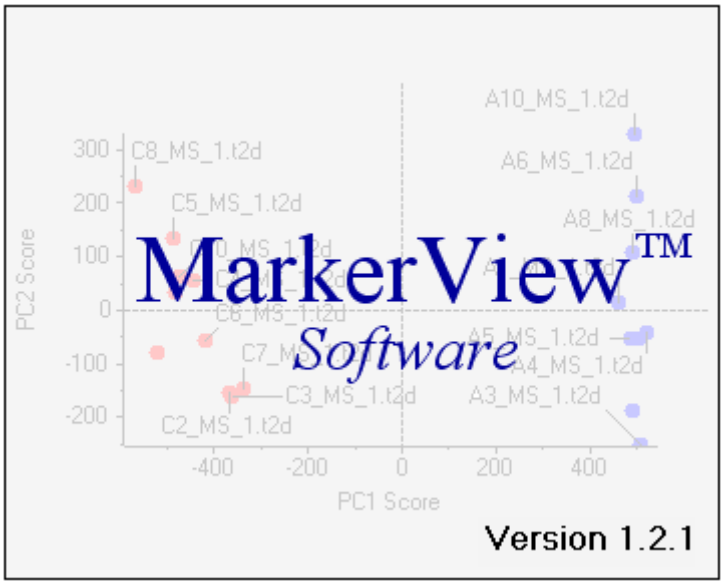


Reference Manual

This document is provided to customers who have purchased AB SCIEX equipment to use in the operation of such AB SCIEX equipment. This document is copyright protected and any reproduction of this document or any part of this document is strictly prohibited, except as AB SCIEX may authorize in writing.

Equipment that may be described in this document is protected under one or more patents filed in the United States, Canada, and other countries. Additional patents are pending.

Software that may be described in this document is furnished under a license agreement. It is against the law to copy, modify, or distribute the software on any medium, except as specifically allowed in the license agreement. Furthermore, the license agreement may prohibit the software from being disassembled, reverse engineered, or decompiled for any purpose.

Portions of this document may make reference to other manufacturers and/or their products, which may contain parts whose names are registered as trademarks and/or function as trademarks of their respective owners. Any such usage is intended only to designate those manufacturers' products as supplied by AB SCIEX for incorporation into its equipment and does not imply any right and/or license to use or permit others to use such manufacturers' and/or their product names as trademarks.

AB SCIEX makes no warranties or representations as to the fitness of this equipment for any particular purpose and assumes no responsibility or contingent liability, including indirect or consequential damages, for any use to which the purchaser may put the equipment described herein, or for any adverse circumstances arising therefrom.

For research use only. Not for use in diagnostic procedures.

The trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners.

AB SCIEX™ is being used under license.

AB SCIEX
71 Four Valley Dr., Concord, Ontario, Canada. L4K 4V8.
AB SCIEX LP is ISO 9001 registered.

Table of Contents

1 Introduction

This document provides a complete description of the functionality available in the MarkerView™ Software. The document describes all of the available menu items, toolbar buttons, pane types and so forth.

While some motivation is given for the actual utility of the various features, this document focuses primarily on describing how the various features work. If you are new to the program it is strongly recommended that you begin with the accompanying User Manual. That document uses a tutorial approach to walk you through the most important parts of the program, commenting extensively on the underlying rationale for the various operations. As its name implies, this document is intended as a reference.

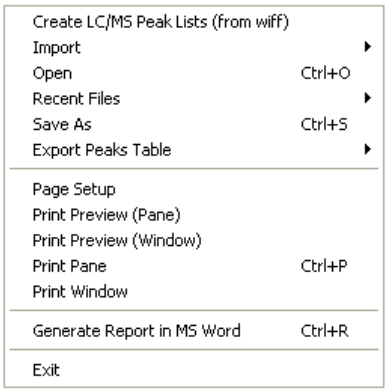
The appendix in section 6 describes to utility applications which are included with the software. Section 7 contains a high-level introduction to principal components analysis. For the benefit of those familiar with version 1.0 or 1.1 of the program, another appendix (section 8) lists the new features added with version 1.2.1.

Note – If you are reading an electronic version of this document, the section numbers appearing in gray text (for example 1) are links which you can click to jump to the referenced section; this also applies to the lines of the Table of Contents.

2 Menu Commands

This section contains an exhaustive description of the top-level menu items available in the MarkerView™ Software arranged in the order the various items appear in the menus. Context menus appearing in the individual pane types are described in subsequent sections.

2.1 File Menu



2.1.1 Create LC/MS Peak Lists (from wiff)

For each specified LC/MS sample, this command creates a *.peaks file containing the peaks which were detected for that sample. This command does not create the aligned Peaks Table and is used in conjunction with the *File→Import→LC/MS Peak Lists* command in section 2.1.2.1.

This option offers additional flexibility as compared to the *File→Import→LC/MS Data* command discussed below. For example you can perform the relatively slow peak finding step once for each sample and then create different aligned Peaks Tables using different groupings of samples; or you can reprocess the same peak lists using different retention time and mass alignment parameters.

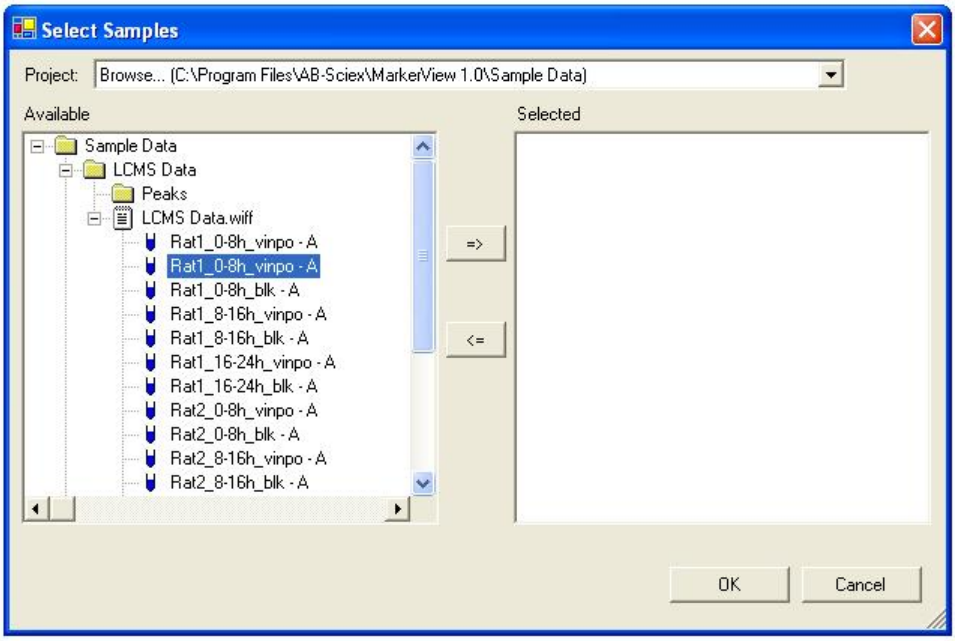
After selecting the command you perform the following steps:

(1) Select the samples to process.

(2) Specify the output folder for the generated *.peaks files – one peaks file will be created for each selected sample. The names of the output peaks files are based on the filenames of the wiff files with the sample index appended.

(3) Specify the various LC/MS peak finding parameters.

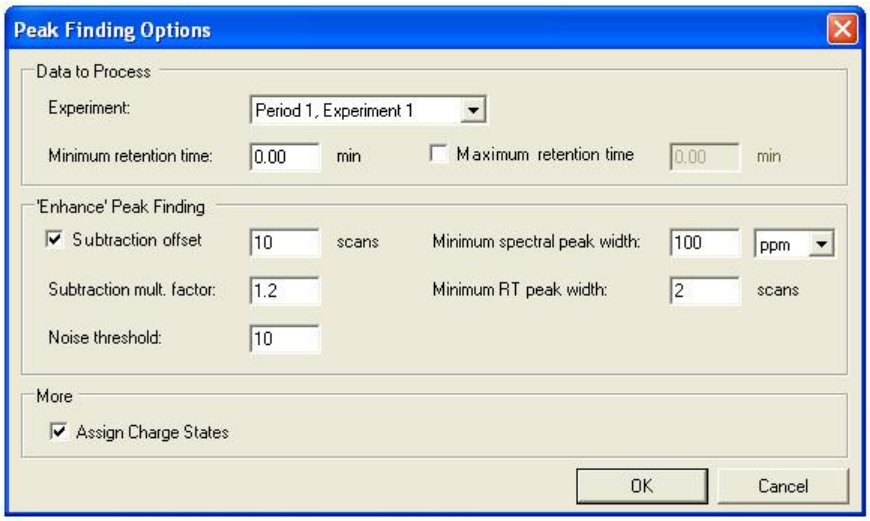
You use the dialog shown below to select the samples to be processed.



A few points related to this dialog:

- You select samples by moving them from the pane on the left-hand side to the one on the right. You can do this in a few ways such as by double-clicking an individual sample, by selecting a sample, data file or folder and clicking the '=>' button or by dragging any of these items from the left-hand pane to the right one. You can also hold the Shift and Control keys to select multiple files, samples or folders before moving them.
- If you accidentally add an unwanted sample, you can use a similar procedure to move it from the right-hand pane back to the one on the left.
- You use the *Project* combobox at the top of the dialog to select a different project. This combobox also contains a *Browse* item which you can choose to select data files which are not associated with any project.

After you click OK in the above dialog, you will be presented with another dialog asking for the output folder for the *.peaks files. After confirming that dialog, you will be presented with a final dialog asking for the peak finding options as shown in the figure below.



The following items control which data is processed:

- *Experiment* – In the case of multiple-period or multiple-experiment data, select the period and experiment combination to process. The program will only process one specific LC/MS experiment at a time.
- *Minimum retention time* – If desired, specify a retention time greater than zero to exclude the start of the runs. This is useful not only to speed processing, but also to avoid finding unwanted peaks in the chromatographic void volume.
- *Maximum retention time* – If desired, check this box and specify the maximum retention time which will be processed to exclude the end of the runs.

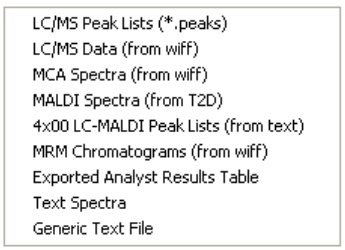
Peaks are found using an algorithm known as 'Enhance' which processes each of the mass spectra for a sample in order of increasing scan number. Masses belonging to the same peak 'cluster' are merged together with a resulting area equal to the sum of all intensities for each mass value in the cluster. The ion with the largest intensity is considered the base peak mass of the cluster. The dialog contains the following items which determine how the peak-finding algorithm operates:

- *Subtraction offset* – If you select this option, each spectrum is first background subtracted by the spectrum the specified number of scans before the current one. For example if this value is 10, the 100'th spectrum will be subtracted by the 90'th spectrum. You should generally use this option since it reduces the chance of finding constant background ions as peaks.
- *Subtraction multiplication factor* – Before subtraction, the background spectrum mentioned above is first multiplied by this value. Using a value somewhat larger than 1.0 is useful to compensate for minor variations in the intensity of background ions.
- *Noise threshold* – In order for a mass value to contribute to a cluster, its intensity in a given spectrum must be larger than this value. It is important that you set this parameter appropriately: setting it too small may cause noise to be identified as peaks and setting it too large will cause real peaks of interest to be missed. We suggest that you examine a few different spectra throughout the run and set the parameter somewhat larger than the chemical noise level. The units for this value are most usually counts/second, however for some time-of-flight data the units are counts – in all cases the absolute magnitude of the parameter corresponds to the observed intensity in the mass spectra themselves.
- *Minimum spectral peak width* – In order for a mass value to contribute to a cluster, its width in a given spectrum must be greater than this value. This value can be specified in absolute units or relative units of parts per million.
- *Minimum RT peak width* – This parameter specifies the minimum width (in scans, not minutes) required in order for a peak cluster to be kept. 'Peaks' narrower than this value are presumed to be noise.

The following item is also available:

- *Assign Charge States* – Select this checkbox if you wish the program to automatically assign charge states to the peaks. If this is not required, for large data sets processing will be faster if this is de-selected. Note that you can choose to assign charge states later (section 2.4.10).

2.1.2 File→Import Menu



2.1.2.1 LC/MS Peak Lists (.peaks)*

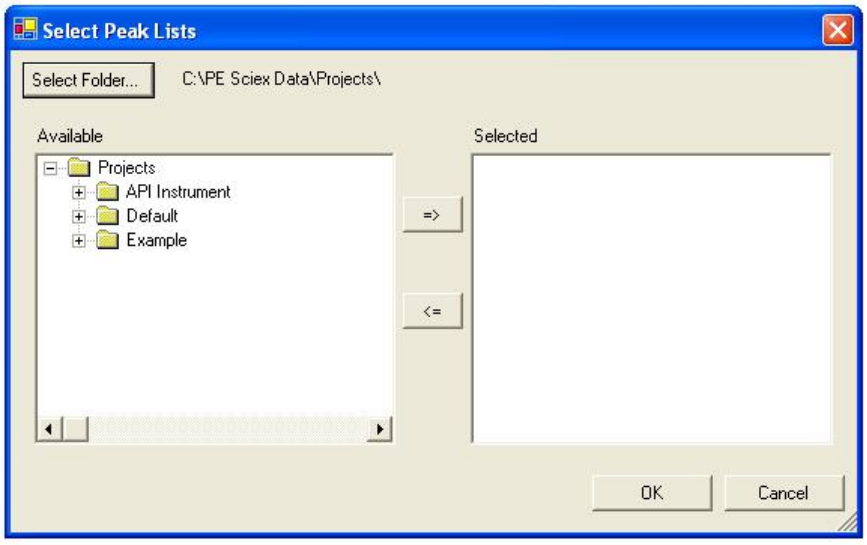
This command creates an aligned Peaks Table using the LC/MS *.peaks files which you specify. These files must have been previously created as described in section 2.1.1. As discussed in that section, the advantage of this two-step process, as compared to the one step *LC/MS Data (from wiff)* command described below, is that peak lists can be reused, for example with different retention time or mass alignment parameters.

After selecting the command you perform the following steps:

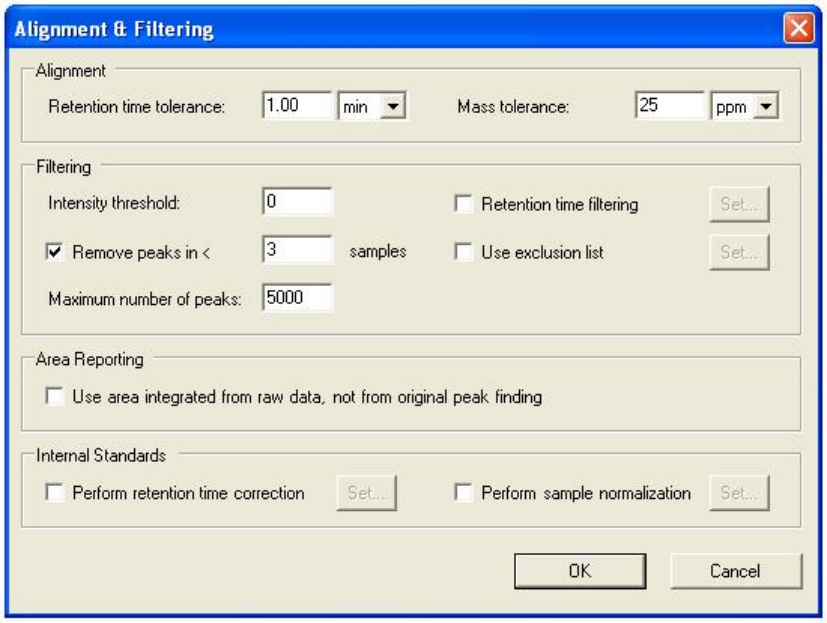
(1) Select the peak lists to process.

(2) Specify the peak alignment and filtering options.

You use the dialog shown below to select the peak lists. This dialog is very similar to that used to select wiff files when creating the original peak lists as described in section 2.1.1. The only significant difference is that you click the *Select Folder* button to choose the directory containing your peak lists.



Once you click OK in the above dialog you are presented with the following dialog asking for the peak alignment and filtering parameters.



The alignment parameters are used to decide if two {mass, RT} peaks found in two samples represent the same underlying chemical component or not. If their retention times and masses are both within the specified

tolerances this is assumed to be the case. Note that this also applies within *one* sample: for example if two isomers with the same mass elute within the specified time tolerance, they are combined into one peak.

- *Retention time tolerance* – Specify the retention time tolerance in either minutes or as a percentage. The percentage option allows a more generous tolerance for late eluting peaks. Note that the tolerance applies to the peaks *after* any retention time correction (if used) as described further below.
- *Mass tolerance* – Specify the mass tolerance in either absolute units or relative units of parts per million.

The following filtering parameters are available:

- *Intensity threshold* – Any peaks present in the *.peaks files with intensity less than the value you specify are discarded. This allows you to use a relatively low *Noise Threshold* when first finding peaks in the original data, with additional filtering performed at this stage. Specifying a threshold lower than the original *Noise Threshold* has no effect.
- *Remove peaks in < n samples* – This option is used to remove peaks which (after alignment) are only present in a few of the samples. The assumption is that such peaks are most likely to represent chemical or other noise and not potentially interesting peaks. That said, it's possible that such peaks may represent incorrectly aligned components if the mass or retention time tolerances are too small, so this option should be used with care.
- *Maximum number of peaks* – This parameter is used to reduce the size of the resulting peak list and is the maximum number of peaks which will be retained. If more than the specified number of peaks are detected, the ones with the smallest intensities are eliminated. You should set this parameter to allow a generous number of peaks to be retained, and only adjust it if you find subsequent data processing to be unreasonably slow.
- *Retention time filtering* – Check this item and click the *Set* button to set additional filtering options. The resulting dialog is described further below in section 2.1.2.1.1.
- *Use exclusion list* – Check this item to exclude unwanted components of known mass and retention time. To edit the list click the *Set* button. The resulting dialog is described in detail under *Options* in section 2.2.8.2.

The following parameter determines how the final peak areas are reported:

- *Use area integrated from raw data, not from original peak finding* – When this item is *not* checked, the peak areas reported by the program are those calculated by the 'Enhance' peak finding algorithm discussed in section 2.1.1. When this item *is* checked the program returns to the raw data for integration using the m/z and retention time ranges for the aligned peaks. The intention is to ensure that all peaks are ultimately treated equally regardless of how the original integration worked. This option slows the import process since the program must return to the raw data, however it can yield more accurate areas, especially for low-level components. In general this option is most useful when processing triple-quadrupole data.

If your samples contain internal standards of known mass and retention time, you can use the options discussed below.

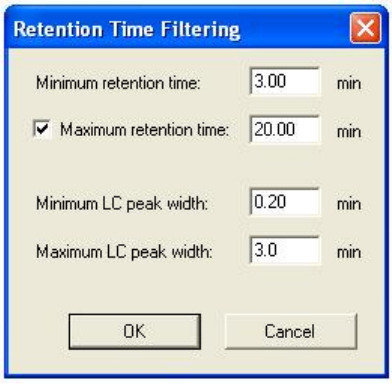
- *Perform retention time correction* – Select this checkbox to adjust the retention times for all peaks found for a given sample using the internal standards for the sample. Note that it is not actually important that the specified 'internal standards' are present at a constant amount in all samples, but merely that they appear in all samples with reasonable intensity.

  The dialog used to specify the internal standards when you click the *Set* button is discussed below in section 2.1.2.1.2.
- *Perform sample normalization* – Select this checkbox to automatically adjust the reported peak areas for each sample using the experimental areas of the internal standard peaks. Unlike retention time correction, these internal standards should be present in all samples at constant concentration. Also

unlike RT correction, if sample normalization is not performed here when the data is first imported, you can always apply it later – see section 2.4.3.2 for details.

2.1.2.1.1 Retention Time Filtering

When you enable the *Retention time filtering* option in the LC/MS *Alignment & Filtering* dialog and click the corresponding *Set* button, the following dialog is presented which allows you to specify additional filtering options as described below.



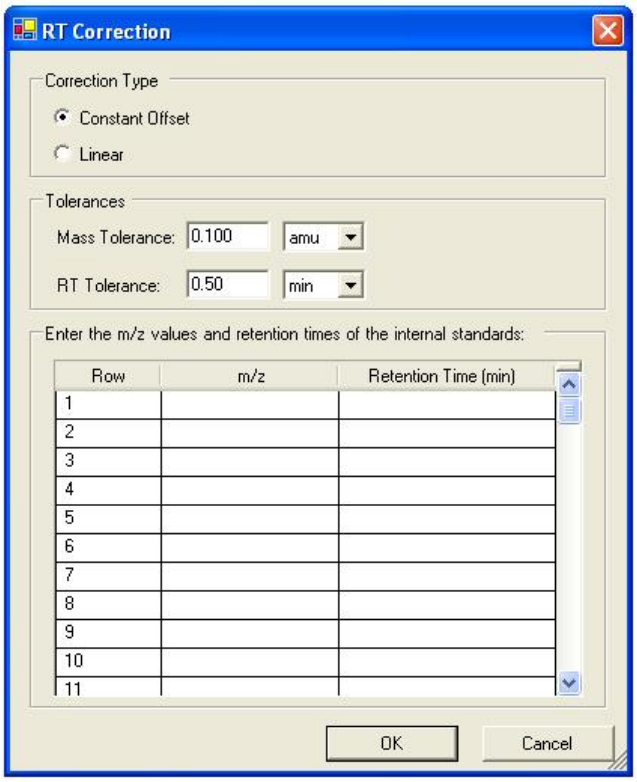
- *Minimum retention time* and *Maximum retention time* – These options allow you to exclude peaks which elute early or late in the chromatographic run. If you do not check the *Maximum retention time* item, the corresponding value is not used and no maximum is applied.
- *Minimum LC peak width* and *Maximum LC peak width* – These options allow you to exclude peaks with very narrow or wide chromatographic profiles.

2.1.2.1.2 RT Correction

When you enable the *Perform retention time correction* option in the LC/MS *Alignment & Filtering* dialog and click the corresponding *Set* button, the following dialog is presented which allows you to define one or more LC/MS internal standards which are used to correct for chromatographic peak shifts.

The new corrected retention times are determined using the expected retention times of the internal standards as explained below. You can review the correction which was applied to each sample by examining the *RT Correction* column of the Samples Table.

Note that for retention time correction, as opposed to sample normalization, it is not essential that the specified internal standards actually be present at a known constant concentration in all samples, but merely that they appear in all samples with reasonable intensity. For example you could use endogenous metabolites which vary in intensity, provided that they appear in all samples



The following items control how retention time correction is performed:

- *Constant Offset* – This option is used to apply a constant offset to the retention times for each sample. The offset for a given sample is the average offset for all internal standards which were located for the sample. If *no* internal standards could be located for the sample, no offset is applied.
- *Linear* – When you specify the *Perform retention time correction* option when importing the data, you use this option to apply a linear correction to the retention times for each sample. A standard linear regression is performed using all internal standards which could be located for the sample. If only one internal standard was found, the program automatically defaults to the *Constant Offset* case described above; if *no* internal standards were found, no correction is applied for the sample.

  When using a linear correction, the best results are obtained if your internal standards cover as much of the retention time range as possible. Using only internal standards which elute within a narrow window will result in excess extrapolation and possible spurious results.

Any peaks found for a given sample within the specified tolerances explained below are assumed to correspond to the internal standards. If the program locates two or more peaks which could potentially correspond to the same internal standard, the peak with the largest response is used.

- *Mass Tolerance* – Specify the mass tolerance in either Da or ppm. This tolerance is used for deciding if a given peak is potentially one of the internal standards.
- *RT Tolerance* – Specify the retention time tolerance in either minutes or as a percentage. The percentage option allows a larger tolerance to be used later in the chromatographic run. You will probably want to specify a somewhat more generous tolerance here than that used for peak alignment since alignment takes place after the retention times have been corrected.

You also specify the m/z values and retention times. Note that you can type directly into these fields or paste.

- *m/z* – Specify the mass/charge of the internal standards in this column.
- *Retention Time* – Specify the retention time in minutes of the internal standards in this column.

2.1.2.2 LC/MS Data (from wiff)

This command creates a new Peaks Table using the specified LC/MS samples. This one-step process is similar to the combined steps of creating peak lists as described in section 2.1.1 and then using them to create a Peaks Table as described in section 2.1.2.1. The disadvantage of this simpler workflow is that since intermediate peaks lists are not generated, you need to re-do all processing from scratch if you wish to re-analyze the data with different retention time and mass alignment parameters.

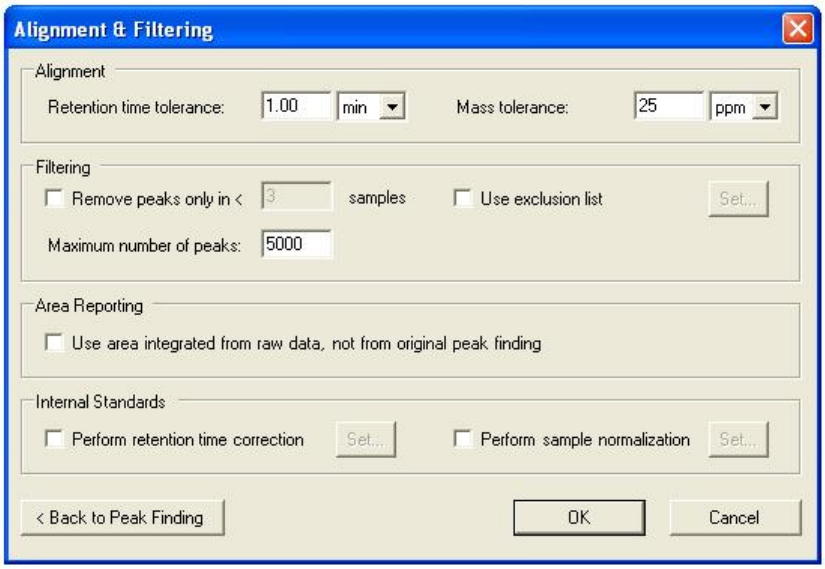
After selecting the command you perform the following steps:

(1) Select samples to process. The dialog you use to select the samples is identical to that described in section 2.1.1 and is not repeated here.

(2) Specify the various LC/MS peak finding parameters. The dialog you use for this is also identical to that described in section 2.1.1.

(3) Specify the peak alignment and filtering options.

The dialog you use to specify the alignment and filtering options is shown below. With the exception of the *Back to Peak Finding* button which is used to return to step (2), all of the fields present in this dialog are also present in the very similar dialog described in section 2.1.2.1, so their description is not repeated here. (The dialog of section 2.1.2.1 contains a few extra parameters which are not required here since you set the intensity threshold and retention time range directly in step (2) above.)



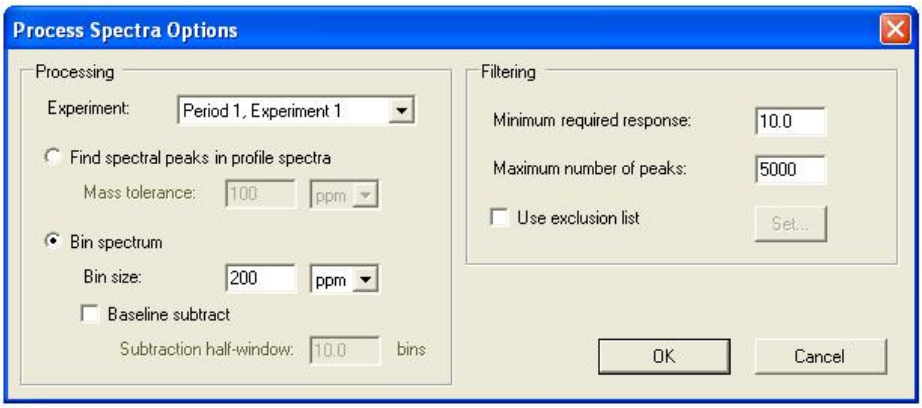
2.1.2.3 MCA Spectra (from wiff)

This command creates a new Peaks Table from a series of MCA analyses where you acquire a single summed spectrum for each sample. After selecting the command you perform the following steps:

(1) Select the samples to process. The dialog you use to select the samples is identical to that described in section 2.1.1 and is not repeated here.

(2) Specify the various alignment and filtering options.

The dialog you use to specify the alignment and filtering options is shown below.



The *Process Spectra Options* dialog contains the following general items:

- *Experiment* – In the case of multiple-period or multiple-experiment data, select the period and experiment combination to process. The program will only process one specific experiment at a time.
- *Find spectral peaks in profile spectra* – Select this option to locate peaks in the spectra using a spectral peak-finding algorithm. This algorithm is similar to that used in the Analyst® Software itself.
  - *Mass tolerance* – When using *the Find spectral peaks* option, peaks found for different samples are considered to represent the same component and are aligned if their masses differ by less than the specified tolerance. You can specify the tolerance in absolute units or relative units or parts per million.
- *Bin spectrum* – Select this option to 'cut' the entire mass range for each spectrum into bins or pieces. The reported area for a given bin is the sum of the intensities for all data points lying within the bin.
  - *Bin size* – When using the *Bin spectrum* option, this parameter specifies the width of the bins in Da or ppm.
  - *Baseline subtract* and *Subtraction half-window* – If this item is checked the program applies a baseline subtraction algorithm before summing the data points for the bins; the half-width used for the algorithm is the specified number of bins times the current bin size. The intention is to prevent any baseline offset from inflating the reported areas. This option is probably most useful for linear-mode MALDI data (see the next section). The subtraction half-window parameter should be set to approximately twice the expected peak width.

The following filtering parameters are available:

- *Minimum required response* – For each aligned peak, if the maximum intensity of the peak for all samples is less than the value you specify, that peak is eliminated. When using *the Find spectral peaks in profile spectra* option, this parameter refers directly to the peak intensity, however when using the *Bin spectrum* option it is the intensity *sum* for the bin.
- *Maximum number of peaks* – This parameter is used to reduce the size of the resulting peak list and is the maximum number of peaks which will be retained. If more than the specified number of peaks are detected, the ones with the smallest intensities are eliminated. You should set this parameter to allow a generous number of peaks to be retained, and only adjust it if you find subsequent data processing to be unreasonably slow.
- *Use exclusion list* – Check this item to exclude unwanted components of known mass. To edit the list click the *Set* button. The resulting dialog is described in more detail under *the Edit→Options* section is section 2.2.8.2.

2.1.2.4 MALDI Spectra (from T2D)

This command creates a new Peaks Table from a series of 4700 or 4800 MALDI spectra, where there is one MS spectrum per sample (this option does not process 4700/4800 LC-MALDI data). You must have previously exported the desired spectra from the Oracle database in the *.t2d format.

After selecting the command you perform the following steps:

(1) Select the t2d files containing the spectra to process. The dialog you use to select the samples is identical to that used to select peaks files when importing LC/MS data – see section 2.1.2.1 for details.

(2) Specify the various alignment and filtering options. This dialog is identical to that described in the previous section – see section 2.1.2.3 for details.

2.1.2.5 4x00 LC-MALDI Peak Lists (from text)

This command creates a Peaks Table from a series of 4700 or 4800 LC-MALDI runs exported in a text format.

Before selecting this command you must use the 4x00's 'Peaks Explorer' program to export a separate peak list for each LC-MALDI run. These text files contain a row for each LC peak detected for the sample. The procedure for generating these files is the following. For each LC run:

- Open the desired spot set in Peaks Explorer
- Use the 'Generate LC/MALDI Heat Map' menu item followed by the 'Generate LC/MALDI Peak List' menu item to create the actual peak list
- Right-click in the 'LC/MALDI Peaks' table and choose the 'Copy all to Clipboard' menu item
- Open a text editor such as Notepad.exe and paste the contents of the clipboard
- Save the file with a '.txt' file extension.

Once the peak lists have been exported, you perform the following steps after selecting the command:

(1) Select the peak lists which you want to process. The dialog used for this is identical to that used to select peaks files when importing LC/MS data – see section 2.1.2.1 for details. Note that the file extension of these peak lists must be '.txt'.

(2) Specify the various alignment and filtering options. This dialog is identical to that described in section 2.1.2.1.

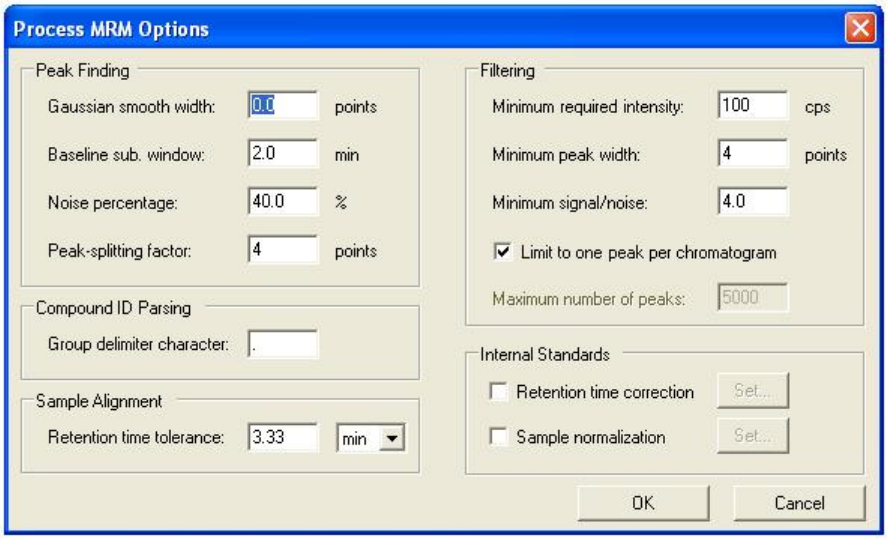
2.1.2.6 MRM Chromatograms (from wiff)

This command allows MRM (multiple reaction monitoring) data acquired on a triple-quadrupole instrument to be directly imported into the program. A chromatographic peak finder is used to find peaks in each chromatogram and they are aligned (for a given transition) using a retention time tolerance.

This command provides a fast way to import MRM data into the program, however note that for most precise work it is probably best to use the Analyst® quantitation module to generate peak areas so that the integrations can be reviewed and adjusted as needed. The Results Tables generated in this way can be imported into the program using the command described in section 2.1.2.7.

After selecting the command you are first prompted for the samples to process using the same sample-selection dialog described in section 2.1.1. Once you have selected the samples the dialog shown below is used to specify the processing options.

Note that the program uses the same MQII algorithm as available in the Analyst® Software to find peaks – for more details on this algorithm and its parameters see the Analyst® documentation



The *Process MRM Options* dialog contains the following items:

Peak Finding:

- *Gaussian smooth width* – Prior to peak detection, each chromatogram is smoothed using a Gaussian smoother with a width at half of the peak height of the specified value. Use a value of zero for no smoothing; otherwise a value between 0.5 and 2.0 is suitable for most data.
- *Baseline sub. window* – This MQII parameter specifies the width of the baseline subtraction window which is applied as the first step of the peak detection process. Enter a value roughly twice the expected peak width.
- *Noise percentage* – This MQII parameter determines the peak detection threshold. Specify a relatively small value (say 20%) to allow small peaks to be detected in the presence of larger peaks in the same chromatogram and a large value (say 80%) to prevent this. The default value of 50% is usually suitable.
- *Peak-splitting factor* – This MQII parameter controls whether adjacent overlapping peaks are found as separate peaks or as a single combined peak. The default value of 3 is usually suitable.

Compound ID Parsing:

- *Group delimiter character* – This field is most useful for proteomics applications. Version 1.5 of the Analyst® software allows compound names or IDs to be associated with each MRM transition. If such IDs are available the *Peak Names* in the resulting Peaks Table will use these values. For proteomics applications it is recommended to use the following naming convention:

  {Peptide Sequence}.{Transition}.{Light or Heavy}

  or

  {Protein Name}.{Peptide Sequence}.{Transition}.{Light or Heavy}

  When present, the '.Light' suffix indicates an analyte and '.Heavy' an internal standards. If this naming convention is followed, the peak *Group* column in the Peaks Table will be automatically populated with the ID where the transition and (if used) analyte or IS indicator is removed. If you are using this convention but with a separator other than a period, you should specify the separator character in this field.

Sample Alignment:

- *Retention time tolerance* – Specify the retention time tolerance in either minutes or as a percentage. The percentage option allows a more generous tolerance for late eluting peaks. Note that the tolerance applies to the peaks *after* any retention time correction (if used) as described immediately below.

Filtering:

- *Minimum required intensity* – Any peaks with peak height less than this value are ignored.
- *Minimum peak width* – Any peaks found by the MQII algorithm which are narrow than this value are ignored.
- *Minimum signal/noise* – Any peaks with a S/N estimate less than this value are assumed to be noise and are ignored.
- *Limit to one peak per chromatogram* – Select this checkbox if you wish exactly one peak to be reported for each MRM transition (for each sample). This makes subsequent data review easier since there are fewer rows in the resulting Peaks Table. When selected, the processing algorithm makes two passes through the data. During the first pass it determines the retention time of the largest peak across all of the samples for each MRM transition separately; during the second pass the peaks near these retention times are integrated. So, this option is useful if it is expected that each MRM transition should contain only one peak of interest and that there are not very large interferences. The option is especially useful for data acquired using the Scheduled MRM™ algorithm.
- *Maximum number of peaks* – This parameter is used to reduce the size of the resulting peak list and is the maximum number of peaks which will be retained. If more than the specified number of peaks are detected, the ones with the smallest intensities are eliminated. You should set this parameter to allow a generous number of peaks to be retained, and only adjust it if you find subsequent data processing to be unreasonably slow.

Internal Standards:

- *Retention time correction* – Select this checkbox to adjust the retention times for all peaks found for a given sample using the internal standards for the sample. Note that it is not actually important that the specified 'internal standards' are present at a constant amount in all samples, but merely that they appear in all samples with reasonable intensity.

  The dialog used to specify the internal standards when you click the *Set* button is discussed below in section 2.1.2.6.1.

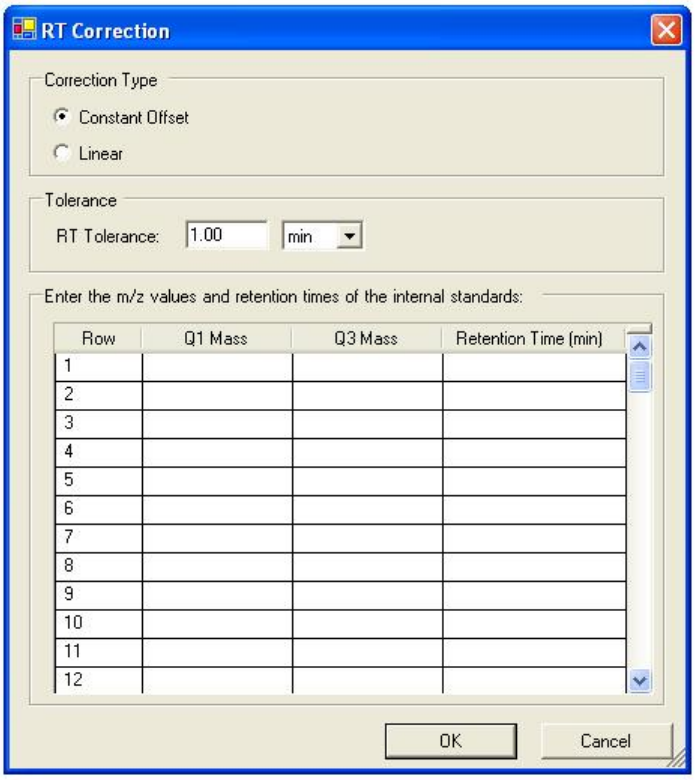
- *Sample normalization* – Select this checkbox to automatically adjust the reported peak areas for each sample using the experimental areas of the internal standard peaks. Unlike retention time correction, these internal standards should be present in all samples at constant concentration. Also unlike RT correction, if sample normalization is not performed here when the data is first imported, you can always apply it later – see section 2.4.3.2.

2.1.2.6.1 MRM RT Correction

When you enable the *Perform retention time correction* option in the *Process MRM Options* dialog and click the corresponding *Set* button, the following dialog is presented which allows you to define one or more MRM internal standards which are used to correct for chromatographic peak shifts.

This dialog is very similar to that shown in section 2.1.2.1.2 for correcting scan LC/MS data, however instead of entering a single mass and a retention time for each standard as in that dialog, you specify both the Q1 and Q3 masses and retention time for each internal standard.

The retention time tolerance is needed in case there is more than one peak found in a given chromatogram. *Note:* You do not explicitly enter a mass tolerance, however an 0.1 Da tolerance is always used when comparing a particular Q1/Q3 pair to masses specified in the dialog.



2.1.2.7 Exported Analyst Results Table

This command allows you to create a new Peaks Table using the information contained in a quantitation Results Table. This allows you to use the application's data analysis tools on quantitative data. The general workflow when using this option is the following:

(1) You use the Analyst® Software to acquire the quantitative data.

(2) You use the Analyst® quantitation tools to create a new Results Table. You use the various tools such as Peak Review to verify the accuracy of the peak integrations in the usual way.

