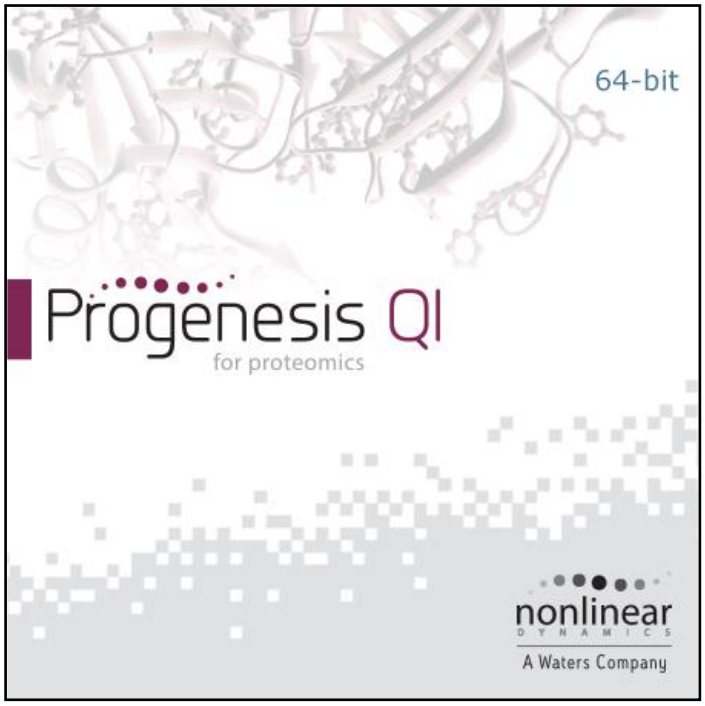


Progenesis QI for proteomics User Guide

Analysis workflow guidelines for DDA data

Waters

THE SCIENCE OF WHAT'S POSSIBLE.™

Contents

Introduction

This user guide takes you through a complete analysis of 6 LC-MS runs with 2 groups (3 replicate runs per group) using the unique Progenesis QI for proteomics workflow. It starts with LC-MS data file loading then Alignment, followed by Analysis that creates a list of interesting peptide ions (peptides) which are explored within Peptide Ion Stats using multivariate statistical methods then onto Protein identity and Reporting.

To allow ease of use the tutorial is designed to start with the restoration of an Archived experiment where the data files have already been loaded. The document covers all the stages in the workflow, initially focusing on the Automatic Processing of the data then the use of the individual analysis stages.

If you are using your own data files please refer to Appendix 1 (page 69) then start at page 7.

How to use this document

You can print this user guide to help you work hands-on with the software. The complete user guide takes about 60 to 90 minutes (dependant on PC spec) and is divided into two sections. This means you can perform the first half focused on LC-MS run alignment and analysis then complete the second half of analysis exploring comparative differences and Protein identity at a convenient time.

If you experience any problems or require assistance, please contact us at support@nonlinear.com

How can I analyse my own runs using Progenesis QI for proteomics?

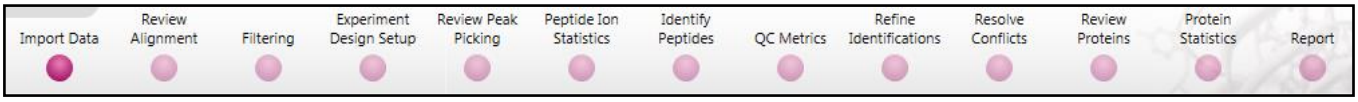
You can freely explore the quality of your LC-MS data using Data Import and then licence your own LC-MS runs using this evaluation copy of Progenesis QI for proteomics. Instructions on how to do this are included in a section at the end of the user guide document. Alternatively if you would like to arrange a demonstration in your own laboratory contact support@nonlinear.com and we will help you.

LC-MS Data used in this user guide

NLD would like to thank Dr Robert Parker and Prof Haroun Shah at the Health Protection Agency, London, UK for providing the example data used in this user guide as well as invaluable discussion on the handling of the data.

Workflow approach to LC-MS run analysis

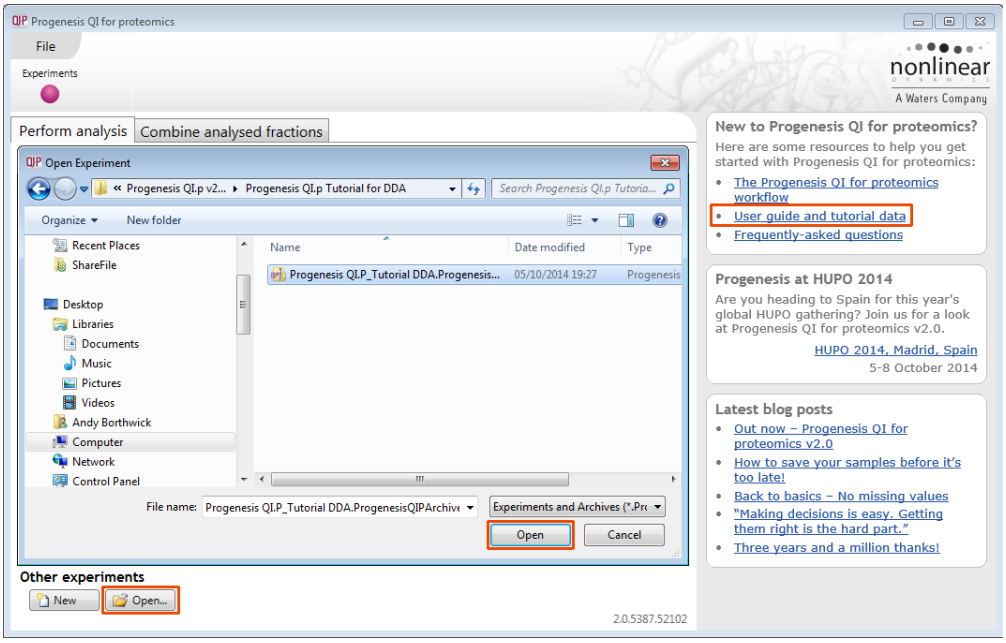
Progenesis QI for proteomics adopts an intuitive **Workflow** approach to performing comparative LC-MS data analysis. The following user guide describes the various stages of this workflow (see below) focusing mainly on the stages from Alignment to Report.



Restoring the Tutorial

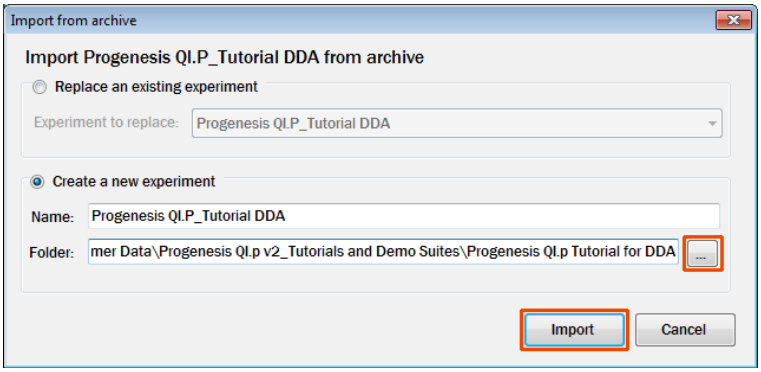
Open Progenesis QI for proteomics and download the Compressed (.zip) Tutorial Archive file from the 'User guide and tutorial' link shown below, placing it in a **new folder** on your desktop. Before restoring the tutorial in the software **you must** first right click on the (.zip) file and extract it to the same folder.

Now you can restore the uncompressed Tutorial archive file. To do this, first locate the **Progenesis QI.P_Tutorial DDA.Progenesis QIP Archive** file using the **Open** button and press Open.

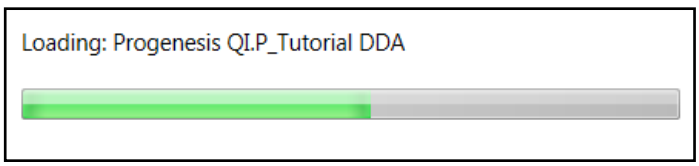


This opens the 'Import from archive' dialog.

Select the **Create a new experiment** option and select the folder in which you placed the archive, using the icon (to the right).



Then press **Import.**

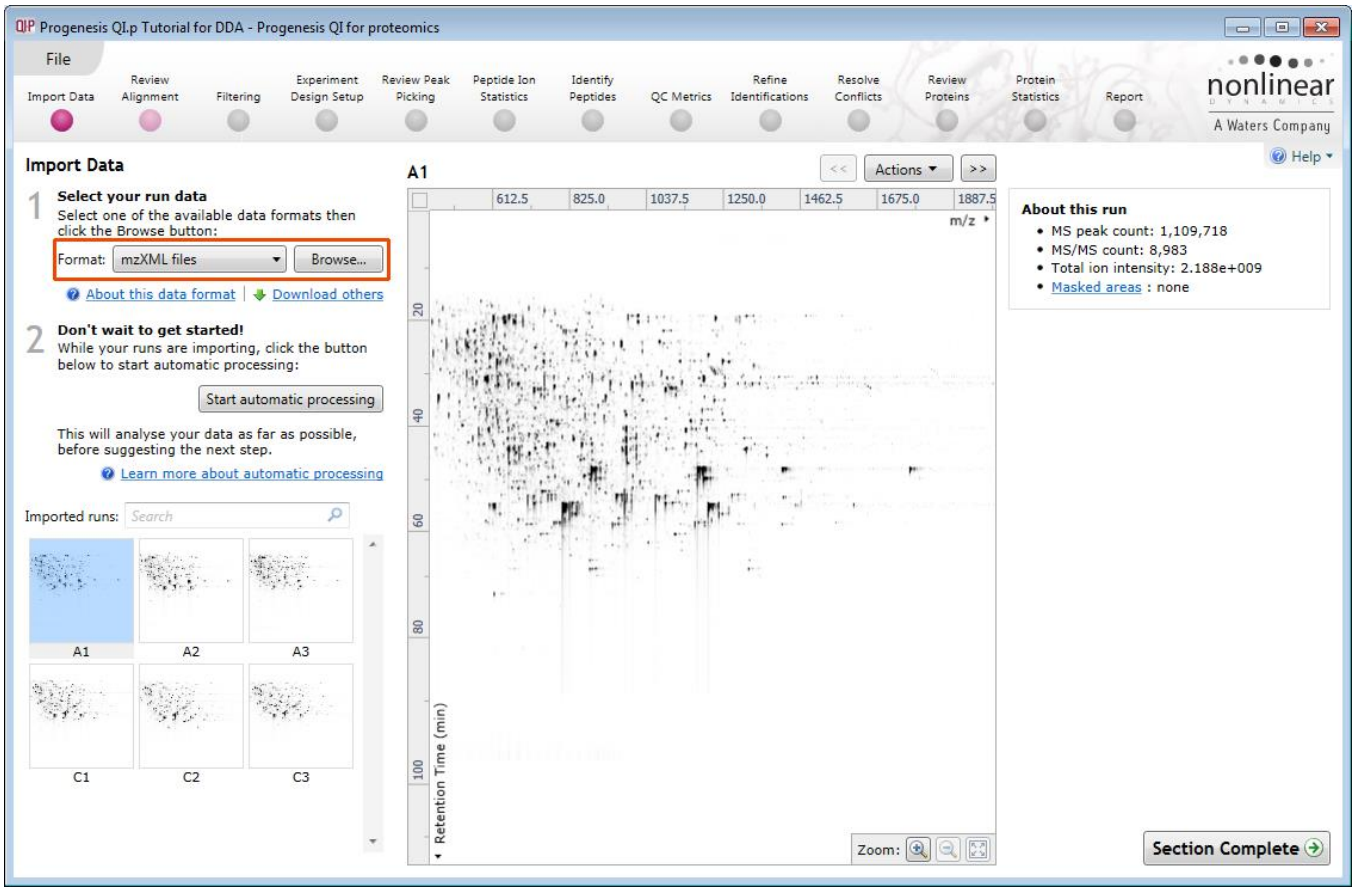


Note: use the **Replace an existing experiment** option if you want to over-write an existing version of the tutorial.

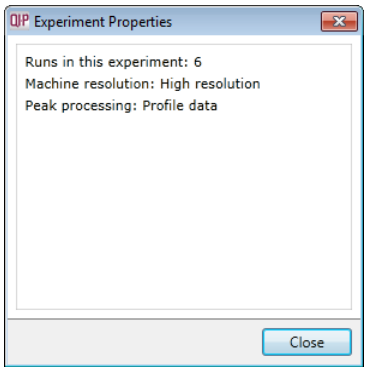
Tip: at each stage in the software there are links to more information and help on the website.

Stage 1: Import Data and QC review of LC-MS data set

The tutorial will now open at the **Import Data** stage (see below).

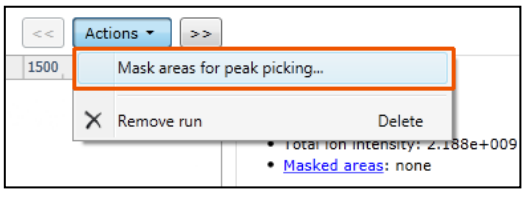


Each data file appears as a 2D representation of the run. At this stage you will be warned if any of the data files have been 'centroided' during the data acquisition and conversion process.



Note: the **Experiment Properties** are available from the File menu**.** These were selected when the experiment was created (see Appendix 1, page 69).

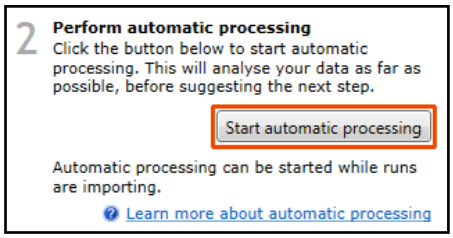
Tip: the **'Exclude areas from selected run'** facility allows you to examine and exclude areas (usually early and/or late in the LC dimension (Retention Time)) that appear excessively noisy due to capture of data during column regeneration (see Appendix 3, page 75). This is not required for this data set.



Note: use the **Remove Run** to remove run(s) from the current experiment.

Now start the Automatic Processing.

Stage 2A: Automatic Processing of your data



The Automatic Processing of your data can be set up and started before the import of your data has been completed by clicking on **Start automatic processing.**

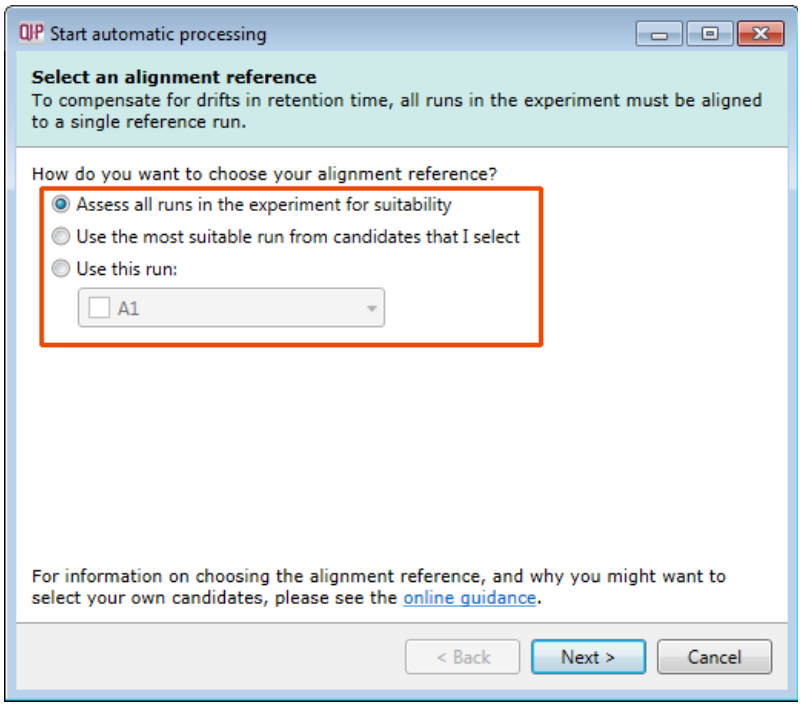
Note: for this tutorial the data has been imported already.

Setup of processing steps in the Analysis Workflow, up to and including Identify Peptides, can be performed in the Automatic Processing wizard, these include:

- Automatic alignment of all runs to a reference run
- Automatic peak picking
- Creating an Experiment design
- Choosing a Quantitation method
- Identification of peptides (only available for MSe and HDMSe data formats)

In this tutorial example you have 6 Data Dependently Acquired (DDA) LC-MS runs, so the automatic identification of peptides is unavailable in the automatic processing.

As the runs have already been imported, click **Start automatic processing** to setup the Automatic processing wizard.



Progenesis QI for proteomics provides three methods for choosing the alignment reference run:

1. **Assess all runs in the experiment for suitability**
   This method compares every run in your experiment to every other run for similarity, then select the run with the greatest similarity to all other runs as the alignment reference.

If you have no prior knowledge about which of your runs would make a good reference, then this choice will normally produce a good alignment reference for you. This method, however, can take a long time for a large number of runs.

2. **Use the most suitable run from candidates that I select**
   This method asks you to make a selection of reference candidates; the automatic algorithm then chooses the best reference from this subset of runs.

This method is appropriate when you have some prior knowledge of your runs suitability as references:
i.e when all the candidate runs are pooled samples or
i.e when all the candidates are from a condition that displays the largest set of common peptide ions.

