MS Xelerator

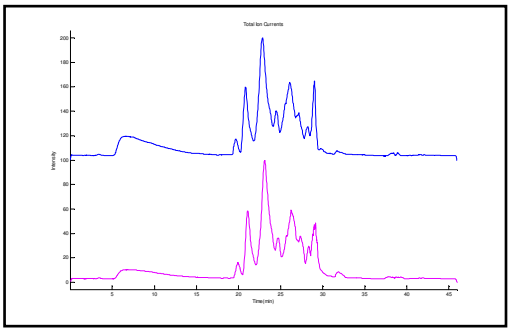
Differential Analysis: Comparing Two Samples

MPeaks Differential Analysis can be used to find differences between two samples. It is a very powerful technique which can be applied to a large number of LC/MS application areas.

Differential Analysis can be performed in a number of ways using MS Xelerator. From the main **Viewer**, one could plot all mass chromatograms in matrix mode and overlay these with mass chromatograms from a reference sample (sort on peak height first). If plotted in negative mode the differences will be much easier to detect. The disadvantage using this method is that you will have to explore a large number of mass chromatograms and very small differences could be missed. The second approach would be to detect all peaks using MPeaks. After this, the MPeaks Viewer could be used to check whether the detected peaks are also present in the reference sample. Both methods are visually based and it will take some time to finish the job. It is not the preferred method for complex samples.

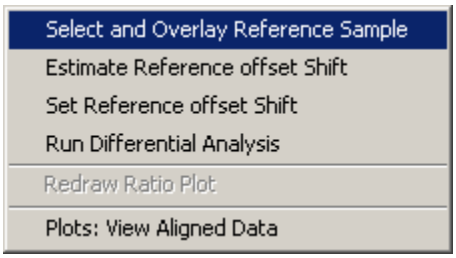
The MPeaks module contains two dedicated algorithms for **Differential Analysis**. Both are extremely fast (10-20 seconds) and offer the possibility to align and shift the data during comparison. The first and default algorithm is the one most powerful. The second algorithm is less restrictive but does a good job for data sets having not too many peaks. To be able to run Differential Analysis the MPeaks table must have been calculated or loaded from disk. Differential Analysis can be started using **Menu > Differential Analysis**. The menu contains a number of items.

The data set used in this example relates to a proteomics application and the task was to compare and find peaks in Sample A that are absent or small in Sample B. The MPeaks result table contains 982 peaks.



Total Ion Current for Sample A and Sample B

Step 1: select and overlay the reference sample. Browse to the location of your reference file, select the sample and press OK.



Before running Differential Analysis you should check to see whether the data sets are well aligned. If the two sets are shifted you will have to apply alignment correction methods. Sort the table on Peak Height and explore a number of mass chromatograms to check for alignment problems.
Figure 1 displays the overlay of one of the bigger peaks from Sample A and the same mass chromatogram from Sample B. The red trace will always be the reference sample. From Figure 1 it can be seen that the reference sample is shifted to the right almost 0.5 minutes. A shift correction will be needed. Check a few

other peaks and see if the same kind of shift occurs. This will not always be the case. Some peaks will shift to opposite directions, or will not be shifted at all.

The goal of exploring the two data sets is to get a good first guess of how much the peaks are shifted.

