Interactive Metabolomics by diffusion NMR: Improving the odds of finding needles in haystacks.

Jonathan Byrne[1], Rasmus Bro[2], Florian Wulfert[1] and Clare A. Daykin[1].

[1]Division of Molecular and Cellular Science, School of Pharmacy, University of Nottingham, Nottingham, UK, NG7 2RD

[2]Dept. of Dairy and Food Science, Faculty of Life Sciences, Copenhagen University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

INTRODUCTION

The Stokes-Einstein equation relates the hydrodynamic radius, r_s, of a molecule to the experimentally observable diffusion coefficient, D.

$$D = \frac{kT}{f} = \frac{kT}{6\pi\eta r_s} \quad (1)$$

NMR spectroscopy can be used to measure molecular diffusion coefficients[1]. Pulsed field gradient stimulated echo pulse sequences (PFGSE) are employed, with the gradient amplitude incremented to give signal attenuation, governed by the equation:

$$I = I_0 \exp(-D\gamma^2 g^2 \delta^2 (\Delta - \delta/3)) \quad (2)$$

where I is the signal intensity at gradient amplitude g, and I_0 is the signal intensity at zero gradient amplitude.

When low molecular weight compounds are non-covalently bound to proteins, the observed diffusion coefficient for the compound will be a weighted average of its free and bound forms. By measuring diffusion by NMR, an estimation of the degree of protein-binding can be made for either low molecular weight endogenous biochemicals or xenobiotics. This type of experiment is commonly referred to as either Diffusion-Ordered Spectroscopy (DOSY) or Diffusion-Edited Spectroscopy, depending on the type of post-acquisition data processing which is applied to the spectra.

DOSY DATA HANDLING

We have developed a graphical user interface, within MATLAB, to process spectral data. This GUI has been used in this work for the importing, baseline correction, chemical shift referencing and diffusion coefficient fitting (see below) of experimental data, as well as for generating simulated data sets for testing the multivariate data analysis methods (MVDA). Bruker DOSY data sets are read in after Fourier transformation and phase correction. The data are formatted for immediate analysis by MVDA, using the Eigenvector PLS toolbox.

METHOD 1: DIFFUSION WEIGHTED NMR

A decay curve of the form of Equation 2 above is fitted at each chemical shift, δ, yielding the diffusion coefficient, D, at that shift. A plot of D against δ (Figure 2) is referred to as a diffusion weighted NMR spectrum.

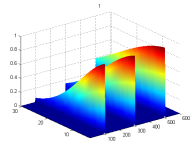


Figure 1

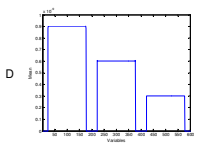


Figure 2

PCA can then be performed on diffusion weighted NMR spectra. Figure 3 shows overlaid diffusion weighted NMR spectra from simulations, with diffusion coefficients ranging from 5.0e-10 to 9.0e-10 m^2/s for single peaks. Figure 4 shows the PCA result (PC1 scores vs. sample).

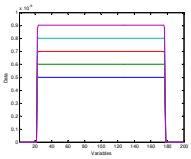


Figure 3

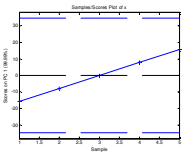
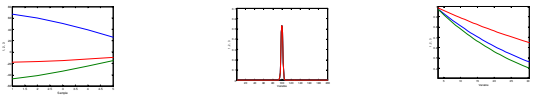


Figure 4

METHOD 2: MULTIWAY ANALYSIS

PARAFAC[2] multiway data analysis can be used to express DOSY data as triads of loading vectors. This gives separation of the samples in a similar way to PCA. Shown below are the results of a three triad PARAFAC analysis for the spectra shown in Figure 3.

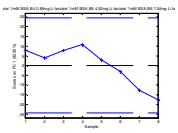


This method of MVDA is ideally suited to the analysis of DOSY data, but is computationally much more expensive than the diffusion weighting approach. The diffusion weighting method is useful for analysis of entire NMR spectra, with PARAFAC then used on regions of the spectra shown to be important in the subsequent PCA.

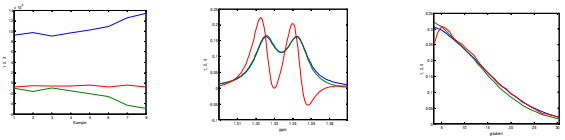
AFFINITY NMR EXPERIMENT

Diffusion NMR experiment (bipolar gradients stimulated echo with CPMG protein signal suppression and WET water suppression) were performed on a 1mM bovine serum albumin/ 40mM lithium lactate aqueous solution at pH=7.4. Ibuprofen sodium salt was added to give ibuprofen:BSA molar ratios of 11, 17, 21, 29, 39, 70, 105 and 142. PCA and PARAFAC analyses were performed on the lactate methyl doublet at 1.33ppm, using data in the range 1.27 to 1.4ppm. The lactate concentration is unchanged between samples.

The loading plot of PC1 from the PCA of diffusion-weighted NMR spectra (right) shows a general linear trend of decreasing score on PC1 with increasing ibuprofen sodium salt concentration.



The results of the three-component PARAFAC analysis (below) support the findings of the PCA.
The contribution from component 1 (blue) increases with increasing ibuprofen concentration, while the contribution from component 2 (green) decreases.
The decay profile of component 1 in Mode 3 is less steep than component 2, which reflects a lower diffusion coefficient. This suggests, intriguingly, that the presence of ibuprofen causes the lactate to bind more readily, and so translate more slowly, than in its absence. Component 3 (red) explains phase shift between samples.



Mode 1 Mode 2 Mode 3

CONCLUSIONS

o Interactive Metabolomics has immense potential as a diagnostic tool. Clustering of sample data can be done non-selectively based on changes in binding behaviour.

o The experimental NMR data used for the analysis must be excellent, requiring optimization of water suppression, protein background suppression and temperature stabilization. Reference deconvolution[3] should be used in the preprocessing to correct lineshape distortion and phase error.

o Future work will involve improving the quality of the experimental data acquired, and adapting the method for studies carried out at physiological temperatures (37°C).

References

1. "Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion, Part 1: Basic Theory" William S. Price, Magnetic Resonance: an Educational Journal, W.S.Price, 1997, **9**, 299 – 336.
2. "PARAFAC: Tutorial and applications", R. Bro, Chemom.Intell.Lab.Syst., 1997, **38**, 149-171.
3. "Reference Deconvolution: A Simple and Effective Method for Resolution Enhancement in Nuclear Magnetic Resonance Spectroscopy", K. R. Metz, M. M. Lam, A. G. Webb, Concepts in Magnetic Resonance, 2000, **121**, 21-42.