MsMetrix

# MsXelerator MSX-Quant: Peptide Quantitation based on Stable Isotope Labeling

## 1. Introduction:

Isotope labelling combined with liquid chromatography-mass spectrometry (LC/MS or LC/MSMS) provides a robust platform for analysing differential protein expression in proteomics research. Efficient analysis of raw data for peptide and protein quantification is a principal challenge in mass spectrometry (MS)-based proteomics. This document describes MSX-Quant; a module part of MsXelerator for relative quantitative analysis of SILAC based high resolution mass spectrometry data. MsX-Quant performs quantitative analysis of large data sets generated from proteomics experiments using various stable isotope-labeling techniques, e.g. SILAC, ICAT, and user-developed labeling methods.

MsX-Quant requests spectral data files imported into MsXelerator (Thermo raw, Waters, Bruker, Sciex, Agilent, mzXML or NetCDF) and a file containing information on the peptides to be quantified. This file can be a user-generated Excel file containing basic peptide information, a search result file from Mascot or an MsXelerator generated peak list. In the latter case no peptide ID information will be available. However, quantitation can still be performed and identification of the peptides of interest could be performed afterwards.
Quantitation is based on ratio analysis of extracted ion chromatograms (XICs/ EICs) from the light and heavy labelled MS data. Accurate and precise quantitation relies very much on the selectivity by which the XICs can be constructed. The output of MsX-quant provides a number of important figures of merit to check the reliability of quantitation. Furthermore, various visualization tools are available so that users can conveniently inspect quantitation results.

To enhance selectivity and sensitivity MsX-Quant also operates on deconvoluted MS data sets (using the Thermo Extract program). Besides a much better selectivity compared to the raw data, all individual charge states (1+, 2+, 3+ etc. ) and all $^{13}C$ isotopes will add up into one MS signal (MH+ or M), thereby increasing sensitivity. Furthermore, quantitation will be much faster as deconvoluted MS data sets are converted to centroided mode and thus much smaller in size.

Figure 1 provides an overview of the MsX-Quant workflow.

This document will explain in detail how to use MsX-Quant. The tutorial is based on one example data set. Below some specific information about the data file used in this document is given:

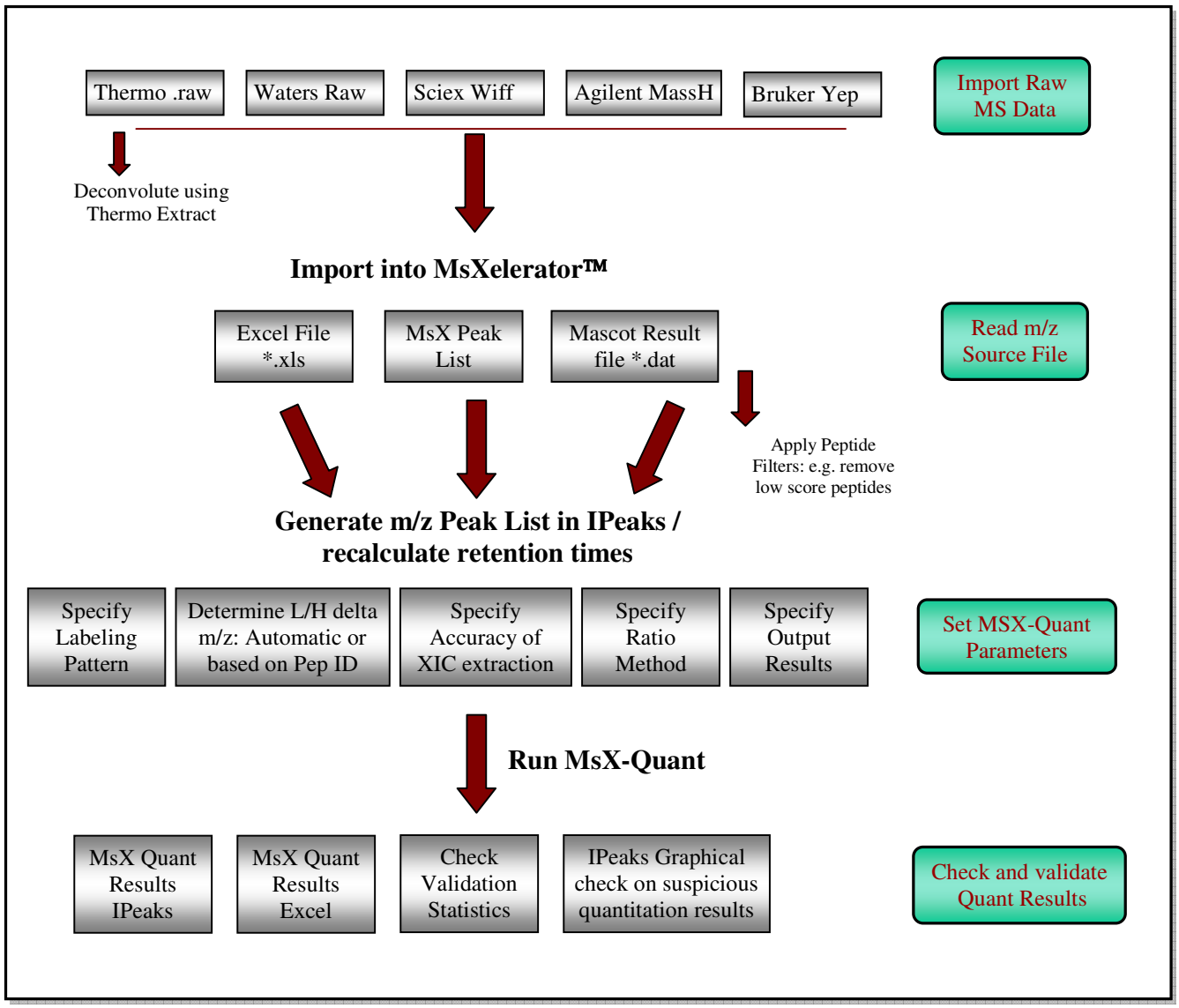



**Figure 1: MsX-Quant Overview**

**Data File:** As an example, a deconvoluted data file from the Orbitrap was used: ot2519_xtract.raw. This file was deconvoluted before using Thermo Xtract software. During the deconvolution process different parameters can be set. For this data file deconvolution to MH+ values was selected.

**m/z Source File:** an Excel m/z source file (MS2-lijst hoge betrouwbaarheid_19-04-2010.xls) will be used to read in all necessary information regarding the peptide IDs, MH+ values, retention times, charge states and possible additional information regarding protein ID.

**Labeling Pattern:** a dimethylation reaction was applied on the protein level. This procedure will add a dimethyl group (unlabelled +28.0313 (C2H6) or labelled 32.0564 (C2H2D4)) to all free NH2 groups (Lysine, but also the N-terminus). The additional N-terminus label is an option that can be specified in MsX-Quant software. After digestion using trypsine, the applied chromatographic method will only retain the N-terminal peptides. The mass difference between the light and heavy isotope doublet will be a multiple

of 4.0251 Da, depending on the number of lysines present in the peptide. To this value we will need to add 4.0251 Da, due to the labeling on the N-terminus position.

**Modifications:** Peptide sequences are specified in the Excel m/z source file. In this example the Excel file is generated based on BioWorks output. MsX-Quant will read the source file and calculate the theoretical MH+ values based on the input sequence and the specified modifications. Modifications are assigned based on fixed letter codes. Currently the following variable and fixed modifications/codes are used:

| **Modification** | **Mass** | **Code:** |
|---|---|---|
| Variable: Oxidation of Methionine | +15.9949 | M* |
| Variable: Deamidation of Asparagine | +0.9840 | N# |
| Fixed: Carbamidomethylation of Cysteine | +57.0214 | C |
| Fixed: Dimethylation on Lysine | +28.0313 | L (light isotope) |
| Fixed: Dimethylation on N-terminus | +28.0313 | Light isotope |

# 2. Importing raw MS data into MsXelerator:

Start MsXelerator and import the sample by selecting: **Menu > File > Import > Xcalibur Raw**. Browse to the raw file and select Open. The Xcalibur raw file will be converted to MsXelerator format. After conversion the sample will be loaded into the Browser (Figure 2). Conversion only has to be done once, the next time the converted file can be loaded directly into the Browser (Menu > File > Open, or load the file using the file history list.

For a general introduction of the Browser and IPeaks please read the appropriate chapters from the manual. The manual can be found in the installation directory or opened from the Browser (Menu > Help > View User Manual).

Figure 2 displays the sample after it has been loaded into the Browser. At the bottom the TIC is plotted, the top window displays the mass spectrum at a selected retention time. The middle plot shows the extracted ion current (EIC) at the selected m/z value. Clicking on any position in the TIC window will extract the mass spectrum. The EIC of the m/z value having the highest intensity in the mass spectrum will be extracted. To automatically zoom in on the m/z range of interest, click on the MS auto zoom icon.

You can set the MS auto zoom range by selecting: Options > MassSpec Zoom Range. Also be sure that raw MS data are displayed.

## Plotting Exact EICs and extracting basic information using the Browser:

As a default, nominal EICs will be extracted and displayed in the Browser. There are different ways to extract EICs at any resolution.

1. One option is to zoom in on the mass peak. Next, select from the menu: Task > Exact Mass Chromatogram Plot. A cursor will be shown. Select the m/z range for extraction graphically. The limits for extraction can be changed after being displayed. Pressing the OK button will create the plot.

2. To automatically extract high resolution EICs use the following procedure: First specify the mass resolution window for extraction: Menu > Options > Set EIC accuracy. The accuracy or window must be set in Da. Default a value of 0.035 will be suggested. Next select: Menu > Options > Use Accurate EIC plotting. If you now click near a certain m/z value in the mass spectrum the High Resolution EIC will be extracted and plotted in the middle window. To overlay two or more accurate EICs, hold **the CtrL key** while clicking (near) any mass peak.

