

An NMR-based metabolomic approach to the analysis of the effects of xenobiotics on endogenous metabolite levels in plants

Nigel J.C. Bailey^{a,*}, Matjaz Oven^{b,**}, Elaine Holmes^c, Meinhart H. Zenk^d and Jeremy K. Nicholson^c

^a SCYNEXIS Europe Ltd. Fyfield Business and Research Park, Fyfield Road, Ongar, Essex CM5 0GS, United Kingdom

^b Leibniz Institut für Pflanzenbiochemie, Abt. Naturstoffbiotechnologie, Weinberg 3, 06120 Halle/Saale, Germany

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

^d Biozentrum-Pharmazie, Universität Halle, Weinbergweg 22, 06120 Halle/Saale, Germany

Abstract. High field ¹H NMR spectroscopy has been employed to obtain, in conjunction with chemometric analysis, information regarding fluctuations in endogenous metabolic profiles for *Crotalaria cobalticola* plant cells following exposure to cobalt chloride.

Such 'metabolomic' type data analysis is often confounded by experimental, environmental or genetic factors that are not correlated to the classifications of interest and serve only to complicate the extraction of meaningful answers from a dataset. This work demonstrates the application of data filtering to remove extraneous data that result from spectrometer variation rather than being correlated with the classes of interest. Samples were analysed from *Crotalaria cobalticola* suspension cell culture following exposure to cobalt chloride using 2 spectrometers. Removal of confounding data due to spectrometer variation resulted in clear separation between control and dosed classes. It was then possible to use the model to determine key changes in biochemical status caused as a result of exposure to cobalt. Branched chain amino acids, succinate and secondary metabolite precursors phenylalanine and tyrosine were all higher in the control samples, whilst choline, glutamate, alanine and lactate were higher in the dosed samples.

Keywords: Cobalt, *Crotalaria cobalticola*, data filtering, ¹H NMR spectroscopy, orthogonal signal correction, pattern recognition, metabolomics, toxicity

Abbreviations:

FID, free induction decay; NMR, nuclear magnetic resonance; OSC, orthogonal signal correction; PCA, principal components analysis; PLS-DA, partial least squares discriminant analysis; TSP, trimethylsilyl propionic acid.

*Corresponding author. Tel.: +44 (0) 1277 367036; Fax: +44 (0) 1277 367099; E-mail: nigel.bailey@scynexis.com.

**New affiliation: School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, United Kingdom.

1. Introduction

The development of novel analytical strategies for deriving information on differential gene function in relation to environmental stressors is essential in order to advance the understanding of the molecular basis of metal tolerance. Whereas genomics and proteomics can provide insights into the potential of a biological system to interact with external perturbations (pharmaceutical/agrochemical compounds, pollutants, environmental effects), metabolomics (defined as the study of the metabolic fluxes in prokaryotes and individual cells or isolated cell complexes) or metabonomics (the study of complex organisms where hundreds of cell types are all performing a multitude of tasks) provide insights into the resulting changes in the metabolic profile of the system that are the ultimate result of such genomic and proteomic influences.

The non-selective, yet highly specific approach of ^1H NMR spectroscopy, where no pre-judgement of the sample is required, offers several advantages with respect to the development of an analytical methodology that is readily transferable between samples from differing applications [1–5], including recent applications in the area of plant metabolomics [6–9].

It is necessary however, to ensure that any analytical approach is suitably robust in its approach and is capable of generating reproducible results ([7,10] and references therein). One potential limitation that may be envisaged is the effect of variation in experimental conditions leading to measurable differences between replicate samples. This variation may come from many different sources (e.g., instrument variation, differences in time of acquisition, extraction procedures, etc.) and in the case of particularly subtle biochemical effects, may lead to confounding variation in a dataset. These potential problems may be exacerbated in the case of multi-site studies, where different spectrometers are used for data acquisition, or where machine time is limited.

One approach to the removal of such confounding data is the application of data filtering, in particular, Orthogonal Signal Correction (OSC) [11,12]. This approach aims to remove any variation within the dataset that is uncorrelated (orthogonal) to the desired classification. The result is a dataset where the remaining variance is that which best describes the dataset in terms of the defined classifications [13].

