technical note

Understanding Metabonomics Mixtures with PCA

L. Scott Ramos, Infometrix, Inc., Suite 250,
10634 East Riverside Drive, Bothell, WA 98011

Abstract

In toxicology studies, alterations to the intrinsic *in vivo* biochemical profile of an animal can be monitored via NMR spectroscopic analysis of urine. Chemometric processing of the NMR data can enhance understanding of the toxic effects; in particular, examination of the principal component analysis (PCA) scores provides insight into these data in a time-efficient manner.

Method

Standardized rats were treated with hydrazine in low and high doses. Urine was collected from these and untreated rats at 24 hour intervals after dosing. ^{1}H NMR spectra of urine samples were obtained at 600 MHz on a Bruker NMR spectrometer. Spectra were added to a KnowItAll® database with intensities normalized but with no further filtering.

Spectral data were truncated to a range of 10 to 0.5 ppm, excluding the range of 4.5 to 6.5 ppm to remove variation due to solvent suppression and the wide urea peak. Data analysis was performed with KnowItAll Informatics System's AnalyzeIt™ MVP, and PCA parameters included normalization to area % followed by mean-centering.

Discussion

An overlay of a collection of the NMR spectra shows the complexity of the samples and the variation in intensity.
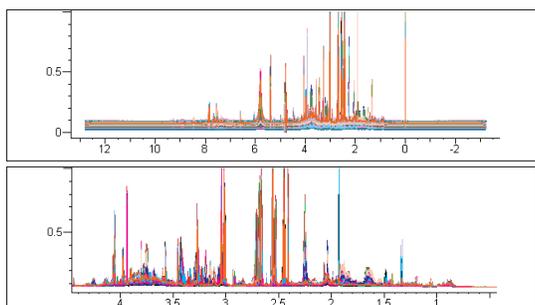
Figure 1. Overlay of NMR spectra. Upper trace: full spectral range; lower trace: diagnostic region.

Following PCA, the scores show that samples treated with the higher dose diverge in a common direction, most control and low dose samples cluster in a single group, and a few low dose samples diverge in a direction very different from the high dose samples.
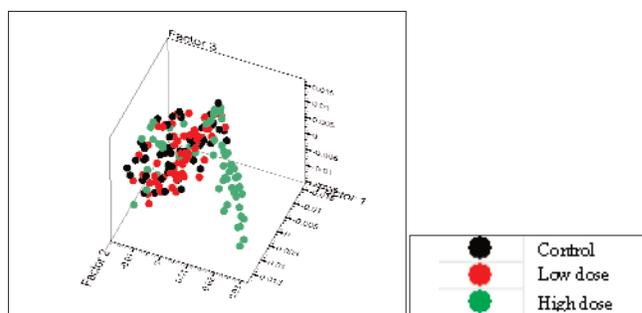

Figure 2. PCA scores of NMR spectra from urine of treated and control animals.

This, however, does not tell the whole story. When the drug induces the largest effect in the animal, the urine carries corresponding metabolites, and the NMR spectrum will differ accordingly. Each individual animal responds in a slightly different time frame, so that the time of maximum response after dosing will vary.

The high dose samples were clearly distinguished from the control and low dose samples. To see whether it was possible to separate the low dose samples from the controls, a subset of the original data was made by selecting the control and low dose properties in the legend. This highlights all samples that fall into these two categories. The samples were transferred to the MineIt™ application in the KnowItAll Informatics System as a hit list, then this hit list was transferred back to AnalyzeIt MVP for PCA analysis. The scores plot was shown and the dosing level was again applied to the plot to distinguish among the control and low dose samples. In addition, the time after dosing property was added as a point size categorization, which allows distinguishing among samples based on time.

BIO-RAD   InfoMetrix

For example, samples collected at different time points are marked in the following figures. The AnalyzeIt MVP legend allows samples categorized by either categorical or continuous property values to be highlighted separately.
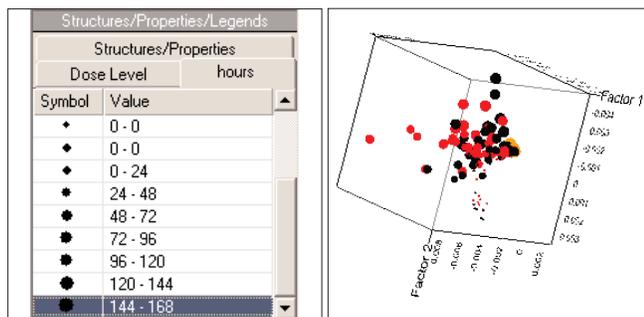

Figure 3. *Legend showing time of collection as selection tool for scores plot sample points, and corresponding scores plot with points highlighted in orange for samples collected after seven days.*

