Chemometrics

Application Note

Chemometrics Applied to a Metabonomic Study of Mouse Urine

Abstract

The objective of metabonomics is to evaluate physiological response of an organism whose metabolism may have been perturbed. For example, an understanding of the effects of a drug administered to a patient can be accomplished by relating changes in a targeted organ to the collection of metabolites excreted by that organ.

Several analytical technologies have been applied to such studies, including ^{1}H-NMR[1], GC/MS[2], and LC/MS[3] and LC with detection by a coulometric array[4]. In each case, the abundance of data that are generated requires application of sophisticated data handling approaches such as the multivariate algorithms of chemometrics.

Introduction

Information provided by metabonomics studies offers an understanding, via a holistic approach, of the variation of metabolites produced in multicellular organisms and their response to systemic changes. To study these complex relationships, information-rich spectroscopies–*e.g.*, NMR or mass spectrometry–are necessary and are particularly valuable when coupled with multivariate analysis by PCA or SIMCA.

Much of the original work in this field was done with ^{1}H-NMR. However, LC/MS enables detection of different compounds than NMR and at lower concentrations. The techniques for LC/MS are straightforward, and instrumentation is fairly ubiquitous. So long as the LC/MS software can prepare data appropriate for the multivariate analysis package, the approach should be accessible to many labs.

In this study, we evaluate, by chemometrics tools, the variations in endogenous urinary metabolic profiles from three strains of mice (black, white and nude) provided by analysis with LC/MS.

Experimental

Liquid chromatography was performed on a Waters® Alliance® 2795 instrument equipped with a 2.1 mm x 10 cm Waters Symmetry® C_{18} 3.5 µm column. The HPLC system was coupled to a Micromass® QTof-Micro™ MS equipped with an electrospray source operating in either positive ion or negative ion mode.

A 20 µL aliquot of diluted mouse urine was injected onto the column, then eluted with a 0-20% (0.5-4 min) and 20-95% (4-8 min) gradient of acetonitrile in 0.1% formic acid. The mass spectrometric data was collected from 100 to 1400 m/z, in positive and negative ion mode.

The LC/MS data were integrated using the Micromass MarkerLynx™ Application Manager software, and the detected peaks from each sample were used to construct a comprehensive list of all components in the analyzed samples. The processed data list was exported for chemometric analysis using Pirouette® version 3.11.

Either 9 or 10 urine samples of each genetic strain were analyzed by LC/MS and chemometrics. Of these, 7 samples of each were retained to form a training set, and the remainder were set aside as an evaluation set.

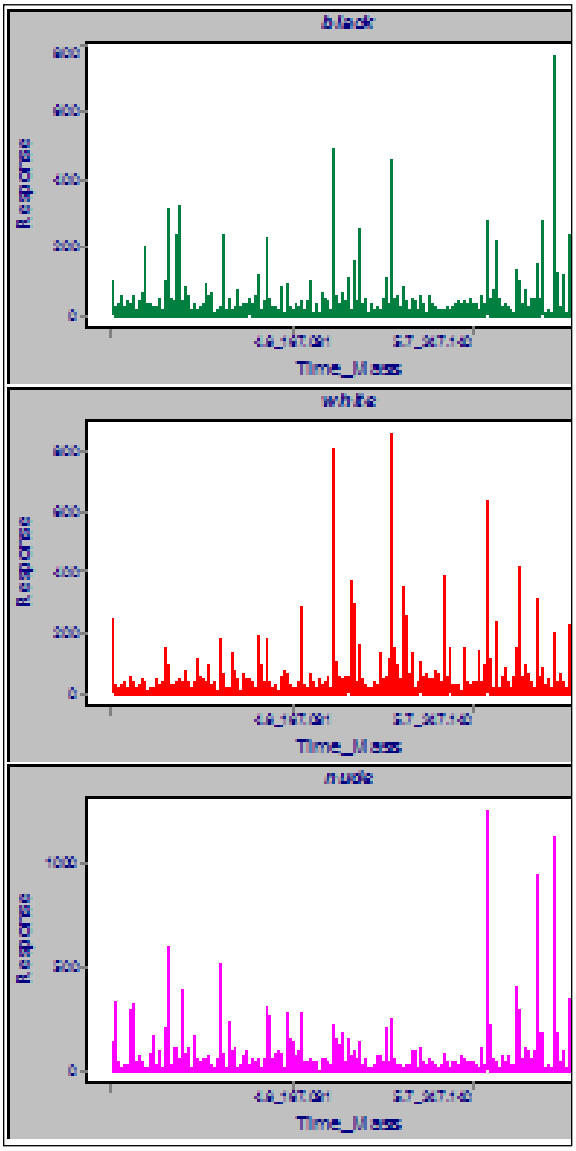

Figure 1. Urinary metabolic profiles (time-mass pairs) of 3 genetic strains of mice

Results

Data from an LC/MS experiment can be saved in a parsimonious format by retaining only values which correspond to peaks, discarding zero responses. Each peak value is associated with a retention time and mass. For comparison of multiple samples simultaneously, a superset of all peaks (time-mass pairs) in all samples must be developed. When displayed graphically, these data do not appear as a chromatographic profile nor a mass spectrum, but multivariate comparisons are still possible. Profiles for the training sets of the three mouse genetic strains are presented in Figure 1.

