MsMetrix

MsXelerator Application: MS Protein Deconvolution of Multiple Charged Ions to estimate Molecular Weight

This document describes a newly implemented algorithm for deconvolution of multiple charged ions from a user selected Mass Spectrum. Two algorithms are currently implemented, one based on multiple charged ions in which the charge ions should be next to each other, e.g. 15^+, 14^+, 13^+, 12^+ etc. The second algorithm does not rely on such a sequence and can be used to solve more complex problems, e.g. mixtures of two or more proteins.

This document describes how to start and use the MS Deconvolution module. The algorithm can be started both from the Brower and MPeaks.

Starting from the Browser:
Figure 1 displays the Browser screen with the Total Ion Current (bottom), extracted mass chromatogram (middle) and the mass spectrum at the selected retention time in the top window. Be sure that the mass spectrum is not displayed in nominal mode. Normally when loading a new file, the MS spectrum is first shown nominal. Just click on the maximum retention time of the extracted chromatogram peak to re-plot the MS spectrum in accurate mode.

Starting the MS Deconvolution algorithm will select only the zoomed part of the mass spectrum. So before starting Deconvolution, zoom in on the part of interest. Start the MS Deconvolution GUI as follow: **Menu > Task > MS Deconvolution.** The GUI will open in a separate window, shown in Figure 2.

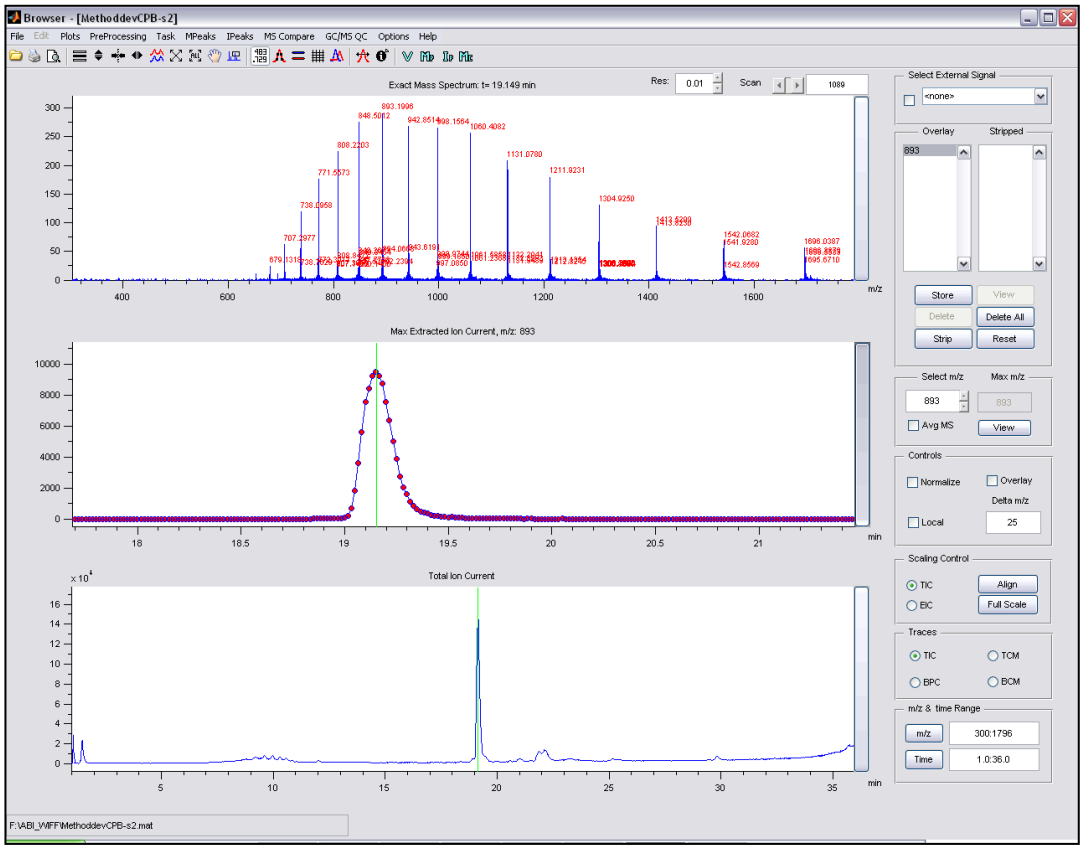


Figure 1: Browser Opening Screen. The selected part of the mass spectrum to be deconvoluted is shown in the top window.

