MsXelerator Differential Analysis:
Finding Differences between Sample and Control

MsMetrix

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 in which it is necessary to compare two samples: Sample/Control, Treated/non-Treated, new Batch/old Batch etc.

MPeaks can also be used to compare peak picking result from a sample to a large number of controls or reference samples. This application is described in another document. Also the comparison of series of samples belonging to different classes or groups is part of another tutorial. Contact MsMetrix for additional documentation.

Comparing and finding differences between two samples can be done in a number of ways, depending on the complexity of the data. MPeaks Differential Analysis can be performed on nominal and accurate Extracted Ion Currents (EICs). The number of peaks in a sample often dictates which method to use. Another point of concern is how well sample and control are aligned along the retention time axis. Several algorithms can be applied to correct for retention time differences between sample and control.
In principle, the Viewer and MsCompare can also be used to find differences between 2 samples. However, these procedures are graphically based and take a long time to complete if many (small) peaks are present in your sample. All Differential Analysis procedures in this document are based on peak picking results from MPeaks on your sample. Once all the chromatographic peaks of interest have been detected we can decide how to proceed.

The MPeaks module contains four dedicated algorithms for **Differential Analysis**. The nominal EIC algorithms and the Table based comparisons are extremely fast (10-20 seconds) and offer the possibility to align and shift the data during comparison. Below a short description of all 4 methods is given:

1. **Correlation Based Matching:** Starting point is a peak picking result table belonging to the sample. Based on correlation statistics, this algorithm determines if peaks from the sample are present in the control. Ratios will be calculated for each entry in the MPeaks result table. The Correlation Algorithm can be combined with a powerful **local alignment** procedure that optimizes the shift correction for **each individually detected peak**. It currently runs in nominal mode only. Alignment correction can also be applied by loading COW (Correlation Optimized Warping) or RPW (Reference Peak Warping) corrected time scales. COW is an alignment procedure based on the Total Ion Current (TIC) or Base Peak Chromatogram (BPC). Using RPW you define a number of reference peaks that should be present in both the sample and control. RPW can be performed in MsCompare. The new time scales can be loaded into MPeaks.
2. **Accurate Peak Matching:** Based on an accurate MPeaks result table (accuracy to be set by the user), this algorithm extracts the EICs of sample and control in accurate mass mode. It then calculates the ratio based on accurate peak height or area. The user is able to define a retention time window to allow for shifts in the control. This algorithm can also be run on COW or RPW aligned sample and control. The COW alignment between sample and control is done as a pre-processing step using the MsCompare module or can be accessed from the Differential Analysis Menu in MPeaks. RPW can only be performed in MsCompare. Accurate Peak Matching can output ratios based on Peak Height and Area. When using peak height only, the algorithm can check for both up- and down regulated peaks. For finding down-regulated peaks, the analysis is reversed. Down-regulated peaks based on peak picking from the control sample that are not present in the current table will be added to the result table.
3. **Differential Table Analysis**: if peak picking result tables are present for both sample and control, we can directly run the algorithm based on a comparison of two result tables. This algorithm can be applied to both nominal and accurate result tables. It offers the user more control compared to method 4. E.g. the user could decide to apply differential analysis to certain peaks only, e.g. monoisotopic peaks having charge 2+ and a specified minimum intensity. The ratios are calculated based on peak height. Matching peaks should elute in the specified retention time window. To correct for chromatographic shifts, the user can perform peak peaking on COW corrected chromatograms.
4. **Nominal Peak Based Matching**: based on the peak picking parameters of the sample this algorithm will run peak picking on the control during the differential analysis. It will then detect and report the differences in nominal

mode. The comparison can be done on COW or RPW corrected EICs if needed. This algorithm currently supports finding both up- and down regulated peaks by using a dual analysis; Sample/Control and Control Sample.

Which algorithm to use depends on the quality of the data, the accuracy, the observed shifts between sample and control and also on the level in which the peaks are present. Very small noisy peaks require a different approach compared to finding large differential peaks. Below a more detailed explanation of the four different algorithms is given.

Correlation Based Matching:

The first algorithm is based on correlation and is named Correlation Based Matching. Each chromatographic peak in the result table of 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 calculated correlation is small, probably no reference peak exists at this location. Using this algorithm, mass chromatograms should be properly aligned.

Differential Analysis using Correlation can be performed in combination with a dedicated alignment algorithm. Besides using a fixed offset correction, the algorithm uses a second parameter to correct for peaks that still show **local shifts**. This so-called dithering parameter will set the window used for local cross correlation. The amount of shifting allowed can be set using the dithering parameter. 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.

Alignment correction can also be applied by loading COW (Correlation Optimized Warping) or RPW corrected time scales. COW is an alignment procedure based on the Total Ion Current (TIC) or Base Peak Chromatogram (BPC).

Correlation Based Matching runs in nominal mode. It performs well if peaks are well separated in a single mass chromatogram. If multiple peaks are present that elute closely in the same EIC, it might be more difficult to use this algorithm, especially if alignment is not perfect. In these cases, it will be difficult to perform a ratio analysis on the correct pair. In this case a good alignment will be necessary, so that a tight window can be used to calculate the ratio. To use Correlation Based Matching, select the first algorithm (A) from the differential analysis menu (see the first tutorial).

Peak Based Comparison in Accurate Mass Mode:

If there are many peaks having the same nominal mass value and the data are from a high resolution instrument, you should try to run the Peak Matching in Accurate Mass Mode. First, the table should be converted to accurate m/z values if not done before. To use Accurate Peak Comparison, select the second algorithm (B) from the differential analysis menu (see second tutorial) and be sure to mark the checkbox **Check in Accurate Mass Mode.** You will have to specify the mass window for accurate EIC extraction (e.g. +/- 0.01 Dalton). By default the algorithm will calculate ratios based on peak height. You can also output ratios based on peak area or both.
Accurate EICs will be extracted for both sample and control. The time width (set in the time window edit box) determines the width of the EIC to be extracted. A Peak based comparison calculates the peak height ratios between accurate EICs based on the maximum intensities found in both EICs. If you need accurate ratios based on area you will have to specify a time window comparable to the width of your peaks.

You should use this algorithm for well aligned samples or in cases that the high resolution mode makes the peaks very specific so that you can use a wider time window. If this is not the case, then first apply COW or RPW alignment. Next run differential analysis on the COW corrected sample. The second example in the document will use this procedure. If the sample EIC has multiple peaks in the applied time-window, then these entries will be marked during output. The user should always check these peaks manually. This algorithm currently supports finding both up- and down regulated peaks by using a dual analysis; Sample/Control and Control Sample.

Peak Based Comparison in Nominal Mass Mode

Based on the peak picking settings of the sample, this algorithm will run MPeaks on the control sample after which a direct comparison between both result tables will be performed. Peak picking for the control sample is performed using slightly more sensitive parameters. 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 retention time of the sample peak. The algorithm uses a default value of 0.30 minutes. If multiple peaks are present in the time window of the control, then the one with the largest intensity is taken for ratio calculation.

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. If strong shifts are present, it is advised to first apply a COW or RPW alignment correction on both samples. In this way you can use fairly small retention time windows for comparison and the problem of multiple peaks can be avoided. This algorithm currently supports finding both up- and down regulated peaks by using a dual analysis; Sample/Control and Control Sample.

Differential Table Analysis

The last algorithm uses the results from two peak picking tables to check which peaks are different and similar. You should run and save the result from peak picking on both the sample and control. Optionally convert the tables to accurate mass values. You can also edit the table to select only the peaks of interest or you can convert to a monoisotopic peak table first. In principle, you can use all of the MPeaks features for filtering before saving the tables.

A differential table analysis is very fast as we do not have to extract (accurate) EICs. The analysis can be performed on both nominal and accurate mass tables. If alignment is problematic, you should run the peak picking **on COW or RPW corrected samples**. This has to be done for both the sample and control separately. Peak Picking must be performed after loading the corrected time scales. The Table comparison is more time consuming as different steps have to be performed:

1. Load the Sample
2. Load the COW aligned time scale if necessary: menu > Pre-Processing > Load COW aligned time scales
3. Run Peak Picking
4. Optionally convert to accurate m/z values or apply the filters you want to use.
5. Save the Result Table

Repeat the above steps for the Control Sample. Next, re-load the Sample into the Browser. Next, start MPeaks and re-load the saved result table. The difference between a direct Table Analysis and Peak Based Matching (algorithm 2) is the option to perform the analysis in accurate mass mode and to apply powerful filters. It does however require more steps to get to the results.

