# Identifying Metabolites in Biofluids

Colin Vitols, Ryan Rosewell

March 2006

*In this note we present a rapid, efficient method for identifying metabolites in biofluid NMR spectra using targeted profiling. Conventional techniques for identifying and quantifying metabolites in such spectra are labor-intensive and error-prone, as positions and linewidths of peaks can vary widely with changes in pH and other solution matrix effects. The length of time and level of operator skill needed to analyze large numbers of these complex spectra (Figure 1) are significant barriers to the widespread application of NMR in metabolomics.*

## *Introduction*

Variability in spectra due to peak position and linewidth changes is often handled by a form of data reduction known as "spectral binning" or "spectral bucketing" [1] [2]. This method involves dividing the spectrum into a number of regions, or bins (K), and considering the total integrated area within each bin for further analysis. The bins can be of a fixed width or they can be of variable size, using manual inspection for the best optimization or automated algorithms for high throughput [3]. Use of spectfral binning is based on the premise that considering regions of the spectra, as opposed to individual data points, can compensate for minor peak shifts and linewidth differences for the same compound across multiple samples (N).

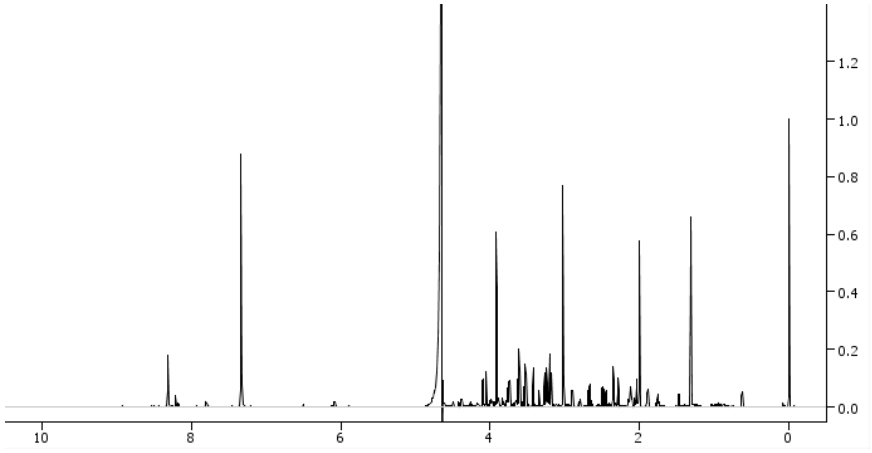


**Figure 1. Analyzing complex NMR spectra like this 800 MHz spectrum of a brain tissue extract can be a daunting task.**

In practice, sensitivity to sample conditions, along with baseline and other spectral distortions that can arise during an experiment, mean that the bin integrations do not necessarily reflect true changes in spectral area [4] [5]. Since pattern recognition techniques such as principal component analysis (PCA) depend on linear combinations of the input spectral bins, artifacts in the integrations of input bins will compromise the analysis. If the metabolites of interest vary subtly, peak position and linewidth differences between samples, as well as instrumental and spectral artifacts, may mask significant changes to their concentrations. Applying statistical tools such as orthogonal signal correction to regression-type experiments has been shown to eliminate components of the data which are not relevant to the analysis, including the effects of physiological and instrumental variation [6].

In spite of these advances in spectral pre-processing algorithms and pattern recognition methods for spectral binning data, little information is available about individual metabolites and their concentrations in the sample. Any statistical treatment of NMR spectra in metabolomics is based on the idea that metabolites are the actual variables of interest. The ideal statistical treatment would directly involve

the concentrations of all metabolites in the samples, since these values represent the underlying physical model that generates the observed NMR spectrum. Such a treatment would allow analysis of compounds either jointly, for total characterization of the mixture, or selectively, for targeted characterizations such as pathway modeling. In this case, there would still be N samples, but each variable K would be the concentration of a single metabolite.

