Perspective

NMR-Based Metabolic Profiling and Metabonomic Approaches to Problems in Molecular Toxicology

Muireann Coen,* Elaine Holmes, John C. Lindon, and Jeremy K. Nicholson*

Department of Biomolecular Medicine, Sir Alexander Fleming Building, Surgery, Oncology, Reproductive Biology and Anesthetics Division, Faculty of Medicine, Imperial College London, London SW7 2AZ, United Kingdom

Received September 13, 2007

We have reviewed the main contributions to the development of NMR-based metabonomic and metabolic profiling approaches for toxicological assessment, biomarker discovery, and studies on toxic mechanisms. The metabonomic approach, (defined as the quantitative measurement of the multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification) was originally developed to assist interpretation in NMR-based toxicological studies. However, in recent years there has been extensive fusion with metabolomic and other metabolic profiling approaches developed in plant biology, and there is much wider coverage of the biomedical and environmental fields. Specifically, metabonomics involves the use of spectroscopic techniques with statistical and mathematical tools to elucidate dominant patterns and trends directly correlated with time-related metabolic fluctuations within spectral data sets usually derived from biofluids or tissue samples. Temporal multivariate metabolic signatures can be used to discover biomarkers of toxic effect, as general toxicity screening aids, or to provide novel mechanistic information. This approach is complementary to proteomics and genomics and is applicable to a wide range of problems, including disease diagnosis, evaluation of xenobiotic toxicity, functional genomics, and nutritional studies. The use of biological fluids as a source of whole organism metabolic information enhances the use of this approach in minimally invasive longitudinal studies.

Contents

1. Introduction to Metabolic Profiling

Broad spectrum metabolic profiling is now recognized as a powerful top-down systems biology tool that can provide a real world link to other omics sciences (*1–4*). The terms metabolomics (*5–7*) and metabonomics (*8*) are widely applied to these types of studies, and the terminology is often used interchangeably. Metabonomics provides a whole-organism biological description of time-related multivariate metabolic response to a treatment. It facilitates the study of the metabolic products and interactions of hundreds of cellular metabolomes (metabolic complements) and fluid compartments, which are unique to each cell type in the body but are coordinated in space and time, and this concept of the interacting metabolomes has been termed the metabonome (*9*). A variety of analytical technologies have been applied to metabolic profiling in toxicology, but most approaches utilize NMR spectroscopy or mass spectrometry as these instrumentalities can capture information on hundreds or even thousands of metabolites in a sample in a single analytical run. To date, there have been more publications reflecting the application of NMR spectroscopy in metabolic toxicology (excluding drug metabolism applications), but modern LC-MS methods are now being successfully utilized in this area (*10–12*), and the balance will change although we consider that there will always be a role for NMR in rapid multivariate metabolic profiling and in metabolic structural elucidation. As NMR spectroscopy (in its own right) comprises a wide range of analytical techniques (and can be applied to biofluids as well as intact tissues),

* To whom correspondence should addressed.

10.1021/tx700335d CCC: $40.75
Published on Web 01/03/2008

we have concentrated our review on the role of NMR spectroscopy in the development of toxicological metabonomics, but we have also considered cognate biomedical applications as they will be of increasing importance in the future.

2. Background: The Early Days of Metabolic NMR spectroscopy

The development of Fourier transform NMR spectroscopy in the late 1960s, the introduction of superconducting magnets in the 1970s, and the consequent sensitivity increases resulted in the first applications of NMR spectroscopy for the metabolic profiling of biofluids and cells. Since then, numerous studies have concentrated on the use of ^{1}H NMR spectroscopy to characterize toxic response to drugs, as reflected in biofluid spectral signatures, and many novel metabolic markers of organ-specific toxicity have been discovered (*13*). The role of metabonomics in particular and magnetic resonance in general in the toxicological evaluation of drugs has developed extensively (*14*). ^{1}H NMR spectroscopy is well suited to the study of toxic events, as a biofluid fingerprint that reflects toxic response can be rapidly achieved without bias imposed by expectations of the type of toxin-induced metabolic changes. Moreover, many of the NMR-detectable metabolites are present at moderate to high concentrations and represent the products and intermediates of many important or hub pathways that are affected by many toxic or disease processes. Early applications of NMR spectroscopy found that quantitative changes in metabolite patterns gave information on the location and severity of toxic lesions, together with insights into the underlying molecular mechanisms of toxicity (*13*, *15*–*17*).

