Retention Time Correction Method Utilizing Unspecified Peaks in MS1 Scans

Philip M. Remeš, Ping Yip, Romain Huguet, Thermo Fisher Scientific, San Jose CA, USA

Abstract

Purpose: Align retention times without using explicit standards. This has several real-time instrument applications, including increasing sample throughput for targeted analysis and m/z recalibration.

Methods: MS1 spectra from a control experiment are stored and compared with spectra during subsequent experiments to perform time and m/z alignment.

Results: Excellent agreement between retention time shifts measured using standards and those estimated by our procedure

Introduction

The elution times of compounds in nano-LC chromatography are notoriously inconsistent, and drifts of 10's of seconds over several days are commonplace (Figure 1). This limits instrument throughput for targeted analysis, as targets must be scheduled with a time buffer to account for the drift. The common way to solve this problem is by spiking in standards and associating their real-time observation with the targets expected to elute at nearby times. A better way is to utilize the already present matrix compounds to align the retention times (Figure 2).

Figure 1. Retention time drift of a peptide over several days

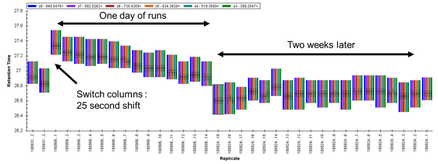
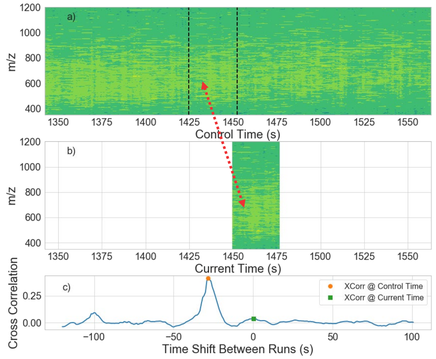


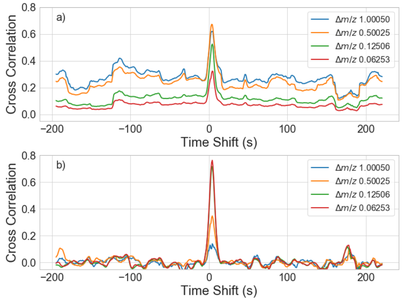
Figure 2. MS1 based retention time correction procedure using HeLa digest background. A control set of spectra (a) are compared with a set of spectra near the current time of elution (b). Cross-correlation of the data sets gives the time shift between experiments (c)



Cross Correlation of Spectra

Cross-correlation is an effective means of comparing sets of spectra, even when the m/z bins of the spectra are quite large (Figure 3a). The spectra can be additionally compressed with a wavelet transform, which not only improves computational throughput, but surprisingly also improves the quality of the cross-correlation. (Figure 3b).

Figure 3. a) Cross-correlations at various m/z bin size b) cross-correlations of wavelet compressed spectra at various m/z bins size



Generating Thresholds for Robust Estimates

For robust estimation of retention time shift, the cross-correlations are curated using data from the control data set. The widths and areas of the auto-correlation maxima are analyzed to produce thresholds (Figure 4).

These thresholds are used together to decide if a cross-correlation should be used for estimation. An exponential filter of good time shift estimates is maintained, and if a cross-correlation is bad, then the last good estimate is used instead of incorporating the new estimate into the filter (Figure 5).

Figure 4. Scatter plot of XCorr width and area with dashed lines indicating their respective thresholds

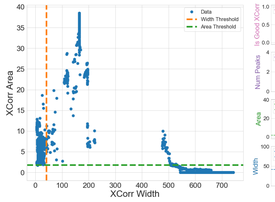
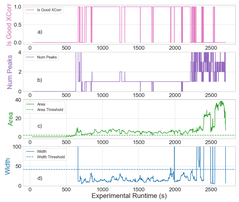


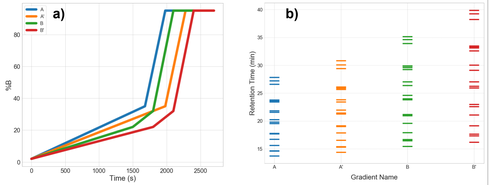
Figure 5. Cross-correlation metrics vs time that are used to decide if a time shift estimation can be made



Validation Experiments

Pierce™ Retention Time Calibration Mixture peptides as well as Promega™ 6x5 reference peptides were spiked into 200 ug of Pierce™ HeLa protein digest standard and analyzed unscheduled during 4x 40-45 minute gradients (Figure 6a). The resulting elution times (Figure 6b) of the standards offer a way to test the retention time estimation procedure.