Targeted profiling is an approach to data reduction that involves comparison to NMR spectral signatures of individual metabolites found in a reference database. This technique works by reducing spectral data to quantified metabolites, providing input variables for pattern recognition tools such as PCA or projection to latent structures (PLS). The approach reduces the dimensionality of the problem space compared to spectral binning, as assignment of all protons in a compound will show all spectral regions correlated to that compound. As a result, a variety of approaches to targeted profiling have recently been developed for both *in vivo* and *in vitro* NMR [4] [7] [8]. This method is most useful when investigating predefined compounds of interest.

## *Targeted Profiling*

Chenomx NMR Suite uses targeted profiling to reduce analysis time, combining advanced analysis tools with a compound library of more than 290 common metabolites. Targeted profiling works with NMR spectra of virtually any complex mixture, including urine, blood serum, saliva and various cell extracts. Data reduction methods such as spectral binning are not necessary; in fact, they reduce the quality of analysis possible with targeted profiling by abstracting away key details of the spectrum.

Before performing an analysis using Chenomx NMR Suite's targeted profiling, it is important to accurately define and characterize the chemical shape indicator (CSI) in a sample using the Processor module, in order to obtain accurate analysis results. After defining appropriate CSI settings, you can move the spectrum to the Profiler module to begin analysis.

One of the strengths of targeted profiling is selectively analyzing for particular compounds. You can consider each compound independently of the rest of the metabolite library, generating results only for compounds of specific interest. You can then export the targeted results for further analysis using statistical or other data analysis software, without the need to completely characterize the entire spectrum.

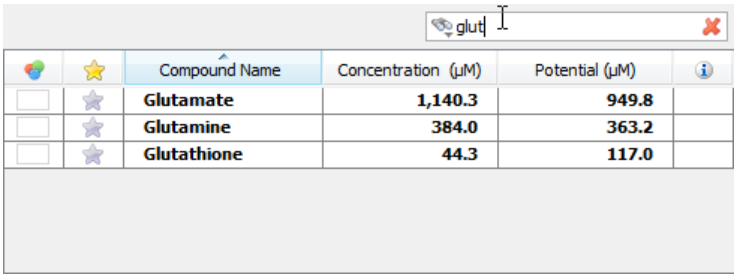


**Figure 2. The compound table can be filtered using text-based quick searches.**

You can select compound signatures to profile in an experimental spectrum in two ways: You can determine a set of compounds in advance, and selectively analyze the spectrum for the selected compounds, or you can select peaks and clusters from the spectrum and correlate them to individual compound signatures.

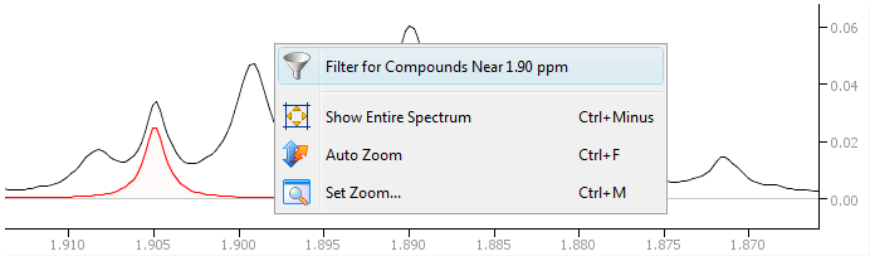


**Figure 3. The frequency filter displays compounds that occur near the selected location in the compound table.**

Predetermining compounds for analysis allows pursuing precisely-defined studies, explicitly targeting a known set of compounds. This approach could be applicable in studying a particular metabolic pathway, monitoring the known metabolites of a prescription medication or even simply focusing a study on only the most commonly-detected metabolites. Quick searches (Figure 2) help you find compounds, display their signatures and fit them to the spectrum.