Examples of early NMR studies of toxins include the effects of exposure to cadmium and mercury salts, which are both potent nephrotoxins (*18*), with acute cadmium exposure also causing profound testicular toxicity. NMR methods were applied successfully to analyze urine for novel metabolites caused by exposure of rats to acute cadmium and mercury dosing in dose−response studies. In the case of mercury, classical patterns of acquired Fanconi syndrome were observed, with markers of proximal tubular injury, including marked amino and organic aciduria, coupled with differential responses of citric acid cycle intermediate excretion (e.g., low citrate with high succinate) consistent with selective inhibition of mitochondrial enzymes such as succinate dehydrogenase and malate dehydrogenase (*15*). These studies showed for the first time that it was possible to capture site-specific severity and mechanistic information simultaneously. Studies by Gartland et al. (*19*) further demonstrated the value of such screening in determining the region-specific toxicity using a variety of experimental nephrotoxins. Many other proximal tubular toxins were studied around this time showing spectroscopic commonalities of time response and recovery from injury (*20*–*23*). The proximal tubular toxicity of para-amino phenol was extensively studied (*24*, *25*) together with the protective effects of buthionine sulfoximine treatment and biliary cannulation (*26*), which showed that this compound worked effectively via a toxic thiol mechanism involving initial glutathione conjugation followed by further metabolism and the generation of a toxic thiol adduct that was transported to the kidney. In addition, the utility of the NMR approach in clinical toxicology was highlighted in a case study involving cutaneous exposure to phenol causing renal failure, the progression and recovery from which was followed by NMR spectroscopy (*27*). This study predates the metabolic trajectory analysis approaches (see below) for studying the development of lesions in experimental toxicology studies that were to become important later in the 1990s. In the case of cadmium toxicity, a similar approach was employed to show that urinary creatine was a highly sensitive early reporter of acute testicular injury with urinary creatine being elevated many hours before lesions became detectable by histopathology (*28*, *29*). In later studies by Timbrell and co-workers (*30*, *31*), it was found that creatinuria could also indicate liver damage, but this was only really indicative in the presence of concomitant taurinuria. Taurinuria was first demonstrated by NMR as a useful urinary marker for liver injury (*32*, *33*), and a wide variety of hepatotoxins are now known to cause taurinuria, although this can be very variable depending on dietary sources (*34*). Other interesting subtoxic, but toxicologically significant metabolic effects of drugs could also be observed using this approach such as sugar acidurias caused by aldose reductase inhibitors (*35*). Ghauri et al. showed that chronic acetaminophen ingestion caused 5-oxoprolinuria in rats and that this could be completely eliminated using dietary methionine supplements indicating that oral drug dosing could deplete sulfur-containing amino acids and disrupt the glutathione cycle (*36*). It was later shown that even at therapeutic doses in humans, fractional 5-oxoprolinuria could be detected after acetaminophen treatment (*37*).

3. Pattern Recognition for Sample Classification and Biomarker Discovery

The use of chemometric methods to analyze complex spectral data sets was perhaps the single most important development in the practical application of metabonomics and has defined the development and progression of the field ever since. The first studies that used PR to classify biofluid samples used a simple scoring system to describe the fluctuating levels of 18 major endogenous metabolites in urine from rats that either were in a control group or had received a specific organ toxin that affected the liver, the testes, the renal cortex, or the renal medulla (*38*, *39*). These studies showed that samples corresponding to different organ toxins mapped into distinctly different regions of the pattern recognition diagrams indicating that site-specific and severity information could be captured directly from the metabolic profile. Various refinements in data analysis were investigated, including taking scored data at three time points after toxin exposure for the nephrotoxins only as well as using a simple dual scoring system (the time and magnitude of the greatest change from control). The maps derived from the full time course information provided the best discrimination between toxin classes, emphasizing the importance of capturing dynamic information in the characterization of toxic lesions. This study was further extended (*40*) to incorporate actual metabolite NMR resonance intensities rather than simple scores. This was carried out for the nephrotoxins in the earlier group plus additional nephrotoxic compounds. A good separation of renal medullary and renal cortical toxins was achieved. In addition, it was possible to differentiate cortical toxins according to the region of the proximal tubule (S1, S2, and S3), which was affected, and also by the biochemical mechanism of the toxic effect. However, it was noted that absolute quantification of metabolites did not necessarily improve PR classification of toxicity over simple scoring systems.