Here we demonstrate the value of applying OSC and NMR based metabolomics in the investigation of metal tolerance and toxicity in plants. In this study, we used the cell suspension culture of a cobalt hyperaccumulator *Crotalaria cobalticola* Duvign. *C. cobalticola* grows endemically in a copper-cobalt belt in the Shaba province of the Democratic Republic of Congo. This plant can accumulate Co^{2+} in its shoots in concentrations as high as $3000\ \mu\text{g/g}$ dry weight [14], while non-accumulating plants normally contain cobalt ions in much lower concentrations (up to $50\ \mu\text{g/g}$ dry weight [15]). Moreover, it was shown recently that *C. cobalticola* can tolerate and accumulate Co^{2+} also when taken into suspension cell culture [16] and it was therefore suitable for the investigation presented here.

2. Experimental

2.1. Cell suspension cultures

Crotalaria cobalticola cell cultures (7 day old culture, obtained from existing cultures at the Institute of Plant Biochemistry, Halle, Germany) were vacuum filtered and washed with sterile water.

Two 1 l Erlenmeyer flasks with 250 ml fresh Linsmaier–Skoog growing media [17] were prepared, and 40 g (fresh weight) cells added to each flask. In addition, one flask contained 10 mM CoCl_2 (dosed culture), whilst the equal amount of water was added to the second flask (control culture). Both flasks

were cultivated under sterile conditions for seven days (gyratory shaker 100 rpm, diffuse light 650 lux, 22°C). After seven days, cells from both flasks were vacuum filtered and washed with sterile water. Filtered cells were then flash frozen with liquid nitrogen and lyophilised.

2.2. Sample preparation

Replicates (approx 20 mg, $n = 10$) of lyophilised cells from control and dosed flasks were weighed out and added to D₂O (1 ml, containing 0.05% w/v TSP as NMR reference). Samples were agitated and then centrifuged at 13 000 rpm for 15 minutes. Supernatant (850 μ l) supernatant was removed and re-centrifuged for a further 10 minutes. Following the second centrifugation, supernatant (700 μ l) was taken for NMR analysis.

2.3. ¹H NMR spectroscopy of cell culture extracts

NMR spectra of the cell extracts were acquired on two Bruker (Bruker GmbH, Rheinstetten, Germany) NMR spectrometers. Although the spectral parameters used were essentially the same on both spectrometers (same number of FIDs, same pulse programme, etc.), specific differences did exist, as would occur if multi-site studies were run, or if spectra were acquired at different times. All the samples (20 in total) were analysed on spectrometer 1, and half of the samples (5 control, 5 dosed, 10 in total) were run on spectrometer 2 as well, giving a total of 30 spectra from the 20 samples.

Spectrometer 1 operated at 600.22 MHz for the ¹H frequency, utilising Bruker BEST™ flow injection technology for sample transfer. All spectra were the result of the summation of 64 free induction decays (FIDs), with data collected into 64k datapoints, a spectral width of δ 20 and an acquisition time of 2.73 seconds.

Spectrometer 2 operated at 600.13 MHz for the ¹H frequency and was fitted with a broadband inverse geometry probe. All spectra were the result of the summation of 64 free induction decays (FIDs), with data collected into 32k datapoints, a spectral width of δ 14 and an acquisition time of 1.95 seconds.

The 90° pulse length was found for the samples on both spectrometers prior to data acquisition. The water signal was suppressed using a standard 1D-presaturation pulse sequence [18]. Prior to Fourier transformation, an exponential line broadening equivalent to 0.3 Hz was applied to the FIDs and spectra were referenced to TSP at δ 0.00.

Assignment of resonances within the spectra was performed based on 2D correlation NMR experiments (data not shown) and through reference to previously published data (e.g., [19]).

2.4. Multivariate data analysis

One-dimensional NMR spectra were reduced to 252 discrete chemical shift regions by digitisation to produce a series of sequentially integrated regions δ 0.04 in width between δ -0.02 and 9.98, using Bruker AMIX software (version 2.0, Bruker GmbH, Germany). The resulting data matrix was exported into Microsoft® Excel 2000 and selected regions removed, i.e. around the residual water signal (δ 4.54–4.98), sucrose (from the media solution, δ 4.18–3.22, δ 5.26–5.18) and TSP (δ -0.02–0.02). The remaining integral regions were normalised to the whole spectrum to remove any concentration effects prior to Principal Components Analysis (PCA) [20].

