Alignment of LC-MS Data Sets for Comparative Protein Expression Profiling Studies

Ian Reah, Ian Morns, Martin O'Gorman, Dave Miller, Wilka Hudson, David Bramwell;
Nonlinear Dynamics, Newcastle upon Tyne, United Kingdom

Introduction

The ability to combine data from multiple LC-MS experiments is required for comparative expression profiling studies. This enables the comparison of different experimental conditions through the use of biological and/or technical replicates, a prerequisite for any statistically valid study.

Figure 1: Statistically valid studies require the use of multiple replications per experimental condition.

In practice the retention time dimension, in particular, often shows shifts and possibly distortions when different runs are compared. The m/z dimension can also show smaller distortions. LC-MS runs need to be aligned to compensate for these forms of positional bias.

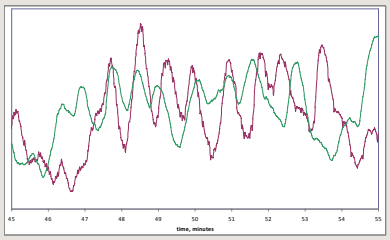


Figure 2: A section of a total ion chromatogram of two LC-MS runs overlaid to illustrate retention time positional bias

A solution to the alignment of LC-MS data is presented. The approach does not reduce the data to the total ion chromatogram, resulting in the ability to adjust for distortions in both retention time and m/z dimensions. It is based on a paired feature detection at the LC-MS level. This is followed by regression analysis in retention time and m/z to produce an alignment grid used to accurately overlay the data. See Figures 3 to 5 below for an example of the alignment results.

Discussion

In traditional LC-MS/MS analysis, information is discarded where data-dependent acquisition is not quick enough to probe every ion or where a confident identification is not produced. This can result in significant protein behaviour being overlooked.

By solving the alignment problem, a statistically robust analysis of LC-MS data can be performed to highlight areas of significant expression behaviour in the experiment.

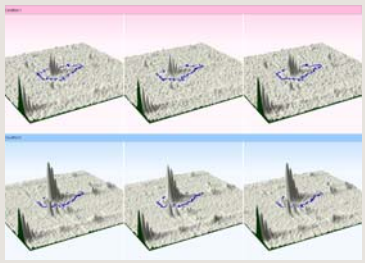


Figure 6: Statistically valid differential analysis of aligned LC-MS data

This can be coupled with targeted MS/MS analysis of only the differential peptides, ensuring that tandem mass spectra for all peptides of interest are acquired. Even if a confident identification does not follow, the experimenter is aware of the interesting behaviour and can investigate further.

Conclusions

The alignment of multiple LC-MS runs is necessary for statistical analysis of experimental data. The algorithm presented here is a robust solution that takes constant and irregular changes into account. It does not reduce the data to a total ion chromatogram or require prior peptide identification to align landmark peptides. The aligned data enables a statistically robust analysis to identify significant protein behaviour which may otherwise have been overlooked.

Example Alignment Results

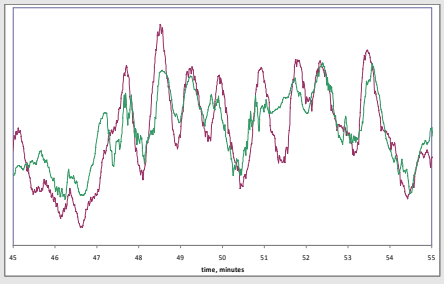
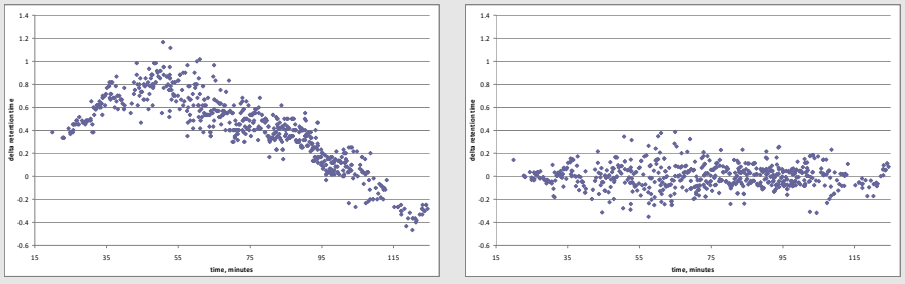


Figure 3: The total ion chromatogram from Figure 2 shown after the alignment of the two LC-MS runs. The alignment of the retention dimensions is now much improved.



a: Differences in retention time before alignment

b: Differences in retention time after alignment

Figure 4: The results of the alignment are displayed by plotting the difference in retention time for all paired features versus the retention time of the reference run. Each point represents a paired feature detected by the software. (a) Before alignment the data has a mean retention time difference of 0.412 minutes with a standard deviation of 0.297. (b) After alignment it has a mean retention time difference of 0.0003 minutes and a standard deviation of 0.101.

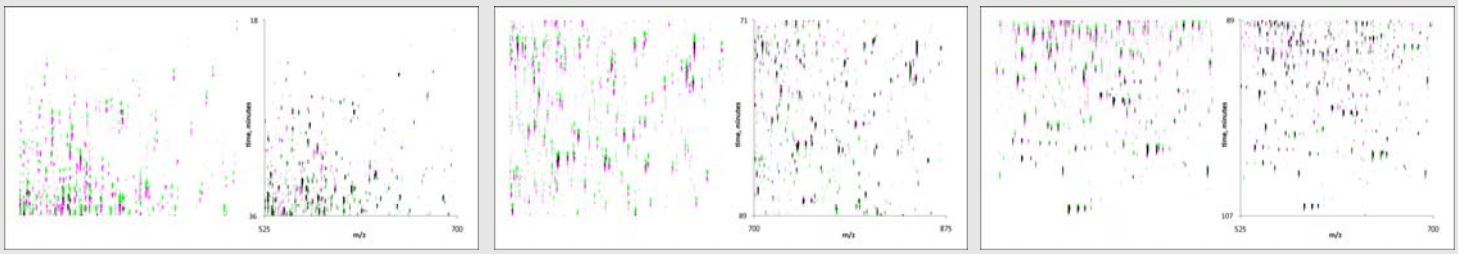


Figure 5: Different areas of the runs are shown overlaid to illustrate the alignment across the range of the data. The LC-MS data is represented in three dimensions – m/z in the horizontal dimension, retention time in the vertical dimension, and the height of the MS signal is represented by the intensity of the colour. The reference run is shown in magenta and the run being aligned in green. In each case, the unaligned data is on the left and the aligned data on the right. When the runs are aligned the colours mix to a grey tone.

Presented at: