MsMetrix

MsXelerator Alignment Tutorial LC/MS: Reference Peak Warping

Introduction:

The MsCompare module of MsXelerator has a number of C**hromatographic Alignment** algorithms each capable of handling situations of different complexity. Alignment algorithms operating on multiple samples that are based on 1D-representations (TIC, BPC etc) are available in MsCompare. 2D-Algorithms are available in the MPeaks Module.

MsCompare Alignment Algorithms:

Dynamic Time Warping (DTW) and Correlation Optimized Warping (COW) are examples of algorithms often used to align 1-D chromatograms. In many cases, these algorithms produce good results, but depending on the complexity of the chromatograms they often fail to solve the alignment problem. Furthermore, COW and DTW use a large amount of memory and CPU time which makes alignment inherently slow, especially for a large number of samples. In general it can be concluded that chromatograms in which the shifts are large and irregular, as often encountered will be difficult to process using COW or DTW. This is also true for chromatograms that are dominated by a large number of peaks or intense background ions. In these cases TICs and BPCs are poor representations of the 2-dimensional data.

Reference Peak Warping:

Reference Peak Warping (RPW) as implemented in **MsCompare** is a method that combines two-dimensional high resolution peak picking with a fast 1-dimensional alignment procedure based on representative peaks that are evenly spread across the retention time axis. RPW consists of a number of steps:

1. In the first step the user selects a reference sample and so-called reference peaks that are nicely spread across the chromatogram. Reference Peaks can be selected manually in MsCompare or read from a so-called list file which contains the m/z values and optionally retention time values. Each of the samples should contain the reference peaks, if possible in more or less equal intensity. **The procedure will align a sample against the reference sample**. The retention times for the reference will therefore not change. Alignment will be repeated for all samples in the project. When selecting the reference chromatogram, try to select a sample in which the peaks are in the middle of the observed shifts!
2. In the second step the user can graphically determine the quality of the selected peaks for all samples. Selected Peaks that are neighbors should be further apart than the observed shifts and be representative of the complete chromatogram.
3. In the third step a spline function is applied to fit a non-linear curve between the observed retention times of reference and sample peaks. The above procedure is applied to all samples in the study. Processing times are typically less than 10 seconds.

   RPW is typically used as a pre-processing step before using more advanced data processing techniques. As only a representative part of all peaks is used for alignment, there will always be peaks that still show shifts after RPW alignment correction. However, in general these shifts will now be fairly small and are more easily handled by algorithms used for further statistical analysis.

In this document a fairly simple procedure will be explained, mainly graphically based. There is another tutorial on CE/MS Alignment using RPW that explains more advanced tools (MsX Demo Reference Peak Warping CEMS II.pdf)

Reference Peak Warping: Manual Selection of the Reference Sample and Reference Peaks:

Figure 1 shows a an example from a small proteomics dataset. Displayed at the bottom are the Total Ion Currents (TIC) for 15 samples in the range 20-80 minutes. The dataset has two groups; Controls and Patients. From the TIC's it is seen that shifts are a real problem when using any further data analysis procedure. We will need to align the data first. The top figure shows the Extracted Ion Chromatograms (EICs) of all samples for a peak having a mass of m/z 627. Plotted are the chromatograms at nominal resolution.

The shifts observed are typically in the order of 1.5 minutes. To measure the shifts, use the Peak Width Measurement Tool from the Icon bar. Measure the width from left to right across the EIC's.
The reference sample selected was sample 12, which seems to be in the middle compared to the other chromatograms. This was checked for a number of peaks.