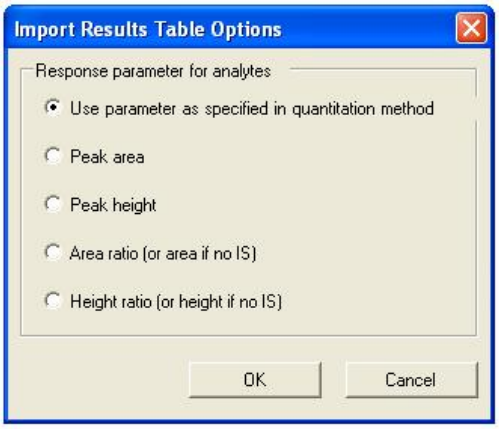
(3) From within the Analyst® Software when the Results Table is active, you use *the Export to Text* command to save the table in a tab-delimited text format. This file will contain the peak heights and areas for all analytes and internal standards as well as other related information.

(4) You start the MarkerView™ Software and select the *Exported Analyst Results Table* menu item.

(5) When prompted, you select the previously exported Results Table.

(6) You specify the desired response parameter (area, height, *etc.*).

The program will create a new Peaks Table with a row or peak for each analyte peak contained in the exported text file. The response parameter reported in the table is determined by your selection from the options dialog shown below.



The following options are available:

- *Use parameter as specified in quantitation method* – The parameter is automatically determined based on the *Regression Parameter* (area or height) specified in the quantitation method and whether or not the analyte has an associated internal standard. For analytes which do *not* have an associated internal standard the peak area or height is used as-is, otherwise the area ratio or height ratio (relative to that internal standard) is used
- *Peak area* – Always uses the analyte peak area.
- *Peak height* – Always uses the analyte peak height.
- *Area ratio (or area if no IS)* – Uses the ratio of the analyte peak area to the internal standard peak area. If the analyte has no associated internal standard, the area is used directly.
- *Height ratio (or height if no IS)* – Uses the ratio of the analyte peak height to the internal standard peak height. If the analyte has no associated internal standard, the height is used directly

2.1.2.8 Text Spectra

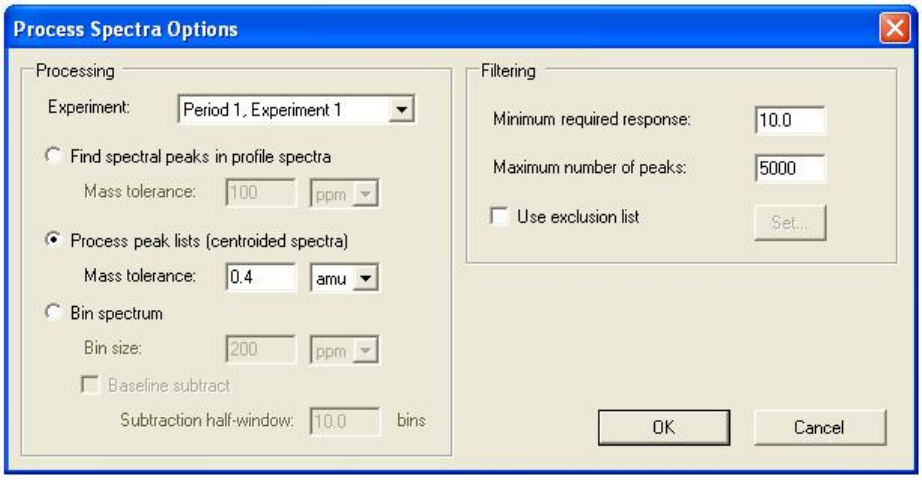
This command creates a new Peaks Table from a series of spectra saved in tab-delimited text format, where there is one spectrum per sample (this option cannot process LC/MS data in text format). This functionality is not intended to be used if your data originates from wiff files or can be exported in the 4700/4800 t2d format; it is provided to allow spectra from other sources to be imported into the program.

The text files must have a '.txt' file extension and should be in tab-delimited mass-intensity format: *i.e.* with the mass/charge at the start of a new line, followed by a tab, followed by the peak intensity or area, followed by a carriage return. The files may contain either profile spectra where all acquired m/z values are present or peak lists.

After selecting the command you perform the following steps:

(1) Select the text files containing the spectra to process. The dialog you use to select the files is identical to that used to select peaks files when importing LC/MS data – see section 2.1.2.1 for details.

(2) Specify the various alignment and filtering options.



The dialog used to specify the options is shown above and is almost identical to the dialog discussed in section 2.1.2.3. The only additional items are the following:

- *Process peak lists (centroided spectra)* – Select this option if you are importing peak lists as opposed to actual profile spectra.
    - *Mass tolerance* – When the above option is used, peaks are considered to represent the same component and are aligned if their masses differ by less than the specified tolerance. You can specify the tolerance in absolute units or relative units or parts per million.

2.1.2.9 Generic Text File

This command creates a new Peaks Table based on the information in a text file containing aligned peaks for a series of samples. Essentially the text file will look like a Peaks Table with a row for each peak and a column for each sample (or vice versa). This functionality is not intended to be used if your data originates from wiff files or can be exported in the 4700/4800 t2d format; it is provided to allow data from other sources to be imported into the program.

These text files should be in the following format

- Fields are separated with tabs with a carriage return at the end of each line.
- The first row *must* start with the text 'Sample Name' and contain the names of the various samples.
- The second row may optionally start with the text 'Sample ID' and contain the sample IDs of the samples.
- The third row may optionally start with the text 'Group' and contain the group or class for the samples.
- All remaining rows should start with the name of the peak followed by the responses for each peak.

The figure below shows an example:

Book1

	A	B	C	D
1	Sample Name	Sample 1	Sample 2	Sample 3
2	Sample ID			
3	Peak 1	946.86	1423.8	894.31
4	Peak 2	117.45	64.537	67.451
5	Peak 3	96.863	67.799	139.22
6	Peak 4	484.9	559.4	425.29
7	Peak 5	110.59	59.412	147.84

You can generate an example file in the correct format by creating a Peaks Table in any other way (by importing LC/MS data, spectra, *etc.*) and then using *the File→Export Peaks Table →Generic Text File* (*or Generic Text File (transposed)*) menu items discussed in section 2.1.6.

It is also possible to transpose these files so that rows become columns and vice versa. This is automatically detected by the program, however in this case you must include both the 'Sample Name' and 'Sample ID' columns.

2.1.3 Open

This command opens a processed data set previously saved in the *.mrkvw format using the *Save As* command. This command is *not* used to process raw data files.

2.1.4 Recent Files

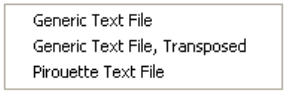
This menu contains sub-menu items for each recently save file (see the next section). Select one of the items to open the corresponding file.

2.1.5 Save As

This command saves the currently active data set to a new *.mrkvw file which you specify. In addition to the actual peak areas for the various peaks and samples, this format also saves related information such as the sample name and group for each sample, which peaks and samples (if any) have been excluded, *etc.*

If you have generated a data set with many peaks excluded, you may wish to save a copy of the data so that you can continue analyzing the data at a later time starting from where you left off.

2.1.6 File→Export Peaks Table Menu



The items in this menu are all used to export the active data to a tab-delimited text file in various formats. This is useful if you wish to perform data analysis outside of the program using third-party or in-house tools.

2.1.6.1 Generic Text File

This command exports to a simple tab-delimited text file containing the sample name, sample ID and group (or class) for each sample, the name of each peak and the response for each sample for each peak. The exact format was described in section 2.1.2.9.

2.1.6.2 Generic Text File, Transposed

This command is similar to the *Generic Text File* command with the difference that the resulting file is transposed so that rows become columns and columns become rows.

2.1.6.3 Pirouette® Text File

This command saves to a specially formatted text file intended for importing into the Infometrix Pirouette program. For more information on Pirouette visit www.infometrix.com.

2.1.7 Page Setup

Sets the current page setup for printing (landscape versus portrait mode, *etc*). All subsequent printing will use this setup.

2.1.8 Print Preview (Pane)

Displays a print preview for the currently active pane.

2.1.9 Print Preview (Window)

Displays a print preview for all of the panes in the active window.

2.1.10 Print Pane

Prints the currently active pane. Tables are printed using as many pages as necessary to print their entire contents; plots are printed on a single page sized to fill the paper.

2.1.11 Print Window

Prints the active window on a single page, sized to fill the paper. If you wish to print only a subset of the panes, you can use the *Hide Pane* toolbar button to first hide the panes which you do not wish to print.

2.1.12 Generate Report in MS Word

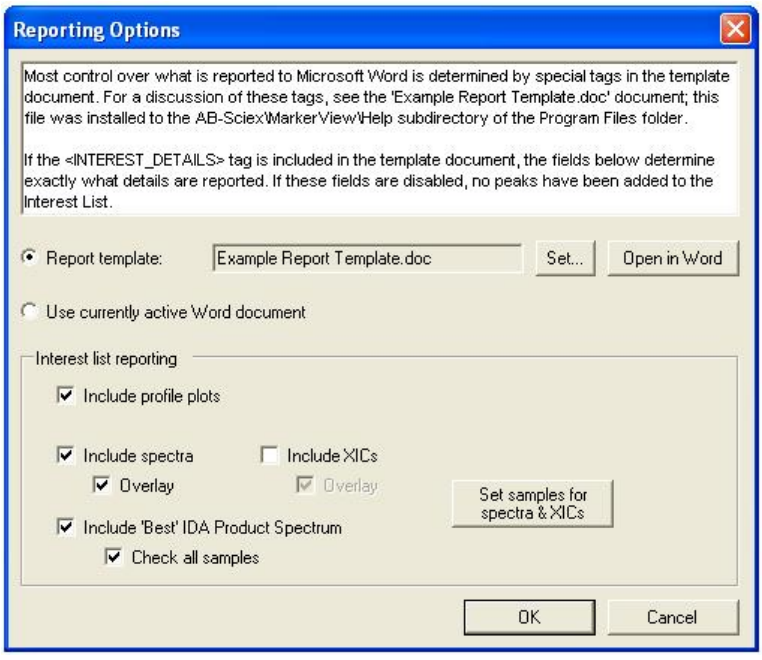
This command is used to create a report in Microsoft Word. You control exactly what is reported by editing the tags in a Word document which is used as a template and by specifying additional options within the MarkerView™ Software itself. The following elements can be reported:

- The name of the saved *.mrkvw file.
- The peak finding and alignment parameters.
- The Samples Table containing the name, group, *etc.* for each sample.
- The list of peaks added to the interest list, if any.
- The list of excluded peaks, if any.
- If PCA data is active, a list of the principal components.
- If PCA data is active, selected scores and loadings plots.
- Additional details for each peak in the interest list. This includes Profile Plots, raw spectra and raw XICs and, if a t-Test Table is active when the report is generated, the t-test t-value, p-value and fold change.

This document does not explain the exact form of the various tags which can appear in the template document. For details see the 'Example Report Template.doc' file which was installed to the AB-Sciex\MarkerView\Help subdirectory of the Program Files folder.

Note: While the application is in the process of generating a report, you should not manually use Word.

Once you select the menu item the dialog shown below is presented.

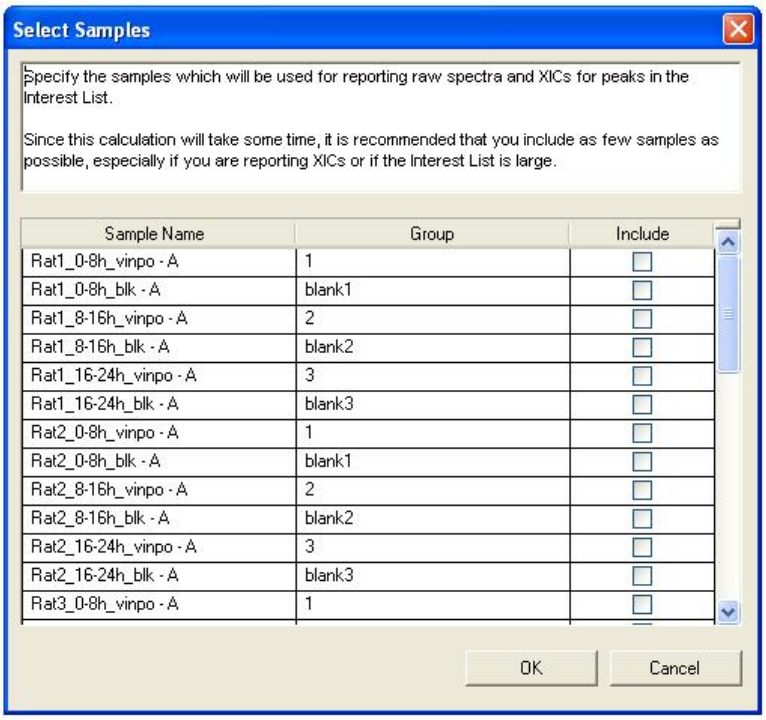


The Reporting Options dialog contains the following items:

- *Report template* – If you choose this option, a new copy of the specified report template is automatically opened in Word and populated.
- *Report template: Set* – Click this button to select the report template containing the special tags. Most likely this will be a modified version of the 'Example Report Template.doc' document.
- *Report template: Open in Word* – Opens the currently specified report template in Word so that you can view or edit it.
- *Use currently active Word document* – If you choose this option, the program assumes that you have previously opened a template in Word and the active Word document will be populated with the report.

If you have added at least one peak to the interest list (see section 4.8) for the active pane, the following items are also available:

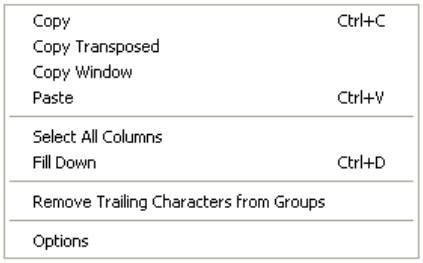
- *Include profile plots* – Select this to include a Profile plot for each peak in the interest list.
- *Include spectra* – Select this to include raw spectra for each peak in the interest list. This option is not available if you have imported text data into the program. Depending on the setting of the *Overlay* checkbox, the spectra for the reported samples are either overlaid in one plot or a separate plot is included for each sample (for each peak).
- *Include XICs* – Select this to include raw extracted ion chromatograms for each peak in the interest list. This option is only available for LC/MS data. The behavior of the *Overlay* checkbox is similar to that described above for spectra.
- *Include 'Best' IDA Product Spectrum* – If you have processed the MS data for a series of IDA runs, check this item to include product spectra (where available) for each peak in the interest list. If the *Check all samples* item is checked, the program checks all available samples for the most intense product spectrum matching the precursor m/z and retention time for each peak, otherwise it uses only the samples specified using the *Set samples for spectra & XICs* dialog discussed below. The former option is more likely to locate a match, however depending on the number of samples it will be slower. If a matching product spectrum (for the specified samples) was not acquired for a particular peak, a message stating that fact is inserted into the Word document.
- *Set samples for spectra & XICs* – If you are reporting raw spectra and/or XICs, click this button to specify which samples are reported. When you do this the dialog shown below is presented. Click the appropriate checkboxes in the *Include* column for all of the samples which you wish to report. As mentioned at the top of the dialog itself, including many samples may considerably slow the reporting process, especially if the interest large is large or if you are reporting XICs.



2.1.13 Exit

Quits the program. If you have unsaved data which you have imported into the program, you will be asked if you wish to save it. However you are *not* asked if you have edited previously saved data (changed sample names, excluded peaks or samples, *etc.*).

2.2 Edit Menu



2.2.1 Copy

When the active pane is a plot, this command copies an image of the plot to the clipboard. When a table is active, the selected portion of the table is copied.

2.2.2 Copy Transposed

When the active pane is a table, this command copies a transposed version of its selected portion to the clipboard.

2.2.3 Copy Window

Copies an image of the entire active window to the clipboard. This is similar to the Microsoft Windows 'Print Screen' functionality except that the window's title bar and the toolbars of its various panes are not included.

2.2.4 Paste

When the active pane is a table with an editable selected area, this command pastes from the clipboard. You can paste into individual cells or columns.

One particular situation in which this is useful is to copy the *Group* column from one Samples Table to another containing the same samples: select the column in the first Samples Table and choose *Copy*, then select the column in the second table and choose *Paste*. This technique can also be used to copy other editable columns (*Sample Name*, *Use*, *etc.*) from one table to another.

2.2.5 Select All Columns

Selects all of the columns of the active table. This is useful if you want to copy the entire contents of the table so that it can be pasted elsewhere.

2.2.6 Fill Down

When the active pane is a table with an editable selected region, this command replicates the information in the first selected row to all subsequent selected rows.

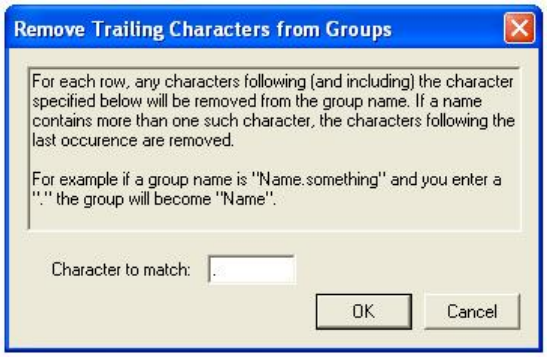
In the example below if the *Fill Down* command is selected the *Use* flag for all peaks will be enabled (since it is enabled for the first row).

Row	Index	Peak Name	m/z	Ret. Time	Group	Use	R
1	1	81.1/13.3 (1)	81.067	13.32		☑	0.(
2	2	83.1/13.7 (2)	83.084	13.65		☐	0.(
3	3	83.1/12.6 (3)	83.084	12.62		☐	0.(
4	4	85.0/12.5 (4)	85.027	12.51		☐	0.(
5	5	89.0/11.5 (5)	89.036	11.51		☐	1.:
6	6	90.6/11.3 (6)	90.616	11.28		☐	0.(
7	7	91.1/11.3 (7)	91.050	11.29	(Monoisotopic)	☑	3.:
8	8	91.1/14.6 (8)	91.051	14.63		☐	0.(

2.2.7 Remove Trailing Characters from Groups

This command removes any characters from the end of the group name – this applies to peaks when a Peaks Table is active and samples when a Samples Table is active. This is perhaps most useful for proteomics applications which use the naming convention mentioned in section 2.1.2.6. For example if the name of a peak group were 'ECOLI.NGEFIEITEK' and if the '.' character were specified the group would become 'ECOLI', allowing you to switch from peptide grouping to protein grouping.

When the command is selected the dialog shown below is presented.



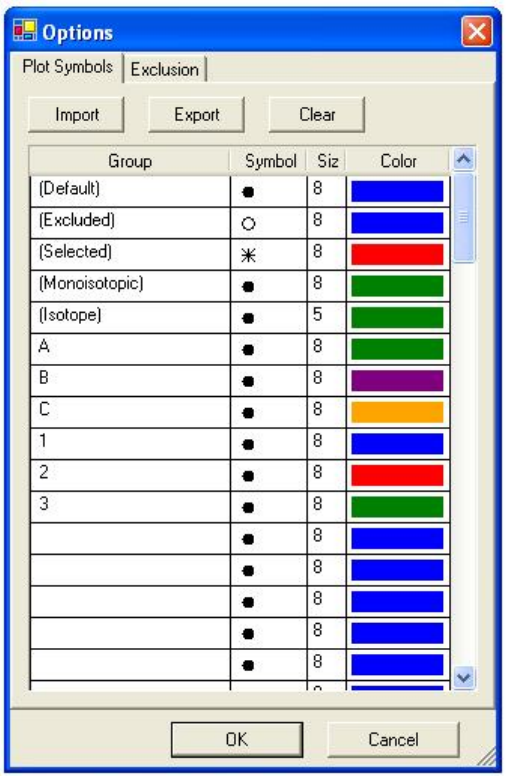
2.2.8 Options

This command displays a tabbed dialog for setting various options. The individual tabs are explained below.

Note: The various options are saved between sessions to the 'Sciex\MarkerView' sub-directory of the Windows 'Application Data' directory. This directory is usually located at 'C:\Documents and Settings*LOGGED_IN_USER*\ Application Data\Sciex\MarkerView' where *LOGGED_IN_USER* is the name of the user logged-into Windows. You can copy all of the program's settings from one computer to another by copying this directory (and the files it contains).

2.2.8.1 Plot Symbols

The *Plot Symbols* tab allows you to associate a particular plot symbol (shape, size and color) with each sample or peak. You do this by specifying the *Group* (or class) in the Samples or Peaks Table (for samples and peaks respectively) and by specifying the corresponding plot symbol here.



The tab contains the following items:

- *Import* – Prompts you for a file containing previously exported plot symbol information and appends the plot symbols from the file to the table. This is useful, for example, if you wish to use the plot symbols defined on one computer on a second computer.
- *Export* – Prompts you for the name of a file and saves the plot symbol information to the file. The *Group*, *Symbol*, *Size* and *Color* are all saved, however note that the first five special groups (discussed below) are not exported.
- *Clear* – Clears all entries from the table with the exception of the first five special groups discussed below. You may want to clear the table before subsequently importing previously exported plot symbols.
- *Group* – Type the names of your groups or classes into this column. You can define as many groups as desired – they do not all need to be present in any one data set. Note that you cannot edit the names of the first five special groups discussed below.
    - *(Default) Group* – The plotting symbol defined for this special group is used when the group for a given sample or peak is not otherwise defined in the table, unless you also include groups named '1', '2', *etc.* as discussed below.

- *(Excluded) Group* – This plotting symbol is used when a sample or peak has been excluded, *i.e.* when its *Use* checkbox has been turned off.
- *(Selected) Group* – This plotting symbol is used when a sample or peak is selected in a graph, for example by clicking directly on the point in the plot or by clicking in a cell of a table linked to the graph.
- *(Monoisotopic) Group* – The plotting symbol defined for this special group is used for peaks with a corresponding group name of '(Monoisotopic)'. The program automatically attempts to assign monoisotopic peaks to this group when processing LC/MS or MS data. If you do not wish monoisotopic and isotope peaks to use a special plotting symbol, you can clear the group names (for example by clearing the name for the first peak in the Peaks Table, selecting the *Group* column, and then choosing the *Fill Down* command from the *Edit* menu).
- *(Isotope) Group* – This plotting symbol is used for peaks assigned to the '(Isotope)' group. The program automatically attempts to assign isotopes peaks to this group when processing LC/MS or MS data. Peaks which cannot be automatically assigned as either monoisotopic or isotope are plotted using the *(Default)* symbol.

- *Symbol, Size and Color* – Set the plotting point symbol, size and color using the comboboxes. If you select the *(Custom)* item from the *Color* combobox you will be presented with the standard Windows Color-Picker dialog allowing you to select a custom color.

If the group name for a particular sample or variable is not otherwise defined, the special *(Default)* plotting symbol will normally be used. However if you explicitly add groups named '1', '2', '3', *etc.* the symbol defined for these groups will be used instead where the '1' symbol is used for the first unspecified group and so forth. For example using the groups specified in the figure above, if there are samples with the following group names:

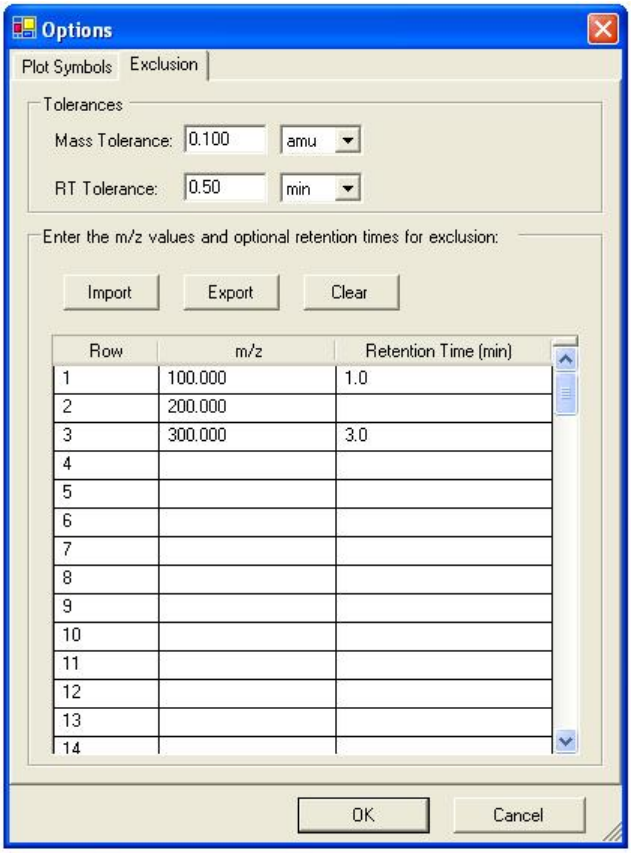
'A', 'B', 'C', 'D', 'E', 'F', 'G'

the 'A', 'B' and 'C' groups will use the explicitly defined symbols, the 'D' group will use the symbol defined for the '1' group (since it is the first undefined symbol in alphabetical order), the 'E' group will use the symbol defined for the '2' group, the 'F' group will use the symbol defined for the '3' group and the 'G' group will use the symbol defined for the *(Default)* group (since no group '4' was defined). The advantage of this scheme is that if you frequently use different group names, you do not need to define plotting symbols for them all, just once for the '1', '2', '3', *etc.* groups.

2.2.8.2 Exclusion

The *Exclusion* tab allows you to specify the m/z values and optional retention times of peaks to remove when importing spectral or LC/MS data. For example you may want to specify the m/z values of known MALDI matrix ions or the m/z values and retention times of known xenobiotic drug metabolites. You enable this feature by selecting the *Use exclusion list* checkbox of the LC/MS and spectral import dialogs.

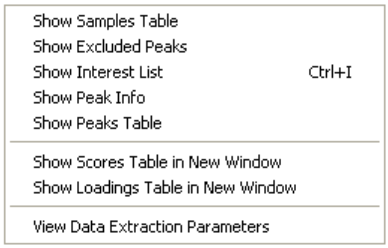
You can populate the exclusion list by typing into the spreadsheet or by importing from a tab-delimited text file as explained below. You can also populate the list by (1) de-selecting the *Use* flag for certain peaks (as explained elsewhere in this document), (2) using the *Show Excluded Peaks* menu item to view the exclusion list and (3) using the menu items in that pane's context menu – see the discussion of the Excluded Peaks Table in section 4.7 for details.



The tab contains the following items:

- *Mass Tolerance* – Specify the mass tolerance in either Da or ppm used for deciding if a given peak is potentially one of the unwanted items in the exclusion list.
- *RT Tolerance* – Specify the retention time tolerance in either minutes or as a percentage. The percentage option allows a larger tolerance to be used later in the chromatographic run. This tolerance is only used when processing LC/MS data.
- *Import* – Click this button to select a tab-delimited text file containing m/z values (in the first column) and optional retention times (in the second column, if present) for appending to the exclusion list. This saves having to manually type the values into the spreadsheet. This is useful in conjunction with the *Export* button so that you can:
  - save an exclusion list which is appropriate for one particular study
  - use a different list when processing data for another study
  - import the original exclusion list if you want to import new data files from the first study
- *Export* – Click this button to save the current exclusion list to a tab-delimited text file of your choosing.
- *Clear* – Click this button to clear the spreadsheet.
- *m/z* – Specify the mass/charge of the peaks to exclude in this column. You must also specify the m/z values of any isotopes if you wish to exclude them too.
- *Retention Time* – If you will be importing LC/MS data, specify the retention time in minutes of the peaks to exclude in this column; if you will be importing spectral data these values are not used. If you leave the retention time empty and subsequently import LC/MS data, the specified m/z will always be excluded for all retention times.

2.3 View Menu



2.3.1 Show Samples Table

This command displays the table of samples associated with the active pane's data. Amongst other things, you use this table to specify the *Group* for each sample. This table is described fully in the Samples Table section of this document in section 4.3.

2.3.2 Show Excluded Peaks

This command displays the table of peaks that have been excluded from analysis for the active pane's data. If no peaks have been excluded the table will be empty. This table is described fully in the Excluded Peaks Table section of this document in section 4.7.

2.3.3 Show Interest List

This command displays the table of peaks that you have manually indicated as being of specific interest for the data associated with the active pane. If you have not marked any as being of interest, the table will be empty. This table is described fully in the Interest List section of this document in section 4.8.

2.3.4 Show Peak Info

This command displays a table showing information for each peak. For example the average and standard deviation of the responses for each peak (for all of the samples) is shown and for LC/MS and MS data the charge state, isotope status and mass defect are shown. This table is fully described in section 4.9.

2.3.5 Show Peaks Table

This command displays the Peaks Table which is initially displayed when new data is processed. This is useful if you have deleted the original Peaks Table, perhaps after performing a PCA or t-test analysis and closing the original window.

2.3.6 Show Scores Table in New Window

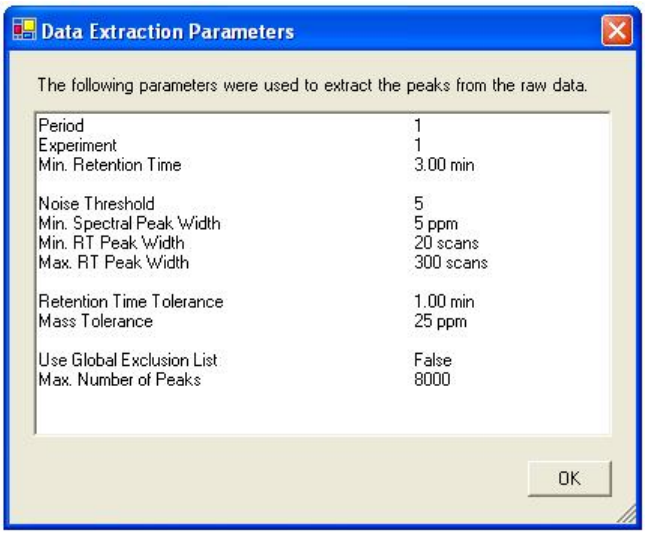
This command displays a table containing PCA scores for each sample in a new window. The command is available when any pane containing PCA data is active (Scores and Loadings Tables and Plots). You will likely only use this command infrequently, however it can be useful if you have deleted the Scores Table from the current PCA display.

2.3.7 Show Loadings Table in New Window

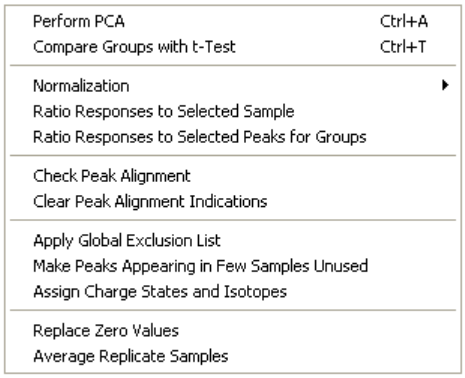
This command displays a table containing PCA loadings for each peak in a new window. The command is available when any pane containing PCA data is active (Scores and Loadings Tables and Plots).

2.3.8 View Data Extraction Parameters

This command displays a dialog showing the parameters which you used when finding and aligning peaks for the active pane's data. An example is shown below.



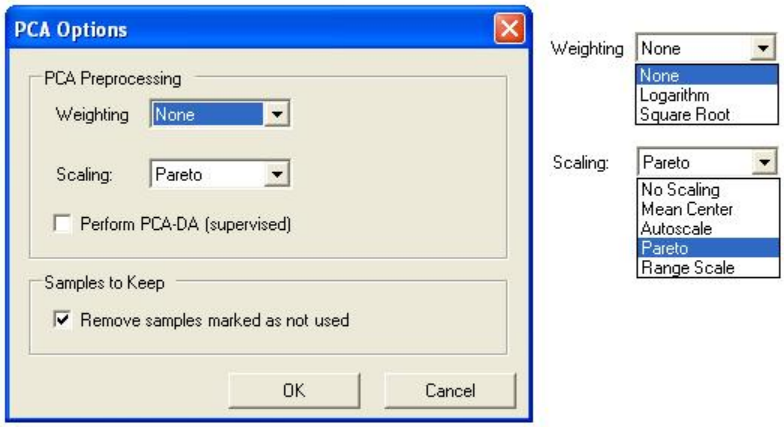
2.4 Analyze Menu



2.4.1 Perform PCA

This command processes the active pane's data using a technique known as principal components analysis (PCA). After selecting the command, you are presented with the dialog described immediately below which allows you to specify various processing options. The resulting display of Scores and Loadings Tables and Plots is described further below.

When you select the command the following dialog is shown.



The dialog contains the following items:

- *Weighting* – This combobox allows you to take either the base-e logarithm or the square root of the peak responses. This step happens before any scaling described below. Most often no weighting is used, however log or square-root weighting can be useful, especially if you use the *No Scaling* option. When using the log option, if the peak response is exactly zero a value of 1.0e-9 is used since log(0) is undefined.
- *Scaling* – This combobox specifies how the peak responses are scaled before PCA is performed. Different scaling methods can reveal different features of the data, so it is recommended that you experiment with this option. The following table describes your available choices:

No Scaling	No scaling of the peak responses is performed. This has the disadvantage that the PCA analysis is largely dominated by the most intense peaks in the data at the expense of more subtle differences. However this option can still be useful to initially focus on major differences.
Mean Center	For each peak, the average response for all samples is calculated and then subtracted from each individual value.
Autoscale	For each peak, the average response for all samples and its standard deviation are calculated. Each value is then subtracted by the average and divided by the standard deviation. This is probably the most common PCA scaling technique since it allows variables of very different response (even ones measured using completely different units) to be compared on an equal footing. In some cases this option may give good separation of samples in the resulting Scores Plot, however the Loadings Plot can often be difficult to interpret since the most important variables do not necessarily stand out.
Pareto	This option is similar to autoscaling, however each value is subtracted by the average and divided by the *square root* of the standard deviation. This is a good initial choice for mass spectrometry data since it prevents intense peaks from completely dominating the PCA analysis, but also allows peaks with good signal/noise to have more importance.
Range Scale	Scales each peak response to a value between zero and one (*i.e.* for each peak the sample with minimum response has its value set to zero, that with maximum response to one).

- *Perform PCA-DA (supervised)* – Check this item to perform PCA discriminant analysis. PCA-DA is a so-called supervised technique meaning that it uses class information – based on the assigned *Sample Group* for the samples – to improve sample separation. Effectively extra weighting is placed on peaks which are best able to separate the known groups.

  Selecting this option will usually improve sample separation in the resulting Scores Plots, sometimes dramatically. However you should be aware that this improved separation *may* be based on noise, *i.e.* on peaks which are randomly more intense in one group as compared to another – you should closely examine any variables in the Loadings Plots responsible for the separation. The analysis is also adversely affected by samples which are incorrectly assigned or are otherwise outliers. For this reason we recommend performing normal PCA first to identify (and correct or remove) any such samples followed by PCA-DA if necessary to improve the separation.

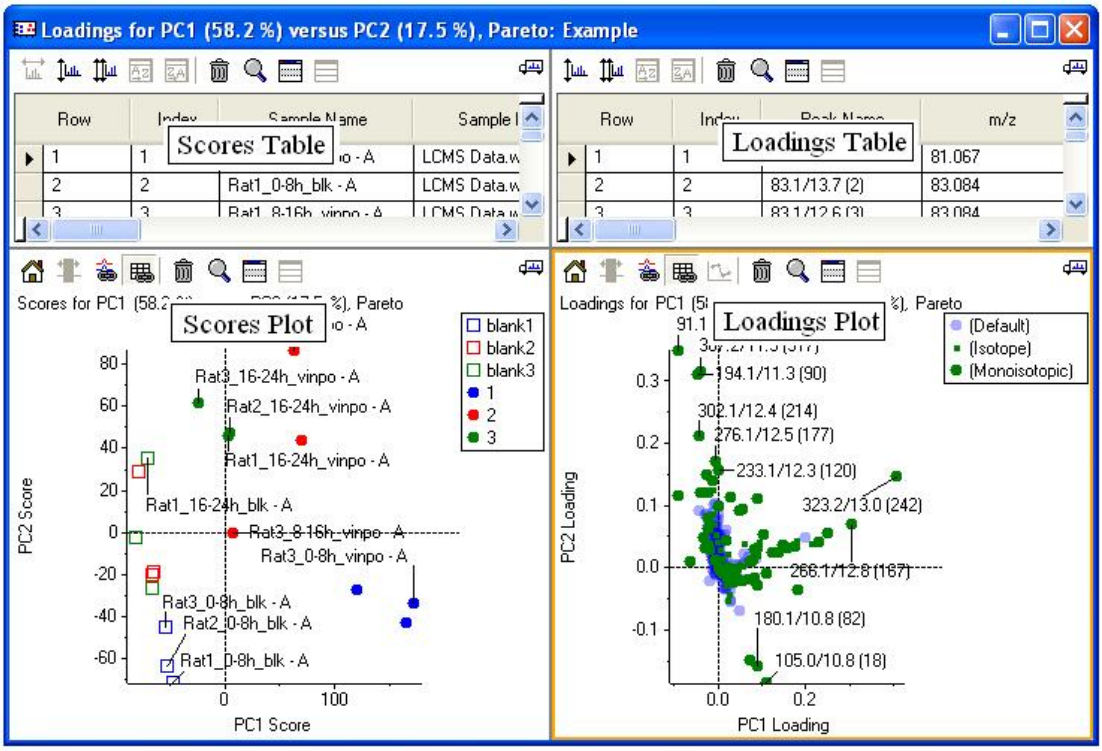
  Note that the PCA-DA option is only available if you have defined at least two different sample groups or classes.

- *Remove samples marked as not used* – If you have unchecked the *Use* flag for one or more samples, this checkbox determines how these samples are treated. If you select this option the unused

samples will be removed completely and will not appear in the subsequent Scores Table; otherwise these samples will appear in that table, but are not used for the actual PCA calculation itself. The latter is useful if you have samples of unknown class and wish to see how they group as compared to other samples of known class, but without affecting the PCA calculation.

Once you click the *OK* button in the PCA dialog, a new window is created containing a Scores Table, a Scores Plot, a Loadings Table and a Loadings Plot as shown in the figure below. If you have a large number of samples or a very large number of peaks, it may take several moments before the window appears.

Note: In the degenerate case that there is only one principal component of any significance, 1D versions of the Scores Plot and Loadings Plot appear rather than the 2D versions shown in the figure. This will happen if you attempt PCA with a data set containing only two samples or perform PCA-DA using exactly two classes.



The Scores Table shows the PCA scores for the samples and the Scores Plot initially plots the scores for the first principal component against the second. Similarly the Loadings Table shows the PCA loadings for the peaks and the Loadings Plot initially plots the loadings for the first PC against the second.

By default these tables and plots are all linked together. For example if you select the PC1 and the PC3 columns of the Scores Table, the corresponding columns of the Loadings Table are automatically selected; and both the Scores Plot and the Loadings Plot update to plot the first PC against the third.

Note: It is important to understand that the data set used for the new window is independent of the original data set used for the calculations. For example if you modify the sample names or groups or uncheck the *Use* flag for samples or peaks, these changes only affect the data set for the new window (as well as any subsequent PCA analyses), but not the Peaks Table or other table used to generate the PCA analysis.

These individual tables and plots are all described fully elsewhere in this document – see the appropriate sections for details.

2.4.2 Compare Groups with t-Test

This command performs a standard t-test for comparing one group of samples to another for the active pane's data. For a given number of samples, the calculated t-value can be converted to a probability that any difference between the two groups of samples is statistically significant (*i.e.* is not simply due to chance).

The program calculates a separate t-value for each peak. For a given peak, the *square* of the t-value for comparing group 1 to group 2 is given as:

$$t^2 = (\langle R_1 \rangle - \langle R_2 \rangle)^2 / (\sigma^2/n_1 + \sigma^2/n_2)$$

where $\sigma^2 = [(n_1 - 1)\ \sigma_1^2 + (n_2 - 1)\ \sigma_2^2] / (n_1 + n_2 - 2)$

and $\langle R_1 \rangle$ and $\langle R_2 \rangle$ are the average responses for the two groups (for the peak), n_1 and n_2 are the number of samples for the groups and σ_1 and σ_2 are the standard deviations of the responses. The quantity σ is sometimes referred to as the pooled standard deviation.

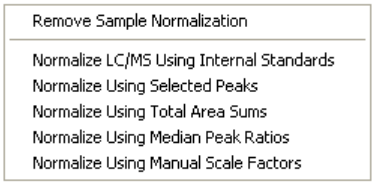
Before selecting the command, you should have previously assigned a group or class to each sample using the Samples Table. When you select the command, the following dialog is presented. After confirming the dialog a t-Test Table is created; this table is discussed in detail in section 4.4. *Note:* If you do not see the dialog, you have previously selected the *Only show this dialog again if the shift key is down* option; hold the Shift key while re-selecting the command.

The dialog contains the following items:

- *Samples per group for 'first to last' comparison* – A t-test is automatically calculated for each possible group comparison: if there are more than two groups, each possible pair-wise comparison is calculated and each group is also compared to all other samples collectively. In addition, the very first acquired samples are compared to the last acquired samples; this comparison is intended to help discover potential drift variables so that they can be removed or ignored. You specify the total number of samples to use for these special 'first' and 'last' groups. Note that this option is not available if there are less than four samples in total.
- *Remove samples marked as not used* – If you have unchecked the *Use* flag for one or more samples, this checkbox determines how these samples are treated. If you select this option the unused samples will be removed completely; otherwise these samples will still be present in the new data set used by that table, but are not used for the actual t-test calculation itself. This can occasionally be useful if you wish to use Profile Plots to visualize how potential outlier samples compare to others, but without having them affect the calculation.
- *Only show this dialog again if the shift key is down* – If you habitually use the same options, you can enable this checkbox so that the dialog is not shown each time you perform an analysis. As the wording implies, you can hold the Shift key while selecting the *Compare Groups with t-Test* command to view the dialog if you have previously enabled this option.