PCA was performed using SIMCA-P 8.0 multivariate data analysis software (Umetrics, Sweden), with mean centering of the data preceding PCA. The output from the PCA analysis consisted of scores plots (giving an indication of the differentiation of the classes in terms of biochemical similarity), and

Table 1

Summary of predictive ability of 3 different OSC models using randomly selected samples for training and test set membership. The percentage correct column gives an indication as to the number of samples classified correctly using PLS-DA (e.g., Fig. 2(c))

	Percentage correct (2PCs)		Percentage of predictions with confidence >99%	
	Control	Dosed	Control	Dosed
Training set	>99%	>99%	≥80%	>99%
Test set	>99%	>99%	≥80%	≥80%

loadings plots, which give an indication as to which NMR spectral regions were important with respect to the classification obtained in the scores plots.

Orthogonal signal correction (OSC) was performed also using SIMCA-P 8.0 in order to remove spectrometer variation. In order to allow validation of the subsequent model, the dataset of 30 samples was split randomly into two groups of 10 and 20 samples, respectively. The group of 20 samples (10 dosed and 10 control) was used as the training set, whilst the remaining 10 samples (5 dosed and 5 control) were used as a test set. For the dummy y matrix that was used to provide information regarding class membership, control samples were assigned the number 1, while dosed samples were 0. Following application of OSC, validation of the model was carried out using partial least squares discriminant analysis (PLS-DA) to predict class membership of the test set, i.e. does the model return the value 1 for a control sample, and the value 0 for a dosed sample. The process of generating an OSC model on randomly chosen samples followed by validation was repeated in triplicate and values indicated (Table 1) are as a result of all 3 models.

3. Results and discussion

Following ^1H NMR spectroscopic analysis of the plant cell extracts, principal components analysis (PCA) was performed, and the resulting scores plot (describing the relative similarity of the samples in multivariate hyperspace) is shown in Fig. 1, with each spectrum represented by a single datapoint. It can be seen that these datapoints are split into 4 sub-groups. The anticipated separation between dosed and control samples is immediately apparent, and is clarified by the dotted line drawn from bottom left to top right in the scores plot. In addition, however, it can be seen that it is possible to draw an additional line (bottom right to top left), which separates the samples according to the spectrometer on which their spectra were acquired. As these spectra were run using the same samples on different spectrometers, it can be seen that the differences between the spectra acquired in different spectrometers were sufficient to result in additional discrimination. Similar separations have recently been reported for data acquired at different field strengths and also when different water suppression techniques were employed [10]. Although this spectrometer variation is, in this case, not problematic as the desired separation (control vs. dosed) was still apparent (and indeed accounts for most of the variation within PC1 while the spectrometer variation appears through visual inspection to be largely in PC2), it was nevertheless decided to attempt to remove this variation by way of orthogonal signal correction (OSC) in order to focus solely on the effects of dosing *C. cobalticola* and to enhance the recovery of biomarkers of cobalt exposure. (PLS-DA was attempted prior to OSC, and whilst improving the separation slightly, the spectrometer variation was still clearly discernible – data not shown.) OSC attempts to filter variation