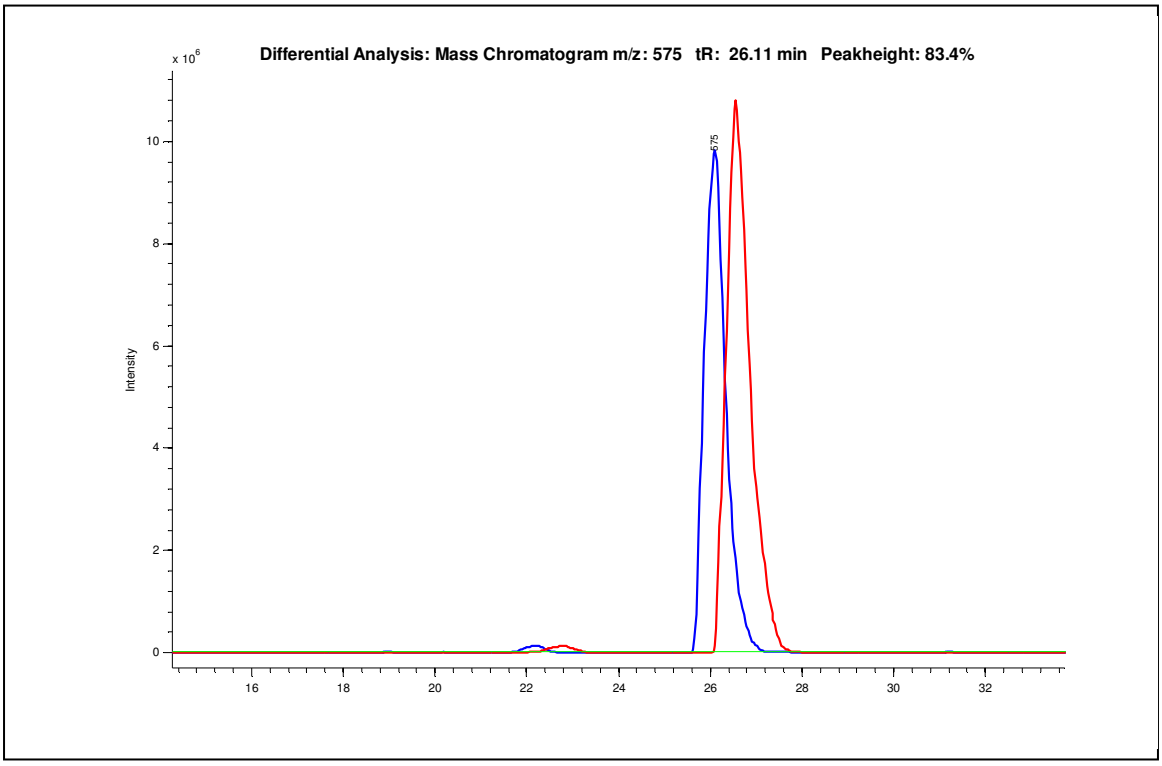
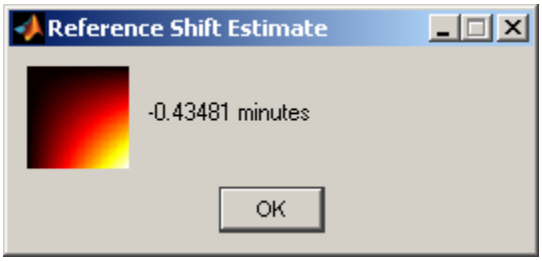


Figure 1: Differential Analysis - overlaying traces to estimate offset time shift between both samples.

The Differential Analysis Tool offers a utility to measure the shift between both peaks. From the menu select: **Differential Analysis > Estimate Reference offset Shift.** The difference between both peak maxima will be returned as a first guess (Offset Shift). The full reference mass chromatogram will be shifted using this offset in a linear way.
The returned offset shift is -0.43581 min., which means that the reference should be shifted to the left to properly align both traces. We will use this value as a best guess.



Select from the menu **Set Reference Offset Shift** and enter the above value in the edit box. Press OK and the view will be updated with the reference trace shifted to the left.

In situations where data sets are well aligned you may skip the above steps. You are now ready to run Differential Analysis. From the menu select: **Differential Analysis > Run Differential Analysis.**

The reference sample was already loaded and is displayed in the reference edit box. To select another sample use the **Browse** button.

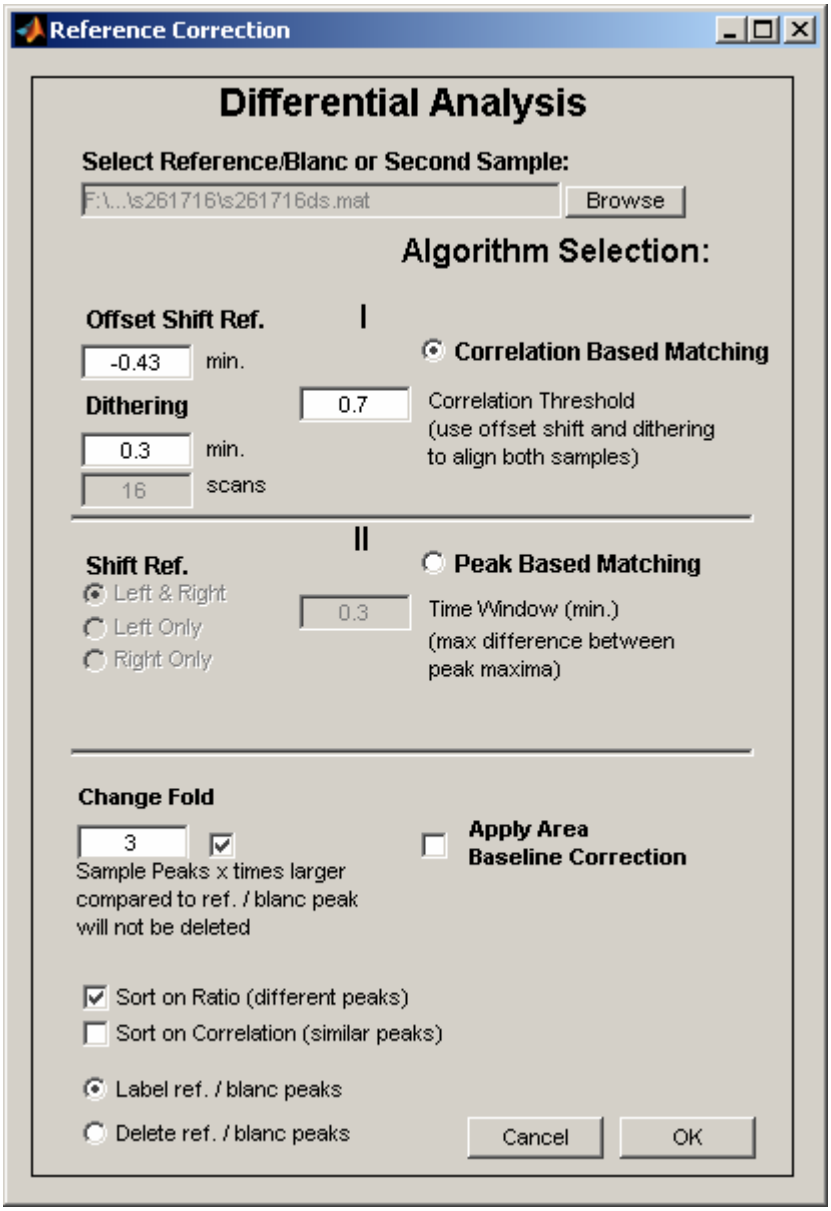


Figure 2: Differential Analysis Window.

Correlation Based Matching:

Differential Analysis uses two different algorithms. The first algorithm is based on correlation and is named Correlation Based Matching. Each peak in the sample will be correlated using the same local region to the reference sample. If the result of cross correlation exceeds a certain **correlation threshold** value (0.7) it will be indicative for the presence of the same peak in the reference sample. If the correlation is small, probably no reference peak exists at this location. Using this algorithm, mass chromatograms should be properly aligned.

Differential Analysis is performed using a dedicated alignment algorithm. Besides an offset shift correction, the algorithm uses a second parameter to allow for locally shifted peaks. A dithering parameter will set the window used for local cross correlation. A value half the offset shift will be a good starting estimate. An optimal alignment will be calculated for each peak from the MPeaks table. After alignment the ratio between sample and reference is calculated. A change fold value will be used to set a threshold for

differentiating peaks. If wanted, peaks having area ratios larger than 3 will not be deleted from the list. All peaks having high correlations and ratios smaller than 3 are present in both samples and not unique. You might want to delete these from the MPeaks table.

Peak Based Matching:

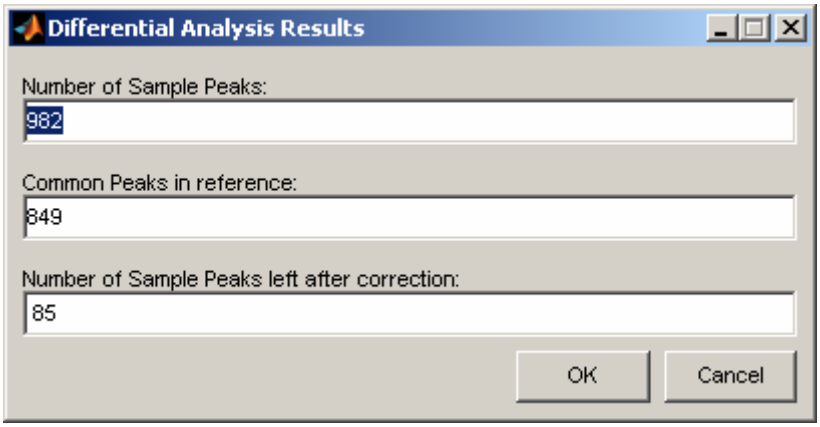
The second algorithm will calculate a full MPeaks table for the reference sample after which a table comparison will be performed. Each peak having the same m/z value is compared to the reference sample. If a reference peak is found, it should elute in a specified window around the first peaks retention time. This parameter is set in the **time window edit box.** The algorithm uses a default value of 0.30 minutes.
Any peak may be shifted in the reference sample set with a maximum of 0.30 minutes. You may restrict the shifting to one direction only (left or right) or use both directions. Using the default value, reference peaks can be shifted both to the left and right.

