XENOBIOTICA , 1999, VOL. 29, NO. 11, 1181–1189

'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data

J. K. NICHOLSON*, J. C. LINDON and E. HOLMES

Biological Chemistry, Biomedical Sciences Division, Imperial College of Science, Technology and Medicine, University of London, Sir Alexander Fleming Building, South Kensington, London SW7 2AZ, UK

Received 5 July 1999

Introduction

The rapid evolution of drug discovery science, fuelled by combinatorial library-based synthesis programmes, has led to increased pressure on the drug safety evaluation process. Once potential drugs have passed the primary biological screening procedures, losses of drug candidate compounds from the product development pipeline (known as 'attrition') need to be minimized. Hence, there is an intensive search for new analytical technologies that will maximize efficiency of lead compound selection based both on efficacy and safety and will minimize overall attrition rates. Current bioanalytical approaches include measurements of responses of living systems to drugs either at the genetic level or at the level of expression of cellular proteins, using so-called genomic and proteomic methods respectively. At present both genomics and proteomics are expensive and labour-intensive, yet potentially are powerful tools for studying different levels of the biological response to xenobiotic exposure. However, even in combination, genomics and proteomics do not provide the range of information needed for an understanding of the integrated cellular function in living systems, since both ignore the dynamic metabolic status of the whole organism. Thus, a new NMR-based 'metabonomic' approach is proposed that is aimed at the augmentation and complementation of the information provided by measuring the genetic and proteomic responses to xenobiotic exposure. Metabonomics is defined as 'the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification'. This concept has arisen from work on the application of ^{1}H-NMR spectroscopy to study the multicomponent metabolic composition of biofluids, cells and tissues over the past two decades (e.g. Nicholson *et al*. 1983, 1985, Bales *et al*. 1984, Gartland *et al*. 1989, Nicholson and Wilson 1989, Moka *et al*. 1998). Also studies utilizing pattern recognition (PR), expert systems and related bio-informatic tools are used to interpret and classify complex NMR-generated metabolic data sets (Gartland *et al*. 1991, Holmes *et al*. 1992, 1994, 1998a, b, Anthony *et al*. 1994, Spraul *et al*. 1997, Beckwith-Hall *et al*. 1998). There is also a significant background to this work in other research fields, notably metabolic control analysis (Kacser and Burns 1973, Kacser 1993, Goodacre *et al*. 1996), and there is a related concept of the 'Metabolome' that represents the total small molecule complement of a cell. However, metabonomics deals with detecting,

* Author for correspondence.

Xenobiotica ISSN 0049–8254 print/ISSN 1366–5928 online © 1999 Taylor & Francis Ltd
http://www.tandf.co.uk/JNLS/xen.htm http://www.taylorandfrancis.com/JNLS/xen.htm

identifying, quantitating and cataloguing the *history* of time-related metabolic changes in an integrated biological system rather than the individual cell. Such multidimensional metabolic trajectories are then related to the biological events in an ongoing pathophysiological process. Here, provided is a brief background to the useful properties of metabonomic data sets and the possible uses of NMR-based metabonomics for toxicological classification and biomarker or surrogate marker identification *in vivo*.

