Insights
Genedata Expressionist
Issue 1

FEATURED ARTICLE

Dominik Mertens, Ph.D., Scientific Account Manager, Genedata AG, Basel, Switzerland

Data Analysis for the Characterization of Biopharmaceuticals: Retention Time Alignment

Huge improvements in mass spectrometry technology have driven advancements in the characterization of biopharmaceutical molecules. As characterization applications gain broader acceptance in biopharmaceutical R&D, organizations increasingly need to deploy solutions in GxP environments.

Biotherapeutics require extensive release testing to confirm that all critical quality attributes fall within acceptable tolerances. As such, there is ongoing debate about how to develop best practices and standard protocols to qualify and validate characterization processes. Ideally, methods developed during the R&D phases can be leveraged for release testing in production. Software platforms capable of supporting both R&D and the QA lab will have to process data from multiple instruments, provide automation as well as flexible reporting, and be ready to run in a regulatory-compliant environment.

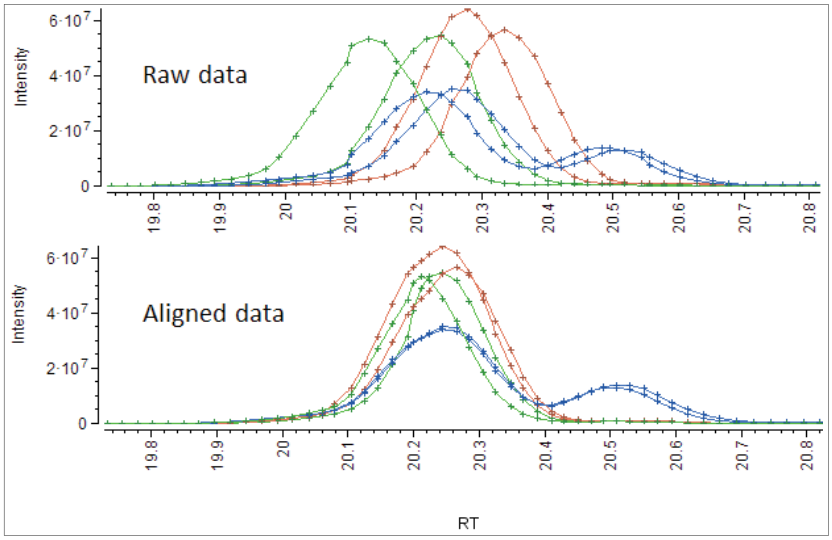

Figure 1: *Raw (above) and aligned data (below)*

In this *Insights* series we describe key data analysis processing steps required for the full automation of MS-based biotherapeutics characterization. In Volume 1 we investigate why retention time (RT) alignment plays an essential role in an automated data analysis pipeline for peptide mapping experiments and how the methods in Genedata Expressionist® for Mass Spectrometry can help to address this challenge.

UTILITY OF RT ALIGNMENTS AS PART OF THE MS PROCESSING WORKFLOW

RT alignment is a mandatory processing step for any LC-MS experiment to support robust quantification of biological molecules across chromatograms. It is well known that factors such as column age strongly affect RT deviations and that differences between columns can have a real impact on long-running projects. RT alignment correction enables the accurate detection and quantification of peaks – even small side peaks due to low abundant components (Figure 1). Therefore, RT alignment is crucial to ensure comparability between processing batches and ultimately for the reliable and reproducible automation of biotherapeutics characterization.

HOW IT WORKS — THE UNDERLYING ALGORITHM

Genedata Expressionist for Mass Spectrometry performs RT alignment using a Pairwise Alignment Based Tree. Unlike so-called Trivial Tree Alignments, Pairwise Alignment Based Tree results will not be affected by the order of the chromatograms and therefore provide consistency among different scientists' results.