Instead of using COW you can also apply a wide retention time window. However, if multiple peaks are present in the same nominal mass chromatogram, the comparison might be difficult. Ratios will then be calculated based on the peak having the largest intensity in the control sample. If you don't want to use COW, you should perform the analysis on accurate mass tables.

To run Differential Table Analysis select: **Menu > Differential Analysis > Find Differences between two Tables.**
Next you will get a question to use a result table belonging to the current sample. Select NO, as we need the result table of a different sample. Browse to the folder where the control sample is located and select the result table belonging to the control (*.mtb). If more than one result table is present, select the one to use. Tables can only be compared in either accurate or nominal mode. Finally, define the Time Window and the Change Fold Factor.
In the results table, the comment field "Ref Peak" will be reported if the peak was also detected and present in the control table. On the right to this comment the ratio will be shown based on Peak Height. You can sort the results using Menu > Differential Analysis > Sort on Ratio.

Example 1: Correlation Based Matching:

The first data set used in this tutorial relates to a proteomics application and the task was to compare and find peaks in Sample A (13DS) that are absent or small in Sample B (16DS). Peak picking was used with default parameters but a selected retention time range. This resulted in the detection of 977 peaks. The user should follow the tutorials on Peak Picking before using Differential Analysis. Read the appropriate chapters in the MsXelerator manual - Chapter 3.3, page 36 on the topic of Peak Picking and Chapter 4 explaining MPeaks Advanced Tasks.

Figure 1 shows the TIC for both sample and control in stacked and overlay mode. As can be seen, the samples show quite a severe shift of about 0.5 minutes over the full retention time range. For this tutorial example we will use nominal peak picking and apply the Correlation based Differential Analysis algorithm in combination with a general off-set correction and a local alignment correction.

Step 1: Start MPeaks and either run the MPeaks peak picking algorithm or load a result table from disk. Sort on Peak Height. The table can be nominal or accurate.

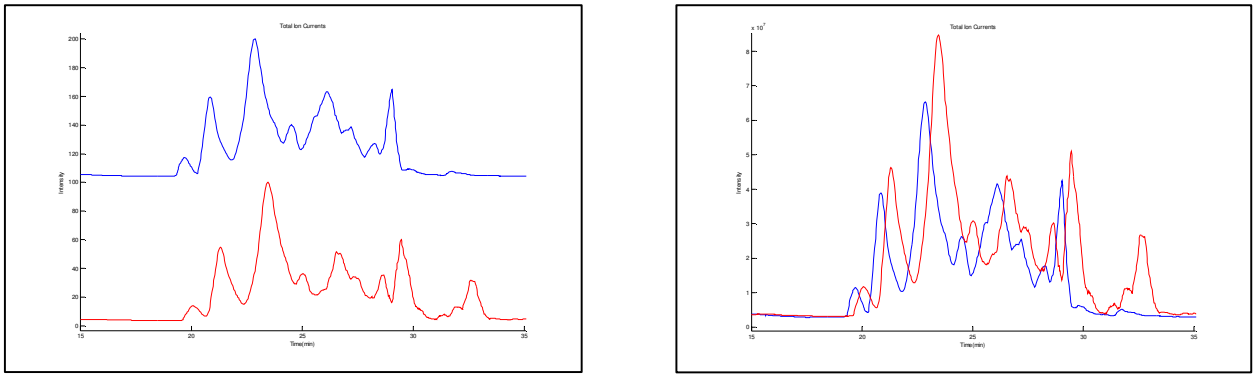


Figure 1: Total Ion Current for Sample A (blue) and Sample B (red) in stacked and overlay plotting mode

Step 2: From the Menu select Differential Analysis > Select and Overlay Ref./Control Sample. Browse to the location of your control sample, select the control sample and press OK. EICs will be plotted in overlay. The sample will be plotted in blue, the control in red. If needed, you can also plot the control sample in mirror mode by selecting Menu > options > plot negative reference.

Using the Correlation based algorithm, you should always check whether the data sets are well aligned. If the two sets are shifted you will **have to** apply alignment correction methods or use a different algorithm. Sort the table on Peak Height and explore a number of the larger peaks to check for alignment problems.

FigureFigure 2 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 2 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 roughly the same amount of shift occurs for a number of peaks. 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 roughly shifted. The local alignment correction will use an advanced cross correlation algorithm to shift all individual peaks to the same retention time. The local alignment will be applied to each peak in the result table. However, if a more or less constant time-shift exists, we can first apply a general time shift first (Off-Set). This will make the local alignment much faster.

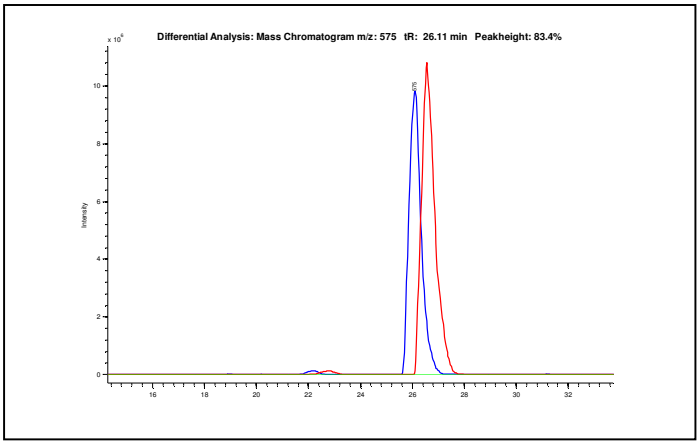
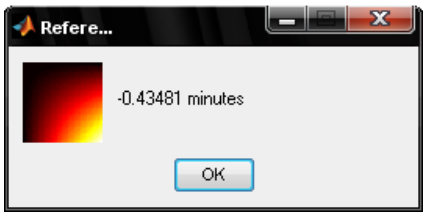


Figure 2: Differential Analysis - overlaying mass chromatograms to estimate offset time shift between both samples.



Step 3: The Differential Analysis Tool offers a utility to measure the shift between both peaks displayed in Figure 2. 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). All reference mass chromatograms will be shifted using this offset in a linear way. The returned offset shift is -0.43581 minutes, which means that the reference should be shifted to the left to properly match the sample peak. We will use this value as a best guess and apply it to all peaks as a default. For many peaks, some misalignment will still exist, but this will be handled in a second step by using a more advanced correction procedure based on cross correlation.

The plot will be updated and shift correction will be applied. Traces are now viewed in aligned mode. You can toggle between aligned and non-aligned viewing by selecting: **Differential Analysis > Plots: view aligned data.** If you need to set the offset correction manually, use **Set Reference Offset Shift** and enter a value in the edit box. To remove the Offset Shift, enter a value of 0. In situations where data sets are well aligned you may skip the above steps. You are now ready to run Differential Analysis.

Step 4: 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. Three algorithms can be selected from the Differential Analysis Window. We will use Algorithm A: Correlation Based Matching, press the radio button for this algorithm.

The previously determined offset shift correction value will be used for Correlation Based Matching and is already entered in the Ref. Offset Shift field, see Figure 3. As explained we will also apply a local alignment correction for each peak individually by using cross correlation (dithering). Each peak is allowed to shift +/- 0.3 minutes (22 scans).

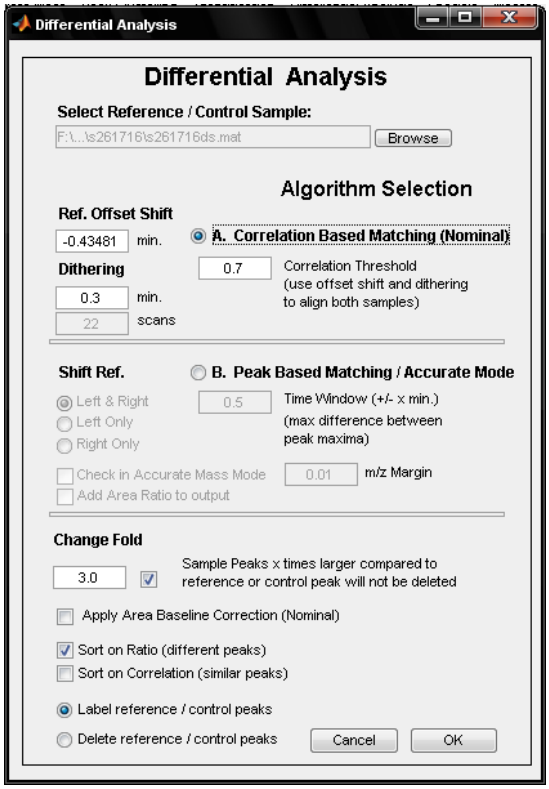


Figure 3: Differential Analysis Main Window

Other Settings:

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). Peaks with a ratio smaller than the Change Fold will be reported as being Control Peaks.