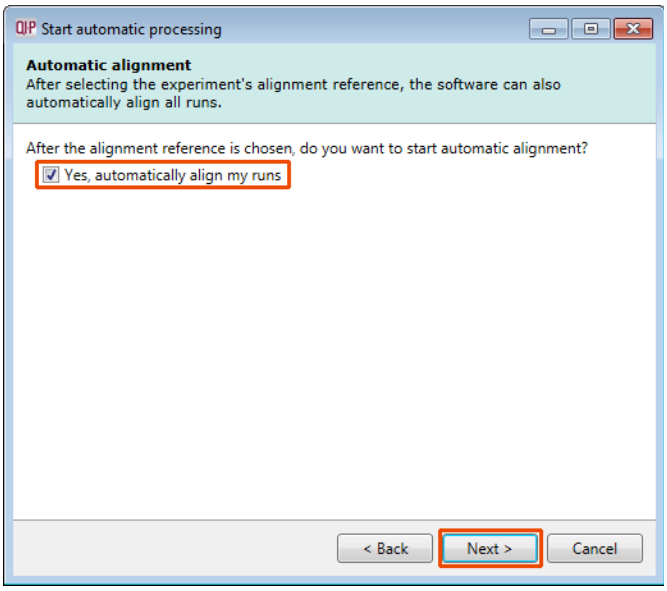
3. **Use this run**
   This method allows you to manually choose the reference run.

Manual selection gives you full control, but there are a couple of risks to note:

- If you choose a pending run which subsequently fails to load, alignment will not be performed.
- If you choose a run before it fully loads, and it turns out to have chromatography issues, alignment will be negatively affected (for this reason we recommend that you let your reference run fully load and assess its chromatography before loading further runs).

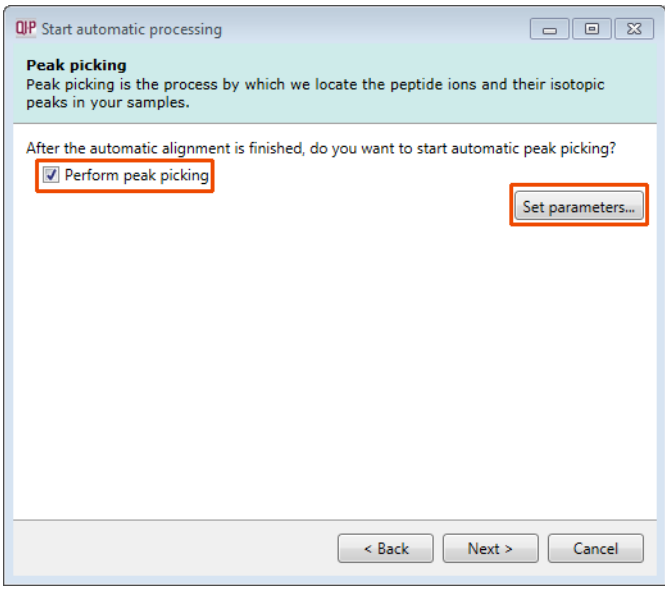
For this tutorial we will select the first option (See Appendix 1, page 69 for more details on using the other options).

You will now be asked if you want to align your runs automatically.



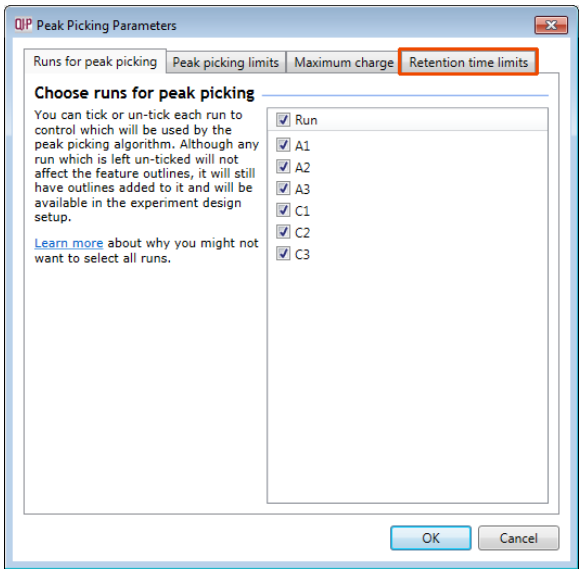
The default is for automatic alignment, click **Next**.

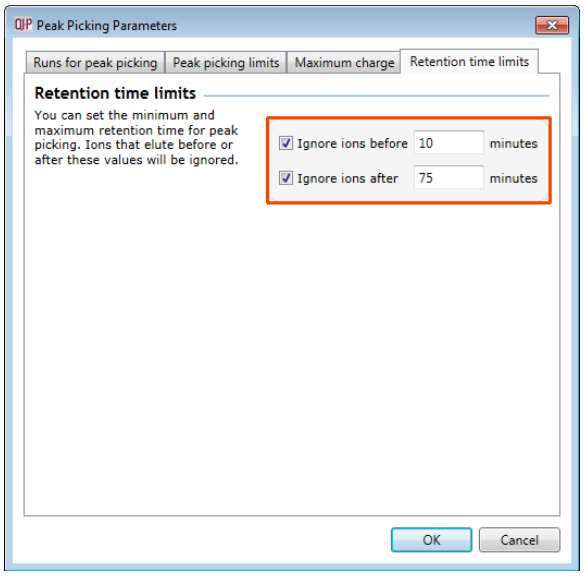
The next page of the processing wizard asks you if you want to **Perform peak picking** and allows you to set appropriate parameters.



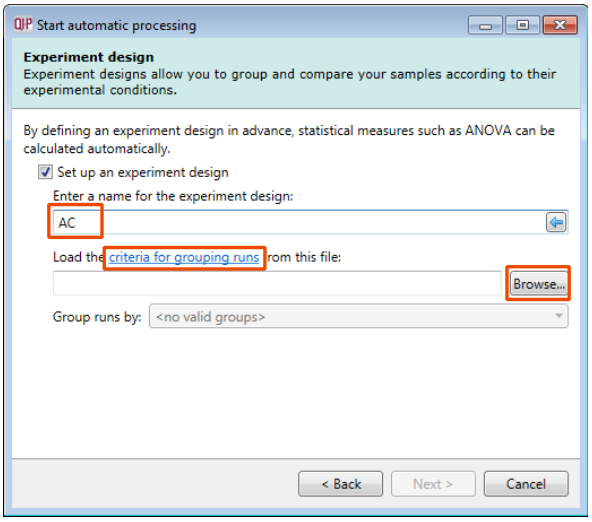
For the purposes of this User guide we will use the default settings for peak picking except in the case of Retention Time limits where we will limit the peak picking to between 10 and 75 min.

Note: for more details on setting Peak Picking parameters refer to the section on Filtering (page 18)
Click on **Set parameters** and select the fourth tab to set **Retention time limits** for the detection. The default limits are displayed in the unticked boxes and detection will use these values whenever the box(s) are unticked. Enter values of 10 and 75 min and tick the boxes as shown below.





Click **OK** to return to the Automatic Processing Wizard and click **Next** to Define an Experiment design.



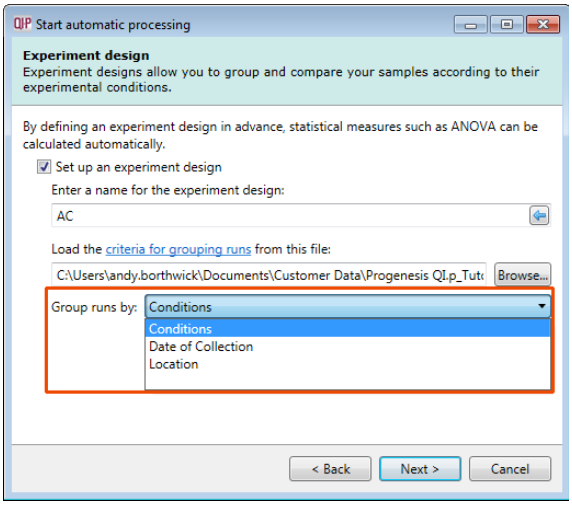
To handle the grouping of your run data you can make use of **sample tracking information** that has been stored in a spread sheet at the time of sample collection and/or preparation.

Note: if a spreadsheet file of your data is not available you can create your experiment designs after the automatic analysis of the runs

For this example there is a **Tutorial Groups.csv** file available with the Tutorial Archive you restored at the beginning of this tutorial exercise.

Sample Name	Conditions	Date of Collection	Location
A1	A	02/03/2011	Fridge A
A2	A	02/03/2011	Fridge A
A3	A	02/03/2011	Fridge B
C1	C	06/03/2011	Fridge B
C2	C	06/03/2011	Fridge B
C3	C	06/03/2011	Fridge A

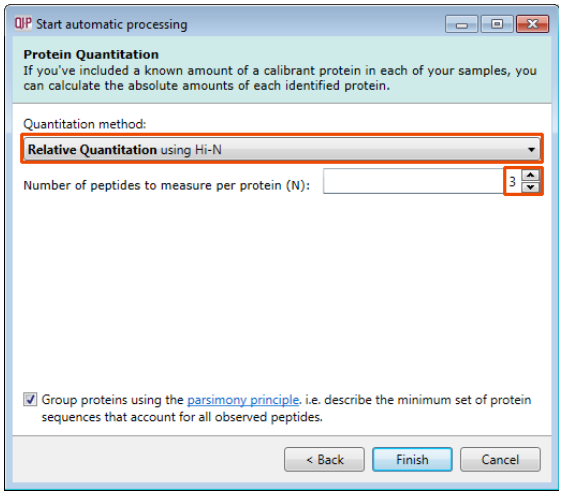
Give the experiment design a name (i.e. AC) and then use the **Browse** function to locate the Tutorial Groups.csv file.
The 'Group runs by' drop down will update to reflect the possible fields in the csv file that you can use to 'group' your runs by: Conditions, Date of collection or Location.

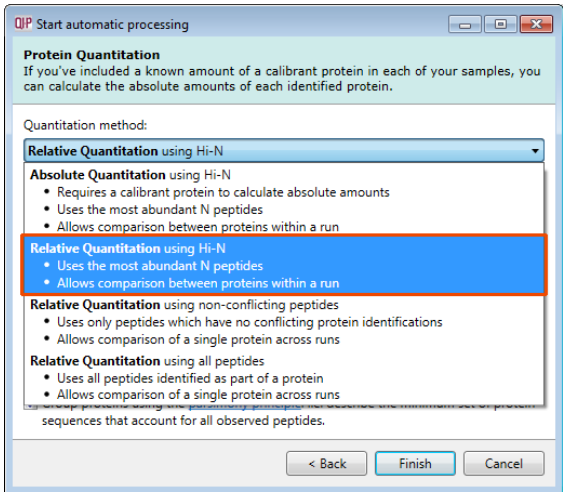


Select Conditions and then click **Next**.

Note: you can create additional experiment designs following the completion of automatic processing.

The Protein Quantitation dialog opens displaying the default method, **Relative Quantitation using Hi-N** which uses up to 3 peptides per protein to compute the relative amount of each 'identified' protein.

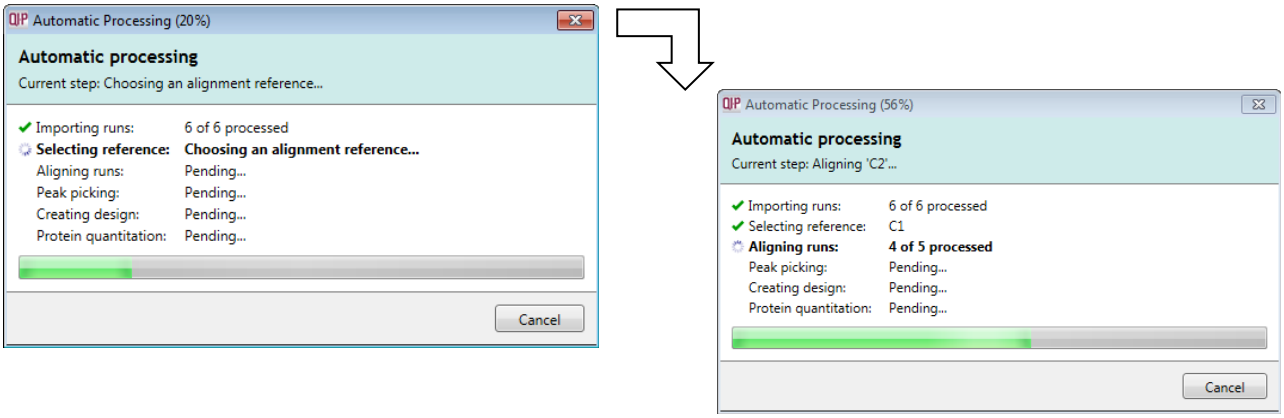




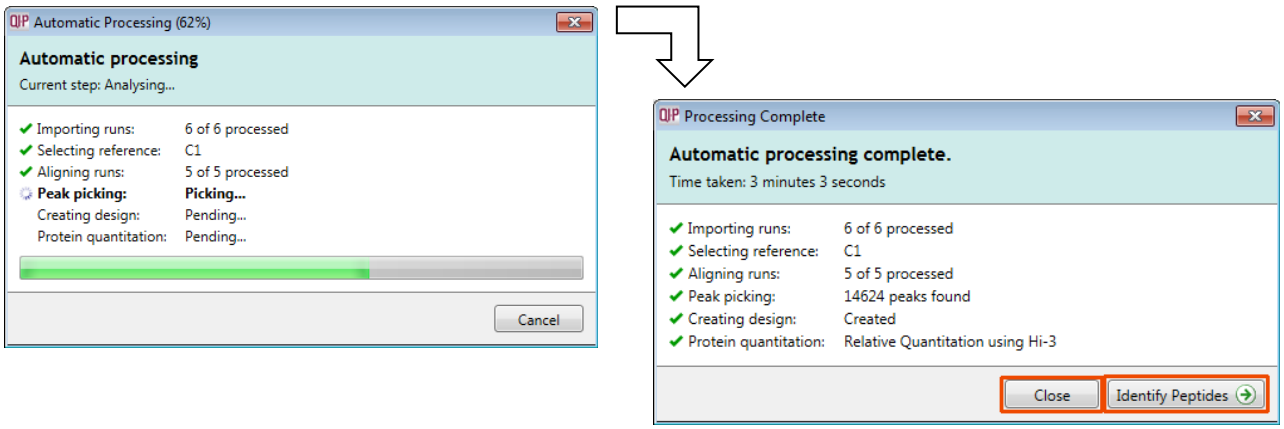
Use the drop down to reveal the alternative methods for protein quantitation

Select the Default option **Relative Quantitation using Hi-N** and click **Finish.**

The Alignment process starts with the automatic selection of C1 as the alignment reference



Once Alignment completes Peak Picking commences



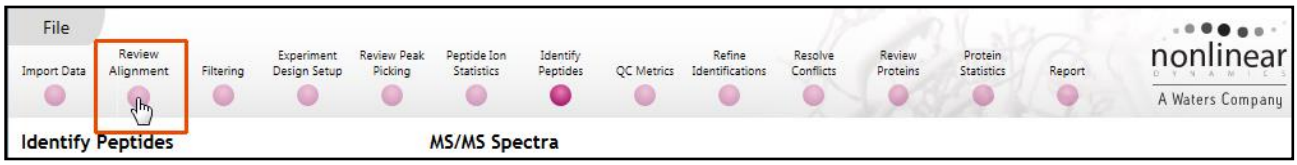
As the whole process proceeds you get information on what stage has been performed and also the % of the process that has been completed.

When Processing completes, depending on what stages you selected to perform, the Wizard displays (on the right) what stage the workflow will open at.

In this example, as the data is DDA it is going to open at **Identify Peptides** if all possible processing steps are selected.

You can either:

- Continue with the analysis, as the Processing dialog is not displaying any warnings, and perform **Identify Peptides**. In which case you can go to page 41
- Open the analysis at Identify Peptides and immediately return to the Review Alignment stage by clicking on it in the Workflow to review alignment quality (page 14)
- Alternatively you can **Close** the dialog. This will **not** move you to a later stage in the workflow but instead will allow you to review the 'Run Chromatography' then navigate systematically through all the stages yourself or jump from stage to stage.