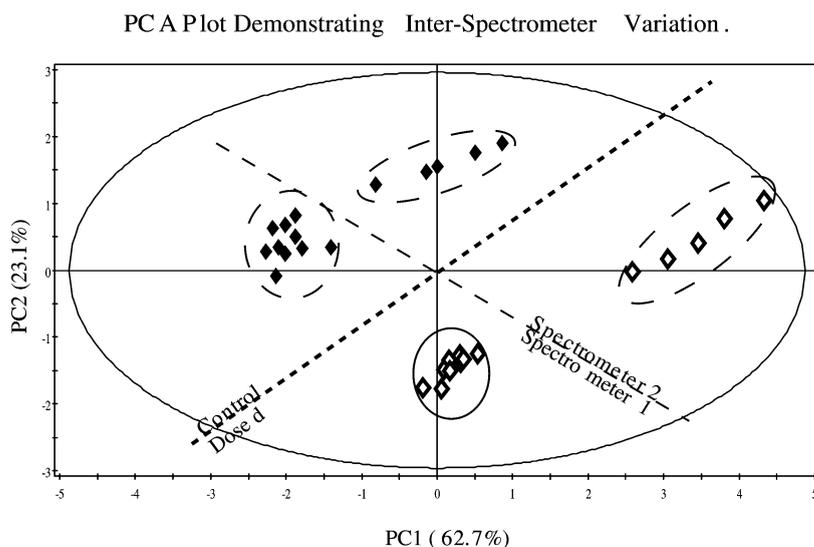


Fig. 1. Scores plot (PC1 v PC2) for control (◆) and dosed (◇) sample groups following PC analysis. The plot displays clear discrimination between the 2 groups, but in addition contains separation due to spectrometer variation.

within the dataset, which is not correlated to the classification of the data i.e., control and dosed classes. In this case therefore, it would be hoped that the OSC process would leave just the data that are correlated to the control and dosed classes and would remove the spectrometer variation, which is not correlated to the dosed vs. control classification. OSC is a supervised method, in that prior knowledge of sample class (a value of 1 was given to the control training samples, and a value of 0 to the dosed training samples) is required in order to perform that analysis. As a result of this however, it is necessary to validate the resultant model with independent samples that were not used in the model making process. The ability of the model (derived using a training set) to predict these unknown samples (the test set) demonstrates that the model has not been overfitted by the training set, and has sufficient robustness to predict class membership of previously unseen samples. Figure 2(a) shows a PCA plot of the 20 samples in a training set, with the data from the two spectrometers being represented as open circles (○) and closed squares (■), respectively. It can be seen that there is now no spectrometer-based discrimination of the data. Secondly, it is apparent that control (on the left) and dosed (on the right) samples are well separated using PC1 only, with much more of the overall dataset variation being present in PC1 in the post-OSC data as compared to the pre-OSC data (80.3% and 62.7% respectively, as indicated in Figs 1 and 2(a)). It was also observed that the control data points were more diffuse than those in the dosed dataset (standard deviations in PC2 of 0.53 and 0.34 for the control and dosed classes, respectively). This is as a result of the toxic insult of the xenobiotic having a greater impact on the metabolic profiles of the cells than the overall variation due to genetic differences between the cells. In effect, metabolic lensing is occurring, where a population of diverging biochemical (and thus metabonomic) profiles is brought back to relative homogeneity through xenobiotic insult.

Validation of the model was accomplished using both PCA and PLS-DA. Figure 2(b) shows the PCA plot containing the 20 training set samples along with the 10 test set samples (training set samples are the filled triangles (▲), test set the open triangles (△)). It can be seen that the model works well, with the entire test set samples being correctly mapped in the multivariate space. This is taken further by the use of PLS-DA to predict classification of the unseen samples, Fig. 2(c). This figure shows the