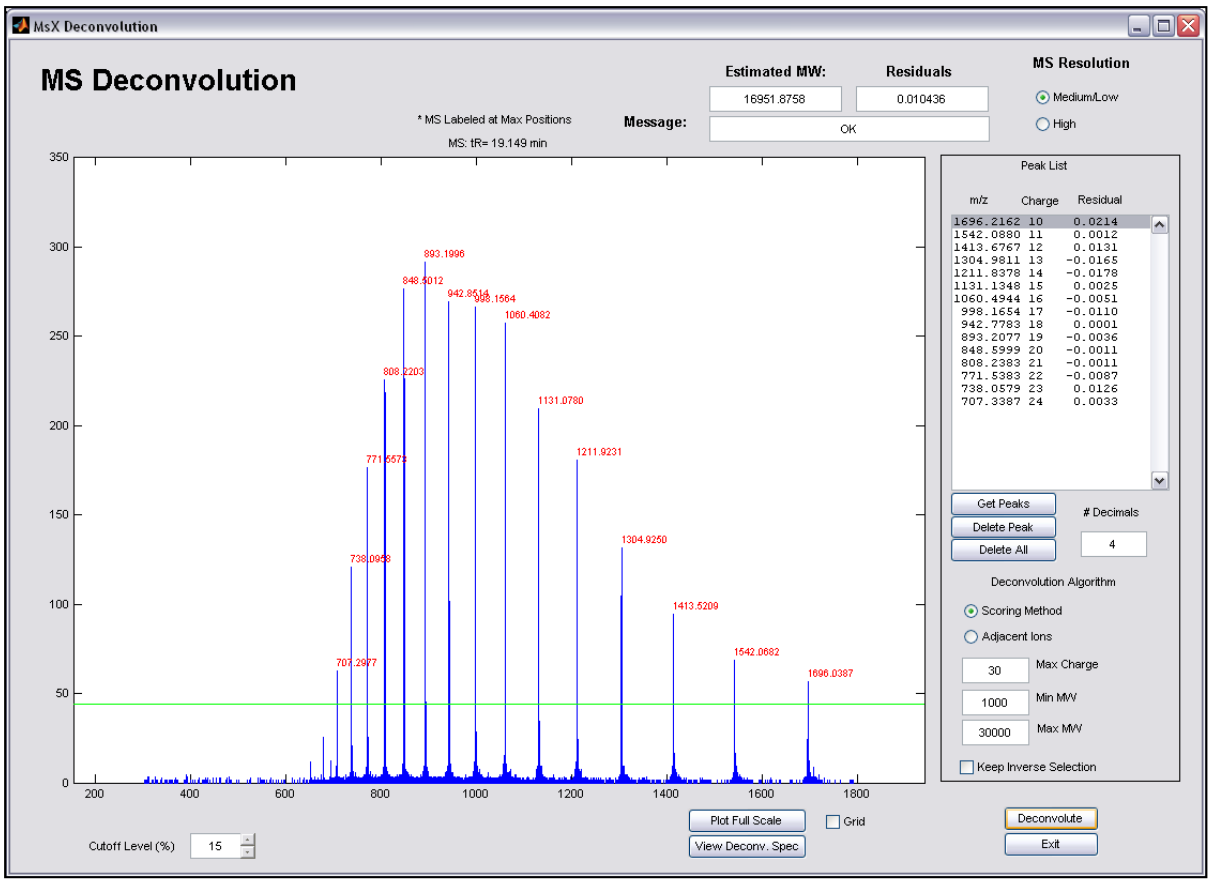


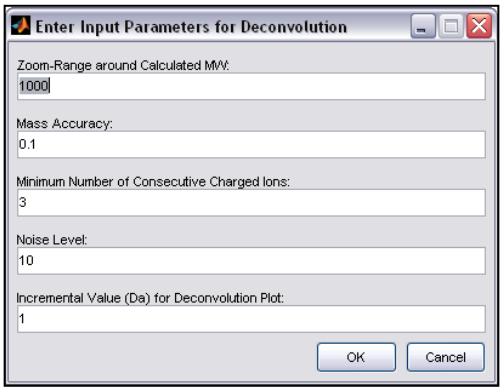
Figure 2: MS Deconvolution Window

Figure 2 shows the selected mass spectrum in the MS Deconvolution GUI. The retention time of the selected mass spectrum is shown at the top of the plot. The peaks are labeled at their maximum m/z value. Depending on the instrument resolution peak maxima can be re-estimated with better precision. As a default, low resolution has been selected from the MS Resolution options (top right of screen). This means that the algorithm will calculate a new average intensity weighted peak position. This will be a more precise value compared to the actual top of the peak.

Press the **Get Peaks** button, and the selected peaks above the threshold level (green line) will be added to the Peak List box. The peaks positions will be re-estimated. You can zoom in on one peak to see the differences. Below the Peak List Window, a few buttons are available to remove selected peaks from the list, or to remove all peaks.
To select peaks from a narrow range, zoom in and press the Get Peaks button. The threshold level will be based on the maximum intensity in this range. To undo, press the **Full Scale** button and press the Get Peaks button again.

The **Cutoff Level (%)** will set a minimum intensity threshold compared to the strongest peak in the display. A default value of 15% is used. You can set this value higher or lower depending on the quality of the spectrum (removing small intensity peaks). Use the slider to set a new value. You can also directly enter a threshold value in the **Threshold Edit Box**. The Peak List box will be updated automatically.

To start deconvolution, press the **Deconvolute button**.



To View the Deconvoluted MS Spectrum, **Press View Deconv. Spec**. You will be asked to enter a few parameters related to the resolution of the deconvoluted spectrum. These parameters are: zoom-range around the calculated MW, Mass Accuracy , the minimum number of charged ions for any MW calculated in the specified range, the Noise threshold based on the raw input spectrum and the resolution of the calculated deconvoluted spectrum. This algorithm is developed to view the MS spectrum close to the calculated

Molecular Weight, not for calculating a spectrum from e.g. 1000 -100.000 Da. Figure 3 displays the Deconvoluted Spectrum +/- 1000 Dalton round the MW value.

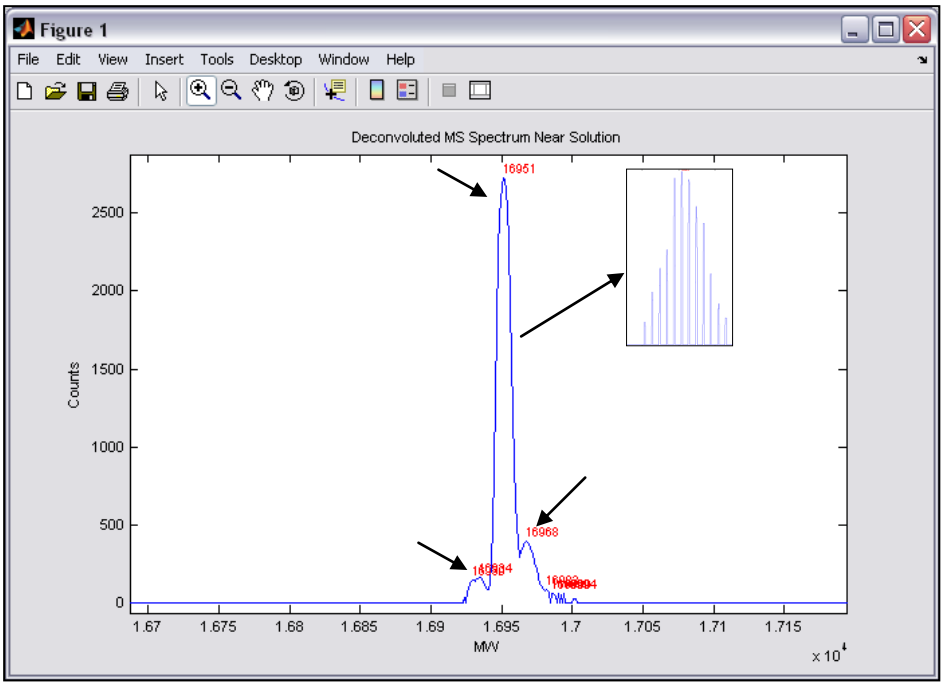


Figure 3: Deconvoluted MS Spectrum near estimated MW.