It is important to understand that the data set used for the new window is independent of the original data set used for the calculations. For example if you modify the sample names or groups or uncheck the *Use* flag for samples or peaks, these changes only affect the data set for the new window (as well as any subsequent analyses), but not the Peaks Table or other table used to generate the t-test.

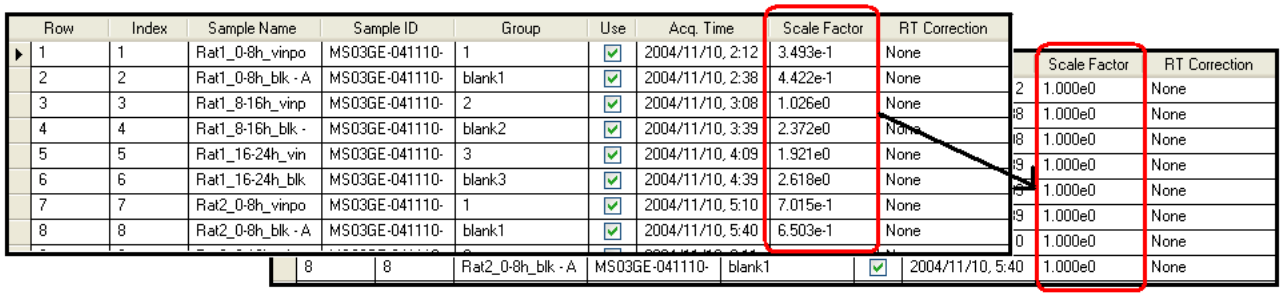
2.4.3 Normalization



2.4.3.1 Remove Sample Normalization

This command removes any normalization or scale factors which have been previously applied to the samples for the active pane's data. You can calculate and apply these normalization factors automatically when samples are first imported into the program, or you can use other commands available in the *Normalization* menu which operate on the active Peaks Table.

When you select the command, a new Peaks Table is created with the normalization factors all set to 1.0. You can verify this by viewing the *Scale Factor* column of the Samples Table for the new data as shown in the figure below. The responses for the various peaks will also be adjusted to reflect the fact that the normalization factors have been removed.



Row	Index	Sample Name	Sample ID	Group	Use	Acq. Time	Scale Factor	RT Correction
1	1	Rat1_0-8h_vinpo	MS03GE-041110-	1	☑	2004/11/10, 2:12	3.493e-1	None
2	2	Rat1_0-8h_blk - A	MS03GE-041110-	blank1	☑	2004/11/10, 2:38	4.422e-1	None
3	3	Rat1_8-16h_vinp	MS03GE-041110-	2	☑	2004/11/10, 3:08	1.026e0	None
4	4	Rat1_8-16h_blk -	MS03GE-041110-	blank2	☑	2004/11/10, 3:39	2.372e0	None
5	5	Rat1_16-24h_vin	MS03GE-041110-	3	☑	2004/11/10, 4:09	1.921e0	None
6	6	Rat1_16-24h_blk	MS03GE-041110-	blank3	☑	2004/11/10, 4:39	2.618e0	None
7	7	Rat2_0-8h_vinpo	MS03GE-041110-	1	☑	2004/11/10, 5:10	7.015e-1	None
8	8	Rat2_0-8h_blk - A	MS03GE-041110-	blank1	☑	2004/11/10, 5:40	6.503e-1	None

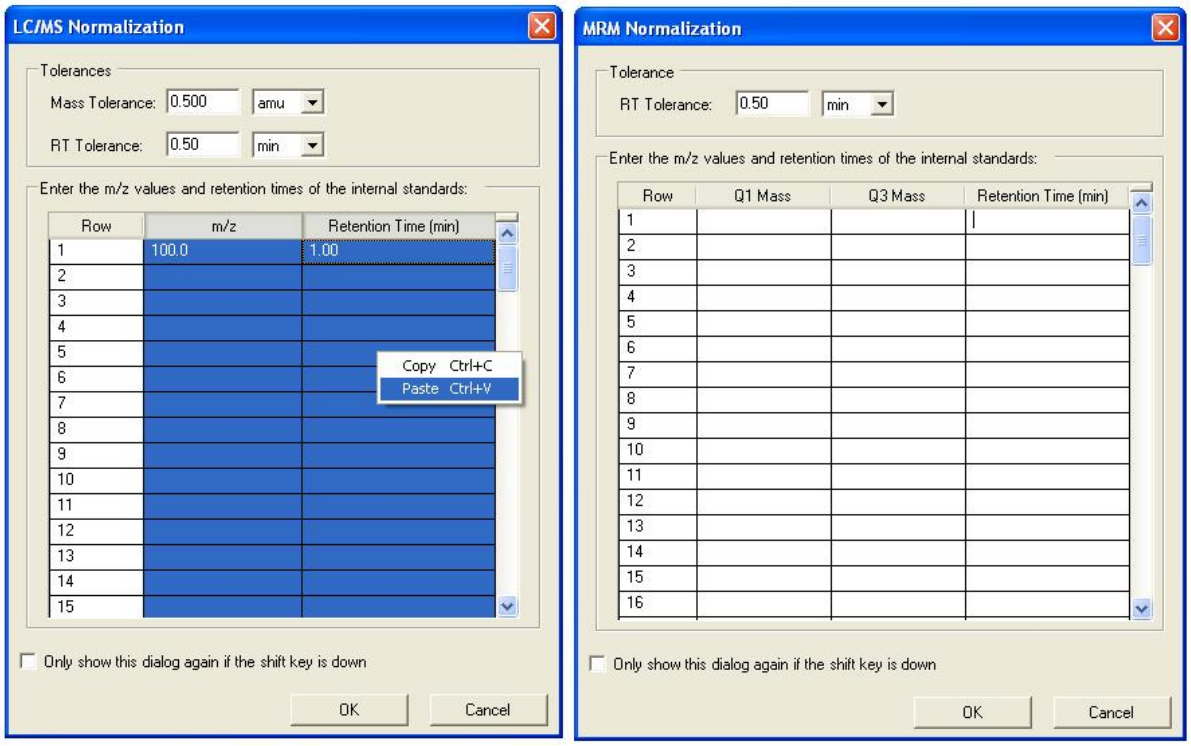
	Scale Factor	RT Correction
2	1.000e0	None
8	1.000e0	None
8	1.000e0	None
9	1.000e0	None
	1.000e0	None
9	1.000e0	None
0	1.000e0	None
8	1.000e0	None

2.4.3.2 Normalize LC/MS Using Internal Standards

This command uses the specified internal standards to calculate and apply new normalization factors for the samples for the active pane's data. This command is useful if you did not normalize the data at the time it was originally imported, or if you wish to normalize using different internal standards.

When you select this menu item one of the dialogs shown below is presented. The left dialog is shown for the usual case of LC/MS data and the right dialog for the case of MRM data. *Note:* If you do not see the dialog, you have previously selected the *Only show this dialog again if the shift key is down* option; hold the Shift key while re-selecting the command.

All of the items in these dialogs are identical to those discussed under *RT Correction* in section 2.1.2.1.2 (for LC/MS) and 2.1.2.6.1 (for MRM) – see those sections for details. In some cases you may wish to use the same internal standards for both retention time correction and for normalization, however in other cases you will want to use different ones. For example for RT correction the internal standards do not need to be present at a constant concentration for all samples, however that is not the case for normalization. If you do wish to use the same internal standards for both purposes, note that you can select columns in the Normalization spreadsheet and use Control-C and Control-V to copy and paste to the RT Correction spreadsheet (or vice versa); or you can use the context menu shown in the figure.



Once you confirm the above dialog a new Peaks Table containing the re-normalized responses is created.

The following procedure is used to calculate the normalization factors:

- Internal standards are located in each of the samples. If more than one peak is within the specified mass and retention time tolerances for a given internal standard, the peak with the largest response in any sample is assumed to represent the internal standard.
- For each internal standard, a normalization factor for each sample is calculated as:

  (Average response for IS for all samples) / (Response for current sample)

  If the particular internal standard does not appear in the sample (*i.e.* it has a response of zero in the Peaks Table), a normalization factor of 1.0 is used.
- The final normalization factor for each sample is the average of the normalization factors for all internal standards.

Note that this command does *not* simply divide the response for each peak by the response for the internal standards. By calculating normalization factors which should generally be close to 1.0 as discussed above, the magnitudes of the responses should not change drastically. After selecting this command, it is a good idea to view the *Scale Factors* column in the Samples Table for the new data to verify that they are all reasonably close to 1.0.

2.4.3.3 Normalize Using Selected Peaks

This command uses the peaks selected in the active Peaks Table to calculate and apply new normalization factors. When you select the command a new Peaks Table containing the re-normalized responses is created.

Before selecting the command, you should select at least one row or peak in the active Peaks Table. These peaks should represent components which you believe should theoretically have constant response for all of your samples; these peaks should also have reasonably good intensity, without being saturated.

The normalization factors are calculated as discussed in the previous section, except that the searching the Peaks Table is not required since you have explicitly indicated which peaks to use. As mentioned there, it is a good idea to view the new *Scale Factors* in the Samples Table for the new data to verify that they are all reasonably close to 1.0.

2.4.3.4 Normalize Using Total Area Sums

This command is used to normalize each sample so that the resulting normalized samples have the same area sum calculated using all peaks. For each sample the total response, ∑area, is calculated where the sum is over all peaks. The scale factor for a given sample is the average total response for all samples divided by the total response for the given sample.

Note that this command does *not* simply divide the response for each peak by the total response for the sample. By calculating normalization factors which should generally be close to 1.0, the magnitudes of the responses should not change drastically.

This normalization technique is useful when all major peaks for all samples are expected to be in common with similar intensity. This technique is not applicable if this is not the case; for example for metabolite ID work, dosed animals may have large xenobiotic metabolite peaks which are not present in the controls samples, and hence potentially very different area sums.

2.4.3.5 Normalize Using Median Peak Ratios

This command is used to normalize each sample to one particular reference sample using the median of the peak areas for the peaks which the two samples have in common. The underlying assumption is that all samples contain many 'background' peaks in common with the reference sample – if this is not the case this command is not applicable.

Before selecting the command you should select the column in the Peaks Table corresponding to the reference sample. For each sample the following calculations take place:

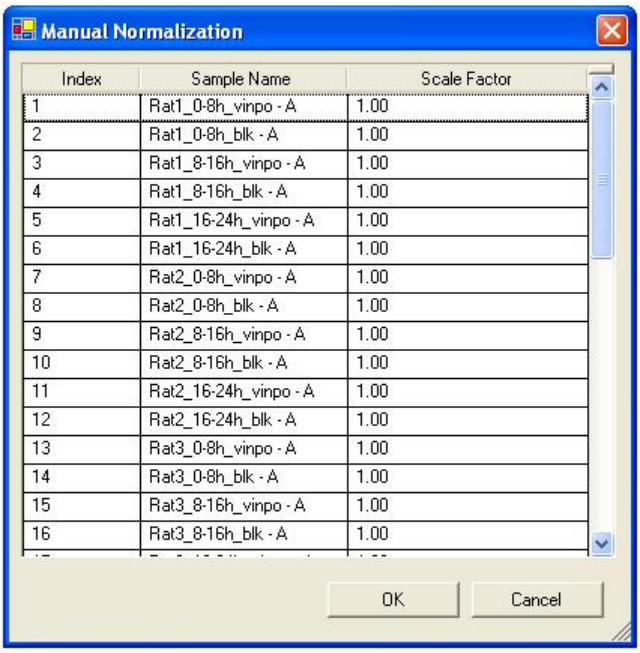
- The ratios of the peak areas (for the sample and the reference sample) are calculated for all peaks which are greater than 1% of the largest peak for each sample.
- The median ratio from the previous step is used as the scaling factor for the sample.
- However if there are less than ten peaks in common between the two samples, normalization is not performed and the scale factor is set to 1.0.

2.4.3.6 Normalize Using Manual Scale Factors

This command allows you to enter specific normalization factors for the samples. This should not be required if the samples contain internal standards detected during the MS run (see sections 2.4.3.2 and 2.4.3.3), however it is useful if you want to normalize using a non mass-spec based analysis.

Once you select the command the dialog shown in the figure below is presented. You can type directly into the *Scale Factor* column or you can select the entire column and hit Control-V to paste from the clipboard (or use the *Paste* command from the context menu).

Note that these are *multiplicative* scale factors. For example if a total creatinine analysis shows that one sample has a response of 3 units and another sample of 5 units, you should *not* enter 3 and 5 as the scale factors, but rather 1/3 and 1/5 (0.33 and 0.20).



2.4.4 Ratio Responses to Selected Sample

This command creates a new Peaks Table for which the absolute peak responses have been replaced by their ratio to a specified sample. This can be useful to directly examine 'fold changes' with respect to a reference sample. Note that the t-test Table also reports average fold changes between the sample groups.

Before selecting the command you should select the column in the Peaks Table corresponding to the reference sample. Note that if the response for a given peak for the reference sample is exactly zero, the ratios for this peak are reported as zero.

2.4.5 Ratio Responses to Selected Peaks for Groups

This command creates a new Peaks Table for which the absolute peak responses have been replaced by their ratio to specific peaks. Each set of peaks with the same *Group* name are processed separately and the ratio is calculated using the selected peak within the group. (If more than one peak is selected for a given group the average response for these selected peaks is used.) In order to use this command you must select at least one peak for each of the different groups.

This command is most useful when MRM data has been analyzed and in particular for proteomics applications which follow the naming convention mentioned in section 2.1.2.6. For example if you have a number of analyte peaks where the name is of the form:

{Protein Name}.{Peptide Sequence}.{Transition}.Light

and internal standards of the form:

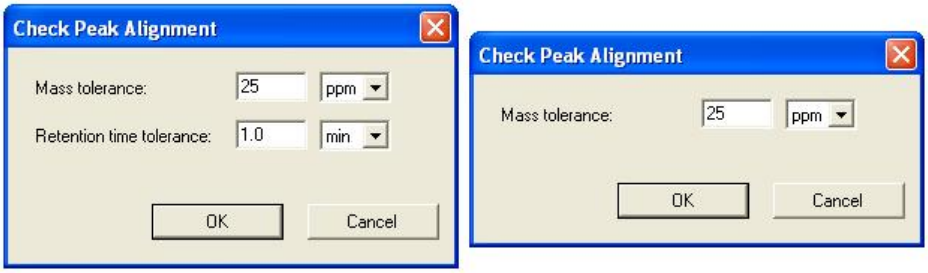
{Protein Name}.{Peptide Sequence}.{Transition}.Heavy

you could first select all of the internal standards and then create a new table in which the ratio of each analyte peptide transition is calculated to its corresponding internal standard. If there were many groups selecting the IS from each group manually would be tedious, however you can use the *Select Peaks for Matching Peak Names* menu item discussed in section 4.2.2.8 (in this example you would specify a pattern of '*.Heavy').

2.4.6 Check Peak Alignment

This command is used to indicate peaks which *may* have been improperly aligned if the original mass tolerance or (for LC/MS data) retention time tolerance were too small. Any peaks within new, larger, tolerances are highlighted in boldface.

When you select the command, the dialog shown below on the left is presented if LC/MS data is active and the one on the right if MS data is active. Specify a mass tolerance and (if appropriate) a retention time tolerance; these tolerances are exactly analogous to the peak alignment tolerances which you supplied when importing the original data. You will want to specify a mass or retention time tolerance (or both) which is somewhat larger than the original alignment tolerance you used when importing the data. Note that you can view the original alignment parameters using the *View Data Extraction Parameters* command discussed is section 2.3.8.



Once you click *OK* in one of the above dialogs, any rows of the active Peaks Table will be redrawn using boldface if there is at least one other peak within the specified tolerance. In the figure below you can see that the peaks in rows six and seven both have similar masses and retention times. They appear in the table as two separate peaks because the original tolerances were sufficiently tight that they were not merged.

You may want to select the rows containing the potentially misaligned peaks and generate an overlaid Profile Plot to help you decide if the peaks are really misaligned or simply represent two distinct components: if for many of the samples peaks with non-zero response appear for one or the other peak, but not for both, this is an indication that they are likely misaligned. You may wish to de-select the *Use* flag for any such misaligned peaks, although if there are many such cases you will probably want to re-import the data using more generous alignment parameters.

Row	Index	Peak Name	m/z	Ret. Time	Group	Use	Rat1_0-8h_vinpo -	R
1	1	81.1/13.3 (1)	81.067	13.32		☑	0.000e0	0.0
2	2	83.1/13.7 (2)	83.084	13.65		☑	0.000e0	0.0
3	3	83.1/12.6 (3)	83.084	12.62		☑	0.000e0	0.0
4	4	85.0/12.5 (4)	85.027	12.51		☑	0.000e0	0.0
5	5	89.0/11.5 (5)	89.036	11.51		☑	1.381e1	1.6
6	**6**	**90.6/11.3 (6)**	**90.616**	**11.28**		☑	**0.000e0**	**0.0**
7	**7**	**91.1/11.3 (7)**	**91.050**	**11.29**	**(Monoisotopic)**	☑	**3.844e2**	**4.0**
8	8	91.1/14.6 (8)	91.051	14.63		☑	0.000e0	0.0

2.4.7 Clear Peak Alignment Indications

This command clears the boldface indications from the active Peaks Table after a previous invocation of the *Check Peak Alignment* command discussed above.

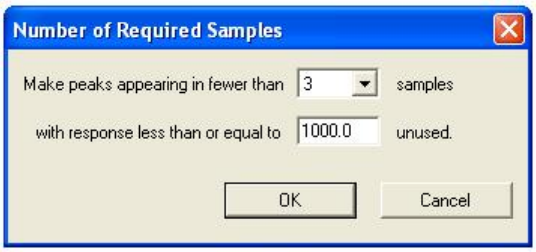
2.4.8 Apply Global Exclusion List

This command unchecks the *Use* checkbox for each peak in the active Peaks Table which is currently present in the global LC/MS or spectral Exclusion list. The global exclusion list is discussed in section 2.2.8.2. This command is useful if you did not specify the exclusion option at the time the data was originally imported or if the exclusion list has subsequently been expanded.

2.4.9 Make Peaks Appearing in Few Samples Unused

This command unchecks the *Use* checkbox for those peaks in the active Peaks Table which only appear in a small number of the samples. This allows you to ignore these peaks in a subsequent PCA or t-test analysis. Depending on the number of replicates and other details of the experimental design, such peaks may be noise, potentially misaligned or very specific to individual samples (but not to their class as a whole).

When you select the command the dialog shown below is presented. Specify the minimum number of samples for which the specified response is required in order for a given peak to remain included. For the example shown in the figure, peaks must appear in three or more samples with a response of more than 1000 in order not to be marked as unused.



2.4.10 Assign Charge States and Isotopes

When MS or LC/MS data is processed by the current version of the program, it automatically attempts to assign the charge state and isotope status for each peak (however it is possible to disable this feature as explained in section 2.1.2.1). The details of this process are explained in the discussion of the Peak Info Table in section 4.9. If you open data processed with version 1.0 or 1.1 of the program or choose to disable automatic charge determination, this information will not be available. In this case you can choose this *Assign Charge States and Isotopes* menu command to calculate it.

Note that the quality of the assignment obtained using this command will usually be at least somewhat poorer than would be obtained by reprocessing the data from scratch. When run during the original peak finding process, the assignment algorithm makes use of information which is not available after the fact.

2.4.11 Replace Zero Values

This command is used to create a new Peaks Table in which values in the active Peaks Table having a reported area of exactly zero are replaced. This may be useful prior to exporting data to certain third-party tools which do not handle zero values well; it can also be useful for removing variance from the data set due to inconsistent peak-finding for small peaks near the signal-to-noise limit.

Peaks with zero area are reported for a sample if the peak-finder finds a peak for one or more *different* samples. Note that this does not necessarily mean that the peak is completely absent in the sample, but merely that the peak-finder did not find a peak. If the peak-finder parameters are set to reasonable values this should not happen for significant peaks, but will still happen for peaks near the signal / noise limit. If there are many such zero values you will probably want to check the peak-finding parameters or possibly consider using the *Use area integrated from raw data* option discussed in section 2.1.2.1.

The following options are available:

- *A specific value* – All zero values are replaced with the specified value.
- *The average for all non-zero used samples* – For each peak, zero values are replaced with the mean of the non-zero responses for the peak.
- *The median for all non-zero used samples* – Similar to the previous item except that the median rather than the mean is used.

- *Minimum value for all non-zero used samples* – For each peak, zero values are replaced with the minimum of the non-zero responses for the peak.
- *The average for non-zero used samples for the group (or specific value if all zero)* – For each peak and sample, zero values are replaced with the mean of the non-zero responses for the peak using the samples from the same group as the current sample. If all responses are zero for the peak and group, the specified value is used instead.
- *The median for non-zero used samples for the group (or specific value if all zero)* – Similar to the previous item except that the median rather than the mean is used.

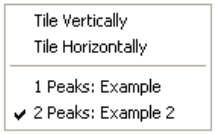
2.4.12 Average Replicate Samples

This command creates a new Peaks Table in which replicate samples in the active Peaks Table are combined. Most often you will probably *not* wish to combine replicates in this way since (for example) observing the degree of scatter of the replicates in a PCA Scores Plot can be useful, however in some cases (especially if there are many samples) you may wish to visually eliminate this additional variance.

Samples are considered to be replicates if they have the same Sample Name as one another. If the sample name ends with a '-' followed by a number, the samples are also considered to be replicates: for example samples named 'Name', 'Name-1', 'Name - 002' would all be combined. Note that if required you can adjust the sample names using the Samples Table (see section 4.3); if you wish to combine samples based on the group name, you can copy and paste the Group column into the Sample Name column from the Samples Table.

Note – After combining replicates in this way, since each sample in the resulting Peaks Table is not associated with one particular data file, it is not possible to view raw spectra or XICs from the combined Peaks Table.

2.5 Window Menu



2.5.1 Tile Vertically

This command arranges any open windows (which have not been minimized) so that they are all beside one another in one row.

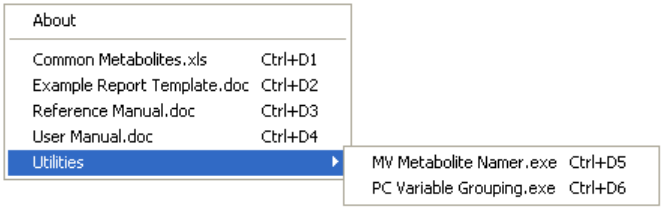
2.5.2 Tile Horizontally

This command arranges any open windows (which have not been minimized) so that they are all above or below one another in one column.

2.5.3 Other menu items for specific windows

Additional menu items appear for each open window showing the name of the currently active pane for the window. Selecting one of these menu items brings the corresponding window to the front and makes it active.

2.6 Help Menu



With the exception of the *About* item, this menu contains an entry for each file in the MarkerView™ Software's *Help* directory, placed by the installer at the AB-Sciex\MarkerView\Help sub-directory of the Program Files folder. The figure above shows the four items which are initially present in this directory – these are explained below.

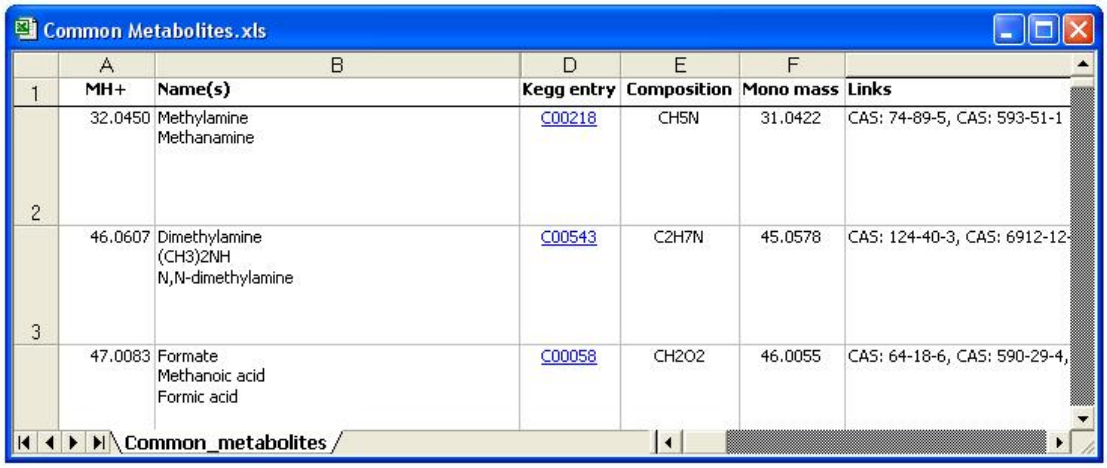
You can also place documents or directories (or shortcuts to them) into this Help folder to have them automatically appear in the menu. When you select these custom menu items, the corresponding document will be opened as if you had double-clicked it from the Windows Explorer.

2.6.1 About

This command displays an 'About' message which shows the version of the MarkerView™ Software as well as copyright and other information.

2.6.2 Common Metabolites

This command opens an Excel document containing the MH^+ and monoisotopic masses of some common human metabolites as shown in the figure.



You can click on the links in the *Kegg entry* column to open an Internet Explorer window showing the structure, pathway and other information for the compound. The links use the KEGG ligand database located at http://www.genome.jp/kegg/ligand.html.

This list is not intended to be complete, but may still be useful. After opening the document you can add new entries and re-save the document for future use.

2.6.3 Example Report Template

This command opens a copy of the document which describes the format for reports – and is itself also a report template. For details on reporting see section 2.1.12.

2.6.4 Reference Manual

This command opens a copy of this document in Microsoft Word.

2.6.5 User Manual

This command opens a copy of the User manual in Microsoft Word.

2.6.6 MV Metabolite Namer

This menu items is used to launch the 'MV Metabolite Namer' application. This is a MarkerView™ Software plug-in utility which assigns specific names to MarkerView peaks based on the mass of a parent drug and expected metabolic transformations. For more details see the discussion in section 6.2.

2.6.7 PC Variable Grouping

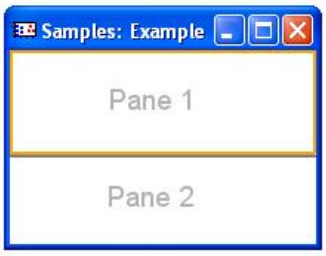
This menu items is used to launch the 'PC Variable Grouping' application. This is a MarkerView™ Software plug-in utility which with the discovery of relationships among MarkerView peaks. For details see section 6.1.

3 Panes

This section contains a discussion of features which apply to all panes, whether tables or plots.

3.1 Active Pane

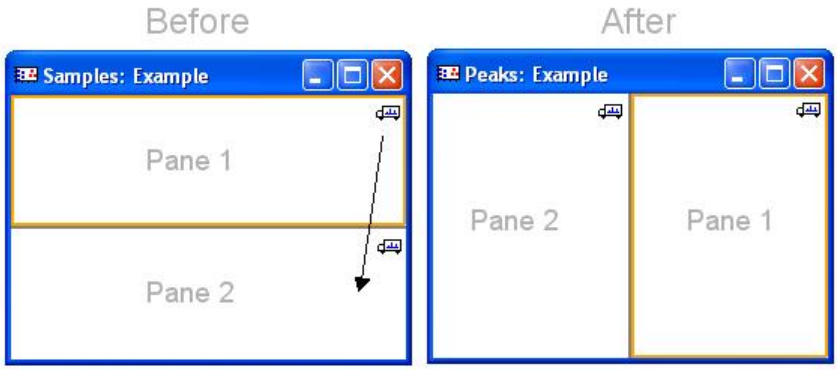
Only one pane is active at any given moment, indicated by the orange border. As shown in the figure below, Pane 1 is currently active. You can activate a pane by clicking anywhere within it. Many menu commands operate on the active pane's data.



3.2 Arranging Panes

This icon appears in the top right corner of each pane and is used to change the relative positions of the panes. You click on the icon in one pane and drag it to the top, bottom, left or right portion of a second pane: depending on where you release the mouse, the first pane will change positions relative to the second. As you drag, one side of the second pane is highlighted in red to indicate the where the first pane will be drawn. The figure below shows the results of dragging the truck from the top pane to the right portion of the bottom pane.

You can also drag panes from one window to another, for example to view two PCA Scores Plots side-by-side using different processing options.



3.3 Status Bar

When any pane is active, the status bar at the bottom of the window (shown in the figure below) updates to show information for the pane's data set as summarized below.

The following fields are present in the status bar:

Samples	The total number of samples present in the data set.
Peaks	The total number of peaks *currently* present. If certain peaks were excluded and a PCA or t-test analysis was subsequently performed, the excluded peaks are removed and are not counted here.
Currently Excluded Peaks	The number of peaks for which the *Use* checkbox has been unchecked.

Interest List Peaks	The total number of peaks which have been added to the interest list (see section 4.8).
Previously Excluded Peaks	The number of peaks for which the *Use* checkbox was unchecked *prior* to the creation of the current pane. These peaks no longer appear in the data set for the pane.
Globally Excluded Peaks	The total number of peaks in the global exclusion list (see section 2.2.8.2). Note that this is not a property of the active data set *per se*.

3.4 Toolbar Buttons

The toolbar buttons described in this section appear in the pane-specific toolbar for all pane types. Additional buttons may also be available, but the buttons described here are always present.

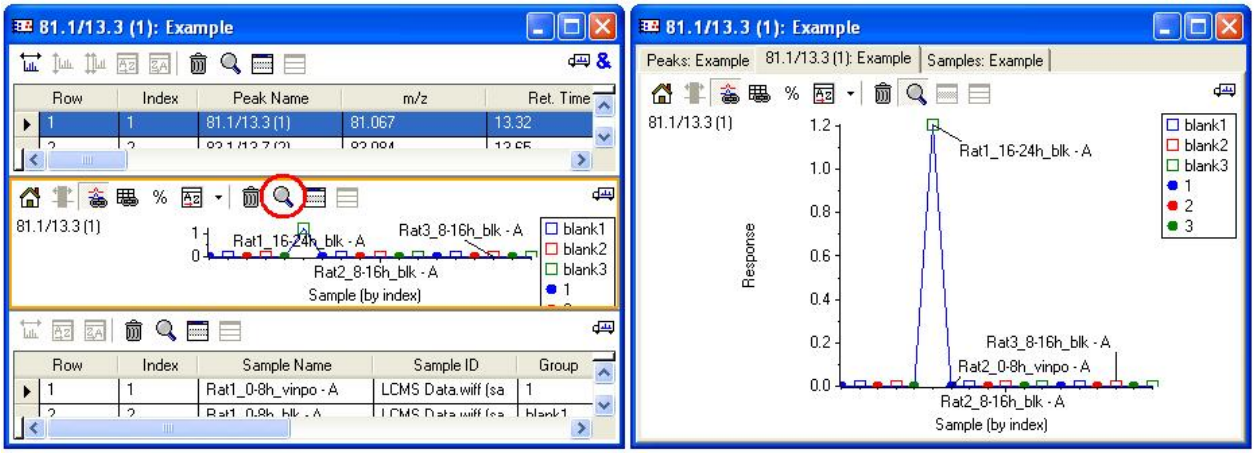
3.4.1 Delete Pane

This button deletes the pane. This is mostly used to remove ancillary plots and tables which were created from the initial table in a window when you are finished with them. If you delete the last pane in a window the entire window is closed.

3.4.2 Zoom Pane

This button causes the pane to fill the entire window (or vice versa). This is useful if there are several panes in the window so that you can temporarily focus on one in detail.

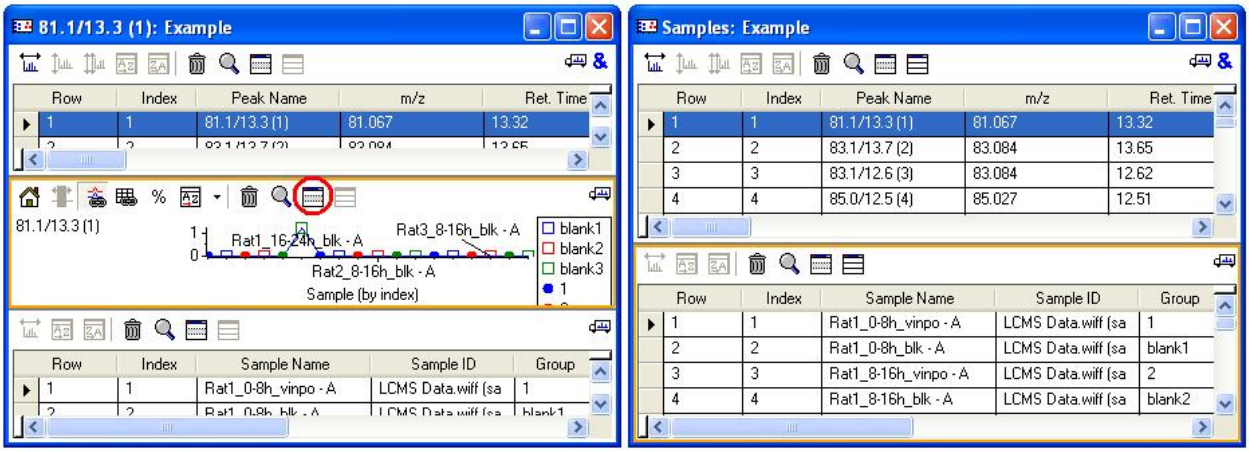
The figure on the left below shows an initial arrangement of panes; if you click *the Zoom Pane* button for the middle pane, the display changes to that shown in the right figure below. When in zoomed mode a separate tab appears at the top of the window for each pane – you switch between panes by clicking the appropriate tab. From zoomed mode, you return to the original view showing all panes by clicking the *Zoom Pane* button a second time.



3.4.3 Hide Pane

This button hides the pane so that other panes in the window fill the available space. This is useful if you wish to view a subset of the panes in some detail, but do not wish to permanently delete other panes.

The figure on the left below shows an initial arrangement of panes; if you click the *Hide Pane* button for the middle pane, the display changes to that shown in the right figure below. You can click the *Hide Pane* button for other panes to hide them too. You can return to the original view by clicking the *Show Hidden Panes* button discussed below.



3.4.4 Show Hidden Panes

This button shows all panes which have been previously hidden. As discussed in the *Hide Pane* section, you can hide one or more panes to view any remaining panes in greater detail. If you have done this, you use the *Show Hidden Panes* button to view all panes.

3.5 Context Menu

All panes contain a context menu which you obtain by clicking the right mouse button within the pane. The various items in these menus are explained in detail for each of the individual pane types in subsequent sections of this document.

4 Tables

This section first discusses general functionality which applies to all tables, followed by a detailed list of each of the various types of table.

4.1 General

4.1.1 Toolbar Buttons

Most of the toolbar buttons discussed in this section are available for all types of table panes, however certain of the buttons used to generate plots are not available for some types of tables. See the discussion of the individual table types for details.

4.1.1.1 Plot Row

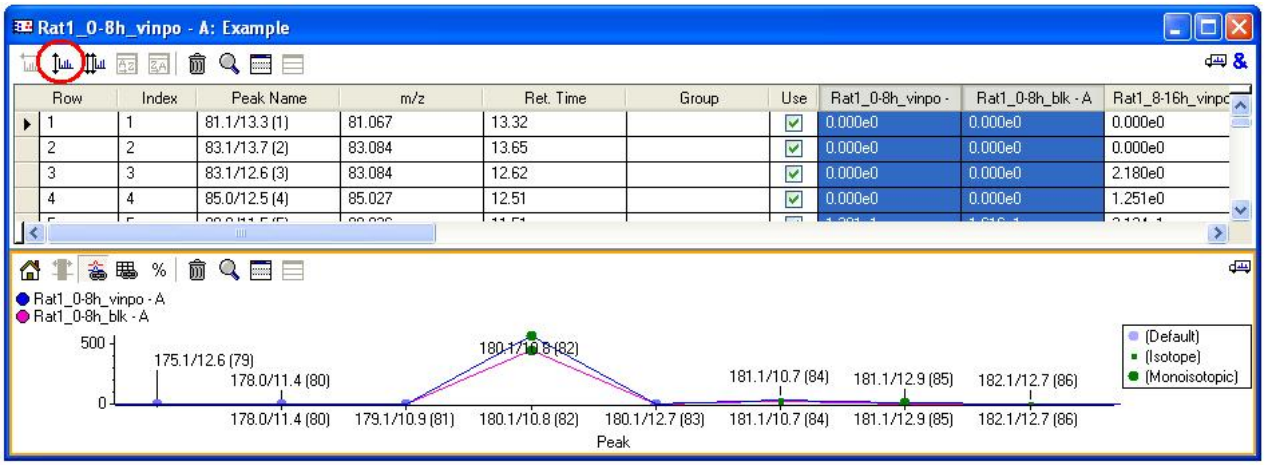
 When available, this button creates a new plot showing the values in the selected row or rows as a function of column number. The resulting plot contains an overlaid trace for each selected row.

For example when performed from a Peaks Table, clicking this button creates a Profile Plot (showing response as a function of sample) with a separate overlaid trace or series for each selected peak.

4.1.1.2 Plot Column

 When available, this button creates a new plot showing the values in the selected column or columns as a function of row number. The resulting plot contains an overlaid trace for each selected column. The plot reflects the order of the rows in the table, so if you sort the table the order of the points in the generated plot will also change.

When performed from a Peaks Table, clicking this button creates a plot of peak response for all peaks with a separate overlaid trace or series for each selected sample as shown in the figure below. (The figure shows the result after zooming the resulting plot.) When performed from a PCA Loadings Table, a plot showing the PCA loadings for each variable is created, and similarly for other table types.

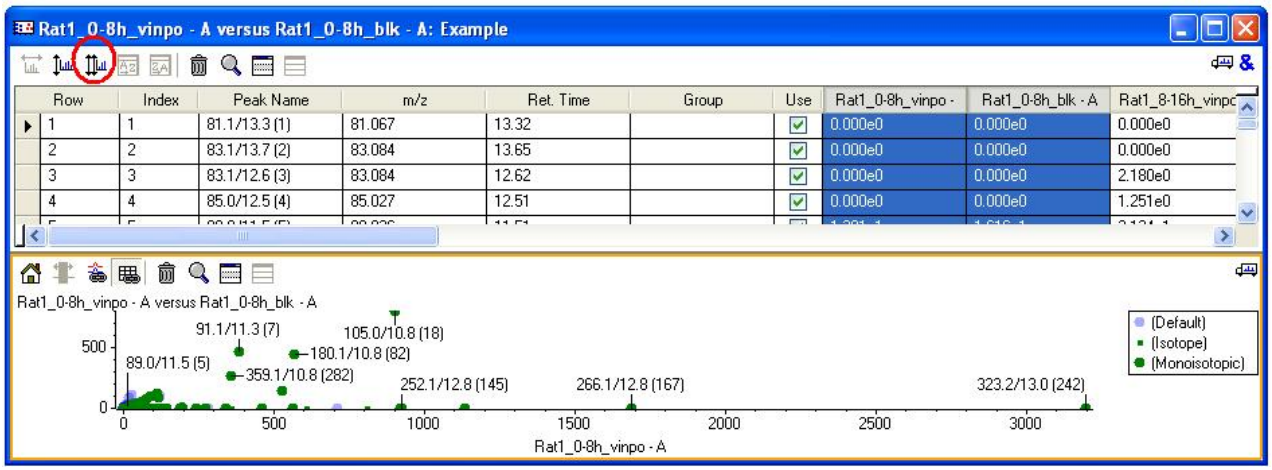


4.1.1.3 Plot Two Columns

 When available and provided that there are exactly two columns selected, this command creates a new display plotting the values in the rows for one column as a function of the other.

When performed from a Peaks Table, clicking this button creates a plot of peak response for two samples as shown in the figure below. This can be a convenient way to visually compare two samples in detail. As you can see in the figure, there are a number of peaks lying near the x-axis which appear in the first sample but not in

the second. When performed from a PCA Scores Table, a plot showing the PCA scores for the two selected principal components is created, and similarly for other table types.



4.1.1.4 Sort Ascending

This button sorts the table so that the values in the selected column are in ascending order. This button is only available when you have selected exactly one column. To remove the effect of the sort operation, select the *Index* column and click the *Sort Ascending* button again.

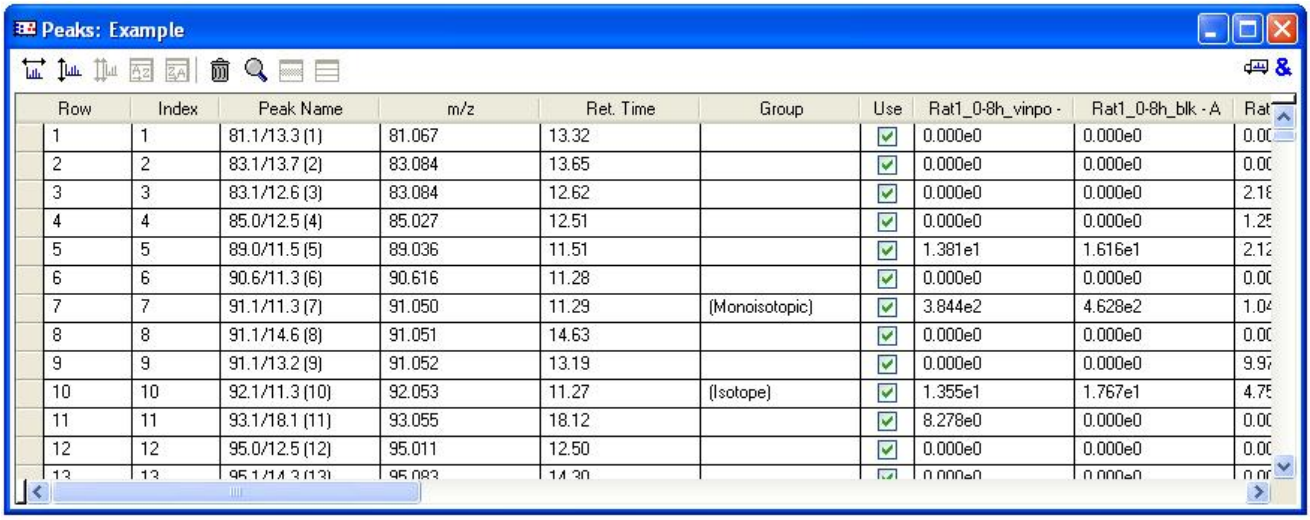
If you have previously generated any column plots *and* these plots are linked to the table, these plots will update to reflect the new sort order. See the Plots discussion in section 5 for how to link plots to tables.