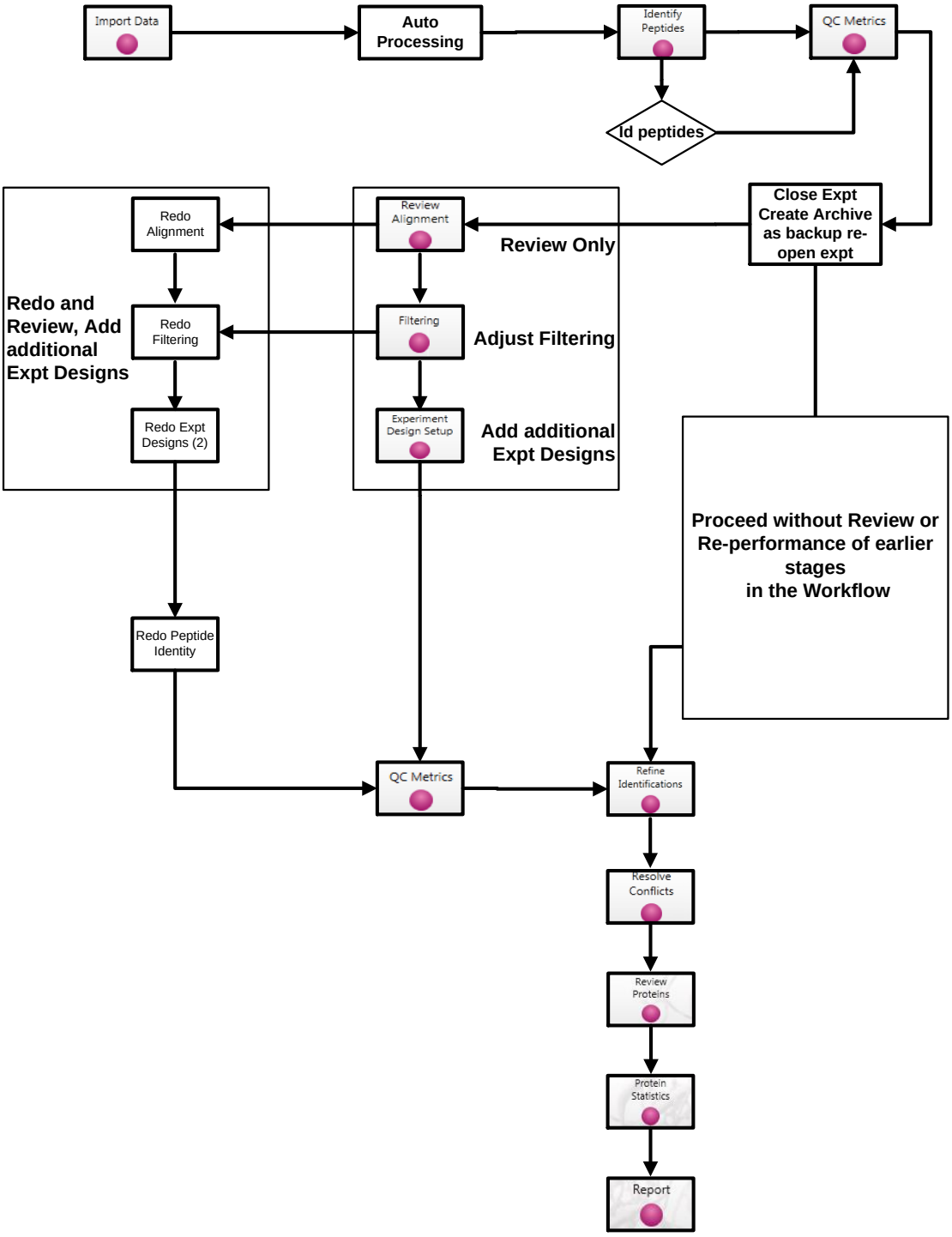
Note: if processing fails to complete successfully there are a number of suggested strategies you can use to proceed with your analysis. These are outlined in Appendix 2 (page 73)

The next stage in this document gives you a graphical view of how to proceed with your analysis following automatic processing.

Stage 2B: After Automatic Processing

When Processing completes, depending on what stages you selected to perform, the Automatic Processing Dialog displays what stage the workflow will open at.

The flow chart gives you an overview of the various steps you can take to proceed with your automatically processed data.



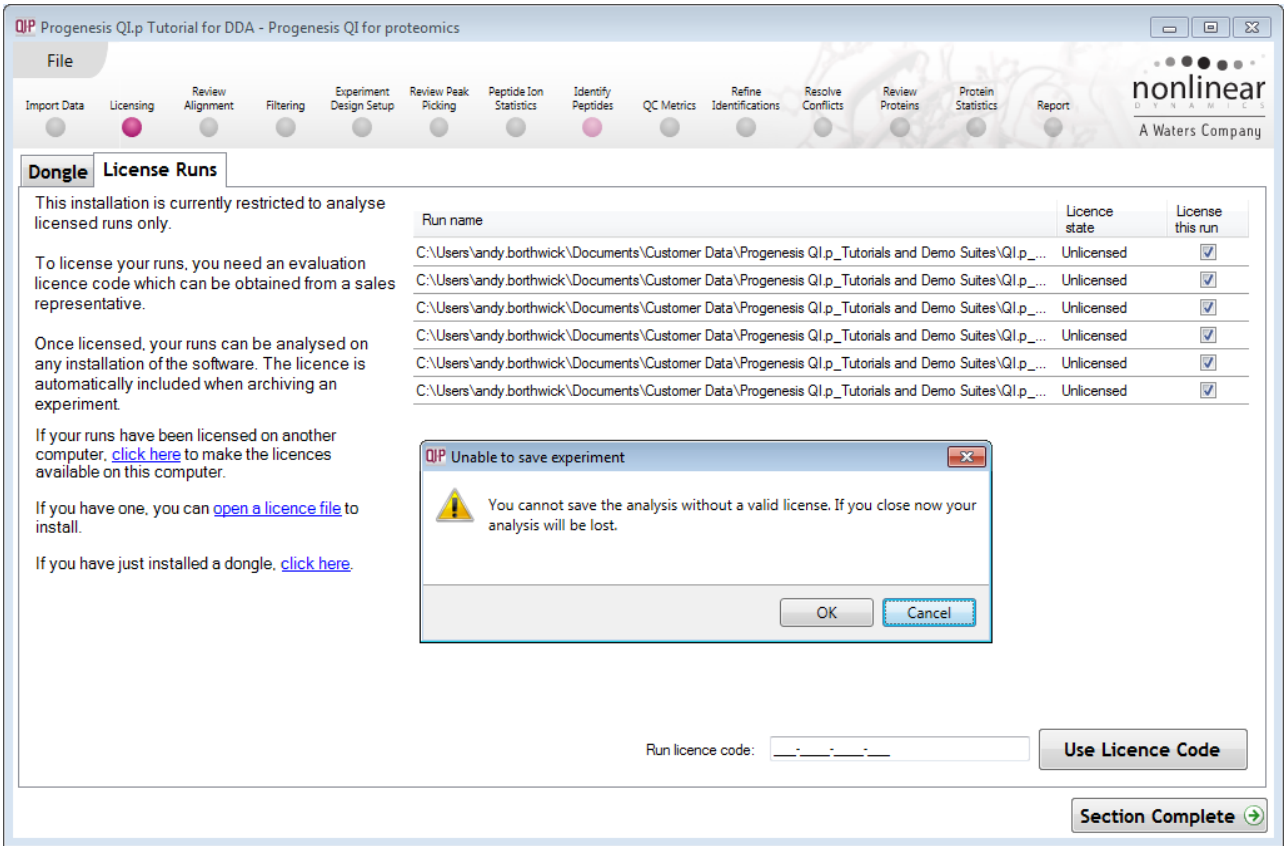
In this example, as the data is DDA it is going to open at Identify Peptides so you can either:

- Continue with the analysis, as the Processing dialog is not displaying any warnings, and perform **Identify Peptides**. In which case you can go to page 41
- Open the analysis at Identify Peptides and return to the Review Alignment stage by clicking on it in the Workflow (page 14).

Stage 3: Licensing

This stage in the analysis workflow will **only** appear if you are using 'Unlicensed' data files to evaluate the software and have no dongle attached.

If you have performed an analysis using Automatic Processing without a valid dongle or do not have the appropriate code to licence your runs, and you close Progenesis QI for proteomics you will be warned that the analysis will be lost.



Note: although the analysis will be lost (Alignment Peak Picking etc) the experiment file and all the 'loaded pre-processed runs are retained'. This allows you to reopen the experiment, when you have a licenced code and/or dongle available, and redo the analysis steps without having to reload the data.

For more details on managing the licencing of your data refer to Appendix 4 (page 76)

If you are using the tutorial archive, this page will not appear as the data files are licensed.

Stage 4: Review Alignment

At this stage Progenesis QI for proteomics Alignment opens displaying the alignment of the runs to the Reference run (C1).

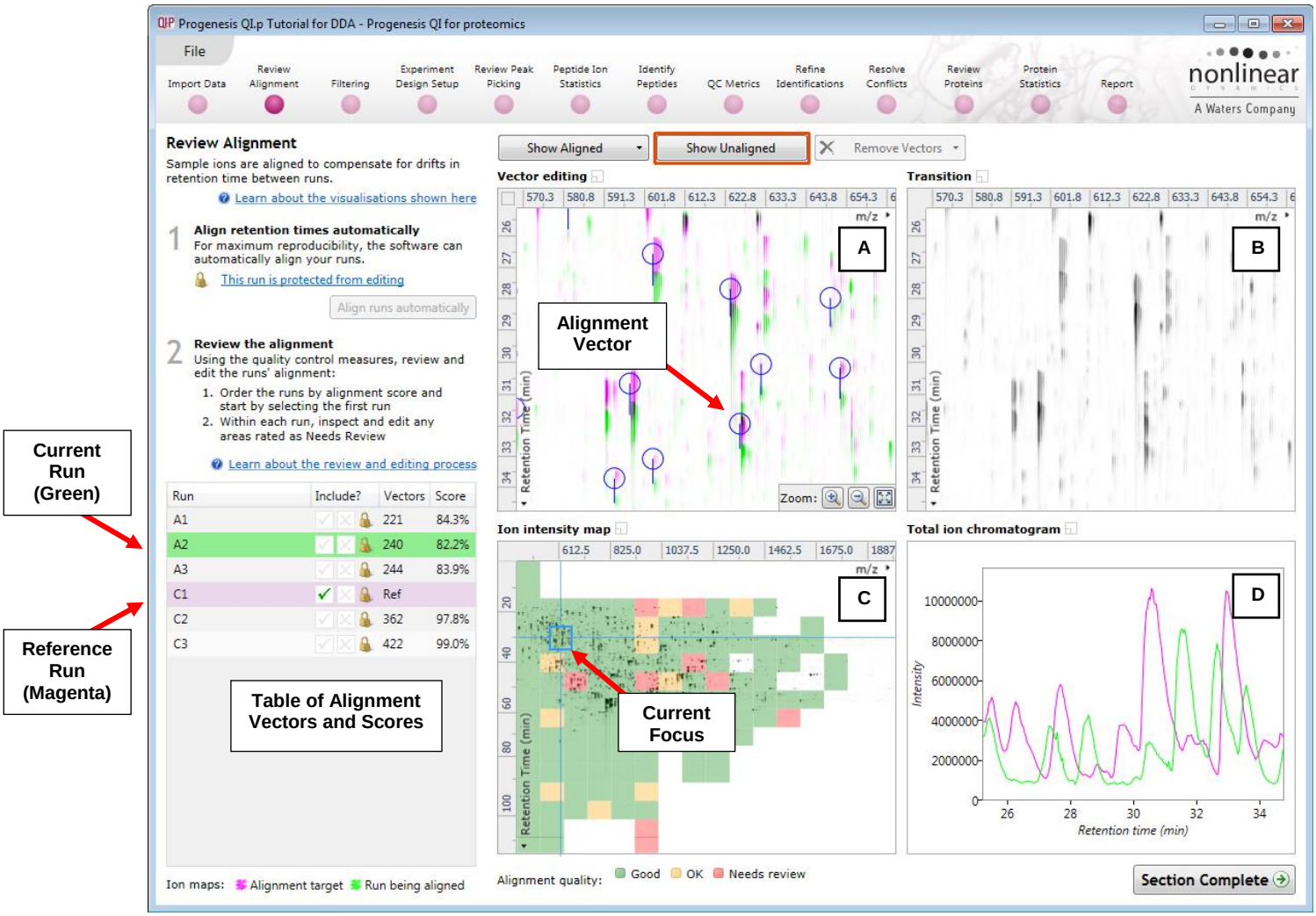
Layout of Alignment

To familiarize you with Progenesis QI for proteomics Alignment, this section describes the various graphical views used in the alignment of the LC-MS runs

Run	Include?	Vectors	Score
A1		221	84.3%
A2		240	82.2%
A3		244	83.9%
C1	✓	Ref	
C2		362	97.8%
C3		422	99.0%

To setup the display so that it looks similar to the one above:

- In the Run table click on Run A2 to make it current. You will now be looking at the alignment of A2 to C1 in the unaligned view. Now drag out an area to review on the **Ion intensity map**. The other 3 views will update to reflect the new focus.



Vector Editing (Window A): is the main alignment area and displays the area defined by the current **focus** rectangle shown in Window C. The current run is displayed in green and the reference run is displayed in magenta. Here is where you can review in detail the vectors and also place the manual alignment vectors when required.

Transition (Window B): uses an **alpha blend** to animate between the current and reference runs. Before the runs are aligned, the peptide ions appear to move up and down. Once correctly aligned, they will appear to pulse. During the process of adding vectors, this view can be used to zoom in and also navigate thus helping with accurate placement of manual vectors.

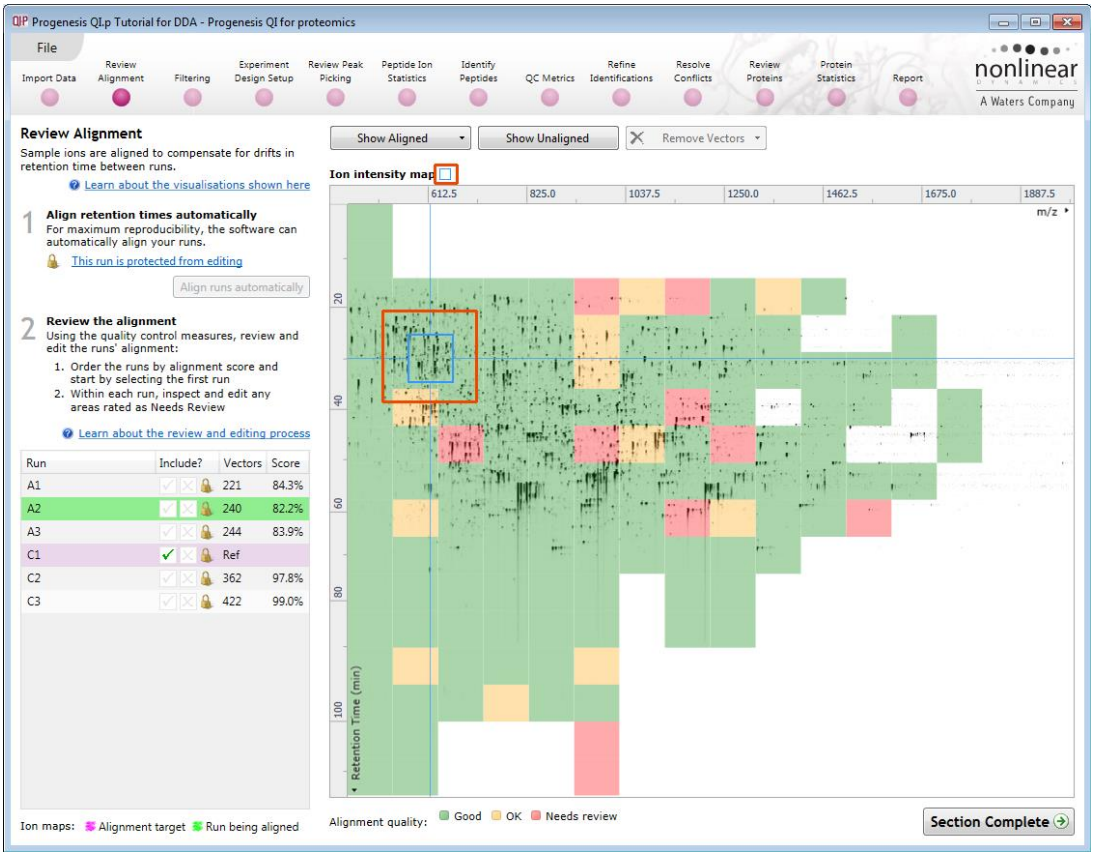
Ion Intensity Map (Window C): shows the **focus** for the other windows. When you click on the view the blue rectangle will move to the selected area. The focus can be moved systematically across the view using the cursor keys. The focus area size can be altered by clicking and dragging out a new area with the mouse. This view also provides a visual quality metric for the Alignment of the runs, from green through yellow to red.

Total Ion Chromatogram (Window D): shows the current **total ion** chromatogram (green) overlaid on the Reference chromatogram (magenta). As the peptide ions are aligned in the **Vector Editing** view the chromatograms become aligned. The retention time range displayed is the vertical dimension of the Focus Grid currently displayed in the **Ion Intensity map** (Window C).

Reviewing quality of alignment vectors

After **Automatic alignment** the number of vectors and Quality Scores will be updated on the **Runs** panel and the vectors will appear (in blue) on the **Vector editing** view.

If the alignment has worked well then the Ion Intensity Map will appear predominantly green and a score of greater than 80% will be reported in the table. Also when **Show Aligned** is selected, in Window A (vector editing) vector length should appear minimal and in Window B (Transition) will show peptide ions pulsing slightly but not moving up and down.



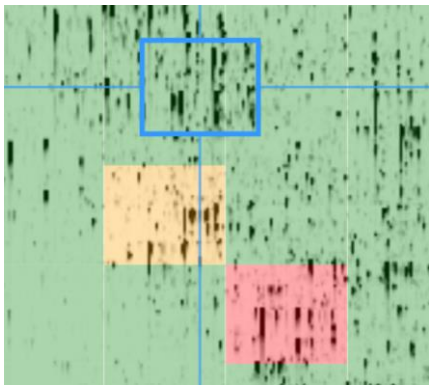
Note: you can use the icon to the right of the panel name to expand or contract each view

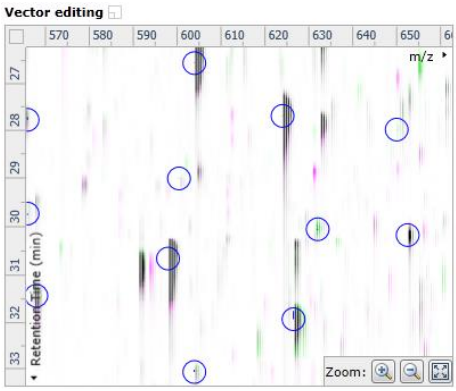
Reviewing Quality of Alignment

At this point the quality metric, overlaid on the Ion Intensity Map as coloured squares, acts as a guide drawing your attention to areas of the alignment. These range from Good (Green) through OK (Yellow) to Needs review (Red). When reviewing individual squares double click on a coloured square to set the focus.

