing a quantifiable colour change in each well plate depending on the substrates and enzymes present. The authors used 700 different assay mixtures to assess metabolism and drug resistance in *E. coli*, making it ideal as a screening tool.

A common factor of the techniques described above, with the notable exception of examinations of biofluids, is that extraction of the metabolites from tissues or cells is required. The concentration of a metabolite in an extraction medium depends on both the metabolite's relative solubility and its tissue concentration. If aqueous and lipophilic metabolites are investigated simultaneously, this requires multiphase extractions, which can be time consuming, and metabolites may even be trapped in the remaining tissue pellet. For example, synaptic glutamate in cerebral tissue is contained within a lipid vesicle and thus may be resistant to aqueous extraction processes.

To circumvent these problems the tissues can be examined directly using MAS ^1H NMR spectroscopy. In this technique the sample is spun at the 'magic angle' (54.7°), reducing towards zero the effects of physical processes that broaden resonances in tissues, such as magnetic susceptibility, dipolar interactions and chemical shift anisotropy. This produces NMR spectra that have line widths close to those obtainable in the solution state (Andrew *et al.* 1959; figure 2).

In § 8 three specific examples of this technique to metabolic profile disease models will be discussed. These examples are from the author's laboratory and in each case the results are placed within the context of the known diseases. Although these animal models represent gene deletions or modifications associated with disease, in all cases the phenotype is markedly milder than that found in humans. In this respect metabolomics allows the assessment of what pathways are affected by these genetic modifications, as well as defining the complex phenotype associated with the disease model. Although not necessarily a silent phenotype, these models often provide a confused message when initially produced.

8. DEFINING THE ACTION OF DYSTROPHIN AND DYSTROPHIN-RELATED PROTEINS THROUGH METABOLITES

While the genetic basis of Duchenne muscular dystrophy (DMD) is well documented, there are still questions remaining as to the role of the various isoform

Figure 2. A comparison of ¹H NMR spectra from an intact piece of liver tissue either (a) static or (b) spun at 6000 Hz to demonstrate the spectral improvements produced by MAS (spectra supplied by Dr Mary Bollard, Imperial College, London (Bollard *et al.* 2000)).

proteins produced by the gene, demonstrating the problem with moving from a genomic to a proteomic description of a disease. There are three full-length isoforms of dystrophin: M-dystrophin found in muscle and, to a small extent, glial cells, P-dystrophin found in the Purkinje cells of the cerebellum and C-dystrophin found in the cortex (Nudel *et al.* 1989) as well as a number of truncated forms of the protein (Lederfein *et al.* 1992). Furthermore, the *mdx* mouse, the most widely used model of the disease, displays little in the way of pathology when compared with human sufferers, suggesting that other proteins may take on dystrophin's role in the mouse but not humans. Previous studies have noted metabolic perturbations in both dystrophic muscle and cerebral tissue (Kemp *et al.* 1993; Even *et al.* 1994; Mokhtarian *et al.* 1996; Tracey *et al.* 1996; McIntosh *et al.* 1998*a,b*), but these have largely been confined to one tissue type or organ, focusing on only a handful of metabolites.

By interrogating ¹H NMR spectra of tissue extracts with PCA, metabolic profiles were quickly established for different dystrophic tissues compared with tissue from control animals (Griffin *et al.* 2001*a,c,e*; figure 3). Dystrophic muscle tissues, including that from the heart and diaphragm, are similarly perturbed in terms of metabolism, and are characterized by an increase in taurine and lactate alongside a decrease in creatine in tissue extracts of dystrophic tissue. This is in keeping with the single protein isoform known to be involved in all these tissue types. Previously, McIntosh *et al.* (1998*a,b*) correlated muscle taurine content to regeneration of muscle tissue in the *mdx* mouse. As well as muscle regeneration, taurine has a number of diverse metabolic roles including osmolyte, cell membrane stabilizer, Ca²⁺ homeostasis and a neurotransmitter (Huxtable 1992), and the primary cause of

increased taurine in dystrophic muscle cells remains to be determined.

Similarly, both spectra from dystrophic cerebellar and cortical tissue could be separated by PCA, confirming the findings of previous researchers examining both sufferers of DMD and the *mdx* mouse that a metabolic deficit is associated with DMD (Tracey *et al.* 1996; Kato *et al.* 1997; Rae *et al.* 1998; Anderson *et al.* 2002; figure 3). The metabolic perturbations detected were different for the cortex and cerebellum, suggesting that the approach was capable of detecting the effects of the different isoforms of the protein.

Dystrophin is associated with the cell membrane, and thus any analysis focusing on aqueous metabolites, and hence the changes in the cytosol, is unlikely to give the complete picture of the metabolic changes associated with dystrophic tissue. MAS ¹H NMR spectroscopy allows the derivation of metabolic profiles from high-resolution spectra using intact tissue. Furthermore, the techniques used to derive these spectra are also sensitive to the environment the metabolites occur in.

Using a conventional pulse sequence, spectra were produced that were dominated by —CH₂CH₂CH₂— and CH₃CH₂— lipid moieties (figure 4). These same lipid moieties readily classified dystrophic and control cardiac tissue using PCA, in keeping with the known progression of dystrophic muscle degeneration (Gillet *et al.* 1993). Waves of muscle necrosis are accompanied by extracellular lipid infiltration, and it is this increased lipid content that is detected in the spectra. To further investigate the lipid resonances, PCA was applied to spectra acquired with either a diffusion edited pulse sequence, more sensitive to slow moving metabolites, or a T₂ edited pulse sequence, more sensitive to mobile metabolites. While

Figure 3. PCA of ^1H NMR spectra of extracts from: (a,b) cardiac; (c,d) cerebral cortex; and (e,f) cerebellum tissue, using either mean centred (a,c,e) or univariate preprocessing (b,d,f). Each symbol represents an NMR spectrum reduced to principal component space. The scale indicates the score of each spectrum along that particular PC. By reference to the contributions, plots of individual PCs the metabolic perturbations associated with dystrophin can also be deduced. Filled squares, dystrophic tissue; open circles, control tissue.

lipid moieties still separated the two tissue types, intriguingly the chemical groups responsible for this separation were different. Dystrophic tissue contained more rotationally free lipids with CH_2CH_2 , CH_3 , $\text{CH}=\text{CHCH}_2$ and COCH_2CH_2 chemical groups and slow-moving lipids with saturated fats containing relatively less COCH_2CH_2 and CH_3 chemical groups (Griffin *et al.* 2001c).

Metabolomics may also have a role to play in following the functional proteomics of the dystrophin family of proteins through metabolites (Griffin *et al.* 2002a). For such roles, where metabolite changes are correlated with protein expression, a supervised pattern recognition technique is required so as to correlate metabolic changes with protein expression. To investigate the action of utrophin, a functionally related protein to dystrophin (Rybakova *et al.* 2002), three mouse models of DMD were investigated using metabolomics: (i) the *mdx* mouse, which does not express dystrophin; (ii) a $\text{Tg}_{\text{full length}}/\text{Dmd}^{\text{mdx}}$ transgenic mouse, termed Fiona, expressing full-length utrophin in skeletal muscle but not heart crossed with the *mdx* mouse

to produce mice lacking dystrophin but having utrophin localized at the sarcolemma (Tinsley *et al.* 1998) and (iii) a mouse lacking both dystrophin and utrophin, termed Gavin.