Genomic and proteomic approaches to drug toxicity assessment

Development of new tools in structural molecular biology has led to an increased understanding of the organization of the genome. This knowledge combined with a massive increase in the ability to identify and sequence genes has led to the point where the entire genome of >20 prokaryotic organisms, e.g. *Archaeoglobus fulgidus* (Klenk *et al*. 1997), has already been sequenced together with one eukaryotic organism with ~ 19000 genes and $>93 \times 10^6$ bp (*Caenorhabditis elegans*; The *C. elegans* Sequencing Consortium 1998). A complete description of the human genome with ~ 80000 genes is probably only a few years away. One of the intellectual products of the molecular biology revolution has been the concept of 'genomics', which is basically a semiquantitative approach to the measurement of gene expression. In the context of drug discovery and for the purposes of toxicological assessment, the genomic approach involves the observation of altered gene expression after drug exposure. The technology involves a new generation of proprietary 'gene chips', which are small disposable devices encoded with an array of genes that respond to extracted cellular mRNA produced after exposure to a foreign compound which has caused the 'switching on' of various genes (Sinclair 1999). Many genes can be placed on a chip array and patterns of gene switching caused by xenobiotic exposure can be monitored rapidly in this way, although at some considerable cost. However, relationships between gene regulation/expression and the integrated function and control of cellular systems (so-called functional genomics) are still far from clear, and will remain so for many years after the complete sequencing of the human genome. The main reason for this is that the vast majority of DNA is non-coding, yet protein coding sequences or genes cannot function as isolated units and can require the presence of neighbouring genes and/or non-coding DNA. The lack of understanding of the biological consequences of altered gene expression has led to the development of proteomics, which is concerned with the semiquantitative measurement of the production of cellular proteins in response to drug exposure and other pathophysiological processes (Anderson *et al*. 1996, Aicher *et al*. 1998, Geisow 1998). Proteomic measurements utilize a variety of technologies, but all involve a protein separation method, e.g. 2D gel-electrophoresis, allied to a chemical characterization method, usually, some form of mass spectrometry (MS). While potentially less expensive than genomics, proteomics is very slow and labour-intensive at present. More importantly, although these measurements may ultimately give profound insights into toxicological mechanisms and provide new surrogate biomarkers of disease, at present it is very difficult to relate genomic and proteomic findings to classical indices of toxicity or toxicological end-points. One simple reason for this is that the current technology and approach precludes the measurement of a detailed time-course of the response to drug exposure or the measurement of responses in a multi-organ system. This may be particularly important for the many known

cases where the metabolism of the compound is a prerequisite for toxicity and especially true where the target organ is not the site of primary metabolism. An example is the case of compounds that form glutathione S-conjugates in the liver that are subsequently processed by β-lyase thus generating reactive intermediates that show ultimate target organ toxicity in the renal proximal tubules (Elfarra *et al*. 1986). There is a need for the development of novel methods that give information of *in vivo* multi-organ functional integrity in real time. NMR-based metabonomics offers one such approach to the generation of this type of information.

NMR-based metabonomics

Foreign compounds may interact with tissue and extracellular components of an animal at a series of organizational levels ranging from changes in genetic expression through protein production and integrated cellular biochemical regulation and control. In such cases there will be alterations detectable at all levels of biomolecular organization and a complete approach to the description of these changes might be termed as 'bionomics' (proposed by Professor Ian D. Wilson). In many cases, drugs exert their toxic effects by interacting directly with genetic material or by inducing the synthesis of drug metabolizing enzymes, which generate toxic products. In such cases genomic and proteomic approaches to toxicity assessment may be useful. However, xenobiotics may act only at the pharmacological level and, hence, may not affect gene regulation or expression. Also significant toxicological effects may be completely unrelated to gene switching or protein synthesis. Exposure to ethanol *in vivo* may switch on many genes, but this does not explain drunkenness! Hence, in many cases facile consideration of genomic and proteomic responses are likely to be ineffective at predicting drug toxicity. However, all drug-induced pathophysiological perturbations result in disturbances in the ratios and concentrations, binding or fluxes of endogenous biochemicals, either by direct chemical reaction or by binding to key enzymes or nucleic acids that control metabolism. If these disturbances are of sufficient magnitude, toxic effects will result that will affect the efficient functioning of the whole organism. In body fluids, metabolites are in dynamic equilibrium with those inside cells and tissues and, consequently, abnormal cellular processes in tissues of the whole organism following a toxic or metabolic insult will be reflected in altered biofluid compositions. In all cases the analytical problem usually involves the detection of 'trace' amounts of analytes in a very complex matrix with many potential interferences. It is critical, therefore, to choose a suitable analytical technique for the particular class of analyte of interest in the biomatrix, for example blood, plasma, urine, bile or organ samples. High-resolution ^{1}H-NMR spectroscopy appears particularly appropriate for investigating abnormal body fluid compositions as a wide range of metabolites can be quantified simultaneously with no sample preparation and 'without prejudice'. Other techniques such as MS may also be useful for generating metabolic data, but differential ionization efficiency in the complex could affect detectability and quantitiation. NMR spectroscopy may also be used effectively to screen for abnormal metabolite profiles in tissue extracts or cell suspensions. It has also been shown that the same approach can be used to investigate the metabolic composition of *intact* tissues using high-resolution magic angle spinning ^{1}H-NMR spectroscopy (Moka *et al*. 1998).

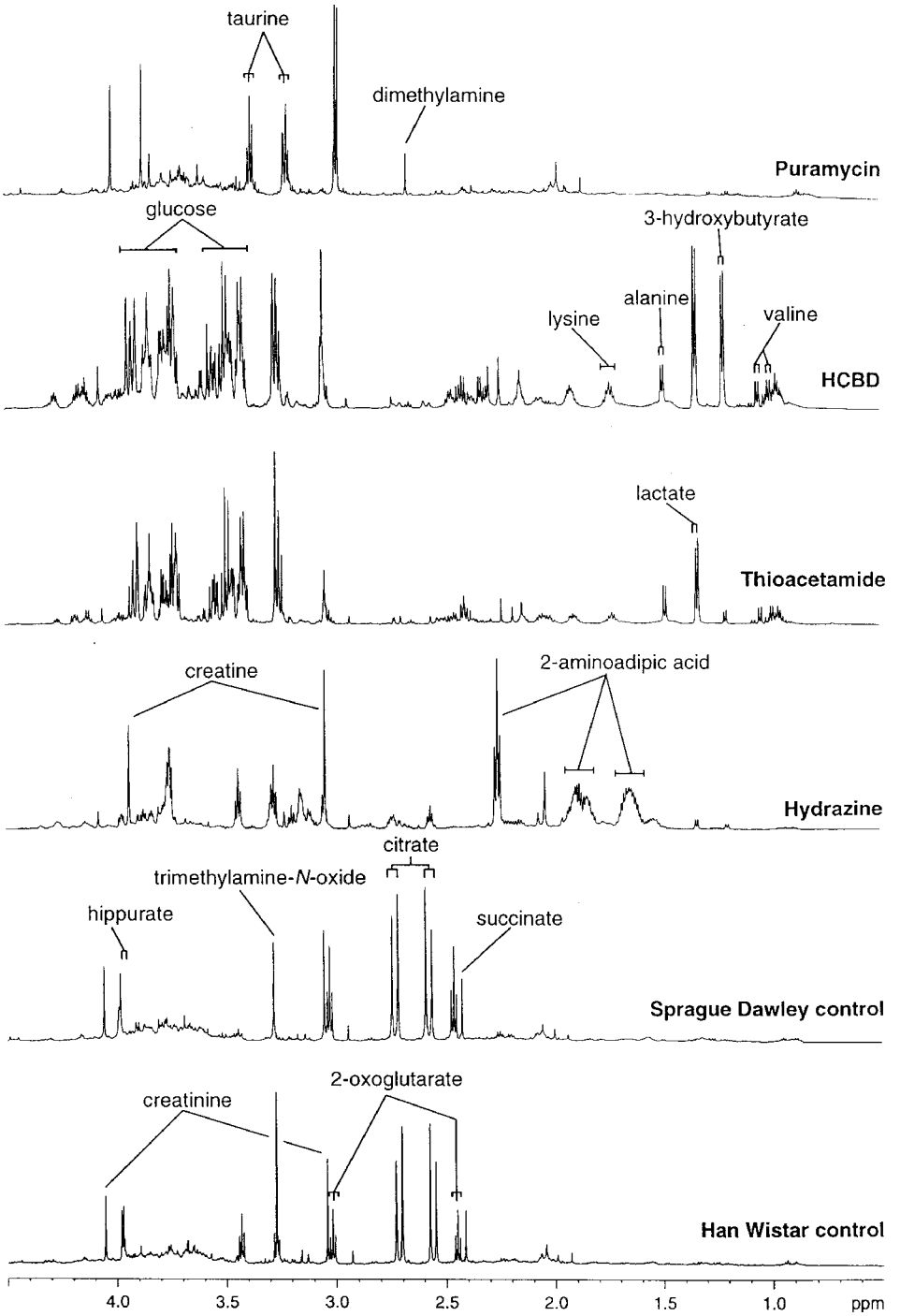

Figure 1. Partial 600 MHz ^{1}H-NMR spectra of a series of urines from the control rat, and those collected 8–24 h after treatment with various model toxins. HCBD, hexachloro-1,3-butadiene.

The exact pattern of endogenous metabolites in body fluids as detected by ^{1}H-NMR spectroscopy depends strongly on the type of toxin to which an animal has been exposed (Nicholson *et al*. 1983, 1985, Bales *et al*. 1984, Gartland *et al*. 1989, Nicholson and Wilson 1989). Each class of toxin produces characteristic changes in the concentrations and patterns of endogenous metabolites in biofluids and this provides information on the sites and basic mechanisms of the toxic process. A typical series of spectra from urine of rat treated with different toxins are shown in figure 1. Bio-analytically, the processes of generating such information is highly efficient, taking only a few minutes per sample and requiring little or no sample pretreatment or reagents. The spectra are very similar in the case of controls (two common models the Han Wistar and Sprague Dawley being shown), but different toxins cause characteristic metabolic perturbations. Because nearly all major classes of metabolic intermediate have characteristic NMR spectra, the technique is very useful for fingerprinting toxin-induced metabolic variations. Thus, ^{1}H-NMR spectroscopic analysis of biofluids has successfully uncovered numerous novel metabolic biomarkers of organ-specific toxicity in the rat, and it is in this 'exploratory' role that NMR as an analytical biochemistry technique excels. For example, changes in the levels of trimethylamine-*N*-oxide, *N*,*N*-dimethylglycine, dimethylamine and succinate are indicative of damage to the renal papilla for which no biochemical biomarkers existed previously (Gartland *et al*. 1989, 1991). Other urinary markers uncovered by ^{1}H-NMR urinalysis include taurine and creatine, which have been correlated with acute liver and testicular toxicity respectively (Nicholson *et al*. 