

Complex Mixture Analysis by NMR Spectroscopic Targeted Profiling

Shahriar Jahanbakht, Patrick Wheeler, Brent Pautler, and Steve Hayward

Advanced Chemistry Development, Inc.

Toronto, ON, Canada

www.acdlabs.com

Introduction

All sorts of mixtures need to be analyzed in science for a variety of reasons. The use of NMR spectroscopy in mixture analysis has greatly increased in recent years, both because of the development of improved methods, and because it can be used to both identify and quantify chemicals from complex mixtures. One specific application that deals with mixtures in difficult matrices is the study of metabolomics. Determining concentrations of various endogenous and exogenous metabolic species in biofluids presents many challenges. For example, it is necessary to be able to identify and quantify peaks for species which may overlap with one another and with strong background signals.

For more general cases, mixture component identification can be achieved by searching spectral libraries, through the use of prediction tools, or with characterization tools such as ACD/NMR Workbook Suite or ACD/Structure Elucidator Suite.

The specific application described here is a targeted approach in which ^1H NMR spectroscopy is employed for quantification. Due to sample complexity such as overlapping peaks, the quantification of a large amount of data remains very time consuming. However, we have developed a semi-automated process using the peak-fitting, macro, and scripting capabilities of NMR Workbook Suite which can be applied to a large series of data in a fast, reproducible manner. Outside of metabolomics, these techniques and tools will also find ready utility in fermentomics and other similar chemometric applications.

Discussion

The process of identifying and quantifying metabolites in biofluid spectra by NMR starts with the initial data collection and processing. As these spectra often contain many thousands of peaks, the data are often simplified by dividing each spectrum into small integral regions called “buckets”. These buckets are subjected to a variety of statistical analyses to determine correlations between the integral values and sample characteristics, e.g., treatment groups. An attempt is then made to identify if the buckets responsible for the correlation belong to a particular metabolite or set of metabolites. The bucket integral values are used as an estimation of the metabolite’s concentration, while information regarding biological pathways is often inferred from this data. However, on the basis of the entire statistical study, these bucket integral values do not necessarily contain only a single peak. Reducing the bucket size can help; however the intelligent bucketing algorithm in ACD/Labs software can circumvent the problem of splitting peak intensities into different buckets. While this algorithm is very robust, alone it does not solve the issue of peak overlap. For this issue, peak deconvolution algorithms such as the one included



in NMR Workbook Suite have the potential to provide better quality integration data for metabolomics. This approach can be applied to both a small area of the spectrum or an entire spectrum, and in batch mode with large data sets.

Our approach is to create a model spectrum which contains the metabolites of interest. The deconvolution algorithm is applied to the model and higher and area of each peaks are recalculated.