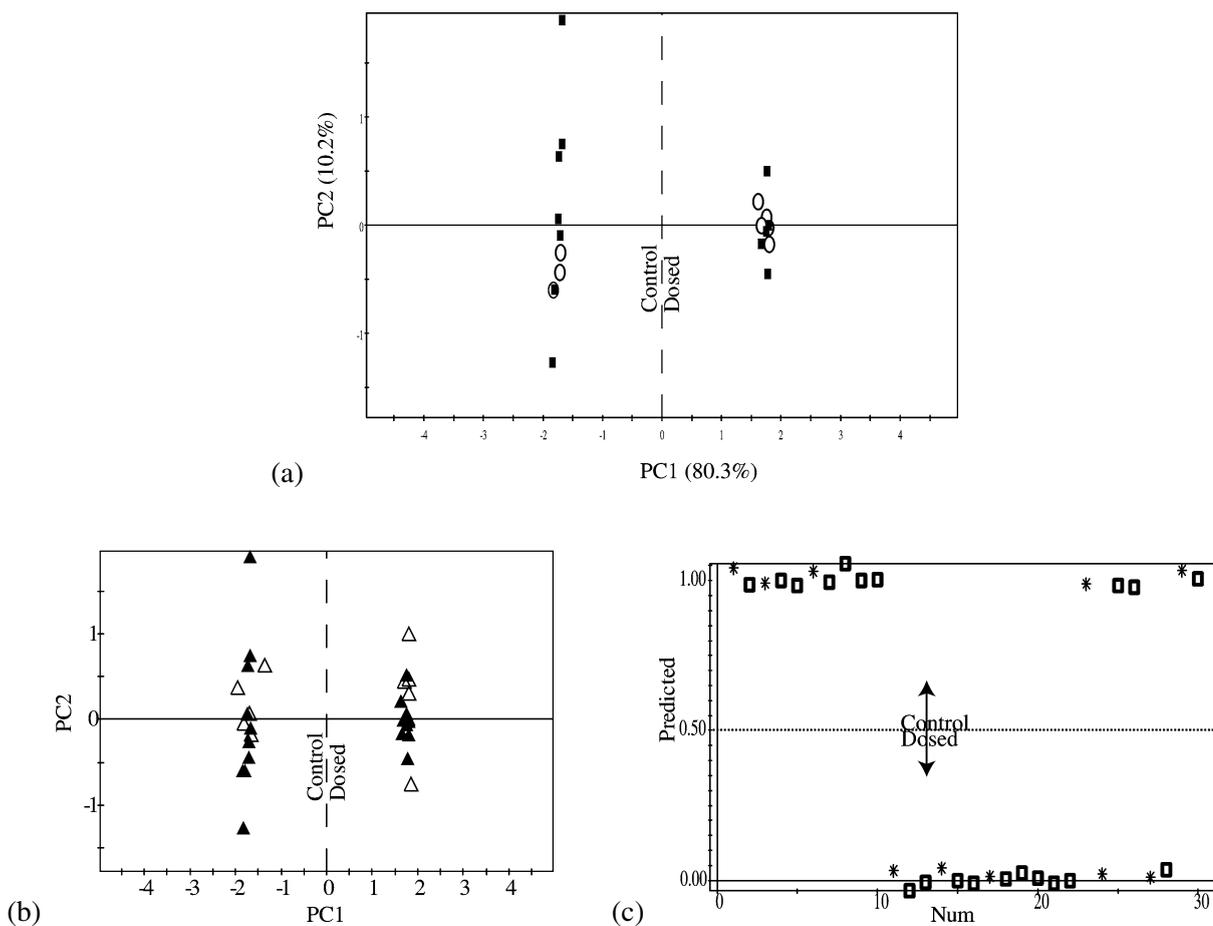


Fig. 2. (a) Post OSC scores plot for control and dosed samples groups following PC analysis. Samples obtained from spectra run on different spectrometers are indicated by open circles (o) and closed squares (■), respectively. All control samples are to the left of the plot, and all dosed samples are to the right. Clear separation between the dosed and control classes was achieved, with $\sim 80\%$ of the variance in the model present in the first PC. (b) Post OSC scores plot for control and dosed samples, using 20 samples for a training set (▲) and 10 samples for a test set (Δ). Samples were randomly assigned to the training or test sets, but all are mapping to the correct area in the plot. (c) Post OSC y -predicted plot following PLS-DA analysis, using 20 samples for a training set (□) and 10 samples for a test set (*). Samples above the line at a value of 0.5 on the y axis are considered to be control samples, and those below, dosed samples.

sample number on the x axis, with the predicted class value on the y axis. Any sample with a predicted value >0.5 (indicated by the black line) is considered a control sample, and anything below the line (i.e. <0.5) is considered a dosed sample. It can be seen that all control samples are above the line, with all the dosed samples below (training set samples are represented by the open squares (□), test set the asterisks (*)). It may also be noted that the inherent variability in the control cells relative to the dosed cells is again observable here, as the range of values given to the control group (0.99 ± 0.04) is larger than that of the dosed group (0.00 ± 0.02).