1989, Gray *et al*. 1990, Sanins *et al*. 1990). Similar approaches can be used using 2D NMR spectroscopy (Nicholson and Wilson 1989). However, the biomarker information in NMR spectra of biofluids is much more subtle and rich than this, as hundreds of compounds representing many pathways can often be measured simultaneously, and it is the overall metabonomic response to toxic insult (occurring over time) that so well characterizes the lesion (Beckwith-Hall *et al*. 1998, Holmes *et al*. 1998a). The most efficient way to investigate these complex multiparametric data is to continue the 1D and 2D NMR metabonomic approach with PR methods.

Pattern recognition and expert system analysis of NMR-generated metabonomic data

A limiting factor in understanding the biochemical information from both 1D and 2D NMR spectra of tissues and biofluids is their very complexity; even 1D ^{1}H-NMR spectra (at 600 MHz or above) of biofluids may contain several thousand resolved lines. The NMR spectrum of a sample under study can be considered as an *n*-dimensional object the dimensions of which could be the concentrations of individual measurable metabolites or more simply the spectral intensity distribution. Thus, the NMR spectrum of the biofluid or tissue provides an *n*-dimensional metabolic fingerprint of the organism based on the sample studied, and this metabolic profile is characteristically changed according to the disease or toxic process. Hence, computer-based PR and expert system approaches have been used to interpret the NMR data obtained in various experimental toxicity states (Gartland *et al*. 1991, Holmes *et al*. 1992, 1994, 1998a, b, Anthony *et al*. 1994, Spraul *et al*. 1997, Beckwith-Hall *et al*. 1998). These statistical tools are very similar to those currently being explored by those in the fields of genomics and proteomics. The