3. Accurate EICs can also be plotted by entering an exact m/z value in the Select m/z edit box (controls on the right).

**Attention:** raw data files can be extremely large (one gigabyte ore more). When accurate EIC extraction is requested using the full retention time range, the extraction might take some time. In these cases use the following procedure to make extraction much faster. First zoom in on the region of the peak of interest in the middle window. Now you can proceed using the above accurate EIC extraction. The EIC will only be extracted using the displayed time range. This procedure is not necessary for data that have been measured or converted to centroided mode, as these files will be relatively small. An overview of the different file sizes to compare between profile and centroided is given below.

| | | |
|---|---|---|
| File Sizes: | Thermo raw profile: | 1.1 Gigabyte |
| | Raw file converted to mzXML: | 2.8 Gigabyte |
| | MsXelerator converted raw file: | 0.9 Gigabyte |
| | | |
| | Thermo raw centroided: | 10.2 Megabyte |
| | Centroided mzXML file: | 5.7 Megabyte |
| | Centroided MsXelerator file: | 2.5 Megabyte |

Figure 2 shows the EIC overlay plot of a light and heavy isotope pair, m/z 2009.015 and m/z 2013.036. For all applications of MsXelerator it will be important to have an estimate of the width of your peaks. The width near the baseline is about 1 minute (22 scans). MsX-Quant needs to determine the retention time of the top of the peak. When the input m/z source file contains retention time information (based on MSMS scans), MsX-Quant will use these retention time as a first approximation. The exact retention time is determined based on this approximate position and the width of the peak.

As can be seen from the plot, the EICs perfectly co-elute. However, dimethyl labeling using deuterium can cause retention time shifts between the light and heavy isotope peaks. This is shown in figure 3 using another light/heavy isotope pair. The heavy isotope peak elutes before the light isotope. The time difference is about 0.2 minutes. For quantitation purposes the non co-eluting property is important information, as some quantitation methods will give wrong results, e.g. if a ratio analysis would be used based on a single scan . In that case the ratio (light/heavy) would be larger than 1. However the EIC based ratio would clearly give a ratio larger smaller than one.

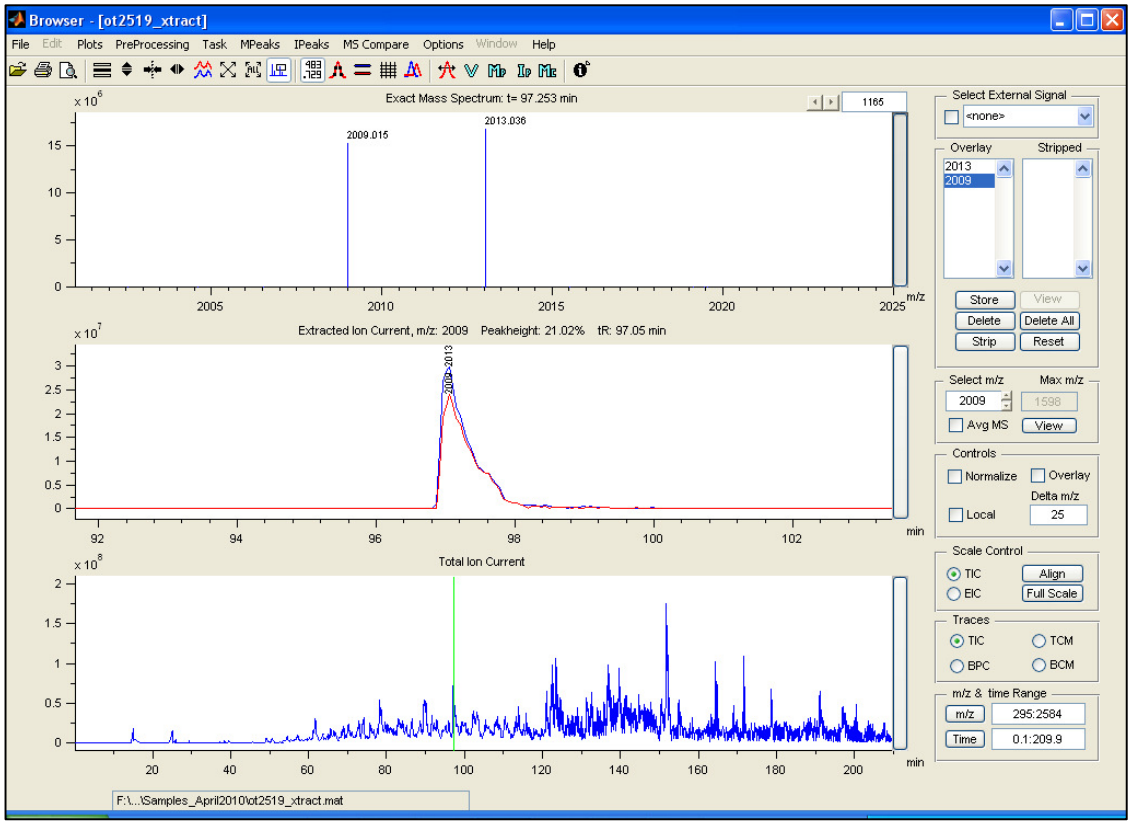
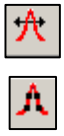



**Figure 2: Sample overview in the Browser. Shown is an EIC overlay plot of the light and heavy isotope pair, m/z 2009.015 and m/z 2013.036. Light and heavy isotope pairs perfectly co-elute, the width of both peaks near the baseline is about 1 minute (22 scans).**

**Measuring EIC peak widths:** to measure the width of a chromatographic peak you can use the Peak Width Measuring tool, available from Menu > Task > PeakWidth, or click on the peak width icon. Select Measure to manually determine the width between two time points. To make the scans visual, toggle the Line Style Icon on the icon bar.



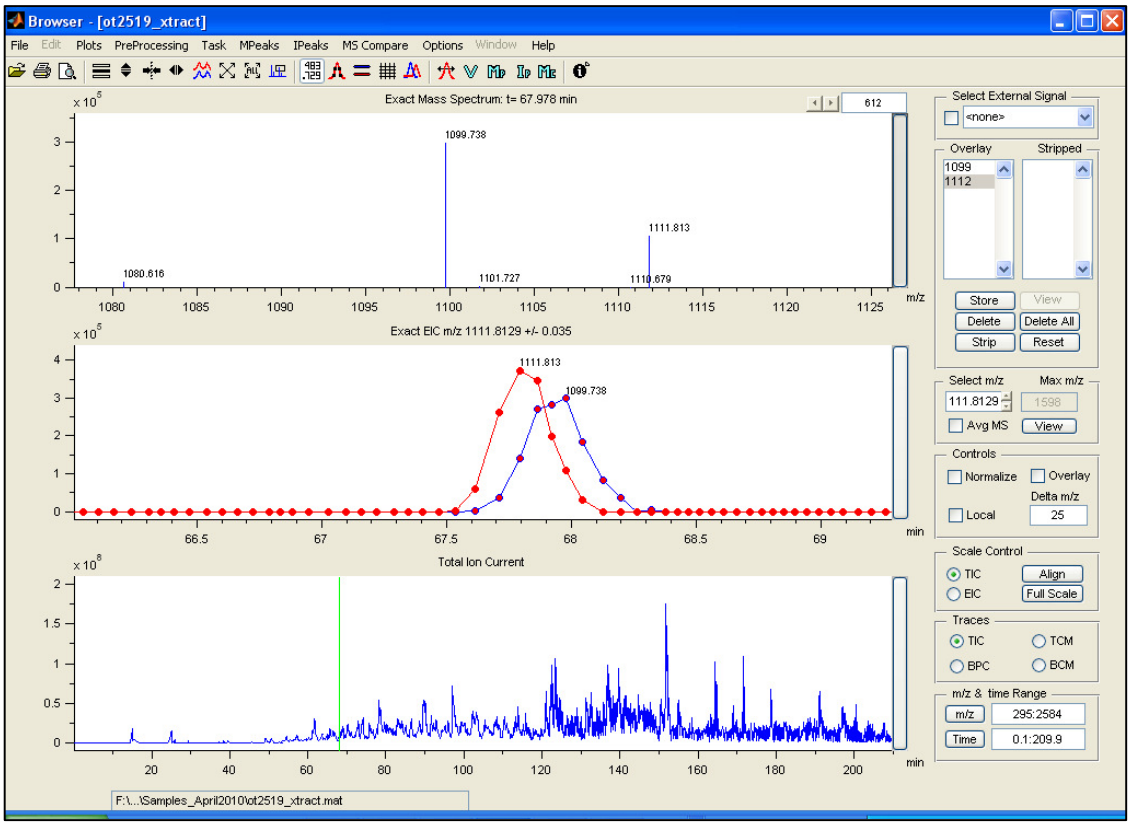



**Figure 3: Sample overview in the Browser. Shown is the EIC overlay plot of a light (blue) and heavy (red) isotope pair, m/z 1099.738 and m/z 1111.813. Light and heavy isotope pairs do not co-elute. Scans are marked.**

# 3. Loading and Conversion of m/z Source Files:

## Introduction:

An m/z source file contains basic information necessary for MsX-Quant. In general these files should contain: Peptide sequence info, information regarding the mass of the peptide (m/z, MH+ or M), retention time, and charge state. Optionally a protein ID number and reference text, e.g. protein description may be present.

Currently MsX-Quant accepts three types of input source files:

1. Excel input file
2. Mascot result file (*.dat)
3. MPeaks peak list (saved result file)

## Mpeaks Peak List:

Quantitation based on an Mpeaks peak list can be performed without sequence information. The user can apply a basic labeling pattern for quantitation. The heavy isotope doublet peak is searched based on the mass of the light peak, the calculated charge and the base labeling pattern. MsX-Quant can be run in a so-called loop. This means that when the base difference between the light and heavy pair (single labelled residue) would be 4.0251 Da, the software can search multiples of this mass difference e.g. 4.02, 8.04, 12.06, 16.08 etc. Quantitation results having the best match, based on the mass accuracy of heavy isotope, will be reported.