The time course of metabolic urinary changes induced by two renal toxins was first investigated in detail by metabonomics using Fisher 344 rats administered a single acute dose of the renal cortical toxin, mercury II chloride, and the renal papillary toxin, 2-bromoethanamine (*41*, *42*). Rat urine was collected for up to 9 days after dosing, and samples were analyzed using high resolution ^{1}H NMR spectroscopy. The onset, progression,

and recovery of the lesions were also followed using histopathology to provide a definitive classification of the toxic state relating to each urine sample, and the geometry of the trajectory generated information relating to the mechanism and sequential targets of the toxin. The concentrations of 20 endogenous urinary metabolites were measured at 8 time points after dosing and mapping methods were used to reduce the data dimensionality. These showed that the points on the plot could be related to the development of, and recovery from, the lesions.

Early pattern recognition studies on NMR data employed a reductionist approach of preselecting metabolite signals of interest. However, the NMR spectral results generated in a metabonomic study yield a unique metabolic fingerprint for each biofluid sample consisting of thousands of overlapping resonances, and measurement of a small set of signals will not reflect the full potential of the spectral profile. If the status of a given organism changes, such as in a diseased state or following exposure to a drug, the unique metabolic fingerprint or signature reflects this change (*2*, *13*). Multivariate statistical methods provide an expert means of analyzing and maximizing information recovery from complex NMR spectral data sets. Detailed inspection of NMR spectra and integration of individual peaks can give valuable information on dominant biochemical changes; however, subtle variation in spectra may be overlooked, and it is difficult to envisage general effects as a function of both dose and time in a large cohort of samples with biological variability. Pattern recognition methods can be used to map the NMR spectra into a representative low dimensional space such that any clustering of the samples based on similarities of biochemical profiles can be determined and the biochemical basis of the pattern elucidated.

The initial objective in metabonomics is to classify a spectrum based on identification of its inherent patterns of peaks and, second, to identify those spectral features responsible for the classification (according to physiological or pathological status), which can be achieved via both supervised and unsupervised pattern recognition techniques. The NMR spectral data is preprocessed, which typically involves Fourier transformation, calibration of the chemical shift scale using an internal reference standard, and phase and polynomial baseline correction. To prepare the NMR data for multivariate modeling, the spectra are often divided into regions (along the chemical shift axis) whose areas are summed to provide an integral so that the intensities of peaks in such defined spectral regions are extracted, a process known as binning. This results in a data matrix (Figure 1a) consisting of rows that reflect observations/samples and columns that represent variables, for example, the spectral integrals of defined bins across the whole spectral width (*43*). Recent advances in chemometric approaches involve the utilization of full resolution NMR data, where each data point in an acquired spectrum is extracted as a variable for modeling. This approach has many advantages, for example, the spectral structure is retained, which enables the NMR user to identify metabolites with ease, and it also avoids searching within bins post data modeling to determine metabolites of discriminatory importance. The use of full resolution NMR data in chemometric modeling will be discussed in greater detail in a later section that encompasses orthogonal-projection on latent structures-discriminant analysis (O-PLS-DA).

Following the above preprocessing steps and the output of a data matrix consisting of samples and their associated variables, normalization is often applied to the rows (spectra). This adjusts spectral intensities so that concentration differences between samples are accounted for such that the samples are more