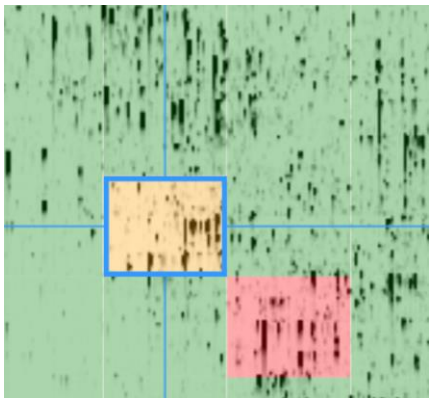
Three example squares are examined here.

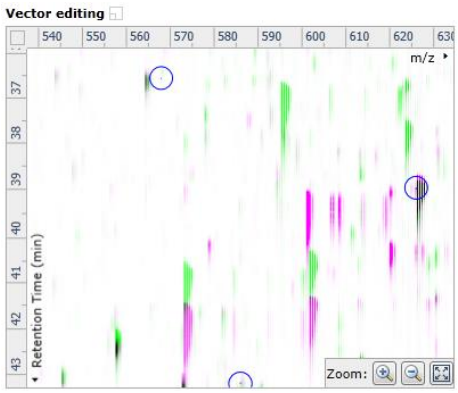
For a ‘green’ square, the majority of the data appears overlapped (black) indicating good alignment. When viewed in the Transition view the data appears to pulse.



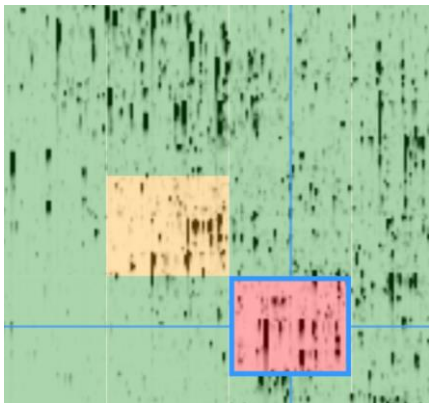


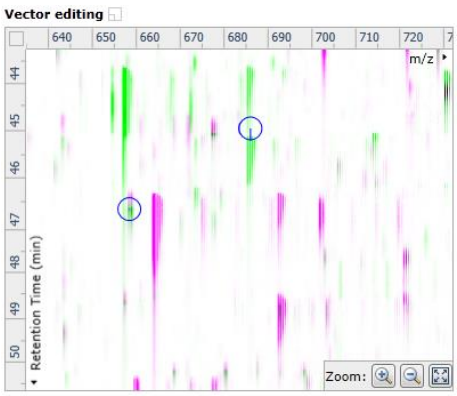
For a ‘yellow’ square, some of the data appears overlapped (black) indicating OK alignment. When viewed in the Transition view some of the data appears to pulse.





For a ‘red’ square little of the data appears overlapped (black) indicating questionable alignment. When viewed in the Transition view little data appears to pulse.

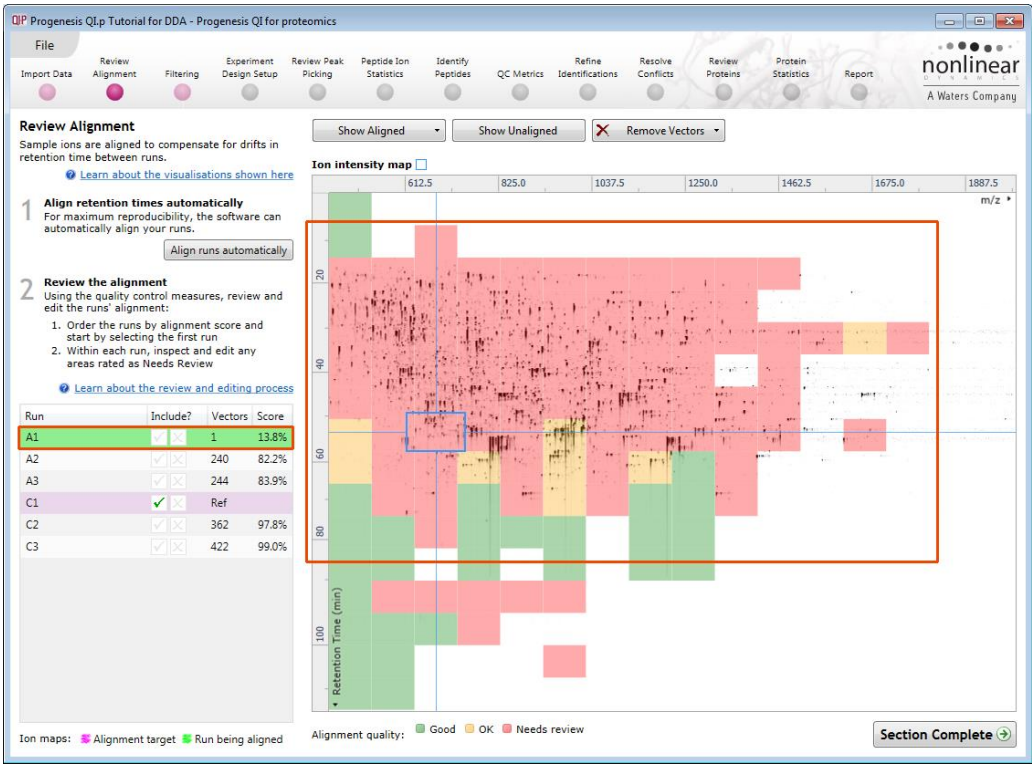




Note: the coloured metric **should be used as a guide**. In cases where there are a few ‘isolated’ red squares this this can also be indicative of ‘real’ differences between the two runs being aligned and should be considered when examining the overall score and surrounding squares in the current alignment.

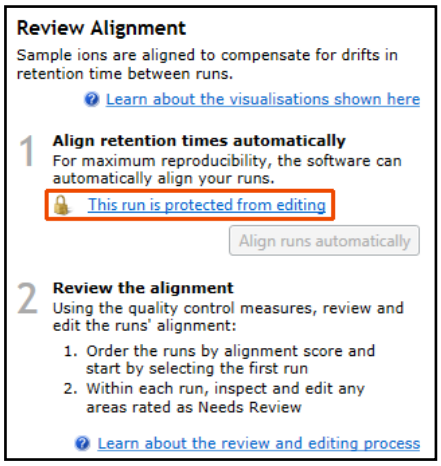
The weighted average of the individual squares gives the overall percentage score for each run alignment.

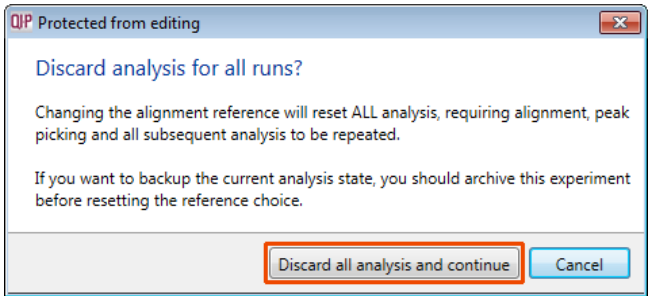
Note: a marked area of red combined with a low score clearly indicates a 'miss alignment' and may require some manual intervention (see Appendix 5, page 77).



Having performed the analysis automatically, in the course of reviewing the quality of alignment you decide that the alignment requires editing then you can unlock the analysis.

To do this click on the **This run is protected from editing** link above the 'greyed out' **Align runs automatically** button.





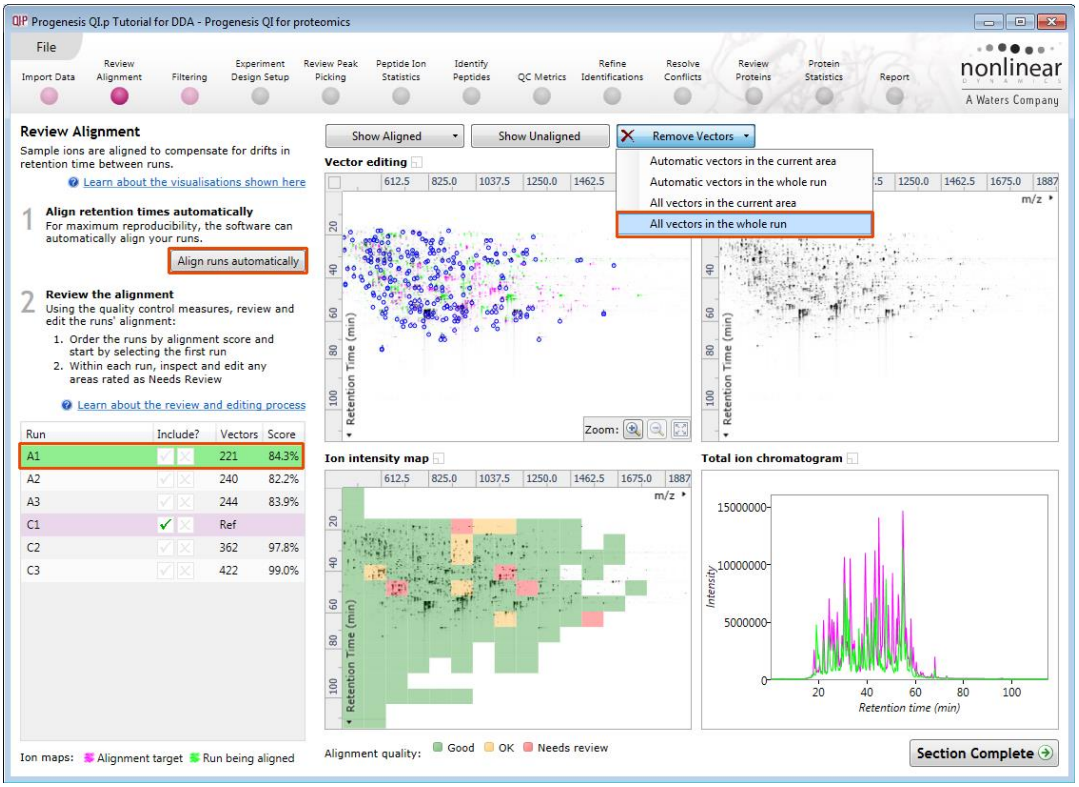
Note: as you click on the link, you will be warned that you are discarding the current analysis (all steps beyond alignment).

Details on editing alignment are described in Appendix 5 (page 77)

The alignment quality of this tutorial data set does not require any manual intervention.

If you have chosen to discard the current automatic analysis and have been exploring the alignment of one or more of the runs using manual vectors (for the purposes of this tutorial) make sure you first remove all manual vectors and then re-perform the Automatic alignment.

To do this for A1, first select Remove 'All vectors in the whole run' and then click Align runs automatically.



Having re-performed the Automatic alignment the Vectors and scores will appear as shown above.

Since you have unlocked your analysis you must now perform the Peak Picking that was originally performed as part of the **Automatic processing**.

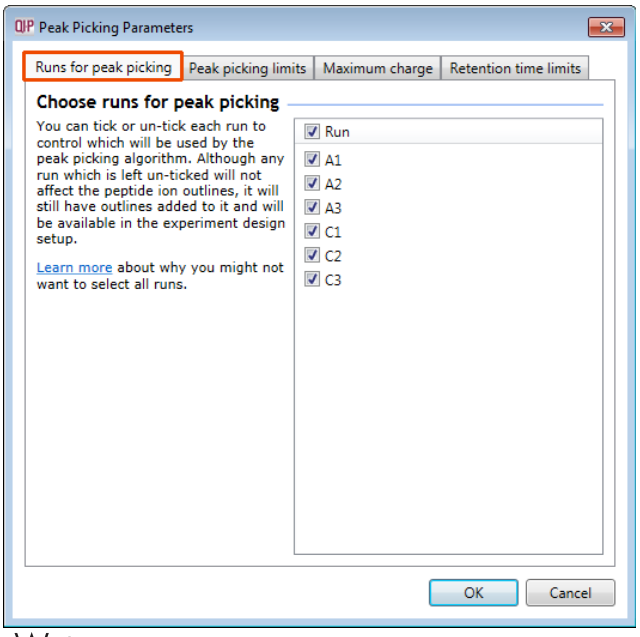
To move to Peak Picking click on **Filtering** on the Workflow or **Section Complete** (bottom right).

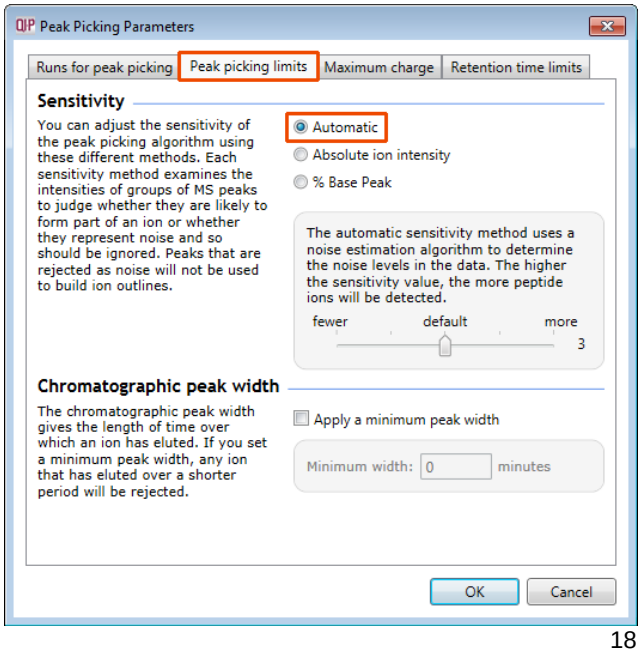
Stage 5A: Filtering

The Peak picking Parameters dialog opens if Picking has not been performed if it has been performed move to section 5B.

Peak Picking Parameters

The Peak Picking Parameters dialog opens, showing all the runs in the current experiment and a tick against each run. This is the default setting, where the peak picking algorithm uses information from all of the runs to contribute to the pattern of peptide ion outlines.





Tip: It may be appropriate **only** to pick peaks that are present in a limited number of your runs. In which case un-tick the runs that you do **NOT** want to contribute to the peptide ion detection pattern. This may be important when one or more of the 'replicate' runs appear noisy due to non-optimal chromatography or sample handling.

Note: peptide ions outlines will be added to 'un-ticked' runs; although these runs will **not** contribute to the peak picking pattern.

Tip: depending on run quality, a suggested minimum number of ticked runs should include at least one replicate of each experimental condition.

The sensitivity of the peak picking can be controlled by adjusting settings under the **Peak picking limits** tab.

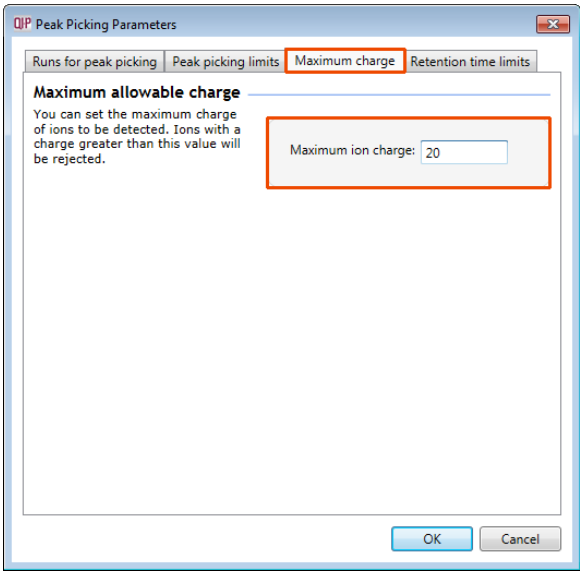
Note: for all 3 **Sensitivity** modes a Chromatographic peak width (Retention time window) for the peaks **can** be set by applying a minimum retention window or peak width in minutes.

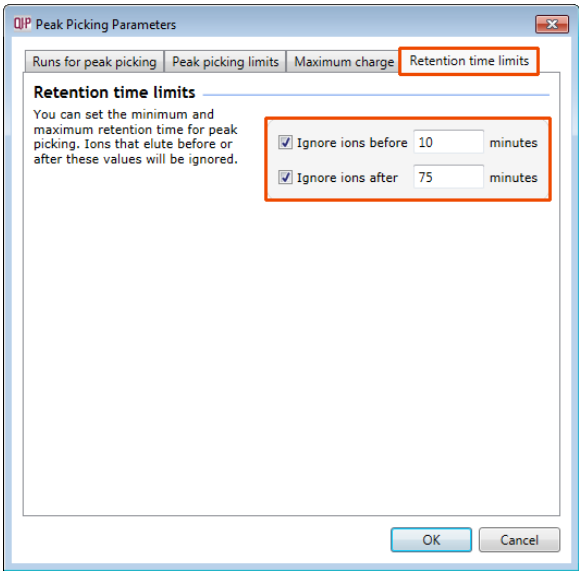
For the runs in this user guide we will use the default settings for the Automatic method and NOT apply a minimum peak width (as shown above).

The third tab allows you to set the **maximum charge** of the peptide ions, which will be detected. The default setting is a charge state of 20. For this example, leave this set as default.

Note: you can either, leave this set as default and remove the high charge state peptide ions at the Filtering stage or you can choose not to detect them in the first place by reducing the charge state threshold here.