## Mascot Result File:

To be done….

## Excel input file:

Table 1 shows a representative part of the Excel input source file with the information needed for quantitation.

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | **Time(s)** | **Reference** | **Protein ID** | **MH+** | **z** | **Peptide** |
| 2 | 80.40 - 80.45 | D-lactate dehydrogenase [Neisseria meningitidis MC58] | NMB0001 | 889.50966 | 2 | M.SASQLLSR.L |
| 3 | 76.35 - 76.41 | transketolase [Neisseria meningitidis MC58] | NMB0002 | 900.52564 | 2 | M.SQLANAIR.F |
| 4 | 147.19 - 147.36 | phosphoglyceromutase [Neisseria meningitidis MC58] | NMB0003 | 935.53779 | 2 | -.MELVFIR.H |
| 5 | 114.11 - 114.18 | DNA polymerase III subunit beta [Neisseria meningitidis MC5 | NMB0004 | 1001.58071 | 2 | -.MLILQAER.D |
| 6 | 146.71 - 146.75 | phosphopyruvate hydratase [Neisseria meningitidis MC58] | NMB0005 | 1019.58791 | 2 | M.SAIVDIFAR.E |
| 7 | 127.97 - 128.10 | outer membrane protein OMP85 [Neisseria meningitidis MC5 | NMB0006 | 1035.54644 | 2 | A.DFTIQDIR.V |
| 8 | 66.81 - 66.84 | hypothetical protein NMB0900 [Neisseria meningitidis MC58 | NMB0007 | 1058.63067 | 2 | A.EN#LSVRRK.I |
| 9 | 77.31 | hypothetical protein NMB0844 [Neisseria meningitidis MC58 | NMB0008 | 1072.57405 | 3 | A.DRIGDLEAR.L |
| 10 | 98.41 - 98.45 | tyrosyl-tRNA synthetase [Neisseria meningitidis MC58] | NMB0009 | 1073.59445 | 2 | M.SVIQDLQSR.G |
| 11 | 89.76 - 89.80 | 50S ribosomal protein L25/general stress protein Ctc [Neis | NMB0010 | 1094.58355 | 2 | M.TYEIQASVR.E |
| 12 | 67.90 - 67.96 | electron transfer flavoprotein, beta subunit [Neisseria meni | NMB0011 | 1099.73688 | 3 | -.MKALVAVKR.V |
| 13 | 72.03 - 72.09 | NADH dehydrogenase I chain F [Neisseria meningitidis MC5 | NMB0012 | 1102.59803 | 2 | S.QTDVIDEVK.T |
| 14 | 93.38 - 93.43 | peptide chain release factor 3 [Neisseria meningitidis MC58 | NMB0013 | 1115.60501 | 2 | M.SQEILDQVR.R |
| 15 | 92.57 - 92.61 | transcription elongation factor [Neisseria meningitidis MC58 | NMB0014 | 1141.71156 | 3 | -.MQKIPLTVR.G |
| 16 | 107.93 - 107.98 | single-stranded binding protein [Neisseria meningitidis MC5 | NMB0015 | 1168.77660 | 3 | M.SLNKVILIGR.L |

**Table 1: Example of Excel source file**

The Excel input file requires a single line of text used for field descriptions. Actual data should start on line 2. Rows for all columns should be equal in size (no missing data allowed). The fields should always appear in the order given below.

- **Column A, Time(s):** Retention time. Retention time should be in minutes and is generally based on a search using MSMS data. Time can be specified as a single value (87.33) or using a range (122.33 – 122.99). The character '-' must separate both time values.
- **Column B, Reference:** description of the protein
- **Column C, Protein ID:** protein ID number
- **Column D, MH+:** mass value; MH+ or m/z

- **Column E, z:** Charge state
- **Column F, Peptide:** peptide sequence (modifications allowed, see page 3)

## Starting IPeaks and importing the m/z source file:

From the Browser start IPeaks. To import the m/z source file select: **Menu > MsX-Quant > Select and Load m/z Source File.**

The following GUI will be displayed, see Figure 4. The window is split into separate sections.

### Section: m/z Source File

Specify the type of input file using the radio buttons. For the demo select the **Excel** file type. When set, press the **Select Source File** button. You can browse to the Excel file belonging to your data file. The source file name will be displayed in the Source File edit box.

### Section: Batch Processing

Batch processing is currently not implemented but reserved for future expansion of MsX-Quant. In principle batch processing will use the same m/z source file, but quantitation will run for a series of samples.

### Section: Excel Precursor Ion Details

This section contains the parameters necessary to convert the source file into an IPeaks Peak List. The following information must be entered:

**Deconvoluted MS file:** mark the Xcalibur deconvolution check box if the raw file was deconvoluted using the Xtract program. If the original raw file will be used please uncheck.

**MH+ or M/Z mass values:** the peptide mass values in the Excel file can be either MH+ values or m/z values. When deconvoluted MS data are used, mass values in the Excel file **must be MH+.**
If the original raw data file is used, you may choose from MH+ or m/z values, whatever is specified in the Excel file.

**Re-Calc Retention Time:** this parameter will always be activated. During m/z source file conversion, the retention times of all peaks will be re-estimated. Different procedures are available.

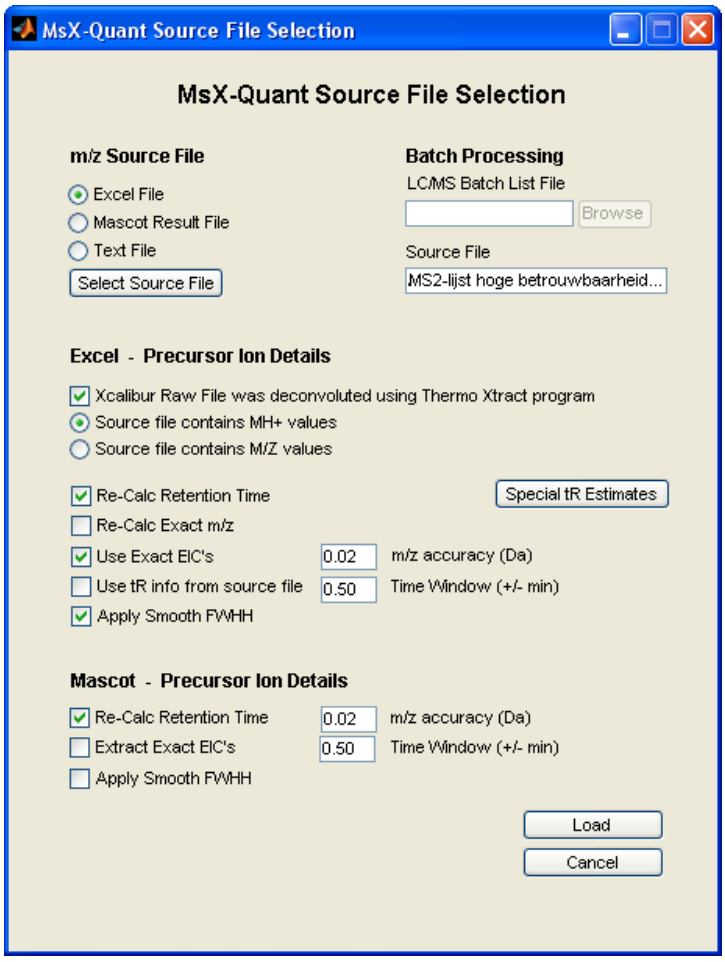



**Figure 4: MsX-Quant Source File GUI**

**Use Exact EICs:** determination of the retention times can be done using nominal extracted EICs or exact (accurate) EICs. It is advised to use Exact EICs for deconvoluted data, since this is fast enough. For raw data based on Mascot, it is better to use nominal EICs. After importing Mascot results you will need to do

some filtering to remove low scoring peptides or remove double (redundant) entries. This will significantly shorten all processing time as the number of peptides may be reduced by 50% or more. Exact EIC extraction will always give a better estimation of the top of the chromatographic peak.

**m/z accuracy (Da):** specify the accuracy for EIC extraction. The default will be a window of 0.02 Da.

**Use tR info from source file:** the source file must always contain retention time information. However, it doesn't have to be used. The retention time information is based on MSMS data and was used to identify the peptides. When the check box is marked, the retention time information from the Excel file will be used to locate the peak in the data. The time window (+/- 0.5 minutes) is used to search for the exact top of the chromatographic peak.
In the case of raw MS data files you will have to use the tR info. In the extracted accurate EIC of a raw MS data file, probably more than one chromatographic peak will be present having equal mass (+/- the specified accuracy). If no retention time info was given, it would be hard to find the correct peak.

When no retention time information is used, the program will extract the EIC over the full time range. The peak having the largest intensity will be used for quantitation. This procedure can be used for deconvoluted data since deconvoluted EICs are in general very selective. Due to the small file size of centroided data, extraction of all accurate EICs over the full time region will not be a problem.

**Time Window (+/- min):** use this window around the retention time (specified in the source file) to extract and re-estimate the retention time for each individual peak.

**Apply Smooth FWHH:** during EIC extraction, the peak's Full Width at Halve Height will be calculated. This value will be used to determine the starting and ending positions for peak area ratio calculation. Small signals can have noisy peaks. Using this option will apply a slight smoothing to the extracted EIC in order to better estimate the FWHH.

**Special tR estimates:** currently this procedure is not in use.

## Section: Mascot Precursor Ion Details

See the explanation using Excel import, the same items apply.

When pressing the **Load button**, the conversion will start. The following calculations will be performed:

- The Excel file will be imported
- All mass values will be read (m/z or MH+). Based on the sequence and the modifications a theoretical MH+ value will be calculated
- The nominal or accurate EIC will be extracted based on the parameters specified
- The retention time at maximum intensity will be calculated
- The maximum intensity and FWHH will be determined
- The calculated peak information together with peptide and protein info will be converted to an IPeaks table and displayed, see Figure 5.