Fine-Structure MS:
If you want to plot the deconvoluted peak with high resolution, you should select a small incremental plot value e.g. 0.01 instead of 1 Dalton. To see the fine-structure for high quality data you also should select a high mass accuracy (peak picking) e.g. a value of 0.01 (see inset for an example of the fine-structure of the main peak). If the mass range is quite large, the calculations can be lengthy.

Algorithms for Deconvolution: 1. Adjacent Ions

In this tutorial we will first use this algorithm. Select the **Adjacent Ions** algorithm.
We need to be sure that the selected ions are all part of the same sequence with charge differences of 1+. No ions must be missing in the selected series. The algorithm will need a minimum of 3 selected peaks. Press the **Deconvolute** button at the lower right of the screen. If successful the algorithm will return the Molecular Weight, the Sum of the Residuals for all peaks, for each individual peak its charge state and the residual for each individual peak. The residuals are the differences between the calculated m/z value for each peak and its measured value. This is based on the estimated MW and the charge states. If residuals are too large you will get a warning that no minimum in the search for an optimum could be found. Check that peaks are really adjacent in charge or use the second algorithm.

Algorithms for Deconvolution: 2. Scoring Algorithm:

This method is more advanced and uses an iterative algorithm to estimate the Molecular Weight. Basic input to this algorithm is a peak list. Compared to the first algorithm no assumptions on adjacent ions regarding charge states are made. You can also use this for more complex situations, e.g. mixtures of two proteins or a low intensity mass spectrum, in which case it is more difficult to select the peaks. The algorithm will need a minimum of 3 peaks.

The algorithm has 3 basic input parameters.

- The maximum charge state (default 30). Change this value based on prior knowledge or experiment.
- The minimum MW, set to 1000 as default
- The maximum MW set to 30000 as default value.

You should adapt these values based on the expected MW. The program will be slower using wider ranges. Very wide ranges can cause memory problems in some occasions.

Based on the input m/z list, the algorithm will calculate/estimate the MW. It will only use the peaks that will give a fairly good minimum in the residual values. When finished, the peak list box will contain the peaks that were used in the optimization. All other peaks will be removed. Before the start of the algorithm all peaks are labeled in red in the mass spectrum plot. After the calculations are finished, the ones used during optimization will be displayed in red; the other ones not part of the protein spectrum will be marked in black.
Be sure to have a good check on the residuals and the determined charge states. In some occasions the algorithm can get stuck into a different minimum by setting charge states as e.g. 22+, 20+, 18+ etc. Clearly this is wrong. Check the residuals, delete the largest one, and run the algorithm again.

Running a more complex spectrum:

Figure 4, displays a more complex spectrum from the same sample. The peaks are close to the noise level and it is not clear which peaks are part of the multiple charge series belonging to the protein.
At the start, about 20 peaks were selected at the threshold level shown. Running the algorithm will return 6 ions, colored in red, as shown in Figure 4. The estimated MW = 7527.9545. Residuals are quite low and charge states are adjacent. One of the residuals is however a little bit higher. Try to delete it from the list and run the algorithm again. Now the MW is estimated to be: 7528.0776. This value is very similar to the first result. Of course the accuracy of the selected peaks will determine the precision of the estimated MW. It might be better to run the algorithm on the top 3-5 most intense ions. You might want to experiment with this option to see how it influences the results.

Deconvolution in case of a mixture:

Figure 4 shows the peaks that were used in the optimization. The ones not used are colored black. Might be that some of these peaks also belong to a protein. How to run deconvolution on the remaining peaks?