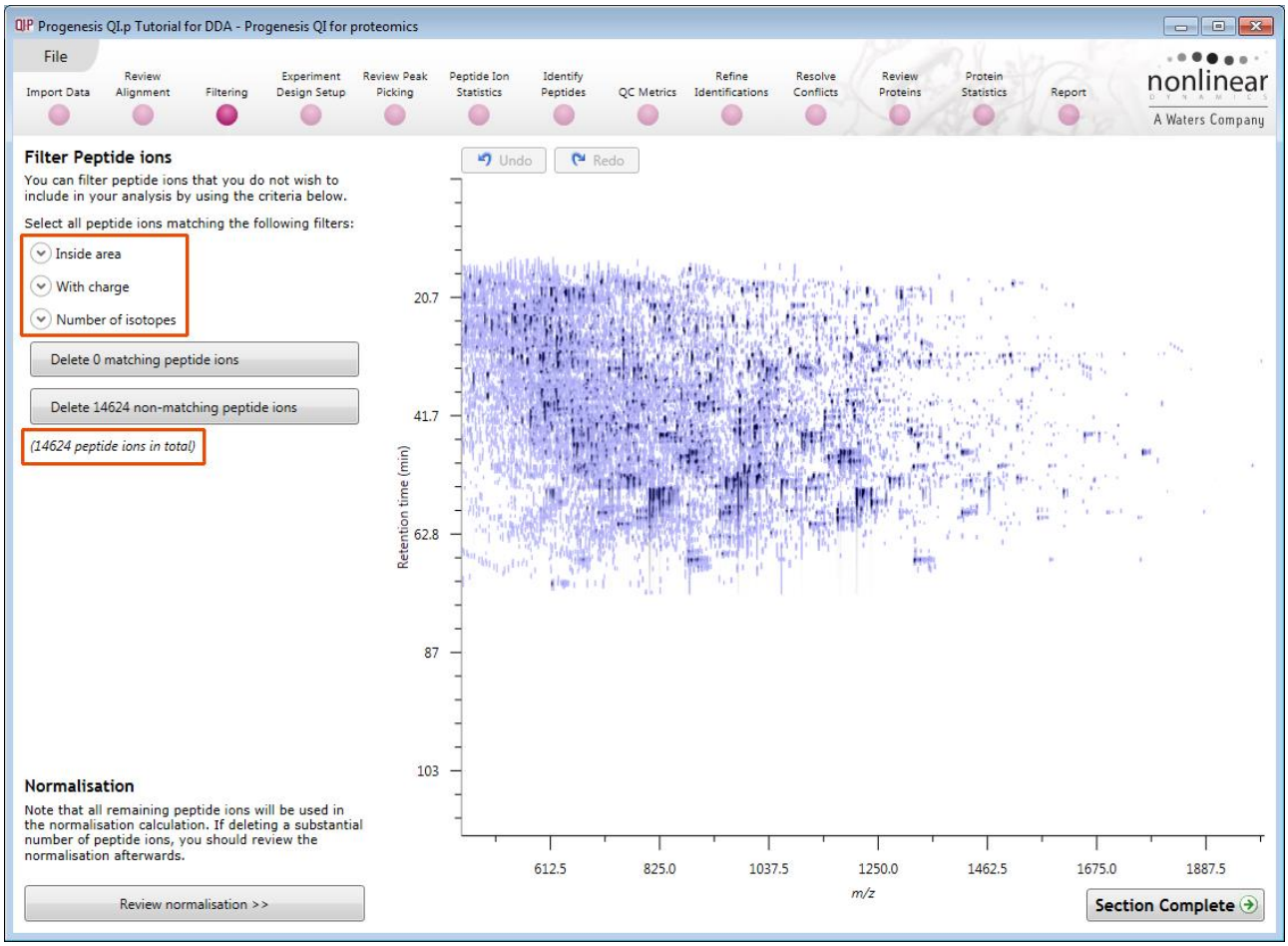
Finally, you can set **Retention time limits** for the detection. The default limits are displayed in the unticked boxes and detection will use these values whenever the box(s) are unticked. Enter values of 10 and 75 min and tick the boxes as shown below.





Click **OK** to start the detection process.

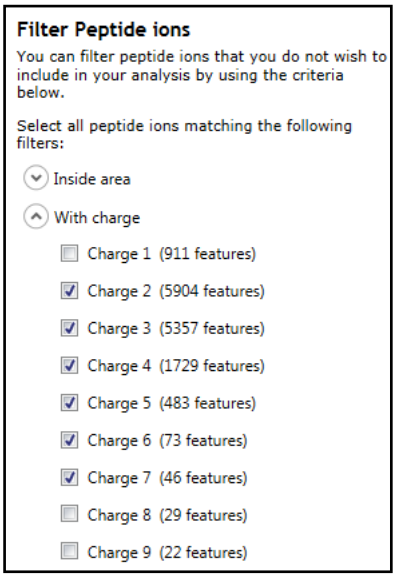
On completion of detection, the Filtering stage will open displaying the number of peptide ions detected, in this example there are 14624 peptide ions.



If required you can remove peptide ions based on position, charge state, number of isotopes or combinations of these peptide ion properties.

As an example, we will filter the peptide ions based on charge ‘charge state’.

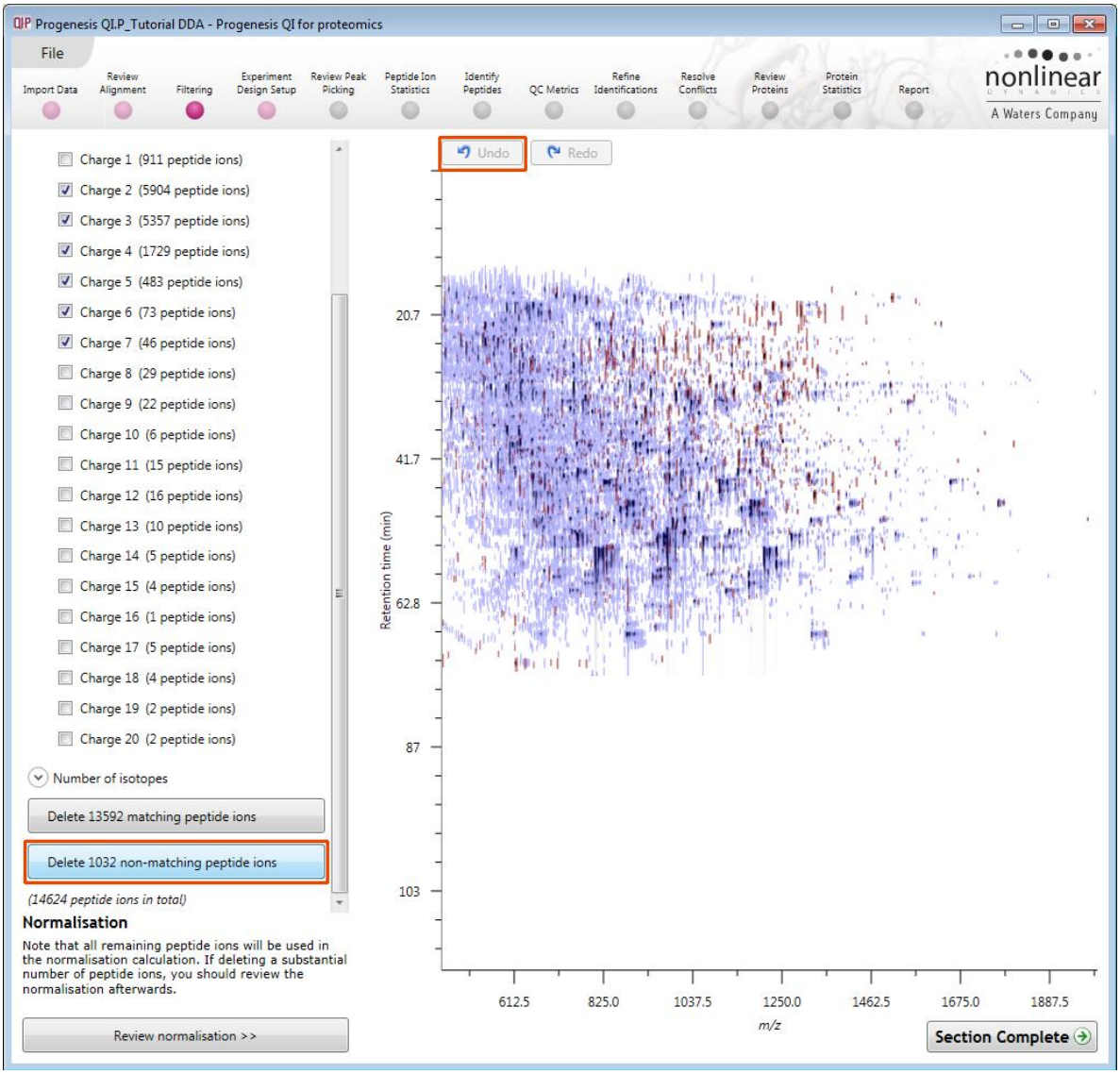
When **With charge** is selected the number of peptide ions present at each charge state is displayed, these can be selected accordingly. In this case we will retain peptide ions with a charge state of 2 to 7.



Area limits, charge state and number of isotopes can be combined to refine the peptide ion selection.

Tip: when filtering on only one property of the peptide ion i.e. charge state, make sure you have 'collapsed' the other filters (see right)

Hence all peptide ions with a charge state of 1 or 8 and above will appear red on the main view as you hold the cursor over the appropriate delete button.



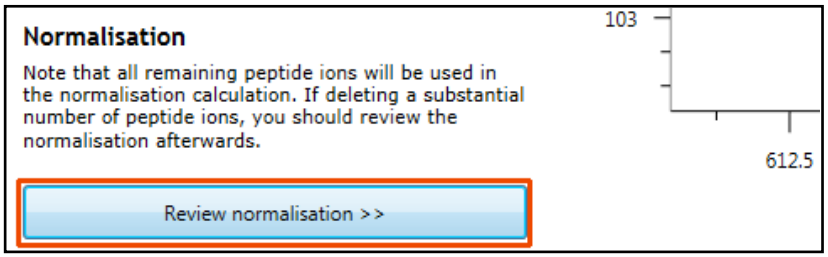
To remove these peptide ions press **Delete 1032 Non Matching Peptide ions.**

You can use the **Undo** button to bring back deleted peptide ions, however, when you move to the next section you will lose the capacity to undo the filter. Before moving on from filtering you can review the normalisation of the data.

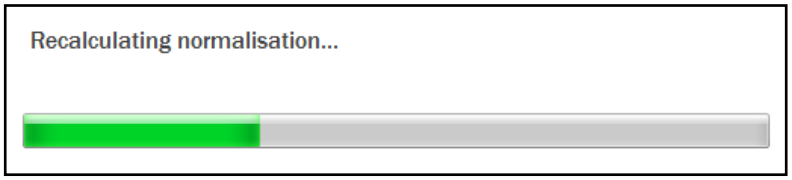
Tip: When you have reached the filtering stage, it is good practice to close the experiment and save an archive. This can be used to restore the unfiltered state if the filtering you have performed is too 'stringent'.

Stage 5B: Reviewing Normalisation

Review normalisation is accessed from the button at the bottom left corner of the filtering page.



If you have filtered out a number of peptide ions from the original detection pattern then the normalisation will update.



The **Review Normalisation** page will open displaying plots for the normalisation of all the peptide ions on each run.

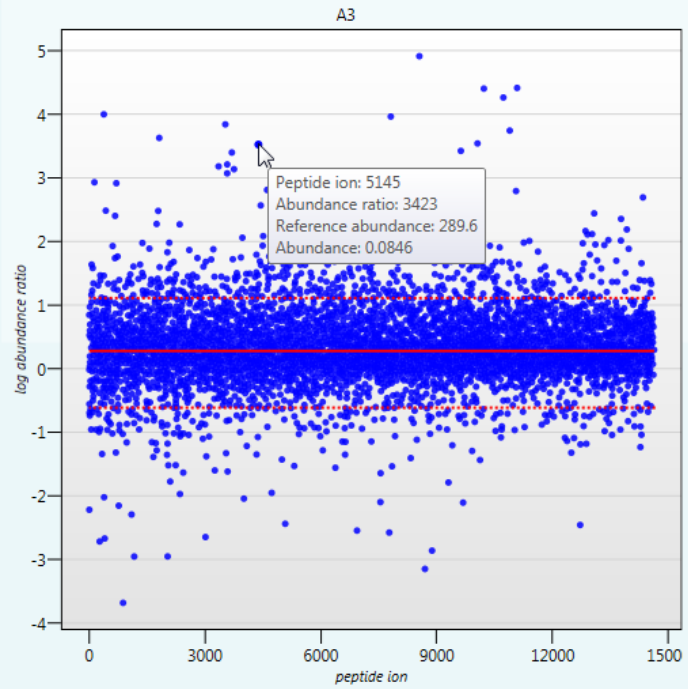
This page in the workflow **does not** allow you to alter the Normalisation of your data but provides you with individual views for each run showing the data points used in the calculation of the normalisation factor for the run.

Alternatively, if you do not believe normalisation is necessary, you can opt to ‘Don’t use any normalisation’ for the rest of the analysis (Normalisation Method tab).

Normalisation factors are reported in the table to the left of the plots.

Calculation of Normalisation Factor:

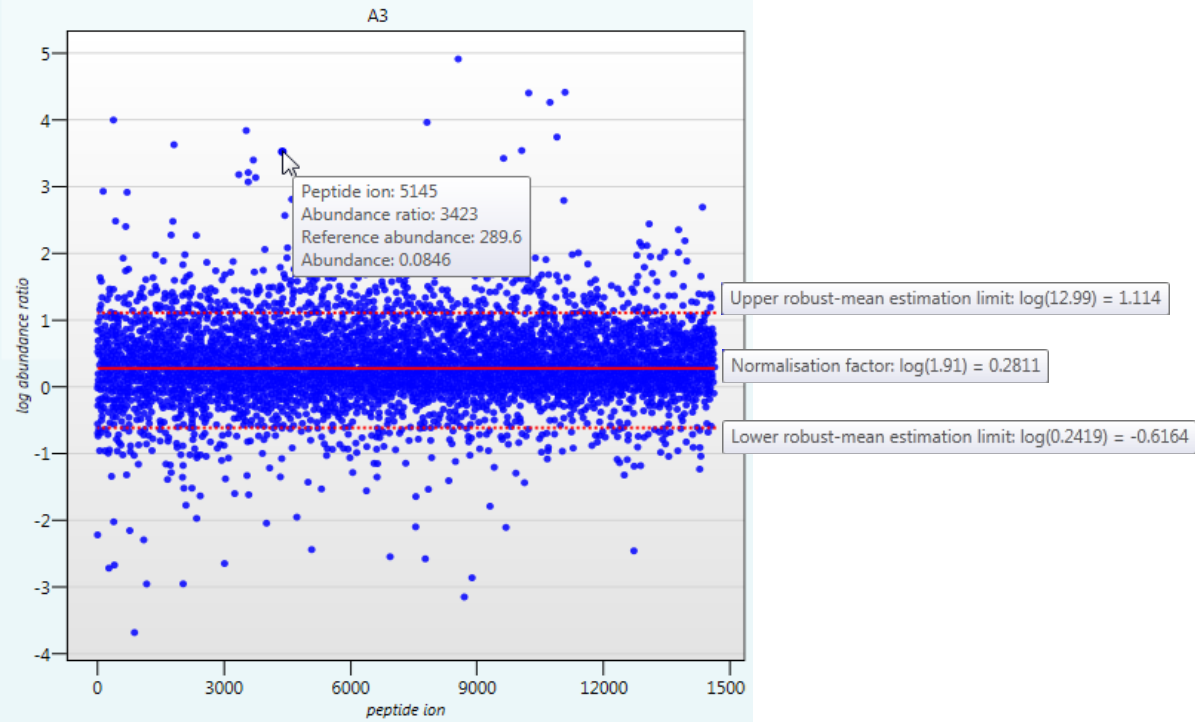
Progenesis QI for proteomics will automatically select one of the runs that is 'least different' from all the other runs in the data set to be the 'Normalising reference'. The run used, is shown above the table of Normalisation factors.



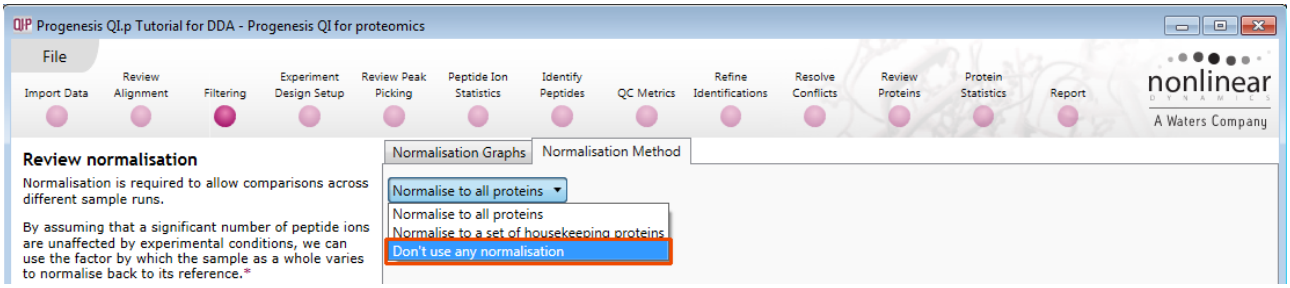
For each sample run, each blue dot shows the log of the abundance ratio for a different peptide ion (normalisation target abundance/run abundance).

The details for individual peptide ions can be viewed as you hold the cursor over the dots on the plot.

On the graph the peptide ions are shown ordered by ascending mean abundance. The normalisation factor is then calculated by finding the mean of the log abundance ratios of the peptide ions that fall within the 'robust estimated limits' (dotted red lines). Peptide ions outside these limits are considered to be outliers and therefore will not affect the normalisation.



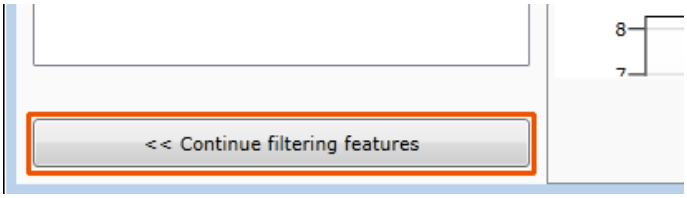
Note: if you do **not** wish to work with normalised data then Select '**Don't use any normalisation'**.



Note: once you have identified a set of proteins, you can then apply the **Normalise to a set of housekeeping proteins** by using this option to locate and select the peptide ions.