Attention, the IPeaks table is not automatically saved. To run different quantitation methods you will only need to import the m/z source file once. Save this file before running MsX-Quant. Menu > File > Save Table, or press the Save Icon on the Toolbar.

Using IPeaks, you can view Mass Chromatograms or Mass spectra for each of the entries in the table. In principle you can use all IPeaks features like sorting, deleting, etc., see the IPeaks chapter in the manual. The IPeaks table after source file conversion will consist of the following fields:

| **IPeaks Field Name** | **Description** |
|---|---|
| Peak or Pair: | peak number |
| m/z: | m/z value (or MH+ value). This value is directly copied from the Excel Source File. This is not the measured m/z value from the MS data!! After completion of MsX-Quant (next step), this value will be replaced by the actual measure value from the MS data file. |

| | |
|---|---|
| Ratio: | all zeros, to be determined using MsX-Quant |
| tR: | recalculated retention time of peak |
| PH: | peak height at calculated retention time |
| % PH: | relative peak height compared to largest peak in list |
| % Area: | not determined during conversion, all set to value of 10.0 |
| Text info: | info extracted from Excel file: Calculated MH+ value (from MsXelerator), charge, peptide sequence, protein ID number and Protein description from Source file |

Additional information can be made visible by clicking on the Toggle Table Info Icon on the Icon Toolbar (far right). Additionally, the following fields will be displayed.

| | |
|---|---|
| Pbase: | not used, all set to value of 10.0 (estimate of number of scans near baseline > peak width) |
| FWHH: | full width at halve height in scans |
| Smooth: | not used, all set to value of 1.0 |
| S/N ratio: | not used, all set to value of 5.0 |

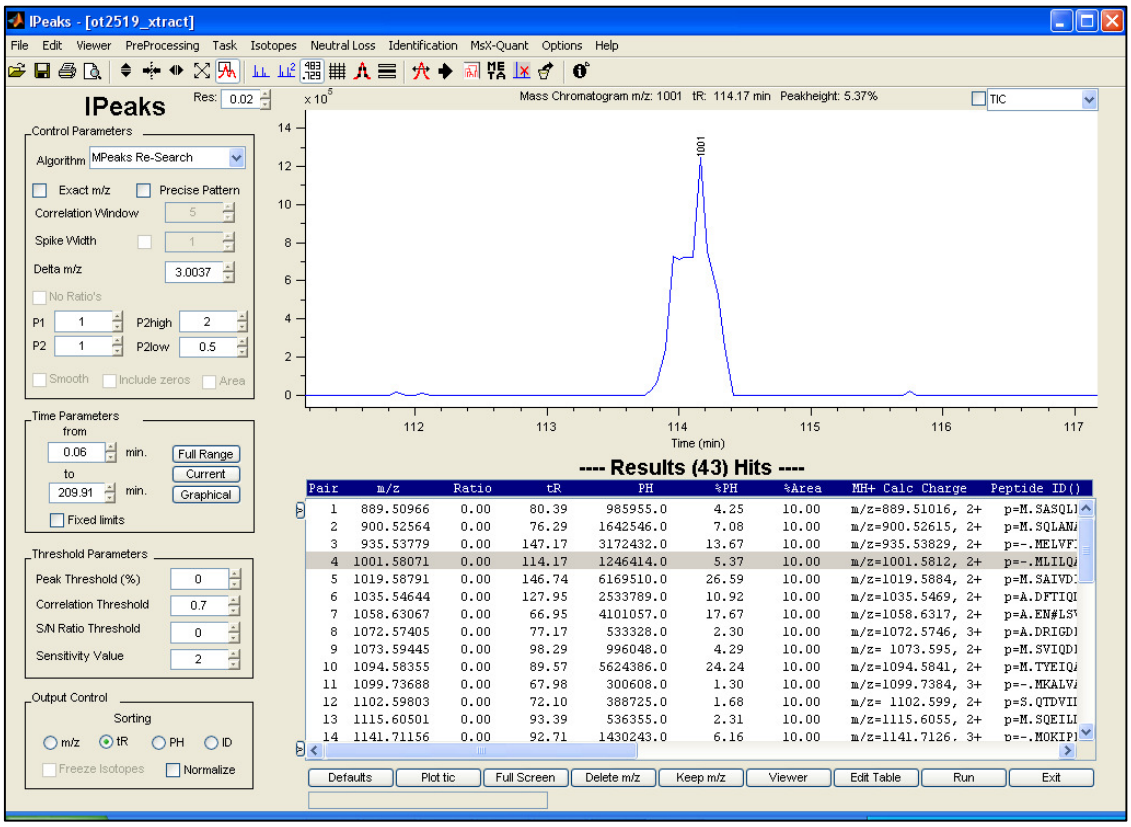



**Figure 5: IPeaks Overview after conversion and import of m/z source file**

## Checking some of the basic conversion parameters in IPeaks:

After conversion we can check if one of the more important parameters - the mass accuracy used during EIC extraction – is correct. Depending on the bias and precision of the instrument, different situations can occur.

In case that a bias exists between theoretical mass and measured mass, using a too narrow mass window might miss the peak. Using a too narrow window on less precise peaks (but no bias) would give a choppy/noisy EIC. Some scans will have no intensity, while others are correct. A too large window means less selectivity, interfering ions could influence EIC extraction which can give incorrect quantitative results.

Some of the entries in the table should be checked regarding the above issue. From IPeaks, activate accurate EIC plotting: Menu > Options > Plot Exact Mass Chromatogram. The mass accuracy can be set from: Menu > Options > Set Mass Accuracy or by entering the value in the Resolution edit box (next to the IPeaks Logo). The have a better zoomed view of the EIC, activate Auto zooming. The EIC will automatically be zoomed using a smaller time region.

Figure 6 shows what happens if the EIC is extracted using a too narrow window. On the left the EIC of m/z 1073.5944 is extracted using a window of 0.02 Dalton. Displayed is the time range 95-101 minutes. On the right the EIC of the same peak is shown but now using a window of 0.009 Da. At some scans no signal is present due to a too narrow mass window. Also the width (FWHH) will be underestimated. Smoothing will only partly correct for this. The area of the peak on the right is underestimated due to missing data. As a consequence, the ratio determination will be incorrect.

The effect of a too large mass window is shown in Figure 7. On the left the EIC is displayed of m/z 1670.975 at a resolution of 0.02. On the right the window is increased to 0.08 Da. An additional peak appears to the left of the peak. The area will be too large and an incorrect ratio will be calculated. For MS data in profile mode and not deconvoluted, the chance of interfering peaks will be even larger

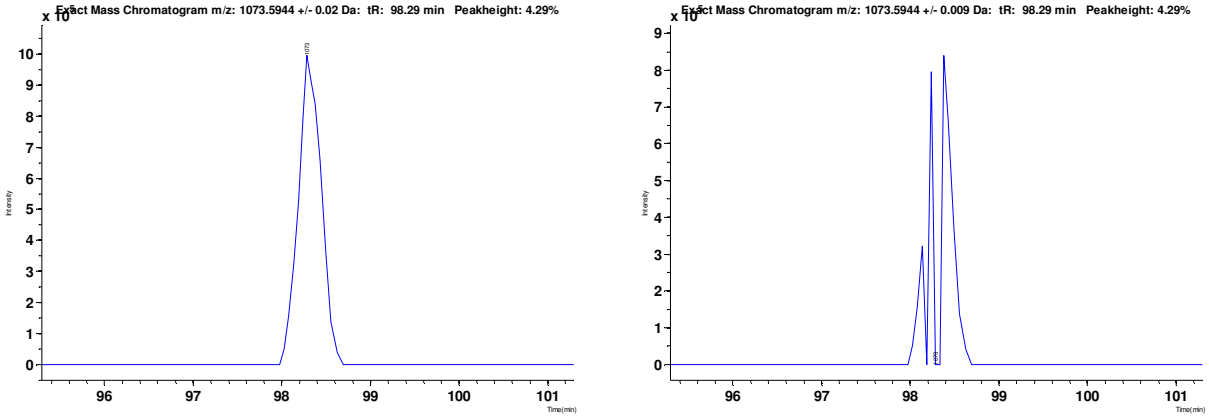



**Figure 6: Example of correct and incorrect mass window for accurate EIC extraction. On the left the EIC of m/z 1073.5944 is extracted using a window of 0.02 Dalton. On the right the EIC of the same peak is shown but now using a narrow window of 0.009 Da. Part of the signal is lost.**

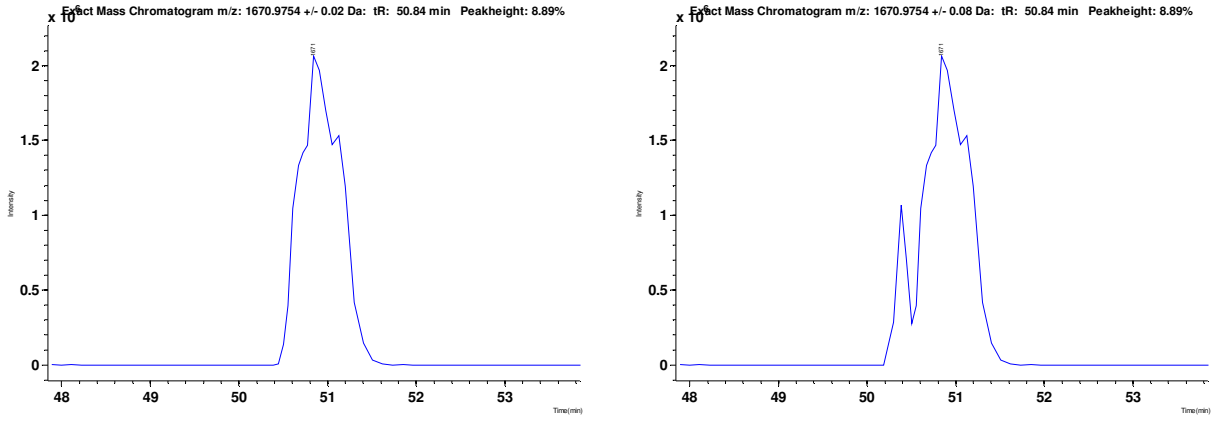



**Figure 7: Example of correct and incorrect mass window for accurate EIC extraction. On the left the EIC of m/z 1670.975 is extracted using a window of 0.02 Dalton. On the right the EIC of the same**

**peak is shown but now using a too wide window of 0.08 Da. On the left of the peak an additional signal appears with m/z 1671.051**

To investigate mass chromatograms and mass spectra in more detail, the selected peak can be copied to the Browser. The Browser offers more flexibility to check peaks interactively (Checking MS spectra on different positions of the peak). To copy the peak from IPeaks to the Browser, use the following procedure:

Press the Browser icon on the Icon Toolbar, or press the 'b' key on the keyboard. The EIC will be copied to the Browser and the mass spectrum at the peak's retention time will be plotted in the Browser EIC window. Zoom in around the EIC. To better view the exact MS data, press Auto zoom MS and also check the exact MS plotting Icon. You can set the MS range displayed from the options menu.

