Analytica Chimica Acta 652 (2009) 173–179

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Spectroscopic correlation analysis of NMR-based metabonomics in exercise science

Gemma M. Kirwan[a], Vernon G. Coffey[b], Julie O. Niere[a], John A. Hawley[b], Michael J. Adams[a,*]

[a] *Applied Chemistry, School of Applied Sciences, RMIT University, Melbourne, Vic., Australia*
[b] *Exercise Metabolism Group, School of Medical Sciences, RMIT University, Bundoora, Vic., Australia*

ARTICLE INFO

Article history:
Received 8 May 2009
Received in revised form 3 July 2009
Accepted 3 July 2009
Available online 8 July 2009

Keywords:
Nuclear magnetic resonance
Metabonomics
Correlation spectroscopy
Statistical total correlation spectroscopy
Covariance matrix
Chemometrics

ABSTRACT

Spectroscopic studies of complex clinical fluids have led to the application of a more holistic approach to their chemical analysis becoming more popular and widely employed. The efficient and effective interpretation of multidimensional spectroscopic data relies on many chemometric techniques and one such group of tools is represented by so-called correlation analysis methods. Typical of these techniques are two-dimensional correlation analysis and statistical total correlation spectroscopy (STOCSY). Whilst the former has largely been applied to optical spectroscopic analysis, STOCSY was developed and has been applied almost exclusively to NMR metabonomic studies. Using a ^{1}H NMR study of human blood plasma, from subjects recovering from exhaustive exercise trials, the basic concepts and applications of these techniques are examined. Typical information from their application to NMR-based metabonomics is presented and their value in aiding interpretation of NMR data obtained from biological systems is illustrated. Major energy metabolites are identified in the NMR spectra and the dynamics of their appearance and removal from plasma during exercise recovery are illustrated and discussed. The complementary nature of two-dimensional correlation analysis and statistical total correlation spectroscopy are highlighted.

1. Introduction

Metabonomics is the measurement of the dynamic metabolic response of living systems to stimuli or modification [1]. The field of study developed from the application of NMR spectroscopy and mass spectrometry as tools for the study of complex biofluids [2,3]. Metabonomics encompasses comprehensive metabolomic profiling in whole organisms and systematic changes due to, for example, diet, lifestyle, and pharmaceutical interventions. In recent years the use and application of NMR-based metabonomics has grown rapidly and NMR techniques are commonly used in profiling urine and blood plasma [4–7], as well as other biofluids such as human cerebrospinal fluid [8], seminal fluid, synovial fluid, digestive fluids, etc. [2–4,9]. NMR is a powerful technique for biochemical analysis and such applications are in contrast to the more traditional use of NMR spectroscopy as a tool for chemical structural elucidation. The spectra provide information on a wide range of low molecular weight metabolites pertaining to biochemical status and physiological processes. Lindon and Nicholson have recently reviewed the field and summarised the status of metabonomics [10].

^{1}H NMR spectra of biofluids, such as urine and plasma, may contain thousands of signals arising from many hundreds of molecules, and a limiting factor in retrieving and understanding the relevant information from biological NMR spectra is their complexity [6]. One common approach to handling this complexity is to consider the NMR signal intensity data from a series of samples as a multi-sample array of metabolite concentrations and to treat it as a statistical object for data reduction and pattern recognition analysis [11]. With increasing spectral resolution (500–950 MHz fields are commonly employed) the complexity of the spectra increases and efficient and effective preprocessing and multivariate analysis methods are of paramount importance.

Signal alignment of spectra recorded from a series of similar samples presents a serious challenge. Averaging, or ‘binning’, data in a pre-selected window of frequencies has been one commonly used technique to reduce the effect of misalignment [3,12] but this can lead to loss of resolution and information. As a result, whole-spectrum analysis is gaining favour with sophisticated spectra alignment algorithms being employed [13–15].

Data analysis in metabonomics makes extensive use of factor analysis methods. Unsupervised pattern recognition using principal components analysis is common and when a dependant variable is present, for example a known concentration of a species or time of recording of each of a dynamic series of spectra, supervised data analysis methods are employed, of which partial least squares (PLS) is the most common [16]. These chemometric techniques are applied via decomposition of a suitable dispersion matrix, of which

---

* Corresponding author. Tel.: +61 3 9925 3358.
E-mail address: mike.adams@rmit.edu.au (M.J. Adams).

0003-2670/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2009.07.011

the covariance matrix, or related correlation coefficient matrix, is the most common.

Since the covariance matrix summarises the statistical interaction between measured variables it is not surprising that many multivariate analysis techniques use covariance, or correlation coefficient, to inform subsequent data analysis.

Correlation analysis in NMR studies was pioneered in the 1970s, and two-dimensional NMR COSY (Correlation Spectroscopy) experiments allow determination of connectivity in a molecule by identifying spin–spin proton coupling [17]. In seeking to apply a similar concept to other spectroscopic techniques, particularly infrared spectroscopy, Noda recognised that the RF pulse sequences used in two-dimensional NMR may be regarded as external perturbations applied to stimulate the system and he developed perturbation-based two-dimensional IR correlation spectroscopy using an externally applied oscillating mechanical perturbation of a sample. To overcome the limitation that the perturbation should be simple sinusoidal, Noda expanded the concept and mathematical basis of the technique to handle an arbitrary form of variable dependence [18] and developed what is referred to as generalized two-dimensional correlation spectroscopy which has been applied to many analytical techniques and a wide variety of analyses, including NMR spectroscopy [19]. For example, in a dynamic series of NMR spectra a significant correlation coefficient value between resonance peaks indicates them as arising from the same or related (in terms of process dynamics) molecular species.