For this experiment, you should leave the **Normalise to all peptide ions** option selected.

Now return to filtering by clicking on the button on the bottom left of the screen



For this example, we **DO NOT** do any additional Filtering so click on **Section complete**.

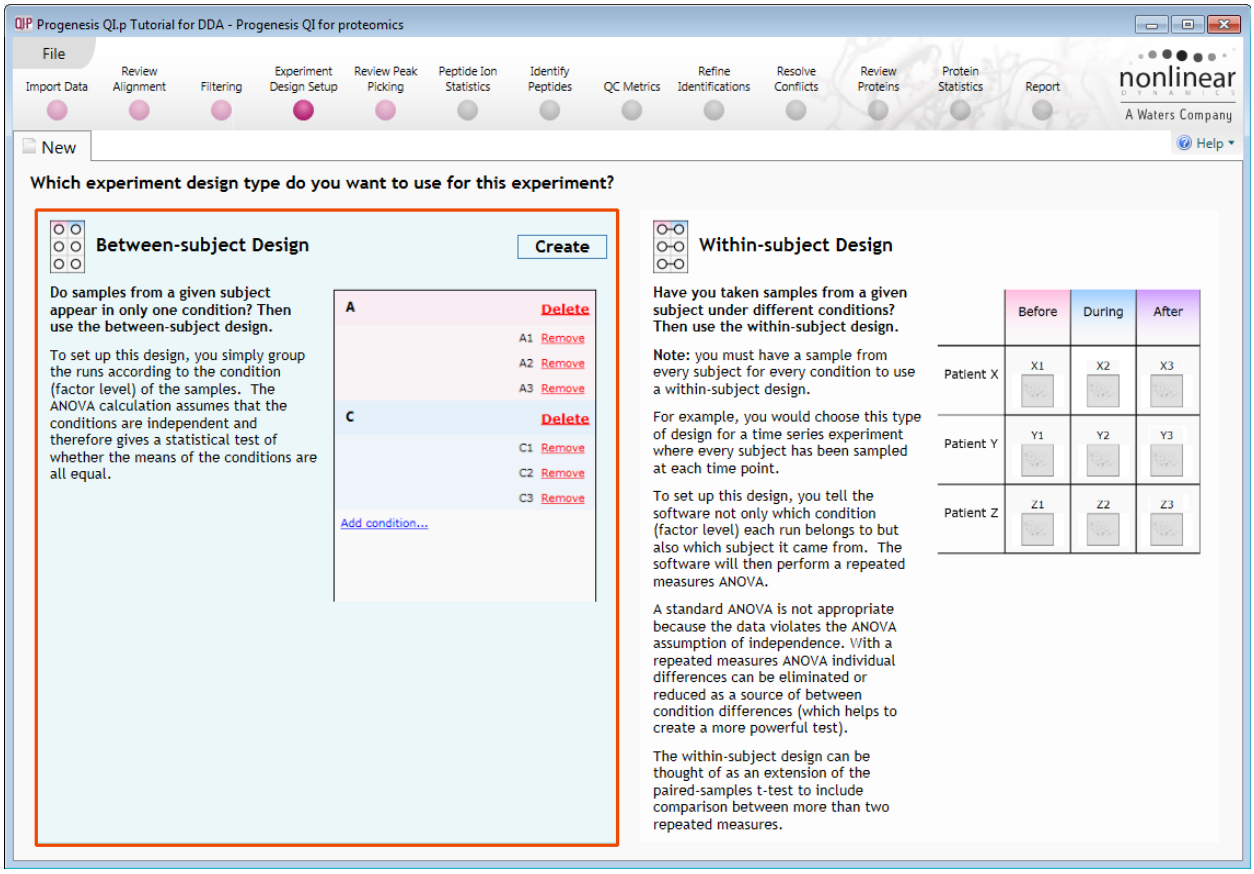
Note: if you do any extra filtering then **Normalisation recalculates** as you move to the next stage in the Workflow**.**

Stage 6: Experiment Design Setup for Analysed Runs

At this stage in the workflow you can setup one or more experimental designs for your data.

There are two basic types of experimental designs:

Between-subject design: here samples from any given subject appear in only one condition. (i.e. control versus various drug treatments).The ANOVA calculation assumes that the conditions are independent and applies the statistical test that assumes the means of the conditions are equal.

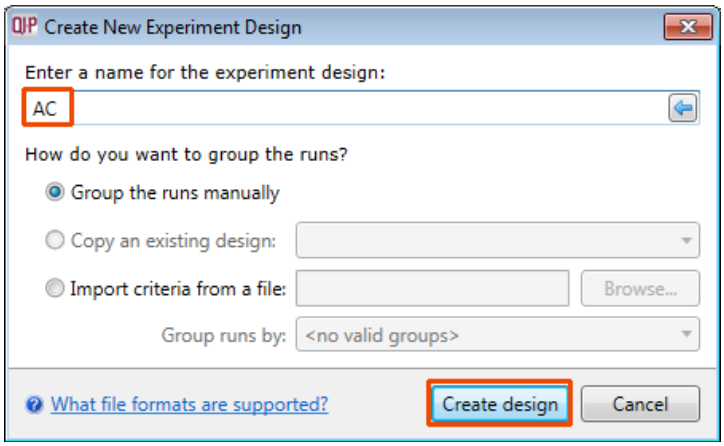


Within-subject design: here samples have been taken from a given subject under different conditions (i.e. the same subject has been sampled over a period of time or after one or more treatments). Here a standard ANOVA is not appropriate as the data violates the ANOVA assumption of independance. Therefore by using a *repeated measures* ANOVA, individual differences can be eliminated or reduced as a source of between condition differences. This within-subject design can be thought of as a extension of the paired samples t-test, including comparison between more than two repeated measures.

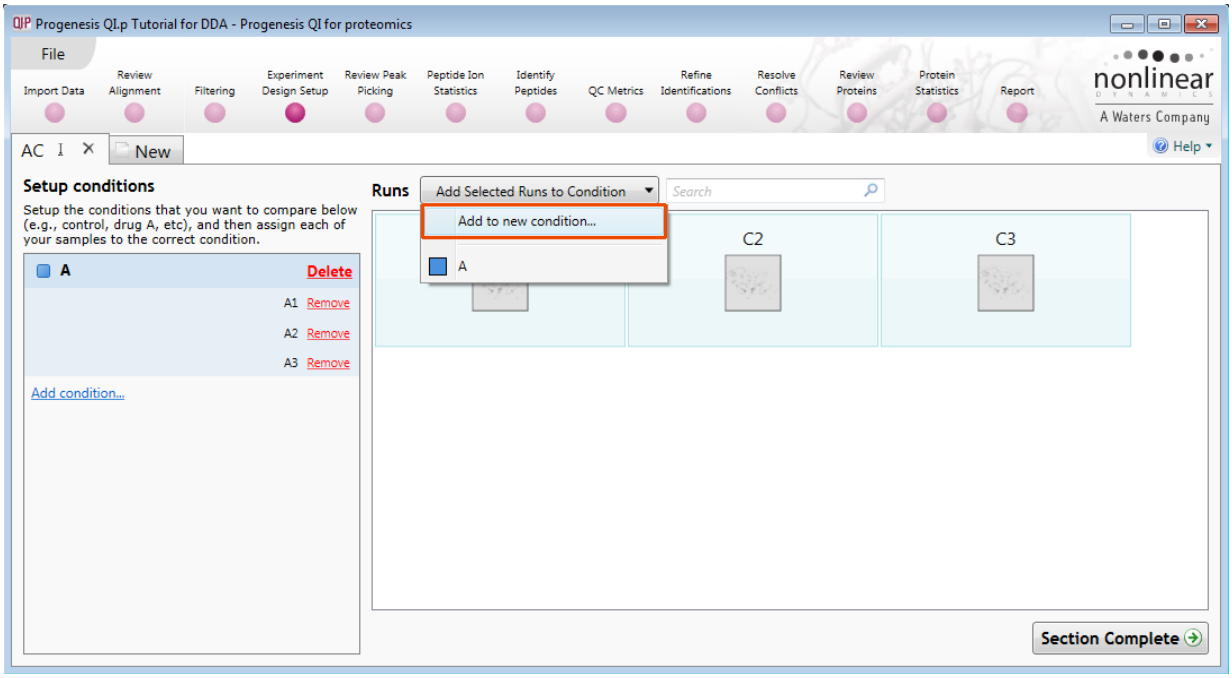
Additional information on how to apply the Within-subject Design is in Appendix 6 page 81

This experiment contains 2 conditions: A and C and uses the **Between-subject design** to group the analysed runs to reflect the Biological conditions in the original study.

To create a new **Between-subject Design** move the cursor over this option and click to open the dialog.

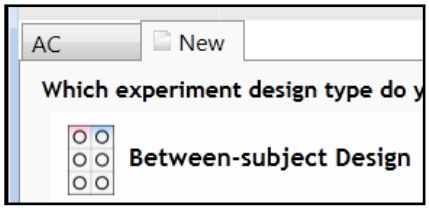


Give the new experimental design a name and then click **Create design**.



To create a new condition

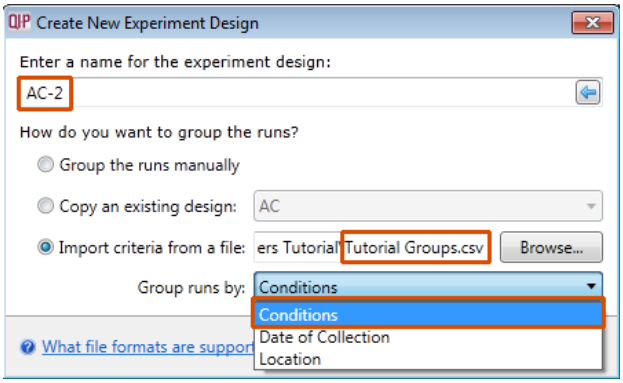
1. Select the runs for the condition by clicking on the required icon in the **Runs** panel, as shown.
2. Press the ‘black triangle’ next to the **Add Selected Runs to Condition** button on the main toolbar.
3. Select **Add to new condition**… from the drop down menu.
4. A new condition will appear in the **Conditions** panel on the left.
5. Rename the condition (e.g. C) by over typing the default name.
6. Repeat steps 1 to 5 until all the required runs are grouped into conditions.



An alternative way to handling the grouping of this set and other larger (and more complex) experimental designs is to make use of **sample tracking information** that has been stored in a spread sheet at the time of sample collection and/or preparation.

For this example there is a **Tutorial Groups.csv** file available in the Experiment Archive you restored at the beginning of this tutorial exercise.

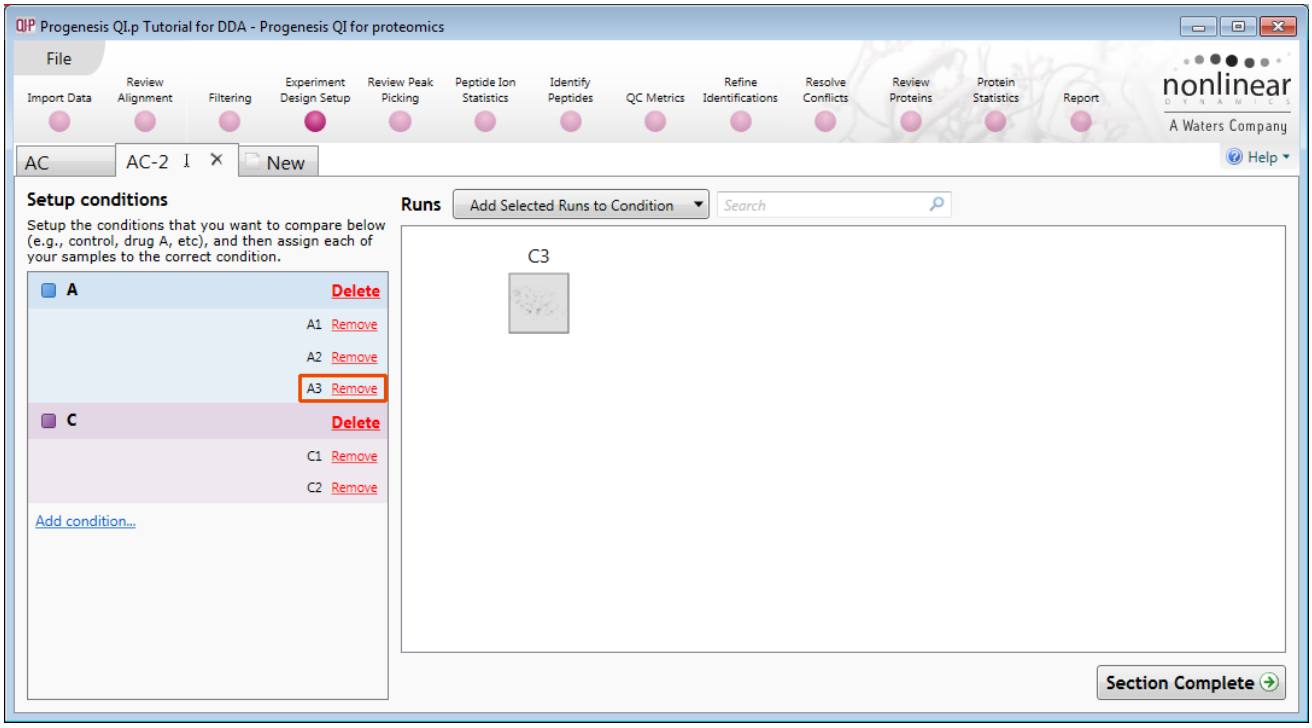
To use this approach select the **Import design from file** option from the New Experiment Design dialog. Then locate the Tutorial Groups file and select what to **Group by**, for example: **Conditions**.



When **Create design** is pressed the new tab refreshes to allow you to adjust the conditions.

Note: currently this second method is the only method of creating an Experiment Design that can be applied when using the Automatic Analysis process. Additional Designs can be applied post Automatic processing.

Use **Delete** on the Conditions panel to remove conditions that are not required in this particular design.



Note: On deleting a condition the runs will reappear in the Runs window.

Note: both designs are available as separate tabs.

To move to the next stage in the workflow, **Review Peak Picking**, click **Section Complete.**

Stage 7: Review Peak Picking and editing of results

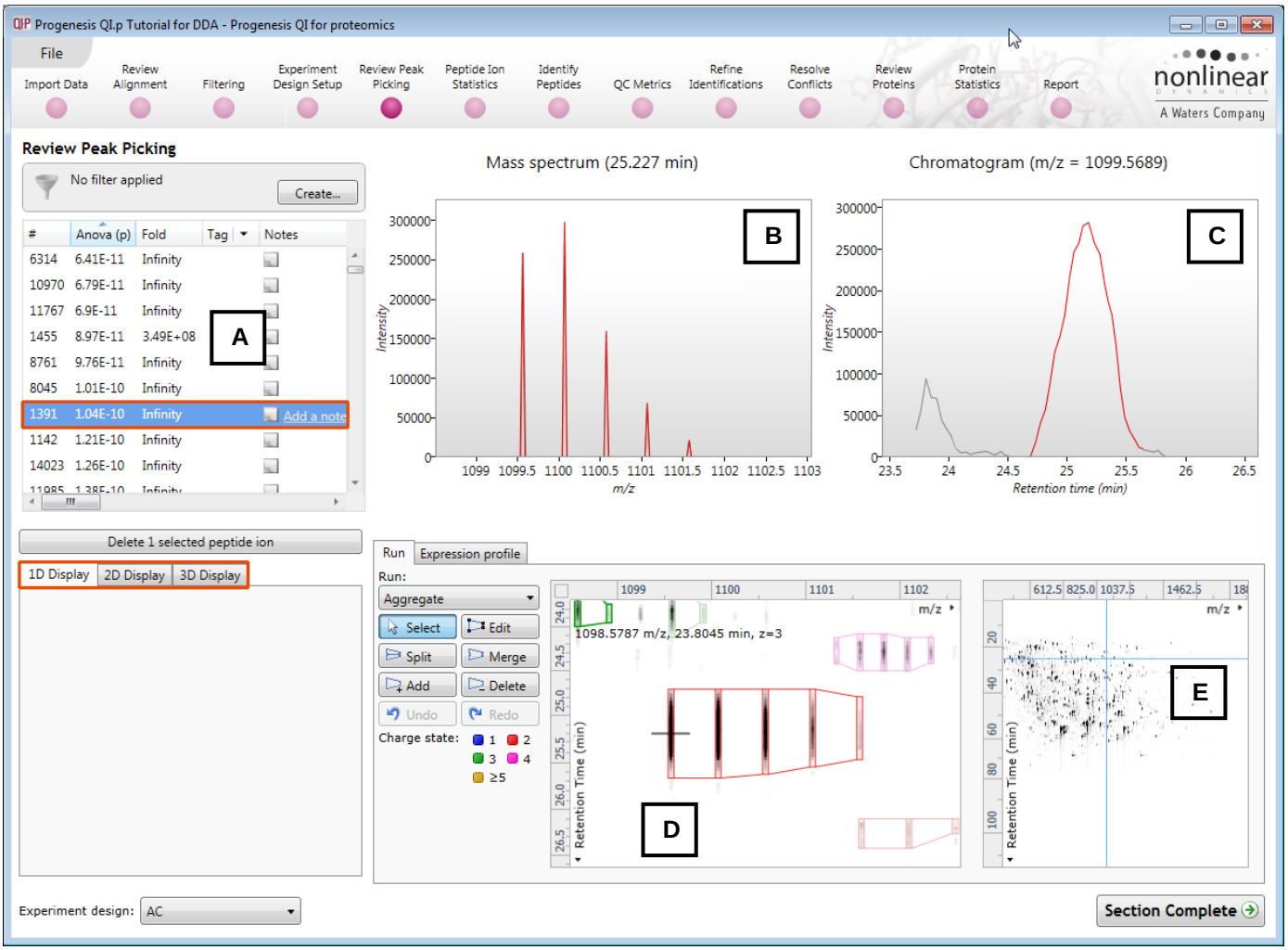
The purpose of this stage in the Workflow is to review the list of peptide ions using the visual tools provided and edit peptide ions if required.