Change Fold:

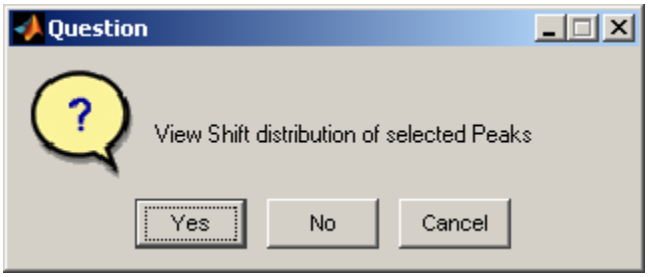
As explained, a Change Fold level may be used. Peaks having ratios larger than this value will not be deleted from the table (similar peaks).

After the calculations have been completed the MPeaks table returns two new column; the correlation value and the ratio between both peak area's. The table can be sorted on Ratio (differences) or Correlation (similarities). Optionally, a baseline correction can be applied if needed. If you only want to keep the "differential" peaks, you should select the **delete peaks radio button**.

The Differential Analysis Results window will give an overview of the number of peaks in the sample, the number of common peaks in the reference sample (peaks in reference with a correlation larger than the threshold) and the number of peaks that will remain if the delete option was selected.



Press **OK**, to continue. You will be asked if summary results on Shift Correction should be plotted.



Press Yes, to view the Shift Results. Figure 3 displays the shift results (offset + dither) in scans for a number of selected peaks. From this Figure it becomes clear that a certain pattern is present. From 20 to 23

minutes the reference peaks are shifted to the right and the shifts increase until about 23 minutes. Beyond this point the shifting decreases, but is still quite large.
The applied offset shift (-0.43 minutes) equals 23 scans. It can be seen that dithering uses at the most 10 additional scans. The dither value of 0.3 minutes (16 scans) seems to be sufficient.

The bottom plot show a histogram of the shift distribution. An average shift of -27 scans is observed between peaks from both data sets.