During foetal development utrophin is found over the entire surface of muscle fibres, but is replaced by dystrophin during development, when utrophin becomes localized to the neuromuscular junction in adult skeletal muscle (Hoffman *et al.* 1989; Khurana *et al.* 1990, 1991). The dystrophic phenotype normally observed in *mdx* mice is absent when muscles overexpress utrophin (Tinsley *et al.* 1998). Furthermore, utrophin has a particularly favourable characteristic in dystrophic tissue: whereas the body may treat dystrophin as a foreign protein in sufferers of DMD, mounting an immune response, increasing the innate expression of utrophin does not have this pitfall (Burton & Davies 2002).

Examining cardiac tissue extracts using supervised pattern recognition, a model was readily built that separated dystrophic and control cardiac tissue from the three

Figure 4. MAS ^1H NMR spectra of cardiac tissue from the *mdx* mouse. The spectra were acquired from the same piece of cardiac tissue using either a Carr, Purcell, Meiboom and Gill (CPMG) pulse sequence with (a) a total T_2 spin echo delay of 40 ms, or (b) a conventional solvent suppressed pulse sequence. The bold type in (a) signifies the position of protons observed in the given resonance. The spin echo delay attenuates the signal from all resonances, but affects motionally constrained metabolites more. In the cardiac spectra the spin echo delay reduces the resonances associated with a number of broad lipid components, in particular those associated with CH_3 and $\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}$ lipid groups. However, the resonances associated with the methyl group of lactate are more pronounced in these spectra.

mouse types described above. This model indicated that an increase in tissue taurine concentration was associated with dystrophic tissue in all three mouse models. Testing the predictive power of these models, a 90% success rate was produced. Thus, the technique could be used to predict the protein expression in new mouse models for DMD through the tissue's metabolomic profile.

In theory, this approach could model a whole series of proteins in terms of metabolite profiles. To investigate the reliability of such a model in diaphragm tissue, spectra of tissue extracts were again used as the basis of a metabolic dataset. The three mouse groups could readily be separated according to their metabolic profiles. Again, as in cardiac tissue, taurine appeared to be a significant biomarker for a failure to express dystrophin. However, the effects of utrophin expression were also apparent, with decreased concentrations of glycerol, creatine and glucose and increased concentrations of short chain fatty acids and β -hydroxybutyrate in diaphragm tissue from $\text{Tg}_{\text{full length}}/\text{Dmd}^{\text{mdx}}$ mice.

The metabolic profile of $\text{Tg}_{\text{full length}}/\text{Dmd}^{\text{mdx}}$ (Fiona) mice was intermediate in terms of the part of the model describing 'dystrophin expression', suggesting that promoting utrophin expression, as occurs in skeletal muscle and diaphragm but not cardiac muscle of $\text{Tg}_{\text{full length}}/\text{Dmd}^{\text{mdx}}$ mice, can negate the metabolic effects of a failure to express dystrophin. This is also supported physiologically in the *mdx* mouse. Tinsley *et al.* (1998) have demonstrated that increased utrophin expression reverses the

Figure 5. Luxophilic cytoplasmic inclusion bodies in the cerebral cortex of the *mnd* mouse: luxol fast blue and cresyl violet staining (magn. $\times 100$). The arrow shows an inclusion body and it is at these sites where oxidative damage is thought to occur.

phenotype characteristic of *mdx* mouse muscle tissue. Thus, this approach may be used for screening for treatments of DMD that cause a temporary change in protein expression (Krag *et al.* 2001), such as gutted adenoviral vectors and plasmid vectors transfecting the missing dystrophin cDNA into a tissue (Cordier *et al.* 2000; Wang *et al.* 2000), aminoglycosides that suppress stop mutations (Barton-Davis *et al.* 1999) and the introduction of stem cells (Gussoni *et al.* 1999). However, perhaps the most intriguing concept is that this may be a general approach for conducting *in situ* proteomics using the high-throughput capabilities of ^1H NMR-derived metabolic profiles.

9. MOUSE MODELS OF CEREBRAL DISORDERS

Both in terms of ^1H NMR spectra *ex vivo* and ^1H magnetic resonance spectroscopy (MRS) *in vivo* the application of pattern recognition techniques to characterize metabolic disorders in brain tissue is well established (Tate *et al.* 1996, 1998; Stoyanova & Brown 2001, 2002). Thus, the application of metabolomics to brain tissue metabolism has effectively already been demonstrated. With the advent of high-resolution MAS ^1H NMR spectroscopy of intact tissues, metabolic profiles can now be generated on biopsy-sized material. This approach for metabolic phenotyping is set to become increasingly popular, especially when compared with relatively more labour-intensive methods of phenotype characterization such as cognitive testing or global histology. Furthermore, the information obtained through metabolic profiles may be more sensitive to the phenotype changes induced by a gene alteration as illustrated by the *mnd* mouse.

The *mnd* mouse is a model of neuronal ceroid lipofuscinosis (NCL3), demonstrating marked progressive cerebral and visual deterioration along with the accumulation of lipofuscin and ceroid deposits (Dawson & Cho 2000; Mole *et al.* 1999; figure 5). These deposits in patients suffering from NCL3 and the *mnd* mouse are caused by impaired formation of the mitochondrial associated membranes (MAMs), decreasing the transport of fats into mitochondria (Vance *et al.* 1997). The depositions of

lipofuscin and ceroid deposits are thought to provide sites where free radicals are produced (Vance *et al.* 1997). Possibly as a direct result of this increase in intracellular free radical production, the mouse also has a profound and progressive vitamin E deficiency in both sera and brain tissue, suggesting that the vitamin E deficiency may in part be responsible for some of the neurological deficits reported.

However, when the vitamin E deficiency was corrected using dietary supplementation, the histopathological features associated with NCL remained (Griffin *et al.* 2002b). Using a bioinformatics approach based on high-resolution solid and solution-state ^1H NMR spectroscopy and PCA, the deficits associated with NCL could readily be defined in terms of a metabolic phenotype. While vitamin E supplementation reversed some of the metabolic abnormalities, in particular the concentration of phenylalanine in extracts of cerebral tissue, it also demonstrated that metabolic deficits associated with NCL were greater than any effects produced from vitamin E supplementation. These deficits included increased glutamate and N-acetyl aspartate (NAA) and decreased creatine and glutamine concentrations in aqueous extracts of the cortex, as well as profound accumulation of lipid in intact cerebral tissue.