The review stage has 4 display modes: 1D, 2D, 3D and Peptide ion Details controlled by the tabs on the bottom left of the display and the expander bar to the right of the table. Each display has multiple views to allow comparative exploration of the detected peptide ions on the aligned LC-MS runs.

Exploring analysed data using the Data displays

Window A: shows the list of peptide ions ranked by the p value for the one way **Anova** using the current grouping.

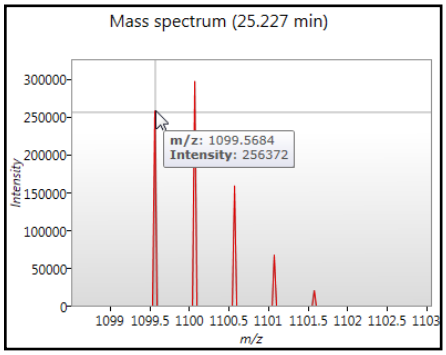
Note: a value of 'Infinity' in the **Fold** column indicates 'Presence/Absence'



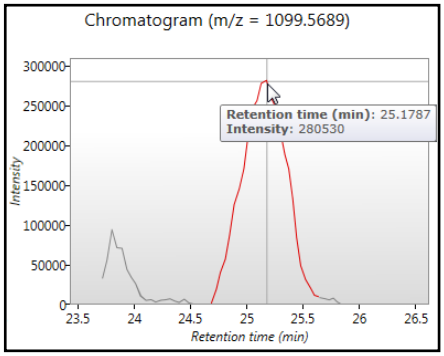
Note: by default all the peptide ions are included in the selection for the next section of the analysis.

To highlight a group of peptide ions drag out a selection on the table.

The 1D Display

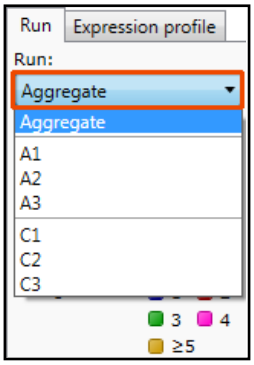


Window B: displays the Mass spectrum for the current peptide ion on the selected Run (in window D). Hold the cursor over the peak to indicate m/z and intensity



Window C: displays the Chromatogram for the current peptide ion on the selected Run (in window D). Hold the cursor over the peak to indicate Retention time and intensity

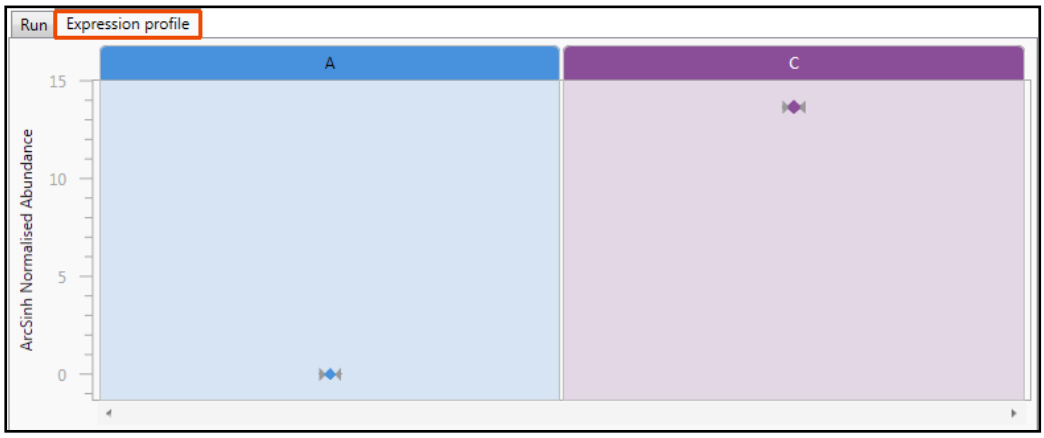
Window D: displays the details of the currently selected run. By default the selected run is an Aggregate of all the aligned runs.

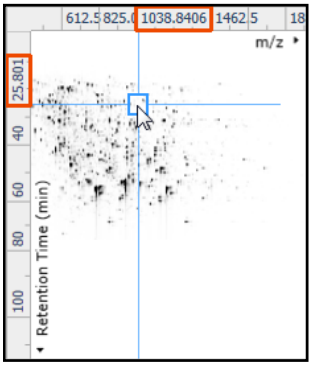


Details of individual runs can be viewed by using the 'Run' link and selecting the run you wish to view.

The peptide ion editing tools are located in this window (see page 32 for functional explanation).

Clicking on the Expression Profile tab in Window D shows the comparative behaviour of the peptide ion across the various biological groups based on group average normalised volume. The error bars show +/- 3 standard errors.



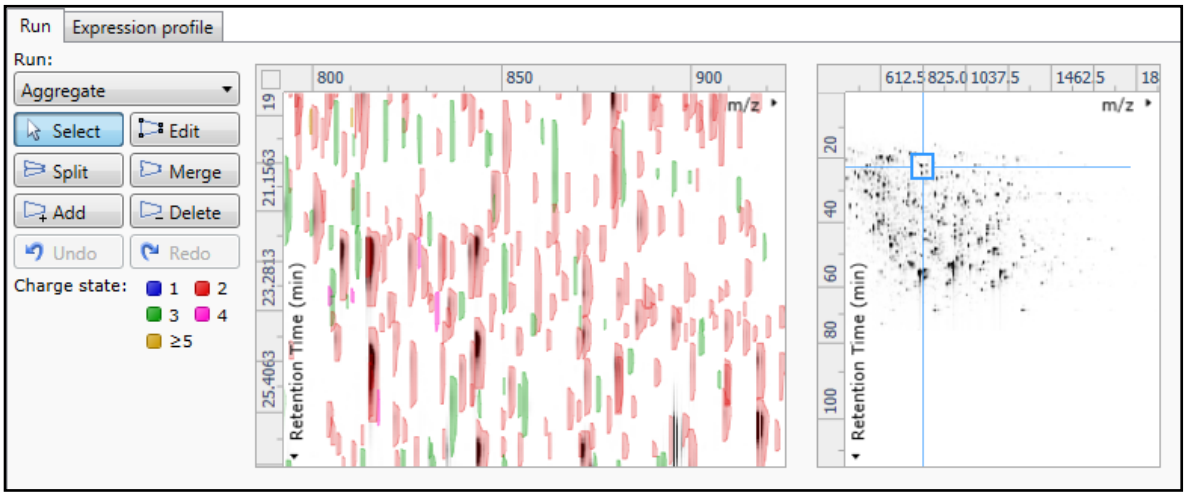


Window E: shows where the current peptide ion is located on the LC-MS run by means of the cross hairs.

To change the current location, click on the image of the run (**note**: the retention time and m/z values update as you move the cursor around this view).

Note: doing this updates the focus of all the other windows.

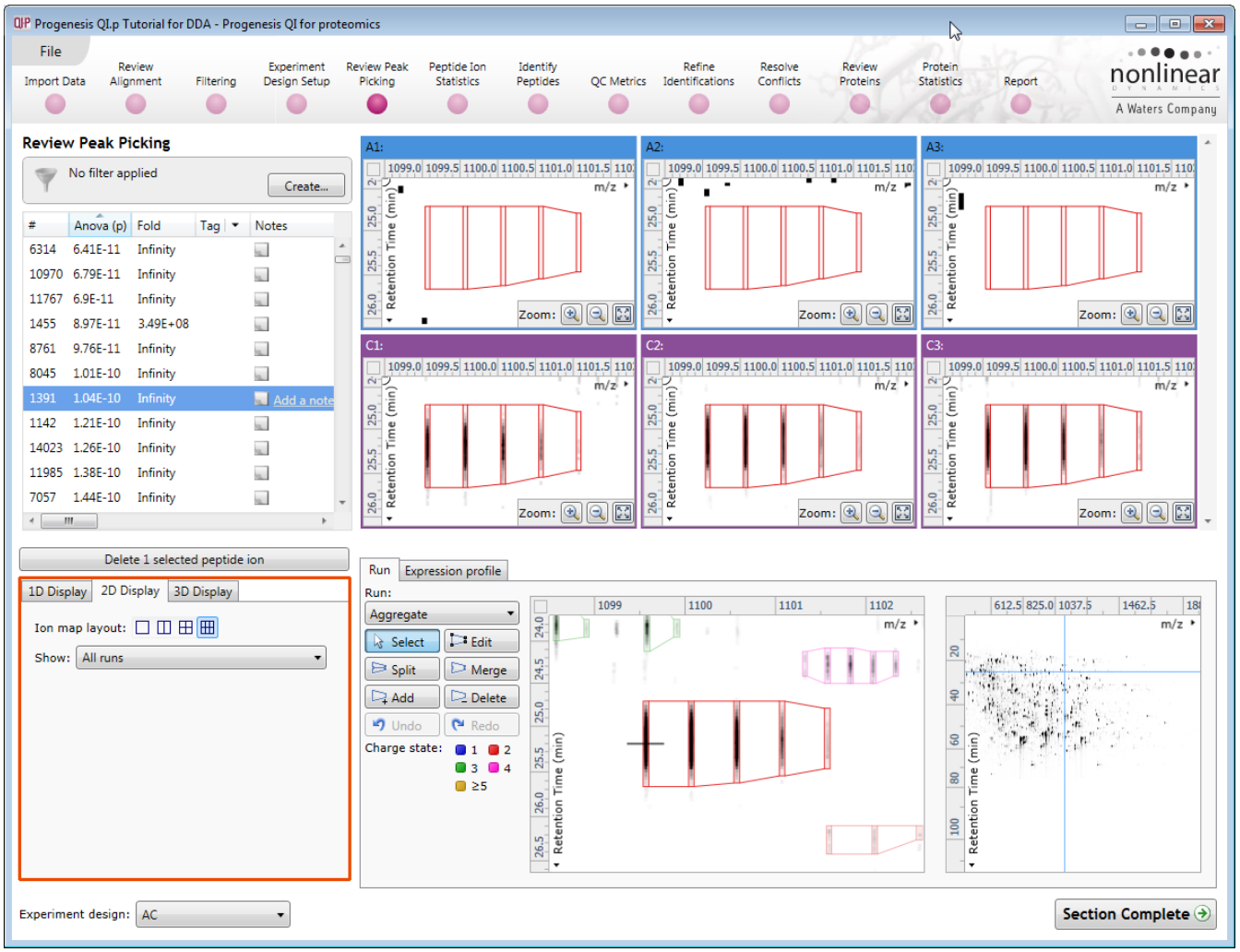
You can also drag out a larger area on this view that will refocus the other windows.



The 2D Display

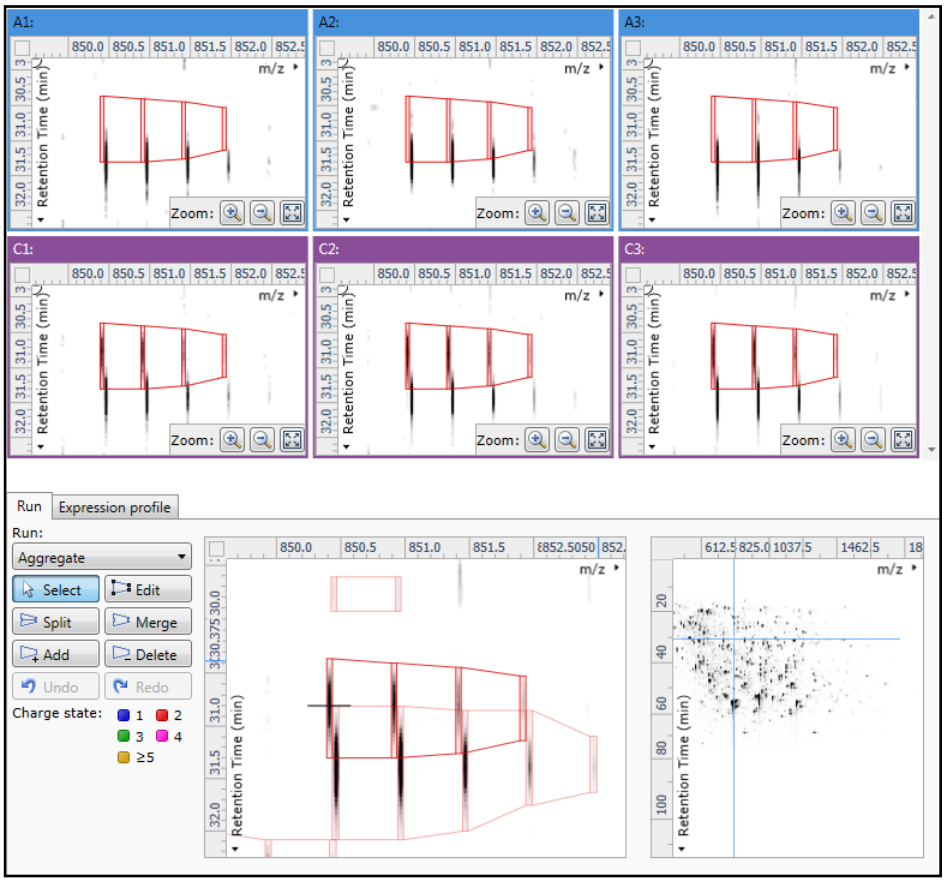
Windows **A, D and E**: perform the same functions across all 4 display modes.

In the 2D Display mode, Window B displays a montage of the current peptide ion across all the aligned LC-MS runs.



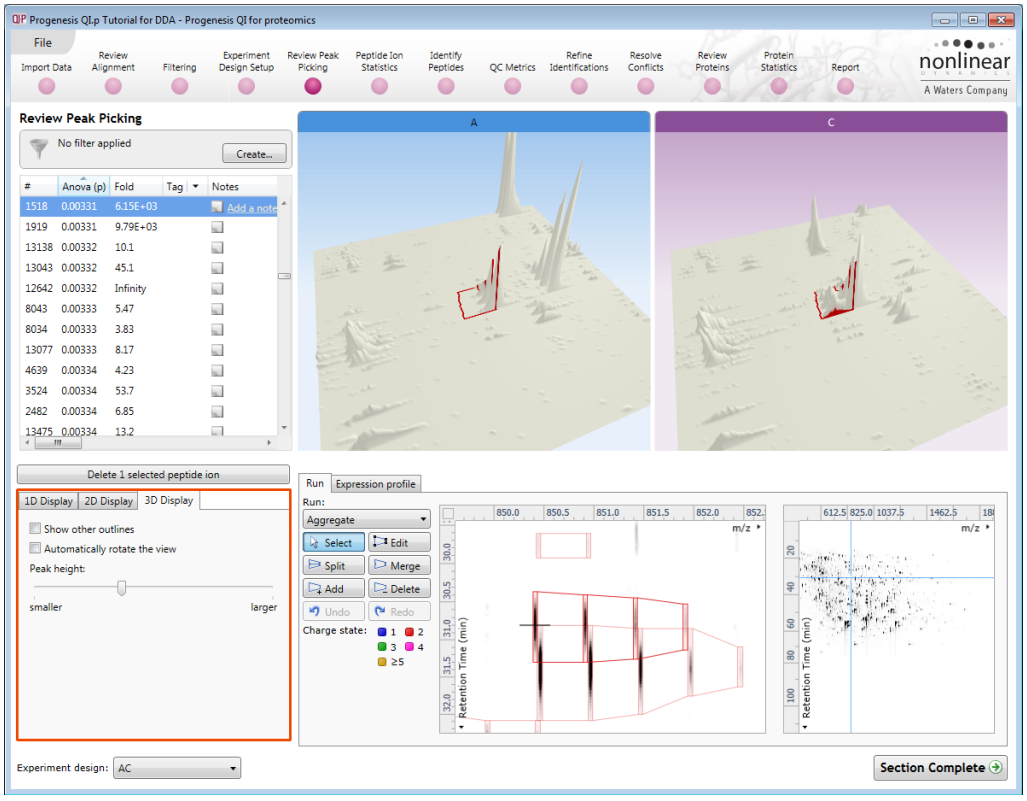
The appearance of the Montage (window B) is controlled by the panel on the bottom left of the display.

Using the the various views in the 2D display one can examine the peptide ion detection in detail to validate the correct detection of even fully overlapping peptide ions as shown above.



The 3D Display

Window B changes into a 3D view by selecting the 3D Montage tab on the bottom left of the display.

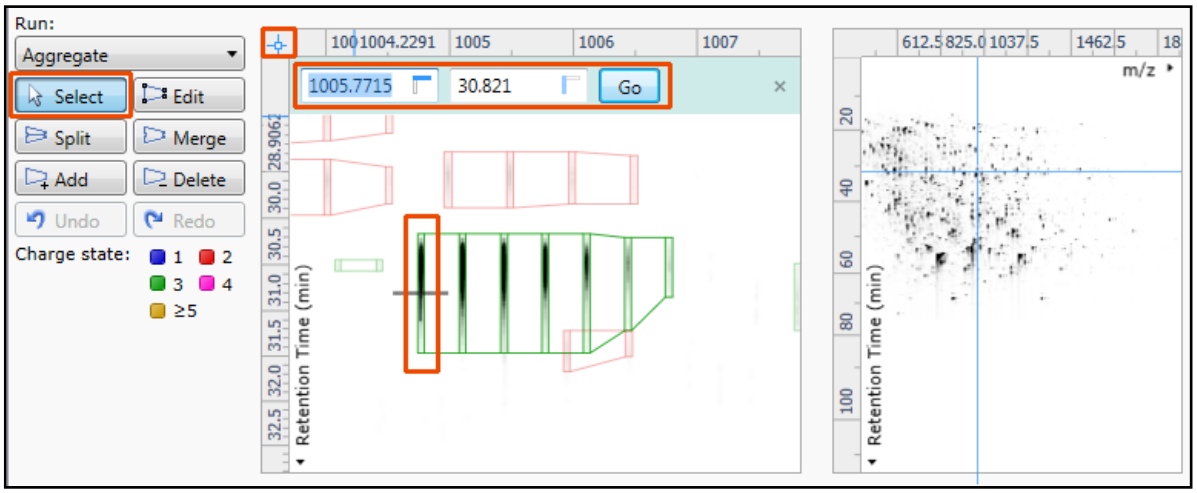


The views can be set to **Rotate** automatically or you can rotate them manually by clicking and dragging them with the mouse.