Clearly, there are differences in the mouse strains which have been captured by the LC/MS analysis. Scores from PCA (principal components analysis) demonstrate that it is possible to distinguish among the 3 strains.

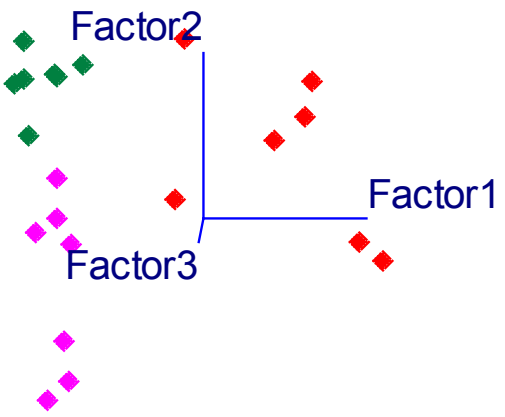

Figure 2. PCA scores of profiles from 3 genetic strains of mice

The SIMCA algorithm (soft independent modeling of class analogy) allows us to create a model from which we can predict category of material, such as to which genetic strain a mouse belongs. The algorithm builds a model by running PCA on each class separately, then compares all samples against these PCA results. Projecting all of the PCA spaces into a single plot is one informative view of the data.

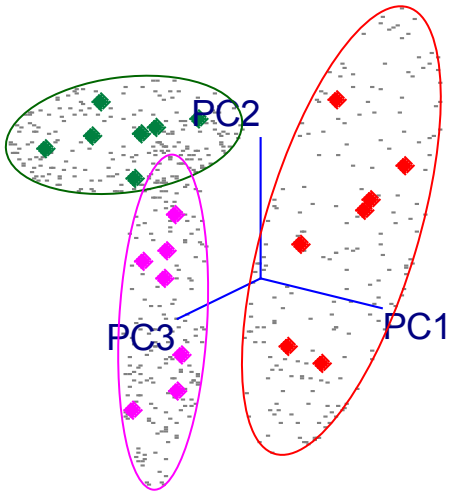

Figure 3. SIMCA projections of the mouse urine profiles, showing the confidence region of each genetic strain category

One of SIMCA's diagnostics is the class distance which measures how well samples are modeled by the different categories. For reliable predictions, values off of the diagonal should exceed 3.

Table 1. SIMCA Class distances

Modeled	Black	White	Nude
Black	0.000	1.847	1.680
White	1.847	0.000	1.375
Nude	1.680	1.375	0.000

This SIMCA model is now ready to use for category predictions of unknown mouse urine. First, the model is applied to the evaluation subset previously set aside (see table at right). All but one of the evaluation samples are correctly predicted.

Sample	Category
1	3
2	1
3	0
4	2
5	1
6	0
7	2

The class projections of these samples is shown below. Sample 3 is found near the boundary of two categories, not properly fitting either, thus the result shown in Figure 4. Ideally, more samples would be analyzed from each category of material to better characterize the natural variation within the categories.

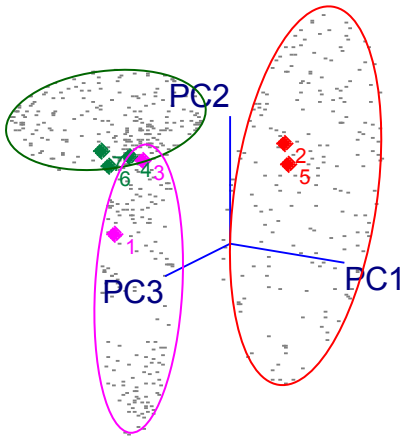

Figure 5. SIMCA projections of profiles of evaluation samples, showing the confidence region of each category from the modeling step. .

Prediction of categories can be handled easily for small numbers of samples in the Pirouette interface. When numbers increase, or when routine predictions are desired, InStep offers a more automated approach.

An InStep™ method describes the decision making logic applied during prediction. Our 3-category problem is a simple case, illustrated below.

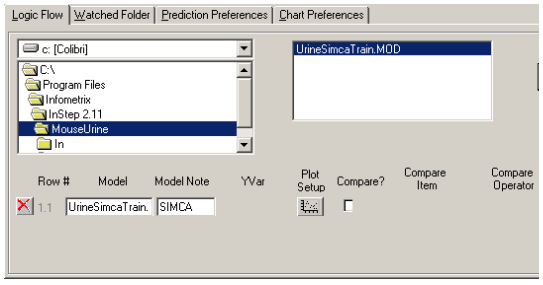

Figure 6. InStep method incorporating a SIMCA model for prediction of mouse genetic strain

InStep permits formatting of its report with great flexibility. For this application, simplicity was chosen for the report items.