Apply Area Baseline Correction:
As we prefer to run the algorithm on Peak Height ratios, this option is not often used.

Sorting:
By default the table is sorted based on the calculated ratio. To keep the order as shown before, do not use the sort option.

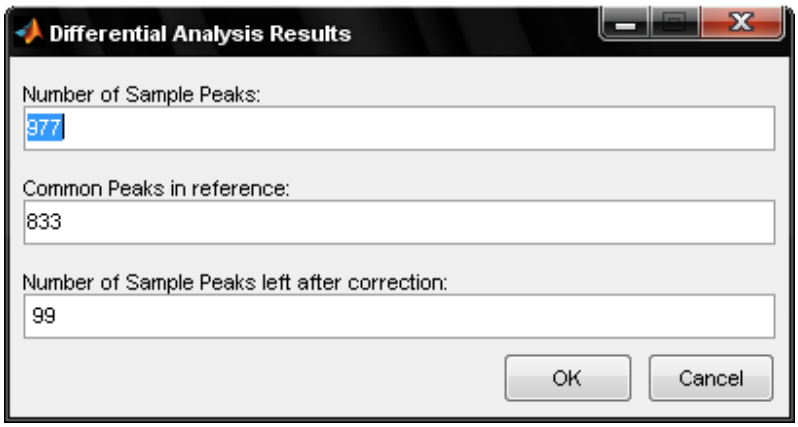
Deleting Control Peaks:
In the case of Algorithm A, the MPeaks table will return two new columns; the correlation value and the ratio between both EIC peak heights. If you only want to keep the "differential" peaks, you should select **Delete Reference / Control Peaks**. The returned table will only show the up-regulated peaks, larger that the Change Fold value. You can also delete peaks manually from the table afterwards.
You can also delete control peaks from the table using the Menu Command: Differential Analysis Menu > Delete Control Peaks.

Cancel:
Press this button to return to MPeaks.

Running the Algorithm:
Press the OK button to run the algorithm. You will be asked to use Peak Heights or Peak Areas for ratio determination. Please use Peak Height, as area determination is often more difficult, especially for small peaks.

Results:

The Differential Analysis result window will give an overview of the number of peaks in the sample, the number of common peaks in the reference sample (peaks in the reference with a correlation larger than 0.7) and the number of peaks that will remain if the delete option would have been used. In total 99 peaks show a Change Fold of 3.0 or larger.



Press **OK**, to continue. You will be asked if summary results on Shift Correction should be plotted. Select No, unless you want to see how the shift correction across the chromatogram looks like.

Press **Yes**, to view the Shift Results. Figure 4 displays the shift results (offset + dithering) 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 shows a histogram of the shift distribution. An average shift of -27 scans is observed between peaks from both data sets.

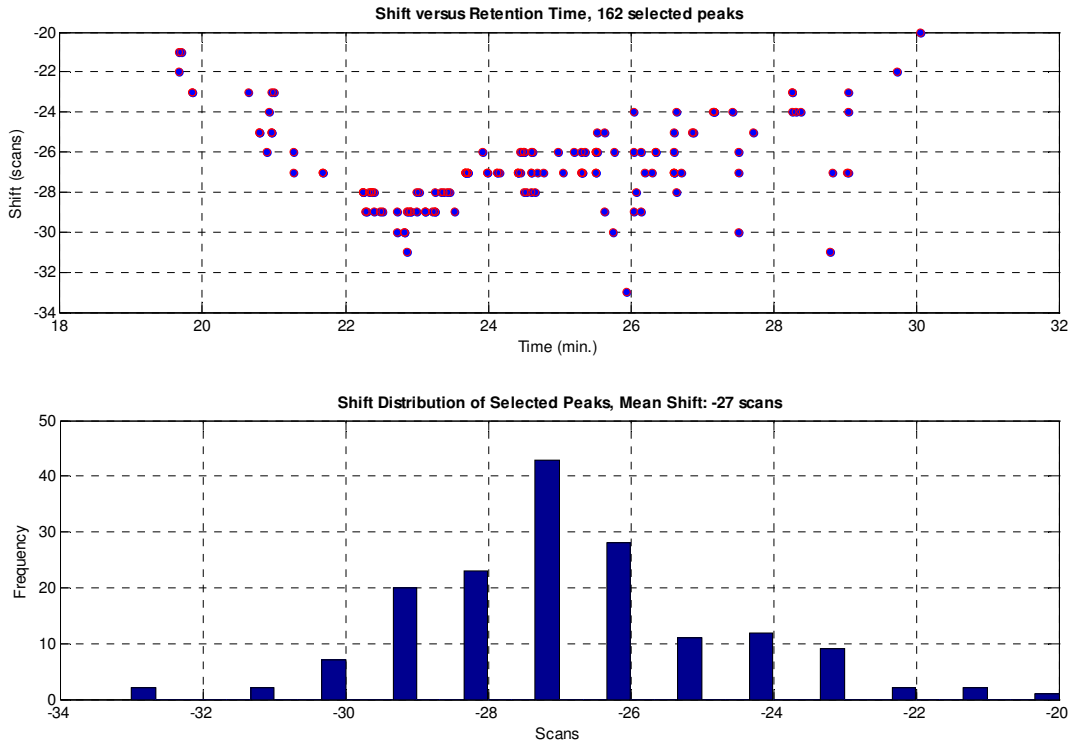


Figure 4: Summary Result Plots for Shift Correction.

You will be returned to MPeaks. The table will be sorted on ratio, see Figure 5.
In all cases a ratio will be calculated even if no peak in the reference sample was found (small correlations). However, sometimes the reference peak intensity will be zero. The ratio value will then be set to a value of 99. To find relevant peaks study both ratio and peak height values. Very small peaks, absent in the reference sample will probably not be very relevant; depending on the type of application.

Figure 5 shows the top peak in overlay with the reference peak, m/z 846 at 26 minutes. 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 -2 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.

Results can be saved to a new MPeaks result table. You can also export the results to a text file or to Excel.

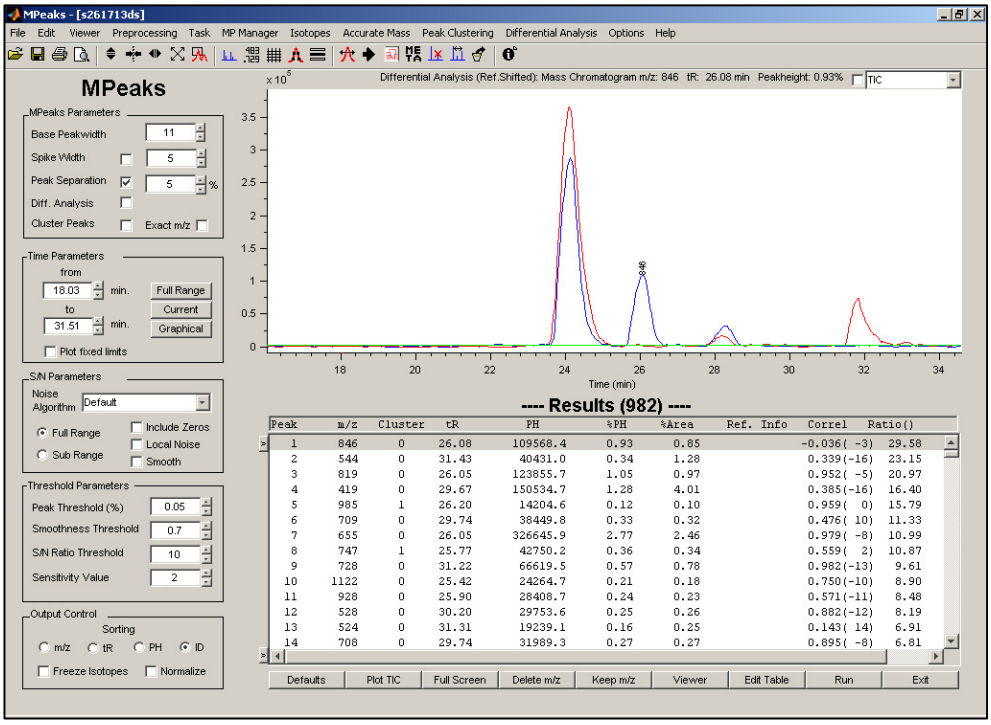


Figure 5: MPeaks Overview after Differential Analysis.