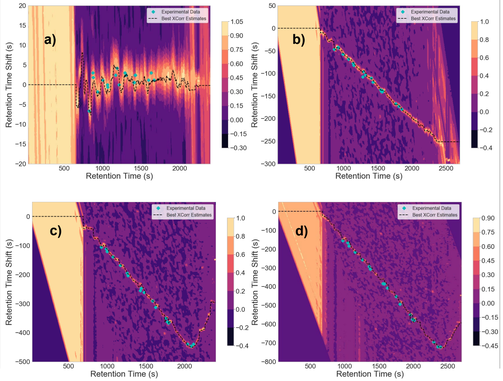
Figure 6. a) LC gradients used to analyze the standards b) Elution times of the standard peptides



Validation Results

The most common application of our method is to align replicate runs that use the same gradient (Figures 1, 7a), however, excellent results were obtained for time shifts of even 750 seconds (Figure 7d).

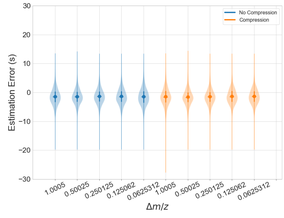
Figure 8. Cross-correlation contour plots of data from gradient A to other conditions, overlaid with time shift estimates from our procedure, and time shift measurements from the spiked standards. The other conditions were: a) replicate of gradient A b) gradient A' c) gradient B d) gradient B'.



Application to Different Sample Matrices

Our procedure was applied to 44 replicate analyses of 39 analytes in digested horse plasma using 40 minute LC gradients, to test whether a different matrix with fewer features than HeLa digest would pose a problem. The estimation errors for various m/z bin sizes and compression had 6 sigma widths smaller than the typical base LC peak width of 12 seconds (Figure 9).

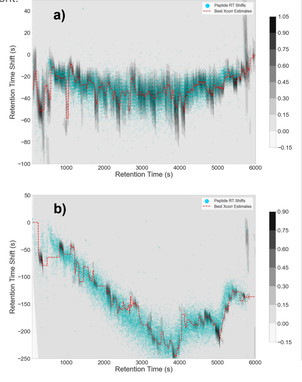
Figure 9. Distributions of retention time estimation errors for analytes in plasma background for various m/z bin sizes, with and without additional wavelet compression of the MS1 spectra.



Comparison of No-FAIMS to FAIMS

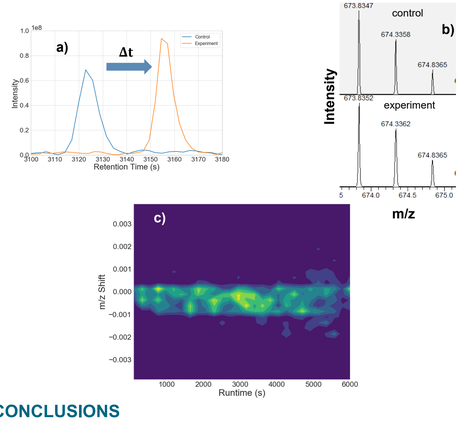
Ideally our procedure will be robust enough that a control set of data can be used for subsequent samples that have subtle differences. For example, the samples may come from different cell lines, or from various stages of diseased versus healthy states. To simulate these sorts of challenges, data from a 110 minute analysis of HeLa digest peptides were compared to a replicate run that utilized a Thermo Scientific™ FAIMS Pro™ Interface operated at -50 V and -65 V compensation voltage. The retention time shifts of common identified peptides were used as controls to demonstrate the ability of our procedure to find shared features for alignment, even when many spectra are quite different.

Figure 10. Cross correlation contour plots comparing a control replicate without FAIMS to replicates a) without FAIMS b) with FAIMS and -50 V compensation voltage. Blue dots represent shifts in retention time from common identified peptides, and the red dashed line is the time shift estimate from our procedure.



m/z Shift Estimation

An interesting application of accurately knowing the time shift between data sets is that spectra can also be compared across the m/z dimension. For example, the data from Figure 10a exhibit a time shift of ~30 seconds at ~3150 seconds (Figure 11a). Subsets of the corresponding spectra (Figure 11b) with $\Delta m/z = 2x10^{-4}$ were cross correlated to obtain the m/z shift between them. Similar analysis of all the MS1 spectra in the runs produces shifts of (-5 +/- 5)x10e-4 Th (Figure 11c). Such a procedure could be used to increase the time span required between instrument m/z recalibrations.



CONCLUSIONS

Robust time correction can be performed using only the background matrix compounds of a sample, without spiking in retention time standards. Future work will include:

- Porting the procedure to the embedded instrument computer
- Attempting to use 30-40 second targeted windows for the analysis of a large number of compounds for many days and LC columns

ACKNOWLEDGEMENTS

We would like to acknowledge Graeme McAlister for the great idea of comparing no-FAIMS to FAIMS data, and Alexander Makarov for helpful discussions about the m/z shift work.

TRADEMARKS/LICENSING

PO65505-EN0519S

ThermoFisher SCIENTIFIC