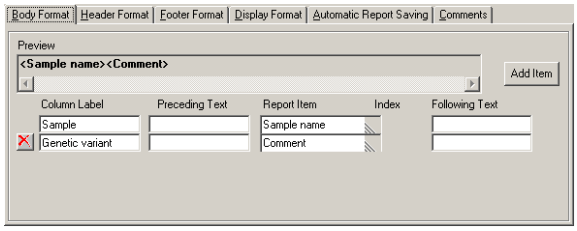

Figure 7. Selection of report body items

The report can be enhanced by adding custom strings based on the results computed by InStep.

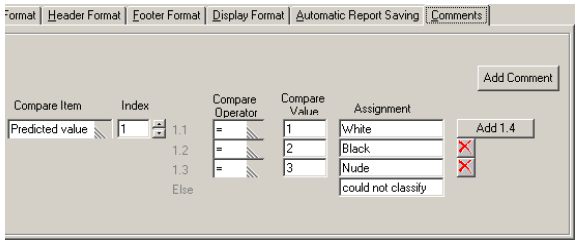

Figure 8. Customizing the report based on results

Finally, InStep is invoked and after processing has started, it will watch a specified folder for incoming files of a designated file type. The file will be processed, then stored in a separate folder, and the report updated, continually, until processing is stopped. This simple report for the evaluation samples is shown below.

Table 2. Basic InStep output showing prediction of evaluation samples

Sample	Genetic strain
N3	Nude
W11	White
N6	could not classify
B1	Black
W15	White
B10	Black

Because InStep has flexibility in string formatting, custom reports are easily developed. As an example, HTML tags can be inserted such that the resulting report can be viewed from a browser, as in Figure 9.

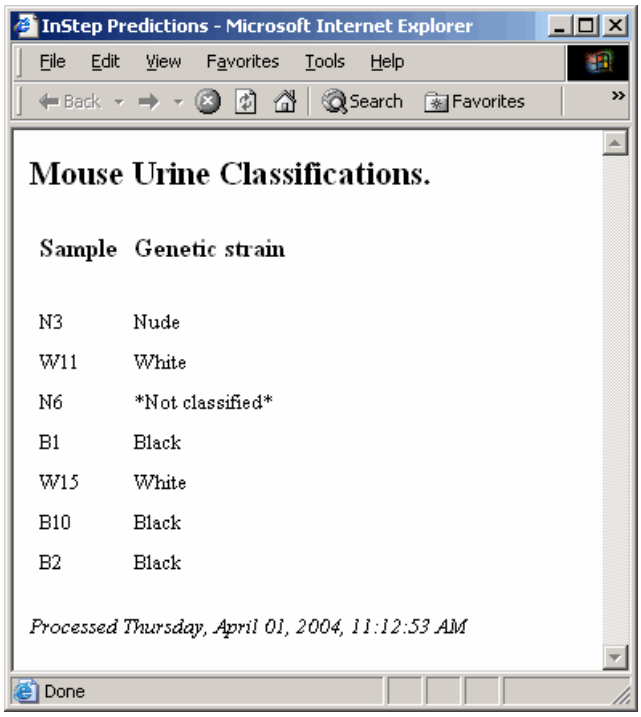

Figure 9. InStep output formatted for and viewed from a web browser.

Conclusions

Multivariate analysis can assist the interpretation of complex spectrometric data generated from metabonomic experiments. Predictions of material type, such as the genetic strain of mice, can be accomplished in an automated fashion, with minimal user interaction.

References

1. E. Holmes, A.W. Nicholls, J.C. Lindon, S. Ramos, M. Spraul, P. Neidig, S.C. Connor, J. Connelly, S.J. Damment, J. Haselden, and J.K. Nicholson. Development of a model for classification of toxin-induced lesions using 1H NMR spectroscopy of urine combined with pattern recognition. *NMR Biomed.*, *11*: 235-244 (1998).

2. P. Jonsson, J. Gullber, A. Nordstrom, M. Kusano, M. Kowalczyk, M. Sjostrom, and T. Moritz. A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal. Chem.*, *76*: 1738-1745 (2004).

3. R. Plumb, J. Granger, C. Stumpf, I.D. Wilson, J.A. Evans and E.M. Lenz. Metabonomic analysis of mouse urine by liquid-chromatography-time of flight mass spectrometry (LC-TOFMS): detection of strain, diurnal and gender differences. *Analyst*,*128*: 819–823 (2003).

4. R. Kaddurah-Daouk, C. Beecher, B.S. Kristal, W.R. Matson, M. Bogdanov and D.J. Asa. Bioanalytical Advances for Metabolomics and Metabolic Profiling. *PharmaGenomics*: January: 46-52 (2004).