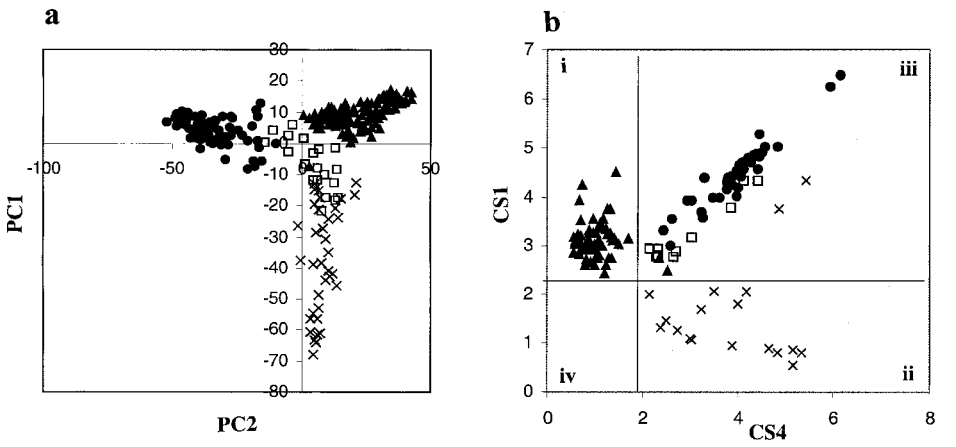

Figure 2. (a) Principal components map of data obtained from rat urines after treatment with lead acetate (□) hydrazine (×) and renal proximal tubular toxins affecting the S3 region (●) and controls (▲). (b) Cooman's residuals plot of test data set using a SIMCA model previously 'trained' using the same spectra shown in (a). Quadrant (i) shows samples unambiguously classified as controls, quadrant (ii) shows 'pure' hydrazine-toxicity classification, quadrant (iii) shows spectra from animals classified as neither control nor hydrazine-treated type, and quadrant (iv) shows an unoccupied field that would indicate mixed hydrazine-toxicity and control classification. In this example, two hydrazine-treated data points are misclassified and two controls are also misclassified as abnormal samples. The lines show the 95% confidence limits of the classifications based on the training set data.

simplest approach is to treat the NMR signal intensity data as a multi-sample array of metabolite concentration or excretion rate scores; it is not necessary to assign the spectrum at this stage as it is treated solely as a statistical object. PR is a general term applied to methods of data analysis that can be used to generate scientific hypotheses as well as *testing* hypotheses by reducing mathematically the many parameters. One of the most useful and easily applied PR techniques is principal components analysis (PCA). Principal components (PC) are new variables created from linear combinations of the starting variables with appropriate weighting coefficients. The properties of these PC are such that (1) each PC is orthogonal (uncorrelated) with all other PC and (2) the first PC contains the largest part of the variance of the data set (information content) with subsequent PC containing correspondingly smaller amounts of variance. Thus, a plot of the first two or three PC gives the 'best' representation, in terms of biochemical variation in the data set in two or three dimensions. Such PC maps can be used to visualize inherent clustering behaviour for drugs and toxins acting on each organ according to toxic mechanism (Nicholson and Wilson 1989, Gartland *et al.* 1991). Such an application of PCA to toxicological mapping of NMR-generated metabonomic data is shown in figure 2a in which there is distinct clustering of data points from the urines of individual animals exposed to different toxins. The position on a PC plot of a sample from a xenobiotic-treated animal is determined purely by its metabolic response as opposed to any other independent knowledge of the compound action; hence, the method is termed 'unsupervised'. Of course, the clustering information might be in lower PC and this also has to be examined. In this simple metabonomic approach a sample from an animal treated with a compound of unknown toxicity is compared with a database of NMR-generated metabolic data and its topographical fit on the PR map is determined (Holmes *et al.* 1998a, b). However, in the real world, toxicological data