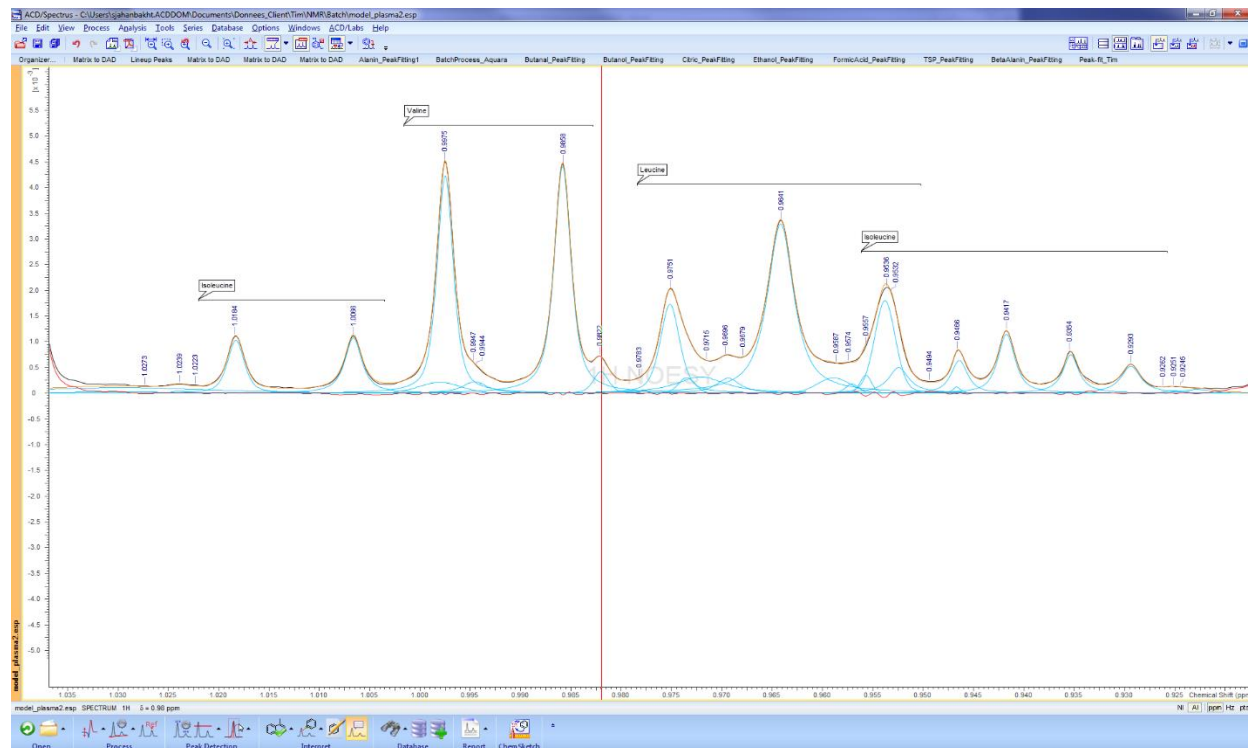


Figure 1. A view of the model. In NMR Workbook Suite the user can prepare the model, and identify the multiplets using the Database Search and/or NMR Prediction tools.

The signals and assignments are then transferred from the model spectrum to the samples in batch mode, and deconvolution applied to the data.

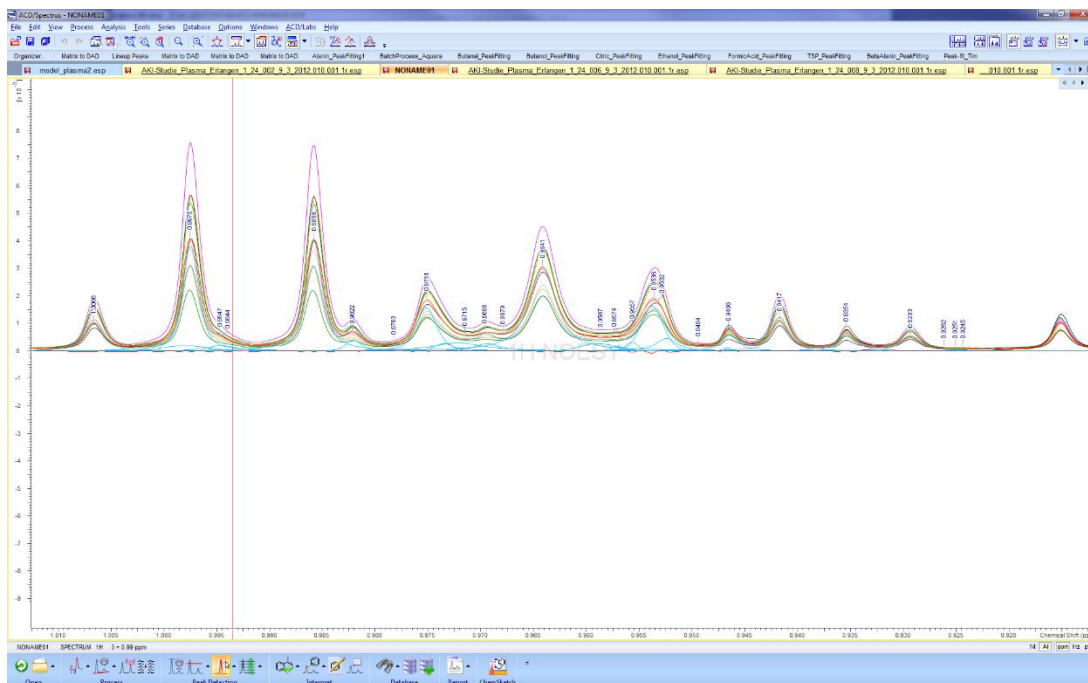


Figure 2. A view of a data series. Peaks and assignments were transferred to the samples from the model.

Next, quantification scripts are applied to calculate absolute area for each multiplet. It is possible to calculate the relative area for each multiplet using an internal reference. Peaks and multiplets can be correlated to quantitate specific mixture components, such as metabolites.