Selecting compounds based on peaks and clusters seen in the spectrum permits a more open-ended approach, treating the analysis of the spectrum as a discovery process. Such an analysis could help discover metabolic implications for a new medication, establish novel biomarkers for disease states, or prepare detailed metabolic profiles for longitudinal research studies. The frequency filter (Figure 3) can provide a helpful shortlist of compounds that may appear in a selected region.

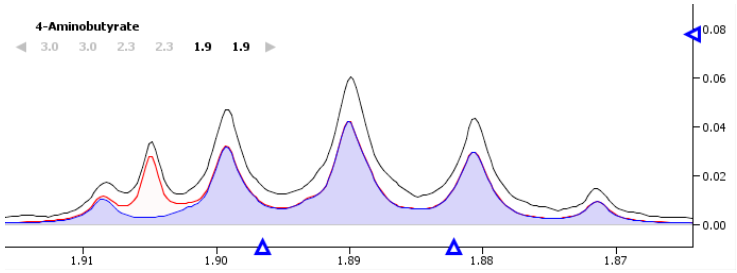


**Figure 4. Visual inspection is often enough to confirm or reject the presence of a compound in a sample.**

## *Fitting Compounds*

The metabolite libraries supplied with Chenomx NMR Suite contain dynamic compound signatures that you can individually fine-tune to provide the best possible fit to an experimental spectrum. You can select and profile compounds through a simple click-and-drag interface. A compound signature initially appears at zero concentration. Dragging the height control upward increases the height of all of the clusters in the compound signature, allowing a direct visual comparison with the corresponding peaks in the experimental spectrum (Figure 4). This is often enough to determine whether the compound is a likely component in the sample.

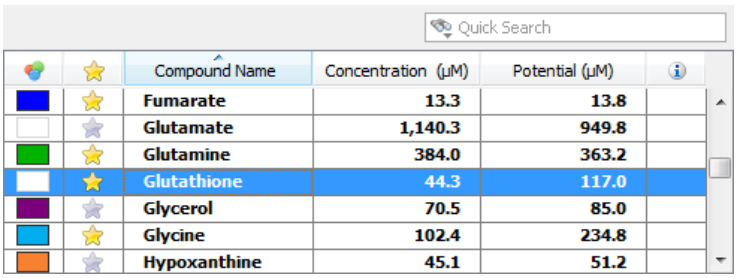


**Figure 5. The compound table displays the results of the fitting process, including concentrations.**

Profiler calculates compound concentrations based on comparison to the CSI settings that you defined for the sample. As the signature height changes, the resulting calculated concentration displayed in the compound table updates dynamically (Figure 5). The initial locations of the clusters in a compound signature are often not exactly matched to their actual locations in the spectrum. These discrepancies are typically due to solution matrix effects, such as sample pH or ionic strength. To compensate, you can move each cluster in a compound signature to a new chemical shift, independently of the other clusters.

As you fit each cluster to the spectrum, the sum line updates to reflect the changes made. For accurate analysis, you should adjust the sum line via changes to individual compounds to match the spectrum as closely as possible. This is especially important in crowded regions, as several compounds may contribute to the overall fit at a given location (Figure 6). As a result, you may need to adjust several compounds to properly adjust the fit in these crowded regions. When the fit is complete, the compound table contains a list of all compounds identified and their respective concentrations.

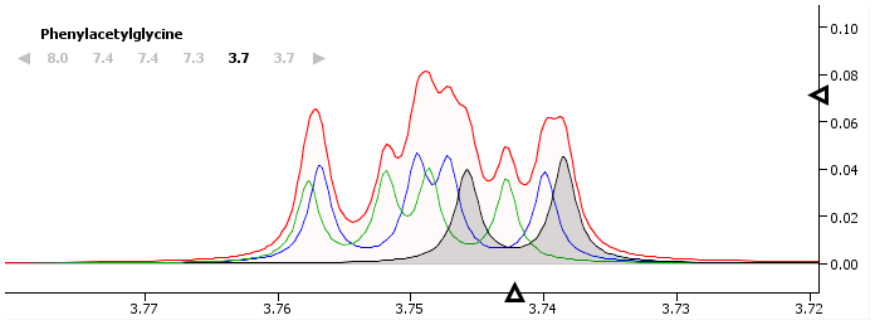


**Figure 6. Strong overlap can yield a sum line (red) that is quite different from the contributing compounds (blue and green).**