4.1.1.5 Sort Descending

The button is the reverse of the previous *Sort Ascending* button: it sorts the table so that the values in the selected column are in descending order. To remove the effect of the sort operation, select the *Index* column and click the *Sort Ascending* button.

4.2 Peaks Table

After importing data or opening previously saved data, a new Peaks Table is created as shown in the figure below. This table contains a row for each detected and aligned peak, and a column for each sample containing the corresponding peak response for that sample.



The table contains the following columns:

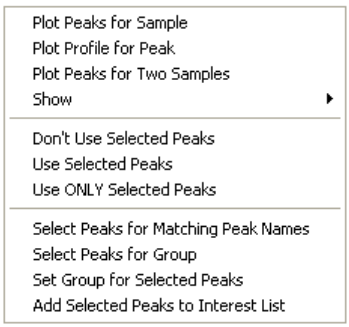
Column	Description
Row	Simply the row number for each row.
Index	The *original* row number. If you have sorted the table this will differ from the current row number. You can un-sort the table by selecting this row and clicking the *Sort Ascending* toolbar button.
Peak Name	The name of the peak for the row. For LC/MS data the name is a combination of the m/z and retention time with the row number appended. For MS data the peak name is simply the m/z.
m/z	The m/z for the peak. This is available for LC/MS or MS data processed by the program, but will display as *N/A* for text data imported using the *Generic Text File* format.
Ret. Time	The retention time in minutes for the peak for LC/MS data; for other data this will display as *N/A.*
Group	The name of the group for the peak. The group determines the plot symbol for plots involving peaks as discussed in section 2.2.8.1. When processing MS or LC/MS data, this column is automatically filled-in with an indication of the peak's isotope status (monoisotopic or isotope) or is left empty if the status could not be determined (this is common for small peaks if no corresponding isotope was detected).
Use	This checkbox allows you to specify whether the peak should be used for subsequent data analysis such as PCA and the t-test. You can disable the peak by clicking directly in the cell or in various shortcut ways discussed elsewhere in this document.
Additional columns	There is an additional column for each sample. The values in the cells for these columns show the response for the given peak and sample combination.

4.2.1 Joining Peaks Tables

The & icon which appears in the upper-right corner of Peaks Tables is used to combine the data for two tables. This allows you to acquire data for a set of samples using two or more different MS experiments and to analyze the combined data set with PCA, as opposed to analyzing each set independently. For example you could acquire both positive and negative mode data for a group of samples (either using on-line polarity switching or using separate injections), import each experiment independently and then combine the results.

You combine two tables by clicking on the icon in one Peaks Table and dragging-and-dropping it on top of a second Peaks Table: the peaks from the first table will be appended to those already present in the second table. It only makes sense to perform this operation if the two tables contain data for exactly the same physical samples in the same order.

4.2.2 Context Menu



4.2.2.1 Plot Peaks for Sample

 This command creates a plot of peak response as a function of peak index or, in the case of MS data, optionally mass/charge. There is a separate overlaid trace for each sample or column which you have selected in the Peaks Table. The resulting plot will reflect the order of the rows in the table, so if you sort the table the order of the points in the generated table will also change. For details about the resulting plot, see the Peaks for Sample Plot in section 5.3.

4.2.2.2 Plot Profile for Peak

 This command creates a plot of peak response as a function of sample. There is a separate overlaid trace in the plot for each peak or row which you have selected in the Peaks Table. These plots can be very useful for visualizing how particular peaks are behaving across the various samples – for details see the Profile Plot discussion in section 5.2. However most frequently you will probably generate these plots from PCA or t-test tables or plots, as opposed to directly from a Peaks Table, simply because if there are many peaks those displays allow you to focus on particular peaks likely to be of specific interest.

4.2.2.3 Plot Peaks for Two Samples

 This command creates a plot of peak response for two samples as was shown in section 4.1.1.3. As mentioned there, this can be a convenient way to visually compare two samples in detail. This plot type is discussed in section 5.4

4.2.2.4 Show Sub-Menu

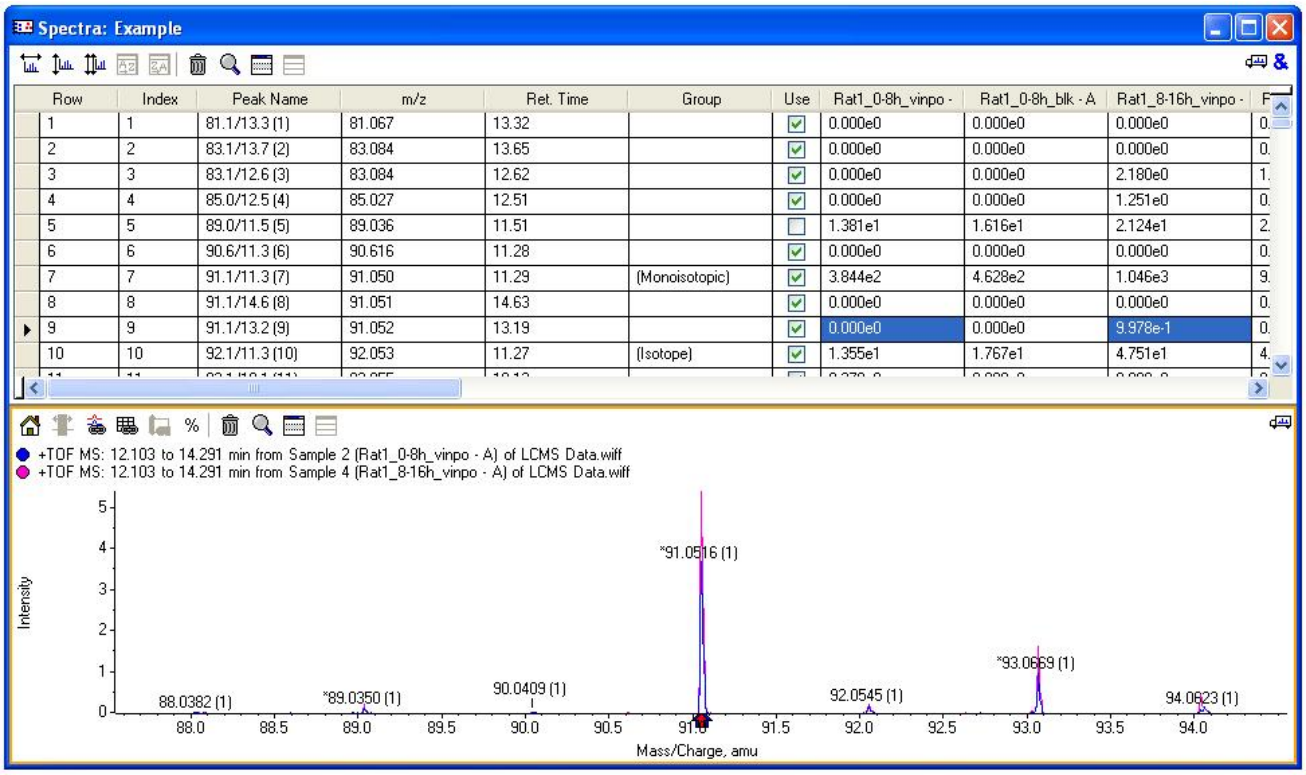
Most of the commands in this sub-menu are used to view raw data; the *Peak List Spectra* command is used to display a 'virtual' mass spectrum using the list of detected peaks. These commands are all explained in detail in the following sections.

4.2.2.4.1 Spectra

This command displays raw mass spectra. This command is available if you have processed LC/MS or MS data, but not if you have imported text data in the *Generic Text File* format. For details on the resulting spectra, see the MS Spectrum Plots discussion in section 5.12.

Most often you will probably not generate spectra directly from the Peaks Table in this way, but rather by making selections in a Profile Plot.

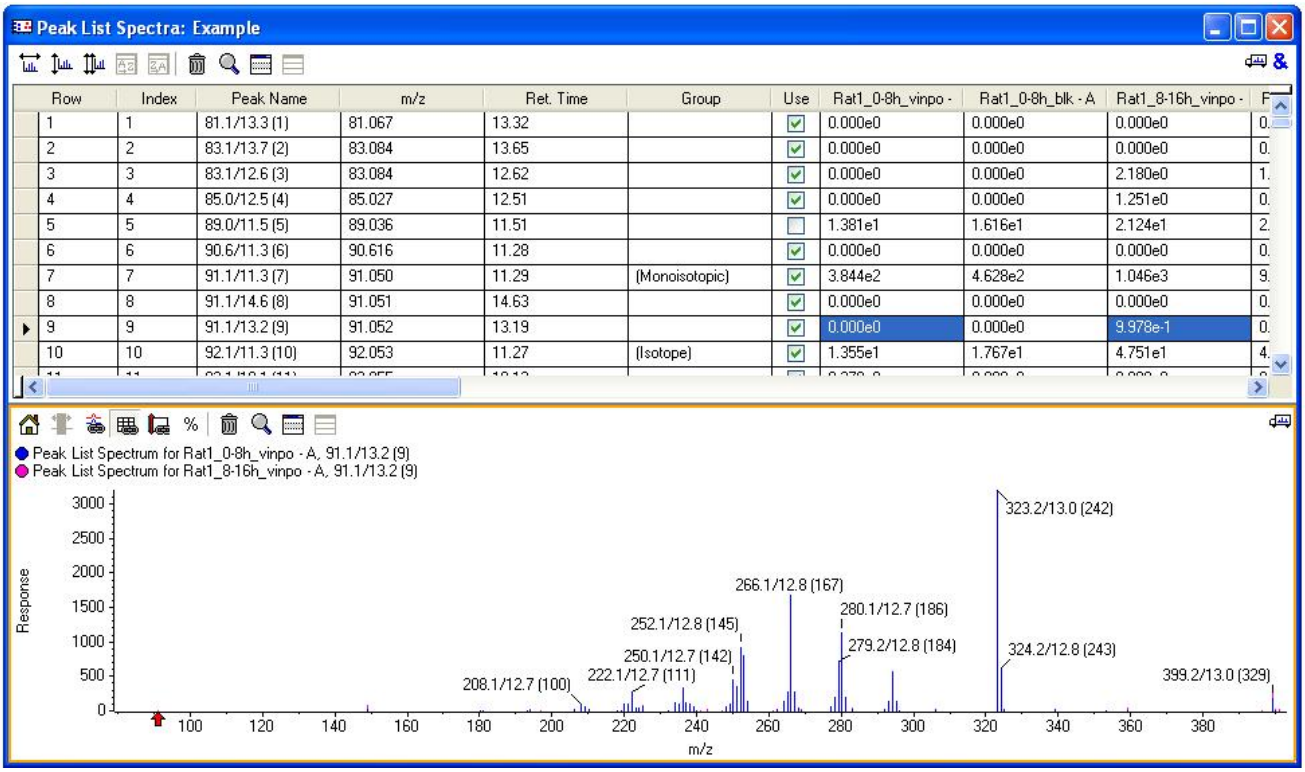
If you select an entire row in the table, a new plot containing overlaid spectra for all samples is created. If you have selected only specific samples for one peak or row, the resulting plot contains a trace for only those samples. For example in the figure below *Show Spectrum* was chosen after selecting the highlighted cells in the table: the resulting spectrum zooms for the selected peak (the ninth one) and contains a trace for just the first and third samples.



4.2.2.4.2 Peak List Spectra

This command displays so-called Peak List Spectra which are 'virtual' mass spectra constructed using the list of detected peaks as opposed to the raw profile data. They are conceptually similar to what would be obtained by showing profile mass spectra as described in the previous section and then centroiding each of those spectra. These plots are described in detail in section 5.13.

If you select an entire row in the table, a new plot containing overlaid Peak List Spectra for all samples is created. If you have selected only specific samples for one peak or row, the resulting plot contains a trace for only those samples. For example in the figure below *Show Peak List Spectra* was chosen after selecting the highlighted cells in the table: the resulting plot marks the selected peak (the ninth one) with an arrow and contains a trace for just the first and third samples. Before the spectra are displayed, a dialog appears asking for the time window (centered on the selected peak) over which the spectrum is calculated as explained in section 5.13.



4.2.2.4.3 XICs

This command displays raw extracted ion chromatograms. This command is only available if you have processed LC/MS data. For details on the resulting XICs, see the XIC Plots discussion in section 5.14.

Most often you will probably not generate XICs directly from the Peaks Table in this way, but rather by making selections in a Profile Plot.

If you select an entire row in the table, a new plot containing overlaid XICs for all samples is created. Since this may be slow, you will probably mostly want to plot XICs for only specific samples. You can do this by selecting these samples (*i.e.* individual cells in the table) for one peak or row before choosing the menu item.

4.2.2.4.4 Contour

This command shows a contour map for the selected sample. A contour plot is a type of heat map with retention time and m/z as axes and the raw intensity as a color-coded value. This command is only available if you have processed LC/MS data. For details on the resulting plot, see the Contour Plot discussion in section 5.16. These plots can be useful if you wish to experiment with different LC/MS peak finding parameters in detail since you can overlay the actual peaks found on the raw data.

4.2.2.4.5 First IDA Product Spectrum

If you have processed the MS data for a series of IDA runs, this command is used to display an IDA product spectrum for the peak corresponding to the selected row of the Peaks Table.

Most often you will probably not generate IDA product spectra directly from the Peaks Table in this way, but rather by making selections in a Profile Plot.

If you select an entire row in the table, all available samples are examined for an IDA product spectrum matching the precursor m/z and retention time of the corresponding peak; otherwise if specific samples are selected (*i.e.* individual cells in the table) only those samples are examined. The very first matching product spectrum is displayed. For details on the resulting plot, see the *IDA Product Spectrum Plot* discussion in section 5.15. If a matching product spectrum (for the samples examined) was not acquired the program displays a message stating that such is the case.

4.2.2.4.6 'Best' IDA Product Spectrum

This command is similar to the *First IDA Product Spectrum* command discussed above except that all samples are always examined for a matching product spectrum and the 'best' one is reported.

In the case that more than one matching spectrum is available, the 'best' spectrum is the one with the largest intensity sum for all fragments (*i.e.* all spectral peaks with the exception of the precursor itself). Although the quality of the resulting spectrum is likely to be higher than for the *First IDA Product Spectrum* case, it may take considerably longer to generate the spectrum since all specified samples must always be examined.

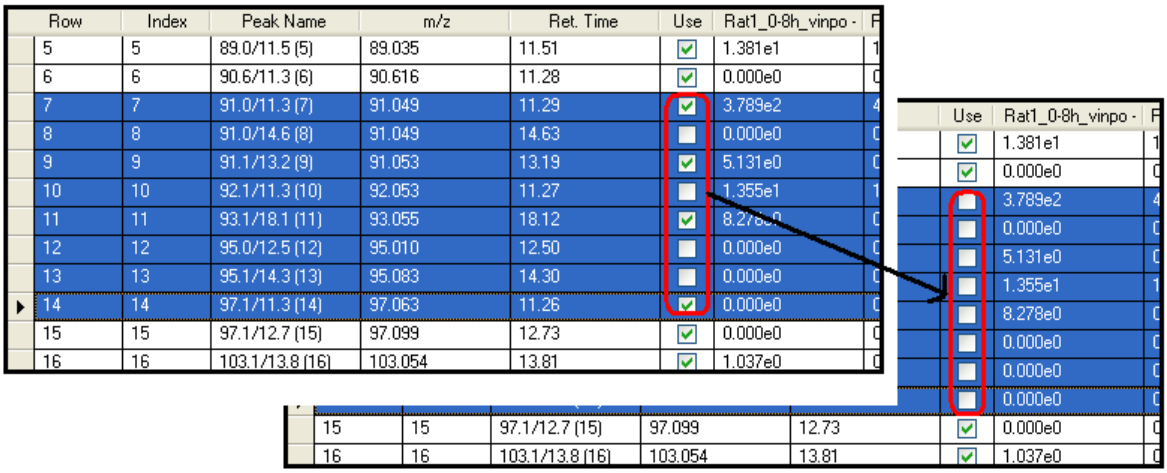
4.2.2.4.7 All IDA Product Spectra

This command is similar to the *First IDA Product Spectrum* command discussed above except that all samples are always examined for matching product spectra and all such matches are reported as overlaid traces in the resulting IDA product spectrum plot.

4.2.2.5 Don't Use Selected Peaks

This command unchecks the *Use* field for any selected peaks. If you wish to de-select this field for many peaks, this is much more convenient than manually clicking in the *Use* cell of the table for each one. Note that you can use the Control key to make disjoint selections in the table, *i.e.* all of the peaks do not need to be immediately adjacent.

The figure below shows the result of selecting this command with the selection range as shown.



Row	Index	Peak Name	m/z	Ret. Time	Use	Rat1_0-8h_vinpo -
5	5	89.0/11.5 (5)	89.035	11.51	☑	1.381e1
6	6	90.6/11.3 (6)	90.616	11.28	☑	0.000e0
7	7	91.0/11.3 (7)	91.049	11.29	☑	3.789e2
8	8	91.0/14.6 (8)	91.049	14.63	☐	0.000e0
9	9	91.1/13.2 (9)	91.053	13.19	☑	5.131e0
10	10	92.1/11.3 (10)	92.053	11.27	☐	1.355e1
11	11	93.1/18.1 (11)	93.055	18.12	☑	8.278e0
12	12	95.0/12.5 (12)	95.010	12.50	☐	0.000e0
13	13	95.1/14.3 (13)	95.083	14.30	☐	0.000e0
14	14	97.1/11.3 (14)	97.063	11.26	☑	0.000e0
15	15	97.1/12.7 (15)	97.099	12.73	☑	0.000e0
16	16	103.1/13.8 (16)	103.054	13.81	☑	1.037e0

Use	Rat1_0-8h_vinpo -
☑	1.381e1
☑	0.000e0
☐	3.789e2
☐	0.000e0
☐	5.131e0
☐	1.355e1
☐	8.278e0
☐	0.000e0
☐	0.000e0
☐	0.000e0
☑	0.000e0
☑	1.037e0

4.2.2.6 Use Selected Peaks

This command checks the *Use* field for any selected peaks. This is essentially the reverse of the *Don't Use Selected Peaks* command discussed above.

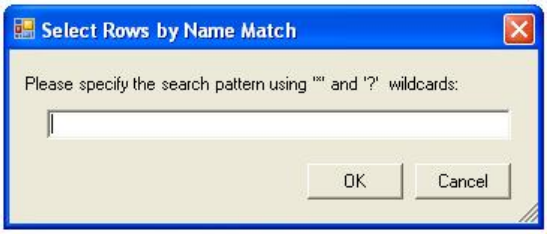
4.2.2.7 Use ONLY Selected Peaks

This command checks the *Use* field for any selected peaks *and* unchecks this field for all other peaks. This command provides a convenient way to focus on only a small subset of the peaks, most likely for subsequent PCA. For example you could sort the Peaks Table based on retention time and then focus only on peaks which eluted within a specific retention time interval.

4.2.2.8 Select Peaks for Matching Peak Names

This command is used to select all rows in the table for which the *Peak Name* field matches the specified pattern. If there are many such rows this is more convenient than manually selected each one individually. This is useful so that you can subsequently perform other operations involving these peaks.

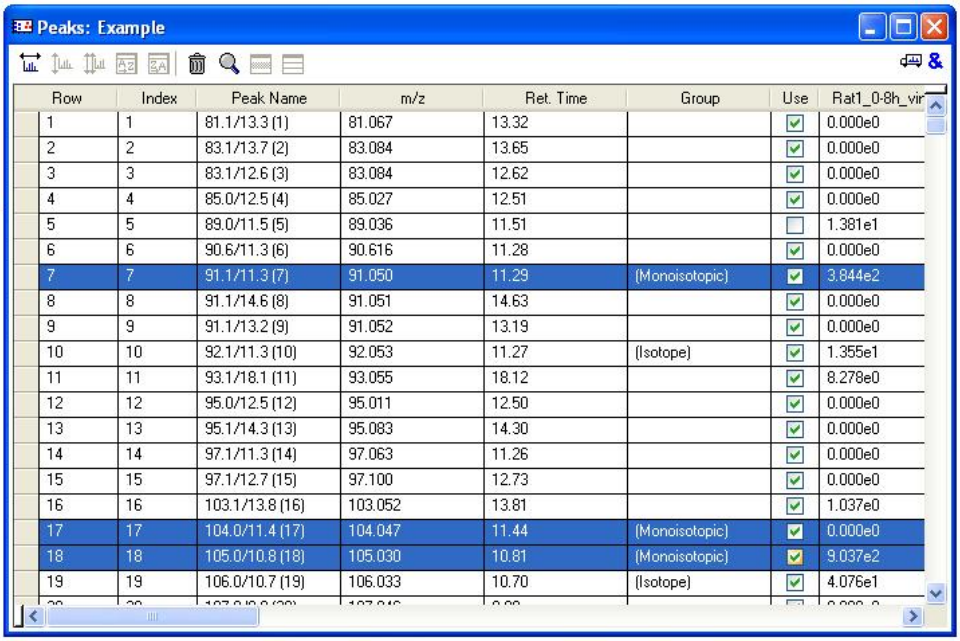
When you select the command the simple dialog shown below is presented. For example if you specify '*.Heavy' (without the quotes) any rows for which the name ends with '.Heavy' are selected; if you specify '*Heavy*' rows are selected for which the name contains 'Heavy'.



4.2.2.9 Select Peaks for Group

This command is used to select all peaks with a particular group name. If there are many such rows, this is more convenient than manually selecting each one individually. This is useful so that you can subsequently perform other operations involving these peaks. For example if you wish to exclude all peaks except those assigned as monoisotopic you can (1) use this command to select only the monoisotopic peaks and then (2) select the *Use ONLY Selected Peaks* command discussed above.

When you select the command, the simple dialog shown below is presented to allow you to select one of the group names currently in use for the table; the '(None)' item corresponds to an empty or undefined group. The figure below shows the results of selecting the '(Monoisotopic)' group.



Row	Index	Peak Name	m/z	Ret. Time	Group	Use	Rat1_0-8h_vir
1	1	81.1/13.3 (1)	81.067	13.32		☑	0.000e0
2	2	83.1/13.7 (2)	83.084	13.65		☑	0.000e0
3	3	83.1/12.6 (3)	83.084	12.62		☑	0.000e0
4	4	85.0/12.5 (4)	85.027	12.51		☑	0.000e0
5	5	89.0/11.5 (5)	89.036	11.51		☐	1.381e1
6	6	90.6/11.3 (6)	90.616	11.28		☑	0.000e0
7	7	91.1/11.3 (7)	91.050	11.29	(Monoisotopic)	☑	3.844e2
8	8	91.1/14.6 (8)	91.051	14.63		☑	0.000e0
9	9	91.1/13.2 (9)	91.052	13.19		☑	0.000e0
10	10	92.1/11.3 (10)	92.053	11.27	(Isotope)	☑	1.355e1
11	11	93.1/18.1 (11)	93.055	18.12		☑	8.278e0
12	12	95.0/12.5 (12)	95.011	12.50		☑	0.000e0
13	13	95.1/14.3 (13)	95.083	14.30		☑	0.000e0
14	14	97.1/11.3 (14)	97.063	11.26		☑	0.000e0
15	15	97.1/12.7 (15)	97.100	12.73		☑	0.000e0
16	16	103.1/13.8 (16)	103.052	13.81		☑	1.037e0
17	17	104.0/11.4 (17)	104.047	11.44	(Monoisotopic)	☑	0.000e0
18	18	105.0/10.8 (18)	105.030	10.81	(Monoisotopic)	☑	9.037e2
19	19	106.0/10.7 (19)	106.033	10.70	(Isotope)	☑	4.076e1

4.2.2.10 Set Group for Selected Peaks

This command sets the *Group* field for any selected peaks. If you wish to set this field for many peaks, this is much more convenient than manually setting each one individually by typing into the corresponding cell of the table. Note that you can use the Control key to make disjoint selections in the table, *i.e.* all of the peaks do not need to be immediately adjacent.

When you select the command, a simple dialog is presented asking you for the name of the group; this dialog is similar to that used to set the group for samples shown in section 4.3.1.7. You can either type directly into the *Group Name* dialog or use the combobox to select a group name which you have previously defined in the *Plot Symbols* tab of the *Options* as discussed in section 2.2.8.1.

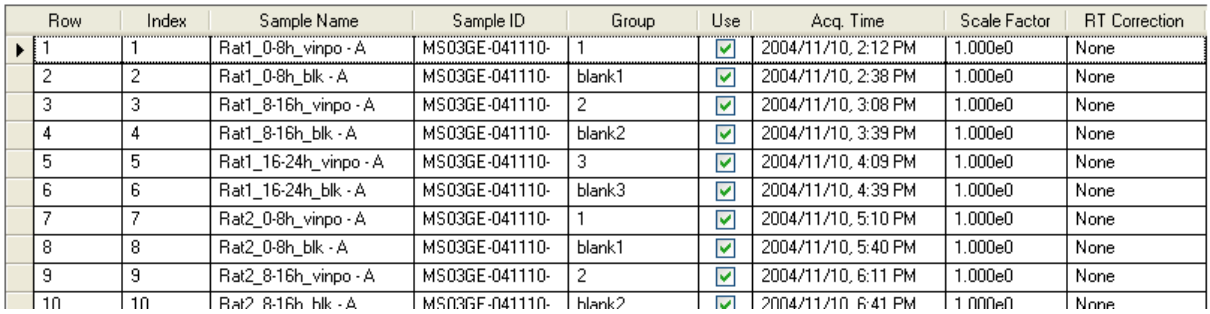
4.2.2.11 Add Selected Peaks to Interest List

This command adds the selected peaks to the Interest List. The main reason for doing this is to subsequently report details for these peaks or possibly to export this interest list for use outside of the application. For details

see the Interest List discussion in section 4.8 and also the *Generate Report in MS Word* command in section 2.1.12.

4.3 Samples Table

The Samples Table contains a row for each sample which you have imported as shown in the figure below. This table is used to set the *Sample Name* and *Group* or class for the various samples as well as to view the *Scale Factor* and other fields. You use the *View→Show Samples Table* menu item to display the table for the active pane's data.

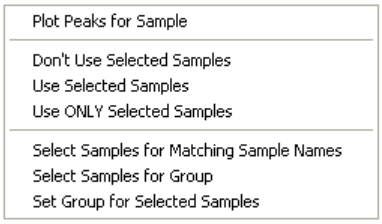


Row	Index	Sample Name	Sample ID	Group	Use	Acq. Time	Scale Factor	RT Correction
1	1	Rat1_0-8h_vinpo - A	MS03GE-041110-	1	☑	2004/11/10, 2:12 PM	1.000e0	None
2	2	Rat1_0-8h_blk - A	MS03GE-041110-	blank1	☑	2004/11/10, 2:38 PM	1.000e0	None
3	3	Rat1_8-16h_vinpo - A	MS03GE-041110-	2	☑	2004/11/10, 3:08 PM	1.000e0	None
4	4	Rat1_8-16h_blk - A	MS03GE-041110-	blank2	☑	2004/11/10, 3:39 PM	1.000e0	None
5	5	Rat1_16-24h_vinpo - A	MS03GE-041110-	3	☑	2004/11/10, 4:09 PM	1.000e0	None
6	6	Rat1_16-24h_blk - A	MS03GE-041110-	blank3	☑	2004/11/10, 4:39 PM	1.000e0	None
7	7	Rat2_0-8h_vinpo - A	MS03GE-041110-	1	☑	2004/11/10, 5:10 PM	1.000e0	None
8	8	Rat2_0-8h_blk - A	MS03GE-041110-	blank1	☑	2004/11/10, 5:40 PM	1.000e0	None
9	9	Rat2_8-16h_vinpo - A	MS03GE-041110-	2	☑	2004/11/10, 6:11 PM	1.000e0	None
10	10	Rat2_8-16h_blk - A	MS03GE-041110-	blank2	☑	2004/11/10, 6:41 PM	1.000e0	None

The table contains the following columns:

Column	**Description**
Row	Simply the row number for each row.
Index	The *original* row number. If you have sorted the table this will differ from the current row number. You can un-sort the table by selecting this row and clicking the *Sort Ascending* toolbar button.
Sample Name	The name of the sample. This name is displayed in plots involving samples and can be edited. For data imported from wiff files the default value is the sample name originally specified when the data was acquired; for data imported from T2D or text files the default value is the filename.
Sample ID	An extra identifier for the sample.
Group	The name of the group or class to which the sample belongs, if known. This field must be assigned in order to perform t-test analyses. The group also determines the plot symbol for plots involving samples.
Use	This checkbox allows you to specify whether the sample should be used for subsequent data analysis such as PCA and the t-test. You can disable the sample by clicking directly in the cell or in various shortcut ways discussed elsewhere in this document.
Acq. Time	For data imported from wiff files, the acquisition date and time at which the data was acquired.
Scale Factor	The multiplicative normalization factor which has been applied for the sample. As discussed in section 2.4.3.2 and the following section, it is a good idea to check these values after performing sample normalization to ensure that the values are reasonable (*i.e.* not grossly different from 1.0).
RT Correction	If retention time correction was applied to LC/MS data at the time it was imported, this column shows the slope and intercept which were used in the equation: Corrected RT = slope * (Original RT) + intercept For example a value of '1.02*RT + 0.21' indicates a slope of 1.02 and an intercept of 0.21 minutes.

4.3.1 Context Menu



4.3.1.1 Plot Peaks for Sample

This command creates a plot of peak response as a function of peak index or, in the case of MS (not LC/MS data) data, mass/charge. There is a separate overlaid trace for each sample or column which you have selected in the Peaks Table. For details about the resulting plot, see the Peaks for Sample Plot discussion in section 5.3.

4.3.1.2 Don't Use Selected Samples

This command unchecks the *Use* field for any selected samples. If you wish to de-select this field for many samples, this is much more convenient than manually clicking in the *Use* cell of the table for each one. Note that you can use the Control key to make disjoint selections in the table, *i.e.* all of the peaks do not need to be immediately adjacent.

4.3.1.3 Use Selected Samples

This command checks the *Use* field for any selected samples. This is essentially the reverse of the *Don't Use Selected Samples* command discussed above.

4.3.1.4 Use ONLY Selected Samples

This command checks the *Use* field for any selected samples *and* unchecks this field for all other samples. This command provides a convenient way to focus on only a small subset of the samples.

4.3.1.5 Select Samples for Matching Sample Names

This command is used to select all rows in the table for which the *Sample Name* field matches the specified pattern. If there are many such rows this is more convenient than manually selected each one individually. This is useful so that you can subsequently perform other operations involving these samples. This is similar to the similar command for Peaks Tables described in section 4.2.2.8 – see that section for an example.

4.3.1.6 Select Samples for Group

This command is used to select all samples with a particular group name. If there are many matching rows, this is more convenient than manually selecting each one individually. This is useful so that you can subsequently perform other operations involving these samples (for example exclude them).

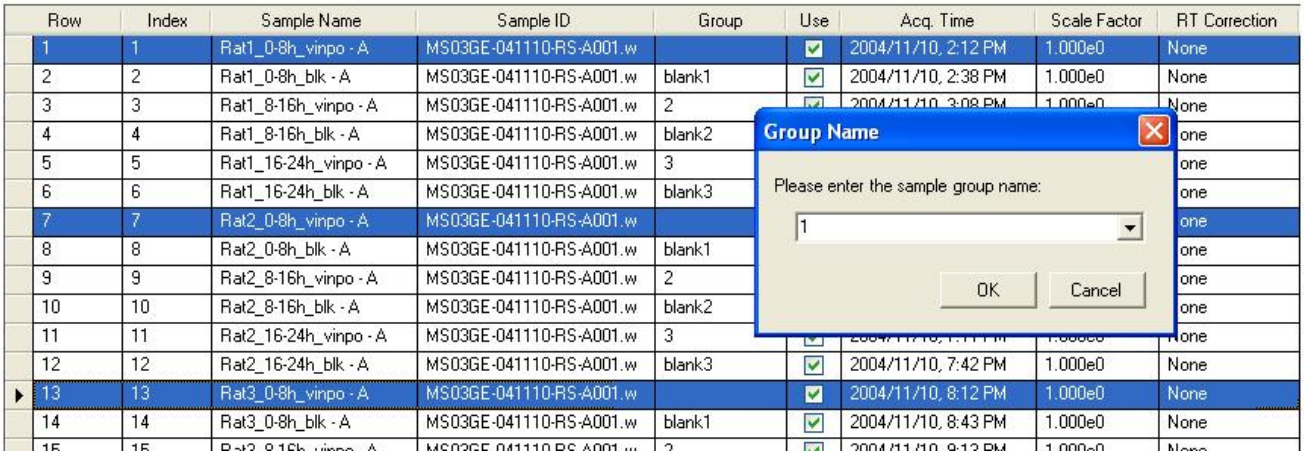
When you select the command, a simple dialog shown is presented to allow you to select one of the group names currently in use for the table; the '(None)' item corresponds to an empty or undefined group. This dialog is similar to that discussed previously for peaks in section 4.2.2.9.

4.3.1.7 Set Group for Selected Samples

This command sets the *Group* field for any selected samples. If you wish to set this field for many samples, this is much more convenient than manually setting each one individually by typing into the corresponding cell of the table. Note that you can use the Control key to make disjoint selections in the table, *i.e.* all of the samples do not need to be immediately adjacent.

When you select the command, a simple dialog is presented asking you for the name of the group. In the figure below the new group name ('1') will be applied to the three selected samples. You can either type directly into

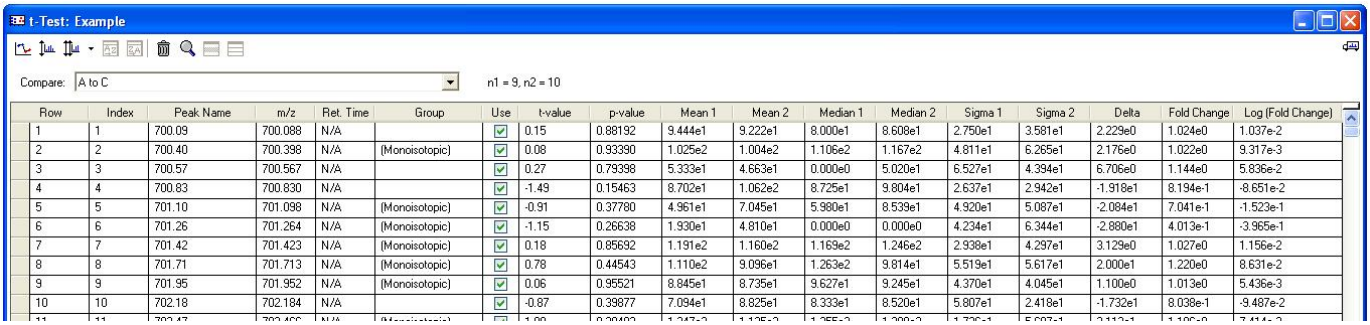
the *Group Name* dialog or use the combobox to select a group name which you have previously defined in the *Plot Symbols* tab of the *Options* as discussed in section 2.2.8.1.



Row	Index	Sample Name	Sample ID	Group	Use	Acq. Time	Scale Factor	RT Correction
1	1	Rat1_0-8h_vinpo - A	MS03GE-041110-RS-A001.w		☑	2004/11/10, 2:12 PM	1.000e0	None
2	2	Rat1_0-8h_blk - A	MS03GE-041110-RS-A001.w	blank1	☑	2004/11/10, 2:38 PM	1.000e0	None
3	3	Rat1_8-16h_vinpo - A	MS03GE-041110-RS-A001.w	2	☑	2004/11/10, 3:08 PM	1.000e0	None
4	4	Rat1_8-16h_blk - A	MS03GE-041110-RS-A001.w	blank2	☑			one
5	5	Rat1_16-24h_vinpo - A	MS03GE-041110-RS-A001.w	3	☑			one
6	6	Rat1_16-24h_blk - A	MS03GE-041110-RS-A001.w	blank3	☑			one
7	7	Rat2_0-8h_vinpo - A	MS03GE-041110-RS-A001.w		☑			one
8	8	Rat2_0-8h_blk - A	MS03GE-041110-RS-A001.w	blank1	☑			one
9	9	Rat2_8-16h_vinpo - A	MS03GE-041110-RS-A001.w	2	☑			one
10	10	Rat2_8-16h_blk - A	MS03GE-041110-RS-A001.w	blank2	☑			one
11	11	Rat2_16-24h_vinpo - A	MS03GE-041110-RS-A001.w	3	☑			None
12	12	Rat2_16-24h_blk - A	MS03GE-041110-RS-A001.w	blank3	☑	2004/11/10, 7:42 PM	1.000e0	None
13	13	Rat3_0-8h_vinpo - A	MS03GE-041110-RS-A001.w		☑	2004/11/10, 8:12 PM	1.000e0	None
14	14	Rat3_0-8h_blk - A	MS03GE-041110-RS-A001.w	blank1	☑	2004/11/10, 8:43 PM	1.000e0	None

4.4 t-Test Table

As discussed in section 2.4.2 you can perform a standard t-test for comparing one group of samples to another. After confirming the resulting dialog, a new t-Test Table is created as shown in the figure below. The table contains a row for each of the original peaks and a column for the various metrics as explained below.



Row	Index	Peak Name	m/z	Ret. Time	Group	Use	t-value	p-value	Mean 1	Mean 2	Median 1	Median 2	Sigma 1	Sigma 2	Delta	Fold Change	Log (Fold Change)
1	1	700.09	700.088	N/A		☑	0.15	0.88192	9.444e1	9.222e1	8.000e1	8.608e1	2.750e1	3.581e1	2.229e0	1.024e0	1.037e-2
2	2	700.40	700.398	N/A	(Monoisotopic)	☑	0.08	0.93390	1.025e2	1.004e2	1.106e2	1.167e2	4.811e1	6.265e1	2.176e0	1.022e0	9.317e-3
3	3	700.57	700.567	N/A		☑	0.27	0.79398	5.333e1	4.663e1	0.000e0	5.020e1	6.527e1	4.394e1	6.706e0	1.144e0	5.836e-2
4	4	700.83	700.830	N/A		☑	-1.49	0.15463	8.702e1	1.062e2	8.725e1	9.804e1	2.637e1	2.942e1	-1.918e1	8.194e-1	-8.651e-2
5	5	701.10	701.098	N/A	(Monoisotopic)	☑	-0.91	0.37780	4.961e1	7.045e1	5.980e1	8.539e1	4.920e1	5.087e1	-2.084e1	7.041e-1	-1.523e-1
6	6	701.26	701.264	N/A	(Monoisotopic)	☑	-1.15	0.26638	1.930e1	4.810e1	0.000e0	0.000e0	4.234e1	6.344e1	-2.880e1	4.013e-1	-3.965e-1
7	7	701.42	701.423	N/A	(Monoisotopic)	☑	0.18	0.85692	1.191e2	1.160e2	1.169e2	1.246e2	2.938e1	4.297e1	3.129e0	1.027e0	1.156e-2
8	8	701.71	701.713	N/A	(Monoisotopic)	☑	0.78	0.44543	1.110e2	9.096e1	1.263e2	9.814e1	5.519e1	5.617e1	2.000e1	1.220e0	8.631e-2
9	9	701.95	701.952	N/A	(Monoisotopic)	☑	0.06	0.95521	8.845e1	8.735e1	9.627e1	9.245e1	4.370e1	4.045e1	1.100e0	1.013e0	5.436e-3
10	10	702.18	702.184	N/A		☑	-0.87	0.39877	7.094e1	8.825e1	8.333e1	8.520e1	5.807e1	2.418e1	-1.732e1	8.038e-1	-9.487e-2

The combobox at the top of the pane is used to select a particular comparison. As explained in section 2.4.2, a t-test is available for each possible pair-wise comparison between groups, for each group compared to all other samples collectively, and for the special 'first to last' comparison. The figure below shows a case for which exactly three groups are defined (*A, B* and *C*).

The area immediately to the right of the combobox ('n1 = 8, n2 = 8' in the figure) shows the number of samples in each of the two groups for the currently selected comparison.