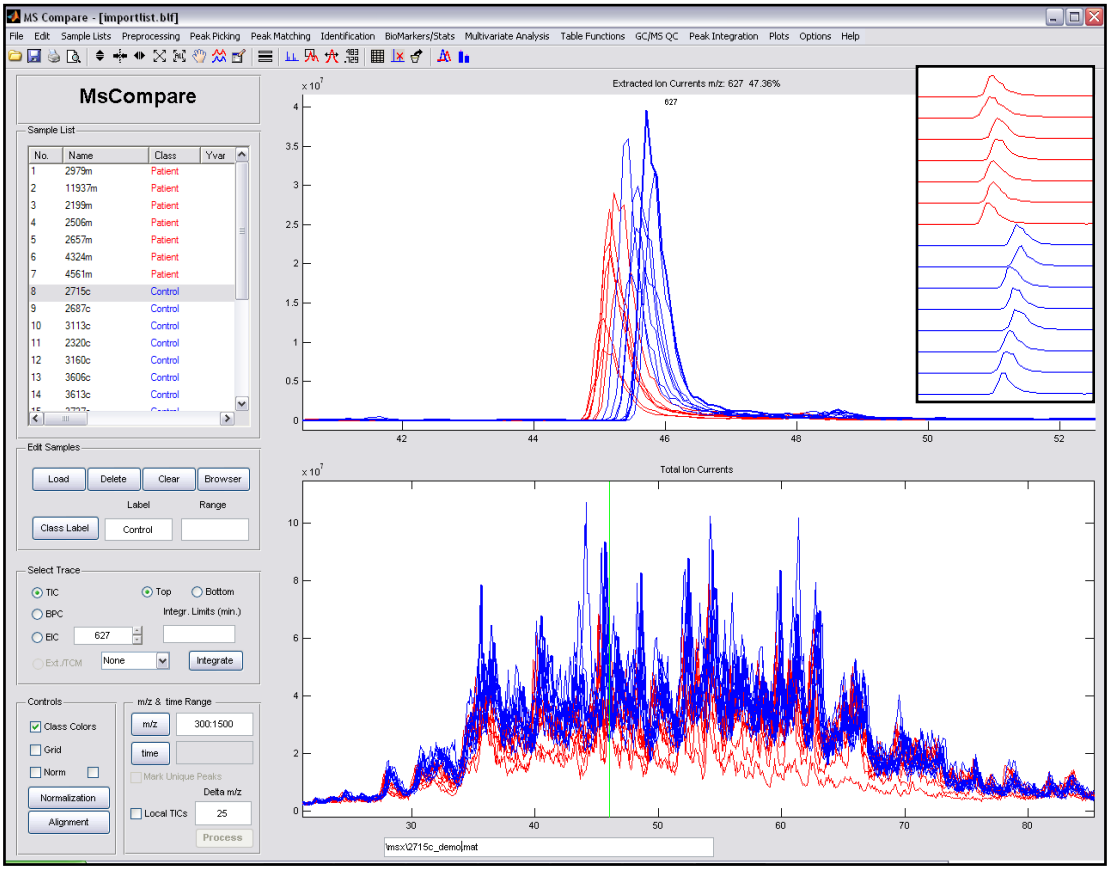


Figure 1: Total Ion Currents of 15 samples showing large shifts. TICs are plotted in overlay mode. The inset shows the EIC's in stacked mode plot. Now you can also see a pattern in the shifts.

Selecting Reference Peaks:

The task is now to select about 10 reference peaks across the chromatogram. Remember; selected peaks should show no interference from nearby peaks with the same mass. Start from left to right and click on a peak in the TIC window. The extracted ion current for the highest m/z value in this scan will be plotted in the top window. To decide whether the peak can be accepted for RPW, you will have to judge two things: 1) do all samples contain this peak and 2) is the peak free from nearby peaks with the same mass. It is not a problem that the EIC shows other peaks, as long as these are further away than the 1.5 minutes (the shift observed).

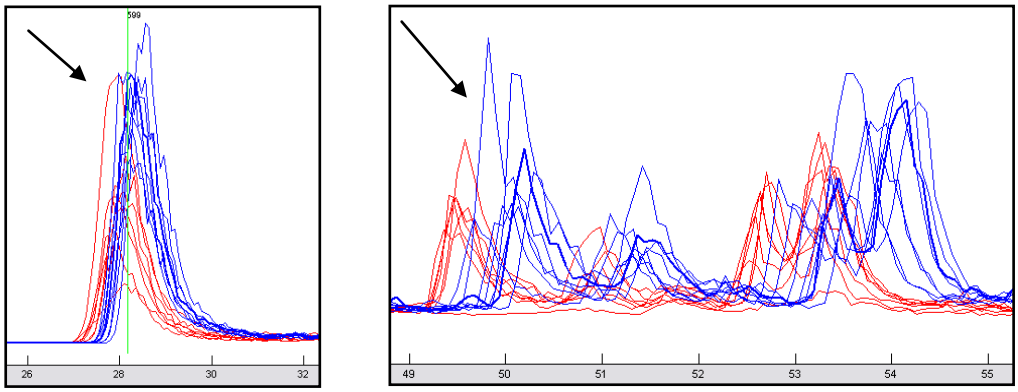
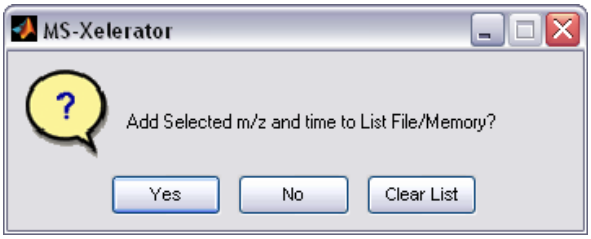