The large accumulation of lipids within *mnd* cerebral tissue is most probably a directly result of faulty MAM production, but the increase in NAA was harder to explain. The acetyl group of NAA is readily hydrolysed and inspection of spectra from intact tissue indicated that the increase in cerebral NAA concentration in the *mnd* mouse resulted from a decrease in the rate of degradation. While further studies are required, provisional results suggest that the metabolic perturbations detected result from a redistribution of a number of enzymes associated with MAMs.

One set of neurological disorders that may be particularly tractable to pattern recognition-based metabolomics are those arising from expansion of nucleotide triplet repeats. The spinocerebellar ataxias are a group of inherited disorders caused by the expansion of CAG repeats translated into polyglutamine expansions. Machado-Joseph disease (MJD) (Mendelian inheritance in man number 109150) is the commonest dominant inherited ataxia disease and is categorized by cerebellar ataxia, spasticity, ophthalmoplegia and dysarthria (Rosenberg 1992; Kawaguchi *et al.* 1994; Takiyama *et al.* 1994; Evert *et al.* 2000). Pathological phenotypes are apparent with a triplet repeat length of 61–84, compared with the normal range of 12–37 (Durr *et al.* 1996). Can metabolomics correlate the triplet repeat length with increasing pathology in models of these diseases?

The protein associated with MJD, ataxin-3, is cytosolic, expressed throughout the body, and in affected individuals forms ubiquitinated intranuclear inclusions (NIs), particularly in neurons of the brain regions targeted by the disease. However, the function of ataxin-3 is still unknown, and in an attempt to define the role of this protein a metabolomic investigation is being conducted in terms of both aqueous extracts and intact tissue. In this study, a mouse model of MJD has been examined that contains a human yeast artificial chromosome (YAC) containing a MJD locus with a triplet expansion length of 86 (Cemal *et al.* 2002). These mice have previously been shown to possess a mild progressive cerebellar deficit and be affected by NI

deposition in the pontine and dentate nuclei. However, metabolic abnormalities have not been previously reported. Using NMR-based metabolomics, a metabolic fingerprint associated with triplet repeat expansions in the MJD gene for both cerebellum and also, intriguingly, cortical tissue was readily produced. Not only did this fingerprint separate mouse tissue from the cerebellum and cortex into those from the disease model and from control animals, but the metabolites responsible for this classification were similar for the two brain regions. These included an increase in glutamine in both regions.

It is possible to extenuate the MJD phenotype in the mouse models further by introducing YACs containing a MJD locus with longer triplet repeat lengths. Lodi *et al.* (1999) found a progressive phenotype in terms of ^{31}P NMR observable high-energy phosphate metabolites in human sufferers of Friedreich's ataxia, also caused by a triplet repeat expansion. It remains to be seen whether a ^1H NMR-derived metabolomic profile will also be able to derive such a progression in the mouse model of MJD.

10. CELL CULTURE SYSTEMS

One of the first applications of ^1H MAS NMR spectroscopy was the investigation of cultured adipocytes and how such cell culture systems could be used to follow cell proliferation in adipocarcinomas (Weybright *et al.* 1998). Despite requiring large spinning rates for these experiments, cultured cells appear viable after MAS NMR spectroscopy (Weybright *et al.* 1998; Griffin *et al.* 2002c), suggesting that the technique could be used to follow temporal changes on the same population of cells. Using this approach in conjunction with pattern recognition, different neural cell types have been distinguished (Griffin *et al.* 2002c). Interestingly, the large lipid content of neuronal cells caused most of the separation from the other cell types in PCA. This suggests that during MRS *in vivo*, where lipid resonances are commonly 'edited out', significant information may be sacrificed concerning the relative distribution of individual cell types.

Metabolomics is equally applicable to other cell culture-based experiments. Ishikawa cells are a human cell line derived from endometrial adenocarcinoma, and being hormone responsive are potentially very useful for investigating drugs that modulate the oestrogen receptor (Carmichael 1998). ^1H MAS NMR spectroscopy of intact cells produced spectra with line widths typically less than 5 Hz, but a number of these resonances were obscured by broad lipid resonances. However, low molecular weight metabolites can be investigated by applying a pulse sequence capable of 'editing out' the more intense lipid signals (figure 6). Examining the action of tamoxifen, a pattern recognition model was built to correlate metabolic changes caused by tamoxifen against dose (Griffin *et al.* 2003). Among the metabolites that contributed to this model were ethanolamine, *myo*-inositol, uridine and adenosine, suggesting alterations in both cell membrane turnover and RNA transcription. Furthermore, the metabolic effects of other oestrogen modulators could be monitored using this partial least squares (PLS) model, effectively scoring these drugs in terms of tamoxifen doses.

Potentially such techniques could be used to monitor drug interactions and subsequent toxicity in any cell

Figure 6. Aliphatic region of two ^1H MAS NMR spectra of endometrial cells acquired using a Carr Purcell Meiboom and Gill pulse sequence with (a) 320 ms and (b) 10 ms total spin echo time. The expanded region shows the effects of T_2 attenuation on the resonances of choline-containing metabolites between 3.2 p.p.m. and 3.3 p.p.m. 1, CH_3CH_2 lipid groups; 2, leucine, valine, isoleucine; 3, lactate; 4, CH_2CH_2 lipid groups; 5, alanine; 6, COCH_2CH_2 lipid groups; 7, $\text{C}=\text{CHCH}_2\text{CH}_2$ lipid groups; 8, acetyl groups; 9, glutamate; 10, creatine; 11, choline; 12, phosphocholine; 13, glycerophosphocholine/ phosphatidylcholine; 14, phosphoethanolamine; 15, taurine; 16, *myo*-inositol.

culture system. The technique is also sensitive to certain chemical and physical changes that are not often considered; for example during molecular biology manipulations. Choline-containing metabolites have been associated with a number of disorders, including malignant cell growth, DMD and multiple sclerosis (Florian *et al.* 1995a,b; Rae *et al.* 1998). During transient transfection of an unknown cloned gene thought to be involved in lipid metabolism using electroporation into hepatocytes, large changes in the relative proportion of choline to phosphocholine and phosphatidylcholine were detected. Under such circumstances it would be tempting to suggest that the gene's function was related to choline metabolism. However, on closer inspection, such changes were found to accompany transient transfection with the LacZ reporter gene and even naked plasmids, demonstrating that the effect arose from the action of electroporation rather than anything related to the gene function (Griffin *et al.* 2001d). Thus, changes in the detection of choline may monitor changes in cell membrane stability.

11. PATTERN RECOGNITION TECHNIQUES FOR METABOLOMICS

As mentioned previously, a central area in metabolomics is the use of pattern recognition tools to decipher the changes associated with a disease process from those associated with normal physiological variation. An integral part to identifying metabolic differences in large datasets, as will become increasingly common in metabolomics, is placing both high- and low-concentration metabolites on an equal footing. For example, while a high-concentration metabolite will have a large numerical variance compared with a low-concentration metabolite, the actual fold

changes may be greater for the lower-concentration metabolite. Thus, measures of metabolite concentrations must be scaled so as to maximize the information obtainable from both high- and low-concentration metabolites. Data are often 'mean centred', where the average of all the spectra is subtracted from each individual spectrum, and hence variance is expressed with respect to the mean spectrum. Using this scaling alone, high-concentration metabolites contribute the most significantly to the description. To increase the representation of low-concentration metabolites, the variance of each integral region can be divided by the variance of the integral region, standardizing the variance for each integral region (referred to as univariate scaling) or by log transformation of the data. Alternatively, intermediate scalings can be used to interrogate the dataset further; for example Pareto scaling where the variance is divided by the square root of variance (Eriksson *et al.* 1999). In this manner both high-concentration and trace metabolites can be identified by subsequent pattern recognition.