To plot the results for a number of EICs, 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.
You can delete peaks from the table manually by pressing the Del key or press the Delete m/z push button. You can also view the mass spectra from sample and control (corrected for shifting). Press the Toggle MassChrom/MassSpec icon on the Icon Toolbar.

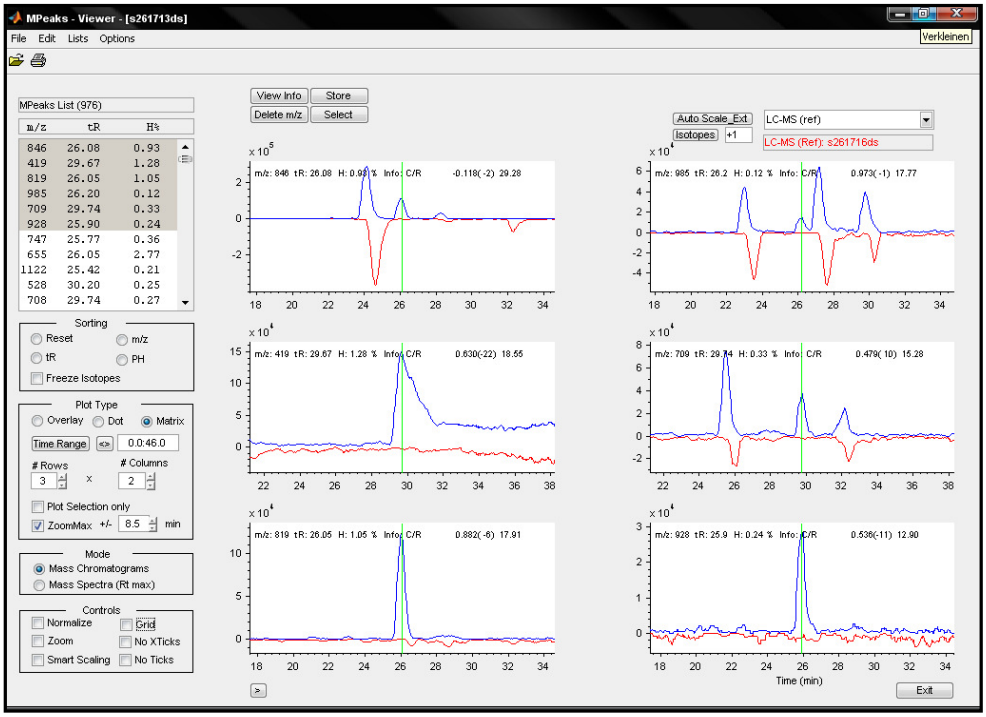


Figure 6: Plotting Differential Analysis Results in Matrix Mode using the MPeaks Viewer (non-aligned view)

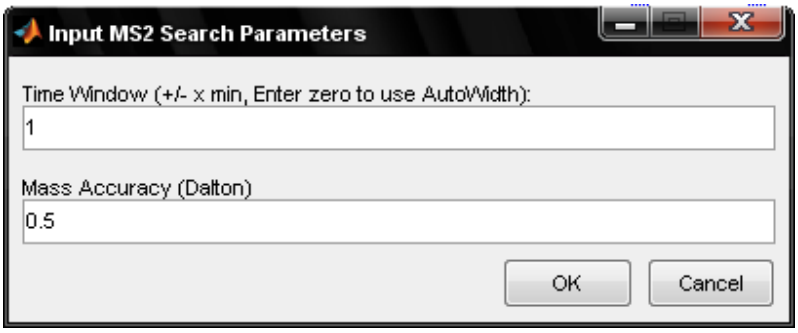
Checking which Differential Peaks do not have MS/MS Spectra:

We would like to test which of the interesting differential peaks in the list have **no MSMS spectra**. It is well know that the MS2 spectra are often based on peaks with high abundances. However, for differential analysis we would like the instrument to focus on differential peaks only, as the common peaks are not of interest. However, many differential peaks will be small having no MSMS spectra. From the differential list in MPeaks the user can check for which of the differential peaks MSMS spectra have been measured. If no MSMS spectrum was measure, you can build an inclusion list of these masses (+time) and re-analyze the sample.

The procedure only works for mzXML based import and the result table should have been converted to accurate monoisotopic peaks. To do this use: Menu > Accurate Mass > Accurate Mass Conversion. Fill in the items of interest. When asked to use narrow or broad range for charge calculation, use broad. Also, you don't have to calculate accurate ratios at this point.
See the Manual (MPeaks – Advanced Tasks).

From the menu select **Task > Label MS2 Scans**. If you get a warning that the MS2 index file is not found and should be created, Press **Yes**. After creating the MSMS index file the following GUI will be presented.

To Search for MSMS spectra belonging to differential peaks, two parameters need to be entered. A time window, in which the MSMS scan should fall, compared to the detected peak. Also the difference between a precursor ion and the m/z value of the detected peak in the table should in theory differ by a small value only. However, we will use a value of 0.50 Dalton. Why? because sometimes it is observed that precursor ions from the MSMS file are actually the ^{13}C isotope peaks.



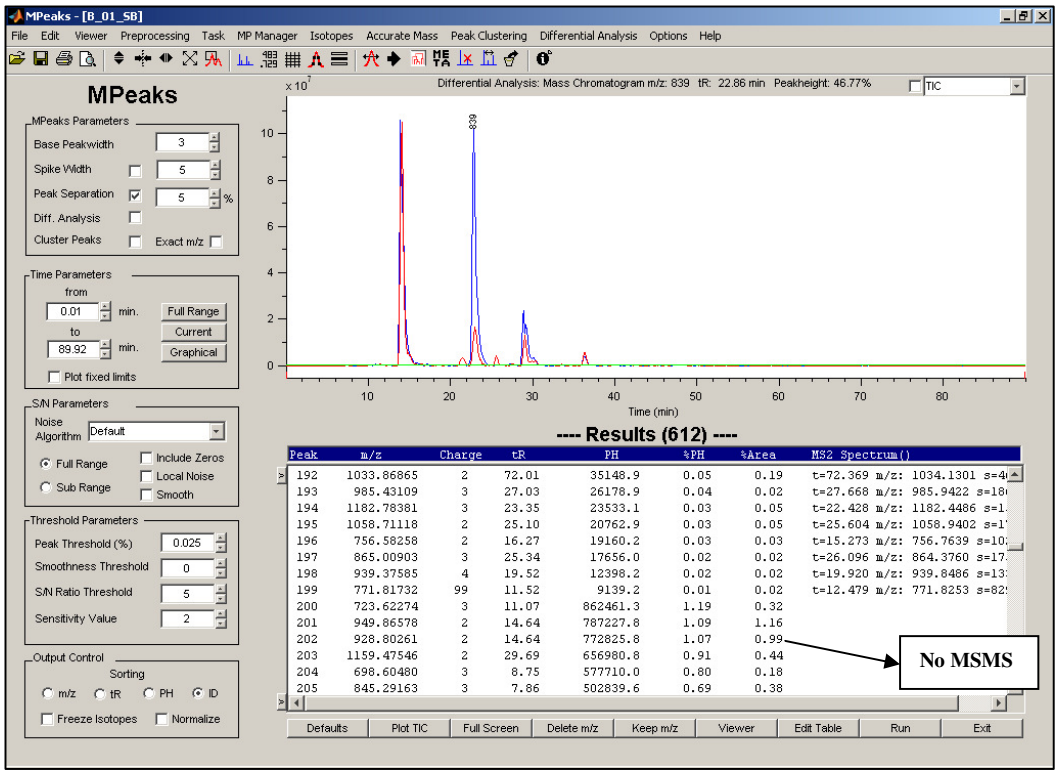


Figure 7: MPeaks Screen after running Differential Analysis and MSMS Labeling. Differential peaks having no MSMS scan are not marked.

The final result table is displayed in Figure 7. You can sort on Peak Height. However, to have all labeled peaks on top of the table, sort on ID (comment field) and scroll down to see where the peaks are having no MSMS Spectrum. These peaks will have no comment field. From a total of 612 peaks, only 199 peaks also have an MSMS spectrum. The other ones, often small, might be the interesting peaks to re-analyze using inclusion lists. The peaks having a label can be deleted as follow. Press “m” to plot the mass spectrum. Select peak number 199 with the mouse, go to the top of the table and then press the first peaks while holding the shift key. All labeled peaks will be marked. Press the Delete m/z button found at the lower part of the screen.

Differential peaks having no MSMS spectra can be saved to text files or Excel, for building an inclusion list.