# 4. Setting MsX-Quant Parameters:

After conversion and loading the m/z source file into IPeaks, you are ready to set and run MsX-Quant. Before running MsX-Quant consider saving the raw IPeaks table as a base table for different Quant runs. To start MsX-Quant select from the Menu: MsX-Quant > Run MsX-Quant. The following Graphical Interface will be shown from which the quant parameters can be set (Figure 8).

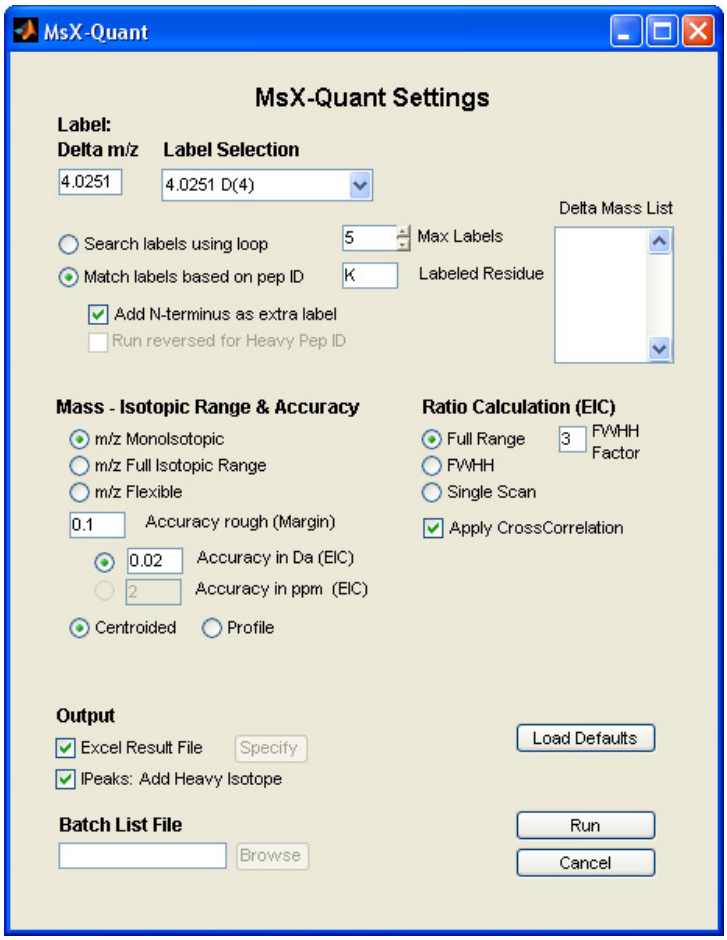



**Figure 8: MsX-Quant Parameter Settings Interface**

Below an explanation of all parameters is given. Settings will be stored, so running MsX-Quant will load the previously used set of parameters.

### *Mass Difference Settings:*

**Label Delta m/z:**
This is the difference in mass between the light and heavy isotope for a single residue and charge 1+. The actual mass difference will be calculated from the number of labelled amino acid residues and the charge of the peptide. The mass difference shown is based on dimethyl labeling in which 4 hydrogen atoms are replaced by 4 deuterium atoms.

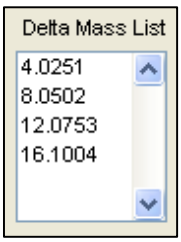
**Label Selection:**
Some predefined labeling patterns can be selected from the list box. Currently available are:

| | |
|---|---|
| 4.0251 | D4 |
| 8.0142 | 13C(6)15N(2) |
| 14N/15N | Number of labelled positions depends on the number of N atoms in the peptide |

**Search Labels using loop:**
The mass difference between light and heavy isotope is depends on the base label mass difference, the charge of the peptide and on the number of labelled residues. If the peptide sequence is not known, e.g. when using Mpeaks source file as input, we do not know the number of residues. In this case MsX-Quant can run the base mass difference in a loop. As an example, if the base mass difference is 4.01, and of loop of 5 iterations is defined, MsX-Quant will search for mass differences of 4.01, 8.02, 12.03, 16.04 and 20.05, relative to the charge of the monoisotopic peak. For each of the mass difference the quantitation will be performed and the one with the best match will be reported. The best match is defined as the iteration number for which the difference in mass between light and heavy isotope peak is smallest.

**Max Labels:**
Set the maximum number of iterations to use when using loop calculations.

**Delta Mass List:**
This edit box will show the mass differences that will be calculated when running MsX-Quant in a loop. This list can be edited to set user defined mass differences. One can use both positive and negative values.




**Match labels based on Pep ID:**
The second and probably preferred choice is to determine the number of labelled residues from the peptide sequence. Select this option to run a single quantitation.

**Labelled Residue:**
Enter the residue letter code used for labeling, e.g. K for Lysine. Currently only one residue can be entered. The next version of MsX-Quant will allow for multiple labelled residues.

**Add N-terminus as extra label:**
If checked the algorithm will add one additional label for the N-terminus. In this example labels were introduced used a dimethylation reaction. This will also label the N-terminus group.

**Run reversed for Heavy Pep ID:**
This is a special case when using quantitation based on a Mascot result file. Some of the identified peptides will relate to the heavy isotope. This is possible when allowing variable modifications in the ID search. If the label was specified as a modification, MsX-Quant will recognize it from the modification code in the Mascot result file. A heavy peptide can subsequently be quantified in reversed mode (searching for the light isotope).

### *Mass – Isotopic Range and Accuracy Settings:*

- **m/z Monoisotopic**
- **m/z Full Isotopic Range**
- **m/z Flexible**

These choices define how the areas based on EICs are extracted.
The default value will be m/z monoisotopic. The area of the monoisotopic mass peak will be extracted and the ratio will be calculated between Light and Heavy isotope.

**m/z full isotopic range** will include the $^{13}C$ isotope peaks. Select this option for higher charged peptides. In these cases ,the monoisotopic peak might only represent a fraction of the total signal. A ratio analysis based on the wider envelope will extract more of the total signal and be more precise. Do not select this option for deconvoluted data. Setting a wider range will also increase the chance of interfering ions being present.

**m/z Flexible;** the monoisotopic extraction will be used for peptides having charge 2+ and 3+, higher charged peptides will use the full isotopic range. Currently the full isotopic range will extract the EIC from one large region, e.g. 1234.12 – 1235.72. No fine tuning making use of split region (depending on charge state) is currently available. This will be introduced in the next version of MsX-Quant.

**Accuracy Rough:**
The specified window defines the mass range in which the **heavy isotope peak** should be present. If no signal is found in the specified region, MsX-Quant will skip the quantitation for this entry. The default value is 0.1 Da, however for RAW data please specify a smaller value, e.g. 0.05. Interfering and larger ions compared to the heavy isotope, can mistakenly be used as the heavy isotope peak.

**Accuracy for EIC extraction:**
**Value in Da or ppm**
This value defines the mass window for accurate EIC extraction in absolute units (Da) or in ppm. See the previous section on accurate EIC extraction for more information. Currently, in this version only the absolute accuracy in Da is available.

**Mass Mode:**
**Centroided or Profile**
Select the type of MS data (centroided or profile) belonging to the LC/MS data file.

### *Ratio Calculation Method (EIC):*

- **Full Range**
- **FWHH**
- **Single Scan**

The selected option defines how the ratio calculation is performed. The area from the light isotope peak can be calculated based on the Full Range, from the middle part of the peak (FWHH) or using a single scan at the retention time maximum. For the heavy isotope the same time region is used. As a default we use the Full Range. In this case N*FWHH is used to set the start and end of a peak. N is defined by the FWHH Factor and is typically set to a value of 3. For non co-eluting peaks you might need to use a value larger than 3, to include the full area for the heavy isotope.

When selecting FWHH, only the middle part of the peak (light isotope) will be used for quantitation. For non co-eluting peaks you should not use this option as the heavy peak will not be extracted correctly.

When selecting Single Scan, the retention time max for the light isotope peak will be used for quantitation. Basically this will be a Peak Height Ratio Calculation at a fixed time position. This method is only to be used for perfectly co-eluting peaks (very fast).

**FWHH Factor:**
This value sets the integration limits for peak area extraction. As a default the value 3 will be used (see above). Non co-eluting peaks might need a larger value, while for perfectly co-eluting peaks a smaller value will do. A large value will increase the chance to include interfering ions. A too small value will be less precise, as part of the signal is not used.

**Apply Cross Correlation:**
One of the validation parameters calculated during quantitation is the Correlation Coefficient (CC) between the EICs of both the light and heavy isotope peak. The CC is a measure of the similarity of both shapes and is size independent. Perfectly co-eluting peaks should have a CC of 0.9-1.0, while shifted peaks or very noisy peaks will have lower CC's.
Lower CC values for peaks having large ratio differences are normal, as one of the EICs might have a low signal. Low CC values for moderately strong peaks having ratios between 0.3 and 3.0 are suspicious and the EICs should be checked in IPeaks. To be checked are the accuracy of EIC extraction, the area time range used for quantitation and the noise level of the EICs. IPeaks has a procedure to manually calculate a ratio. You will have the opportunity to replace the calculated value by the manually determined ratio.