Both unsupervised and supervised techniques can be used to derive metabolic profiles. To investigate the innate variation in a dataset, unsupervised techniques such as PCA (Hotelling 1933), hierarchical cluster analysis (HCA), K-means clustering or self-organizing maps (Kohonen 1982) could be used. However, where specific questions are being posed, supervised techniques such as prediction to latent structures through PLS (Eriksson *et al.* 1999), genetic programming (Kell 2002), discriminate function analysis (Lachenbruch 1975), support vector machines (Christianini & Shawe-Taylor 2001) and neural networks may be more appropriate. PLS, the regression extension of PCA, can also be used as a means of data filtering, referred to as orthogonal signal correction. Variation that is orthogonal to the trend of interest is removed using PLS. For all the supervised techniques it is necessary to test the robustness and predictability of the models produced, although the biological function of the metabolites identified may also indicate the success of the particular pattern recognition tool in producing a metabolic profile of a given disease.

To implement these pattern recognition tools there are several good commercially available software packages including SIMCA (PCA and PLS tools; Umetrics, Sweden), PIROUETTE (HCA and PCA; Informetrix, USA) and MATLAB (a range of multivariate tools found in various tool boxes; Mathworks, USA). Furthermore, for an excellent review on pattern recognition methods, albeit for use with microarray data, the reader is referred to Valafar (2002).

12. A THEORETICAL BASIS FOR METABOLOMICS

So why does metabolomics work as a functional genomics approach? This is a trivial question if the mutation or modification being investigated involves an inborn error of metabolism or an enzyme crucial to a particular metabolic pathway. Such mutations produce classical metabolic responses such as perturbations in steady-state metabolite concentrations, and altered flux rates in pathways connected by this enzyme. Also mutations that have a profound effect on the cellular environment will also have an effect on the metabolome. For example, the failure to express dystrophin in dystrophic muscle tissue produces an influx

of calcium, which in turn affects a host of metabolic pathways, providing an explanation as to why metabolic profiles may be particularly suited for following models of DMD.

However, Teusink *et al.* (1998) provide an explanation as to why the technique has much broader applications by showing that their FANCY approach has a firm theoretical basis in metabolic control analysis (MCA). In MCA (Fell 1996), metabolic flux is modelled in terms of control coefficients that measure either fractional changes in flux rates or metabolite concentrations as a result of some perturbation (Kacser & Burns 1973). For this approach either changes in flux rate or metabolite concentrations are measured, with the latter often proving to be technically simpler and more sensitive to the perturbation. Importantly, MCA has shown that the flux rate through a particular pathway is rarely solely determined by one enzyme and as such a perturbation at one point is propagated throughout the network. Thus, the multivariate analysis that is a part of metabolomics measures indirectly these concentration control coefficients, and hence should be a widely applicable tool (Raamsdonk *et al.* 2001).

Metabolism appears to consist of many small modules that are highly connected into a hierarchy where a few metabolites such as pyruvate and acetyl-CoA connect many such modules (Ravasz *et al.* 2002). Furthermore, for 43 organisms the clusterings follow a simple power law, suggesting that for even complicated metabolic networks, the topology is scale-free and highly connected. A perturbation at one point will be rapidly transferred to many of the key nodes, and if these nodes are represented by metabolites of sufficient concentration for NMR to detect them, then it appears reasonable that metabolomics will be able to detect these perturbations.

13. FUTURE DIRECTIONS AND CHALLENGES FOR NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY-BASED METABOLOMICS

Although spinning speeds used in MAS NMR spectroscopy at present do not appear to produce significant damage in tissue samples, as superconducting magnets move to increased field strength and spectrometers attain higher observation frequencies, samples will potentially be spun at higher speeds to remove spinning side bands from the region of interest. At some point tissue degradation and the heating effect caused by spinning the sample will become significant. Thus, there is a need to develop pulse sequences that can operate at slower speeds but still provide high-resolution MAS ^1H NMR spectra. Wind *et al.* (2001) and Hu *et al.* (2002) have recently described a pulse sequence capable of operating with a sample spinning speed of only 1–4 Hz, suggesting that high-resolution spectra for metabolic profiling and fingerprinting will be obtainable with minimal tissue damage.

Meanwhile, the development of *in vivo* MRS continues, and with improved coil design and better localization pulse sequences, spectra are obtainable that are comparable in resonance line width to those obtainable using ^1H MAS NMR spectroscopy, particularly for fatty tissues (figure 7). Thus, there will be wider scope for metabolomic studies *in vivo*.

Figure 7. A ^1H MAS NMR spectrum of an intact glioma removed from a rat brain obtained at (c) 600 MHz compared with two *in vivo* pulse MRS sequences ((a) stimulated echo acquisition mode (STEAM) and (b) localization by adiabatic selective echo refocusing (LASER)) examining the same glioma *in vivo* at 400 MHz. Spectra were provided by Professor R. A. Kauppinen, University of Manchester.

Even prior to these advances, the potential of being able to detect metabolites in tissues on a comparable scale to those used by histologists makes ^1H MAS NMR spectroscopy a desirable technique for pathology. If the technique can be fully automated, a routine step in histopathology will be sending sections to be analysed by both histology and ^1H MAS NMR spectroscopy, providing another tier in the systems approach to functional genomics.

Looking at the wider picture, perhaps the most intriguing application of metabolomic and metabonomic techniques may come in comparative functional genomics of model organisms. To an extent this has already begun in yeast (Raamsdonk *et al.* 2001), aided by such depositories as EUROSCARF (European *Saccharomyces cerevisiae* archive for functional analysis) (Winzeler *et al.* 1999) containing the *ca.* 6000 mutant strains of yeast (Goffeau *et al.* 1996). However, functional genomics involving other species could benefit from such FANCY approaches. For example, a nematode worm (*Caenorhabditis elegans*) model of DMD exists, where the dystrophin analogue dys-1 has been knocked out (Gieseler *et al.* 2000). Given the success of metabolomics as a tool to probe dystrophic tissue in mice, the *C. elegans* model could be used to screen for small molecules that could

promote upregulation of dystrophin or its related proteins in this animal using metabolomics. This would provide a relatively cheap and high-throughput technique for treating this terminal disease.