Different time points were selected, and the corresponding scores plots are shown in Figure 4. Note how the orange points, which correspond to the times indicated, move away over time from the central cluster, which contain the control samples and the predose sample collection times, toward the outer reaches of the scores. This trajectory of points as a function of time correlates to the increasing effect of metabolizing the drug until recovery. Because each animal responds in a slightly different way, the trajectories are not coincident, but the trends are usually similar. When the drug is eliminated, after six days, all sample points cluster together again, although at a point offset from the original, predose position (see Figure 3).
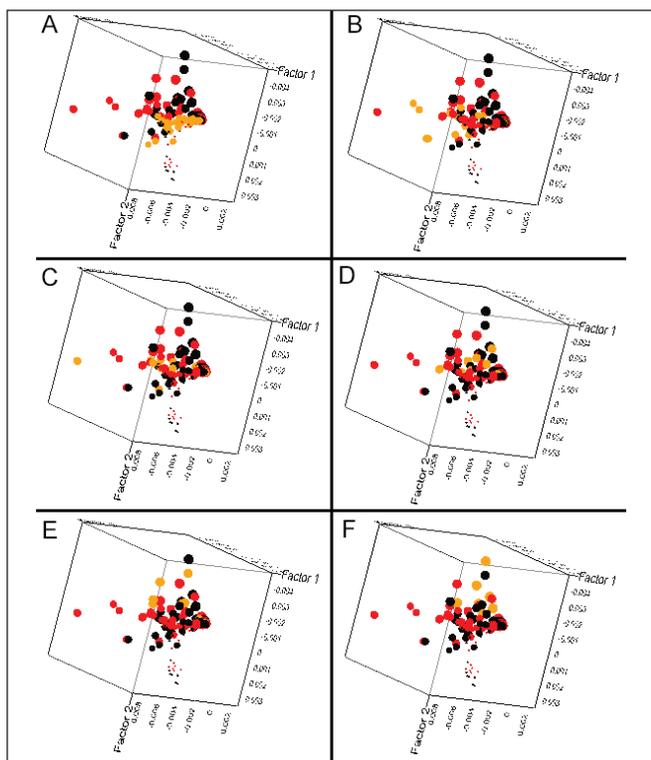

Figure 4. *PCA scores where points in orange indicate samples collected at (a) 24h, (b) 48h, (c) 72h, (d) 96h, (e) 120h, and (f) 144h.*

We can use the dispersion of points in the scores plot to help us understand the differences in the NMR spectra. Points at the extremes should exhibit characteristic spectra that relate to a particular effect. Three points have been highlighted (see Figure 5) among the scores and their corresponding spectra are shown in Figure 6.
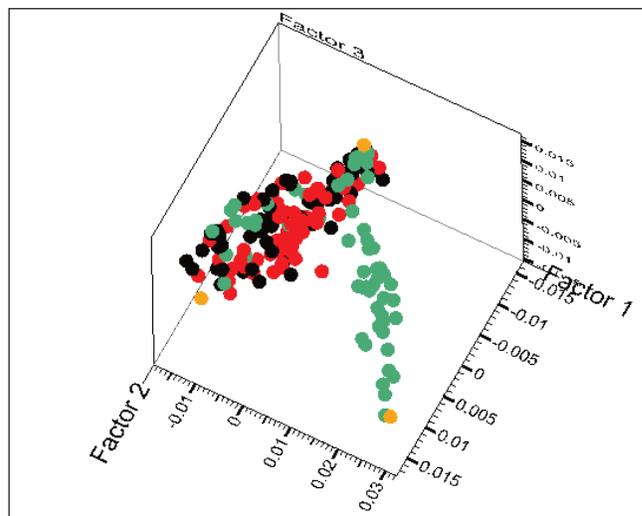
Figure 5. *Three points in the PCA scores plot are highlighted in orange and represent the extremes of the scores space.*

The first profile shows a typical control sample with the resonances from citrate, 2-oxoglutarate and succinate present in the center. The second profile exhibits the classic pattern of markers for hydrazine hepatotoxicity, namely taurine, creatine and 2-aminoadipate. Finally, the last profile shows the classic pattern of contamination with the largest peak being acetate; and the two peaks to the right are lactate and ethanol.
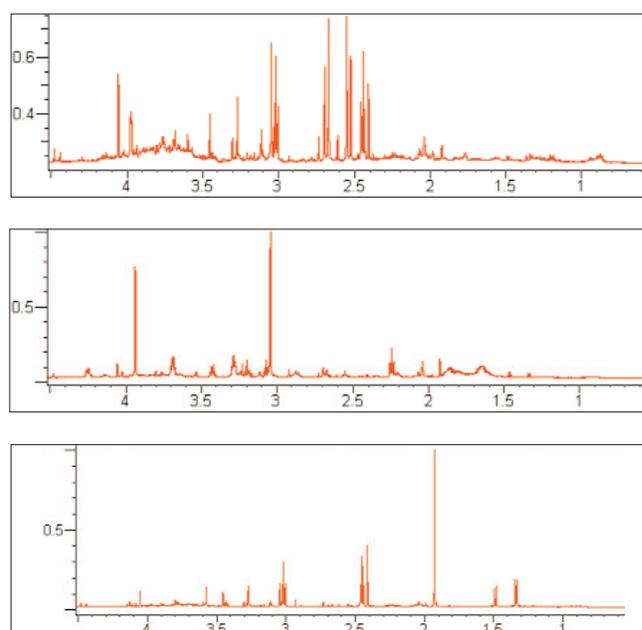
Figure 6. *Spectra from points at extremes in the scores plot, representing different features in the database.*

Conclusions

Combining the capabilities of a database with principal component analysis gives the analyst the capability of evaluating trends in NMR quickly and efficiently. The PCA work leads directly to the identification of key profiles, simplifying the effort of sifting through hundreds or even thousands of spectra. In addition, this analytical approach is an efficient means of identifying underlying biochemical events (control, toxicity, time-related change, etc.). The study of toxins or drugs in development can benefit substantially.

Acknowledgement

The data and background for this application note were supplied by Dr. Andrew Nicholls of GlaxoSmithKline, London.