Improved Procedure for Automatic Alignment of Strongly Overlapping Peak-Regions in High-resolution 1H NMR spectra

Q. Zhao[1], R. Stoyanova[2], N. Clarke[3], I. Pelczer[3], and T. R. Brown[1]

[1]Columbia University, New York, NY, United States, [2]Fox Chase Cancer Center, Philadelphia, PA, United States, [3]Princeton University, Princeton, NJ, United States

Introduction:

Experimental/instrumentally-induced variations in peak positions are often obstructive to the application of pattern recognition (PR) techniques in the high-resolution spectral data analysis. Previously, we have proposed Principal Component Analysis (PCA)-based routine, which aligns the spectra using a reference peak and phases the entire dataset uniformly [1], and further we refined the procedure for local adjustment of frequency shifts only in peak regions, where these shifts occur [2]. However, these approaches are based on the assumption that the peaks are reasonably separated. Here we present an automatic routine that successfully aligns peaks in strongly overlapping regions and in lack of a single reference peak. The developments are implemented in a software tool, *HiRes* [3], which integrates all the necessary data preprocessing and correction steps together with PR techniques. It is freely available for research purposes at http://hatch.cpmc.columbia.edu/highresmrs.html.

Methods:

35 spectra from horse blood plasma were collected using a Varian 600MHz spectrometer with relaxation-edited spin-lock and water-suppression acquisition sequence. The raw spectral data was processed in *HiRes*. A typical 600 MHz ^{1}H NMR spectrum from blood plasma ([0.5-4.5] ppm) is shown in Figure 1. In absence of a single reference peak the methyl resonance of the alanine (1.46ppm, a doublet, indicated by the arrow) is utilized for global alignment of the data. Alanine is a resonance usually with negligible chemical shift variance and little overlap with other peaks, therefore suitable for the alignment purpose. A new utility is added in *HiRes* to automatically detect a doublet structure and align the spectra globally accordingly. Subsequently the analysis is concentrated on the glucose spectral region, indicated by a box on the ^{1}H NMR spectrum. PCA is performed in this region and sub-regions of resonances are interactively selected based on the derivative shapes in the second Principal Component (PC). The spectra are then aligned locally within each region. At each step the PCs are automatically updated and the user can monitor the changes, occurring in the dataset.

Results:

The first four PCs of the 'glucose' region before any alignment are presented in the left-most box in Figure 1. A dramatic improvement after the alanine alignment can be seen in the PCs in the center box – the first PC now encompasses 96.4% of the total variance (increased from 77.4%). There are 6 distinct regions (indicated by the circles) in the second PC, entirely represented by derivative shapes, which indicate presence of frequency shifts in these regions. The right-most box shows the PCs after the local alignment within these regions. The percentage of the first PC is further improved to 98.8% and it is evident that the frequency shifts have been effectively removed, with the second and higher PCs now showing amplitude related variations.

Discussion/Conclusion:

The proposed approach substantially improves the data quality prior to the application of PR techniques by removing the undesirable experimental variations that often can mask subtle spectral changes. With *HiRes*, it is highly interactive and intuitive. This can significantly simplify the process of pattern discovery and biomarker identification.

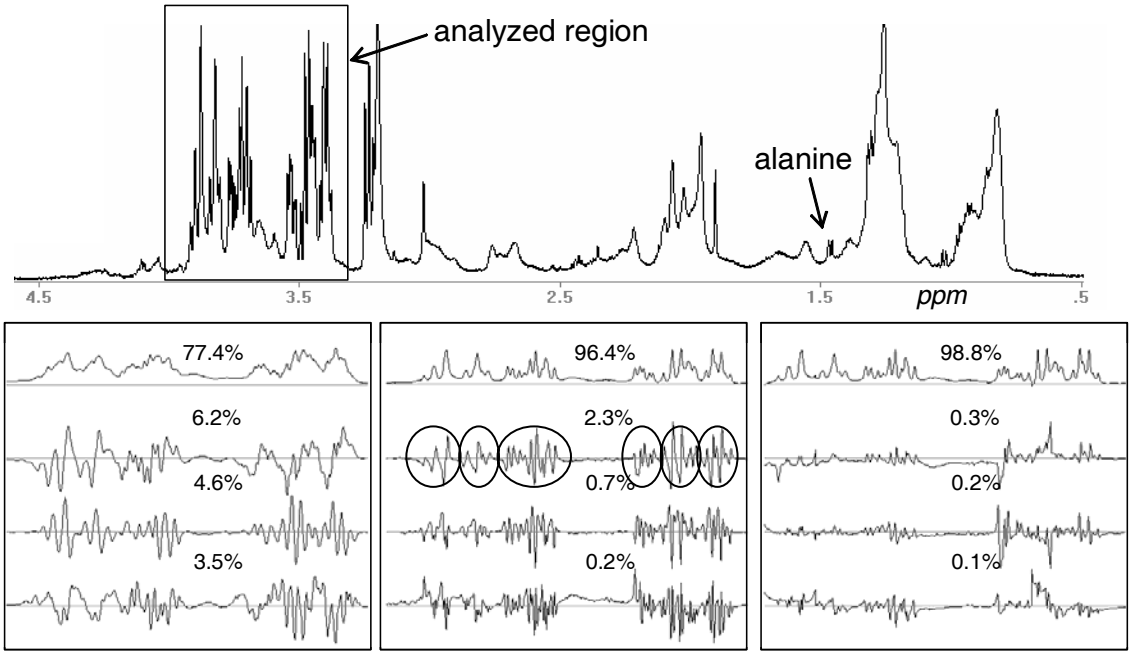

Figure 1. (left) Spectral region [0.5-4.5]ppm of a typical 600 MHz ^{1}H NMR spectrum from blood plasma. The alignment procedure is illustrated in the [3.35- 4.0]ppm region, predominately containing resonances of glucose. The first four PCs of this region, together with their corresponding eigen values are presented in the boxes below in the following order: before alignment, after global alignment based on the methyl resonance of the alanine, and after final local alignment of the individual peak-regions in the glucose region.

Acknowledgement: National Institutes of Health DK070301

Reference: [1] Stoyanova R., Brown T.R., J. of Mag. Res. 154, 163-175, 2002; [2] Stoyanova R., Nicholls A.W., et al, J. of Mag. Res., 170, 329-335, 2004; [3] Zhao Q., Stoyanova R. et al, Bioinformatics, 22, 2562-2564, 2006