Example 2: Peak Based Matching in Accurate Mass Mode:

In this second example we will use Accurate Mass Differential Analysis, with and without Correlation Optimized Warping correction. The example is applied to Differential Analysis of a Peptide Mapping study. The Sample and Control relate to Trypsin digested Avastin. The initial sample was oxidized using 3.3% tBHP. We want to compare the oxidized sample versus the initial sample (control). Measurements were performed using Agilent LC1200 system with a Waters Q-TOF Premier. All samples were measured in triplicate.

Figure 8 shows the Sample Overview in the Browser. For a small peak, m/z 619.3, the mass spec and the EIC is plotted. For peak picking in MPeaks is will be important to estimate the width of the peaks in scans. The scans are made visible in Figure 8. The width (near the baseline) is about 30 scans. You can count the dots or use the Width Measurement Tool on the Icon Toolbar. For MPeaks we will use 11 scans as the minimum peak width.

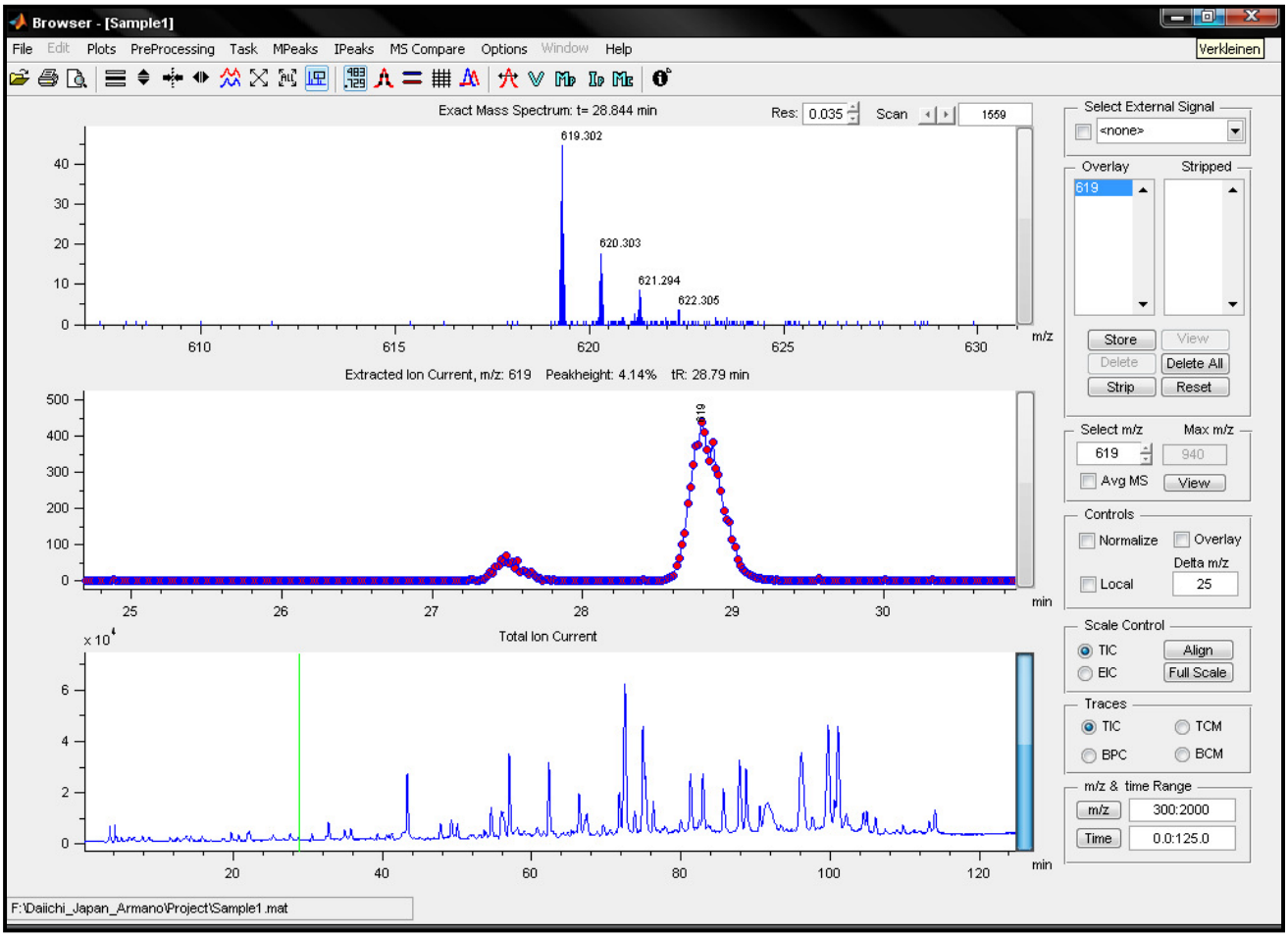


Figure 8: Sample Overview in the Browser.

Peak Picking:

Start MPeaks and load default values. This sample contains many peaks. For the demo we will therefore set the peak threshold to a level of 0.5%, ten times higher compared to the default value of 0.05%. This means that smaller peaks will be deleted from peak picking. Peak Picking will be done in nominal mode. To convert the peaks automatically to accurate m/z values, mark the **Exact m/z** check box in the Peak Picking Parameters section. Accurate m/z values will be calculated based on a single scan at the peak max retention time. Mpeaks has different options to calculate accurate m/z values. Check the section MPeaks Advanced Tasks from the MsXelerator Manual (Chapter 4.2).

Press the Run button from the buttons below the Table region. A total of 2248 peaks will be detected. At this stage, the table includes all isotopes. You can examine peaks by clicking on the entries in the table. You will be asked if plotting should be done in accurate mass mode (e.g. m/z 619.31 +/- 0.01 Da). The resolution used for plotting EICs can be set from the resolution edit box at the top left of the Figure, or from the Options menu. You can easily toggle between EIC and Mass Spec plotting by pressing the Toggle Mass Spectrum icon button on the Toolbar. On the Toolbar you will also find icons to use Auto Zooming and Exact /Nominal EIC plotting.
For large data sets measured in Profile Mass mode, be sure to use **Auto Zooming**, otherwise the extraction might take some time. For this data set a runtime of more than 2 hours was used. See the MsXelerator Manual for an overview of general features for MPeaks and the Browser.

Press the 'h' button to get an overview of all active keyboards keys for zooming, scaling, plotting etc.

As we will only be interested in the larger differential peaks we will first reduce the table by removing all peaks having an absolute intensity less than 20 counts. This seems a fair value for the Q-TOF instrument.
Press Edit Table. An overview will be shown of the basic Peak Characteristics (m/z range, tR range, Peak Height, Peak Height %, Area%, FWHH etc. To remove the small peaks, enter the value 20 in the PH (Peak Height) min box on the left. Press enter. The Num Peaks box at the bottom of the Figure will be change to 535 Peaks. If you want to redo press the Restore MPeaks Table icon from the Icon Bar. Alternatively, you can set a higher Peak Height % Threshold level directly from the MPeaks settings (Threshold Parameters Section).

We are not ready yet, because we want the differential analysis to be performed on monoisotopic peaks. For peaks having charge states 1+ to 4+ this seems to be a good approach. For higher charge states and larger m/z values, it is often better to use the largest isotope peak. The current table can be converted to monoisotopic peaks. At the same time the charge states will be calculated. From the Menu select: Accurate Mass > Accurate Mass Conversion. The Accurate Mass Setting window (Figure 9) will be shown. For this data set select the appropriate settings (Keep Mono, Profile Mode, Q-TOF, Calc. Charge State and recalculate Peak Height. We will not calculate accurate areas. The Accurate Mass Conversion can also be run directly from a nominal mass result table.

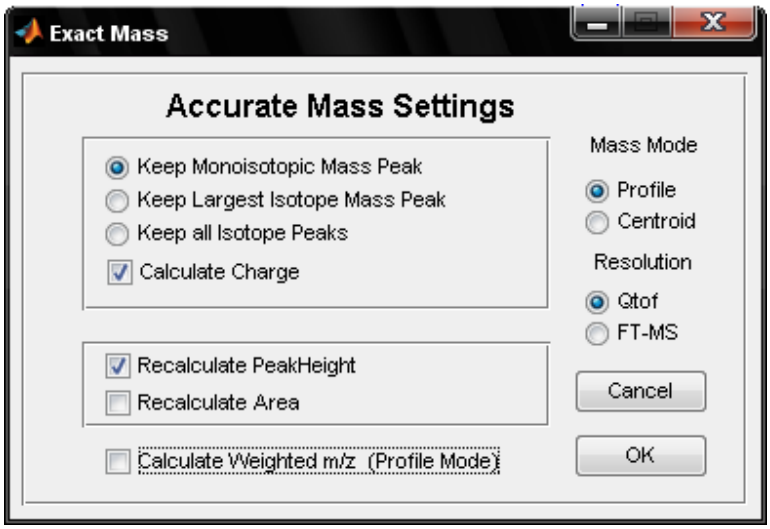


Figure 9: Accurate Mass Settings Window.