The process of producing a model using the training set and validating it with test set data was repeated 3 times (with samples allocated to either the training or test set randomly) to obtain an overall indication of robustness of the dataset as a whole. A summary of the predictive ability of the models is

shown in Table 1. Overall, it can be seen that the models produced are capable of predicting the correct classification for the vast majority of the samples, with a high level of confidence.

Following the creation of a model from which spectrometer variation has been filtered, it was possible to determine the identity of those spectral regions, and thus the biochemical species responsible for, the differences between the control and dosed classes. In the PCA, the new principal components are made up of linear combinations of the original variables to produce the scores plot. By determining which of the original variables were most highly correlated with the new principal components (the “loadings”), it was possible to obtain an indication of those loadings that were most important in the separation between control and dosed sample classes. It can be seen in Fig. 3 that a series of adjacent variables around 0.98 have a negative value in the loadings plot. As the control samples had negative values for PC1 (Fig. 2(b)), the variables around 0.98 are thus correlated with the control samples, whilst the positively correlated dosed samples (Fig. 2(b)) are correlated with those variables that also have a positive correlation with PC1 in Fig. 3. Thus it can be seen that branched chain amino acids, succinate and secondary metabolite precursors phenylalanine and tyrosine are all higher in the control samples, whilst choline, glutamate, alanine and lactate are higher in the dosed samples. The function of induced compounds can only be hypothesised at this stage. Choline concentration could be elevated due to the increased phospholipase D activity. This enzyme hydrolyses phosphatidylcholine into choline and phosphatidic acid, which is a known lipid second messenger. It was reported that phosphatidic acid concentration increases within minutes in plants under stress such as wounding, pathogen elicitors, osmotic and oxidative stress, or exposure to ethylene and abscisic acid [21]. Alternatively, choline is also a direct precursor of glycine betaine, a well-known plant osmoprotectant [22,23]. It was shown recently that increased choline concentration is crucial for glycine betaine biosynthesis in transgenic tobacco plants [23]. Elevated concentrations of lactate could occur due to the impact of cobalt ions on cell respiration machinery. Hence, it was shown previously that cobalt complexes strongly inhibit Ca^{2+} uptake in rat mitochondria [24]. Levels of glutamic acid and alanine could have increased due to increased protein biosynthesis. It should be also noted that no changes in cysteine and citrate concentrations were observed in this study. However, it was reported recently that the concentration of these two compounds rose dramatically in *C. cobalticola* cells when they were exposed to cobalt ions, as determined by an enzymatic assay, HPLC and a spectrophotometric method [16]. It is possible that NMR spectroscopy is not sensitive enough to detect differences in concentrations of cysteine and citrate in *C. cobalticola* cell extracts.

This work demonstrates how the combination of NMR and chemometrics allows the elucidation of biochemical differences between control and dosed sample groups following the administration of cobalt to plant cell cultures of *C. cobalticola* suspension cell culture. In addition, however, it focuses on the resolution of potential problems occurring during data acquisition, specifically, that of spectrometer variation. Instrument and or experimental variation is a phenomenon that is very often uncontrollable to a certain extent. With instrument complexity on the scale of NMR spectrometers, it is particularly inevitable that such variation is going to occur. Depending on the magnitude of the desired observation in a particular experiment, experimental variation may or may not become an issue. This paper demonstrates however, that when its presence is noted, it may be removed through use of a data filtering method, such as OSC which then allows data analysis to proceed normally. As shown in this paper, it is necessary to ensure sufficient validation of a filtered model is performed due to the supervised nature of the technique. Although the data in this example was at worst mildly affected by experimental variation, the post-OSC data clearly shows an improvement in the data which can only help in the subsequent analysis and interpretation of such data.