For labeling procedures giving rise to non co-eluting peaks (deuterium based), the CC will in many cases be quite small, because the peaks have shifted and are thus different. In these cases you should apply Cross Correlation instead of normal Correlation. The Cross Correlation Coefficient (CCC) will automatically take into account the shifting of the peaks. After shift-correction, the CC is calculated. In this case a low CC value is probably related to noisy peaks, bad peak shapes or the shifts might be very large.
The default range for cross correlation is 15 scans. In the example this converts to 0.85 minutes.

Examples how to interpret the CC and other validation parameters will be given in the next chapter.

### *Output:*

**Excel Result File:**
Check this item when results should be written to Excel. The default output name will be msx_quant_result.xls and the file can be found in the data directory where the LC/MS date file is located. You will be asked to open this file after the run has completed. The Excel result file will contain more information compared to the IPeaks output.

**IPeaks: add Heavy Isotope:**
The input table consists of the light isotope peak only. Checking this option will add the heavy isotope peak to IPeaks. This makes if possible to automatically create overlay EICs of the light and heavy doublet pairs.

**Batch List File:**
Not used. To be implemented in near future for running MsX-Quant on series of samples.

**Load Defaults:**
Press this button to load default parameters for all settings. Default settings are currently based on the example used in the demo.

**Run:**
Starts MsX-Quant execution.

**Cancel:**
Leave the GUI and return to IPeaks.

# 5. Output Results and Validation Checks:

At the end of the run you will be asked if the Excel output file should be opened. Select yes; you will be able to compare results from Excel and IPeaks simultaneously. After the run, the IPeaks result screen will be similar to the one shown in Figure 9.

**Attention:** the IPeaks result table after quantitation is not automatically saved.

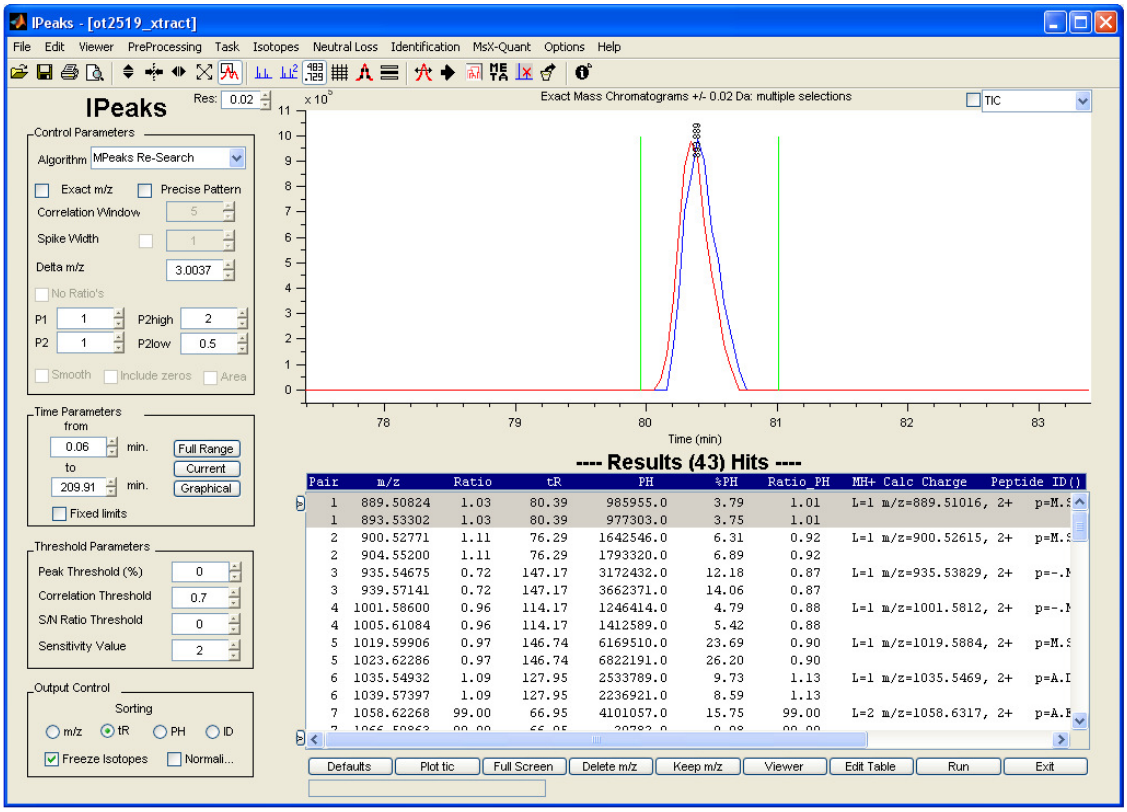



**Figure 9: IPeaks output screen after MsX-Quant run has completed**

In figure 9, the heavy isotope peaks have been added to the result table and the peaks will behave as pairs. Clicking on an entry will overlay the EICs of both peaks. If exact masschromatogram plotting was off, please turn it on. You probably want to activate Auto zooming and plot Peak Area Integration limits. To plot the limits select Menu > Options > Plot Peak Width Markers. The limits will be based on the FWHH range and the FWHH factor specified during MsX-Quant.

To switch to the mass spectrum view, press the 'm' key or select the Mass Spectrum Toggle icon from the Icon Toolbar. You will probably want to view both isotope peaks in zoom mode. If the mass range plotted (default 25 Da) is too small, you can use a wider MS range; Menu > Options > MS Zoom Range.

## IPeaks MsX-Quant Result Table:

**Recalculated m/z values:**
After the run, the m/z values have been updated. They are now based on the actual values found in the MS data and are taken from the m/z value belonging to the retention time maximum. Probably small deviations will exist between found and calculated m/z values. The Excel output will calculate the deviations in ppm

between found and calculated masses for both the light and heavy isotope. It belongs to one of the validation criteria.

You can optionally recalculate the average exact mass values across the FWHH of each peak, Menu > Task > Accurate Mass Conversion. You will be asked to specify how to calculate the exact mass, Single Scan or based on FWHH. For FWHH, an intensity weighted average m/z value will be calculated. In the case of profile data, the weighted average of the profile m/z peak is calculated first, followed by a weighted average across the chromatographic peak.

If the heavy isotope peak is not found, its table value will be set to the mass of the light isotope peak. The ratio value will be listed as Nan (Not a Number), and the CC will be set to a value of 0.00.
Figure 10 shows that one of the entries could not be quantified (entry 26). The mass spectrum is also shown. It seems that we are dealing with a heavy isotope peak. This means that the m/z value taken from the m/z source file was incorrect and relates to the heavy labelled peptide.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 26 | 1421.78723 | NaN | 131.24 | 17667584.0 | 67.84 | NaN | L=1 m/z=1421.7999, 2+ p=D.Q |
| 26 | 1421.78723 | NaN | 131.24 | 0.0 | 0.00 | NaN | |
| 27 | 1440.74683 | 0.86 | 65.83 | 2684696.0 | 10.31 | 0.81 | L=2 m/z=1440.7416, 3+ p=-.M |
| 27 | 1448.79431 | 0.86 | 65.83 | 3316668.0 | 12.74 | 0.81 | |
| 28 | 1507.79456 | 0.95 | 91.54 | 4018418.0 | 15.43 | 0.80 | L=2 m/z=1507.7937, 3+ p=-.M |
| 28 | 1515.84045 | 0.95 | 91.54 | 5008359.0 | 19.23 | 0.80 | |

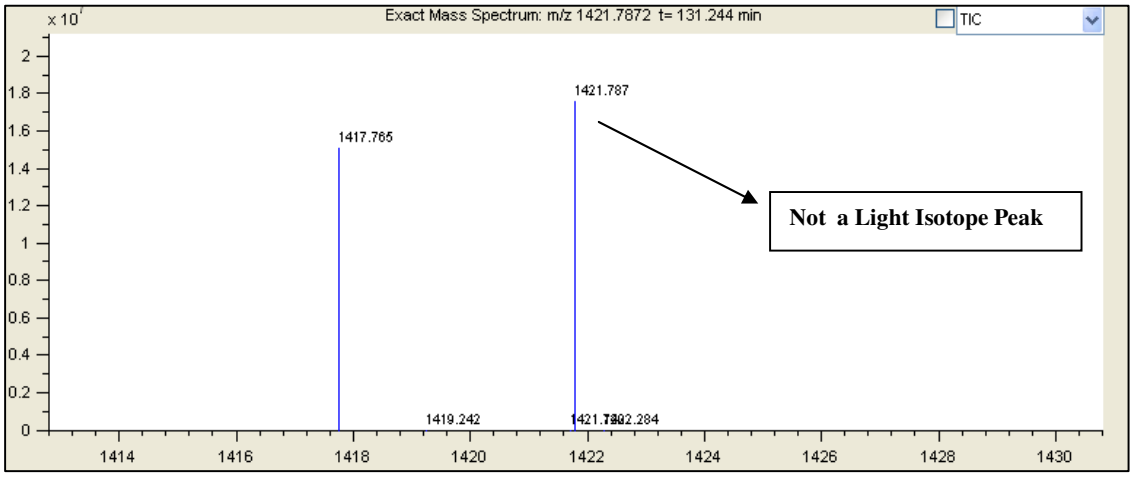



**Figure 10: Example output showing for Peak 26 (m/z 1421.787). No heavy isotope peak could be found. From the mass spectrum it is concluded that entry 26 relates to the heavy peptide and not the light isotope.**

**Area and Peak Height Ratios:**
After the MsX-Quant run, the Area Ratio will be listed in the third column. The seventh column displays the peak height max ratio. It is a simple ratio from the maximum peak height of the light and heavy isotope, calculated over the integration range. Large ratios exceeding the value 99 will be listed in the table as 99.