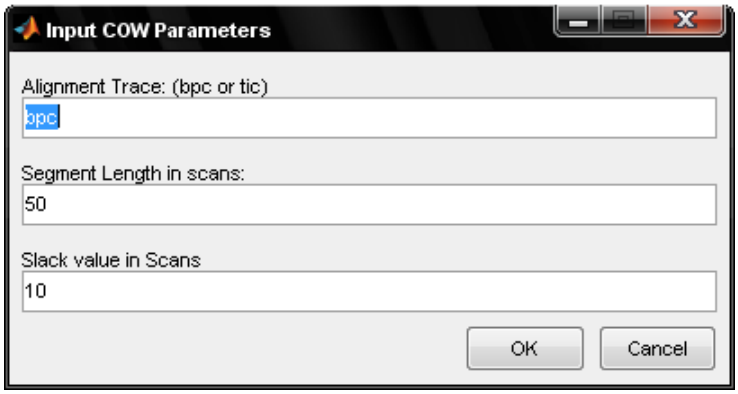
Attention: when using Differential Analysis in Proteomic applications it is sometimes better to keep the largest peak from the Isotope Cluster instead of the mono-isotopic peak. For high masses and charged ions (4+ and up) the mono-isotopic peak is often much smaller compared to the largest isotope peak. A comparative analysis might be more precise using the largest signal.

Press OK to start the conversion. When asked to use a broad or narrow range for charge state determination, use Broad. For SILAC based data having difficult isotopic patterns, it is often better to use a narrow range for charge state determination. Do not calculate accurate areas when asked, since at this moment we will not use the areas. After completion the table contains 280 entries and will be sorted on m/z value. Save e this table so that it can be used at a later moment (Menu > File > Save Result Table, enter a comment), or press the save table button.

Aligning the Sample and Control using COW:

Sort the Table on Peak Height. Next overlay the control sample. Examine a few peaks. It appears that most of the peaks in the control sample are shifted to the right, some are not shifted. As we will run the high resolution differential analysis algorithm we can either use a time window of about 0.5 minutes or use the COW algorithm to fix the alignment problem. In the demo we will apply COW correction and also run the analysis with no alignment correction. From the menu select: Differential Analysis > Run COW Alignment. The following window will appear. Generally, we will apply COW to BPC traces. You might want to use the TICs instead.
COW is applied by dividing the chromatogram in different sections. Each section has a length equal to the segment length in scans. You should know approximately how many sections to use. You can measure a part of the chromatogram that includes a view peaks. The Slack value is related to the amount of peak shifting in scans. Too small values will not solve the alignment problem. Try the basic settings suggested below to start with. COW is computationally intensive. It will be replaced by faster procedures in the future. You can also apply Reference Peak Warping which is extremely fast.

After completion of the calculations, a window will be displayed showing the original BPC traces of sample and control in the top. The lower window shows the COW corrected traces. Corrected time scales will always be saved to disk (and not automatically applied to sample and control yet). For long chromatograms, zoom in using the mouse in a part of interest. Next, press the spacebar to plot the aligned traces in equal scale. Check if alignment is successful. If not try, other parameters. If COW is not able to do a good job, you should use the High Resolution Differential Analysis using a wider time window.



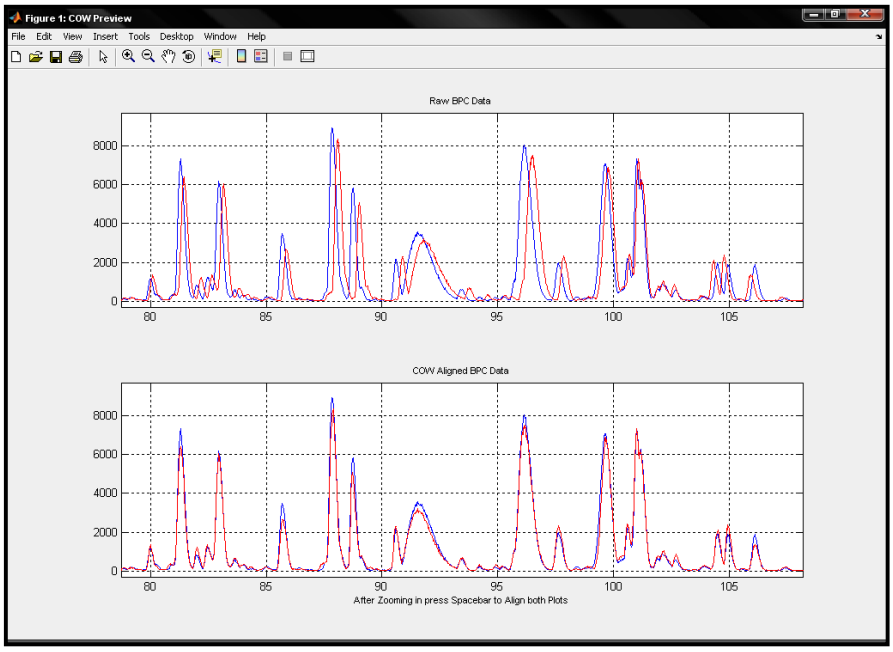


Figure 10: Raw (top) BPC and COW corrected BPC traces

From Figure 10, it can be concluded that COW Alignment appears to be quite successful.

Attention: COW alignment can be more difficult in proteomic studies where the TIC or BPC is much overloaded, due to the enormous amount of peaks present. In these cases it is better to use Reference Peak Warping on "user created TICs". This alignment tool is available from the MsCompare module.

Running Differential Analysis:

First we will run the analysis using no alignment correction. Start Differential Analysis using: Menu > Differential Analysis > Run Differential Analysis. In this case select Algorithm B; Peak Based Matching. Be sure the check the **Check in Accurate Mass mode** checkbox. Also specify the mass window for EIC extraction. For the Q-TOF a value of 0.05 Dalton is appropriate. On FT-MS instruments you might want to use a higher resolution of 0.01 Da. Do not forget to specify the **Time Window.** The width of this window should be entered in minutes. It should be based on the expected shifts between peaks from sample and control. If shifts are significantly, a too small value will miss the top of the control peak for calculating the ratio. A too wide range might include the wrong peak if multiple peaks are present in the control at the specified resolution.

Accurate EICs will be extracted for both the sample and control. The time width determines the width of the EIC to be extracted. A Peak based comparison calculates the peak height ratios between accurate EICs based on the maximum intensities found in both EICs. If you need accurate ratios based on area you will have to specify a time window comparable to the width of your peaks. Press OK, and the calculation will start. The summary overview will tell you how many peaks are different. In total 8 peaks were detected having a ratio of 3.0 or larger.

You can sort the result table on ratio. Menu Diff. Anal. > Sort on Ratio. The sorted table is shown in Figure 12.

The calculated ratios are displayed at the end of the Table. If the ratio is smaller than the specified Change Fold limit, the peak is marked as being a control peak (present in the control sample).

Attention:
Sometimes, a * is added to the comment field. This means that at the specified resolution and within the time window that was applied, more than one peak is present in the Sample. The user should check that the correct pair has been used in the ratio calculation.

You can remove the Control Peaks and save the results to a new MPeaks table.