Perhaps the biggest advantage of metabolomics is the technique's relative cheapness. A transcriptional approach is currently two to three orders of magnitude more expensive than a metabolomic investigation, assuming a similar repetition number. Furthermore, the preliminary evidence suggests that the reproducibility of metabolomics is greater than transcriptomics, further reducing costs in a metabolomic approach to functional genomics. While the level of information obtained through metabolomics will be far more difficult to interpret because the link between protein and metabolite is less well defined than that between mRNA and protein expression, a first pass of an organism's metabolome is also feasible in terms of time. Using flow probe technology 12 NMR samples an hour can routinely be run in a fully automated manner. Thus, at least in terms of a first pass, metabolomics may be a quick and relatively cheap process to profile an organism in terms of its complete functional genomic interactions.

The author thanks Dr M. Bollard of Imperial College London and Professor R. Kauppinen of Manchester University for supplying spectra, and Dr N. Bailey of Scynexis, and Fraser Griffin for useful discussions. Work in the author's laboratory is supported by The Royal Society.

REFERENCES

- Allen, J., Davey, H. M., Broadhurst, D., Heald, J. K., Rowland, J. J., Oliver, S. G. & Kell, D. B. 2003 High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nature Biotechnol.* **21**, 692–696.
- Anderson, J. L., Head, S. I., Rae, C. & Morley, J. W. 2002 Brain function in Duchenne muscular dystrophy. *Brain* **125**, 4–13.
- Andrew, E. R., Bradbury, A. & Eades, R. G. 1959 Removal of dipolar broadening of NMR spectra of solids by specimen rotation. *Nature* **183**, 1802.
- Bailey, N. J., Cooper, P., Hadfield, S. T., Lenz, E. M., Lindon, J. C., Nicholson, J. K., Stanley, P. D., Wilson, I. D., Wright, B. & Taylor, S. D. 2000 Application of directly coupled HPLC-NMR-MS/MS to the identification of metabolites of 5-trifluoromethylpyridone (2-hydroxy-5-trifluoromethylpyridine) in hydroponically grown plants. *J. Agric. Food Chem.* **48**, 42–46.
- Barton-Davis, E. R., Cordier, L., Shoturma, D. I., Leland, S. E. & Sweeney, H. L. 1999 Aminoglycoside antibiotics restore dystrophin function to skeletal muscles of *mdx* mice. *J. Clin. Invest.* **104**, 375–381.
- Beckwith Hall, B. M., Nicholson, J. K., Nicholls, A. W., Foxall, P. J., Lindon, J., Connor, S. C., Abdi, M., Connelley, J. & Holmes, E. 1998 Nuclear magnetic resonance spectroscopic and principal components analysis investigations into biochemical effects of three model hepatotoxins. *Chem. Res. Toxicol.* **11**, 260–272.
- Bochner, B. R., Gadzinski, P. & Panomitos, E. 2001 Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res.* **11**, 1246–1255.
- Bollard, M. E., Garrod, S., Holmes, E., Lindon, J. C., Humpfer, E., Spraul, M. & Nicholson, J. K. 2000 High resolution ^1H and ^1H - ^{13}C magic angle spinning NMR spectroscopy of rat liver. *Magn. Reson. Med.* **44**, 201–207.
- Brindle, J. T. (and 10 others) 2002 Rapid and non-invasive diagnosis of the presence and severity of coronary heart disease using ^1H -NMR-based metabolomics. *Nature Med.* **8**, 1439–1445.
- Buchholz, A., Hurlbaush, J., Wandrey, C. & Takors, R. 2002 Metabolomics: quantification of intracellular metabolite dynamics. *Biomol. Engng* **19**, 5–15.
- Bundy, J. G., Spurgeon, D. J., Svendsen, C., Hankard, P. K., Osborn, D., Lindon, J. C. & Nicholson, J. K. 2002 Earthworm species of the genus *Eisenia* can be phenotypically differentiated by metabolic profiling. *FEBS Lett.* **521**, 115–120.
- Burton, E. A. & Davies, K. E. 2002 Muscular dystrophy: reason for optimism? *Cell* **108**, 5–8.
- Carmichael, P. L. 1998 Mechanisms of action of antiestrogens: relevance to clinical benefits and risks. *Cancer Invest.* **16**, 604–611.
- Castrillo, J. I., Hayes, A., Mohammed, S., Gaskell, S. J. & Oliver, S. G. 2003 An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry* **62**, 929–937.
- Cemal, C. K., Carroll, C. J., Lawrence, L., Lowrie, M. B., Ruddle, P., Al-Mahdawi, S., King, R. H., Pook, M. A., Huxley, C. & Chamberlain, S. 2002 YAC transgenic mice carrying pathological alleles of the MJD1 locus exhibit a mild and slowly progressive cerebellar deficit. *Hum. Mol. Genet.* **11**, 1075–1094.
- Christianini, N. & Shawe-Taylor, J. 2001 *An introduction to support vector machines*. London: Cambridge University Press.
- Cordier, L., Hack, A. A., Scott, M. O., Barton-Davis, E. R., Gao, G., Wilson, J. M., McNally, E. M. & Sweeney, H. L. 2000 Rescue of skeletal muscles of gamma-sarcoglycan-deficient mice with adeno-associated virus-mediated gene transfer. *Mol. Ther.* **1**, 119–129.
- Dawson, G. & Cho, S. 2000 Batten's disease: clues to neuronal protein catabolism in lysosomes. *J. Neurosci. Res.* **60**, 133–140.
- De Luca, V. & St Pierre, B. 2000 The cell and developmental biology of alkaloid biosynthesis. *Trends Pl. Sci.* **5**, 168–173.
- Devaux, P. G., Horning, M. G. & Horning, E. C. 1971 Benzyl-oxime derivatives of steroids; a new metabolic profile procedure for human urinary steroids. *Analyt. Lett.* **4**, 70–82.
- Durr, A. (and 13 others) 1996 Spinocerebellar ataxia 3 and Machado-Joseph disease: clinical, molecular, and neuropathological features. *Ann. Neurol.* **39**, 490–499.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N. & Wold, S. 1999 *Introduction to multi- and megavariate data analysis using projection methods (PCA & PLS)*. Umeå, Sweden: Umetrics.
- Evans, G. A. 2000 Designer science and the 'omic' revolution. *Nature Biotechnol.* **18**, 127.
- Even, P. C., Decrouy, A. & Chinot, A. 1994 Defective regulation of energy metabolism in *mdx*-mouse skeletal muscles. *Biochem. J.* **304**, 649–654.
- Evert, B. O., Wullner, U. & Klockgether, T. 2000 Cell death in polyglutamine diseases. *Cell Tissue Res.* **301**, 189–204.
- Fan, T. W. 1996 Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. *Prog. Nucl. Magn. Reson. Spectrosc.* **28**, 161–219.
- Fell, D. A. 1996 *Understanding the control of metabolism*. London: Portland Press.
- Fiehn, O. 2001 Combining genomics, metabolome analysis and biochemical modeling to understand metabolic networks. *Comp. Function Genomics* **2**, 155–168.
- Fiehn, O. 2002 Metabolomics: the link between genotypes and phenotypes. *Pl. Mol. Biol.* **48**, 155–171.
- Fiehn, O. 2003 Metabolic networks of *Cucurbita maxima* phloem. *Phytochemistry* **62**, 875–886.
- Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Tretheway, R. N. & Willmitzer, L. 2000 Metabolite profiling for plant functional genomics. *Nature Biotechnol.* **18**, 1157–1161.