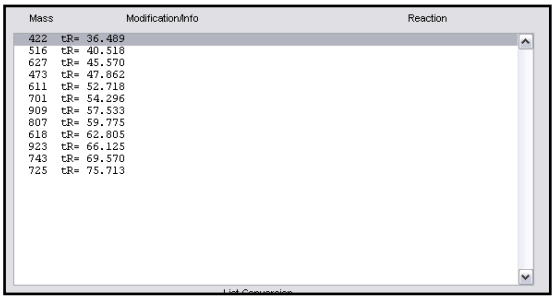
Figure 2: Selected Peaks: Left- high quality peak (present for all samples and free from interference), Right: low quality peak (peak shows interference from a peak to the right (2-3 minutes apart). Also, one of the peaks has low intensity compared to the other peaks.

Figure 2 shows two examples of peak selection. On the left a proper peak is selected. The right peak should not be selected because of the nearby interfering peaks. Furthermore, for one of the EIC's the peaks seems to be missing.

When the quality of the peak is OK, it must be stored to memory. Press the **RPW Store Button** on the Icon Tool Bar. The question below will be asked: Press **Yes** to add the Peak to the other selected Peaks. If you want to clear all peaks in memory, select Clear List.



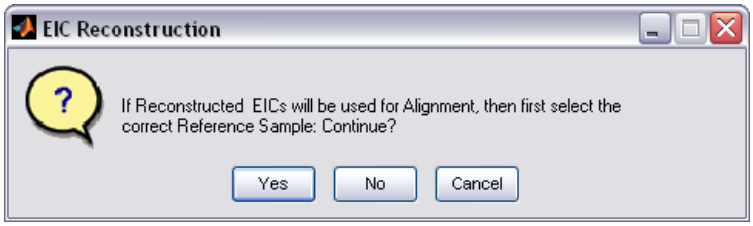
When you are ready with the selecting peaks across the chromatogram, you can view the selected peaks by opening the List file. From the Menu select: **Edit > Edit Peak Tracking List File**. The file will be opened. Part of the file is displayed in the figure below. The m/z values are shown together with the retention times for 12 selected peaks. Please save the file. When you are not happy with the selection, you can easily delete one or more of the reference peaks.



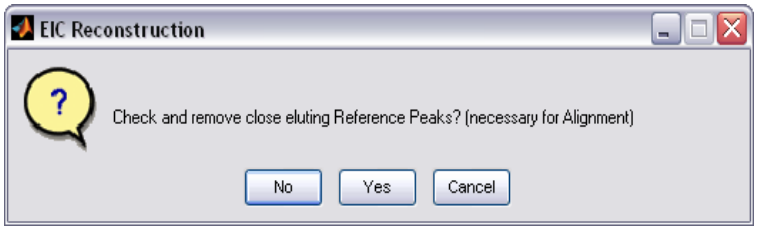
Attention: If you have real Reference Peaks spiked to your samples, you can directly create a list file and enter the m/z values for the known reference peaks and their expected retention time positions.

Reconstruction of Chromatograms Based on Reference Peaks:

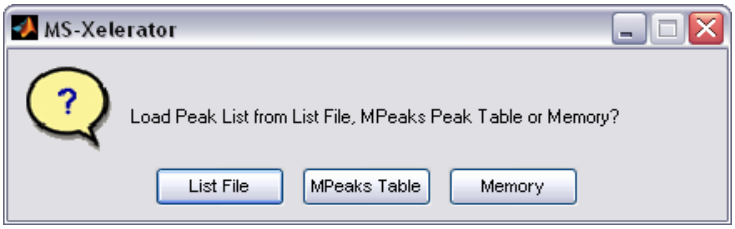
Next, we will start the RPW Alignment by selecting: **Menu > PreProcessing > Alignment: Ref. Peak Warping**. Before staring, be sure that your reference sample (in this case sample 12) is the active sample in the Sample Listbox. You will get a number of questions, starting with the message below.