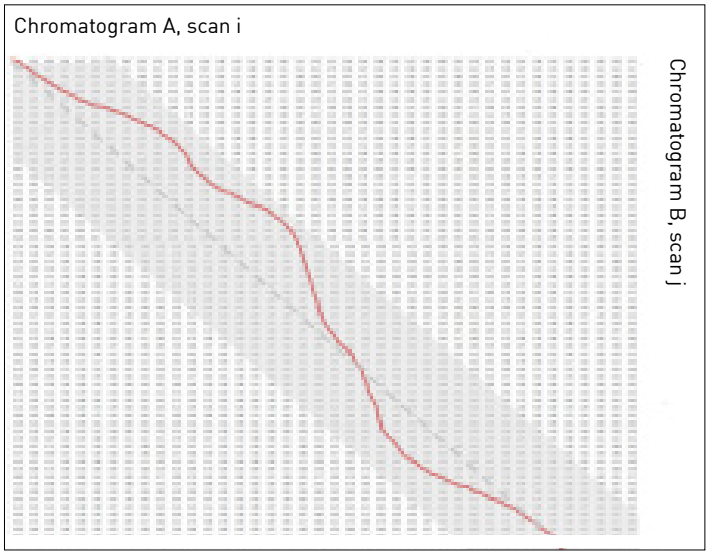

Figure 2: *Correlation matrix, the optimal path is shown in red*

The similarity of a pair of two chromatograms is measured by computing the correlation of all their spectra, resulting in a correlation matrix as illustrated in Figure 2. No reference peaks are required as all biological signals from the chromatograms are used to create an accurate alignment. The RT alignment algorithm detects the optimal path (red) close to the diagonal, which itself represents the path without RT alignment. The RT shifts will be measured by the distance between the diagonal and the optimal path. To avoid computing the correlation matrix for all possible spectrum pairs, an RT window is set to limit the maximum RT shifts. This approach of non-linear shifts is essential to perform a precise RT alignment of many chromatograms from samples under different conditions and with multiple replicates.

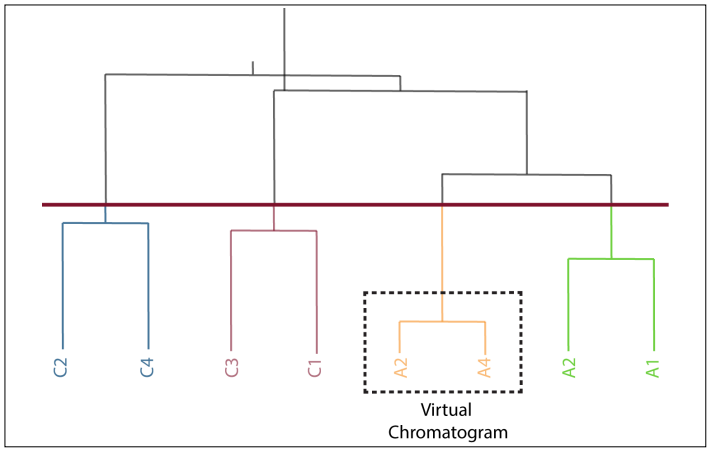

Figure 3: *RT alignment using a Pairwise Alignment Based Tree*

First, the algorithm computes a virtual chromatogram by applying RT shifts to the two most similar chromatograms (Figure 3). Second, it computes the similarity between the newly created virtual chromatogram and all remaining real chromatograms, and uses the best pair to create an additional virtual chromatogram. After iterative processing of these two steps, the final average chromatogram is computed, including all RT shifts of the input chromatograms.

EVALUATING THE PERFORMANCE

The RT alignment activity of Genedata Expressionist for Mass Spectrometry generates RT correction curves of the input chromatograms (Figure 4). This allows scientists to directly see the maximum RT shifts that have been applied to the input data. If the correction curves are close to the maximum RT search window, it is likely that the algorithm was not able to find the optimal path and the activity should be re-run with a larger search interval to identify the optimal RT alignment.

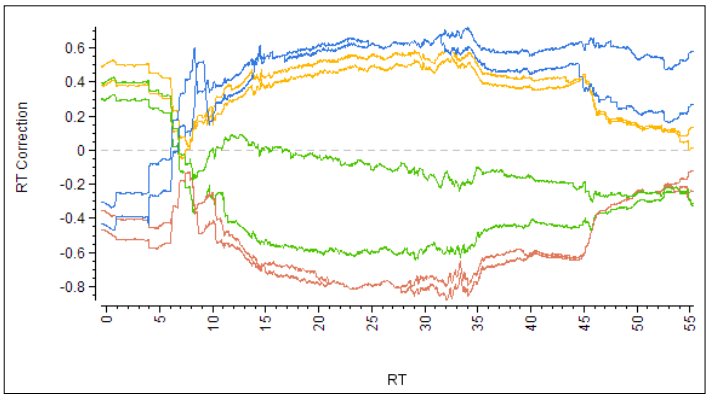

Figure 4: RT correction curves of the input chromatograms

SUMMARY

The RT alignment available in Genedata Expressionist for Mass Spectrometry is an effective, fast, and consistent method for making even large numbers of chromatograms comparable. Effective RT alignment is an essential prerequisite for reliable identification and quantification. Accurate alignment allows even small overlapping peaks from different samples to be detected and quantified precisely. Scientists can investigate the outcome of the alignment and will be supported by different visualization tools such as RT correction curves and cluster views to fully assess the source and extent of chromatogram shifts.