- Florian, C. L., Preece, N. E., Bhakoo, K. K., Williams, S. R. & Noble, M. D. 1995a Characteristic metabolic profiles revealed by ^1H NMR spectroscopy for three types of human brain and nervous system tumours. *NMR Biomed.* **8**, 253–264.
- Florian, C. L., Preece, N. E., Bhakoo, K. K., Williams, S. R. & Noble, M. D. 1995b Cell type-specific fingerprinting of meningioma and meningeal cells by proton nuclear magnetic resonance spectroscopy. *Cancer Res.* **55**, 420–427.
- Garrod, S., Humphreys, E., Connor, S. C., Connelly, J. C., Spraul, M., Nicholson, J. K. & Holmes, E. 2000 High resolution ^1H NMR and magic angle spinning NMR spectroscopic investigation of the biochemical effects of 2-bromoethanamine (BEA) in intact renal and hepatic tissue. *Magn. Reson. Med.* **45**, 781–790.
- Gavaghan, C. L., Holmes, E., Lenz, E., Wilson, I. D. & Nicholson, J. K. 2000 An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse. *FEBS Lett.* **484**, 169–174.
- Gieseler, K., Grisoni, K. & Segalat, L. 2000 Genetic suppression of phenotypes arising from mutations in dystrophin-related genes in *Caenorhabditis elegans*. *Curr. Biol.* **10**, 1092–1097.
- Gillet, B., Doan, B. T., Verre-Serrie, C., Barbere, B., Berenger, G., Morin, S., Koenig, J., Peres, M., Sebillé, A. & Beloeil, J. C. 1993 *In vivo* 2D ^1H NMR of *mdx* mouse muscle and myoblast cells during fusion: evidence for a characteristic signal of long chain fatty acids. *Neuromusc. Disord.* **3**, 433–438.
- Goffeau, A. (and 15 others) 1996 Life with 6000 genes. *Science* **274**, 546–567.
- Goodacre, R., Shann, B., Gilbert, R. J., Timmins, E. M., McGovern, A. C., Aisberg, B. K., Kell, D. B. & Logan, N. A. 2000 The detection of the dipicolinic acid biomarker in *Bacillus* spores using Curie-point pyrolysis mass spectrometry and Fourier transform infrared spectroscopy. *Anal. Chem.* **72**, 119–127.
- Griffin, J. L., Walker, L., Garrod, S., Holmes, E., Shore, R. F. & Nicholson, J. K. 2000a Metabolic differences exemplified by urinary profiles and renal tissue in the bank vole (*Clethrionomys glareolus*), the wood mouse (*Apodemus sylvaticus*), the white toothed shrew (*Crocodyura suaveolens*) and a strain of laboratory rat using ^1H NMR. *J. Comp. Biochem. Physiol. B* **127**, 357–367.
- Griffin, J. L., Walker, L. A., Troke, J., Osborn, D., Shore, R. F. & Nicholson, J. K. 2000b The initial pathogenesis of cadmium induced renal toxicity. *FEBS Lett.* **478**, 147–150.
- Griffin, J. L., Williams, H. J., Sang, E., Clarke, K., Rae, C. & Nicholson, J. K. 2001a Metabolic profiling of genetic disorders: a multitissue ^1H nuclear magnetic resonance spectroscopic and pattern recognition study into dystrophic tissue. *Anal. Biochem.* **293**, 16–21.
- Griffin, J. L., Walker, L. A., Shore, R. F. & Nicholson, J. K. 2001b Metabolic profiling of chronic cadmium exposure in the rat. *Chem. Res. Toxicol.* **14**, 1428–1434.
- Griffin, J. L., Walker, L., Shore, R. F. & Nicholson, J. K. 2001c High-resolution magic angle spinning ^1H -NMR spectroscopy studies on the renal biochemistry in the bank vole (*Clethrionomys glareolus*) and the effects of arsenic (As^{3+}) toxicity. *Xenobiotica* **31**, 377–385.
- Griffin, J. L., Mann, C., Scott, J., Shoulders, C. & Nicholson, J. K. 2001d Choline containing metabolites during cell transfection: an insight into magnetic resonance spectroscopy detectable changes. *FEBS Lett.* **509**, 263–266.
- Griffin, J. L., Williams, H. J., Sang, E. & Nicholson, J. K. 2001e Abnormal lipid profile of dystrophic cardiac tissue as demonstrated by one- and two-dimensional magic-angle spinning ^1H NMR spectroscopy. *Magn. Reson. Med.* **46**, 249–255.
- Griffin, J. L., Sang, E., Evens, T., Davies, K. & Clarke, K. 2002a Metabolic profiles of dystrophin and utrophin expression in mouse models of Duchenne Muscular dystrophy. *FEBS Lett.* **530**, 109–116.
- Griffin, J. L., Muller, D., Woograsingh, R., Jowatt, V., Hindmarsh, A., Nicholson, J. K. & Martin, J. E. 2002b Vitamin E deficiency and metabolic deficits in neuronal ceroid lipofuscinosis described by bioinformatics. *Physiol. Genomics* **11**, 195–203.
- Griffin, J. L., Bollard, M., Nicholson, J. K. & Bhakoo, K. 2002c Spectral profiles of cultured neuronal and glial cells derived from HRMAS ^1H NMR spectroscopy. *NMR Biomed.* **15**, 375–384.
- Griffin, J. L., Pole, J. C., Nicholson, J. K. & Carmichael, P. L. 2003 Cellular environment of metabolites and a metabonomic study of tamoxifen in endometrial cells using gradient high resolution magic angle spinning ^1H NMR spectroscopy. *Biochim. Biophys. Acta* **1619**, 151–158.
- Gussoni, E., Soneoka, Y., Strickland, C. D., Buzney, E. A., Khan, M. K., Flint, A. F., Kunkel, L. M. & Mulligan, R. C. 1999 Dystrophin expression in the *mdx* mouse restored by stem cell transplantation. *Nature* **401**, 390–394.
- Hall, R., Beale, M., Fiehn, O., Hardy, N., Sumner, L. & Bino, R. 2002 Meeting report: plant metabolomics: the missing link in functional genomics strategies. *Pl. Cell* **14**, 1437–1440.
- Hoffman, E. P., Beggs, A. H., Koenig, M., Kunkel, L. M. & Angelini, C. 1989 Cross-reactive protein in Duchenne muscle. *Lancet* **2**, 1211–1212.
- Holmes, E., Foxall, P. J., Nicholson, J. K., Neild, G. H., Brown, S. M., Beddell, C. R., Sweatman, B. C., Rahr, E., Lindon, J. C. & Spraul, M. 1994 Automatic data reduction and pattern recognition methods for analysis of ^1H nuclear magnetic resonance spectra of human urine from normal and pathological states. *Anal. Biochem.* **220**, 284–296.
- Holmes, E. (and 10 others) 1998 The identification of novel biomarkers of renal toxicity using automated data reduction techniques and PCA of proton NMR spectra of urine. *Chemometrics Intell. Lab. Syst.* **44**, 245–255.
- Horning, E. C. & Horning, M. G. 1971 Human metabolic profiles obtained by GC and GC/MS. *J. Chromatogr. Sci.* **9**, 129–140.
- Hotelling, H. 1933 Analysis of a complex of statistical variables into principal components. *J. Educational Psychol.* **24**, 417–441.
- Hu, J. Z., Rommereim, D. N. & Wind, R. A. 2002 High-resolution ^1H NMR spectroscopy in rat liver using magic angle turning at a 1 Hz spinning rate. *Magn. Reson. Med.* **47**, 829–836.
- Huxtable, R. J. 1992 Physiological actions of taurine. *Physiol. Rev.* **72**, 101–163.
- Kacser, H. & Burns, J. A. 1973 The control of flux. *Symp. Soc. Exp. Biol.* **27**, 65–104.
- Kato, T., Nishina, M., Matsushita, K., Hori, E., Akaboshi, S. & Takashima, S. 1997 Increased cerebral choline-compounds in Duchenne muscular dystrophy. *Neuroreport* **8**, 1435–1437.
- Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M. & Akiguchi, I. 1994 CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nature Genet.* **8**, 221–228.
- Kell, D. B. 2002 Metabolomics and machine learning: explanatory analysis of complex metabolome data using genetic programming to produce simple, robust rules. *Mol. Biol. Rep.* **29**, 237–241.
- Kemp, G. J., Taylor, D. J., Dunn, J. F., Frostick, S. P. & Radda, G. K. 1993 Cellular energetics of dystrophic muscle. *J. Neurol. Sci.* **116**, 201–206.