are more complex as lesions develop and resolve in real time and, hence, there are time-related changes in NMR-detected metabolic profile (Holmes *et al*. 1992, Beckwith-Hall *et al*. 1998). Also, it is more rigorous to compare effects of xenobiotics in the original *n*-dimensional NMR metabonomic space. Hence, as an alternative approach and to develop automatic toxicity classification methods, it has proved efficient to use a 'supervised' approach to NMR data analysis. Here, a 'training set' of NMR metabonomic data is used to construct a mathematical model that predicts correctly the class of each sample. This training set is then tested with independent data ('test set') to determine the robustness of the computer-based model. These models are sometimes termed expert systems, but may comprise systems based on a range of different mathematical procedures such as principal components, artificial neural networks and rule induction. In all cases the methods allow the quantitative description of the multivariate boundaries that characterize and separate each class of xenobiotic in terms of their metabolic effects. Certain supervised methods, such as SIMCA (soft independent modelling of class analogy; Kowalski *et al*. 1986) also allow a level of probability to be placed on the goodness of fit. Using such systems a sample can be classified as belonging to a single class of toxicity, to multiple classes of toxicity (more than one target organ) or to no class. The latter case would indicate deviation from normality (control) based on the training set model but having a dissimilar metabolic effect to any toxicity class modelled in the training set (unknown toxicity type). An example of an expert systems based classification of toxicity data is shown in figure 2b. In this simple illustrative case SIMCA models were constructed for both control rat urines and for rat urines from hydrazine-dosed animals using a training set of NMR data. The Cooman's residuals plot shown in figure 2b demonstrates that the majority of the test controls and test hydrazine-treated spectra are correctly classified and S3 type renal cortical toxins and lead acetate (which causes a range of renal, haemopoeitic and hepatotoxic effects) are all correctly classified as neither control nor hydrazine type. By building an exhaustive series of models it is possible to use SIMCA and other methods to provide classification probabilities for a wide range of toxicity types.

The metabonomic expert systems currently under construction in our group can be considered to operate at three distinct levels of pathophysiological discrimination:

1. Classification of the sample or organism as 'normal or abnormal' according to metabonomic criteria derived from a large database of controls (this will be a useful tool in the control of NMR spectrometer automation using sequential flow injection NMR spectroscopy; Spraul *et al*. 1997).
2. Classification of the target organ for toxicity and site of action within the tissue.
3. Identification of the biomarkers of toxic effect and toxic mechanism classification for the compound under study.

Interestingly, these levels of classification or discrimination would also apply even if data were derived from genomic or proteomic studies and similar arguments could be applied to clinical diagnostic screening procedures. As the size of toxicological databases increases together with improvements in rapid throughput of NMR samples (300 samples per day per spectrometer is now possible with the first generation flow injection systems), more subtle expert systems will be necessary using techniques such as 'fuzzy logic', which permits greater flexibility in decision boundaries between classes. Using the metabonomic methods described above, it has already been possible to develop a prototype expert system for classification

at level 1, and has also effected level 2 classification procedures for a range of toxicological endpoints and target organs. The level 3 classification poses more complex problems in terms of expert system development, but detailed biomarker information can already be obtained from inspection of the PC loadings (Holmes *et al*. 1998b).

In conclusion, there is a vast range of biochemical, toxicological and clinical chemical problems that can be addressed using metabonomics based on high-resolution ^{1}H-NMR spectroscopy of biomaterials. At present even simple ^{1}H-NMR experiments on whole biofluids can generate substantial amounts of metabolic data that can give surprisingly detailed insight into the biochemical processes in the whole organisms and the investigation of species differences in terms of toxicological biomarkers. The numbers of applications of metabonomics is bound to increase in parallel with ongoing developments in instrumentation and techniques. In particular, the development of computer-based PR and expert systems for data analysis is expected to make major contributions to the advancement of NMR-based metabolic science. Other important areas accessible to metabonomic investigation include studies on biochemical consequences of genetic modification, e.g. in 'knock-out animals', investigations into effects of environmental pollutants, for clinical evaluation of drug therapy and efficacy, and the investigation of idiosyncratic toxicity in man. Finally, it should soon be possible to combine genomic, proteomic and metabonomic data sets into comprehensive 'bionomic' systems for the holistic evaluation of perturbed *in vivo* function.