After you have profiled all the compounds that you wish, you may find further information useful to place the identified compounds in context. You can access detailed information about a particular compound via the reference panel which contains structures, alternate names, CAS-RN, and links to external databases, including KEGG Ligand [http://www.genome.ad.jp/kegg/ligand.html], ChEBI [http://www.ebi.ac.uk/chebi/] and PubChem [http://pubchem.ncbi.nlm.nih.gov/].

For further analysis, you can export results from single or multiple spectra as a tab-delimited text file (.txt) for import into statistical analysis software such as Umetrics. You can also import the analysis from a profiled spectrum into a new spectrum in the same dataset as a starting point for your analysis.

## *Conclusion*

Targeted profiling in Chenomx NMR Suite is a versatile technique that may be used both as a primary research tool and as a component of a larger workflow. Existing spectra can be imported directly, or processed in one of several major NMR processing packages and imported afterward. PCA can be applied to the concentration data instead of the raw spectrum, providing principal components as specific compounds instead of just important regions that must later be identified. With targeted profiling, identification of compounds is an integral part of the analysis process, instead of a time-consuming post-processing step.

## *Acknowledgements*

Chenomx would like to thank Brent McGrath and Dr. P. Silverstone in the Department of Psychiatry at the University of Alberta for providing data for this application note.

## *References*


[1] KPR Gartland, CR Beddell, JC Lindon, and JK Nicholson. **1991.** *Application of pattern recognition methods to the analysis and classification of toxicological data derived from proton nuclear magnetic resonance spectroscopy of urine*. *Molecular Pharmacology.* 39:629-642

[2] ML Anthony, BC Sweatman, CR Beddell, JC Lindon, and JK Nicholson. **1994.** *Pattern-recognition classification of the site of nephrotoxicity based on metabolic data derived from proton nuclear magnetic resonance spectra of urine*. *Molecular Pharmacology.* 46:199-211

[3] CD Eads, CM Furnish, I Noda, KD Juhlin, DA Cooper, and SW Morrall. **2004.** *Molecular factor analysis applied to collections of NMR spectra*. *Anal Chem.* 76:1982-1990

[4] DJ Crockford, HC Keun, LM Smith, E Holmes, and JK Nicholson. **2005.** *Curve-fitting method for direct quantitation of compounds in complex biological mixtures using H-1 NMR: Application in metabonomic toxicology studies*. *Anal Chem.* 77:4556-4562

[5] BC Potts, AJ Deese, GJ Stevens, MD Reily, DG Robertson, and J Theiss. **2001.** *NMR of biofluids and pattern recognition: Assessing the impact of NMR parameters on the principal component analysis of urine from rat and mouse*. *J Pharm Biomed Anal.* 26:463-476

[6] BM Beckwith-Hall, JT Brindle, RH Barton, M Coen, E Holmes, and JK Nicholson. **2002.** *Application of orthogonal signal correction to minimise the effects of physical and biological variation in high resolution 1H NMR spectra of biofluids*. *Analyst.* 127:1283-1288

[7] SW Provencher. **2001.** *Automatic quantitation of localized in vivo H-1 spectra with LCModel*. *NMR in Biomedicine.* 14:260-264

[8] FJ Bamforth, V Dorian, H Vallance, and DS Wishart. **1999.** *Diagnosis of inborn errors of metabolism using 1H NMR spectroscopic analysis of urine*. *J Inherit Metab Dis.* 22:297-301