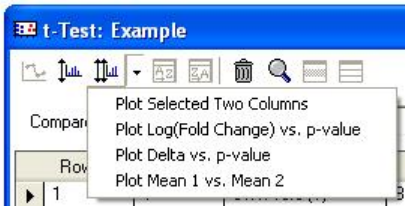
The *Row, Index, Peak Name, m/z, Ret. Time, Group* and *Use* columns are identical to those discussed for the Peaks Table in section 4.2. The table contains the following additional columns:

Column	Description
t-value	The raw t-value for the selected comparison. The formula used to calculate this value was given in section 2.4.2.

p-value	The p-value is the probability that any difference between the two groups for the selected comparison is due to chance only. So, values close to one indicate that the difference is almost certainly just due to chance and values close to zero indicate that the difference is *not* likely to be due to chance. *(Note:* For those familiar with version 1.0 or 1.1 of the program, this is essentially the reverse of the percentage probability reported by those versions: p-value = 1 - Percent_Probability/100.)
Mean 1	The mean response for the first group for the selected comparison.
Mean 2	The mean response for the second group.
Median 1	The median response for the first group for the selected comparison.
Median 2	The median response for the second group.
Simga 1	The standard deviation of the responses for the first group for the selected comparison.
Sigma 2	The standard deviation of the responses for the second group for the selected comparison.
Delta	The difference between the two averages (Average 1 - Average 2).
Fold Change	The ratio of the two averages (Average 1 / Average 2). (If the second average is zero, a value of 'Infinity' is displayed.)
Log (Fold Change)	The base-10 logarithm of the Fold Change. (If the first average is zero a value of '-Infinity' is displayed and if the second average is zero a value of 'Infinity' is displayed.)

4.4.1 Plot Two Columns Toolbar Button

Clicking the *Plot Two Columns* toolbar button plots the two currently selected columns against one another in the usual way for tables as discussed in section 4.1.1.3. As shown in the figure below, for t-test tables this button has an associated drop-down menu which is used to quickly generate common plots without having to manually first select the columns. The various menu items are explained below and these two-column t-test plots are described in section 5.6.



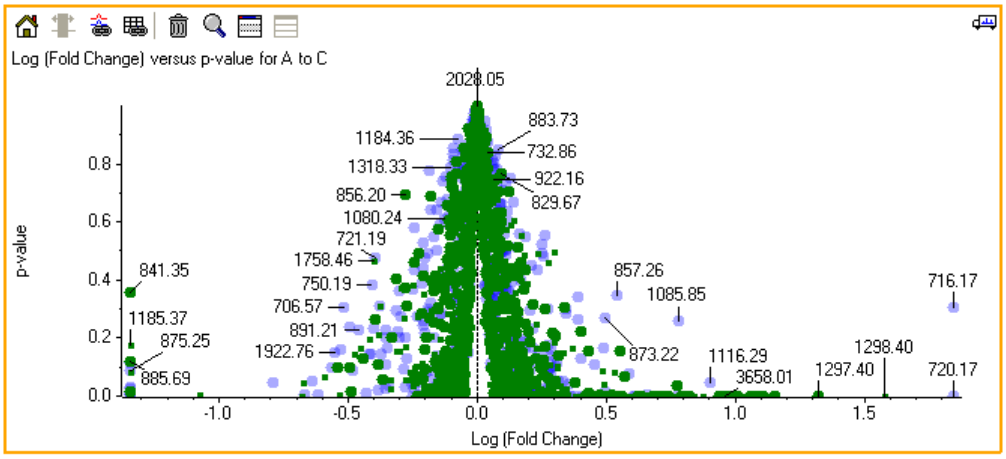
4.4.1.1 Plot Selected Two Columns

This command is available if exactly two columns are selected and is equivalent to directly clicking the button.

4.4.1.2 Plot Log(Fold Change) vs. p-value

This command plots Log (Fold Change) as the x-value and p-value as the y-value for the currently selected group as shown in the figure below. This plot allows you to focus on peaks with small p-values and largish (absolute) fold changes.

Note: As described in the table above, peaks can have a Log (Fold Change) of ±Infinity if the average for either group is exactly zero. These peaks are still plotted, but at the very left or right of the graph offset some distance from the other points. For example in the figure below the peak on the far right with m/z of 716.17 does **not** have a calculated Log (Fold Change) near 1.8, but rather of +Infinity.



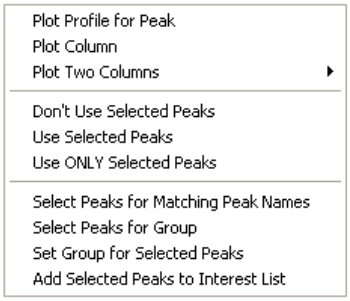
4.4.1.3 Plot Delta vs. p-value

This command plots the difference between the two groups (delta) as the x-value and the p-value as the y-value.

4.4.1.4 Plot Average 1 vs. Average 2

This command plots the average values for the two groups for the selected comparison against one another. This display is useful for identifying peaks which have a ratio significantly different from the average. Conversely it is useful for locating peaks which may be useful to use as internal standards for normalization.

4.4.2 Context Menu



With the exception of the first three items (described below), the remaining items are identical to those discussed for the Peaks Table – see the sub-sections of section 4.2.2 for details.

4.4.2.1 Plot Profile for Peak

This command creates a plot of peak response as a function of sample. There is a separate overlaid trace in the plot for each peak or row which you have selected in the table. This command is identical to that for the Peaks Table discussed in section 4.2.2.2, however a few additional words are in order here.

These plots can be very useful for visualizing how particular peaks are behaving across the various samples, especially if you focus on peaks likely to be of specific interest. For example:

- For a particular comparison of interest, sort the table so that peaks with the lowest p-values appear at the top of the table.
- Select the first row of the table and select *the Plot Profile for Peak* command.
- Click back in the table and use the up and down arrow keys to scroll through the peaks. Since the Profile Plot is linked to the table by default, it will automatically update to show the profile for the selected row of the table.

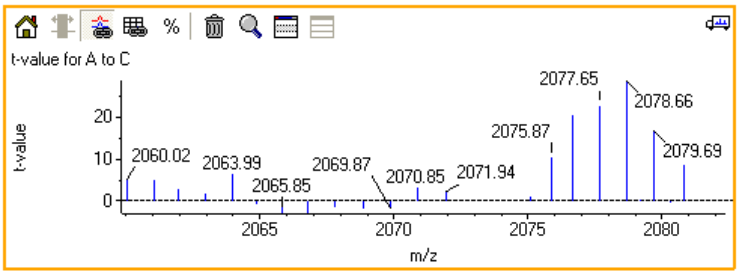
Since the p-values are unsigned, the above procedure will intermingle peaks which are larger in the first group with those which are larger in the second group. If you instead sort on the t-value column, the peaks which are

larger in one group will appear at the top of the table and those which are larger in the other group will appear at the bottom (depending on the sort direction).

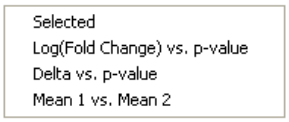
4.4.2.2 Plot Column

This command creates a new plot showing the values in the selected column or columns as a function of row number or mass. A separate overlaid trace is created in the resulting plot for each of the originally selected columns. These plots are discussed in section 5.5.

This command allows you to visualize the t-values, p-values and other metrics. For example in the zoomed figure below it can be seen that the peak near 2075 Da and its isotopes is much larger in group *A* than in other group and that the peak near 2065 shows the reverse behavior.



4.4.2.3 Plot Two Columns Sub-Menu



The commands in this sub-menu are identical to those described above in section 4.4.1.

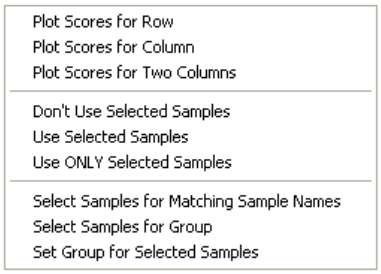
4.5 PCA Scores Table

As discussed in section 2.4.1 you can perform a principal components analysis for the active pane's data. As mentioned there, a new window is created in which one of the panes is a PCA Scores Table. As shown in the figure below the Scores Table contains a row for each of the original samples and a column for the scores for the various principal components.

Row	Index	Sample Name	Sample ID	Group	Use	PC1 (58.2 %)	PC2 (17.5 %)	PC3 (10.1 %)
1	1	Rat1_0-8h_vinpo - A	LCMS Data.wiff (sa	1	☑	1.657e2	-4.262e1	7.144e0
2	2	Rat1_0-8h_blk - A	LCMS Data.wiff (sa	blank1	☑	-4.722e1	-7.109e1	-2.083e1
3	3	Rat1_8-16h_vinpo - A	LCMS Data.wiff (sa	2	☑	6.938e1	4.408e1	-2.549e1
4	4	Rat1_8-16h_blk - A	LCMS Data.wiff (sa	blank2	☑	-6.491e1	-1.856e1	-2.840e1
5	5	Rat1_16-24h_vinpo - A	LCMS Data.wiff (sa	3	☑	2.781e0	4.590e1	-3.564e1
6	6	Rat1_16-24h_blk - A	LCMS Data.wiff (sa	blank3	☑	-6.959e1	3.545e1	-5.182e1
7	7	Rat2_0-8h_vinpo - A	LCMS Data.wiff (sa	1	☑	1.197e2	-2.707e1	5.273e0

The *Row, Index, Sample Name, Sample ID, Group* and *Use* columns are identical to those discussed for the Samples Table in section 4.3. There is an additional column containing the PCA scores for the samples for each of the significant principal components. The title of these columns indicates the percentage of the total variance amongst all the samples which is accounted for by the principal component. In the case that the PCA-DA option was used, the titles of the columns are D1, D2, *etc.* rather than PC1 and so forth to indicate that these are discriminant components.

4.5.1 Context Menu



With the exception of the first three items (described below), the remaining items are identical to those discussed for the Samples Table – see the sub-sections of section 4.3.1 for details. The remaining commands used for generating plots were also discussed in the sub-sections of section 4.1.1, so only a brief description is included here.

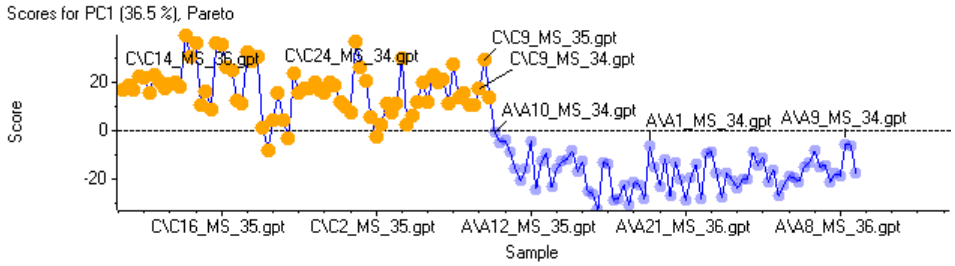
4.5.1.1 Plot Scores for Row

This command creates a new plot showing the PCA scores for all principal components for the selected sample or samples. A separate overlaid trace is created for each of the selected samples. You will probably use these plots only infrequently, however they may be useful if you wish to study one sample in detail.

4.5.1.2 Plot Scores for Column

This command generates a plot displaying the PCA scores for the selected PC column or columns for the samples. A separate overlaid trace is created in the resulting plot for each of the originally selected columns. These plots are discussed in section 5.8.

Most often you will probably plot two principal components as a function of one another, rather than just one as here, since this will provide additional separation power. However at least for cases in which any one PC does a good job of separating the samples, plotting that one PC, as in the figure below, may also be useful.



4.5.1.3 Plot Scores for two Columns

This command generates a standard PCA scores plot showing the scores for all samples for the two selected principal components. These plots are discussed in section 5.7

The initial PCA window containing the Scores Table automatically creates a Scores Plot for the first two principal components. This plot is also automatically linked to the table, so if you select any two PC columns of the table the Scores Plot will update. So, you really only need to use this command to generate a new Scores Plot if you delete the automatically generated plot or if you wish to display two simultaneous plots for different principal components. In this latter case you will need to un-link one of the plots from the table, otherwise both plots will automatically show data for the same principal components.

Be certain to explore the Scores Plots for cases other than PC1 and PC2.

4.6 PCA Loadings Table

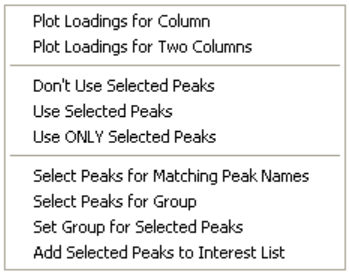
As discussed in section 2.4.1 you can perform a principal components analysis for the active pane's data. As mentioned there, a new window is created in which one of the panes is a PCA Loadings Table. As shown in the

figure below, the Loadings Table contains a row for each of the original peaks and a column for the loadings for the various principal components.

Row	Index	Peak Name	m/z	Ret. Time	Group	Use	PC1 (58.2 %)	PC2 (17.5 %)	PC3 (10.1 %)
1	1	81.1/13.3 (1)	81.067	13.32		☑	-1.321e-3	2.236e-3	-5.688e-3
2	2	83.1/13.7 (2)	83.084	13.65		☑	-2.759e-3	8.587e-3	-1.027e-2
3	3	83.1/12.6 (3)	83.084	12.62		☑	-2.567e-3	1.139e-2	-2.265e-2
4	4	85.0/12.5 (4)	85.027	12.51		☑	3.468e-5	1.939e-2	-6.010e-3
5	5	89.0/11.5 (5)	89.036	11.51		☑	-9.871e-3	3.110e-2	-9.375e-3
6	6	90.6/11.3 (6)	90.616	11.28		☑	1.292e-3	5.884e-3	-2.362e-3
7	7	91.1/11.3 (7)	91.050	11.29	(Monoisotopic)	☑	-8.919e-2	3.492e-1	-3.677e-2
8	8	91.1/14.6 (8)	91.051	14.63		☑	-1.616e-3	2.735e-3	-6.958e-3
9	9	91.1/13.2 (9)	91.052	13.19		☑	-4.804e-4	6.637e-2	-3.822e-3
10	10	92.1/11.3 (10)	92.053	11.27	(Isotope)	☑	-2.005e-2	8.447e-2	-5.030e-3
11	11	93.1/18.1 (11)	93.055	18.12		☑	8.242e-3	-7.042e-3	2.054e-3

The *Row, Index, Peak Name, m/z, Ret. Time, Group* and *Use* columns are identical to those discussed for the Peaks Table in section 4.2. There is an additional column containing the PCA scores for the samples for each of the significant principal components. The title of these columns indicates the percentage of the total variance amongst all the samples which is accounted for by the principal component. In the case that the PCA-DA option was used, the titles of the columns are D1, D2, *etc.* rather than PC1 and so forth to indicate that these are discriminant components.

4.6.1 Context Menu



With the exception of the first two items (described below), the remaining items are identical to those discussed for the Peaks Table – see the sub-sections of section 4.2.2 for details.

4.6.1.1 Plot Loadings for Column

This command generates a plot displaying the PCA loadings for the selected PC column or columns for the peaks. A separate overlaid trace is created in the resulting plot for each of the originally selected columns – see section 5.11 for a discussion.

Most often you will probably plot two principal components as a function of one another, rather than just one as here, however this display can be useful if you have decided that a particular principal component is important and wish to focus attention on that one PC.

4.6.1.2 Plot Loadings for Two Columns

This command generates a standard PCA loadings plot showing the loadings for all peaks for the two selected principal components – see section 5.10 for a discussion of these plots.

The initial PCA window containing the Loadings Table automatically creates a Loadings Plot for the first two principal components. This plot is also automatically linked to the table, so if you select any two PC columns of the table the Loadings Plot will update. So, you really only need to use this command to generate a new Loadings Plot if you delete the automatically generated plot or if you wish to display two simultaneous plots for different principal components. In this latter case you will need to un-link one of the plots from the table, otherwise both plots will automatically show data for the same principal components.

4.7 Excluded Peaks Table

This table contains those peaks which have been excluded from analysis for the active pane. As mentioned in section 2.3.2, this table is obtained using the *View->Show Excluded Peaks* command. Note that this is not a global list, but is simply a list of those peaks excluded for the currently active data set.

The following figure shows an example table. Note that if no peaks have been excluded the table will be empty.

Row	Index	Peak Name	m/z	Ret. Time	Group	Current
1	1	81.1/13.3 (1)	81.067	13.32		☑
2	2	83.1/13.7 (2)	83.084	13.65		☑
3	3	83.1/12.6 (3)	83.084	12.62		☑
4	4	85.0/12.5 (4)	85.027	12.51		☑
5	5	89.0/11.5 (5)	89.036	11.51		☑
6	6	90.6/11.3 (6)	90.616	11.28		☑

With the exception of the last column, the other columns are identical to those discussed for the Peaks Table in section 4.2.

The *Current* column indicates whether the corresponding peak still exists in the underlying data set. If this field is checked, the *Use* flag for the peak has been turned off, but the peak is still present in the related Peaks Table. If this field is not checked, the *Use* flag was *previously* turned off and PCA or a t-test was subsequently run; in this case these peaks will not be present in the PCA Loadings Table or the t-Test Table. In the example above the first three peaks have been removed from the PCA or t-Test data set. If you wish to subsequently re-include these peaks, you need to revert to the original Peaks Table.

4.7.1 Context Menu

4.7.1.1 Add All Peaks to Global Exclusion List

This command adds all of the peaks in the Excluded Peaks Table to the global exclusion list described in section 2.2.8.2. This allows you to exclude these peaks up front when importing LC/MS or MS data in the future. If there are more than a few excluded peaks, this is much more convenient than manually typing into the global list.

4.7.1.2 Add Selected Peaks to Global Exclusion List

This command adds the peaks whose rows have been selected to the global exclusion list described in section 2.2.8.2. This is similar to the previous command, except that you can control exactly which peaks are added to the global list. For example the selected peaks may be drug metabolites or background ions which you wish to automatically exclude when importing new data into the program.

4.7.1.3 Clear Global Exclusion List

This command clears all peaks from the global exclusion list described in section 2.2.8.2. It is a shortcut to using the *Clear* button in the exclusion dialog itself.

4.7.1.4 Select Peaks Matching Peak Names

This command is used to select all peaks for which the peak name matches a specified pattern as described in section 4.2.2.8.

4.7.1.5 Select Peaks for Group

This command is used to select all peaks with a particular group name as discussed for the Peaks Table in section 4.2.2.9.

4.7.1.6 Set Group for Selected Peaks

This command sets the *Group* field for any selected peaks as discussed for the Peaks Table in section 4.2.2.10.

4.7.1.7 Add Selected Peaks to Interest List

This command adds the selected peaks to the Interest List. The main reason for doing this is to subsequently report details for these peaks or possibly to export this interest list for use outside of the application. For details see the Interest List in the next section and also the *Generate Report in MS Word* command in section 2.1.12.

4.8 Interest List Table

This table contains those peaks which you have explicitly added to the interest list for the active pane. Note that this is not a global list, but is specific to the currently active pane's data. The table is obtained using the *View→Show Interest List* menu item.

The main reason for adding peaks to the interest list is to subsequently report details for these peaks or possibly to export this interest list for use outside of the application. For details on reporting see the *Generate Report in MS Word* command in section 2.1.12.

The following figure shows an example table. Note that if you have not added any peaks to the interest list the table will empty.

Row	Index	Peak Name	m/z	Ret. Time	Group	Charge	Mono.	Mass	Excl.	Comment
1	1	91.1/11.3 (7)	91.050	11.29	(Monoisotopic)	1	☑	90.042	☐	
2	2	194.1/11.3 (90)	194.077	11.28	(Monoisotopic)	1	☑	193.069	☐	
3	3	302.1/12.4 (214)	302.125	12.43	(Monoisotopic)	1	☑	301.117	☐	
4	4	387.2/11.3 (317)	387.154	11.29	(Monoisotopic)	1	☑	386.146	☐	

The *Row, Index, Peak Name, m/z, Ret. Time* and *Group* columns are identical to those discussed for the Peaks Table in section 4.2. The *Charge, Mono. and Mass* columns are discussed in section 4.9 for the Peak Info Table. The *Excl.* column is checked if the peak is excluded from analysis for the data set. The *Comment* column allows you to specify a comment for the peak, either at the time the peak is added to the interest list (as described below) or later by typing directly into the cells. As discussed in the reporting section, you can include this comment in reports.

You can add peaks to the interest list from a number of different locations. For example you can select peaks of interest in a Peaks Table, a t-Test Table, a PCA Loadings Table or an Excluded Peaks Table and select the *Add Selected Peaks to Interest List* command from the context menu. You can perform a similar operation using the context menu of a Profile Plot or a PCA Loadings Plot.

When you choose any of these commands the dialog shown below is presented by default. Any comment which you specify will be used for all of the peaks (if there is more than one) which you are adding to the interest list. You can also set the comment later by typing directly into the cells of the table. If you do not wish to be prompted by this dialog each time you add peaks to the interest list, choose *the Only show this dialog again if the shift key is down* checkbox and click *OK*.

4.8.1 Context Menu

4.8.1.1 Clear Interest List

This command removes all peaks from the interest list.

4.8.1.2 Remove Selected Peaks from Interest List

This command removes any peaks whose rows have been selected from the interest list.

4.8.1.3 Set Comment for Selected Peaks

This command allows you to set the comment for all peaks whose rows have been selected. If you wish to set the same comment for many peaks, this is more convenient than typing into each of the corresponding *Comment* fields one by one. When you select the command, a dialog similar to that shown above asking you for the comment is presented.

4.8.1.4 Select Peaks for Matching Peak Names

This command is used to select all peaks for which the peak name matches a specified pattern as described in section 4.2.2.8.

4.8.1.5 Select Peaks for Group

This command is used to select all peaks with a particular group name as discussed for the Peaks Table in section 4.2.2.9.

4.8.1.6 Set Group for Selected Peaks

This command sets the *Group* field for any selected peaks as discussed for the Peaks Table in section 4.2.2.10.

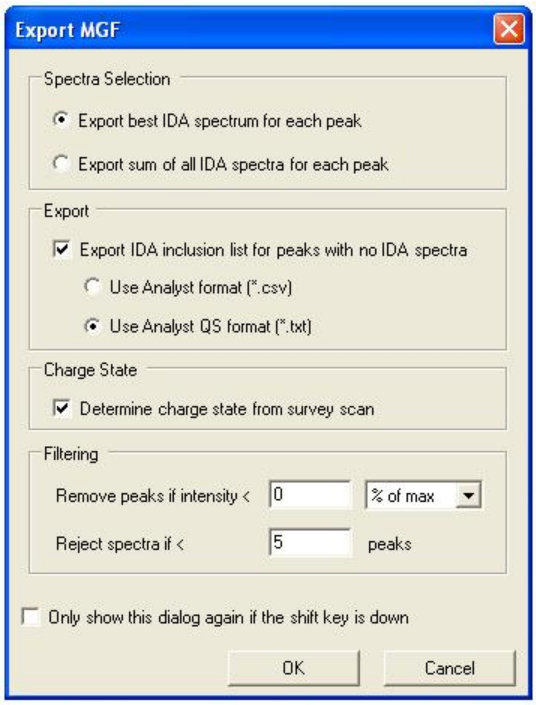
4.8.1.7 Export Analyst IDA Inclusion List

This command first prompts you for the name of a file; it then saves the interest list in format suitable for directly importing into the Analyst® IDA method editor as an inclusion (or exclusion) list. If you are *not* using the Analyst® QS software you should select the *Analyst Inclusion/Exclusion List (*.csv)* format (the default); otherwise you should select the *Analyst QS Inclusion/Exclusion List (*.txt)* format.

4.8.1.8 Export Mascot MGF for Available IDA Spectra

If you have processed the MS data for a series of IDA runs, this command creates a text file containing a peak list for the (available) IDA MS/MS spectra for each peak added to the interest list. The file is saved in '.mgf' (Mascot Generic Format) for use with the Mascot protein search engine. Other search engines such as X!Tandem are also able to process files in this format. This command is only applicable to product spectra of peptides.

If an IDA dependent spectrum was not triggered for a particular peak for at least one of the samples, that peak is skipped. When you select the command the dialog shown below is presented. If you click the *OK* button you are then presented with a standard File Save dialog asking for the name of the .mgf file to create.



The dialog contains the following items:

- *Export best IDA spectrum for each peak* – In the case that IDA was triggered for a particular peak in the interest list more than once (for all of the samples), select this option to export the peak list for the one 'best' spectrum. The 'best' spectrum is the one with the largest intensity sum for all fragments (*i.e.* all spectral peaks with the exception of the precursor itself).
- *Export sum of all IDA spectra for each peak* – In the case that IDA was triggered for a particular peak in the interest list more than once (for all of the samples), select this option to export the peak list for the spectrum obtained by summing them all together. This option will generally give the highest quality result, although if spectra from different peptides with similar parent m/z values and retention time were acquired (possibly for different samples), the summed spectrum will be a mixture of those for these different peptides.
- *Export IDA inclusion list for peaks with no IDA spectra* – If this option is checked, an Analyst® inclusion list is created which contains an entry for each peak for which IDA did not trigger (for any of the samples). This is similar to the *Export Analyst IDA Inclusion List* functionality discussed in section 4.8.1.7, however in this case the inclusion list does not contain those peaks for which IDA product spectra are already available. The inclusion list is placed in the same directory as the specified 'mgf' file with the same filename.

  If you are *not* using the Analyst® QS software you should select the *Use Analyst format (*.csv)* option, otherwise the *Use Analyst QS format (*.txt)* format.
- *Determine charge state from survey scan* – If this option is checked the survey scan is examined for each IDA spectrum and the charge state of the parent peptide is determined and written to the file. If this item is unchecked, charge state information is not written and the charge states to use should be manually specified in the Mascot search options.
- *Remove peaks if intensity < x* – Any peaks with intensity smaller than the specified value are presumed to be noise and are removed from the peak list. You can specify the threshold as either a percentage of the most intense peak in the peak list or in absolute counts.
- *Reject spectra if < n peaks* – Any peak list containing less than the specified number of peaks is not included in the exported file.

- *Only show this dialog again if the shift key is down* – If you habitually use the same options, you can enable this checkbox so that the dialog is not shown each time you select it. As the wording implies, you can hold the Shift key while selecting the command to view the dialog if you have previously enabled this option.

4.9 Peak Info Table

This table contains general information for each peak including charge state, isotope status, mass defect and statistical information. It also provides a way to select or reject peak based on these criteria. The table is obtained using the *View→Show Peak Info* menu item, an example of which is shown in the figure below.

When MS or LC/MS data is processed, the program automatically attempts to assign the charge state and isotope status for each peak. For many peaks this automatic assignment is not possible, for example for small peaks for which the corresponding isotope may have been below the peak detection threshold. Especially for noisy peaks, it is possible for the assignment algorithm to assign a different charge state or isotope status for the same aligned peak for different samples; in this case the results from the sample with the largest response for the peak are reported.

If you open data processed with version 1.0 or 1.1 of the program, charge state and isotope status will not be available until you select the *Assign Charge States and Isotopes* menu item from the *Analyze* menu (see section 2.4.10).

Row	Index	Peak Name	m/z	Ret. Time	Group	Use	Charge	Mono	Mass	Mass Defect	Mean	Median	Sigma	%RSD	Min	Max	Samples > 0
1	1	81.1/13.3 (1)	81.0674	13.32		☑	N/A	■	N/A	0.067	6.700e-2	0.000e0	2.842e-1	4.243e2	0.000e0	1.206e0	1
2	2	83.1/13.7 (2)	83.0836	13.65		☑	N/A	■	N/A	0.084	4.362e-1	0.000e0	1.177e0	2.698e2	0.000e0	4.683e0	3
3	3	83.1/12.6 (3)	83.0845	12.62		☑	N/A	■	N/A	0.084	9.320e-1	0.000e0	2.206e0	2.367e2	0.000e0	7.806e0	4
4	4	85.0/12.5 (4)	85.0266	12.51		☑	N/A	■	N/A	0.027	1.219e0	0.000e0	1.998e0	1.639e2	0.000e0	6.542e0	7
5	5	89.0/11.5 (5)	89.0356	11.51		☑	N/A	■	N/A	0.036	1.486e1	1.356e1	9.993e0	6.725e1	1.058e0	3.944e1	18
6	6	90.6/11.3 (6)	90.6164	11.28		☑	N/A	■	N/A	0.616	7.813e-2	0.000e0	3.315e-1	4.243e2	0.000e0	1.406e0	1
7	7	91.1/11.3 (7)	91.0503	11.29	(Monoisotopic)	☑	1	☑	90.0425	0.050	8.753e2	9.211e2	3.278e2	3.745e1	2.503e2	1.433e3	18
8	8	91.1/14.6 (8)	91.0507	14.63		☑	N/A	■	N/A	0.051	1.003e-1	0.000e0	4.253e-1	4.243e2	0.000e0	1.805e0	1
9	9	91.1/13.2 (9)	91.0516	13.19		☑	N/A	■	N/A	0.052	1.212e1	4.736e0	1.636e1	1.350e2	0.000e0	5.092e1	13
10	10	92.1/11.3 (10)	92.0528	11.27	(Isotope)	☑	1	☐	91.0449	0.053	4.085e1	4.177e1	1.892e1	4.632e1	6.767e0	7.710e1	18
11	11	93.1/18.1 (11)	93.0547	18.12		☑	N/A	■	N/A	0.055	4.599e-1	0.000e0	1.951e0	4.243e2	0.000e0	8.278e0	1
12	12	95.0/12.5 (12)	95.0106	12.50		☑	N/A	■	N/A	0.011	2.949e-1	0.000e0	9.512e-1	3.225e2	0.000e0	3.849e0	2

The *Row, Index, Peak Name, m/z, Ret. Time* and *Group* columns are identical to those discussed for the Peaks Table in section 4.2. The table contains the following additional columns:

Column	Description
Charge	The charge state assigned for the peak. If this could not be determined 'N/A' is displayed.
Mono.	Indicates whether or not the peak was assigned as monoisotopic – a checkmark indicates a monoisotopic peak, an unchecked box indicates an isotope, and an indeterminate state (for example the first row in the figure above) indicates that this could not be determined.
Mass	The actual mass based on the mass/charge and the assigned charge state, *assuming* that the charging agent is hydrogen. If the charge state could not be calculated this displays as 'N/A', otherwise it is calculated as: Mass = charge * ($m/z \pm M_H$) where M_H is the mass of hydrogen and the '-' sign is used for positive mode and the '+' for negative mode. Note that for an isotope peak this is the mass of the actual isotope, *not* of the corresponding monoisotopic peak.
Mass Defect	The mass defect. Depending on the setting of the *Signed Mass Defect* option, this is either with respect to the nearest integral mass or the truncated mass as discussed below in section 4.9.1.4.
Mean	The mean response for the peak for all of the samples.
Median	The median response for the peak for all of the samples.

Sigma	The standard deviation of the response of the peak for all of the samples.
%RSD	The percent relative standard deviation of the response for the peak for all of the samples, *i.e.* 100% * Sigma / Mean.
Min	The minimum response for the peak for all of the samples.
Max	The maximum response for the peak for all of the samples.
Samples > 0	The number of samples for which the peak was detected.

4.9.1 Context Menu

The items in the second and third groups above are identical to those discussed for the Peaks Table – see the sub-sections of section 4.2.2 for details. The remaining items are described below.

4.9.1.1 Plot Profile for Peak

This command creates a plot of peak response as a function of sample. There is a separate overlaid trace in the plot for each peak or row which you have selected in the Peaks Table. This command is identical to that for the Peaks Table discussed in section 4.2.2.2.

4.9.1.2 Plot Column

This command creates a new plot showing the values in the selected column or columns as a function of row number or mass. A separate overlaid trace is created in the resulting plot for each of the originally selected columns. Note that only columns to the right of the *Mass* column can be plotted. These plots are discussed in section 5.17.

4.9.1.3 Plot Two Columns

This command plots the two currently selected columns against one another in the usual way for tables as discussed in section 4.1.1.3. Note that only columns to the right of the *Mass* column can be plotted. These plots are discussed in section 5.18.

4.9.1.4 Signed Mass Defect

This option controls how the mass defect is reported. If this item is checked the reported defect is with respect to the closest integer to the m/z value; otherwise it is with respect to the next lower integer. For example for an m/z value of 100.8 the mass defect is reported as -0.2 if this option is checked and as 0.8 otherwise. Note that the defect is reported for the m/z value, not the actual mass. This option is remembered and automatically applied the next time you display the Peak Info Table.

5 Plots

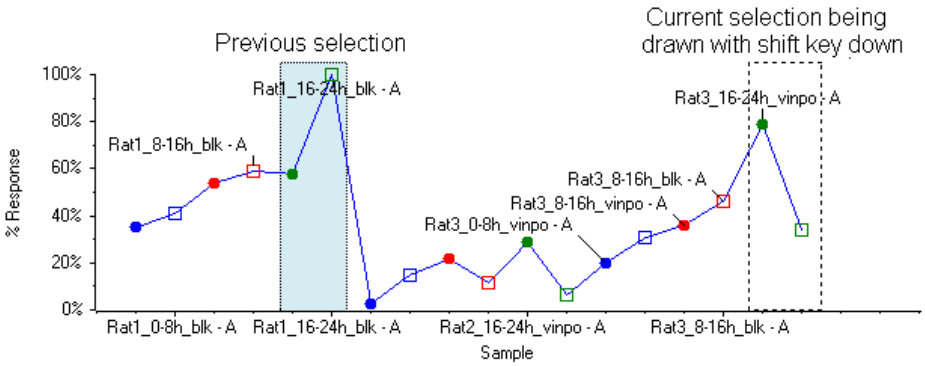
This section first discusses general functionality which applies to all plots, followed by a detailed list of each of the various types of plot.

5.1 General

5.1.1 Selections

You can select a region in a plot by clicking and dragging within the plot area. If you hold the Shift key while dragging any existing selections are kept, otherwise they are cleared and only the new selection will appear. The figure below shows a previous selection as well as a selection which is in the process of being defined.

Certain data processing operations apply to the selected region. For example once you have made a selection, you can use the *Zoom* toolbar button to make that range fill the entire plot area. Other operations apply to the selected range of data points. For example in the figure below, once the second selection has been finished and the mouse released, there will be four selected points: the *Rat_16-24h_blk-A* sample and the sample immediately before it and the *Rat3_16-24h_vinpo-A* sample and the one immediately after it.

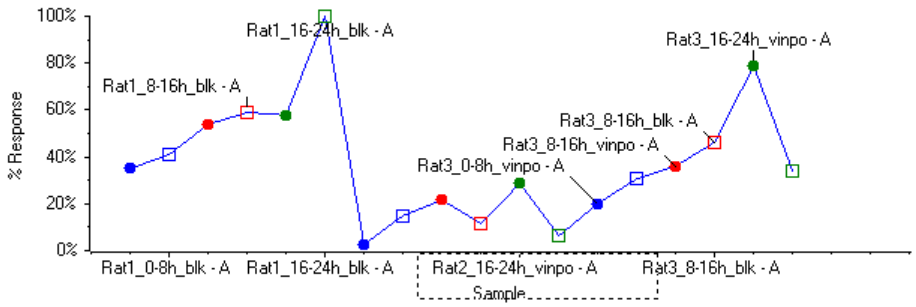


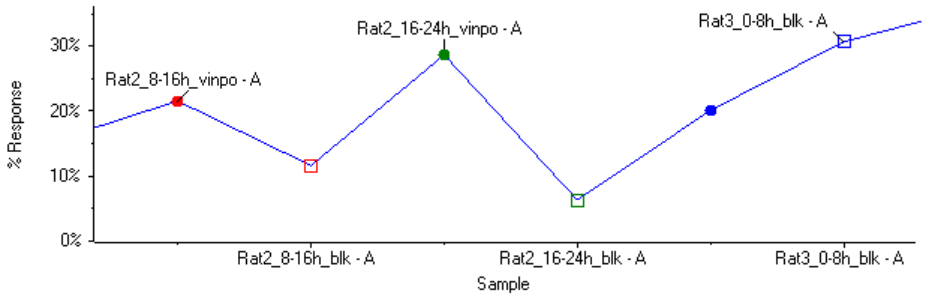
5.1.2 Zooming

Plots are usually zoomed by clicking and dragging in the x-axis or, more rarely, the y-axis. The top figure below shows a zoom region which is the process of being defined in the x-axis. If the mouse is released at the current location the x-axis will be zoomed as shown in the lower figure below. Note that when zooming the x-axis, the y-axis is automatically zoomed too so that all of the vertical plot area is used; if for some reason you do not want this behavior, you can hold the Control key while releasing the mouse. Once you have zoomed in this way, you can click once in the axis while holding the Shift key to undo the previous zoom operation.

If you double-click in the x-axis the plot will be returned to the home view in which all data is visible (in both the x and y directions). The behavior is similar if you double-click in the y-axis, except that only the y-direction is affected.

You can also zoom by drawing a selection and using the *Zoom* toolbar button or context menu command.





5.1.3 Plot Symbols

The plotting symbol for a particular data point is determined by the *Group* associated with the sample or peak as discussed in the *Plot Symbols* section of the *Options* in section 2.2.8.1. As discussed there, default plot symbols are used for MS and LC/MS data based on the isotope status of the peak.

When displaying a plot legend, you can double-click one of the items in the legend to select all of the points in the plot for the corresponding group; you can also single-click on the color spot to the immediate left of the item.

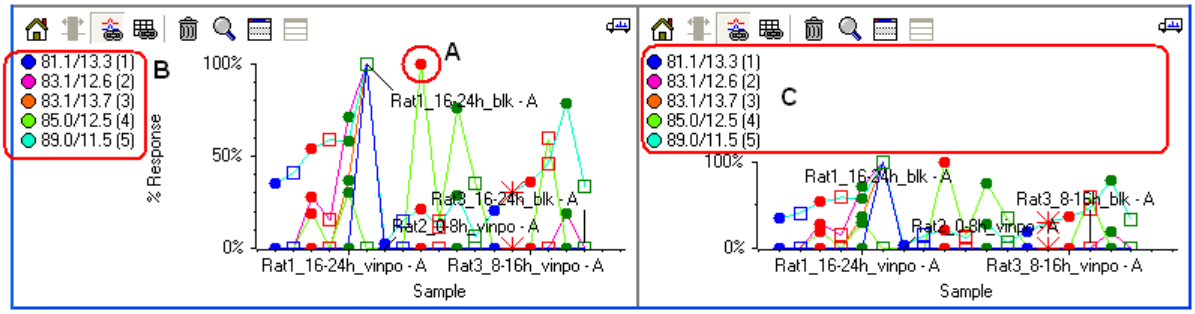
5.1.4 Overlaid Traces

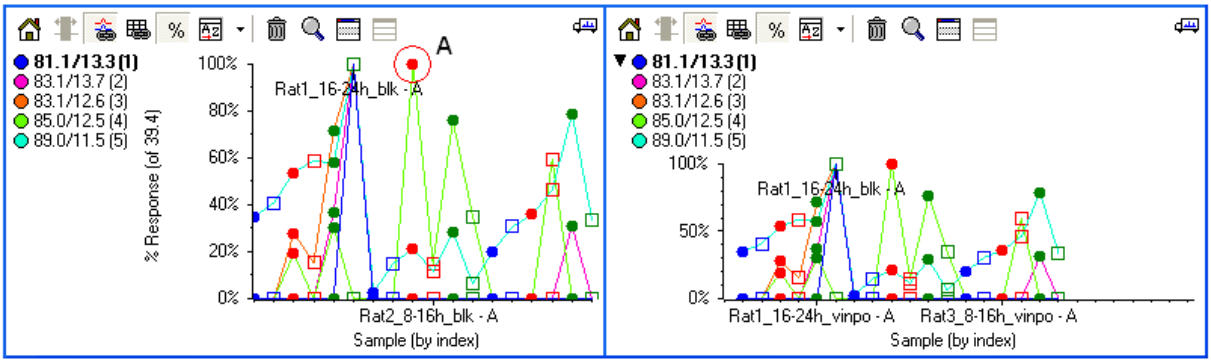
When a plot contains multiple data series or traces, only one of those traces is considered to be 'active' at any given moment. The active trace is the only one which is labeled and certain operations apply only to this trace.

You can activate a particular trace in a few different ways. Perhaps the easiest way is to click once directly on one of the plot symbols, for example the circled point *A* in the figure below. Note that you must click directly on a data point, not on the line joining two points. You can also activate a particular trace by double-clicking anywhere within the title line for the desired trace or by single-clicking within the color spot to the left of a title.

The plot displays a multi-line title either above the plot area or to its left (as specified in the *Display* options discussed in section 5.2.2.4.4). Note that if there are a very large number of overlaid traces all of the titles will not be visible. In this case you can click within the title and drag up or down (*i.e.* while keeping the mouse pressed) to scroll.

When the title is position above the graph you can view either all of the titles (or as many as are visible within the space available) or just the title for the active data by clicking the triangle icon (▼ or ▶) in the upper-left corner.





5.1.5 Toolbar Buttons

All of the toolbar buttons discussed here are available for all types of plots. Certain plot types contain additional items which are described in the appropriate sections further below.

5.1.5.1 Home View

 If the plot has been zoomed, this button returns it to the 'home' view; that is to say the view in which both the x and y axes show their default ranges and all available data is visible. You can also double-click in the x-axis to home the plot.

5.1.5.2 Zoom

 This button zooms the plot so that the selected region fills the entire available space. Before selecting this item you need to make a selection in the plot by clicking and dragging. You can also zoom by clicking and dragging directly in the x-axis (or y-axis) of the plot.

5.1.5.3 Link Graph Axes

 This button determines how the x-axis of the plot should behave when there are multiple plots of the same type. If this item is selected or pressed for at least two plots, zooming one of the plots will automatically cause the other plots to show the same range. This is convenient if you have multiple plots of the same type, probably in the same window, and want to compare between them.

5.1.5.4 Link to Table

 This button determines how plots should update when you make new selections in the table used to generate the plot. If this item is selected or pressed and you change the selection in the table, the plot will automatically update to show the data corresponding to the new selection. This item is permanently disabled for plots which did not originate from a table but were instead created from another plot.