Acknowledgements

The authors thank Professor Ian Wilson, AstraZeneca Pharmaceuticals, Alderley Park, (who also coined the term 'bionomics') and Professor Jeremy Everett, Pfizer Pharmaceuticals, Sandwich, for helpful discussions on the background and basic philosophy of the work on metabonomics over several years.

References

Aicher, L., Wahl, D., Arce, A., Grenet, O. and Steiner, S., 1998, New insights into cyclosporine A nephrotoxicity by proteome analysis. *Electrophoresis*, **19**, 1998–2003.

Anderson, N. L., Taylor, J., Hofmann, J. P., Esquer-Blasco, R., Swift, S. and Anderson, N. G., 1996, Simultaneous measurement of hundreds of liver proteins: application in assessment of liver function. *Toxicologic Pathology*, **24**, 72–76.

Anthony, M. L., Sweatman, B. C. Beddell, C. R., Lindon, J. C. and Nicholson, J. K., 1994, Pattern recognition classification of the site of nephrotoxicity based on metabolic data derived from high resolution proton nuclear magnetic resonance spectra of urine. *Molecular Pharmacology*, **46**, 199–211.

Bales, J. R., Higham, D. P., Howe, I., Nicholson, J. K. and Sadler, P. J., 1984, Use of high resolution proton nuclear magnetic resonance spectroscopy for rapid multi-component analysis of urine. *Clinical Chemistry*, **30**, 426–432.

Beckwith-Hall, B. M., Nicholson, J. K., Nicholls, A., Foxall, P. J. D., Lindon, J. C., Connor, S. C., Abdi, M., Connelly, J. and Holmes, E., 1998, Nuclear magnetic resonance spectroscopic and principal components analysis investigations into biochemical effects of three model hepatotoxins. *Chemical Research in Toxicology*, **11**, 260–272.

Elfarra, A. A., Jakobson, I. and Anders, M. W., 1986, Mechanism of S-(1,2-dichlorovinyl) glutathione-induced nephrotoxicity *Biochemical Pharmacology*, **35**, 283–288.

Gartland, K. P. R., Beddell, C., Lindon, J. C. and Nicholson, J. K., 1991, The application of pattern recognition methods to the analysis and classification of toxicological data derived from NMR spectroscopy of urine. *Molecular Pharmacology*, **39**, 629–642.

Gartland, K. P. R., Bonner, F. and Nicholson, J. K., 1989, Investigations into the biochemical effects of region-specific nephrotoxins. *Molecular Pharmacology*, **35**, 242–251.

Geisow, M. J., 1998, Proteomics: one small step for a digital computer, one giant leap for humankind. *Nature Biotechnology*, **16**, 206.

GOODACRE, R. RISCHERT, D. J., EVANS, P. M. and KELL, D. B., 1996, Rapid authentication of animal cell lines using pyrolysis mass spectrometry and autoassociative artificial neural networks. Cytotechnology, **21**, 231–241.

GRAY, J., NICHOLSON, J. K., CREASY, D. M. and TIMBRELL, J. A., 1990, Studies on the relationship between testicular toxicity and urinary and plasma creatine concentration. *Archives of Toxicology*, **64**, 443–450.

HOLMES, E., BONNER, F. W., SWEATMAN, B. C., LINDON, J. C., BEDDELL, C. R., RAHR, E. and NICHOLSON, J. K., 1992, NMR spectroscopy and pattern recognition analysis of the biochemical processes associated with the progression and recovery from nephrotoxic lesions in the rat induced by mercury (II) chloride and 2-bromoethanamine. *Molecular Pharmacology*, **42**, 922–930.

HOLMES, E., FOXALL, P. J. D., NICHOLSON, J. K., NEILD, G. H., BROWN, S. M. BEDDELL, C., SWEATMAN, B. C., RAHR, E., LINDON, J. C., SPRAUL, M. and NEIDIG, P., 1994, Automatic data reduction and pattern recognition methods for analysis of ^{1}H nuclear magnetic resonance spectra of human urine form normal and pathological states. *Analytical Biochemistry*, **220**, 284–296.