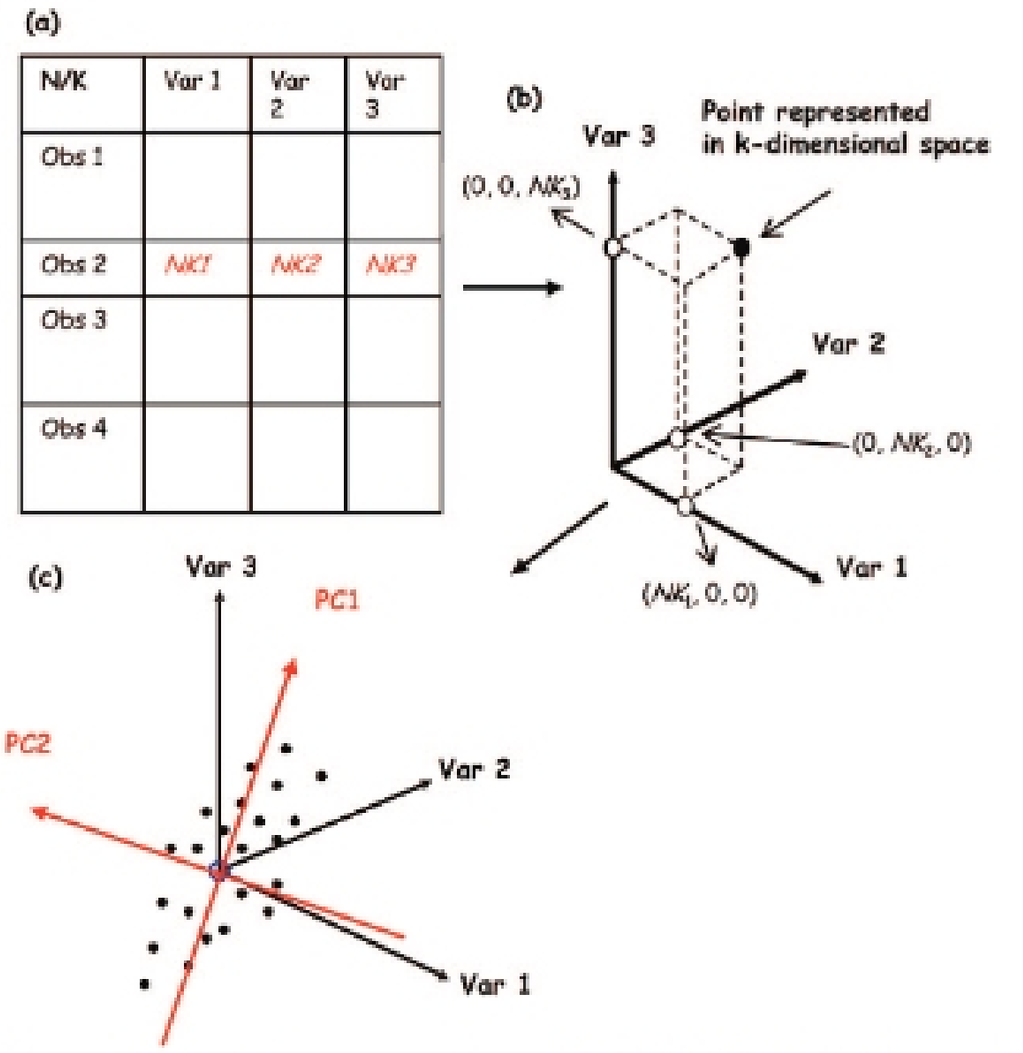

Figure 1. Principal components analysis. (a) Stylized data matrix consisting of N observations (spectra, $N = 4$) and K variables (spectral regions, $K = 3$). (b) Representation of the three variables placed in a 3D Cartesian coordinate system. (c) All observations in the data matrix are placed in 3D space, and the computed principal components are shown as vector arrows. Key: Obs, Observation; Var, Variable. (Reprinted with permission from Coen, M., and Kuchel, P.W. (2004) Metabonomics based on NMR spectroscopy. *Chem. Aust. 6*, 13–17. Copyright 2004 The Royal Australian Chemical Institute Inc.)

directly and reliably comparable. A commonly applied normalization method known as normalization to total area or constant sum sets the total spectral area of each spectrum to unity; therefore, the intensities of all data points are expressed relative to this. However, many other approaches are routinely used, and metabonomic studies that have investigated the effects of normalization routines on data modeling have been reported in the literature (*44*, *45*). Scaling is the final preprocessing step typically applied to NMR spectral data prior to chemometric modeling and is a column operation that aims to reduce the noise in the data and hence improve model interpretability, for example, each column in a matrix can be set to have unit variance or a mean of zero (*44*, *46*).

Principal components analysis (*47*) has been widely used in metabonomic studies and is an unsupervised approach in that it allows inherent clustering behavior of samples to be ascertained with no *a priori* knowledge of sample class membership. PCA reduces the dimensionality of a data set as it allows multidimensional data vectors to be projected onto a hyperplane of lower dimensions (typically 2 or 3), with this projection explaining as much of the variation as possible within the data. As previously introduced, the NMR data consists of a matrix of N observations (spectra) and K variables (spectral regions) (Figure 1a) so that a variable space of K dimensions is created. Each variable represents a numerical value on one coordinate axis, and each observation is placed in K-dimensional space. This situation is depicted in Figure 1b for the simple case of three variables in which all observations are added to the one coordinate system (Figure 1c), and then the first principal component (PC) is calculated by a standard method. The first PC (PC1) is a linear combination of the original input variables, and it describes the largest variation in the data set (Figure 1c). The second PC (PC2) is then calculated, and this is orthogonal to PC1 and describes the next highest degree of variation in