A New Approach to Sample Class Segregation Based on Application of Bayesian Spectral Decomposition to Metabonomic NMR Data Sets

R. Stoyanova[1], J. C. Lindon[2], J. K. Nicholson[2], T. R. Brown[3]

[1]Division of Population Studies, Fox Chace Cancer Center, Philadelphia, Pennsylvania, United States, [2]Biological Chemistry, Imperial College, London, United Kingdom, [3]Hatch Center for MR Research, Columbia University, New York, NY, United States

Introduction

NMR-based metabonomic analysis of biofluids and tissues combines high-field NMR spectroscopy with chemometric methods and has proved to be a valuable technique in characterizing and predicting the nature and target of toxicity. This emerging technology is complimentary to genomic and proteomic analyses. All three disciplines generate large data-rich matrices, necessitating the application of appropriate multivariate statistical algorithms. The major goal of these techniques is classifying the biological samples based on differences in treatment. Here we present a novel approach for class separation, Bayesian Spectral Decomposition (BSD)[1], which, in addition to classification of the samples, provides a direct link between the spectral changes caused by the experimental variables. BSD assumes each sample to be a mixture of components with individually associated basic spectral shapes so that each spectrum is the sum of varying amounts of these basic spectra. Both the spectral shapes and their magnitudes are simultaneously determined. The magnitudes of the spectral shapes are related to other information available for the samples, such as time and dose of administered toxin, etc.

Materials and methods

BSD is applied to ^{1}H NMR spectra of urine from Han Wistar rats in a hydrazine experiment[2]. Hydrazine causes histopathological alterations in the rat liver, which are evident within 24 h following its administration. The rats were assigned to four dose groups: control, 75, 90, and 120 mg/kg. Urine samples were collected pre-dose (-8 to 0h), 0-8h, 8-16h, 16-24h, 24-32h, 48-56h, 72-80h, 96-104h, 120-128h, and 144-152h post-dose. One-dimensional ^{1}H NMR spectra of urine were measured at 600.13 MHz on a Bruker DRX-600 spectrometer. The integral of creatinine was used to normalize the data for the amount of material in each sample. A procedure for individual peak alignment[3] was applied to the spectral region between 2.20 and 3.63 ppm (8051 points). BSD was applied to the real part of the spectra.

A common issue for classification is the number N of sought patterns, features, classes, which in many cases is uncertain. The advantage of BSD is that the posterior probability, which is related to the product of the likelihood and the prior is low, if the model becomes unnecessary complicated. Bayesian model selection is essentially a quantitative statement of Occam razor principle: when two models fit the data equally well, prefer the simpler model. We describe results for N = 2, 3 and 4.

Results

The solutions for N=2 are presented in Figure 1. Their magnitudes in each dose-group, as a function of time, are presented in the upper graph, and the identified spectral patterns – below. The difference between the treated and control rats is clear: while the first shape (black diamonds) maintains almost a constant high level in the untreated rats, the magnitude of the second (open circles) is very low throughout the experiment. Alternatively, significant changes in the magnitudes of these shapes occur for the treated animals: the magnitude of the first shape decreases in response to hydrazine and it recovers after 36 hours (with the exception of the 120 mg/kg treated rat, which had to be sacrificed prior to acquiring a full series of spectra), while the second shape exhibits a reciprocal behavior during the experimental cycle. The first spectral pattern contains Krebs cycle intermediates: citrate, succinate. The second is comprised of 2-aminoadipic acid, taurine and creatine. For N=3 a new pattern is identified, which is associated with the 'aberrant' behavior in response to treatment. In contrast to the "aberrant" spectral pattern associated with a severe toxicity response (Fig. 1), this pattern maintains a very similar behavior for all doses in the treated rats, suggesting that it may be rather associated with biological processes that are only moderately responsive to the treatment. Finally, for N=4 the 'normal' spectral pattern is split in two: the one consists entirely of 2-oxoglutarate, and the second is comprised of the remaining Krebs cycle intermediates. The observed differences in the temporal behavior of these patterns suggests that distinct anti-toxicity and/or regenerating response mechanisms are triggered in a sequentially temporal manner following treatment with hydrazine.

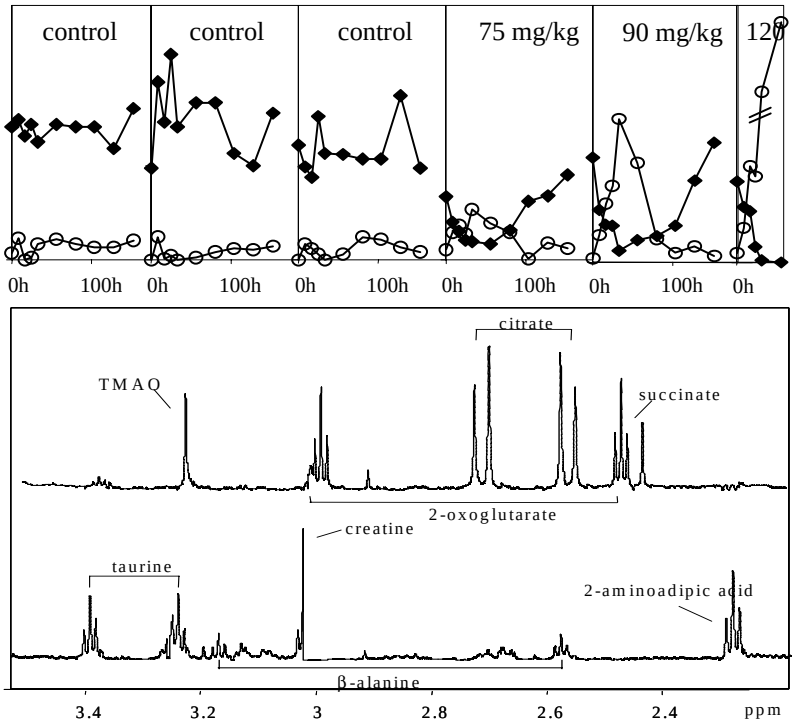

Figure 1. (upper) Magnitude of two spectral patterns across the treated and not treated rats as a function of time: open circles are associated with the aberrant pattern (last two points of treatment with 120 mg/kg are off the scale of the graph) and filled diamonds are associated with the normal urine pattern; (lower) spectral patterns, identified by BSD: normal (upper) and aberrant (lower).

Conclusions

The presented results demonstrate the potential of BSD in combination with thorough preprocessing of the data to identify the underlying basic spectral shapes and connect them quantifiably with the experimental variables. With more than the expected two classes (control and treatment) an intricate picture of the underlying metabolic processes emerge. Connecting this information with known biochemical pathways will produce thorough understanding of the basic toxicological or other processes.

References

[1]Ochs, M.F., Stoyanova, R.S., Arias-Mendoza, F., and Brown T.R. *J. Magn. Reson.* **137**, 161-176 (1999)

[2]Nicholls, A.W., Holmes, E., Lindon, J.C., Shockcor, J.P., Farrant, R.D., Haselden, J.N., Damment, S.J., Waterfield, C.J., Nicholson, J.K. *Chem.Res.Toxicol.,* **14**, 975-987 (2001)

[3]Stoyanova, R.S, Nicholls, A.W., Lindon, J.C., Nicholson, J.K., Brown T.R. *ISMRM* (2001).