NMR-based metabonomics has largely been pioneered by Nicholson and co-workers who developed a technique for aiding identification and assignment of multiple NMR peaks from the same molecule in a complex biofluids mixture [20,21]. The method, Statistical Total Correlation Spectroscopy (STOCSY), displays co-linearity of intensity variables from a series of NMR spectra, so that correlations between resonances arising from the same molecule can be identified [7,22].

Sasic has recently applied two-dimensional correlation analysis of ^{1}H NMR metabonomics data from rat urine collected at specific time intervals following administration of a trial drug [23]. Both binned and high-resolution data were examined, and covariance and correlation coefficient maps produced to aid interpretation of the data.

In the current study, human blood plasma samples, collected from subjects following vigorous exercise and ingestion of a controlled diet, have been investigated by ^{1}H NMR spectroscopy and the data subjected to STOCSY and generalised two-dimensional correlation analysis. Recently, Pederson et al. have reported the effects of glycogen-depleting exercise and subsequent carbohydrate and caffeine ingestion on rates of post-exercise muscle glycogen accumulation [24]. Plasma samples collected during a comparable study replicating the methods of Pedersen and co-workers [24] are analysed here to illustrate the merits and features of two-dimensional correlation analysis and demonstrate the potential for ^{1}H NMR analysis in exercise biochemistry.

2. Theory

Two-dimensional correlation spectroscopy provides graphical representation of the quantitative relationships between signal intensity at all pairs of spectral variable over the range of the dependant, perturbation variable. In the case of a dynamic series of spectra this perturbation variable is represented by the known time intervals at which discrete samples are collected or, in a continuous study, the spectra are recorded during the experiment. The methods consider not the specific individual spectra themselves but the dynamics and changes associated with spectral features. Changes in data are visualised and investigated by examination of spectral maps.

Noda treated the correlation between signals as a complex number comprising two orthogonal components known as the synchronous, $\mathbf{\Phi}$, and asynchronous, $\mathbf{\Psi}$, correlation intensity [18]. These components are defined by,

$$\Phi = \bar{X}^T \cdot \bar{X} \tag{1}$$

$$\Psi = \bar{X}^T \cdot N \cdot \bar{X} \tag{2}$$

where $\bar{X}$ is a matrix of corrected spectra, recorded at $i = 1$ to n sequential time intervals over $\delta = 1\text{–}m$ spectral variables (chemical shift values). Generally, $\bar{X}$ is the matrix of mean-centred spectra and Φ is thus the covariance matrix indicating the simultaneous changes of the spectral intensities observed at any pair of spectral variables. Visual inspection of Φ will indicate correlated, positive and negative, and uncorrelated spectral features. The diagonal vector of Φ is referred to as the autocorrelation spectrum comprising the so-called auto-peaks. For mean-centred data this diagonal is simply the vector of variance values for the series of spectra.

The asynchronous matrix, Ψ, represents sequential or unsynchronised variations and is obtained from the cross product between the original mean-centred data and orthogonalised data. This is achieved in Eq. (2) by the Hilbert-Noda matrix, $\mathbf{N}$ (size $n \times n$) that serves to extract out of phase portions of the signals. It is defined by

$$N_{i,k} = \begin{cases} 0 & \text{if } i = k \\ 1/\pi(i-k), & \text{otherwise} \end{cases} \tag{3}$$

The asynchronous matrix, $\mathbf{\Psi}$, serves to indicate sequential, but not simultaneous, changes of spectral intensities measured at two different spectral variables. The presence of a non-zero element in $\mathbf{\Psi}$ means that that the spectral dynamics at the pair of variables are not in linear relationships [25]. The diagonal of $\mathbf{\Psi}$ is thus zero by definition. Noda has provided a graphical interpretation of $\mathbf{\Phi}$ and $\mathbf{\Psi}$ and their complementary nature [26].

It is usual to display both Φ and Ψ as two-dimensional contour plots of covariance between the measured intensities at the spectral variables. However, visual interpretation of contour maps of the covariance, synchronous and asynchronous, matrices is difficult with high-resolution spectra, such as the ^{1}H data used here. The matrices are relatively large (typically a few thousand elements square) and sparse, and relationships between molecular species may be far apart on the chemical shift axis. This problem is overcome by selecting discrete slices from the matrices. A slice, $\mathbf{\Phi}(\delta 1,\delta)$ or $\mathbf{\Psi}(\delta 1,\delta)$, defines the synchronous or asynchronous relationships between a single peak (characteristic of a selected molecule appearing at δ1) and all other peaks, across all δ. The selected molecule, or peak, is sometimes referred to as the ‘driver’ peak [27].

The covariance matrix, Φ, as defined in Equation 1 also lies at the heart of STOCSY, which in addition makes use of the correlation coefficient matrix between spectral variables, $\boldsymbol{R}$. The correlation coefficient between recorded spectra and the dependant variable, time, $\rho(\delta,t)$, can also be useful.

$$\boldsymbol{R} = \tilde{X}^T \cdot \tilde{X} \tag{4}$$

$$\rho_{(\delta,t)} = \tilde{X}^T \cdot \tilde{t} \tag{5}$$

where $\tilde{X}$ represents the matrix of standardised (auto-scaled) spectral intensity values and $\tilde{t}$ the standardised time variable. The values from a slice, $\boldsymbol{R}$ ($\delta 1$, δ), from the correlation coefficient matrix indicate the degree of correlation between the selected variable and all others and have been displayed by colour coding the correlation coefficient values and projecting these on the NMR spectrum or covariance slice [20]. In this study, the correlation coefficient slices are displayed as separate plots for clarity alongside the covariance, synchronous and asynchronous, slice plots.