1. Run the algorithm on the full list, but check that the **Keep Inverse Selection** box is checked. Deconvolution will be performed like before. You will get a message that the List Box will now contain the non-fitted peaks from the first run.
2. Run the algorithm as second time on the remaining peaks (the Keep Inverse Selection box will be automatically set un-checked).
3. The results are shown in Figure 5, the estimated MW is 7004.7564 and the series of ions is displayed in the List Box.
4. Repeat this procedure if necessary for even more complex situations.

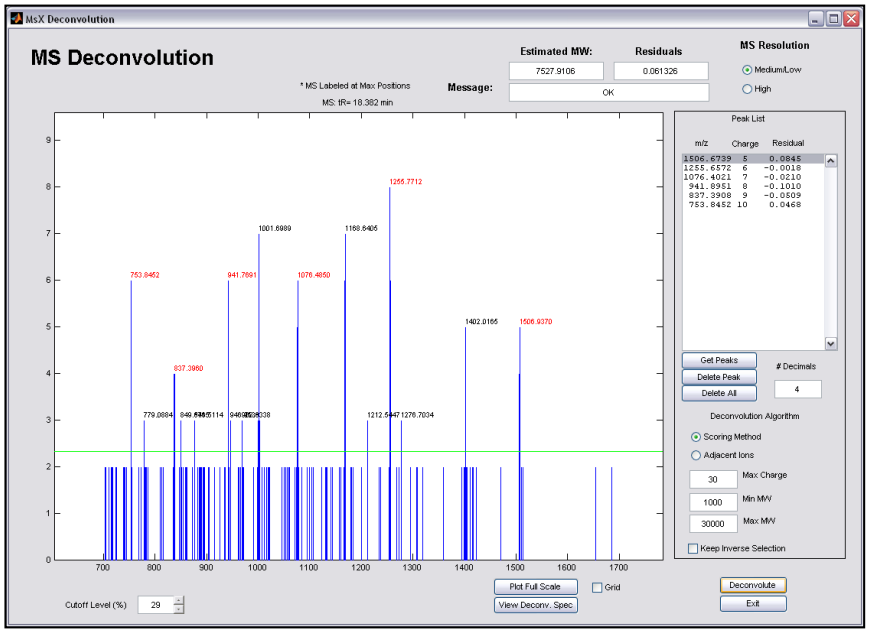


Figure 4: MS Deconvolution applied to a more complex low intensity mass spectrum

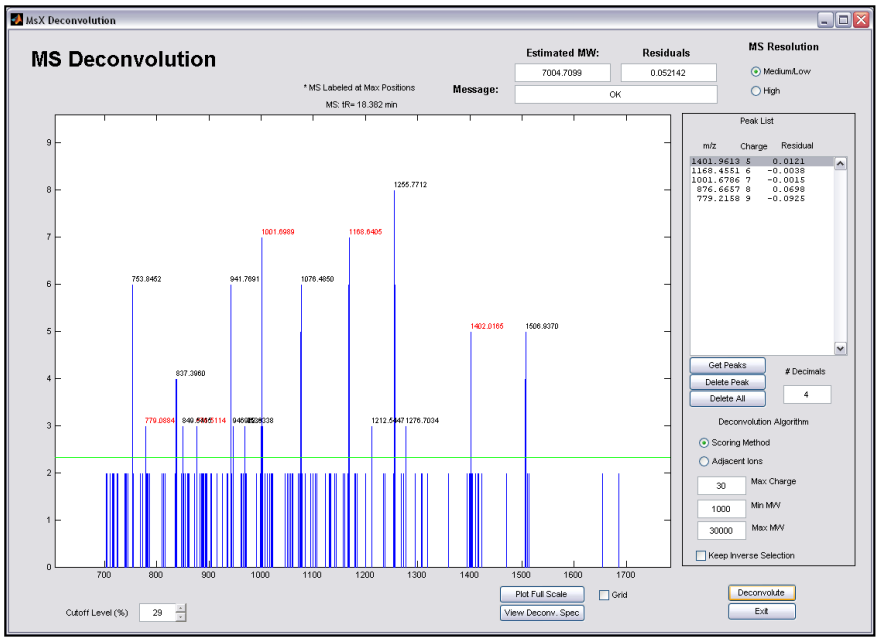


Figure 5, MS Deconvolution Solution when run in Inverse mode to detect and estimate second protein in mixture

Figure 6 shows the Deconvoluted MS spectrum in the region 6500-7800 Da. The two calculated MW are clearly detectable, other small peaks are false positives. The noise level was set to a value of 2 counts.