HOLMES, E., NICHOLLS, A. W., LINDON, J. C., RAMOS, S., SPRAUL, M., NEIDIG, P., CONNOR, S. C., CONNELLY, J., DAMMENT, S. J. P., HASELDEN, J. N. and NICHOLSON, J. K., 1998a, Development of a model for classification of toxin-induced lesions using ^{1}H-NMR spectroscopy of urine combined with pattern recognition. *NMR in Biomedicine*, **11**, 1–10.

HOLMES, E., NICHOLSON, J. K., NICHOLLS, A. W., LINDON, J. C., CONNOR, S. C., POLLY, S. and CONNELLY, J., 1998b, Identification of novel biomarkers of renal toxicity using automatic data reduction techniques and PCA of proton NMR spectra of urine. *Chemometrics and Intelligent Laboratory Systems*, **44**, 251–261.

KACSER, H., 1993, Recent developments beyond metabolic control analysis. *Biochemical Society Transactions*, **23**, 387–391.

KACSER, H. and BURNS, J. A., 1973, The control of flux. In D. D. Davies (ed.), *Rate Control of Biological Processes. Symposium of the Society for Experimental Biology*, Vol. 27 (Cambridge: Cambridge University Press), pp. 65–104.

KLENK, H. P. *et al.*, 1997, The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*. *Nature*, **390**, 364–370.

KOWALSKI, B., SHARAF, D. and ILLMAN, D., 1986, *Chemometrics* (New York: Wiley).

MOKA, D., VORREUTHER, R., SHICHA, H., HUMPFER, E., LIPINSKI, M., SPRAUL, M., FOXALL, P. J. D., NICHOLSON, J. K. and LINDON, J. C., 1998, Biochemical classification of kidney carcinoma biopsy samples using magic angle spinning ^{1}H-NMR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, **17**, 125–132.

NICHOLSON, J. K., BUCKINGHAM, M. J. and SADLER, P. J., 1983, High resolution proton NMR studies of vertebrate blood and plasma. *Biochemical Journal*, **211**, 605–615.

NICHOLSON, J. K., HIGHAM, D., TIMBRELL, J. A. and SADLER, P. J., 1989, Quantitative ^{1}H-NMR urinalysis studies on the biochemical effects of acute cadmium exposure in the rat. *Molecular Pharmacology*, **36**, 398–404.

NICHOLSON, J. K., TIMBRELL, J. A. and SADLER, P. J., 1985, Proton NMR spectra of urine as indicators of renal damage: Mercury nephrotoxicity in rats. *Molecular Pharmacology*, **27**, 644–651.

NICHOLSON, J. K. and WILSON, I. D., 1989, High resolution proton NMR spectroscopy of biological fluids. *Progress in NMR Spectroscopy*, **21**, 449–501.

SANINS, S. M., TIMBRELL, J. A., ELCOMBE, C. R. and NICHOLSON, J. K., 1990, Hepatotoxin-induced hypertaurinuria: a proton NMR study. *Archives of Toxicology*, **64**, 407–411.

SINCLAIR, B., 1999, Everything 's great when it sits on a chip: a bright future for DNA arrays. *The Scientist*, **13**, 18–20.

SPRAUL, M., HOFMANN, M., ACKERMANN, M., NICHOLLS, A. W., DAMMENT, S. J. P., HASELDEN, J. N., SHOCKCOR, J. P., NICHOLSON, J. K. and LINDON, J. C., 1997, Flow injection ^{1}H-NMR spectroscopy combined with pattern recognition: implications for rapid structural studies and high throughput biochemical screening. *Analytical Communications*, **34**, 339–341.

THE *C. elegans* SEQUENCING CONSORTIUM, 1998, Genome sequence of the nematode *C. elegans*. *Science* [special issue], **11 December**, 2041–2046.