- Keun, H. C., Beckonert, O., Griffin, J. L., Richter, C., Moskau, D., Lindon, J. C. & Nicholson, J. K. 2002 Crogenic probe ^{13}C NMR spectroscopy of urine for metabonomic studies. *Anal. Chem.* **74**, 4588–4593.
- Khurana, T. S., Hoffman, E. P. & Kunkel, L. M. 1990 Identification of a chromosome 6-encoded dystrophin-related protein. *J. Biol. Chem.* **265**, 16 717–16 720.
- Khurana, T. S., Watkins, S. C., Chafey, P., Chelly, J., Tome, F. M., Fardeau, M., Kaplan, J. C. & Kunkel, L. M. 1991 Immunolocalization and developmental expression of dystrophin related protein in skeletal muscle. *Neuromusc. Disord.* **1**, 185–194.
- Klose, J. (and 12 others) 2002 Genetic analysis of the mouse brain proteome. *Nature Genet.* **30**, 385–393.
- Kohonen, T. 1982 Self-organizing formation of topologically correct feature maps. *Biol. Cybern.* **43**, 59–69.
- Kose, F., Weckwerth, W., Linke, T. & Fiehn, O. 2001 Visualizing plant metabolomics correlation networks using clique-metabolite matrices. *Bioinformatics* **17**, 1198–1208.
- Krag, T. O., Gyrd-Hansen, M. & Khurana, T. S. 2001 Harnessing the potential of dystrophin-related proteins for ameliorating Duchenne's muscular dystrophy. *Acta Physiol. Scand.* **171**, 349–358.
- Lachenbruch, P. A. 1975 *Discriminate analysis*. New York: Hafner Press.
- Lederfein, D., Levy, Z., Augier, N., Mornet, D., Morris, G., Fuchs, O., Yaffe, D. & Nudel, U. 1992 A 71-kilodalton protein is a major product of Duchenne muscular dystrophy gene in brain and other non-muscle tissue. *Proc. Natl Acad. Sci. USA* **89**, 5346–5350.
- Lindon, J. C. (and 25 others) 2003 The role of metabonomics in toxicology and its evaluation by the COMET project. *Toxicol. Appl. Pharmacol.* **187**, 137–146.
- Liu, C.-J., Blout, J. W., Steele, C. L. & Dixon, R. A. 2002 Bottlenecks for metabolite engineering of isoflavone glycoconjugates in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 14 578–14 583.
- Lodi, R., Cooper, J. M., Bpooradley, J. L., Manners, D., Styles, P., Taylor, D. J. & Schapira, A. H. 1999 Deficit of *in vivo* mitochondrial ATP production in patients with Friedreich ataxia. *Proc. Natl Acad. Sci. USA* **96**, 11 492–11 495.
- McIntosh, L. M., Garrett, K. L., Megeney, L., Rudnicki, M. A. & Anderson, J. E. 1998a Regeneration and myogenic cell proliferation correlate with taurine levels in dystrophin- and MyoD-deficient muscles. *Anat. Rec.* **252**, 311–324.
- McIntosh, L. M., Baker, R. E. & Anderson, J. E. 1998b Magnetic resonance imaging of regenerating and dystrophic mouse muscle. *Biochem. Cell Biol.* **76**, 532–541.
- Mole, S. E., Mitchison, H. M. & Munroe, P. B. 1999 Molecular basis of the neuronal ceroid lipofuscinoses: mutations in CLN1, CLN2, CLN3 and CLN5. *Hum. Mutat.* **14**, 199–215.
- Mokhtarian, A., Decrouy, A., Chinnet, A. & Even, P. C. 1996 Components of energy expenditure in the *mdx* mouse model of Duchenne muscular dystrophy. *Pflügers Arch.* **431**, 527–532.
- Moolenaar, S. H., Engelke, U. F. & Wevers, R. 2003 Proton nuclear magnetic resonance spectroscopy of body fluids in the field of inborn errors of metabolism. *Ann. Clin. Biochem.* **40**, 16–24.
- Nicholson, J. K., Hingham, D. P., Timbrell, J. A. & Sadler, P. J. 1989 Quantitative high resolution ^1H NMR urinalysis studies on the biochemical effects of cadmium in the rat. *Mol. Pharmacol.* **36**, 398–404.
- Nicholson, J. K., Holmes, E. & Lindon, J. 1999 Metabonomics. *Xenobiotica* **29**, 1181–1189.
- Nicholson, J. K., Connelly, J., Lindon, J. C. & Holmes, E. 2002 Metabonomics: a platform for studying drug toxicity and gene function. *Nature Rev. Drug Discov.* **1**, 153–161.
- Nicholson, J. K. & Wilson, I. 2003 Understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nature Rev. Drug Discov.* **2**, 668–676.
- Nudel, U., Zuk, D., Einat, P., Zeelon, E., Levy, Z., Neuman, S. & Yaffe, D. 1989 Duchenne muscular dystrophy gene product is not identical in muscle and brain. *Nature* **337**, 76–78.
- Oliver, S. G. 1996 From DNA sequence to biological function: the new voyage of the Beagle. *Biochem. Soc. Trans.* **24**, 291–292.
- Oliver, S. G. 2002 Functional genomics: lessons from yeast. *Phil. Trans. R. Soc. Lond. B* **357**, 17–23. (DOI 10.1098/rstb.2001.1049.)
- Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. 1998 Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **16**, 373–378.
- Pichersky, E. & Gang, D. R. 2000 Genetics and biochemistry of secondary metabolites: an evolutionary perspective. *Trends Pl. Sci.* **5**, 439–445.
- Raamsdonk, L. M. (and 12 others) 2001 A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnol.* **19**, 45–50.
- Rae, C., Scott, R. B., Thompson, C. H., Dixon, R. M., Dumughn, I., Kemp, G. J., Male, A., Pike, M., Styles, P. & Radda, G. K. 1998 Brain biochemistry in Duchenne muscular dystrophy: a ^1H magnetic resonance and neuropsychological study. *J. Neurol. Sci.* **160**, 148–157.
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabasi, A. L. 2002 Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555.
- Robertson, D. G., Reily, M. D., Sigler, R. E., Wells, D. F., Paterson, D. A. & Braden, T. K. 2000 Metabonomics: evaluation of nuclear magnetic resonance (NMR) and pattern recognition technology for rapid *in vivo* screening of liver and kidney toxicants. *Toxicol. Sci.* **57**, 326–337.
- Roessner, U., Wagner, C., Kopka, J., Trethewey, R. N. & Willmitzer, L. 2000 Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. *Pl. J.* **23**, 131–142.
- Rosenberg, R. M. 1992 Machado–Joseph disease: an autosomal dominant motor system degeneration. *Mov. Disord.* **7**, 193–203.
- Rybakova, I. N., Patel, J. R., Davies, K. E., Yurchenco, P. D. & Ervasti, J. M. 2002 Utrophin binds laterally along actin filaments and can couple costameric actin with sarcolemma when overexpressed in dystrophin deficient muscle. *Mol. Biol. Cell* **13**, 1512–1521.
- Schoolnik, G. K. 2002 Microarray analysis of bacterial pathogenicity. *Adv. Microb. Physiol.* **46**, 1–45.
- Shockcor, J. P. & Holmes, E. 2002 Metabonomic applications in toxicity screening and disease diagnosis. *Curr. Top. Med. Chem.* **2**, 35–51.
- Smith, L. L. 2001 Key challenges for toxicologists in the 21st century. *Trends Pharmacol. Sci.* **22**, 281–285.
- Stoyanova, R. & Brown, T. R. 2001 NMR spectral quantification by principal component analysis. *Nucl. Magn. Reson. Biomed.* **14**, 271–277.
- Stoyanova, R. & Brown, T. R. 2002 NMR spectral quantification by principal component analysis. III. A generalized procedure for determination of lineshape variations. *J. Magn. Reson.* **154**, 163–175.
- Styles, P., Soffe, N. F., Scott, C. A., Cragg, D. A., Row, F., White, D. J. & White, P. C. J. 1984 A high resolution NMR probe in which the coil and preamplifier are cooled with liquid helium. *J. Magn. Reson.* **60**, 397–404.
- Sze, D. & Jardetzky, O. 1994 High-resolution proton NMR studies of lymphocyte extracts. *Immunomethods* **4**, 113–126.