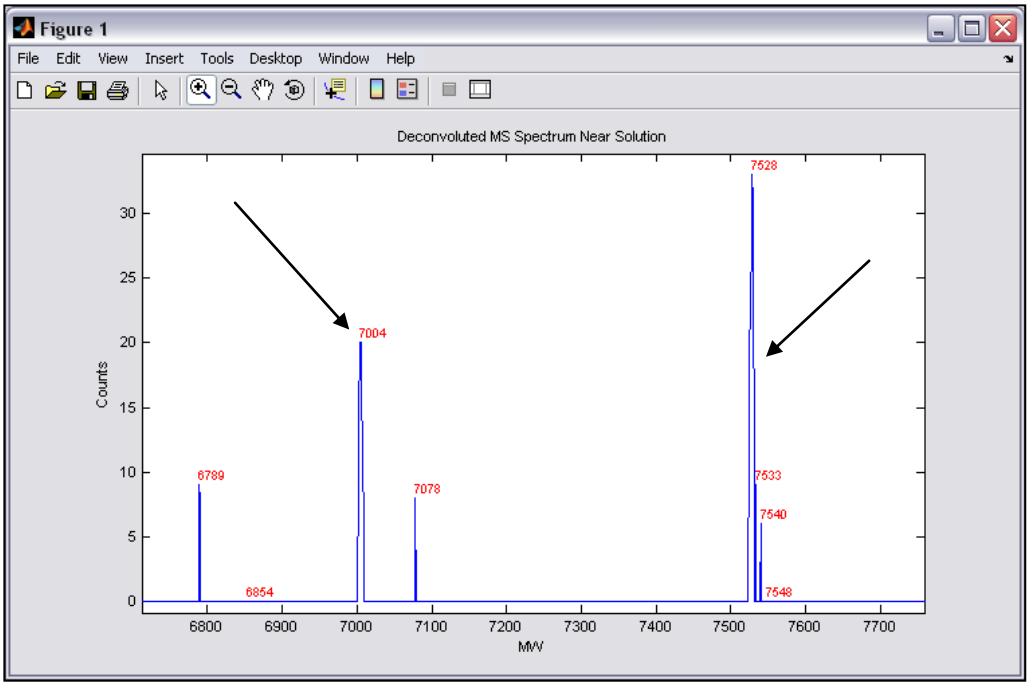


Figure 6: Deconvoluted MS Spectrum, region 6600-7800 Da.

Running Deconvolution from MPeaks

Deconvolution can be started from MPeaks using the following procedure:
Run MPeaks. You can for each peak in the table plot it mass spectrum by pressing the ‘m’ key on the keyboard or select the toggle mass chromatogram/mass spectrum button from the Icon Bar.

To start the Deconvolution, first zoom in on the part of interest. Next press the**‘d’** key on the keyboard, or select **Menu > Specials > MS Deconvolution.**

Some tips when using MPeaks:
As the MPeaks algorithm will try to detect all significant chromatographic peaks in your data, the list will also include ^{13}C isotopes, adducts and fragments. As for the data in this tutorial, the MPeaks list will contain all the charged ions. These can be recognized as co-eluting chromatographic peaks.

A trick would be to sort the table on retention time and see which peaks are part of the protein spectrum of charged peaks. Alternatively, you can use the Clustering Algorithm to group co-eluting peaks. After an MPeaks run, select **Menu > Peak Clustering**. See the details for this algorithm in the manual in chapter 4.6
Basically, the algorithm will try to group co-eluting peaks in a cluster. Within each cluster, the largest peak will be on top of the list. If needed, you can delete all other peaks part of each cluster. After Clustering, the third column of the MPeaks table will contain the cluster number. Try to look for significant clusters and run Deconvolution from this list.

Alternatively, you can create a DOT plot from the Viewer and check which peaks form a co-eluting group graphically.

==========