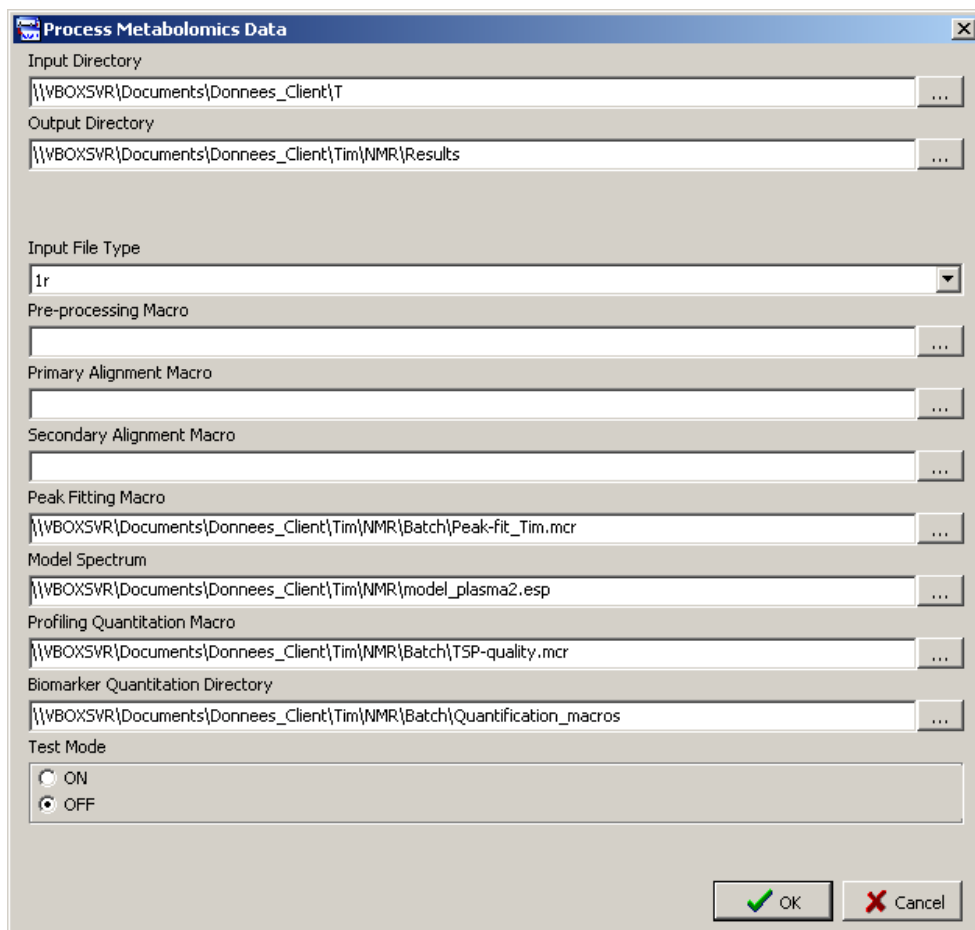


Figure 3. The Batch Mode interface for processing spectral components.

The results are presented in an Excel table (Figure 4).

File	tas	FID	TSP	Alanin_Quality	Alanin_rel	Alanin	Glucose_Quality	Glucose_rel	Glucose
angen\ng		10	2373058560	1.06557	0.014011	33249775	1.024806	0.094014	223101088
langen\ng		10	2373104384	1.065318	0.014062	33370353	1.024779	0.094227	223610584
angen\ng		10	2470617600	1.039624	0.020517	50689712	1.027985	0.11911	294274192
langen\ng		10	2472076288	1.03976	0.020429	50502104	1.027973	0.119219	294717776
angen\ng		10	2293802496	1.069346	0.016986	38963658	1.025381	0.081689	187378056
langen\ng		10	2299908096	1.071292	0.017164	39476762	1.025359	0.081674	187843048
angen\ng		10	2398321408	1.044678	0.02363	56672812	1.032382	0.097723	234370864
langen\ng		10	2392332032	1.046699	0.023871	57107770	1.032399	0.098166	234846736
angen\ng		10	2301764352	1.092352	0.017123	39412080	1.011525	0.089302	205552304
langen\ng		10	2301522944	1.092065	0.017174	39527264	1.011514	0.089515	206020864
angen\ng		10	2384862976	1.075839	0.031641	75460376	1.030552	0.119711	285494320
langen\ng		10	2383528192	1.076649	0.031817	75836580	1.030527	0.119975	285962944
angen\ng		10	2424466176	1.057077	0.022554	54680550	1.033445	0.058447	141701728
langen\ng		10	2425392128	1.05722	0.022697	55049676	1.033392	0.0586	142126960
angen\ng		10	2316644352	1.058646	0.020779	48137018	1.031173	0.060672	140555840
langen\ng		10	2316005376	1.058754	0.020936	48488990	1.031117	0.060888	141017112
angen\ng		10	2316712192	1.031728	0.022474	52064694	1.027184	0.065091	150796080
langen\ng		10	2310954240	1.032952	0.022674	52397590	1.027152	0.065456	151265848
angen\ng		10	2325526784	1.03321	0.022625	52614096	1.030084	0.058591	136255000
langen\ng		10	2326567424	1.033295	0.022799	53044420	1.029977	0.05879	136779768

Figure 4. A view of the results in Excel, simplified for legibility



For verification purposes, all processed spectra are then saved in both ACD/Labs and JCAMP data formats.

Conclusion

The process of quantification metabolites in biofluid spectra can be substantially automated with the use of ACD/NMR Workbook Suite and its add-ons.