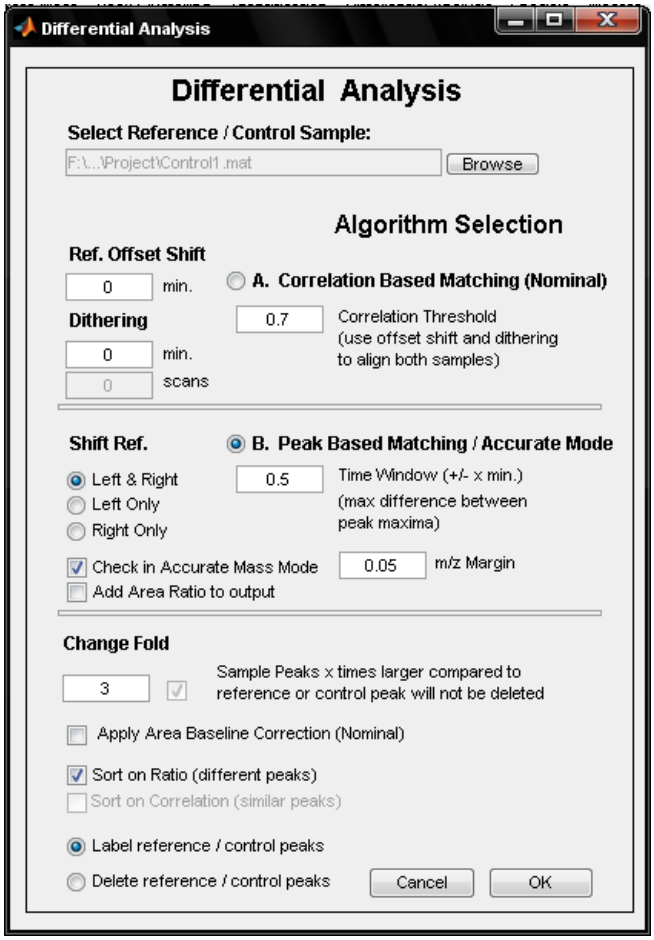


Figure 11: Differential Analysis GUI set for High Resolution Peak Matching

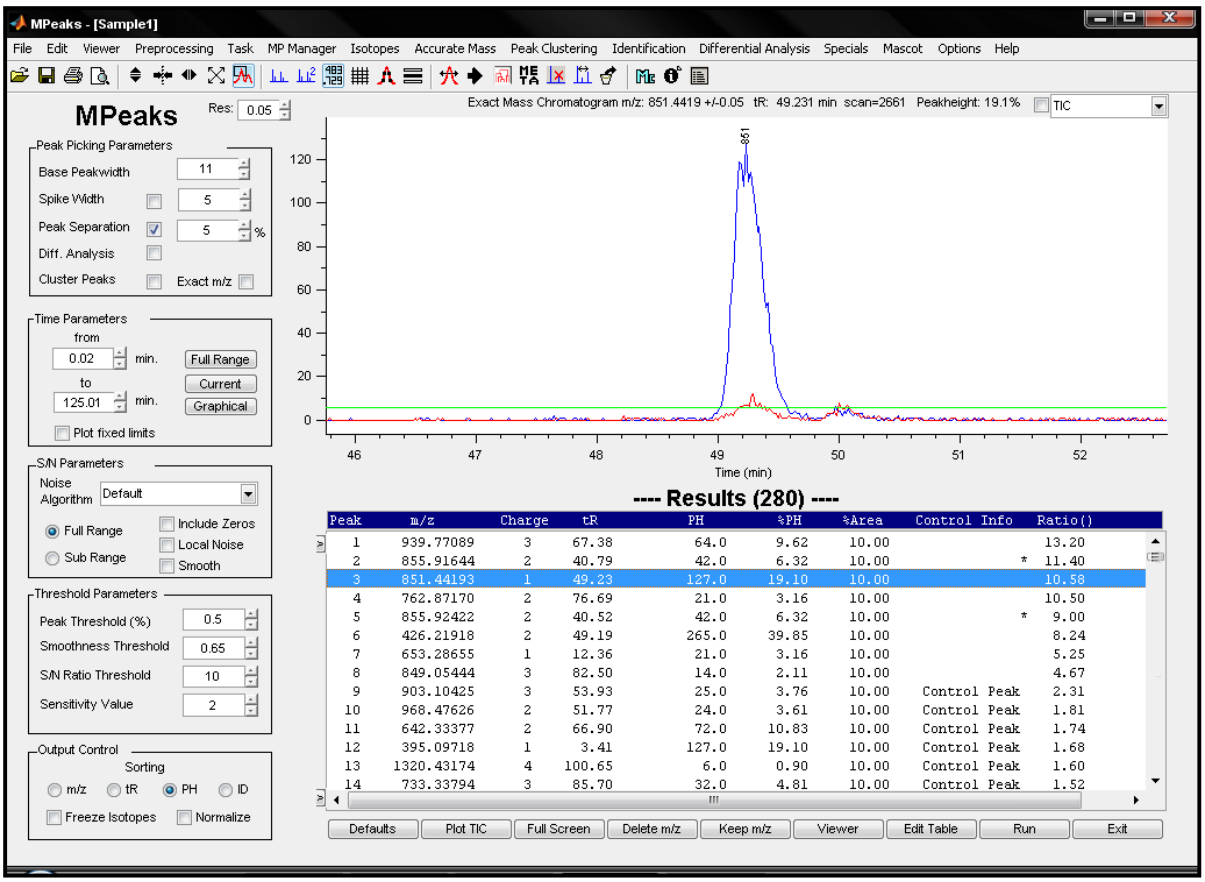


Figure 12: Differential Analysis Results obtained in High Resolution mode. Table sorted on ratio.

Running Differential Analysis on COW Corrected Samples:

In the next demo we will run Differential Analysis based on the Time Corrected (COW) sample and control. The COW corrected time scales from Figure 10 have been saved to disk. To apply the new time scales select: **Menu > Differential Analysis > Load COW time scales**. You will get a question to confirm loading the new time scales. The plots will be updated and viewing of the Aligned data will be activated. Too switch between aligned and non-aligned plotting change this option in the Differential Analysis Menu.

Now run the same procedure, **but use less wider limits regarding the time window, e.g. 0.2 minutes**. Press OK. The result shows that 7 peaks are found to be differential. So one of the peaks from the previous un-corrected run seems to be different. The peak is displayed in the original time scale in Figure 13.

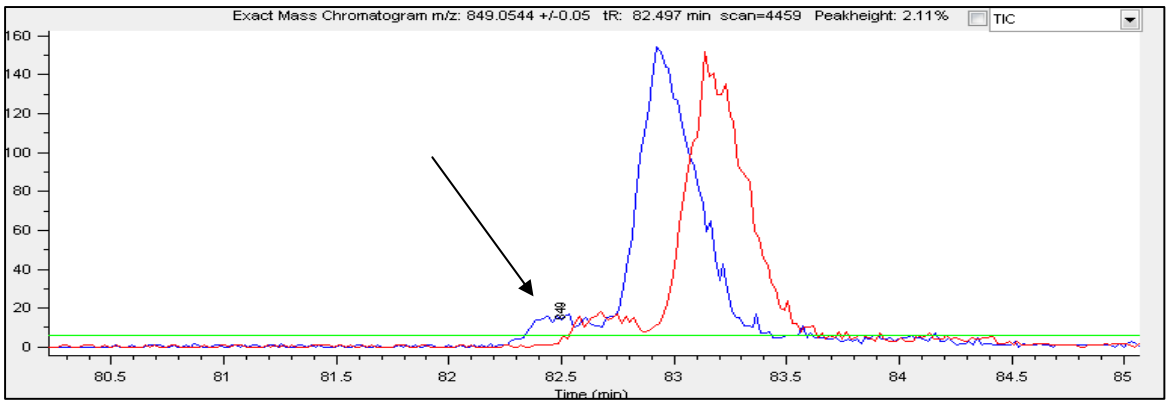


Figure 13: small peak (left shoulder on blue trace) having different result between un-corrected and corrected time scales
In the original time scales the peak ratio is calculated using a time window of 0.5 minutes. The plot shows that at this resolution the peak has a larger interfering peak on the right. Therefore the ratio was not correctly calculated. In the corrected time frame, the ratio is taken from a smaller range using aligned traces. Now the ratio will be correct. Also the table will not display any *, because when using a smaller time window of 0.2 minutes no double peaks are present in the sample. When the add Area Ratio to Ouput option was selected, the area ratio will also be reported. In general ratios based on area and peak height should be comparable. If this is not the case, you should look into the results in more detail.

You can of course also load the COW corrected time scales and apply any of the other differential analysis algorithms. In the case of a direct table comparison (the result table of the control is stored on disk), you should load the COW scales before running Peak Picking, otherwise the retention times in the table will not be corrected.

Plotting the MS Spectra of Sample and Control

To plot the MS spectra of both samples, click the MS button on the toolbar. Use Auto Zooming to automatically zoom in on the MS peak. You can set the default MS zoom range from the Options Menu. After zooming in, press the Shift A key combination to plot the MS spectrum of the control using the same scale. When COW was applied, the MS spectrum of the control sample will be extracted based on the corrected time scale. The title of the MS plot of the control sample will display the original time scale.