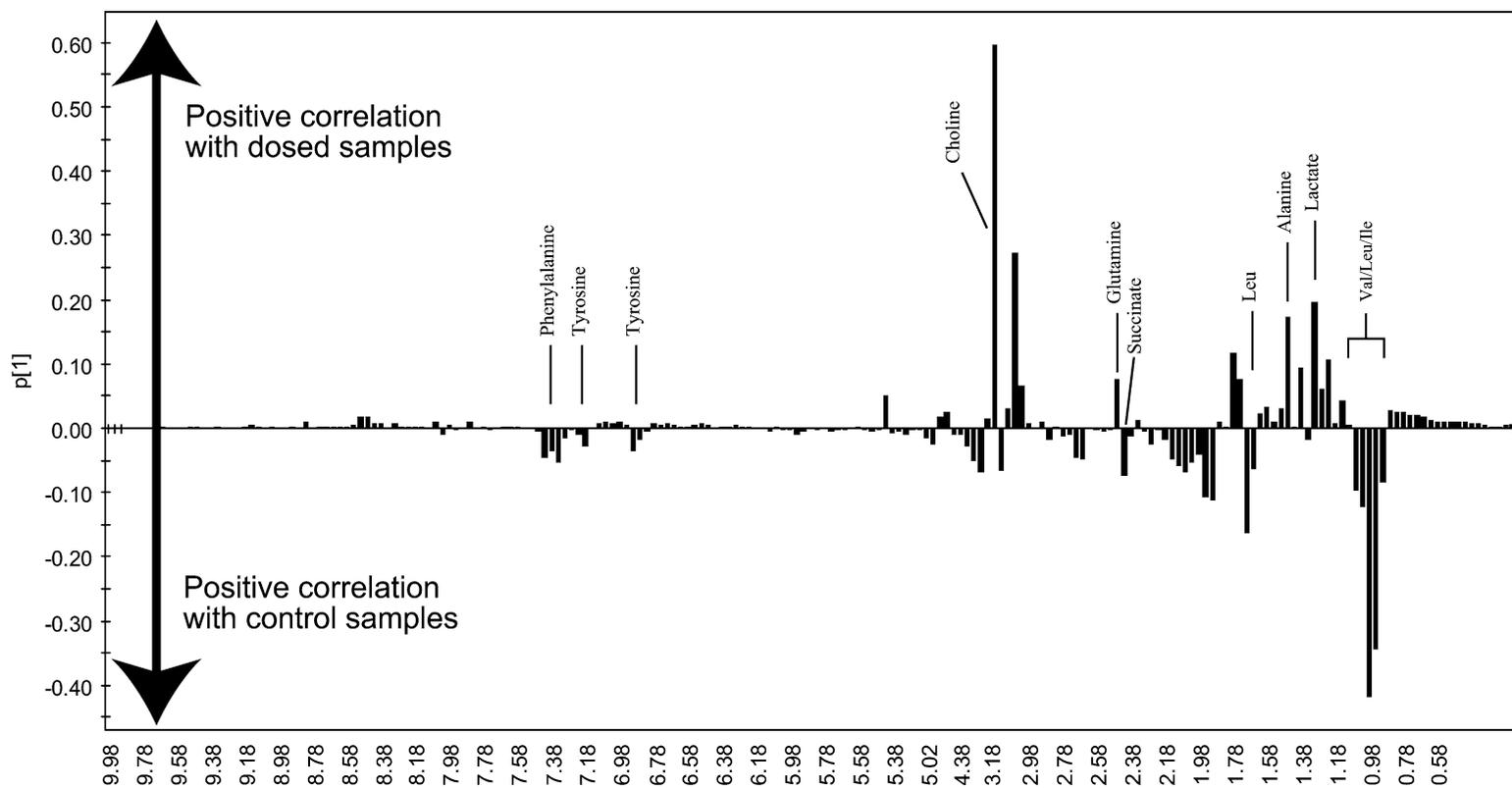


Fig. 3. Loadings column plot for dosed and control showing PC1 only. This plot allows elucidation of the chemical entities that are key in separating the control and dosed groups following PCA. Variables with a large positive value are positively correlated with the dosed group, whilst those with negative values are positively correlated with the control group.

Overall, this work shows that the combination of NMR, chemometrics and, where needed, data filtering, is a potentially powerful and useful addition to the techniques currently employed in metabolomic studies. In addition, the methodology may have wider generic utility in applications drawn from proteomics or genomics where more subtle changes may be masked by confounding data.

Acknowledgement

The authors wish to thank the European Science Foundation Plant Adaptation Programme for providing financial assistance to NJCB to allow completion of this work.