Note that for certain types of plots this item is pressed by default, whereas for other plot types it is not.

5.1.5.5 Percentage Y-Axis

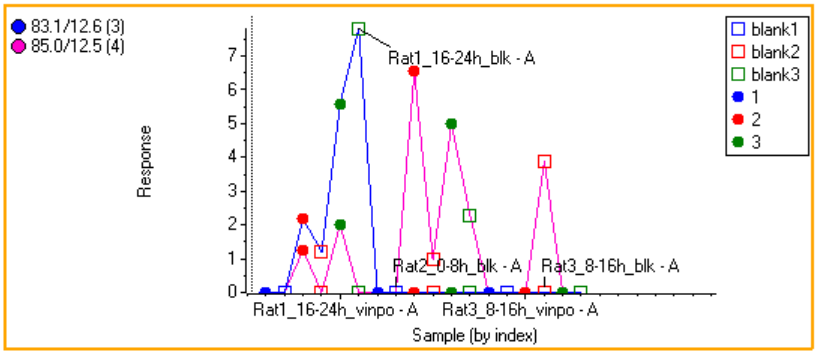
 This button determines the y-axis scaling: when checked, overlaid traces are scaled so that the maximum value for each trace is at 100%. Using a percentage y-axis is convenient if the absolute magnitude of the traces is very different. It is also useful for visually assessing the degree of correlation of overlaid traces.

5.2 Profile Plot

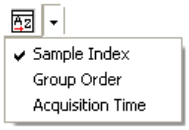
Profile Plots display peak response as a function of sample for a given peak. As shown in the figure you can overlay traces for multiple peaks within the same plot. These plots are very useful for discovering how particular

compounds vary for your samples. As mentioned in *the Check Peak Alignment* discussion in section 0, they are also useful for verifying possible peak alignment problems.

You can create a Profile Plot in a number of different ways as discussed in detail elsewhere in the document. For example you can select particular rows in a Peaks Table and use the *Plot Profile for Peak* context menu item or toolbar button. More likely you will create these plots from the displays which help you focus on peaks likely to be of particular interest, such as the t-test table or PCA Loadings Plot.



5.2.1 Sort Order Toolbar Button



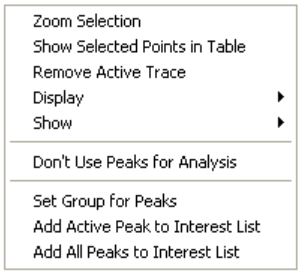
The items in this menu control the order in which the samples appear in the plot. Note that your last selection from this menu is remembered and automatically applied to future Profile Plots which you create.

If you select the *Sample Index* option the samples appear in the order in which they were originally imported. If you select the *Group Order* option they appear in order of alphabetical group name (within one group they appear in sample index order); you must assign group names in order to use this option. The *Acquisition Time* option is only available for data imported from wiff files: it orders the samples from earliest acquired to latest. This item is also not explicitly available if the acquisition order matches the *Sample Index* order.

Clicking the button directly (as opposed to selecting the menu items) automatically advances to the next possible sort. For example if the *Acquisition Time* item is not available because it matches the *Sample Index* order, clicking the button toggles between *Sample Index* and *Group Order*.

Ordering samples by *Group Order* will generally make it easiest to visualize any differences between groups. The *Acquisition Time* option, or the *Sample Index* option if you originally imported the samples in the same order in which they were acquired, can be very useful for revealing systematic trends in the data.

5.2.2 Context Menu



5.2.2.1 Zoom Selection

 This command zooms the plot so that the selected region fills the entire available space. Before selecting this item you need to make a selection in the plot by clicking and dragging. You can also zoom by clicking and dragging directly in the x-axis (or y-axis) of the plot.

5.2.2.2 Show Selected Points in Table

If the plot was created directly from a Peaks Table, this command selects the cells in the table which correspond to the currently selected samples. If the plot was created in another way such as from a t-Test Table this command is not available.

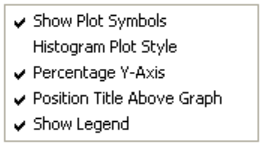
This command can be useful as a quick way to locate particular cells in the table. It can also be useful if you have created other plots from the Peaks Table (perhaps spectra or XICs) and wish to update them: if these other plots are linked to the Peaks Table, they will automatically update when the cells are selected in the table.

5.2.2.3 Remove Active Trace

When there is more than on overlaid trace in the plot, this command removes the one which is currently active.

This can be useful if you have created an overlaid Profile Plot from a PCA Loadings Plot using a number of selected peaks. If you subsequently wish to add *most* of these peaks to the interest list, you can remove the traces for the ones you do not wish to add and then use *the Add All Peaks to Interest List* command described below.

5.2.2.4 Display Sub-Menu



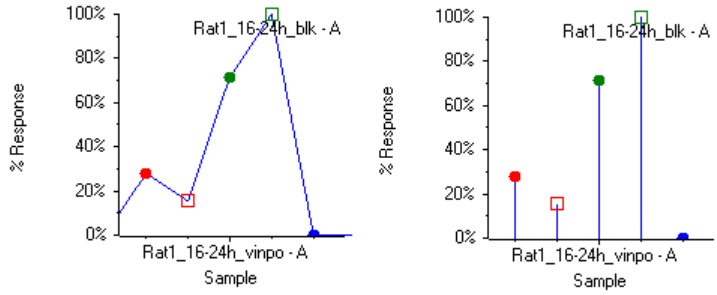
This menu contains a number of options for adjusting the exact appearance of the plot. Note that your last selections from this menu are remembered and automatically applied to future Profile Plots which you create. The same is true for the *Display* options for other plot types, however each different type of plot uses a different set of defaults.

5.2.2.4.1 Show Plot Symbols

This command allows you to display or hide plot symbols.

5.2.2.4.2 Histogram Plot Style

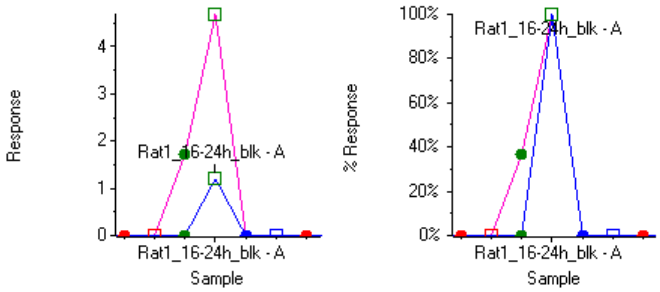
When this item is checked, connecting lines are not drawn between samples; instead a line is drawn from the point to the x-axis. The figure below illustrates the difference.



5.2.2.4.3 Percentage Y-Axis

 When this item is checked, overlaid traces are scaled so that the maximum value for each trace is at 100%. The figure below illustrates the difference.

Using a percentage y-axis is convenient if the absolute magnitude of the traces is very different. It is also useful for visually assessing the degree of correlation of overlaid traces.

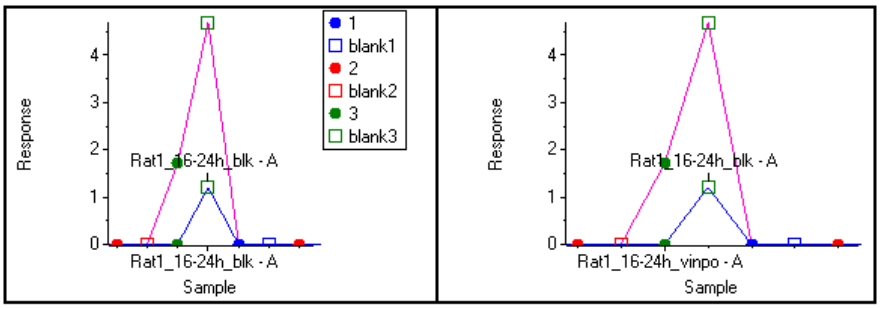


5.2.2.4.4 Position Title Above Graph

This command determines whether the plot title is drawn above the plot or to the left of the plot. This is illustrated in the figure in section 5.1.4.

5.2.2.4.5 Show Legend

This command determines whether or not a legend explaining the various sample plot symbols should be drawn to the right of the plot. This item is only available if you have associated *Group* names with the samples.



5.2.2.5 Show Sub-Menu

The commands in this sub-menu are all used to view raw data.

5.2.2.5.1 Show Spectra

This command displays the raw mass spectra and is available if you have processed LC/MS or MS data, but not if you have imported text data in the *Generic Text File* format. For details on the resulting spectra, see the MS Spectrum Plots discussion in section 5.12.

If you have made a selection in the Profile Plot, a new plot containing overlaid spectra for just those specific samples is created. If you have not made a selection, there are overlaid spectra for all samples.

5.2.2.5.2 Show Peak List Spectra

This command displays so-called Peak List Spectra which are 'virtual' mass spectra constructed using the list of detected peaks as opposed to the raw profile data. These plots are described in detail in section 5.13.

If you have made a selection in the Profile Plot, a new plot containing overlaid Peak List Spectra for just those specific samples is created. If you have not made a selection, there are overlays for all samples.

5.2.2.5.3 Show XICs

This command displays raw extracted ion chromatograms and is available if you have processed LC/MS data. For details on the resulting XICs, see the XIC Plots discussion in section 5.14.

If you have made a selection in the Profile Plot, a new plot containing overlaid XICs for just those specific samples is created. If you have not made a selection, there is an overlaid XIC for all samples. In general you will probably want to select just a few samples since calculating XICs for all samples may take a long time.

5.2.2.5.4 IDA Product Spectrum commands

If you have processed the MS data for a series of IDA runs, these commands are used to display IDA product spectra. They are identical to those described previously for the Peaks Table in sections 4.2.2.4.5, 4.2.2.4.6 and 4.2.2.4.7 – see those sections for details. For details on the resulting plots themselves, see the *IDA Product Spectrum Plot* discussion in section 5.15.

If you have made a selection in the Profile Plot, only the selected samples are examined for potential matching IDA spectra; if you have not made a selection all samples are used. Especially when using the 'Best' or 'All' options, you may wish to select a limited range of samples for best performance.

5.2.2.6 Don't Use Peaks for Analysis

This command unchecks the *Use* field for the peaks. If there is more than one overlaid trace this applies to all overlaid peaks, not just the currently active one.

5.2.2.7 Add Active Peak to Interest List

This command adds the peak for the active trace to the Interest List. The main reason for doing this is to subsequently report details for these peaks or possibly to export this interest list for use outside of the application. For details see the Interest List discussion in section 4.8 and also the *Generate Report in MS Word* command in section 2.1.12.

5.2.2.8 Add All Peaks to Interest List

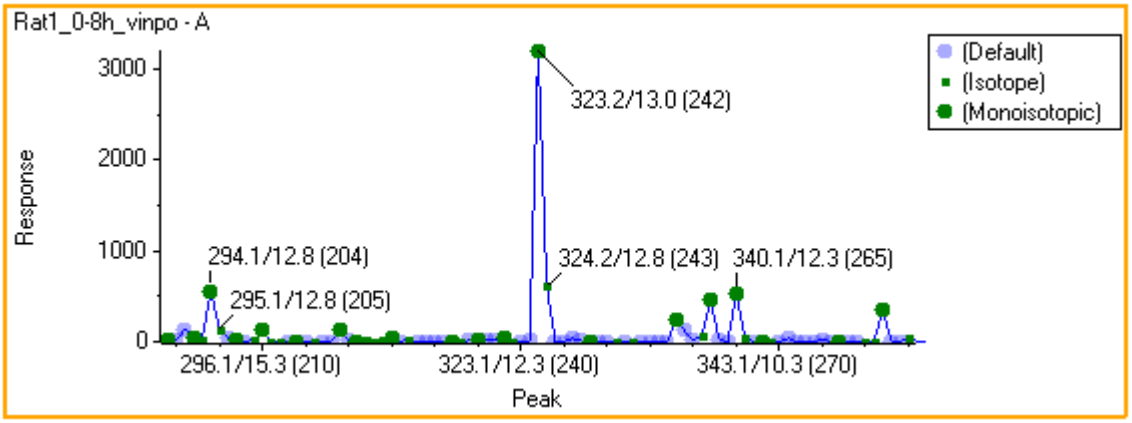
This command adds the peaks for the overlaid traces to the Interest List.

5.3 Peaks for Sample Plot

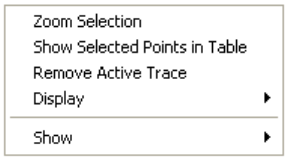
These plots display peak response as a function of peak index as shown in the figure below. Tools such as the t-test and PCA are more appropriate for finding differences between groups of samples, but these plots can be useful for visualizing high-level differences between individual samples. For example you can overlay traces for multiple samples in one plot or use multiple plots in separate panes to look for obvious visual differences.

As discussed elsewhere, you can create these plots by selecting one or more columns in a Peaks Table or one or more data points in a PCA Scores Plot and then using the *Plot Peaks for Sample* context menu item.

For MS and LC/MS data the x-axis represents peak index or mass/charge depending on the *Mass/Charge X-Axis* option as discussed below in section 5.3.1.1.



5.3.1 Context Menu



The figure above shows the various items which are available in the context menu. With the one exception described immediately below, all of these items, including the items in the *Display* and *Show* sub-menus, are also present in the context menu for Profile Plots. See the various sub-sections of section 5.2.2 for a detailed discussion.

5.3.1.1 Display – Mass/Charge X-Axis

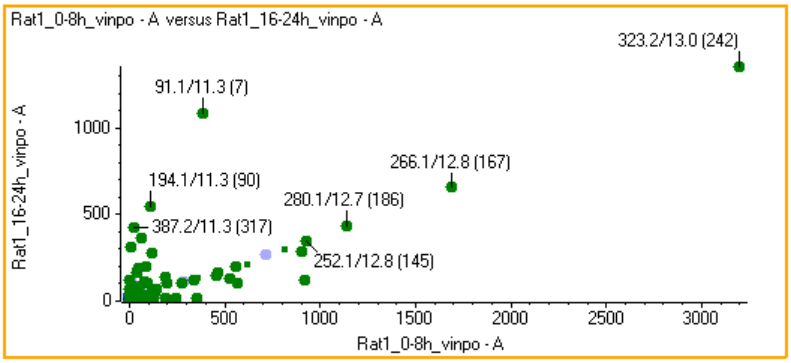
This option of the *Display* sub-menu controls whether the x-axis for MS and LC/MS data represents peak index or mass/charge.

When this option is not selected, the axis is labeled with the actual names of the peaks; in other words the plot is really a type of category plot. This may be more obvious if you use the *Display→Histogram Plot Style* context menu item to view the plot without connecting adjacent points with lines. Note that when using the mass/charge option for LC/MS data, distinct peaks with the same m/z but different retention time will be plotted using the same x-value.

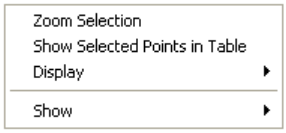
5.4 Peaks for Two Samples Plot

These plots display the peak responses for two samples as shown in the figure below. This provides a convenient way to visually compare two samples in detail. As you can see in the figure, most of the larger peaks appear to be present in both samples with a roughly constant ratio; however the *91.0/11.3* peak (amongst others) is relatively much larger in the first sample.

You create these plots by selecting the two columns of a Peaks Table for the desired samples and then choosing the Plot Peaks for Two Samples context menu item or toolbar button.



5.4.1 Context Menu

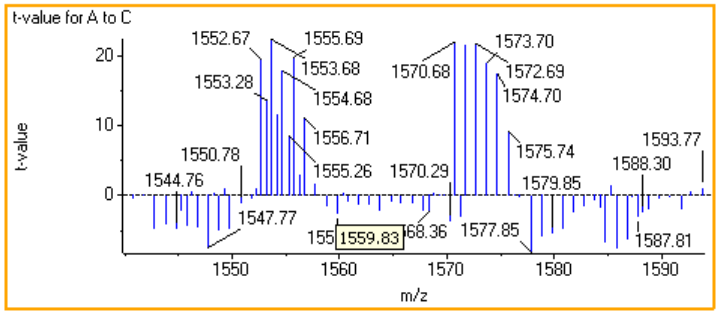


The figure above shows the various items which are available in the context menu. All of these items are also present in the context menu for Profile Plots and were described in detail in section 5.2.2.

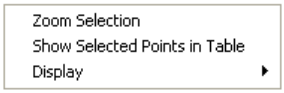
5.5 t-Test Plot (for One Column)

These plots display the various t-test metrics (t-value, p-value, *etc.*) against peak index or mass/charge for one or more columns of a t-Test Table. These columns represent the comparison of two particular groups of samples to one another.

You create these plots by selecting one or more columns in a t-Test Table and selecting the *Plot Column* context menu item or the corresponding toolbar button as discussed in section 4.1.1.2. The figure below shows the raw t-value plotted against m/z. Plotting t-values, as opposed to p-values, allows for more effective use of the dynamic range since they do not cluster at 0.0 and 1.0. In the zoomed figure below it can be seen that the peak near 1553 Da and its isotopes is much larger in group *A* than in group *C*.



5.5.1 Context Menu

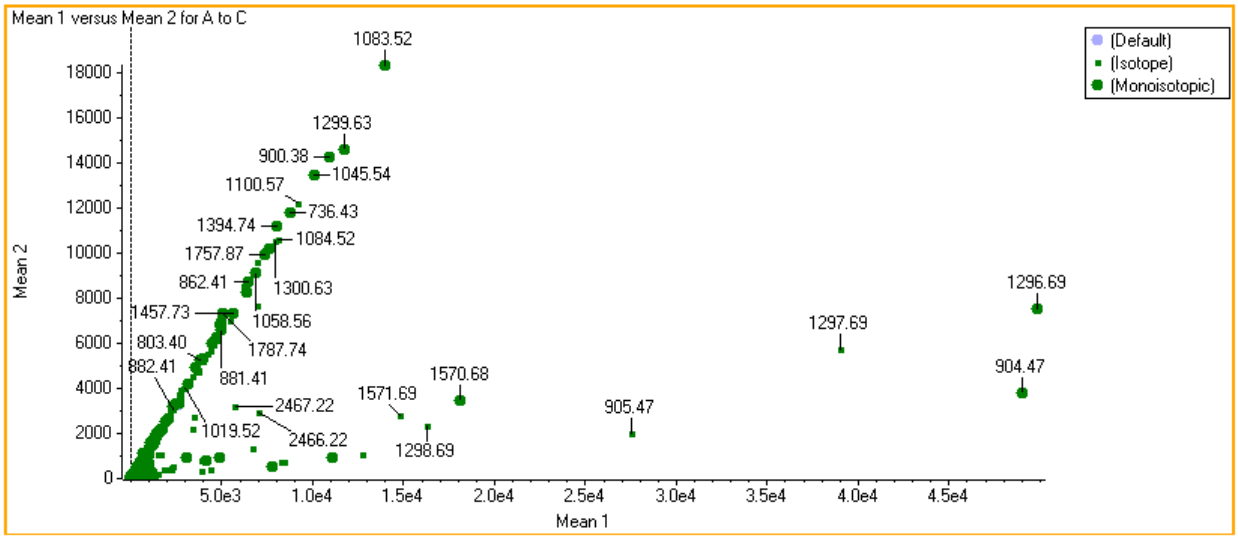


The figure above shows the various items which are available in the context menu. With one exception, all of these items, including the items in the *Display* sub-menu, are also present in the context menu for Profile Plots – see the various sub-sections of section 5.2.2 for a detailed discussion. The *Mass/Charge X-Axis* option of the *Display* sub-menu is discussed in section 5.3.1.1.

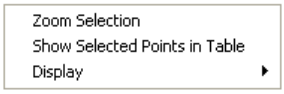
5.6 t-Test Plot (for Two Columns)

These plots display any two t-test metrics (t-value, p-value, *etc.*) as a function of one another.

You create these plots by selecting exactly two columns in a t-Test Table and selecting the *Plot Two Columns* toolbar button or menu item; there are also shortcuts for creating commonly used plots as discussed in section 4.4.1. The figure below plots the average values for each peak for each of the two groups (*A* and *C* in this case). It can be seen immediately that many of the peaks have similar response for the two groups, but that there are a number of peaks (1296.69, *etc.*) which are much larger in the first group.



5.6.1 Context Menu

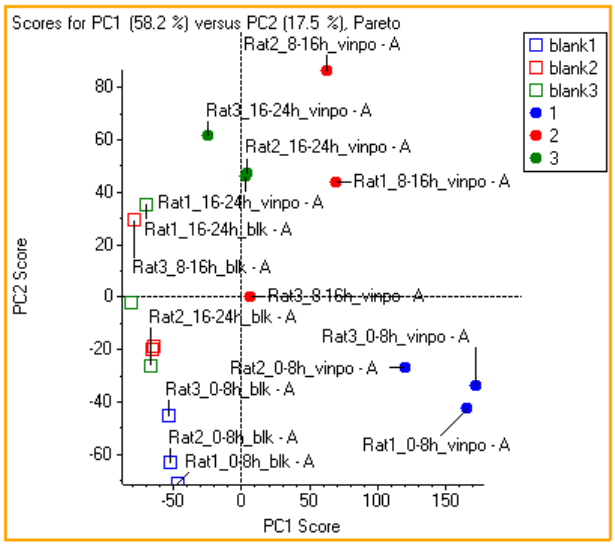


The figure above shows the various items which are available in the context menu. All of these items, including the items in the *Display* sub-menu, are also present in the context menu for Profile Plots – see the various sub-sections of section 5.2.2 for a detailed discussion.

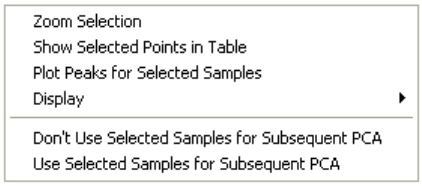
5.7 Scores Plot (for Two PCs)

These plots display PCA scores for all samples for two principal components as shown in the figure below. If you perform PCA you will use these plots extensively to visualize how your various samples group together. For example the figure below indicates that the three samples belonging to group *1* – in the lower-right corner of the graph – are well separated (mostly due to PC1) from all other samples.

You create these plots by selecting two PC columns in a PCA Scores Table and choosing *Plot Scores for Two PCs* from the context menu or by using the corresponding toolbar button as shown in section 4.5.1.3. That said, the initial PCA window containing the Scores Table automatically creates a Scores Plot for the first two principal components which is linked to the table. So, you really only need to use this command to generate a new Scores Plot if you delete the automatically generated plot or if you wish to display two simultaneous plots for different principal components.



5.7.1 Context Menu



The *Zoom Selection and Show Selected Points in Table* commands are also present in the context menu for Profile Plots – see the various sub-sections of section 5.2.2 for a detailed discussion of these commands. The remaining commands are described below.

5.7.1.1 Plot Peaks for Selected Samples

Creates a plot of peak response as a function of peak index or, in the case of MS (not LC/MS data) data, mass/charge. There is a separate overlaid trace for each selected sample. These plots were discussed in section 5.3.

5.7.1.2 Display

5.7.1.2.1 Show Legend

This command determines whether or not a legend explaining the various sample plot symbols should be drawn to the right of the plot. This item is only available if you have associated *Group* names with the samples.

5.7.1.2.2 Show Labels

This command determines whether the data points should be automatically labeled using the Sample Name field.

5.7.1.3 Don't Use Selected Samples for Subsequent PCA

De-selects the *Use* field for the selected samples. The net effect is identical to un-checking the *Use* field for the appropriate samples in the PCA Scores Table. As explained for the *Plot Symbols* tab of the *Options* dialog in section 2.2.8.1, the point symbol for these unused samples will change to that specified for the *Excluded* group (by default a hollow circle).

This command is useful if you wish to exclude outlier samples from a subsequent PCA analysis. Note that the existing PCA display will not change: you need to perform a new PCA analysis.

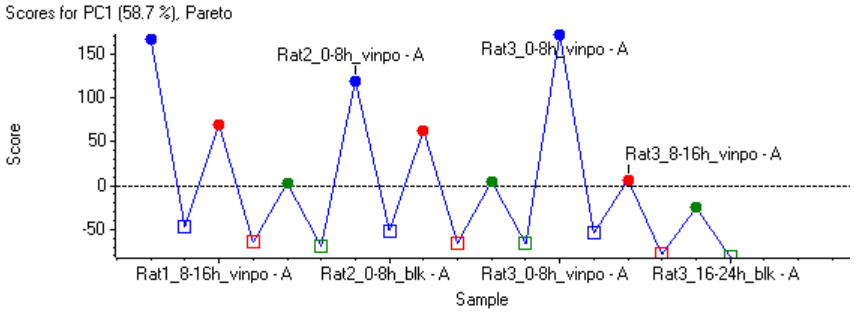
5.7.1.4 Use Selected Samples for Subsequent PCA

Selects the *Use* field for the selected samples: this is the reverse of the command discussed immediately above. The net effect is identical to checking the *Use* field for the appropriate samples in the PCA Scores Table.

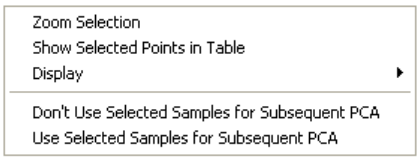
5.8 Scores Plot (for One PC)

These plots display PCA scores for one specific principal component for all samples as shown in the figure below. Most often you will probably plot two principal components as a function of one another, rather than just one as here, since this will provide additional separation power. However at least for cases in which any one PC does a good job of separating the samples, plotting that one PC may be useful. For example in the figure the samples represented by the filled blue circles have the largest PC1 scores and separate from the other samples. These 1D plots are also used for the degenerate case in which only a single principal component is available.

You create these plots by selecting a PC column in a PCA Scores Plot and choosing *Plot Scores for PC* from the context menu or by using the corresponding toolbar button as discussed in section 4.5.1.2.



5.8.1 Context Menu



The figure above shows the various items which are available in the context menu. All of these items have been described in previous sections.

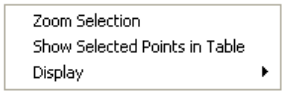
The *Zoom Selection* and *Show Selected Points in Table*, as well as the items in the *Display* sub-menu, are also present in the context menu for Profile Plots. See the various sub-sections of section 5.2.2 for a detailed discussion. The *Don't Use Selected Samples for Subsequent PCA* and *Use Selected Samples for Subsequent PCA* commands were discussed for Scores Plots involving two principal components in the previous section.

5.9 Scores for Sample Plot

These plots display PCA scores for all principal components for a specific sample. You will probably use these plots only infrequently, however they may be useful if you wish to study one sample in detail.

You create these plots by selecting the row corresponding to the desired sample in a Scores Table and selecting *Plot Scores for Sample* from the context menu or by using the toolbar button as discussed in section 4.5.1.1.

5.9.1 Context Menu



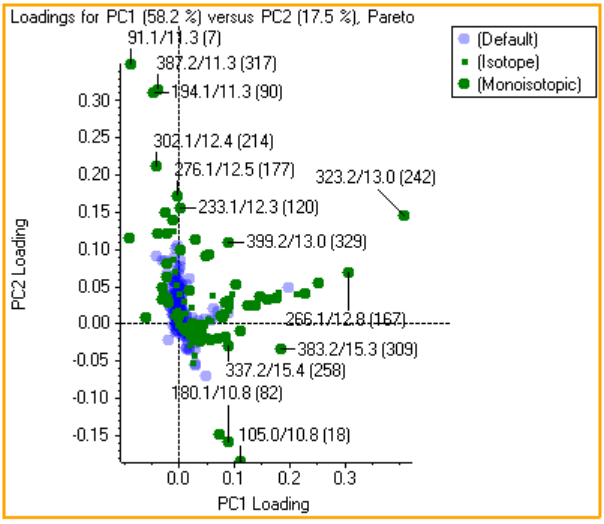
The figure above shows the various items which are available in the context menu. All of these items, including the items in the *Display* sub-menu, are also present in the context menu for Profile Plots – see the various sub-sections of section 5.2.2 for a detailed discussion.

5.10 Loadings Plot (for Two PCs)

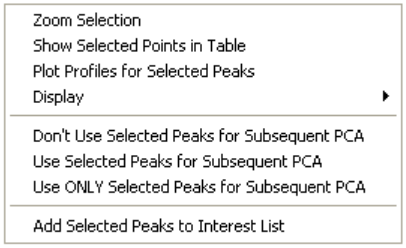
These plots display PCA loadings for all samples for two principal components as shown in the figure below. If you perform PCA, you will use these plots extensively to visualize which peaks or variables are responsible for separating your samples in a corresponding Scores Plot.

You will probably also use the *Plot Profiles for Selected Peaks* toolbar button or context menu item discussed further below to visualize how particular peaks are behaving for your samples

You create these plots by selecting two PC columns in a PCA Loadings Table and choosing *Plot Loadings for Two PCs* from the context menu or by using the corresponding toolbar button as mentioned in section 4.6.1.2. That said, the initial PCA window containing the Loadings Table automatically creates a Scores Plot for the first two principal components which is linked to the table. So, you really only need to use this command to generate a new Loadings Plot if you delete the automatically generated plot or if you wish to display two simultaneous plots for different principal components.



5.10.1 Context Menu



The *Zoom Selection*, *Show Selected Points in Table* and *Display - Show Legend* commands are also present in the context menu for Profile Plots. See the sub-sections of section 5.2.2 for a detailed discussion of these commands. The remaining commands are described below.

5.10.1.1 Plot Profiles for Selected Peaks

This command creates a plot of peak response as a function of sample (note that there is also a corresponding toolbar button). There is a separate overlaid trace in the plot for each selected peak. These plots can be very useful for visualizing how particular peaks are behaving across the various samples – for details on the resulting plot, see the discussion of Profile Plots in section 5.2.

Once you have created a plot using this command:

- If you select a different group of peaks and choose the command again, a new Profile Plot is created. However if you hold the Shift key while selecting the menu item, the previous Profile Plot is updated. This saves having to delete the original plot if you do not want both.
- If you single-click directly on a peak in the Loadings Plot, the Profile Plot will update to show the profile for that peak only. This is a convenient way to quickly view profiles for different peaks.
- If you hold the Shift key while single-clicking directly on a peak in the Loadings Plot, any peaks already present in the Profile Plot are left in place and the peak for the clicked point is added.

5.10.1.2 Don't Use Selected Peaks for Subsequent PCA

This command de-selects the *Use* field for the selected peaks. The net effect is identical to un-checking the *Use* field for the appropriate peaks in the PCA Loadings Table. As explained for the *Plot Symbols* tab of the *Options* dialog in section 2.2.8.1, the point symbol for these unused samples will change to that specified for the *Excluded* group (by default a hollow circle).

You will probably use this command frequently to exclude peaks which you believe are not relevant so that PCA can be re-run without them. For example if you wish to observe the effect of a drug on endogenous metabolites, you will probably want to exclude the xenobiotic drug metabolites. Note that once you have excluded peaks in this way, you can use the commands in the context menu of the Excluded Peaks Table to add them to the global exclusion list; this will allow you to exclude these peaks up front when importing data in the future.

5.10.1.3 Use Selected Peaks for Subsequent PCA

This command selects the *Use* field for the selected peaks: this is the reverse of the command discussed immediately above. The net effect is identical to checking the *Use* field for the appropriate peaks in the PCA Loadings Table.

5.10.1.4 Use ONLY Selected Peaks for Subsequent PCA

This command selects the Use field for the selected peaks *and* de-selects this field for all other peaks.

This command is useful if you wish to re-run PCA using only a targeted set of the peaks to discover how well this subset is able to separate your samples. For example you might select only those peaks somewhat near the center of the Loadings Plot to ignore all peaks which are around the outside edges (*i.e.* all peaks which contribute significantly to the loadings for the two PCs being plotted).

5.10.1.5 Add Selected Peaks to Interest List

This command adds the selected peaks to the interest list. The main reason for doing this is to subsequently report details for these peaks or possibly to export this interest list for use outside of the application. For details see the Interest List, discussed in section 4.8, and also the *Generate Report in MS Word* command in section 2.1.12.

As mentioned elsewhere in this document, it is important that you understand that when PCA is performed on (for example) a Peaks Table, the PCA window uses a *copy* of the original data in the Peaks Table. So if you do the following:

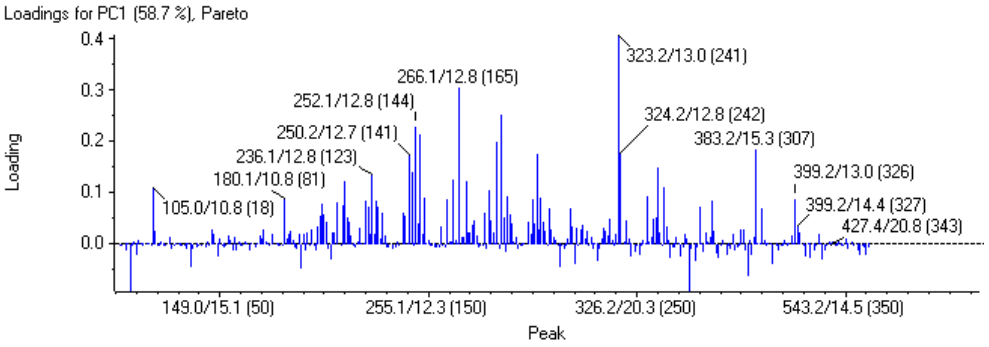
- Use PCA to process data in a Peaks Table (or another PCA window)
- Add selected peaks to the interest list
- Close the PCA window and return to the original Peaks Table

the peaks will not have been added to the interest list for the Peaks Table: they were added to the interest list for the now closed PCA window. Similarly, if you added peaks to the interest list for the Peaks Table before performing PCA, they would also be present in the interest list for the PCA window; but if you added them to the Peaks Table after performing PCA, the PCA window would not be affected.

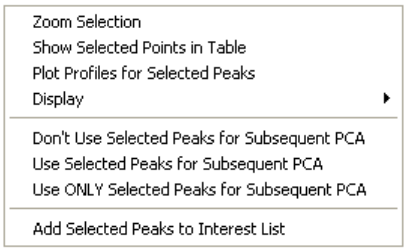
5.11 Loadings Plot (for One PC)

These plots display PCA loadings for one specific principal component for all samples as shown in the figure below. Most often you will probably plot two principal components as a function of one another, rather than just one as here, however this display can be useful if you have decided that a particular principal component is important and wish to focus attention on that one PC. These 1D plots are also used for the degenerate case in which only a single principal component is available.

You create these plots by selecting a PC column in a PCA Loadings Plot and choosing *Plot Loadings for PC* from the context menu or by using the corresponding toolbar button as described in section 4.6.1.1.



5.11.1 Context Menu



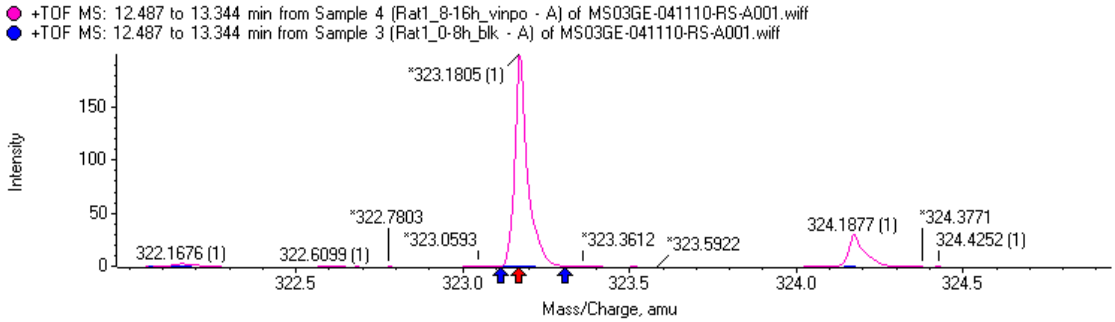
The figure above shows the various items which are available in the context menu. All of these items have been described in previous sections.

The *Zoom Selection* and *Show Selected Points in Table* commands were described for Profile Plots in section 5.2.2. With one exception, the items in the *Display* sub-menu were described in section 5.2.2.4. The *Mass/Charge X-Axis* option of the *Display* sub-menu was discussed in section 5.3.1.1. The remaining items were discussed for Loadings Plots involving two principal components in the previous section.

5.12 MS Spectrum Plots

These plots display raw mass spectra as shown in the figure below. They are only available if you have processed LC/MS or MS data, but not if you have imported text data in the *Generic Text File* format. These plots are very useful for verifying that apparent differences are real and are not a result of incorrect peak finding or poor peak alignment.

You can create these plots in a number of different ways as discussed elsewhere in this document. Most frequently you will probably make selections in a Profile Plot and use its *Show Spectra* context menu item. In order to generate these plots, you must have access to the original data files used when the data was imported. If you have moved these files, the program will prompt you to locate them by displaying a standard *Open File* dialog.



For LC/MS data the reported spectrum is averaged over the retention time range detected for *all* samples which aligned to the selected peak: the time range is from minimum detected LC peak start for all samples to the maximum detected LC peak end for all samples. So, the same retention time range is used to calculate the average for all samples, not a specific retention time range for each sample.

For a description of the format of the labels (the meaning of the '*', *etc.*) see the *Mass (charge)* discussion in section 5.12.2.2.2. The m/z values marked with blue arrows (⬆) correspond to the peak start and peak end detected by the peak finder when the data was imported for *all* samples. In other words the left arrow corresponds to the peak start for the sample with the very lowest detected start and vice versa for the right arrow.

For LC/MS data or for MS data imported using a peak-finding approach, the red arrow in the m/z axis (⬆) marks the centroid location determined by the peak finder when the data was first imported. Note that this reported m/z value refers to the aligned sample with largest intensity, so in general the arrow will not be exactly correct for other samples. Also note that for LC/MS data, the m/z value reported for the peak will not exactly match that used in the labeling of the plot, even for the sample with largest intensity: the program does not directly detect peaks in this averaged spectrum. For MS data imported using a binning approach, the arrow simply marks the location of the bin which will not in general correspond to a peak.

The plot is initially zoomed so that only a window around the arrow is visible, however you can return to the Home view to see the entire m/z range in the usual way. However for LC/MS data is essential to understand that apparent differences between the samples may be misleading for any peaks other than the one indicated with the arrow. The retention time range averaged should always be reasonable for all samples for the marked peak, but it may well happen that this time range captures most of a *different* m/z peak for one sample, but only a portion of it for another sample.

As explained below, depending on the *Multiply by Sample Scale Factor* setting, note that the intensities may have been multiplied by the sample scale factors.

5.12.1 Update Peaks Only Toolbar Button

 In addition to the standard toolbar buttons discussed in section 5.1.5, spectra and XICs have one additional button. This item determines how the plot should update when it is linked to the Peaks Table which was used to create it. This is only applicable when the plot is linked (using the  button).

If this button is pressed, the plot will continue to show spectra or XICs for the samples used to create the plot initially, regardless of the actual samples currently selected in the table – in other words only the peaks are linked to the table, not the samples. This allows you to make 'lazy' selections in the Peaks Table by selecting an entire row without causing all samples to also be used.

5.12.2 Context Menu

The *Zoom Selection* and *Remove Active Trace* commands are also present in the context menu for Profile Plots – see the sub-sections of section 5.2.2 for a detailed discussion of these commands. The remaining commands are described below.

5.12.2.1 Display Sub-Menu

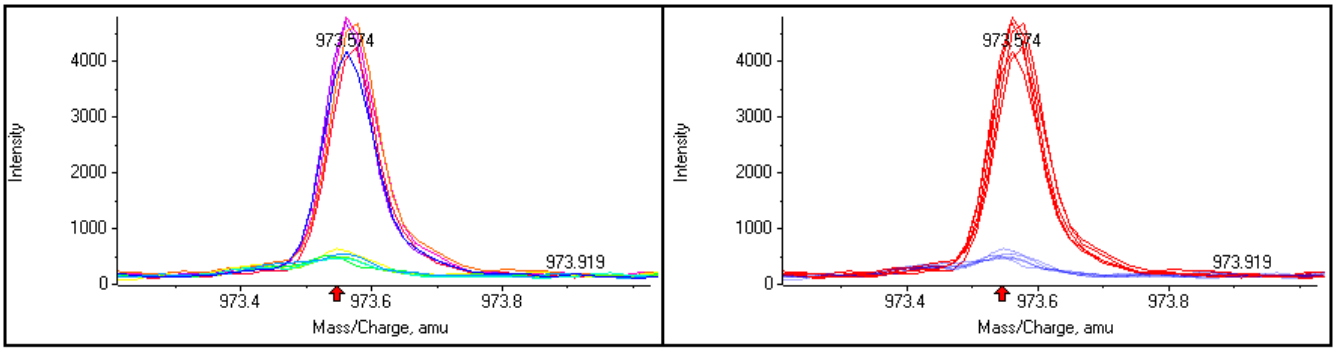
5.12.2.1.1 Percentage y-Axis

When this item is checked, overlaid traces are scaled so that the maximum value for each trace is at 100%. This was illustrated in section 5.2.2.4.3.

5.12.2.1.2 Use Group Colors for Traces

By default, each overlaid spectrum is plotted using a color determined by the program. If you select the *Use Group Colors for Traces* option, the color for each spectrum is taken from the plotting symbol for the group to which the corresponding sample belongs – see section 2.2.8.1 for a discussion of the plotting symbols.