**Number of detected labels:**
At the beginning of the comment field the number of labels (residues) found has been added. It will be displayed as L=1, or e.g. L=2. In this example the number of residues was directly taken from the peptide sequence.

**(Cross) Correlation Coefficient:**
The CC or CCC will be displayed in column 10. However this column is not directly visible, but is hidden by the comment field. To make it visible, press the Toggle Table Info Icon on the Toolbar. 
Values below 0.9 might be suspicious and should be validated manually by examining the EICs. Figure 11 displays a pair having a CCC of 0.815. Is seems that especially the heavy isotope has an irregular shape. This will to a large extend effect the CCC.

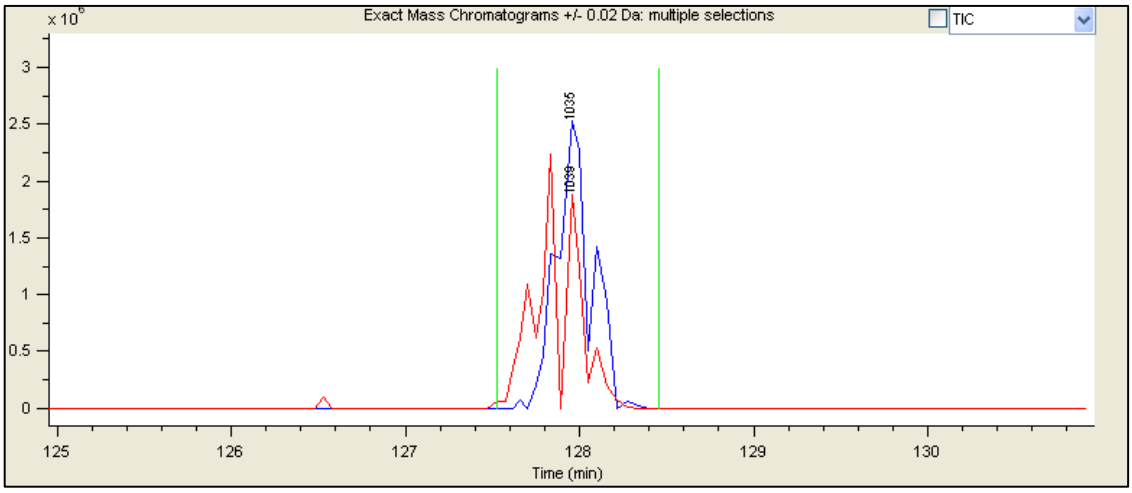



**Figure 11, example of light and heavy EIC peaks having low Correlation Coefficient (0.815). Signal ratio is comparable.**

**Manually calculating Ratios:**
If, for some reason, the calculated ratio seems incorrect, you can manually calculate it from the displayed EICs. Be sure to select the correct accuracy for plotting. To integrate manually, first zoom in on the EICs displayed.

Next select; Menu > Task > Determine Ratios /PH/ Area from Plot.

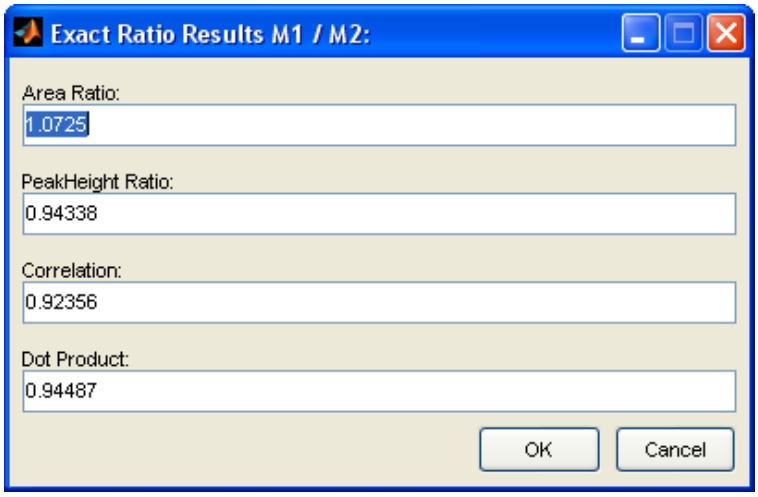



**Figure 12: Manual Peak Integration**

A crosshair cursor will be shown. Click to the left and right of the peak to set the integration time range. After that you will be asked to calculate the Ratio from M2/M1 or M1/M2. M1 relates to the first plotted peak (blue), M2 relates to the second EIC (red). Please select M1/M2 (the second option). An overview will be shown in which the ratios based on area and peak height are displayed. Also the Correlation Coefficient (normal CC) and the Dot Product are calculated as a measure of peak shape similarity (Figure 12).

**MIDAR Functions:**
Some functionality of IPeaks can be used after running MsX-Quant. These functions can be found in another part of IPeaks, but will in near future be made accessible from the MsX-Quant Menu.
MIDAR (Mass Isotope Distributions of Amino Acid Residues) functions can be found under: Menu > Isotopes > MIDAR. The following functions can also be applied to MsX-Quant output.

Delete Zero ratios from the result table
Delete Nans from the result table
Delete Low Correlated Peaks
Set Ratio Threshold
Set Ratio Threshold > exclude equal (keep only large up- or down-regulated peaks)
Sort on Ratio
Plot Smoothed Mass Chromatograms
Create Area ratio versus Peak Height Ratio Plot (Line or Scatter Plot)

To read more about MIDAR and its functions see the manual page 133-142.

**Making a copy of the Current Table:**
Any operation performed on the MsX-Quant result table will be permanent. You can reload the original table if the table was saved. Alternatively, make a copy of the table in memory. Press **Task > Store Current IPeaks Table.** You can now easily retrieve the table by selecting: **Task > Restore IPeaks Table.** Alternatively press the Restore IPeaks Table Icon on the Toolbar.

## MsX-Quant Excel Output Description:

You will have the option to write results to Excel directly after quantitation. Some additional output is written to Excel, not found in the IPeaks table. A few additional validation parameters will be calculated. Currently these are not present in the IPeaks table, but will be implemented in near future together with newly developed validation parameters.

Table 2 shows part of the Excel output from the first 31 entries.

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MsX-Quant:10-Aug-2010 Input File:ot2519_xtract.mat | | | | | | | | | | | | | | |
| No. | Sequence | ID | m/z L | m/z H | dm/z_L (ppm) | dm/z_H (ppm) | tR | N_label | N_AA+1 | Ratio_A | Ratio_H | Correl | Charge | Intensity |
| 1 | M.SASQLLSR.L | NMB0001 | 889.5082 | 893.533 | -2.16 | -0.36 | 80.391 | 1 | 1 | 1.03 | 1.01 | 0.996 | 2 | 985955 |
| 2 | M.SQLANAIR.F | NMB0002 | 900.5277 | 904.552 | 1.73 | -0.89 | 76.289 | 1 | 1 | 1.11 | 0.92 | 0.931 | 2 | 1642546 |
| 3 | -.MELVFIR.H | NMB0003 | 935.5468 | 939.5714 | 9.04 | -0.47 | 147.17 | 1 | 1 | 0.72 | 0.87 | 0.764 | 2 | 3172432 |
| 4 | -.MLILQAER.D | NMB0004 | 1001.586 | 1005.611 | 4.79 | -0.26 | 114.165 | 1 | 1 | 0.96 | 0.88 | 0.859 | 2 | 1246414 |
| 5 | M.SAIVDIFAR.E | NMB0005 | 1019.599 | 1023.623 | 10.46 | -1.27 | 146.736 | 1 | 1 | 0.97 | 0.9 | 0.861 | 2 | 6169510 |
| 6 | A.DFTIQDIR.V | NMB0006 | 1035.549 | 1039.574 | 2.34 | -0.43 | 127.954 | 1 | 1 | 1.09 | 1.13 | 0.815 | 2 | 2533789 |
| 7 | A.EN#LSVRRK.I | NMB0007 | 1058.623 | 1066.599 | -8.52 | -69.61 | 66.95 | 2 | 2 | >99 | >99 | 0.917 | 2 | 4101057 |
| 8 | A.DRIGDLEAR.L | NMB0008 | 1072.574 | 1076.598 | -0.47 | -0.75 | 77.174 | 1 | 1 | 0.86 | 0.71 | 0.983 | 3 | 533328 |
| 9 | M.SVIQDLQSR.G | NMB0009 | 1073.604 | 1077.628 | 8.16 | -0.98 | 98.286 | 1 | 1 | 1.03 | 1.03 | 0.99 | 2 | 996048 |
| 10 | M.TYEIQASVR.E | NMB0010 | 1094.58 | 1098.603 | -3.9 | -1.84 | 89.567 | 1 | 1 | 0.79 | 0.63 | 0.985 | 2 | 5624386 |
| 11 | -.MKALVAVKR.V | NMB0011 | 1099.738 | 1111.813 | -0.77 | 0.01 | 67.978 | 3 | 3 | 0.97 | 0.81 | 0.97 | 3 | 300608 |
| 12 | S.QTDVIDEVK.T | NMB0012 | 1102.594 | 1110.559 | -4.43 | -76.42 | 72.103 | 2 | 2 | 4.15 | 4.12 | 0.891 | 2 | 388725 |
| 13 | M.SQEILDQVR.R | NMB0013 | 1115.601 | 1119.626 | -4.4 | -0.07 | 93.391 | 1 | 1 | 1.07 | 1.07 | 0.995 | 2 | 536355 |
| 14 | -.MQKIPLTVR.G | NMB0014 | 1141.716 | 1149.764 | 2.82 | -1.83 | 92.715 | 2 | 2 | 0.82 | 0.68 | 0.967 | 3 | 1430243 |
| 15 | M.SLNKVILIGR.I | NMB0015 | 1168.784 | 1176.832 | 5.42 | -2.21 | 107.769 | 2 | 2 | 0.89 | 0.81 | 0.968 | 3 | 2360986 |
| 16 | M.GQKINPTGFR.I | NMB0016 | 1173.675 | 1181.721 | 1.18 | -3.64 | 72.824 | 2 | 2 | 0.93 | 0.63 | 0.937 | 2 | 4078155 |
| 17 | M.TTLHFSGFPR.V | NMB0017 | 1190.644 | 1194.666 | 10.47 | -3.13 | 115.215 | 1 | 1 | 1.08 | 0.93 | 0.978 | 3 | 3312515 |
| 18 | M.SIKVAINGFGR. | NMB0018 | 1217.73 | 1225.779 | -5.05 | -1.42 | 107.096 | 2 | 2 | 0.91 | 0.86 | 0.928 | 3 | 1656405 |
| 19 | A.AGVHVEDGWAR. | NMB0019 | 1224.611 | 1228.632 | -0.95 | -2.94 | 86.486 | 1 | 1 | 0.88 | 0.82 | 0.981 | 3 | 3528085 |
| 20 | M.ARKTPISLYR.K | NMB0020 | 1260.776 | 1268.826 | -1.75 | -0.79 | 73.567 | 2 | 2 | 0.95 | 1.05 | 0.974 | 3 | 1341205 |
| 21 | M.AIVKMKPTSAGF | NMB0021 | 1342.827 | 1354.898 | 2.14 | -3.32 | 57.438 | 3 | 3 | 0.78 | 0.56 | 0.93 | 2 | 2869298 |
| 22 | -.MNGKYYYGTGR. | NMB0022 | 1365.663 | 1373.71 | 1.06 | -2.51 | 70.644 | 2 | 2 | 0.87 | 0.75 | 0.932 | 3 | 1999221 |
| 23 | M.AKKISILVGSLP | NMB0023 | 1368.929 | 1381.002 | -0.66 | -1.84 | 100.944 | 3 | 3 | 0.74 | 0.54 | 0.805 | 3 | 1588484 |
| 24 | M.AAKDVQFGNEVF | NMB0024 | 1389.75 | 1397.796 | 1.43 | -3.52 | 74.173 | 2 | 2 | 0.94 | 0.71 | 0.908 | 2 | 7413490 |
| 25 | M.SVTVETLENLEE | NMB0025 | 1417.766 | 1421.788 | 8.61 | -1.6 | 131.315 | 1 | 1 | 0.96 | 0.89 | 0.944 | 2 | 15654912 |
| 26 | D.QYVVLLRN#TSV | NMB0026 | 1421.787 | NA | -8.91 | NA | 131.244 | 1 | 1 | NA | NA | 0 | 2 | 17667584 |
| 27 | -.MKPNIHPDNYR. | NMB0027 | 1440.747 | 1448.794 | 3.63 | -1.88 | 65.834 | 2 | 2 | 0.86 | 0.81 | 0.95 | 3 | 2684696 |
| 28 | -.MQNSTTEFLKPE | NMB0028 | 1507.795 | 1515.84 | 0.57 | -2.84 | 91.54 | 2 | 2 | 0.95 | 0.8 | 0.975 | 3 | 4018418 |
| 29 | M.SETENQALTFAK | NMB0029 | 1550.817 | 1558.864 | -0.42 | -1.74 | 86.926 | 2 | 2 | 0.9 | 0.73 | 0.979 | 2 | 2747204 |
| 30 | M.ADNYVIWFENLE | NMB0030 | 1567.806 | 1571.831 | 9.97 | -0.05 | 176.875 | 1 | 1 | 0.99 | 0.91 | 0.978 | 2 | 2920450 |
| 31 | A.DVSLYGHIKAGV | NMB0031 | 1648.907 | 1656.954 | 10.07 | -1.93 | 137.073 | 2 | 2 | 0.78 | 0.89 | 0.89 | 2 | 23205120 |