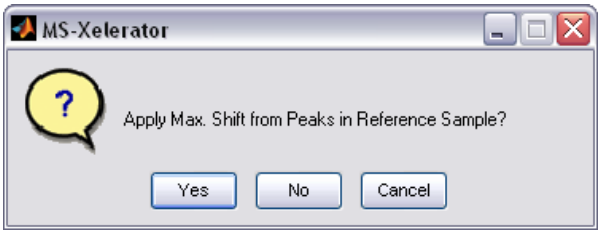
The next question relates to how far apart the selected peaks are from each other. Actually this question is from the Advanced Alignment Procedure using an MPeaks Table as input. So, skip it by **selecting No.**



Next, you will be asked if the reference peaks should be read from a list file (if saved), from an MPeaks Peak table (Advanced Alignment Procedure) or from memory. **Select Memory.**



After this, you will get a question to apply a Maximum Shift. The Alignment procedure will match reference peaks in the Reference Sample with peaks in all other samples. The detection of peaks can be restricted to a maximum time window. Reference Peaks from the others samples should not have shifted more than the value to be specified. This option is especially important if the EIC of the selected m/z values has more than one peak. You don't want the wrong peak to be detected. Detection in the samples will take place in the specified window and the largest peaks found in this window will be used.



Select **Yes,** and specify a value **of 1.5 minutes**. The **Reference Traces will be reconstructed** and displayed. You will be asked if the peaks should be labeled by the m/z value or a character. **Select m/z value**. Next question is, if the peaks

should be connected between the samples by lines. Select **Yes**, this will give you the opportunity to more clearly see the shifts. The next question on replacement of the TIC's can be answered by **No**.

The reconstructed Traces based on the Reference Peaks will be shown in stacked mode, with the peaks labeled and connected. You can clearly see how shifts change from sample to sample. Is seems that a jump takes place when going from the last sample in group 1 to the first sample in group 2. Might be that the sample run order was not random between groups. From a statistical point of view, this is not a good procedure. As it concerns the Alignment, Figure 3 below looks OK. Peak Picking was done properly and no strange features are present. The peaks used as reference peaks are at least 2.5 minutes apart and the spread across the chromatogram is good.

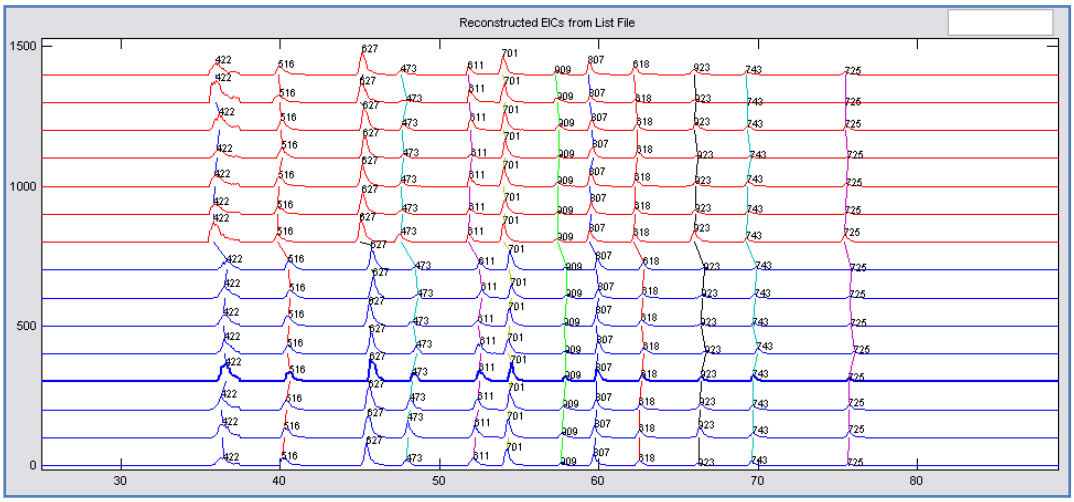
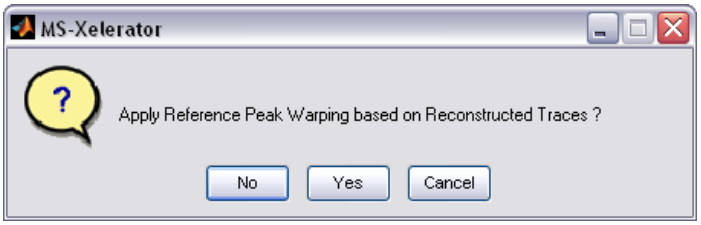


Figure 3: Reconstructed TIC's based on the reference peaks, showing the shifts between the samples

Next question is if you want to apply Reference Peak Warping on the Reconstructed Traces shown in Figure 3. If the Peak Picking is OK, and no other problems are visible, then select **Yes**. If you observe that wrong peaks are selected, or peaks are missed or not present in all samples, then select Cancel and change the problematic reference peaks. You might need to delete some. This is best done from the saved List File.