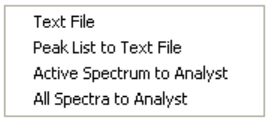
This option is useful if you are comparing several spectra each from different groups since it allows you to easily determine which spectra belong to which groups. For example in the figure below on the left it is not immediately obvious that all of the spectral peaks from one group are larger than those in another, whereas the figure on the right (with the option enabled) makes this clear – all of the smaller peaks on the right are in blue and the larger ones are in red. If you want to distinguish individual samples, you should disable this option since the default color scheme uses a different color for each trace.



5.12.2.1.3 Multiply by Sample Scale Factor

If you have normalized the responses for the various peaks, using internal standards or otherwise, this command determines whether the intensities of the spectra or XICs should also be multiplied by the scale factor. If this item is not checked, the raw data is displayed as-is. If you have previously verified that the scale factors are reasonable, you may wish to check this item.

5.12.2.2 Export Sub-Menu



5.12.2.2.1 Text File

This command saves the {mass/charge, intensity} data points for the active spectrum (in the case that there is more than one overlaid trace) to a text file of your choosing. The mass/charge value is written as the first column and the intensity as the second column, separated from the first column by a tab; there is a separate line terminated with a carriage return for each data point.

5.12.2.2.2 Peak List to Text File

This command saves the peak list for the active spectrum to a tab-delimited text file of your choosing. There is an initial header row with an additional row for each peak. The following columns are written:

Column Name	Description
Mass/Charge	The mass/charge for the peak.
Area	The peak area.
Height	The height of the peak. This will be somewhat less than the raw intensity if there is a baseline offset.
Width	The width of the peak from peak start to peak end.
Width at 50%	The width of the peak at an intensity equal to half of the peak height.
Resolution	An estimate of the resolution for the peak. Note that this estimate may not be meaningful for peaks with low intensity.
Charge	The charge state for the peak. A value of '0' indicates that the charge could not be determined.
Monoisotopic	'Yes' is the peak was called as the monoisotopic peak of an isotope cluster *or* if the charge state could not be determined; 'No' otherwise.
Mass (charge)	The mass/charge with the charge state following in brackets. This is the format used for labeling spectral plots. If the charge state could not be determined it is not written. An initial '*' indicates that the peak was called as the monoisotopic peak of an isotope cluster *or* that the charge state could not be determined. For example '*100.00 (2)' indicates a monoisotopic peak with a mass/charge of 100 and a charge state of 2; '100.50 (2)' indicates an isotope with an m/z of 100.5 and a charge of 2; '*100.00' indicates a peak with an m/z of 100 for which the charge state could not be determined.

5.12.2.2.3 Active Spectrum to Analyst

This command opens a copy of the currently active spectrum in the Analyst® Software. This is useful so that you can explore the data further using the various tools available there. Note that this command is only available for data originating from an Analyst® wiff file.

5.12.2.2.4 All Spectra to Analyst

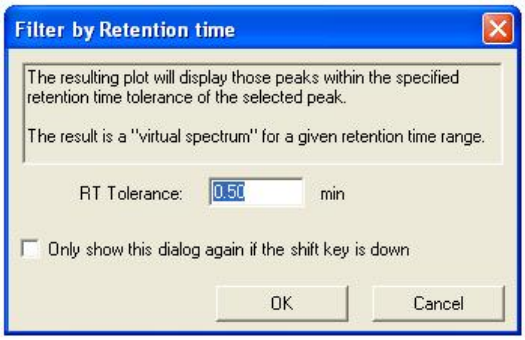
This command opens a copy of all of the overlaid spectra in the Analyst® Software.

5.13 Peak List Spectrum Plots

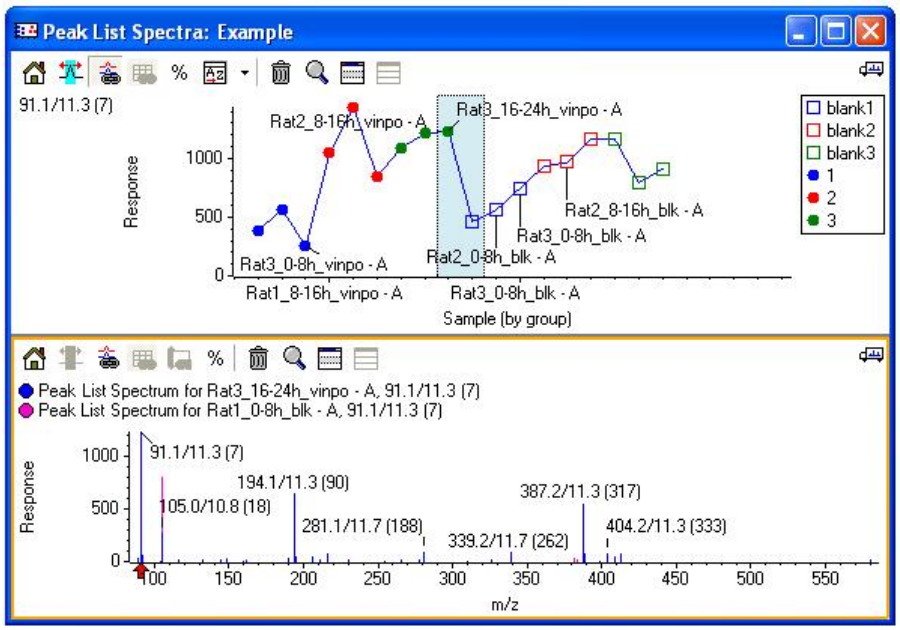
For LC/MS data, these plots display so-called spectral peak lists as 'virtual' mass spectra constructed using the list of detected peaks as opposed to the raw profile data. They are conceptually similar to what would be obtained by showing profile mass spectra for a particular peak as described in the previous section and then centroiding each of those spectra.

These displays are useful in the same situations as it is useful to examine the raw profile spectra (for example looking for fragments, adducts or multiple-charge states of a peak of interest, *etc.*), however they can be generated much more quickly. They can also be displayed if the original data file is not available. Of course the raw spectra themselves need to be examined when judging the accuracy of the peak-finding process itself.

After selecting the *Show Peak List Spectra* command, the dialog shown in the figure below is presented. The resulting Peak List Spectra will contain a data point for each peak which is within the specified retention time tolerance of the peak of interest.



The figure below shows a Profile Plot for which two samples have been selected in the top pane and the resulting Peak List Spectrum in the bottom pane. A separate overlaid trace was created for each selected sample (if no samples were selected, a trace would have been generated for each available sample). The m/z value corresponding to the peak is marked with a red arrow in the x-axis of the plot.



5.13.1 Context Menu

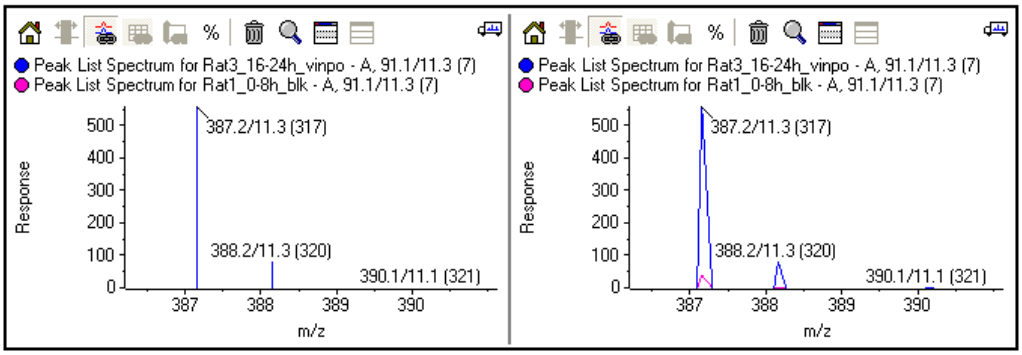
The *Zoom Selection* and *Remove Active Trace* menu items were both discussed for Profile Plots – see the sub-sections of section 5.2.2 for a detailed discussion of these commands

5.13.1.1 Display Sub-Menu

These commands were all discussed in the sub-sections of section 5.2.2, however the *Histogram Plot Style* option behaves somewhat different and is described below.

5.13.1.1.1 Histogram Plot Style

When this item is checked, a vertical line is drawn from each data point to the x-axis as shown in the left-hand pane in the figure below. When this item is unchecked, each data point is drawn as a triangle as shown in the right-hand figure below. The base of the triangle lies on the y=0 line and extends from the lowest start mass detected for *all* samples which aligned to the selected peak to the corresponding highest mass. The advantage of this plotting style is that the active trace doesn't obscure the others.



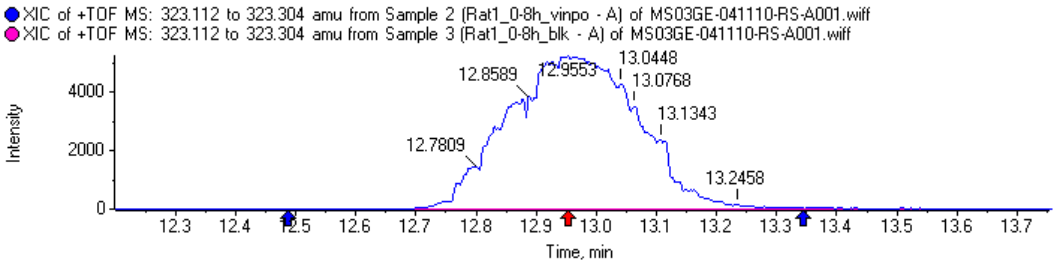
5.13.1.2 Export Sub-Menu

All of the menu items in this sub-menu are described for MS Spectra in section 5.12.2.2.

5.14 XIC Plot

These plots display raw extracted ion chromatograms for LC/MS data as shown in the figure below. Similarly to Spectrum Plots, these plots are very useful for verifying that apparent differences are real and are not a result of incorrect peak finding or poor peak alignment.

You can create these plots in a number of different ways as discussed elsewhere in this document. Most frequently you will probably make selections in a Profile Plot and use its *Show XICs* context menu item. In order to generate these plots, you must have access to the original data files used when the data was imported. If you have moved these files, you will be prompted to locate them by displaying a standard *Open File* dialog.



The mass range used for the XIC calculation corresponds to the spectral peak start and peak end detected by the peak finder when the data was imported for *all* samples. In other words the starting m/z of the XIC window corresponds to the spectral peak start for the sample with the lowest detected starting m/z and vice versa for the ending m/z of the XIC window. So, the same mass range is used to calculate the XIC for all samples, not a specific mass range for each sample

The retention times marked with blue arrows ( ) correspond to the retention time range detected for *all* samples which aligned to the selected peak: the left arrow is the overall minimum detected LC peak start for all samples and the right arrow is the overall maximum detected LC peak end.

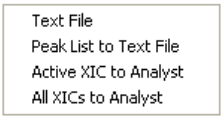
The red arrow in the retention time axis ( ) marks the LC peak apex determined by the peak finder when the data was first imported. Note that this reported value refers to the aligned sample with largest intensity, so in general the arrow will not be exactly correct for other samples. Also note that even for the sample with largest intensity, the arrow may not correspond exactly to the peak apex, especially if the background subtraction option was selected during the peak finding stage; the reported retention time refers to the background subtracted data in this case.

The plot is initially zoomed so that only a window around the arrow is visible, however you can return to the Home view to see the entire retention range in the usual way.

5.14.1 Toolbar and Context Menu

Most of the toolbar buttons and context menu items are essentially identical to those for Spectrum Plots – see section 5.12 for details. The exceptions are described below.

5.14.1.1 Export Sub-Menu



5.14.1.1.1 Peak List to Text File

This command saves the peak list for the active chromatogram to a tab-delimited text file of your choosing. There is an initial header row with an additional row for each peak. The following columns are available:

Column Name	Description
Time	The retention time in minutes for the peak.
Area	The peak area.
Height	The height of the peak. This will be somewhat less than the raw intensity if there is a baseline offset.
Width	The width of the peak in minutes from peak start to peak end.
Width at 50%	The width of the peak in minutes at an intensity equal to half of the peak height.

5.14.1.1.2 Active XIC to Analyst

This command opens a copy of the currently active XIC in the Analyst® Software. This is useful so that you can explore the data further using the various tools available there. Note that this command is only available for data originating from an Analyst® wiff file.

5.14.1.1.3 All XICs to Analyst

This command opens a copy of all of the overlaid XICs in the Analyst® Software.

5.15 IDA Product Spectrum Plot

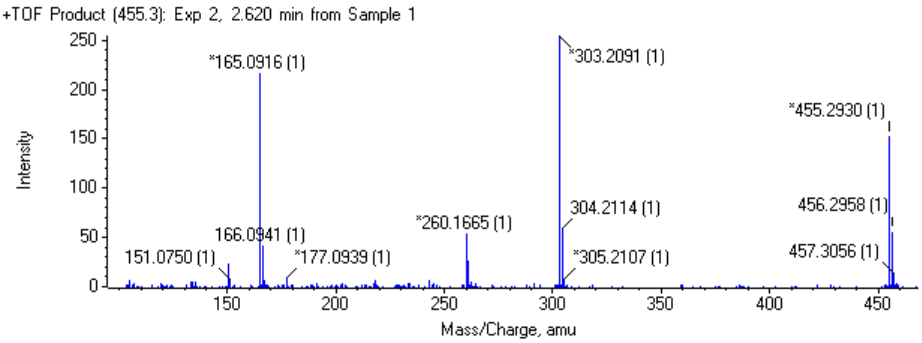
These plots display raw product spectra obtained during an IDA analysis and are available if you have processed the MS data for a series of IDA runs.

These plots are useful for manual structural interpretation. As explained below you can export the spectra to text files, possibly for importing into the Analyst® Software for subsequent data processing or, for the case of peptide product spectra, to an MGF file for protein database searching.

You create these plots by selecting from the context menu of a Profile Plot or, less commonly, a Peaks Table – see sections 4.2.2.4.5, 4.2.2.4.6 and 4.2.2.4.7 for details. As discussed in those sections, the intention is to allow you to quickly locate a product spectrum (if acquired) corresponding to a specific peak of interest. There are options for displaying all matching product spectra or an individual one. An IDA dependent spectrum is considered as a match for a particular peak if its retention time is within the entire range of the aligned peak detected in the MS data (*i.e.* from the earliest peak start to the latest peak end) and if its precursor mass/charge is within 0.4 Da of the m/z of the MS peak.

You can also create these plots by adding selected peaks to the interest list (see section 4.8) and using the *Include 'Best' IDA Product Spectrum* reporting option as discussed in section 2.1.12.

In order to generate these plots, you must have access to the original data files used when the data was imported. If you have moved these files, the program will prompt you to locate them by displaying a standard *Open File* dialog.



The above figure shows an example product spectrum. For a description of the format of the labels (the meaning of the '*', *etc.*) see the *Mass (charge)* discussion in section 5.12.2.2.2. The mass range is initially zoomed so that peaks less than 1% of the base peak near the beginning or end of the range are not visible, however you can return to the Home view to see the entire mass range in the usual way.

5.15.1 Toolbar and Context Menu

Most of the toolbar buttons and context menu items are identical to those for MS Spectrum Plots – see section 5.12 for details. The two additional items are described below.

5.15.1.1 Sum Overlaid Spectra

If you have used the *All IDA Product Spectra* command (section 4.2.2.4.7) to generate overlaid product spectra, this command is used to sum those spectra into a single trace. This is useful for increasing the signal/noise of peaks in weak spectra.

5.15.1.2 Export -> Mascot MGF Context Menu Item

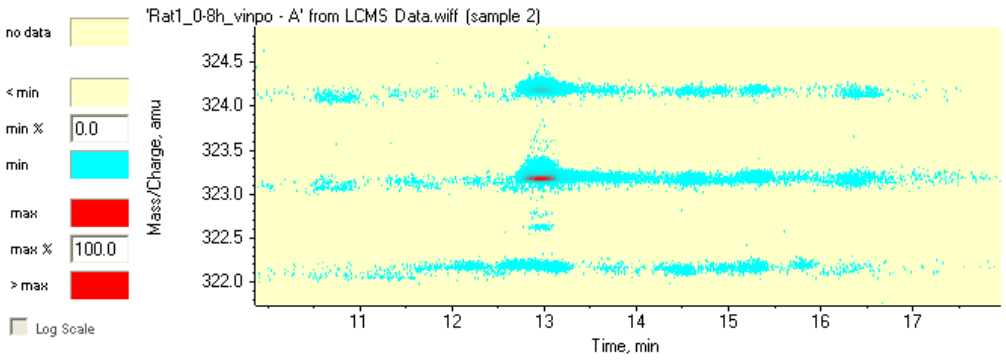
This command is used to save the peak list to an '.mgf' (Mascot Generic Format) text file for use with the Mascot protein search engine. Other search engines such as X!Tandem are also able to process files in this format. This command is only applicable to product spectra of peptides.

Note that the observed precursor mass/charge is written to the file and overrides any value set manually in the Mascot search options. However the charge state is *not* written to the file and should be specified in the search options.

5.16 Contour Plot

These heat map plots display retention time on the x-axis, mass/charge on the y-axis and raw intensity as a color coded value for LC/MS data. These plots are particularly useful for reviewing the performance of the peak finder; this is explained in detail in the accompanying User Manual. These plots are generated by choosing the *Show -> Contour* context menu item when a Peaks Table is active; you must select exactly one column (or cell) in the table representing the sample of interest before selecting the command.

The figure below shows an example contour plot along with the tools used to adjust the color mapping. You interact with these plots in the same way as for other plots as discussed in section 5.1.

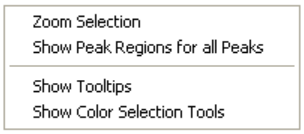


5.16.1 Color Selection Tools

The tools shown to the left of the plot in the figure above are used to adjust the exact appearance of the image. If you are focusing on a limited range of the data at a given moment, you can map a particular color range to a subset of the intensity range of the data as explained in the table below.

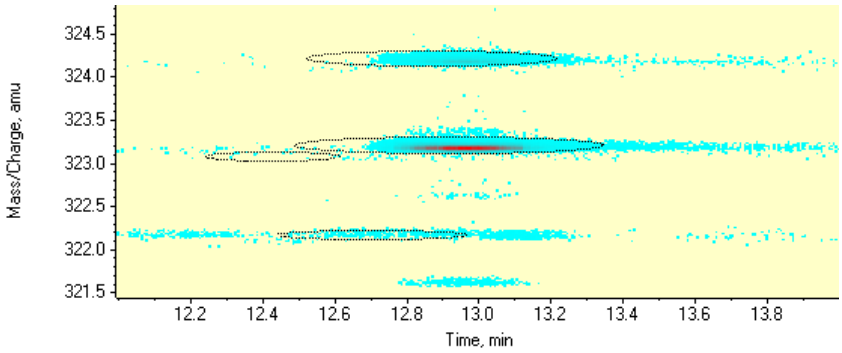
Item	Description
no data	Color used for data points which have an intensity of exactly zero (*i.e.* those which are missing).
< min	Color used for data points which have an intensity greater than zero, but less than the specified minimum intensity. If this is set equal to the *no data* color, peaks which are below the specified *min %* will appear to vanish.
min %	The intensity threshold for switching from the *<min* color to the *min* color. This value is expressed as a percentage of the maximum intensity for the entire data set.
min	Color used for data points which have an intensity equal to the specified minimum. The color for data points with an intensity greater than this value, but less than the specified maximum intensity, is determined by interpolating between this color and the *max* color.
max	Color used for data points which have an intensity equal to the specified maximum.
max %	The intensity threshold for switching from the *max* color to the *>max* color. This value is expressed as a percentage of the maximum intensity for the entire data set.
> max	Color used for data points which have an intensity greater than the specified maximum. Since there is a total of 256 colors available, reducing this *>max* value will enhance the visibility of small peaks – since all available colors can be used for them alone.
Log Scale	If this item is checked the intensity mapping using a logarithmic rather than a linear scale. Note that this is only applicable if the 'min %' is set to a value larger than zero.

5.16.2 Context Menu



5.16.2.1 Show Peak Regions for all Peaks

This command determines whether or not ellipses indicating the regions for which peaks were found should be added to the display as shown in the figure below. These regions can be especially useful for evaluating the peak finding process.

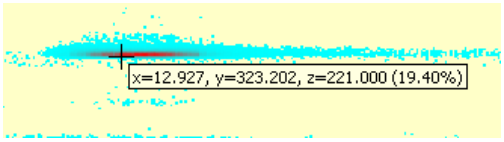


If you have imported more than one sample, the displayed peak regions refer to the **aligned** data, and not simply to the peaks found for the sample currently displayed. For example the circled region shown below for the peak with m/z near 324.2 appears to start before the actual peak itself: this is because this same peak eluted earlier in at least one of the other samples. Similarly, the peak with m/z near 323.1 and a retention time of approximately 12.4 minutes is present in at least one sample, but not in the current one.

Also note that the reported peak areas do not correspond directly to all of the intensity within the circled areas, even if only one sample was imported. The ellipses indicate the extent of the peaks in the retention time and m/z directions, but only those data points within that region which are above the specified *Noise Threshold* are actually included in the reported area.

5.16.2.2 Show Tooltips

This command displays a tooltip showing the retention time (*x*), mass/charge (*y*) and intensity (*z*) for the data point directly under the mouse as shown in the figure below.



5.16.2.3 Show Color Selection Tools

This command determines whether the tool palette used to adjust the colors and thresholds is displayed. Hiding these tools allows you to devote as much of the available space as possible to the data itself.

5.17 Peak Info Plot (for One Column)

These plots display the various metrics in a Peak Info Table (mass defect, average response, *etc.*) against peak index or mass/charge for one or more columns. Note that only columns to the right of the *Mass* column can be plotted.

You create these plots by selecting one or more columns in a Peak Info Table and selecting the *Plot Column* context menu item or the corresponding toolbar button as discussed in section 4.9.1.2

5.17.1 Context Menu

All of the items available in the context menu are also present in the context menu for Profile Plots – see the various sub-sections of section 5.2.2 for a detailed discussion.

5.18 Peak Info Plot (for Two Columns)

These plots display any two metrics in a Peak Info Table (mass defect, average response, *etc.*) as a function of one another. Note that only columns to the right of the *Mass* column can be plotted.

You create these plots by selecting exactly two columns in a Peak Info Table and selecting the *Plot Two Columns* toolbar button or menu item as discussed in section 4.9.1.3.

5.18.1 Context Menu

With one exception, all of the items available in the context menu are also present in the context menu for Profile Plots – see the various sub-sections of section 5.2.2 for a detailed discussion. The *Mass/Charge X-Axis* option of the *Display* sub-menu is discussed in section 5.3.1.1.

6 Appendix: Utility Applications

6.1 PC Variable Grouping

This section describes a utility for discovering relationships among MarkerView™ peaks. Peaks which are correlated with one another can be manually or automatically assigned to groups and the resulting groups can be manually inspected from within MarkerView.

For a detailed discussion of the technique used to group the variables see:

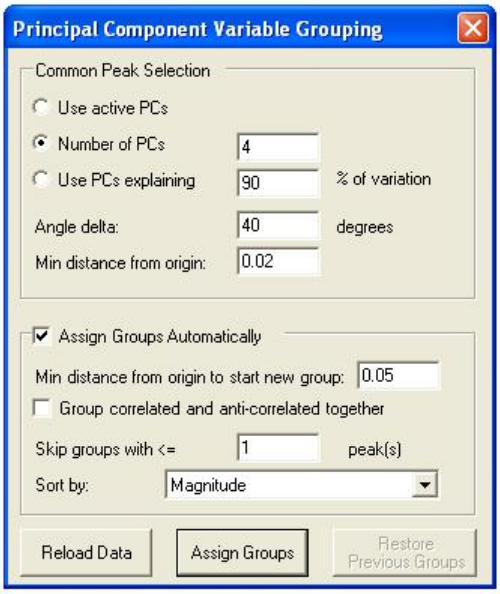
> *Dimensionality Reduction and Visualization in Principal Component Analysis*
> Anal. Chem., 2008, **80** (13), pp 4933-4944

Which is available for download as a pdf file from:

> http://pubs.acs.org/doi/abs/10.1021/ac800110w

The utility is started by selecting the 'PC Variable Grouping.exe' application from the *Help* menu. When the program starts it displays the form shown in the figure below. A few general comments:

- Before starting the utility, a PCA display must be active (most likely a Loadings Plot).
- You can leave the utility open, switching between it and the MarkerView™ software, however you must click the *Reload Data* button if you wish to process a different PCA data set from the one originally active when the utility was started.



The following items in the *Common Peak Selection* section apply when assigning groups manually or automatically.

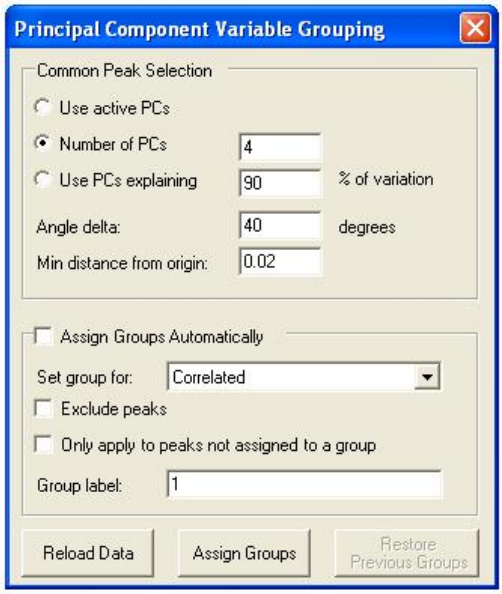
- *Use active PCs* – Select this option to use the two principal components which are currently displayed in the active PCA loadings plot. (If a loadings plot is not active, this item is unavailable.) Using the active PCs is intuitive since the groups will correspond directly to those which can be visually discerned from the loadings plot, however potentially important variation in other PCs will of course not be taken into account.
- *Number of PCs* – Select this option to use the specified number of principal components, starting with PC1.
- *Use PCs explaining* – Select this option to use as many principal components as are required to account for the specified percentage of the total variation, starting with PC1.

- *Angle delta* – Specify the size of the angle which defines the variables belonging to a given group. You may need to experiment with this setting since small values will split what is probably wanted as a single group into multiple groups.
- *Min distance from origin* – Use this setting to prevent variables with small loadings for the PCs from being assigned to a group. Including too many such variables makes review of the groups difficult.

Select *Assign Groups Automatically* to assign variables to a group.

- *Min distance from origin to start new group* – Use this setting to prevent variables with small loadings for the PCs from starting a new group. Variables closer to the origin than this value, but farther than the *Min distance for group* (from the *Common Peak Selection* section), can be assigned to an existing group, but will not themselves serve to define a new group.
- *Group correlated and anti-correlated together* – Check this item to combine both correlated and any inversely correlated peaks into the same group as one another. If this item is not checked, inversely correlated peaks define a separate group.
- *Skip groups with <= n peak(s)* – If this item is set to a value larger than zero, groups must have more than the specified number of peaks in order to be retained and reported. Groups without the required number of peaks are assigned to a common group called with the name 'Too few peaks'. If you use this option, you should review this special group to ensure that no important peaks were excluded.
- *Sort by: Magnitude* – The variable with largest size (for the PCs being used) is always assigned to group '1', as are any correlated variables satisfying the specified criteria. If this option is selected group '2' will contain the next largest peak (and the other variables which correlate with it) and so forth.
- *Sort by: Angle to first group* – If this option is selected group '1' is assigned as described in the previous point, group '2' as the next group closest to group '1' (as measured by their angle in PC space) and so forth. This can be useful when reviewing the groups since adjacent groups will be most similar to one another.

If you do not check *Assign Groups Automatically*, the dialog reconfigures as shown in the figure below. This option is used to assign a group for the variables correlated with those currently selected in the active loadings plot. If desired, once you have assigned a group, you can select a different set of points in the loadings plot and assign another group. If the loading plot is not active or if it contains no selected variables, this option is not available.



- *Set group for* – Most often you will probably use the *Correlated* option to assign a group to those variables which behave similarly to the selected variable(s). The *Anti-correlated* option assigns a

group to those variables with inverse behaviour and the *Both* option assigns a group for both sets of variables.

- *Exclude peaks* – Enable this option to exclude all variables which will be assigned to the group.
- *Only apply to peaks not assigned to a group* – Enable this option to prevent peaks which have been previously assigned to a group from being re-assigned; otherwise all peaks satisfying the criteria will be assigned.
- *Group label* – Enter the label which you wish to use for the group.

The form contains the remaining items:

- *Reload Data* – If you wish to process a different PCA data set from the one originally active when the script was started, you must click this button so that the utility can load the new data.
- *Assign groups* – Click this button to assign groups to the variables using the specified options.
- *Restore Previous Groups* – If you have previously assigned groups to some or all of the variables, click this button to undo the last operation. Note that it is only possible to undo a single operation. Also note that any changes to the groups made manually within the MarkerView™ application after groups were assigned using the script will be lost.

6.1.1 Using Peak Groups

This section contains a few general comments regarding the 'Peak Group' feature available within the software. For more details see the discussion elsewhere in this document.

- Peaks Tables, Loadings Tables and t-test Tables all contain a column titled *Group*. You can manually type the name of a group for selected variables (or paste into the cells) in addition to using the utility described in this document.
- Plot symbols are used in loadings plots for peaks or variables which have been assigned to a group in the same way as for samples in scores plots. The *Plot Symbols* tab of the *Options* dialog is used to associate plotting symbols with a particular group for *both* samples and peaks.
- The utility assigns groups using integer names starting at 1 – in addition to the special 'Too few peaks' group mentioned above. For this reason it is useful to assign plot symbols for a reasonable number of integers starting at 1. This can either be done manually, as mentioned in the previous point, or by using the *Import* button in the *Plot Symbols* tab of the *Options* dialog and selecting the provided file called 'Default PCVG Plot Symbols.ptsym' (installed to the 'MarkerView\Bin' directory). This file defines 42 symbols in addition to the 'Too few peaks' symbol. Note that if more than 42 groups are assigned the plot symbols will cycle; for example the symbol for group '43' will use the same symbol as for group '1'.
- If a group has been assigned for at least one peak, Loadings Plots will display a legend showing the groups with the corresponding plotting symbols. (If the legend does not appear, make sure that the *Show Legend* item is selected from the context menu's *Display* sub-menu).

  If you ouble-click the color spot to the left of a particular item in the legend, all variables for that group are selected in the Loadings Plot (you can also double-click on the text). Since the plot is linked to its Loadings Table and to its last Profile Plot (assuming that one has been displayed), the table and Profile Plot will also update. So, you can display a Profile Plot containing all overlaid variables for a particular group by (1) generating a profile plot for *any* variable and then (2) clicking the appropriate item in the legend. Also, since the plot is linked to the Loadings table, you can (1) click a particular group in the legend and then (2) use the table's context menu to exclude the group, *etc.*

  Note that if the legend contains more items than can be displayed, you can scroll by clicking within the legend and dragging.

6.2 MV Metabolite Namer

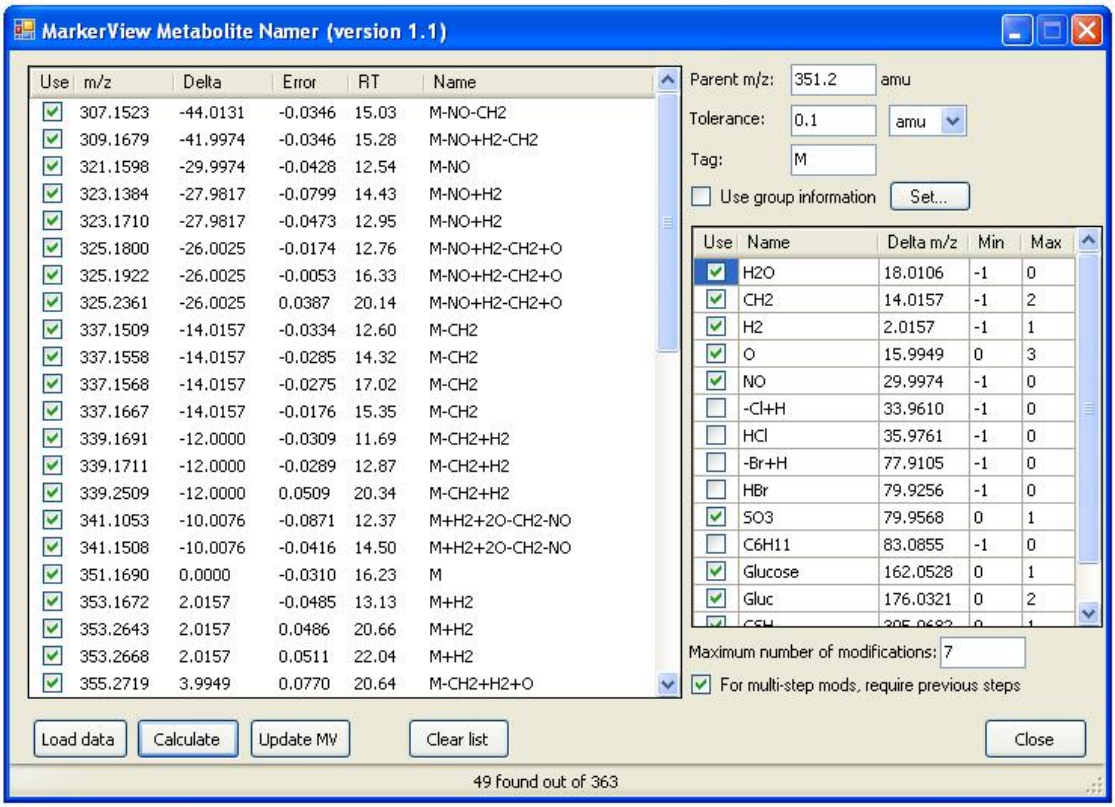
This section describes the use of the 'MV Metabolite Namer' application. This is a MarkerView™ software plug-in utility which assigns specific names to peaks based on the mass of a parent drug and expected metabolic transformations. These potential identifications are based on the masses of the peaks.

When the utility is first started the window shown below is presented (although note that results are not shown in the output table on the left until you click the *Calculate* button). The various items in the window are explained in detail below, however a typical workflow is the following:

- Open the desired data set in the MarkerView™ software and ensure that it is the currently active document.
- Start the utility by selecting it from the *Help* menu.
- Enter the *Parent m/z* (not molecular weight).
- Specify a suitable mass tolerance and other parameters.
- Click the *Calculate* button.
- Click the *Update MV* button if you wish to open a copy of the original data set with names assigned for the matched peaks.

When you click the *Calculate* button, the program checks each peak in the data set for a match (within the specified mass tolerance) to combinations of gains or losses of the specified modifications. All combinations of modifications are considered, but constrained by the min and max ranges for each modification and the maximum total number of simultaneous modifications. A few additional constraints can also be set as discussed below.

If a given peak matches a particular combination, it is listed in the output table along with a name showing the modification. If more than one possible combination of modifications matches a particular peak, the combination using the fewest modifications is reported; if multiple combinations still match, the one with the smallest mass delta is reported.



The window contains the following items:

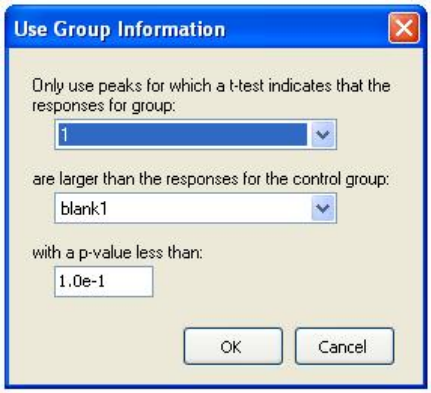
- *Output table* – This table is populated with matches when you click the *Calculate* button. For each match the table includes (in order): a *Use* field which allows you to use the proposed modification or not, the mass/charge of the matching peak, the theoretical mass shift due to the corresponding metabolic transformation, the mass difference between the peak and the expected mass, the retention time of the peak, and a name for the transformation. The name is constructed by appending the name of each individual transformation (along with a '+' to indicate gain or a loss to indicate loss and its count) to the specified *Tag*.
- *Parent m/z* – Specify the mass to charge of the parent drug – note that this is **not** the molecular weight. In certain cases you may wish to specify the m/z of a fragment of the parent – see the discussion of the *Tag* item below for details.
- *Tag* – Specify the text which is used as the root for the names of the proposed potential metabolites. Each name is constructed by appending the name of each individual transformation (with its count) to the specified tag. If you are dealing with a drug which fragments easily, you may wish to do this:
    a. Specify the m/z of the actual parent and enter a suitable tag (*e.g.* 'Par')
    b. Click the Calculate button.
    c. Specify the m/z of the fragment and enter a suitable tag (*e.g.* 'Frag').
    d. Click the Calculate button.

  In this case the matches for the fragment will be appended to the previously calculated list for the parent (also see the *Calculate* item below).
- *Tolerance* – Specify the mass tolerance to use when attempting to match the mass of a peak with a possible transformation.
- *Use group information* – Check this item so that matching only considers peaks with responses statistically higher in a dosed group as opposed to a control group. See the discussion of the *Use Group Information* dialog below for details. Note that if this item is grayed-out, you need to click the *Set* button to specify the groups.
- *Set* – Click this button to set the groups used when the *Use group information* checkbox is selected. See the discussion of the *Use Group Information* dialog below for details. If this button is grayed-out, there is either no currently loaded data or the data does not specify suitable groups.
- *Transformations table* – This table contains an entry for each modification which you wish to consider. When looking for transformations, the utility will consider *all* possible combinations of these modifications – up to the specified *Maximum number of modifications*. For each modification you specify:
    a. *Use* – If checked the modifcation is considered, otherwise not. This allows you to include less common modifications in the list and to easily decide to use them or not.
    b. *Name* – the name of the modification; this is used in the construction of the name of matching modifications. If you wish to delete a particular transformation you can set the name empty (you do not need to remove the information for the other fields).
    c. *Delta m/z* – The mass shift associated with the modification.
    d. *Min* – The minimum number of times which this modification can be applied. Specify a negative value to allow for losses and a value of zero if only gains should be considered. It is also possible to specify a positive value, but this is only useful if you wish to force this modification to *always* be present.
    e. *Max* – The maximum number of times which this modification can be applied. Specify a positive value to allow for gains and a value of zero if only losses should be considered.

  The transformations list is automatically saved for use the next time you run the program, but only when you click the *Calculate* button.
- *Maximum number of modifications* – Specify the maximum number of modifications which will be considered; this option is used to prevent an excessive number of unlikely transformations from being used. Note that this is not the number of *different* transformations, so M+2O counts as two modifications, not one.

- *For multi-step mods, require previous steps* – This option is used to prevent transformations from being considered unless a match was found for all transformations from previous steps. For example when this item is checked, M+CH2+O will not be considered unless a match is found for *both* M+CH2 and M+O.
- *Load data* – Click this button to load the currently active MarkerView™ data set. Note that the utility will automatically load the active data when it starts, so you only need to click this button if MarkerView contained no open documents when the utility was started or if you open new data in MarkerView without re-launching the utility.
- *Calculate* – Click this button to search for matches subject to the various constraints (mass tolerance, *etc.*). If you have previously clicked the *Calculate* button (without subsequently clicking the *Clear* button), the previous matches will be cleared *unless* you have changed the *Tag* text. You may wish to change the tag to list matches for both the parent drug and a fragment as discussed above in the *Tag* section.
- *Update MV* – Click this button to open a new copy of the data set. The resulting peak names will be updated for any matches. Once the data is open you can perform the usual operations (PCA, *etc.*).
- *Clear list* – Click this button to clear the output table.
- *Close* – Click this button to quit the utility.
- *Status area* – When you click the calculate button, a message will appear stating how many peaks were matched and the total number of peaks in the data set. Error messages may also be reported here.

When you click the *Use group information: Set* button in the main window, the dialog shown in the figure below is presented. The options set using this dialog are used in conjunction with the *Use group information* checkbox so that matching only considers peaks with responses statistically higher in a dosed group as opposed to a control group. The intention is to prevent endogenous background peaks from being spuriously reported as potential metabolites.



When the option is enabled, a t-test is performed for each peak using the samples for the two groups specified in the dialog. If the resulting p-value that the average response for the first group is larger than that for the second group is larger than the specified value, the peak is unlikely to be a metabolite and it is never assigned as such.

If you enter a relatively high p-value (say 0.5), it is quite likely that some of the reported matches will still correspond to background material. If you specify a low value (say 0.01) this is still possible, although much less likely.

In order to use this option, both groups should ideally contain multiple samples and it is required that at least one of the groups contains two or more samples. If there is only one group or none of the groups has more than one sample, the *Use group information: Set* button is grayed-out.

7 PCA Tutorial

This section attempts to provide a non-mathematical introduction to principal components analysis or PCA.

PCA is a data analysis technique which transforms an original set of variables into a new set, generally with many fewer significant variables. In fact the main purpose is to reduce the number of variables since it is much easier to understand and to examine a relatively small number of variables as compared to the potentially thousands which may result from an LC/MS analysis – with the ultimate goal of finding variables which distinguish groups of samples from one another. This reduction in dimensionality is possible because many of the original variables will be redundant in the sense that they are correlated with one another or else be of negligible intensity.