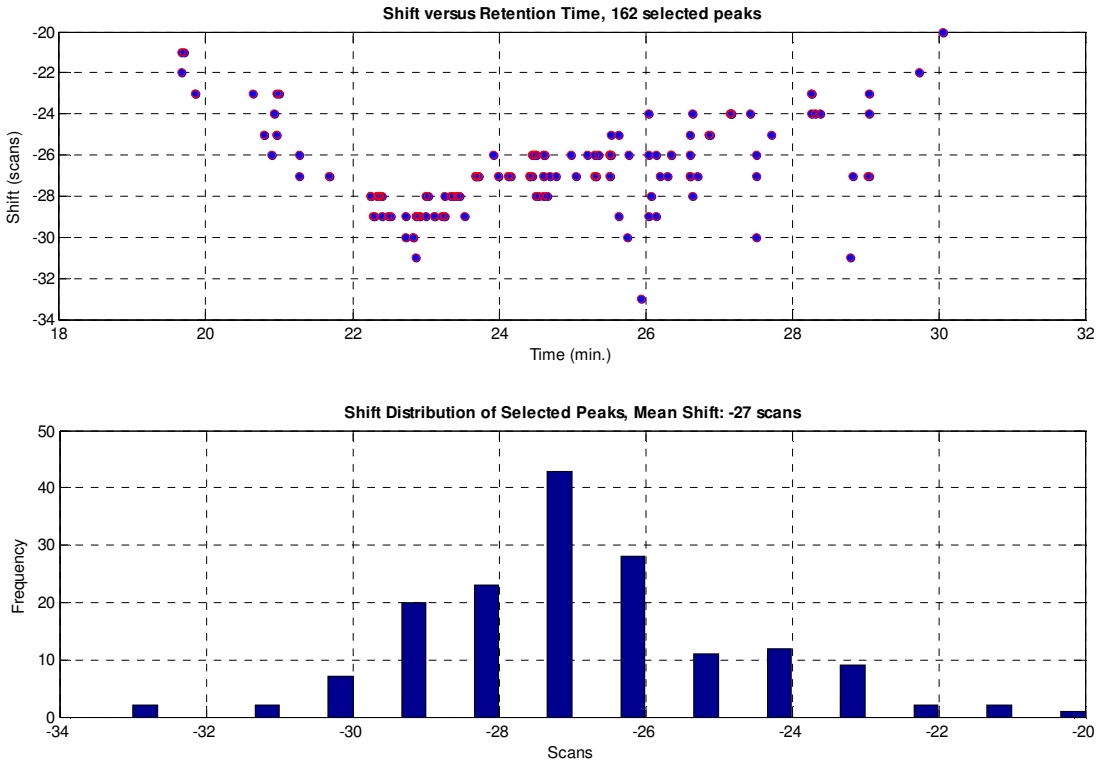
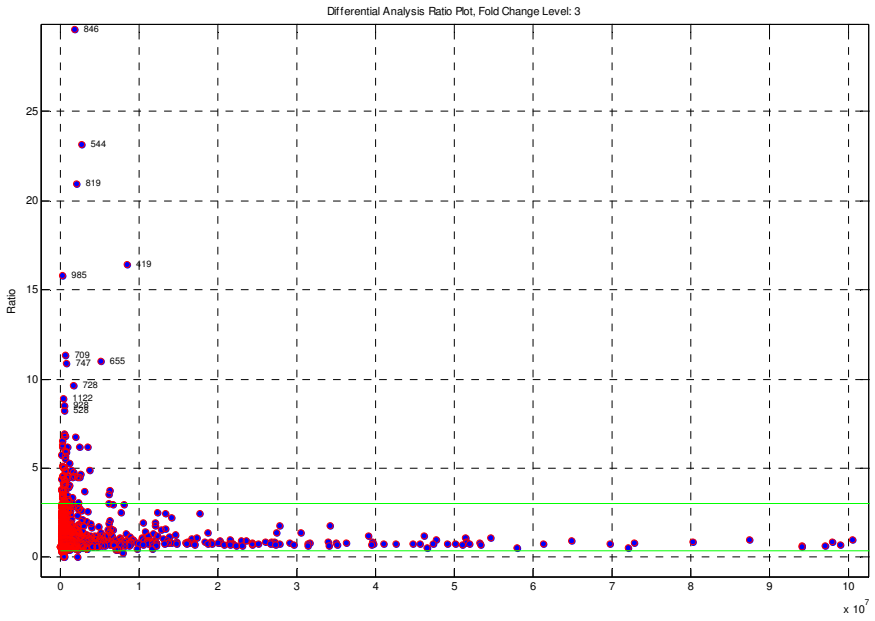


Figure 3: Summary Result Plots for Shift Correction.

You will be returned to MPeaks and a **Ratio-Area Differential Plot** will be the shown, which is basically a scatter plot between peak area (x-axis) and the peak area ratios between sample and reference peak (y-axis). Peaks may be labeled using the mouse. The peak will be highlighted in the table. To label a group use the mouse and hold the shift key. To enlarge the plot press the "f" key (Full Screen Plot). From Figure 3 we can conclude that 31 peaks in the sample have ratios between 5 and 30. The largest peak from this selection has a peak area of 2.5%, all others are much smaller.

In all cases a ratio will be calculated even if no peak in the reference sample was found (small correlations). However, sometimes the "reference area" will be zero or negative. The ratio value will then be set to a value of 99. To find relevant peaks study both ratio and area values. Very small peaks, absent in the reference sample will probably not be very relevant, this may be quite dependent on the type of application.

Figure 5 shows the top peak in overlay with the reference peak. The ratio is 29.5 and the area is 0.85%. The correlation is very bad (-0.036), which means that no reference peak could be matched. The dithering cross



correlation returns an optimal shift value of -3 scans. For peaks having low correlations the dither value has no real meaning. Only in case of matching peaks in the reference sample an optimal shift value can be calculated with real meaning.

Figure 4: Area – Ratio Plot showing Differential Analysis results.