The result of RPW alignment will be shown in a separate window, and you will get a question if you want to save the RPW results. Select **Yes.** The RPW results are shown in Figure 4 below. The alignment for the reference peaks is perfect. The assumption is that all other peaks not used as reference will now be much better aligned. This is not a 100% guarantee; you should check other peaks after the alignment as well. However, RPW should have solved most of the problems. The small shifts that are still present can be handled by the Peak Picking or Peak Matching algorithms in MsCompare.

In Figure 4, you can zoom in on a small retention time range, and press the space bar to plot the aligned peaks in the same time window. You can save the figure, for reporting purposes.

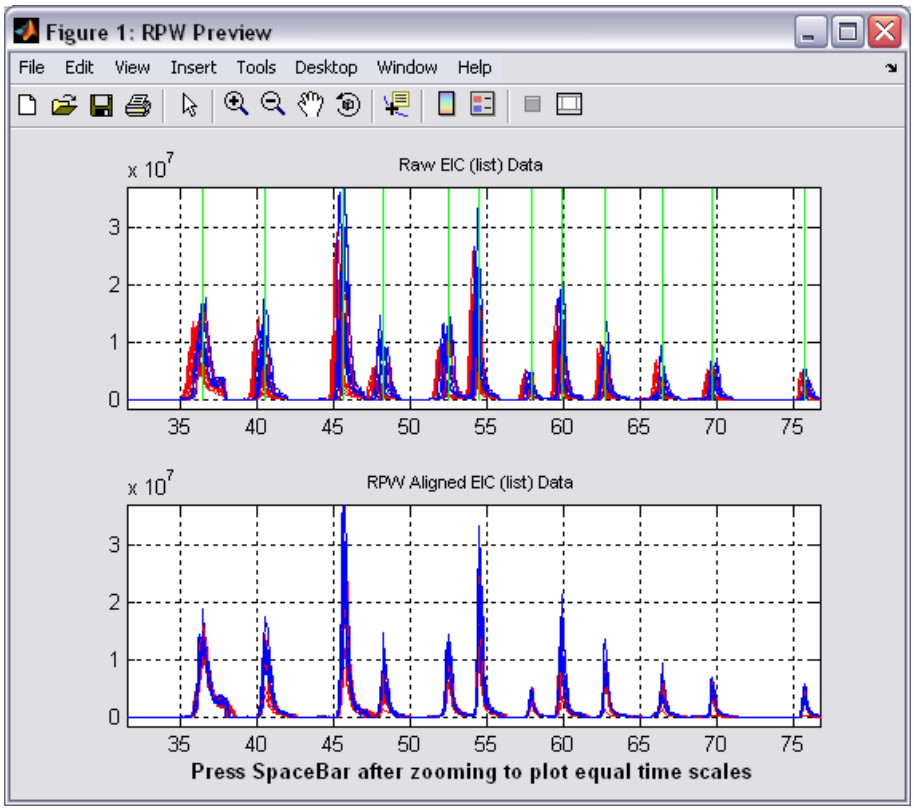


Figure 4: Reference Peak Warping overview: Top - Raw Data of Reconstructed TIC's based on reference peaks. Bottom, aligned peaks. Peak Positions of reference peaks are marked in the top window.

After saving the results, renew the TIC plot in MsCompare by selecting the TIC radio button. Check other peaks not used for alignment. The new retention time scales for all samples are saved to disk. The next time you start MsCompare with the same project, you will be asked if you want to use the RPW aligned data or the original data. In cases that you saved the results, but you are still not happy, you should clear all data from MsCompare, re-load the original data and start over with the Alignment procedure. Alignment can be based on groups of samples, as long as the reference sample is still the same. Sometimes it is difficult to work with many samples, especially if the alignment is complex. Then try to use smaller groups of samples. However, be sure that the reference sample is always present in each group.

For Advanced Reference Peak Warping Alignment see the Tutorial Document II using CE/MS Data. Compared to the other Alignment methods available in MsCompare: e.g. **COW**, Reference Peaks warping is the preferred method of alignment as it is fast and easy to use. COW is much slower, will use the full retention time region and sometimes has difficulties aligning problematic samples. Very complex samples, having thousands of peaks can be handled by RPW, because a small selection of representative peaks is used.

One important assumption of RPW is that the elution order of the reference peaks does not change from sample to sample. Most of the time, the order will be the same. Otherwise, select a different reference peak.

Sometimes, a sample does not have a reference peak. Then, select another peak, which is probably much smaller, but is present in all samples. Be sure the peak is selective. In the next version of RPW a method will be introduced that can handle missing peaks as well.

In case of very difficult cases, please contact MsMetrix for assistance.