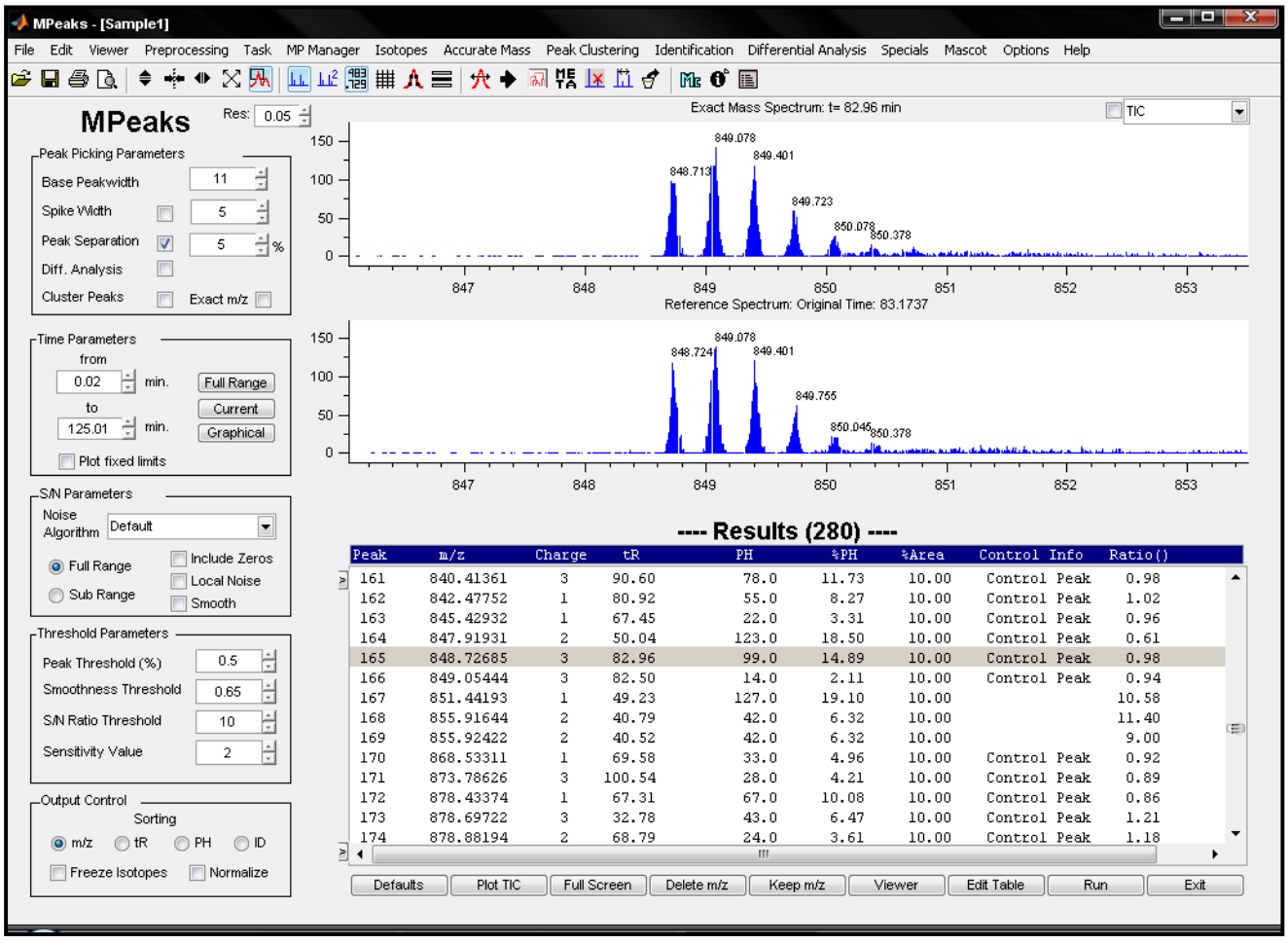


Figure 14: MS spectra of Sample and Control

Additional: Use Mascot Precursor ions for Peak Picking

In this document Peak Picking was performed by MPeaks based on Full Scan MS data. However, the module also has the possibility to perform peak picking on Mascot result files. This means that the precursor ions from Mascot Identified Peaks can be used for peak picking and differential analysis. However, the precursor time information is often not related the correct peak maximum retention time of the chromatographic peak. The algorithm will therefore recalculate the real retention times and is able to remove all other redundant MSMS precursors having the same mass. This type of Peak Picking can be accessed from: **Menu > Mascot > Import Mascot Result File, followed by a high resolution recalculation of the retention times.** After that, all available procedures can be used as shown in this document.

Finding Down regulated Peaks:

The above procedure (sample/control) was used to find up- regulated peaks in the sample (oxidized form). Peak picking was therefore performed on the sample. Using the above procedure you will often find down-regulated peaks too. However, you cannot find down-regulated peaks that are completely absent in the sample, simply because the peak picking was performed on the sample. To find down-regulated peaks in the control (or unique peaks) when using Algorithm A, you should reverse sample and control (perform peak picking on the control) and repeat the procedure. You can directly load a new sample from MPeaks, if needed.

New: Finding up- and down-regulated peak in one run:

For algorithm B (Nominal or accurate Peak Picking algorithm), it is now possible to detect both up- and down-regulated peaks. This algorithm will perform a dual analysis on both Sample/Control and Control Sample. Down-regulated peaks that are found in the second step and not present in the result table are added to the result table. The algorithm currently only operates on peak height ratios.

At the end of the Sample/Control check you will get the question below. Press yes, to run a reversed check. The reversed check will perform a peak picking on the control, convert to accurate mass values and will add to the table all unique peaks from the Control that have larger intensities compared to the smallest peak in the result table. The algorithm will give you a message how many peaks from the control sample will be added.

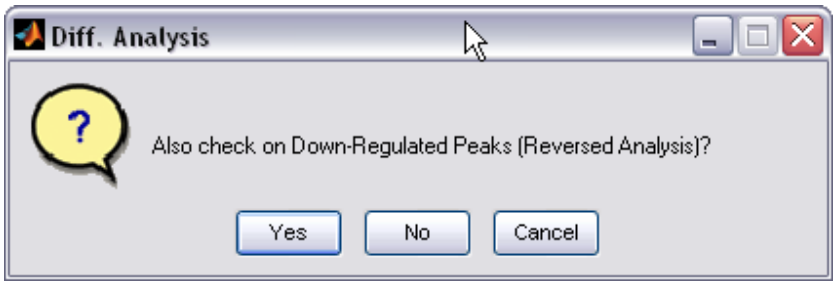


Figure 15 displays part of the result table after merging the peaks from the control sample. The first 250 peaks originate from peak picking on the sample. In total 62 peaks from the Control sample were added (Origin Field). These peak will also have a comment (Not in Table), as these were not found in the original table. The ratio will always be calculated as Sample/Control.

To reduce the list to only up- or down-regulated peaks, select: **Menu > Differential > Analysis > Delete Control Peaks**. Enter a ratio threshold, e.g. 3.0. Next select if this value should be used for both up- and down-regulated peaks. Press Up & Down. Peaks having a ratio larger than 3.0 or smaller than 0.333 will be kept in the result table, all others are removed.

Peak	m/z	Cluster	tR	PH	%PH	%Area	Ref Info	Ratio	Origin()
239	1102.03882	0	85.72	27.0	4.04	3.59		1.29	:Sample
240	560.34094	0	42.74	88.0	13.15	4.08		0.95	:Sample
241	439.21799	0	15.80	103.0	15.40	3.15		1.00	:Sample
242	891.95386	0	93.48	46.0	6.88	3.94		0.83	:Sample
243	1199.30054	0	99.63	32.0	4.78	4.23		1.06	:Sample
244	848.72412	1	82.98	118.0	17.64	3.64		0.98	:Sample
245	1976.00745	0	113.30	27.0	4.04	3.63		1.03	:Sample
246	1163.66370	0	71.71	57.0	8.52	3.22		1.07	:Sample
247	919.93134	0	83.64	49.0	7.32	3.24		1.00	:Sample
248	379.12256	0	3.41	93.0	13.90	1.14		0.45	:Sample
249	300.69391	0	20.74	122.0	18.24	2.64		1.02	:Sample
250	706.35590	0	30.53	79.0	11.81	2.21		0.96	:Sample
251	419.24341	0	55.80	41.0	6.33	3.91	Not in Table	0.42	:Control
252	560.28851	0	75.25	61.0	9.41	5.07	Not in Table	0.82	:Control
253	565.95618	0	50.04	67.0	10.34	2.64	Not in Table	0.64	:Control
254	588.33948	0	81.31	90.0	13.89	6.28	Not in Table	1.18	:Control

Figure 15: Part of the result table after running a down-regulated check. Peaks from the control sample that are unique are added to the result table.

In case you need any help on Differential Analysis
contact:
support@msmetrix.com

==========