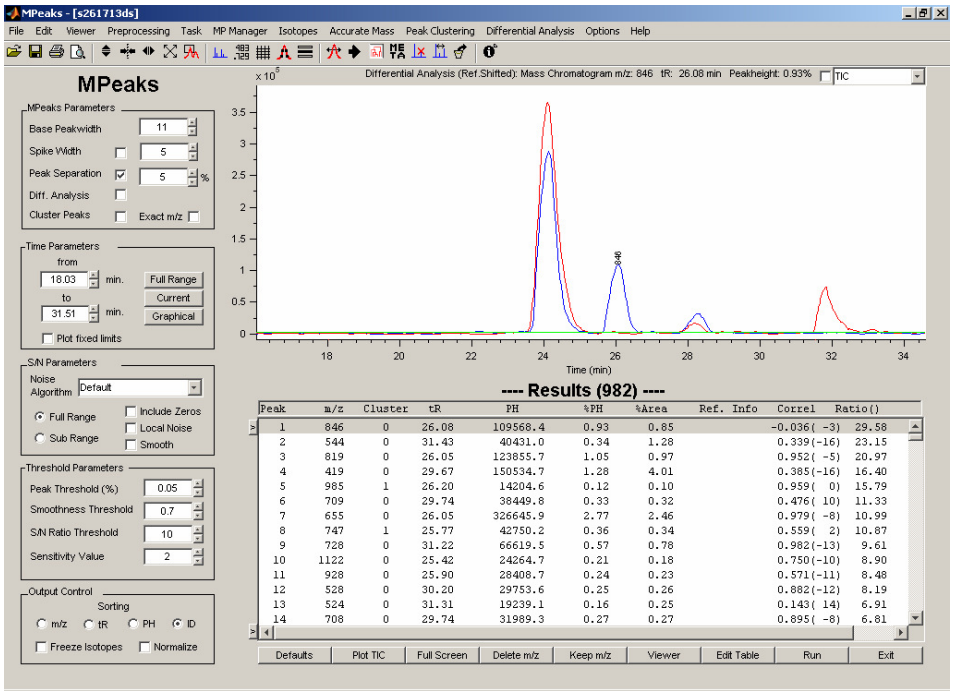


Figure 5: MPeaks Overview after Differential Analysis.

To plot the results, switch to the Viewer and explore the peaks in negative overlay mode (**Figure 6**). Results are shown for the first six peaks having the largest ratios (raw data, not aligned). The peak heights vary from 0.12% to 1.05% compared to the largest peak in Sample A.

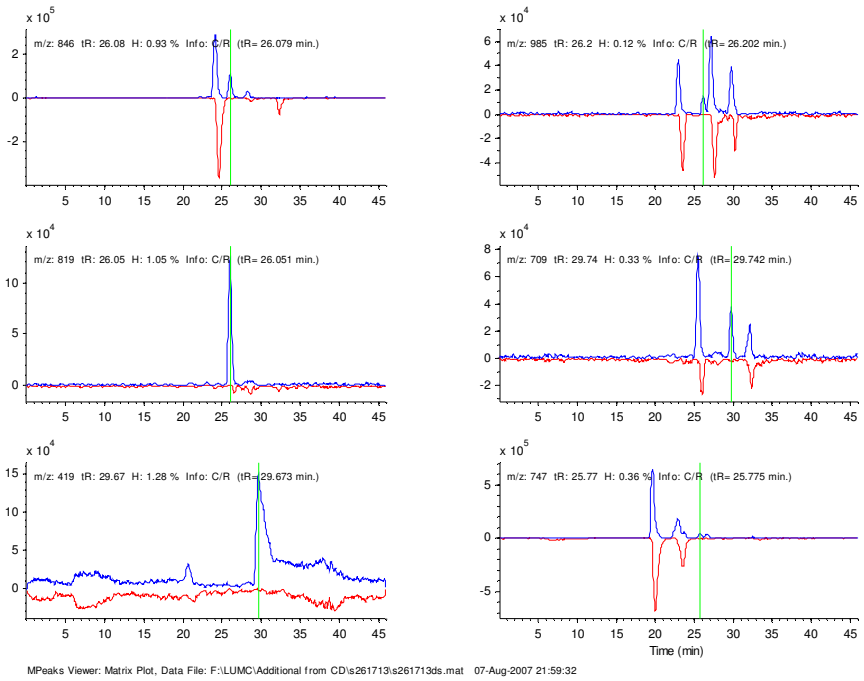


Figure 6: Differential Analysis Results

MsMetrix