Software Tool for Comprehensive Assessment and Interpretation of Metabolomic Data

Q. Zhao[1], R. Stoyanova[2], S. Du[1], P. Sajda[1], T. R. Brown[1]

[1]Columbia University, New York, NY, United States, [2]Fox Chase Cancer Center, Philadelphia, PA, United States

Introduction: NMR spectroscopy can provide a wide range of information for metabolic characterization of biological samples. However, the quantity and complexity of spectroscopic data obtained from metabolomic studies has made data interpretation very difficult. The challenges stem not only from the need to apply standard processing procedures to large sets of spectra and connect them with auxiliary information about the samples, but also the frequent necessity to invoke pattern recognition (PR) methods to ensure optimal information retrieval. Most existing software deals only with a few of the aspects of this process and in general acts on a single spectrum.

Here we present a software tool, High Resolution Spectroscopy (*HiRes*), to address all these challenges and provide comprehensive analysis of metabolic datasets. It combines standard spectral processing routines, data correction functions, techniques for reducing information complexity of multi-spectral dataset such as Principal Component Analysis (PCA) [1], and tools for PR, such as Non-Negative Matrix Factorization (NMF) [2], which extracts a set of spectral features with direct physical interpretation that can be used for identifying meaningful biochemical effects. These spectral features are also correlated with auxiliary information from the samples, such as biological end-points, drug toxicity or disease diagnosis.

Methods: *HiRes* is a software developed using C++ on Windows platform. Figure 1 below shows a screenshot of *HiRes* processing a series of Bruker spectra from urine samples. We use this figure to introduce the main functionalities from four aspects:
Data importation and organization: HiRes allows automatic data importation of NMR spectra acquired from different instruments such as Bruker and Varian spectrometers. Spectral parameters can be manually specified for any other data format. Series of spectra are imported in a single step, together with any additional information associated with the data, such as drug dose, measurement time, from a separate tab-delimited text file. All the spectra and associated information are organized in a project and listed in a table (window A) that can be easily sorted by each column.
Data display: Multiple spectra are grouped together and labeled for easy browsing (window B); each individual spectrum can also be viewed separately (window C). Spectra can be selectively displayed by specifying property filters, such as sample ID, drug dose, etc. Spectra are shown in either real, imaginary or magnitude mode with different unit types such as PPM, Hz, or Ms.
Spectral processing: HiRes provides various interactive data processing procedures (window C), which can be either applied to a single spectrum or all the spectra. These functions include: constant and linear phase correction (manually or automatically); temporal filtering (Lorentzian or Gaussian); baseline correction using convolution difference or linear fitting; PPM origin adjustment; zero filling; spectral normalization using number of acquisitions or single reference peak; preliminary peak alignment using maximum point and advanced alignment using PCA, either globally to the entire spectrum, or locally within the selected region.
Spectral pattern recognition: After the spectra are properly prepared using above preprocessing and correction procedures, different analyzing methods, such as PCA and NMF, can then be applied to any user-selected spectral region of interest. The results such as PCs and scores are displayed in an analysis window (window D). In the score plot, each point represents an individual spectrum; by clicking on any point, the user can easily go back to the associated spectrum for viewing, removal or further correction if needed. Each time a spectrum is modified, the analysis results will be automatically updated. This direct interaction between the preprocessing stage and the analyzing stage greatly facilitates the multi-iteration process.

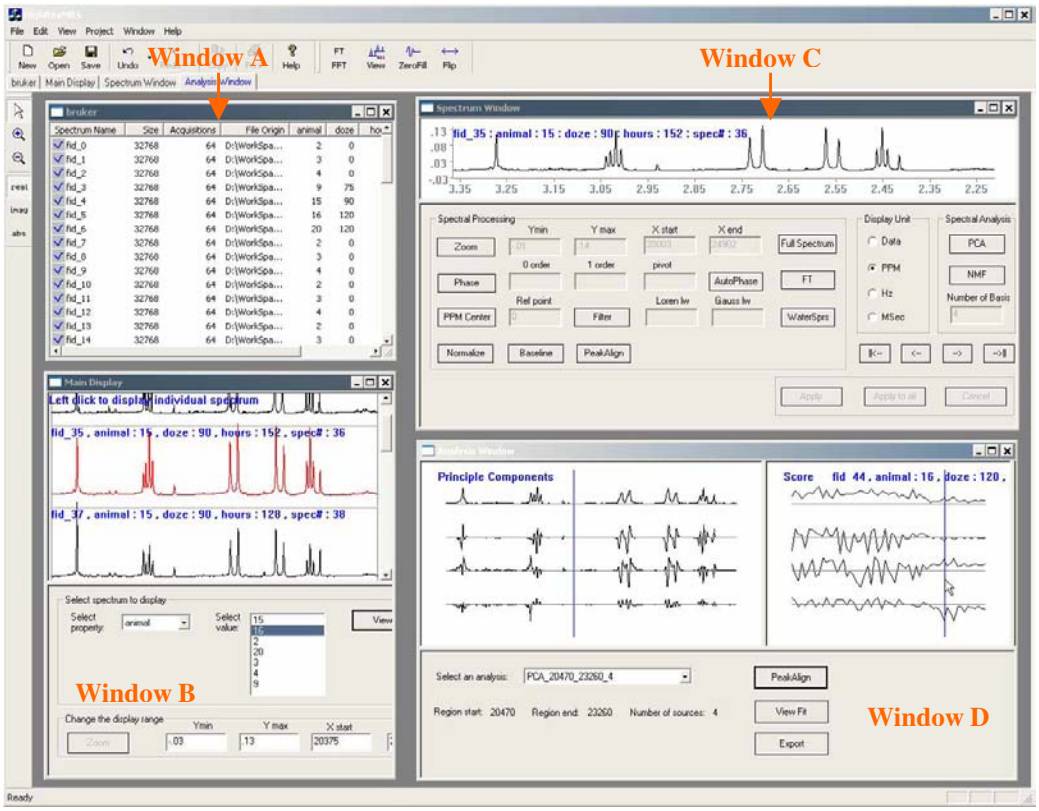

Figure 1 Screenshot of *HiRes*

Discussion: *HiRes* contains extensive multi-spectral processing functions, auxiliary information management, and sophisticated spectral pattern recognition algorithms. By integrating these components together and providing multi-directional interaction among them, it addresses the challenges for assessment and interpretation of large metabolomic data sets, greatly simplifying this otherwise lengthy and difficult process.

Acknowledgement: National Institutes of Health DK070301

Reference: [1] Stoyanova, R., Brown, T.R., NMR in Biomed. 14: 271-7 (2001). [2] Sajda, P., Du, S., Brown, T., Stoyanova, R., Shungu. D., Mao, X., Parra, L., IEEE Transactions on Medical Imaging 23(12): 1453 65 (2004).