**Table 2: Excel output from MsX-Quant**

Below a description is given for all columns in the Table:

- **Column A: No.:** Peak entry number
- **Column B: Sequence:** Peptide sequence
- **Column C: ID:** Peptide ID number
- **Column D: m/z L:** m/z value of light isotope (based on MS data)
- **Column E: m/z H:** m/z value of heavy isotope (based on MS data)

- **Column F: dm/z_L:** delta mass in ppm between mass of measure light isotope and calculated mass based on sequence
- **Column G: dm/z_H:** delta mass in ppm between measured and calculated heavy isotope
- **Column H: tR:** retention time of peak
- **Column I: N_label:** number of detected labelled residues. If mass difference determination is based on Peptide ID, this column will be a copy of the next column (column J). If mass difference was based on loop calculations, this column will contain the best iteration match
- **Column J: N_AA+1**: number of labelled residues determined from sequence. If N-terminal option was checked the title will be N_AA+1, else N_AA
- **Column K: Ratio_A:** Area ratio between light and heavy isotope
- **Column L: Ratio_H:** Peak Height ratio between light and heavy isotope
- **Column M: Correl:** (Cross) Correlation Coefficient between light and heavy EIC
- **Column N: Charge:** Charge of peptide, copied from the input info
- **Column O: Intensity**: Peak Height of the light isotope at the calculated retention time
- **Column P: Reference**: Protein Reference info copied from the input info

Entries in the table for which the heavy isotope peak could not be found are marked with NA. This applies to the fields: m/z H, dm/z_H, Ratio_A and Ratio_H. In these cases, the correlation coefficient will be set to zero.

Compared to the IPeaks table, two additional validation parameters are calculated, the mass difference in ppm between found and calculated (based on sequence) mass for both the light and heavy isotope. From the table shown, it can be seen that the precision for the light isotope peaks are all below 10 ppm. For the heavy isotope peaks, two entries have large deviations; no. 7 and 12. The seventh entry also has a very large ratio.

**Both entries are examined in depth using IPeaks.**

Entry 7: plotting the Mass Spectrum of Entry 7 shows that this peak (m/z 1058.62) is not the light but the heavy isotope. The algorithm does however find a small peak in the expected heavy position (m/z 1066.599, within the accepted accuracy of +/- 0.1 Da). As it seems, this is just a spurious peak. Therefore its ratio is very high and the calculated difference in ppm (-69.6) is way off. A negative deviation means the measured m/z value is smaller compared to the theoretical calculated m/z value. The correlation is quite high, because the elution profile of m/z 1066.599 co-elutes with the light isotope.

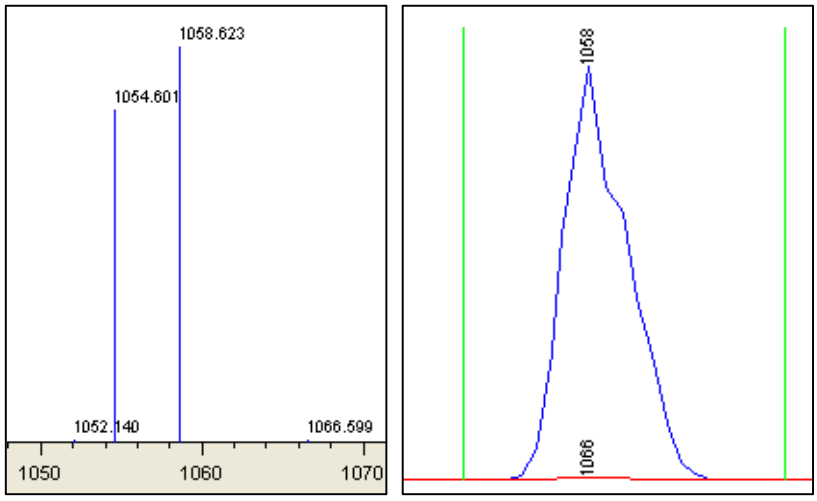



**Figure 13: mass spectrum and EICs of m/z 1058.623 and m/z 1066.599. It appears that the source file has an entry related to the heavy isotope peak.**

Entry 12, has a large deviation (-76.4 ppm) between measured and calculated mass, the ratio value is 4.12 (demo sample was mixed in a 1:1 L/H ratio) and the correlation coefficient is on the edge, 0.891. The pair is not to be trusted due to the large deviation from the expected mass. The found m/z value is lower in mass compared to the calculated mass. Figure 14 displays the Mass Spectrum and EICs.

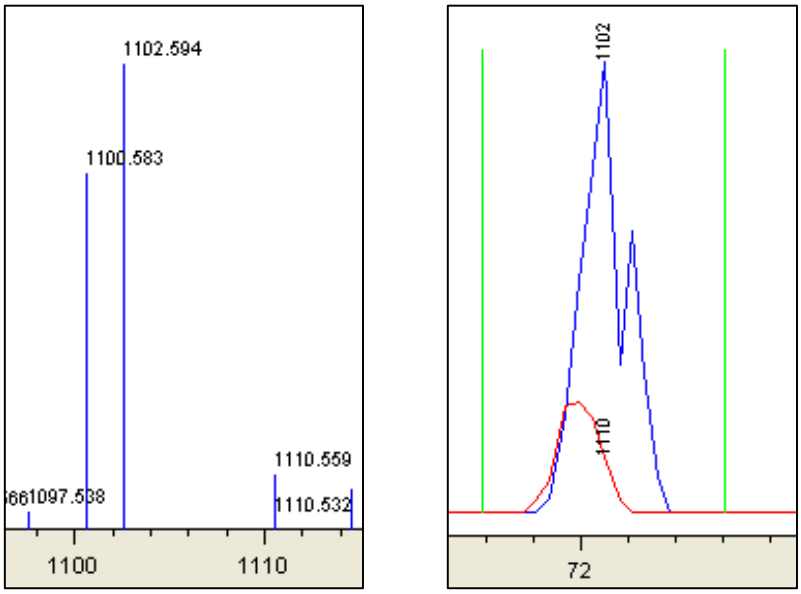



**Figure 14: mass spectrum and EICs of m/z 1102.594 and m/z 1110.559. The heavy m/z peak has a large deviation from the theoretical value (-76.4 ppm) and is not to be trusted.**

Any questions, remarks or suggestions regarding MsX-Quant should be send to support@msmetrix.com

**MsMetrix**