- Takiyama, Y., Oyanagi, S., Kawashima, S., Sakamoto, H., Saito, K., Yoshida, M., Tsuji, S., Mizuno, Y. & Nishizawa, M. 1994 A clinical and pathologic study of a large Japanese family with Machado-Joseph disease tightly linked to the DNA markers on chromosome 14q. *Neurology* **44**, 1302–1308.
- Tate, A. R., Crabb, S., Griffiths, J. R., Howells, S. L., Mazucco, R. A., Rodrigues, L. M. & Watson, D. 1996 Lipid metabolite peaks in pattern recognition analysis of tumour *in vivo* MR spectra. *Anticancer Res.* **16**, 1575–1579.
- Tate, A. R. (and 14 others) 1998 Towards a method for automated classification of ^1H MRS spectra from brain tumours. *Nucl. Magn. Reson. Biomed.* **11**, 177–191.
- Teusink, B., Walsh, M. C., van Dam, K. & Westerhoff, H. V. 1998 The danger of metabolic pathways with turbo design. *Trends Biochem. Sci.* **23**, 162–169.
- Tinsley, J., Deconinck, N., Fisher, R., Kahn, D., Phelps, S., Gillis, J. M. & Davies, K. 1998 Expression of full-length utrophin prevents muscular dystrophy in *mdx* mice. *Nature Med.* **4**, 1441–1444.
- Tracey, I., Dunn, J. F. & Radda, G. K. 1996 Brain metabolism is abnormal in the *mdx* mouse model of Duchenne muscular dystrophy. *Brain* **119**, 1039–1044.
- Tweeddale, H., Notley-McRobb, L. & Ferenci, T. 1998 Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ('metabolome') analysis. *J. Bacteriol.* **180**, 5109–5116.
- Valafar, F. 2002 Pattern recognition techniques in microarray data analysis. *Ann. NY Acad. Sci.* **980**, 41–64.
- Vance, J. E., Stone, S. J. & Faust, J. R. 1997 Abnormalities in mitochondria-associated membranes and phospholipid biosynthetic enzymes in the *mnd/mnd* mouse model of neuronal ceroid lipofuscinosis. *Biochim. Biophys. Acta* **1344**, 286–299.
- Wang, B., Li, J. & Xiao, X. 2000 Adeno-associated virus vector carrying human minidystrophin genes effectively ameliorates muscular dystrophy in *mdx* mouse model. *Proc. Natl Acad. Sci. USA* **97**, 13 714–13 719.
- Waters, N. J., Holmes, E., Williams, A., Waterfield, C. J., Farrant, R. D. & Nicholson, J. K. 2001 NMR and pattern recognition studies on the time-related metabolic effects of alpha-naphthylisothiocyanate on liver, urine, and plasma in the rat: an integrative metabonomic approach. *Chem. Res. Toxicol.* **14**, 1401–1412.
- Waters, N. J., Holmes, E., Waterfield, C. J., Farrant, R. D. & Nicholson, J. K. 2002 NMR and pattern recognition studies on liver extracts and intact livers from rats treated with alpha naphthylisothiocyanate. *Biochem. Pharmacol.* **64**, 67–77.
- Weybright, P., Millis, K., Campbell, N., Cory, D. G. & Singer, S. 1998 Gradient, high-resolution, magic angle spinning ^1H nuclear magnetic resonance spectroscopy of intact cells. *Magn. Reson. Med.* **39**, 337–345.
- Wind, R. A., Zhi Hu, J. & Rommereim, D. N. 2001 High-resolution ^1H NMR spectroscopy in organs and tissues using slow magic angle spinning. *Magn. Reson. Med.* **46**, 213–218.
- Winzeler, E. A. (and 51 others) 1999 Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906.
- World Health Organisation 1992 *Environmental health criteria 134: cadmium*. Geneva: WHO.