References

- [1] J.K. Nicholson, J. Connelly, J.C. Lindon and E. Holmes, *Nature Rev. Drug Disc.* **1** (2002), 153–160.
- [2] J. Nicholson, J. Lindon and E. Holmes, *Xenobiotica* **29** (1999), 1181–1189.
- [3] D. Robertson, M. Reily, R. Sigler, D. Wells, D. Paterson and T. Braden, *Tox. Sci.* **57** (2000), 326–337.
- [4] J.G. Bundy, D. Osborn, J.M. Weeks, J.C. Lindon and J. Nicholson, *FEBS Lett.* **500** (2001), 31–35.
- [5] C.L. Gavaghan, E. Holmes, E. Lenz, I.D. Wilson and J. Nicholson, *FEBS Lett.* **484** (2000), 169–174.
- [6] N.J.C. Bailey, M. Oven, E. Holmes, J.K. Nicholson and M.H. Zenk, *Phytochem.* **62** (2003), 851–858.
- [7] M. Defernez and I.J. Colquhoun, *Phytochem.* **62** (2003), 1009–1017.
- [8] K.-H. Ott, N. Aranibar, B. Singh and G.W. Stockton, *Phytochem.* **62** (2003), 971–985.
- [9] J.L. Ward, C. Harris, J. Lewis and M.H. Beale, *Phytochem.* **62** (2003), 949–957.
- [10] H.C. Keun, T.M.D. Ebbels, H. Antti, M.E. Bollard, O. Beckonert, G. Schlotterbeck, H. Senn, U. Niederhauser, E. Holmes, J.C. Lindon and J.K. Nicholson, *Chem. Res. Tox.* **15** (2002), 1380–1386.
- [11] S. Wold, H. Antti, F. Lindgren and J. Ohman, *Chemometr. Intell. Lab.* **44** (1998), 175–185.
- [12] B.M. Beckwith-Hall, J.T. Brindle, R.H. Barton, M. Coen, E. Holmes, J.K. Nicholson and H. Antti, in: *PLS and Related Methods – Proceedings of the PLS'01 International Symposium*, V.E. Vinzi, C. Lauro, A. Morineau and M. Tenenhaus, eds, Centre International de Statistique et d'Information appliques, Paris, France.
- [13] J.B. Hauksson, U. Edlund and J. Trygg, *Mag. Res. Chem.* **39** (2001), 267–275.
- [14] A. Baker, S. McGrath, R. Reeves and J. Smith, in: *Phytoremediation*, N. Terry and G. Banuelos, eds, CRC Press, Boca Raton, Florida, 1999, pp. 85–107.
- [15] P. Peterson and C. Girling, in: *Effect of Heavy Metal Pollution in Plants*, N. Lepp, ed., Applied Science Publishers, London and New Jersey, 1981, pp. 213–278.
- [16] M. Oven, E. Grill, A. Golan-Goldhirsh, T.M. Kutchan and M. Zenk, *Phytochem.* **60** (2002), 467–474.
- [17] Linsmaier and Skoog, *Physiol. Plant.* **18** (1965), 100–127.
- [18] J. Jeener, B. Meier, P. Bachmann and R. Ernst, *J. Chem. Phys.* **71** (1979), 4546–4553.
- [19] T.W.M. Fan, *Prog. NMR Spectrosc.* **28** (1996), 161–217.
- [20] L. Eriksson, E. Johansson, N. Kettaneh-Wold and S. Wold, *Introduction to Multi- and Megavariate Data Analysis using Projection Methods (PCA and PLS)*, Umetrics AB, Umeå, Sweden, 1999.
- [21] T.T. Munnik, *Plant Sci.* **6** (2001), 227–233.
- [22] T. Chen and N. Murata, *Curr. Opin. Plant Biol.* **5** (2002), 250–257.
- [23] S. McNeil, M. Nuccio, M. Ziemak and A. Hanson, *Proc. Natl. Acad. Sci. USA* **98** (2001), 10001–10005.
- [24] J. Unitt, K. Boden, A. Wallace, A. Ingall, M. Coombs and F. Ince, *Bioorg. Med. Chem.* **7** (1999), 1891–1896.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