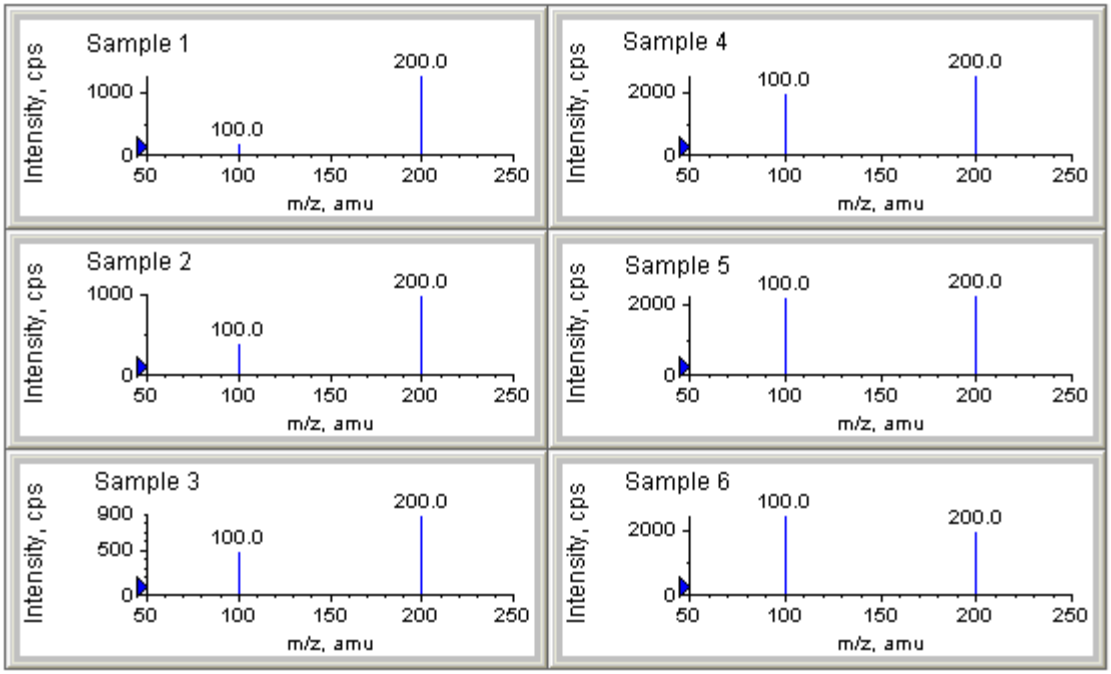
As a very simple example, consider a data set consisting of the peak areas for three isotopes each for two different compounds for a number of samples. The total number of peaks or variables is six, however the isotope peaks for each of the respective compounds will be exactly correlated (within experimental error), so two variables are really sufficient to describe the data. Assuming that the areas of the two compounds themselves are uncorrelated, PCA will find a first new variable which is a linear sum of the isotopes for the first compound and a second new variable for the isotopes of the second compound.

As discussed further below, one important property of PCA is that the first new variable (or the first principal component or PC1) accounts for the largest difference between the samples; the second principal component accounts for the second largest difference (or, if you prefer, the largest remaining difference) and so forth. So each principal component accounts for less of the differences, and is less important in this sense, than the previous PC. In practice, even with a large number of variables, only the first few principal components are likely to be of interest.

In the isotope example above there are technically six principal components, however only the first two will account for the significant 'real' differences between the samples. The remaining PCs will account for differences between the samples due to noise – since the ratio of the isotopes will not be exactly identical between samples.

7.1 Geometric Interpretation of PCA

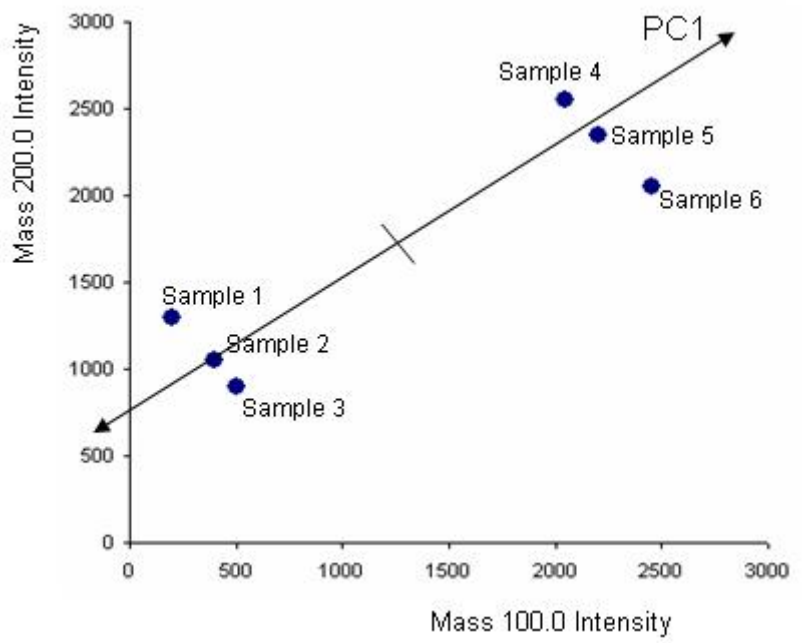
The figure below shows hypothetical spectra for six samples. Based on the absolute intensity and on the intensity ratios of the two ions, the samples on the left are clearly similar to one another and different from those on the right.



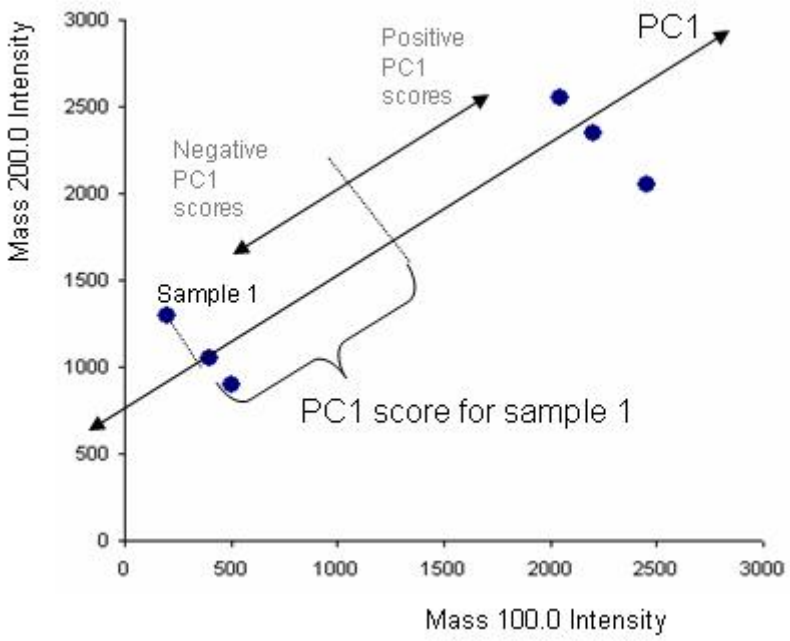
The table below shows the same information in tabular format. The first row represents the peak height for the peak with m/z of 100.0 Da and the second row for the peak with m/z of 200.0 Da. This is essentially the same information which is shown in MarkerView™ Peaks Tables.

Peak	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Mass 100.0	200	400	500	2000	2250	2500
Mass 200.0	1300	1000	900	2600	2300	2000

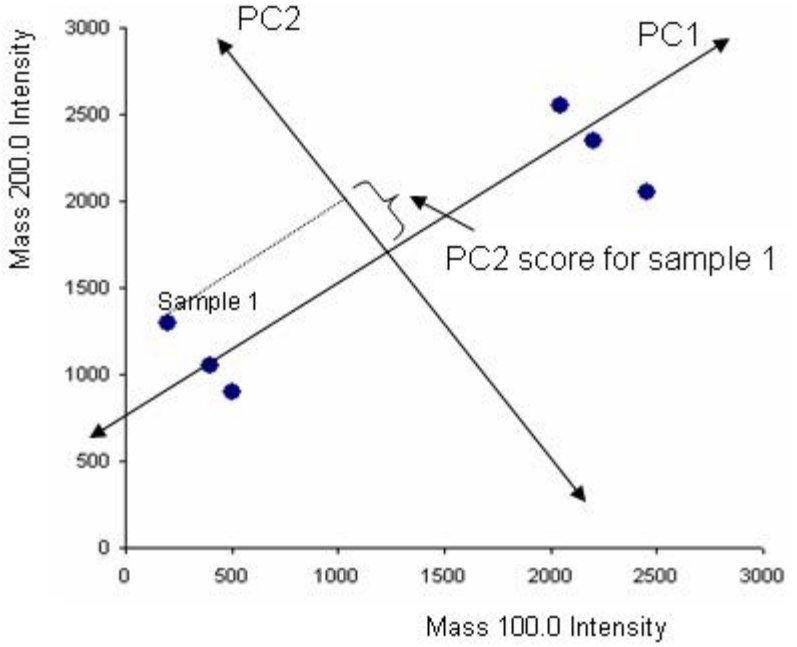
The figure below plots the intensities for the two peaks. For example 'Sample 1' has an intensity of 200 cps for the peak with m/z of 100.0 Da and 1300 cps for the m/z 200.0 peak. The first principal component is the straight line which explains the most variation in the data; in this two-dimensional case this is similar to a standard linear regression.



The *score* for each sample is its distance along PC1 as shown in the next figure. When mean centering is used (discussed in the next section), the score is relative to the average for all samples (the midpoint of the PC1 line) and hence is negative for samples to the left of the average position and positive for samples to the right. For example 'Sample 1' has a PC1 score of roughly -1300.

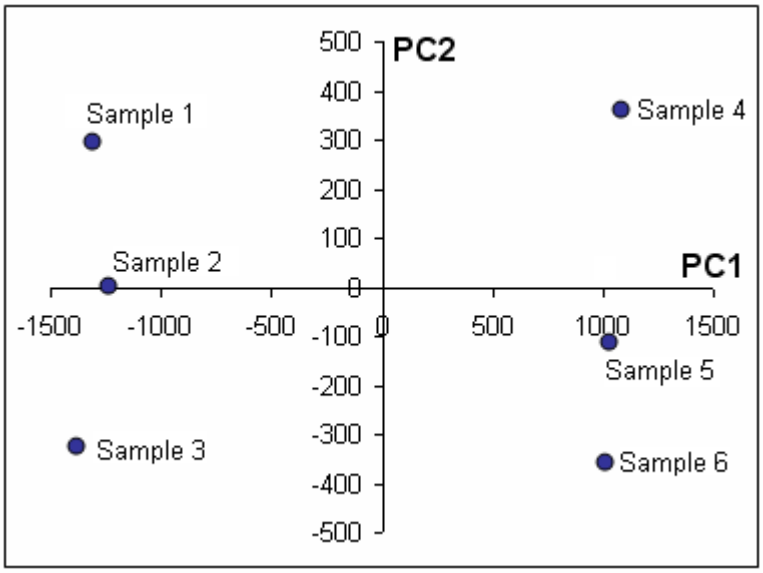


The second principal component is the straight line which explains the most remaining variation; it will always be perpendicular to PC1. The figure below shows PC2; Sample 1 has a score for this PC of approximately 300.



The figure below shows the two-dimensional scores plots. The (PC1, PC2) scores for each of the six samples are plotted. In this simple example the PC1 score separates the samples into the two expected groups and explains most of the variation – note that the PC1 scores are larger than the PC2 scores. In this case PC2 can be interpreted as representing noise.

The more closely together particular samples cluster in the scores plot, the more similar they are. The ideal situation is one in which the samples for each class (diseased versus healthy, *etc.*) are grouped close to one another, but far away from the samples for the other classes.



The principal component vectors shown in the previous figures can be expressed in terms of the original variables. For this example:

PC1 vector $\propto$ 1.0 * (Mass 100.0 vector) + 1.0 * (Mass 200.0 vector)

PC2 vector $\propto$ -1.0 * (Mass 100.0 vector) + 1.0 * (Mass 200.0 vector)

These vectors can be scaled so that their total length is 1.0:

PC1 vector = 0.707 * (Mass 100.0 vector) + 0.707 * (Mass 200.0 vector)

PC2 vector = -0.707 * (Mass 100.0 vector) + 0.707 * (Mass 200.0 vector)

The *loading* for a particular variable for a given principal component is the contribution it makes to these scaled vectors. For example the mass 100.0 component has a loading of 0.707 against PC1 and -0.707 against PC2. The loading indicates how important is the contribution made by a variable to one of the principal components. The loadings plots shown in the *Simple Data Set* section below illustrate the practical importance of this.

The scores are in fact calculated from the loadings. For example the PC2 score for a sample is given as:

PC2 score = (Mass 100.0 loading for PC2) * (Mass 100.0 intensity for sample) +
(Mass 200.0 loading for PC2) * (Mass 200.0 intensity for sample)

or:

PC2 score = -0.707 * (Mass 100.0 intensity) + 0.707 * (Mass 200.0 intensity)

Note that the score for a particular sample can be high either because it has a high intensity for a variable with a positive PC loading (the m/z 200.0 peak for PC2 in this case), or because it has a low intensity for a variable with a negative PC loading.

7.2 Scaling

Most often the raw data is scaled before PCA is run. This is especially important in cases where the variables being compared have different response units, perhaps in a social sciences study. For mass spectrometric data all of the responses will correspond to peak intensities are areas and hence have the same units, so scaling is not absolutely essential, but is still usually useful.

If the original data is not scaled in any way the variables with the largest response will tend to dominate the PCA analysis. For example a very large variable which is approximately constant for all samples may dominate PC1. This is rarely desired so, at the least, data is almost always mean centered.

For each response for each variable and sample *mean centering* subtracts the average response for that variable for all samples. For example the responses for the 100.0 m/z variable discussed in the previous section for all

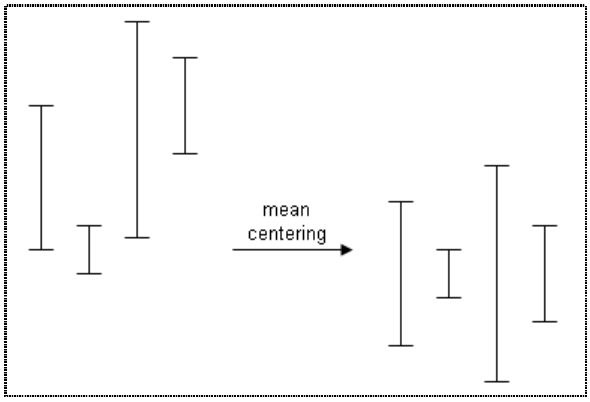
samples are:

{200, 400, 500, 2000, 2250, 2500}.

The average response for this variable is 1308 so the mean centered responses are:

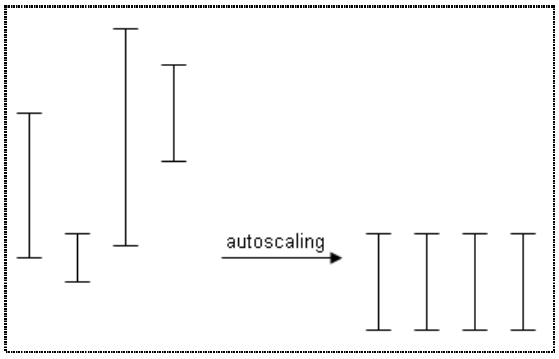
{-1108, -908, -808, 692, 942, 1192}

This is illustrated pictorially in the figure below where each vertical lines represents the range of a different variable (four in this case). After scaling each variable retains its original range, but its average value is centered at zero.

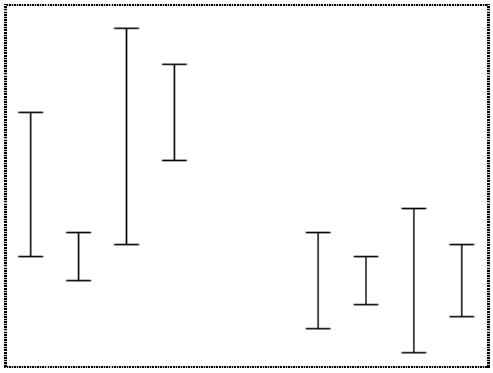


With mean centering variables which span a large range (*i.e.* those with the largest variance) will be weighted the most heavily by PCA. Often this is a perfectly reasonable way to scale the data, however consider the case with the following two variables: one with an average response of 100.0 and one with an average response of 1.0, both with a 10% variation. After mean centering, the absolute variations are 10.0 and 0.1. So the larger variable can still dominate the analysis, even though the relative variations are the same.

With *autoscaling*, illustrated pictorially below, the data is first mean centered and then divided by the standard deviation for the variable. The net effect is that each scaled variable has a mean of zero and a variance of one, hence this technique is also called scaling to unit variance or UV. This is probably the most common way of scaling data since each variable is treated equally in the sense that larger relative changes in response will receive the most weighting.



Pareto scaling is a technique intermediate between mean centering and autoscaling. With this form of scaling the data is first mean centered and then divided by the *square root* of the standard deviation for the variable. The net effect is that larger variables receive more importance than with autoscaling, but less than with mean centering alone. This is illustrated in the figure below where the resulting variances are equal to the respective square roots of the original variances. Although one needs to experiment in individual cases, we have found that pareto scaling is very often a good chose for processing mass spectrometry data, presumably since larger peaks will generally have better signal/noise than smaller ones and allowing them to have some extra weight, as compared to autoscaling, is helpful.



7.3 Simple Data Set

You may find it helpful to explore a simple artificial data set to gain an intuitive understanding of how to interpret PCA results. The figure below shows a data set which was created with Excel in the format required for directly importing into the MarkerView™ software. A few points to note about this data set:

- There are ten samples, five in a group arbitrarily called 'A' and five in a group called 'C'. These group names were chosen to correspond to those in section 3 of the User Manual. If you follow the instructions in section 3.5 the point symbol used for the samples will correspond to that shown in the figures further below.
- As discussed below, there are different types of named variable: One is constant from sample to sample, some contain random noise at different levels, some are 'real' in the sense that they are at different levels between the two groups.

Simple Data.xls

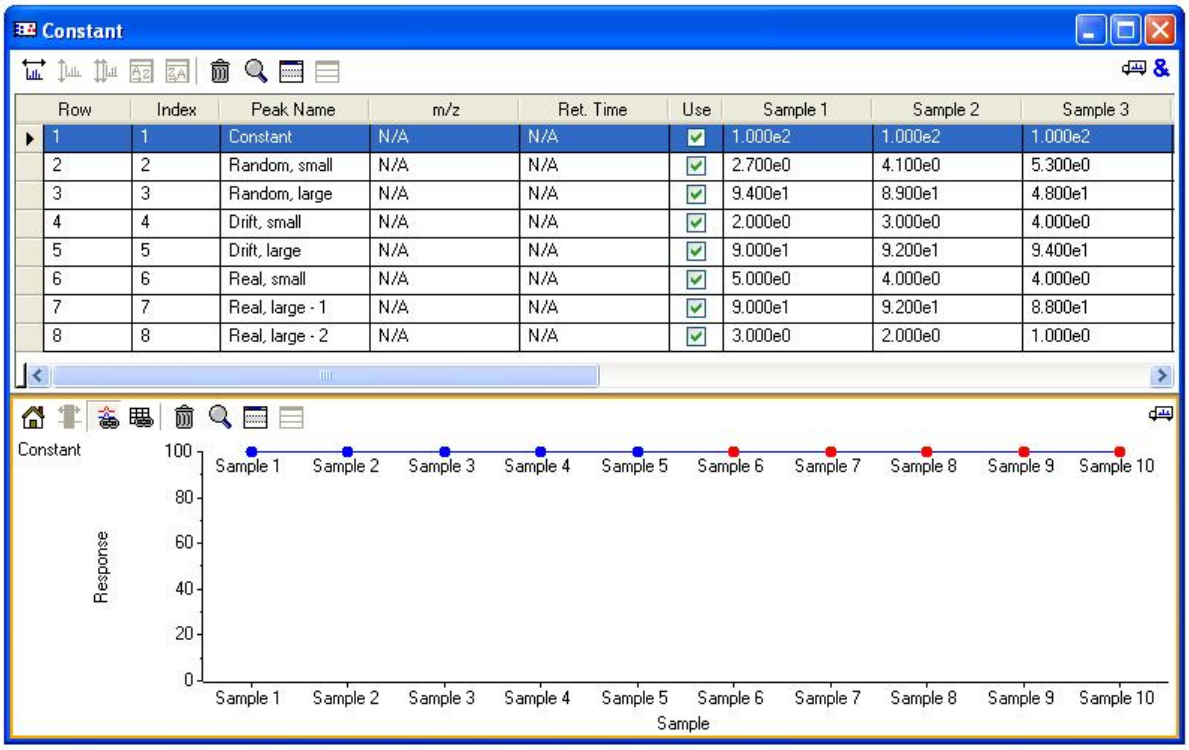
	A	B	C	D	E	F	G	H	I	J	K
1	*Sample Name*	**Sample 1**	**Sample 2**	**Sample 3**	**Sample 4**	**Sample 5**	**Sample 6**	**Sample 7**	**Sample 8**	**Sample 9**	**Sample 10**
2	*Group*	*A*	*A*	*A*	*A*	*A*	*C*	*C*	*C*	*C*	*C*
3	**Constant**	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
4	**Random, small**	2.70	4.10	5.30	3.10	4.30	0.90	8.10	7.50	4.20	6.80
5	**Random, large**	94.00	89.00	48.00	94.00	77.00	21.00	24.00	88.00	66.00	55.00
6	**Drift, small**	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00
7	**Drift, large**	90.00	92.00	94.00	96.00	98.00	100.00	102.00	104.00	106.00	108.00
8	**Real, small**	5.00	4.00	4.00	5.00	6.00	9.00	11.00	12.00	10.00	9.00
9	**Real, large - 1**	90.00	92.00	88.00	90.00	91.00	105.00	107.00	104.00	106.00	107.00
10	**Real, large - 2**	3.00	2.00	1.00	3.00	2.00	105.00	107.00	104.00	106.00	107.00
11											

A text version of the above file is included with this document. To import this data into the program:

- From the *Import* sub-menu of the main *File* menu select the *Generic Text File* item.
- Browse for the example file called 'Simple Data.txt' and select it. This file is installed to the 'MarkerView\Sample Data' directory.

Start by exploring the nature of the variables themselves. For example:

1. Select a row in the peaks table by clicking in the row header to the left of the row number and click the *Plot Profile for Peak* icon ( ). The resulting Profile Plot for the constant variable is shown in the figure below.

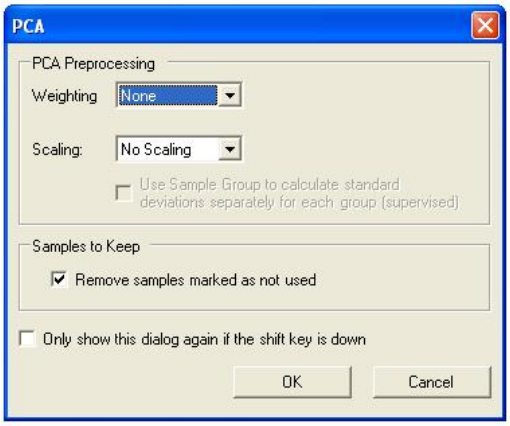


2. Click the **Link to Table** button () in the Profile Plot pane; this causes the plot to automatically update as new variables are selected in the Peaks table.
3. Select the remaining rows in the Peaks Table one at a time and examine the Profile Plot. Note that when the table is active (i.e. with the orange border) you can use the up and down arrow keys to change rows quickly.

   The 'drift' variables are intended to simulate a continuous change in the response of a background peak as discussed in the User Manual. The 'real' variables all show a clear difference in response between the two groups.

Now explore the results of running PCA with different scaling techniques. A few examples are given below.

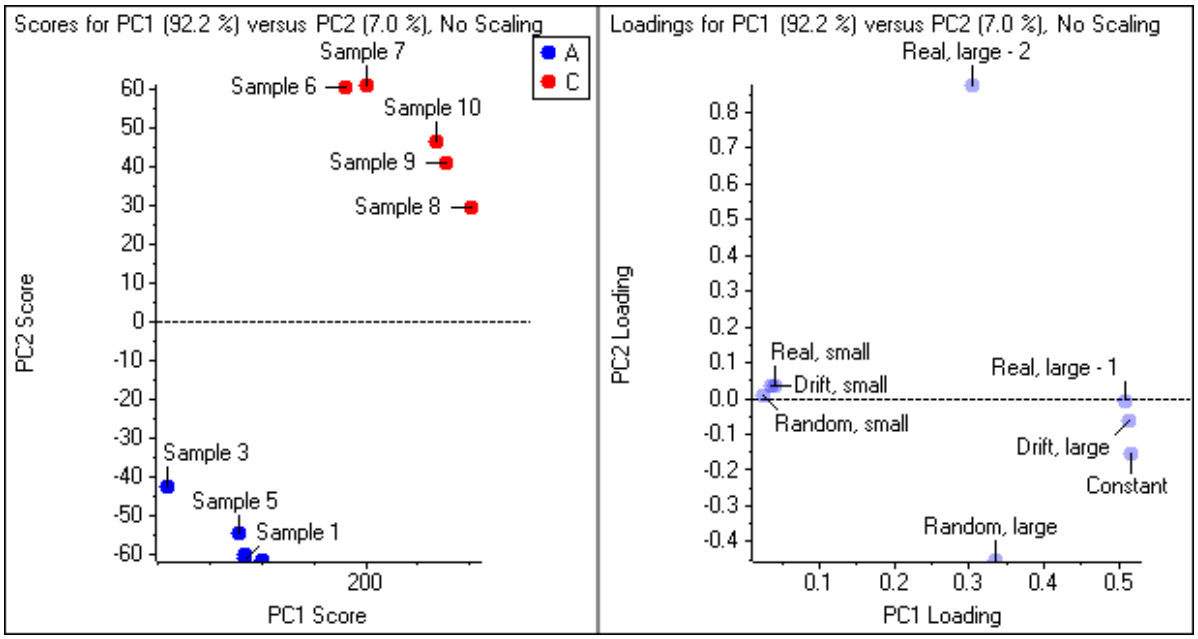
1. Select the *Perform PCA* menu item from the *Analyze* menu and choose no weighting and no scaling in the resulting dialog.



2. The figure below shows the resulting scores and loadings plots. The two groups of samples are clearly well separated from one another in both the PC1 and PC2 directions.

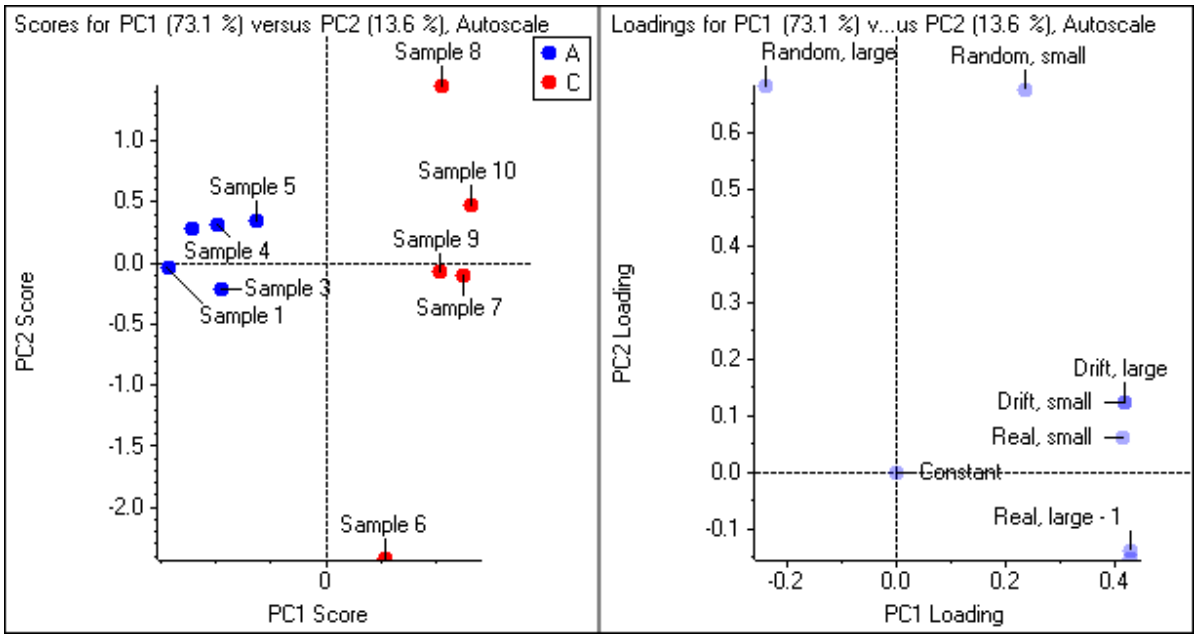
Note that the variables with large PC1 loading are the ones with largest intensity – in fact the 'Constant' variable which provides no separation between the groups is in the same region of the plot as the 'Real, large – 1' variable, which does provide separation, since they both have peak responses of about 100. For a real data set with many peaks, it would be difficult to distinguish the useful large variables from the others.

Also note that all of the variables with low intensity cluster together near the origin. For real data with many peaks, it would probably be impossible to notice the 'Real, small' variable from the other small variables which cluster nearby.



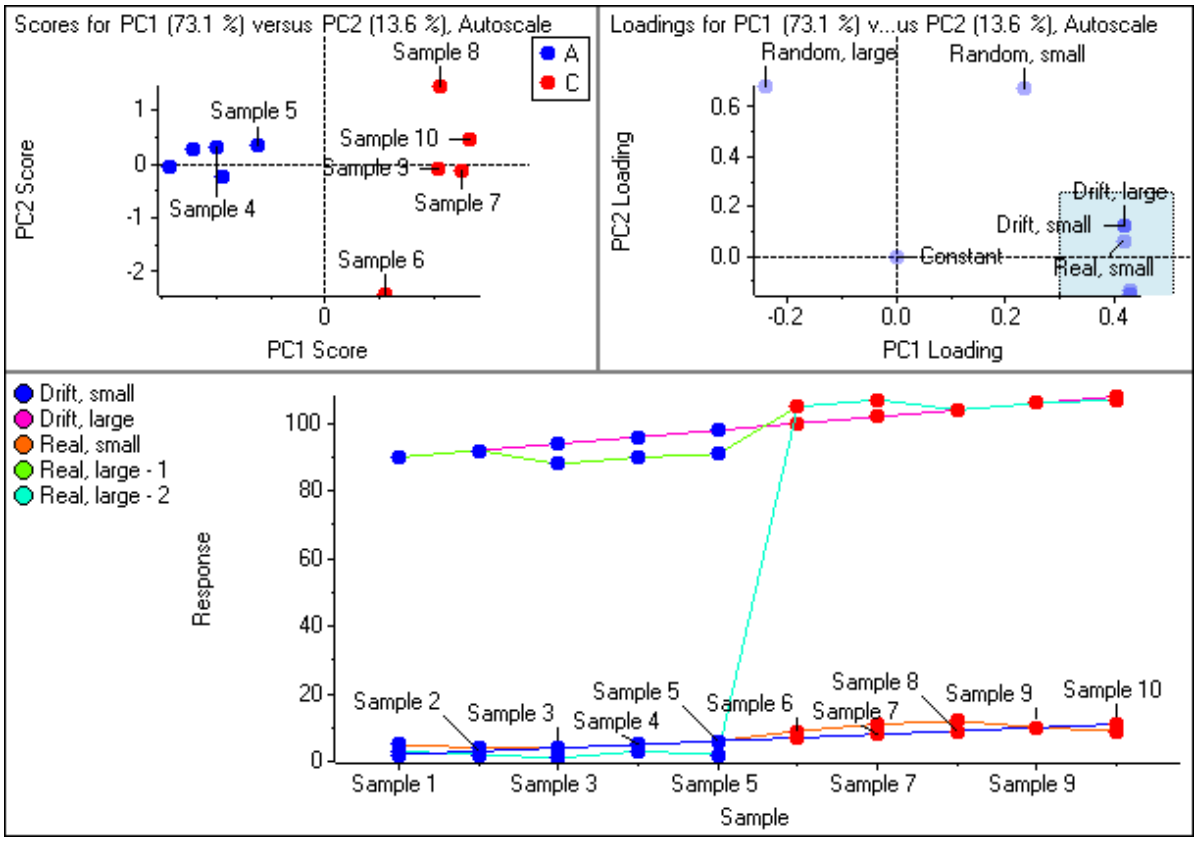
3. Repeat the PCA analysis with the *Autoscale* option. As shown in the figure below PC1 separates the samples. In this case the 'Constant' variable has no effect and is positioned at the origin, PC1 has large loadings for the variables which are different between the two groups (regardless of intensity) and PC2 has large loadings for the noise variables.

   Although the plot may make it appear that the separation in the PC1 direction is comparable in magnitude to the scatter in the PC2 direction, note that PC1 actually accounts for more than five times as much of the total variance (73.1% / 13.6%) as PC1.



4. Make a selection which covers the variables in the lower-right portion of the loadings plot and select *Plot Profiles for Selected Peaks* from the right-click context menu. The result is shown in the figure below.

   All of these variables have a lower response for the 'A' group of samples than for the 'C' group, including the 'Drift' variables which show a continuous change as opposed to the more step-like response of the other variables. This kind of change in background variables is relatively common with MS data. For real data, acquiring the samples in random order will prevent these drift variables from correlating with those representing real differences between groups. These variables will therefore hopefully cluster away from the variables of most interest in the loadings plot.



5. Although the 'Random, small' variable has a large PC2 loading and hence is well separated from the variables in the lower-right corner of the loadings plot, it does have a reasonable PC1 score of more than 0.2. Click directly on this point to cause the Profile Plot to update to display the profile for this variable.

   Although it is a random variable, this variable happens to have a larger average response for the group 'C' samples and hence contributes to the separation of the groups. With a sufficiently large number of variables such as is usually obtained with MS data, such random variables which fortuitously help to separate the groups will occur. Even if they are very small in intensity, the autoscaling approach will tend to cluster these variables near the 'real' ones, potentially making those variables more difficult to discern. (As an aside, the software allows the raw MS spectra and/or XICs to be examined to help you decide on a variable's likely significance based on a visual examination of the signal to noise.)

6. Examine some of the other principal components. For example click on the column heading for the PC2 column in the Scores Table and drag to also select the PC3 column. Both plots should update to show the scores and loadings for these principal components.

   The samples no longer separate in the scores plot based on the 'A' and 'C' groups. The loadings plot shows that the only variables with significant PC2 or PC3 scores are the 'random' ones. This is typical of PCA analyses: The first few principal components (although just one in this simple example) will represent valid physical differences between the samples and the remaining components will represent noise and variations in individual samples.

Each principal component explains less of the total differences between the samples than the previous one. The percentage of the total differences explain by each PC is shown in the column heading of the Scores and Loadings Tables (and also in the titles of the plots). In this example these percentages are:

PC1	73.1 %
PC2	13.6 %
PC3	8.7 %
PC4	3.7 %
PC5	0.8 %

For this example there are actually eight theoretical principal components (the minimum of the number of variables and number of samples), however the software does not show PCs which explain less than 0.5% of the total variation as being almost certainly of no significance.

7. Repeat the PCA again choosing the *Pareto* scaling option. The variables responsible for separating the two groups again have large PC1 loadings. As compared to autoscaling, the main differences is that the most intense variables have the largest PC1 loadings (whether 'Drift' or 'Real') while the less intense variables appear nearer to the origin – although still well separated from it.

   Although each data set is different, pareto scaling often seems to do the best job of allowing the most interesting variables (*i.e.* those which are both intense and separate the groups) to be most quickly found.

8. Repeat the PCA using the *Mean Center* scaling option. The variable with the largest absolute difference from its average ('Real, large - 2') dominates PC1 and the next largest ('Random, large') dominates PC2.

9. Repeat the PCA, experimenting with the *Logarithm* weighting option for the various scaling options. For example when pareto scaling is used, 'Sample 6' is a clear outlier with logarithmic weighting due to the fact that it has a low response for both of the variables important for PC2 ('Random, small' and 'Random, large').

   Although the different weightings can reveal different features of the data, their use is generally not required.

8 Appendix: New Features

8.1 New Features From Version 1.0 to 1.1

For the benefit of those familiar with version 1.0 of the program, the more important changes to the program for version 1.1, along with links to the relevant sections of this document, are listed below. Note that relatively minor changes are not included.

- The Use *area integrated from raw data, not from original peak finding* option of the *Alignment & Filtering* dialog (discussed in section 2.1.2.1) can lead to more accurate peak areas for low-level components. This is especially true for quadrupole data as compared to time-of-flight data.
- The *Baseline Subtraction* option (section 2.1.2.3) is useful when importing linear-mode MALDI spectra.
- The *4x00 LC-MALDI Peak Lists* import command discussed in section 2.1.2.5 is used to create a Peaks Table from a series of 4700 or 4800 LC-MALDI runs. This is done by importing peak lists previously exported using the 4x00's 'Peaks Explorer' program.
- The *MRM Chromatograms* import command discussed in section 2.1.2.6 allows MRM data to be directly imported into the program – a chromatographic peak finder is used to find peaks in each MRM chromatogram.
- The *Process peak lists* option of the *Process Spectra Options* dialog of section 2.1.2.8 is used to import spectral peak lists (from text files) as opposed to profile spectra.
- If you have processed the MS data for a series of IDA runs, you can display product spectra corresponding to specific MS peaks of interest; these spectra are automatically zoomed to exclude 'empty' regions at the start or end of the acquired mass range. The main discussion for this feature is in section 5.15; the way to report these spectra is in section 2.1.12.
- The various plot symbol information (group name, shape, size and color) can be exported and subsequently re-imported (section 2.2.8.1). This is useful if you wish to use plot symbols defined on one computer on a second computer.
- You can use custom colors for plotting symbols as discussed in section 2.2.8.1.
- The PCA-DA option when performing PCA (section 2.4.1) can be very useful for improving sample separation when there are known groups or classes.
- The *Normalize Using Median Peak Ratios* command (section 2.4.3.5) is useful for normalizing sample response when internal standards are not available.
- The *Normalize Using Manual Scale Factors* command (section 2.4.3.6) is used when you have specific normalization factors, most likely obtained from a non mass-spec-based analysis.
- The *Ratio Responses to Selected Sample* command (section 2.4.4) allows you to directly examine 'fold changes' with respect to a reference sample.
- The status bar described in section 3.3 shows information for the active pane's data set.
- You can save the mass/charge and retention times for all peaks in the Interest List in a format suitable for directly importing as an Analyst® IDA exclusion list. The intention is to allow product spectra to be acquired for these peaks during a second run (see section 4.8.1.7).
- You can now optionally disable peak labeling for Scores Plots (section 5.7.1.2.2).
- Both the {x, y} data points and their corresponding peak lists can be exported to tab-delimited text files for spectra and XICs as described in section 5.12.2.2. For the new IDA product spectral plots, you can also export in Mascot '.mgf' format.

8.2 New Features From Version 1.1 to 1.2.1

For the benefit of those familiar with version 1.1 of the program, the more important changes to the program for version 1.2 are listed below.

- For MS and LC/MS data, charge state and isotope status are automatically assigned for each peak. This is described in section 4.9. The default plotting symbols for peaks are based on the isotope status as described in section 2.2.8.1. When opening data processed with version 1.0 or 1.1 of the program, you can assign this information using the command discussed in section 2.4.10.
- The t-test Table has been considerably reworked to display additional metrics and plots. The command to generate the table is discussed in section 2.4.2, the table itself in section 4.4, and the related plots in sections 5.5 and 5.6.
- For LC/MS data, 'Peak List Spectra' are available. These are 'virtual' mass spectra constructed using the list of detected peaks and a retention time window, as opposed to the raw profile data (section 5.13).
- A new Peak Info Table is available which contains general information for each peak (section 4.9).
- Mass spectra and XICs originating from Analyst® wiff files can be sent directly to the Analyst® Software; this is useful so that the data can be explored further using the various tools available there (section 5.12.2.2 for spectra and 5.14.1.1 for XICs).
- It is now possible to define specific plotting symbols for variables as opposed to just samples (section 2.2.8.1).
- To potentially save having to define many custom plot symbols for different group names, you can define plot symbols once for groups named '1', '2', '3', *etc.* to use these symbols when plotting if a specific symbol is not otherwise defined (section 2.2.8.1).
- The 'Enhance' LC/MS peak-finding algorithm has been modified to calculate centroid masses in the same way as the Analyst® spectral peak finding algorithm. The net result is that spectra from wiff files will have more accurate mass/charge values reported. In addition, the algorithm was modified to allow it to more reliably find closely eluting isomers as separate peaks. Note that for these reasons the results obtained with this version of the application will not exactly match those obtained with previous versions.
- If the MS data has been processed for a series of peptide IDA runs, a Mascot mgf file can be created containing a peak list for each of the (available) MS/MS spectra for each peak added to the interest list (section 4.8.1.8).
- A new normalization option, *Normalize Using Total Area Sums*, has been added (section 2.4.3.4).
- You can use the Shift and Control keys to select multiple files or samples in the Select Samples dialog (section 2.1.1).
- For greater ease of use, the command to sort Profile Plots has moved from the context menu to the toolbar and has some additional functionality (section 5.2.1) and the command to generate Profile Plots for selected peaks is now available on the toolbar as well as the context menu.

The following functionality was added for the minor version 1.2 to 1.2.1 upgrade:

- It is no longer necessary to install a version of the program which corresponds to a specific version of the Analyst® software – in fact it is not required to have the Analyst® software installed to process LC/MS data in wiff files.
- Processing of MRM data (especially data acquired with the Scheduled MRM™ algorithm) has been enhanced as discussed in section 2.1.2.6. Also see sections 2.2.7 (working with group names), 2.4.5 (calculating group area ratios) and 4.2.2.8 (selecting peaks matching a given pattern in the name).
- A shortcut list of recently saved files is available (section 2.1.4).
- The 'Peak Name' field in the Peaks Table is now editable.