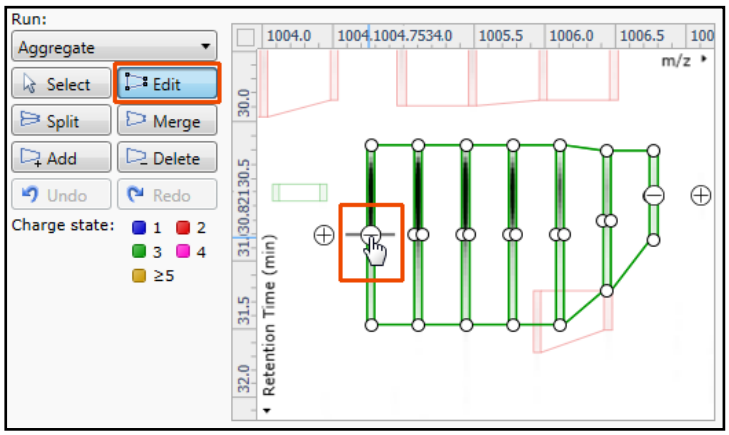
Editing of peptide ions in the View Results stage

As an example of using the editing tools which are located on the left of the **Run** view, we will remove and add back the 'monoisotopic peak' for the detected peptide ion selected below. A peptide ion can be selected from the 'Peptide ions' list or located using the various views.

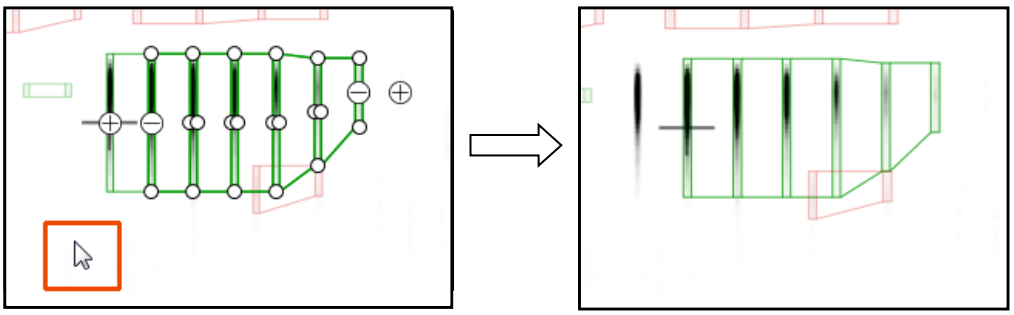
1. Locate the peptide ion at approx 1005.7 m/z and 30.8 min using the **Go To Location** tool (at the intersection of the m/z and RT axis).



2. Select the **Edit** tool and click on the peptide ion to reveal the 'edit handles'

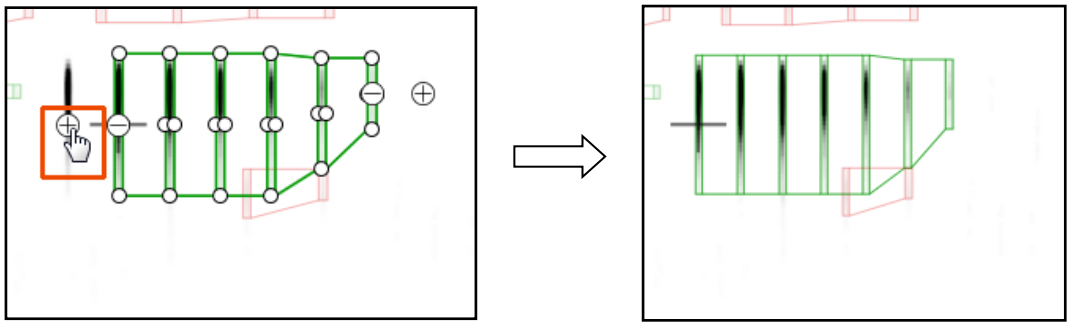


3. Click on the 'minus' handle over the monoisotopic peak to remove it.



4. Click outside the boundary of the peptide ion to update the view.

5. To add a peak to an existing peptide ion, ensure that **Edit** is selected then click inside the peptide ion to reveal the handles.

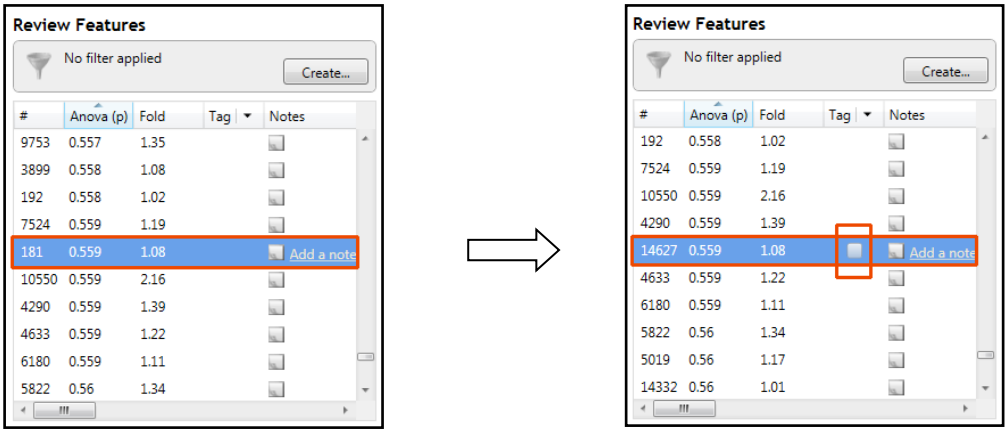


6. Click on the 'plus' handle on the peak to add it.

7. Then click outside the peptide ion to update the view.

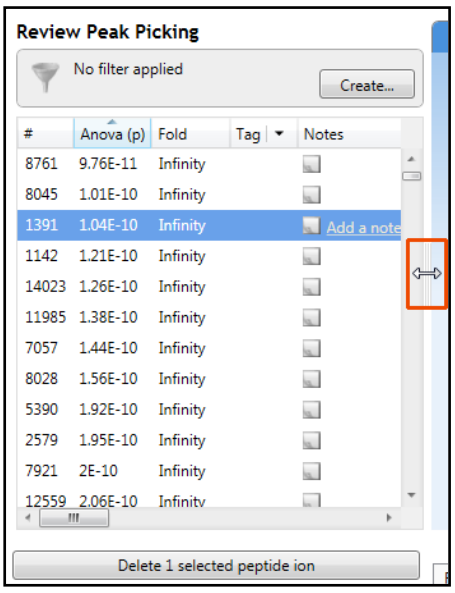
8. **Note**: If you are not satisfied with the editing use the **Undo** button and retry.

9. **Note**: that a tag is automatically added to the edited peptide ion in the table and the peptide ions id number is changed to the next available one at the end of the list.



The other tools: split, merge, add and delete behave in a similar fashion and their use can be combined to achieve the desired results.

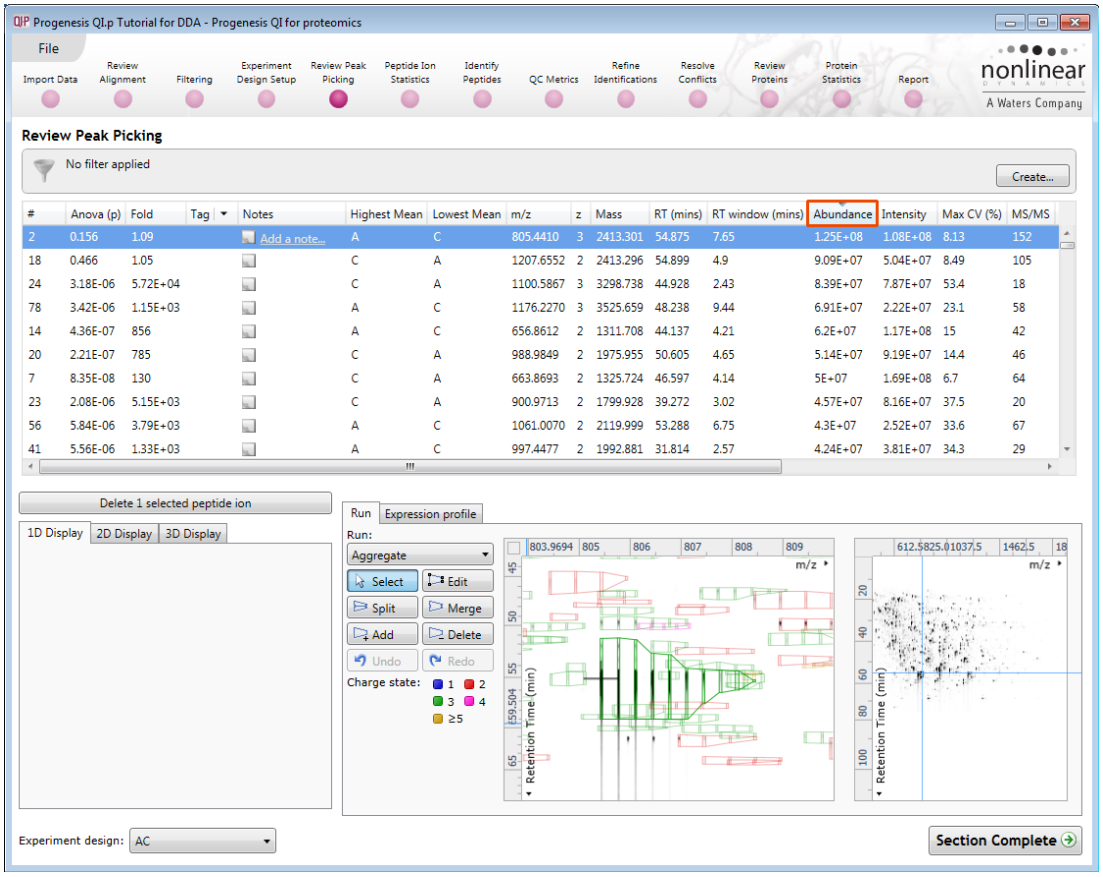
Selecting and tagging peptide ions for Peptide Ion Statistics



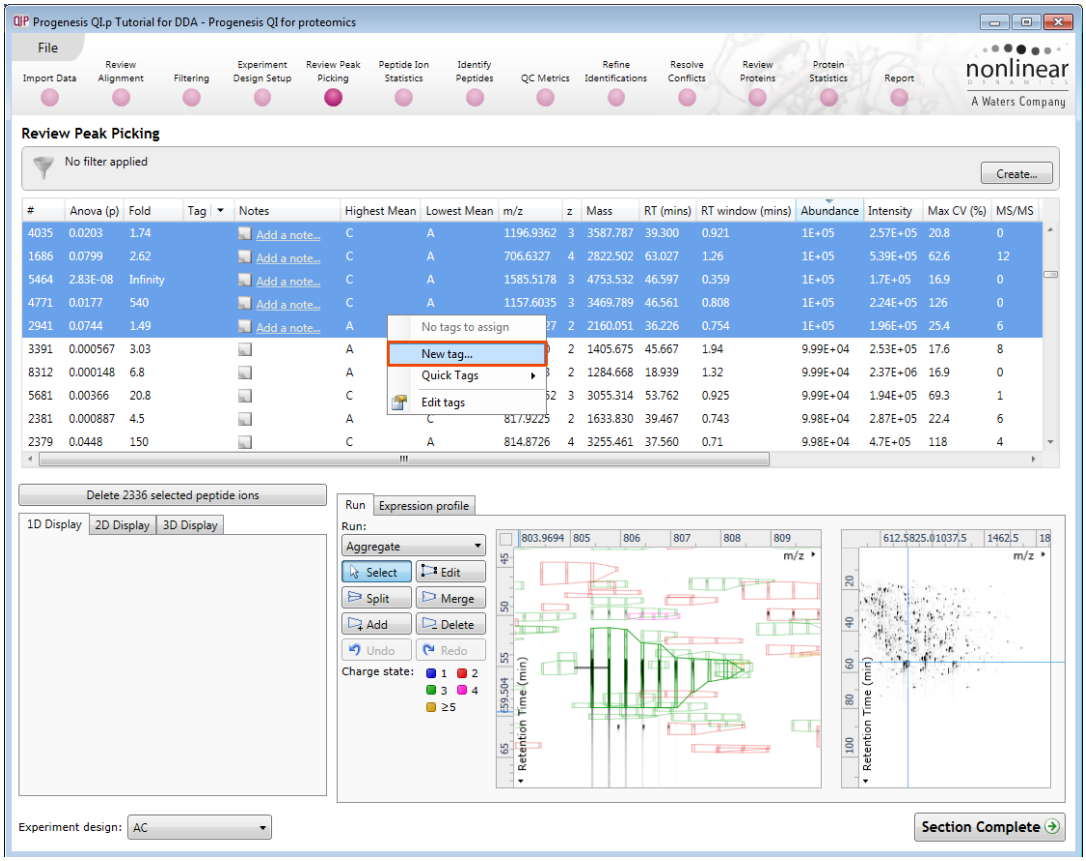
There are a number of ways to 'refine' your 'Ranked List' of analysed peptide ions before examining them with the Statistical tools in **Peptide Ion Statistics**. These make use of simple 'Selection' and 'Tagging' tools that can be applied to the various groupings created in Stage 6 (page 25). An example is described below.

First expand the 'Peptide ions' table to show all the details by clicking on the 'Splitter Control' to the right of the Review Peptide ions table.

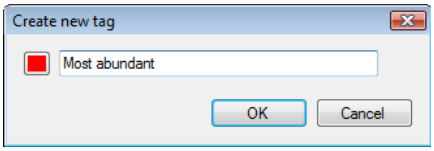
Then order on **Abundance** and select all peptide ions with an Abundance of 1E+05 and greater, (the exact number is not important).



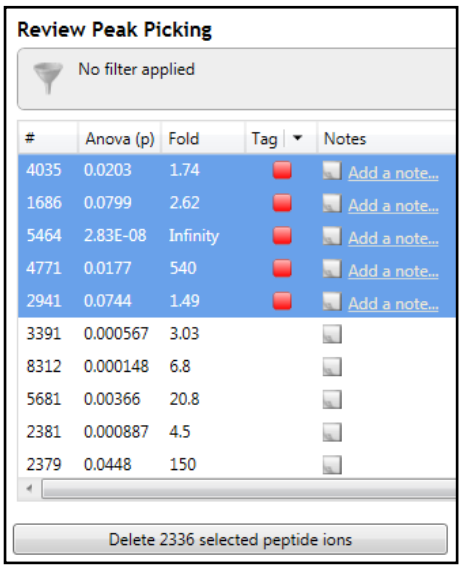
With the peptide ions still highlighted right click on them and select 'New Tag'.



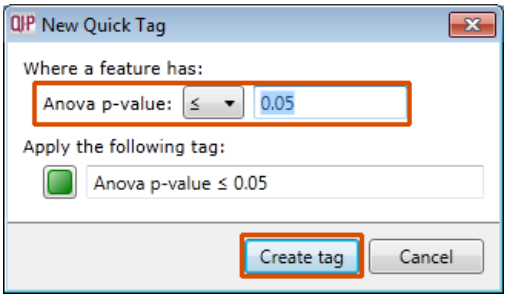
Give the Tag a name. i.e. **'Most abundant'**.

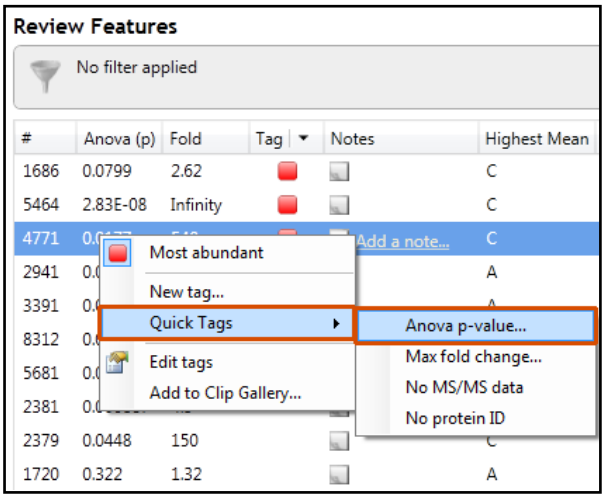


On clicking **OK** the Tag is added to the peptide ions highlighted in the table (signified by a coloured square).

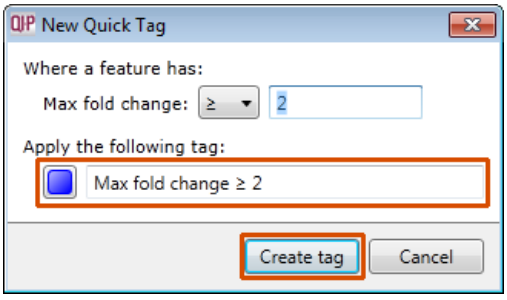


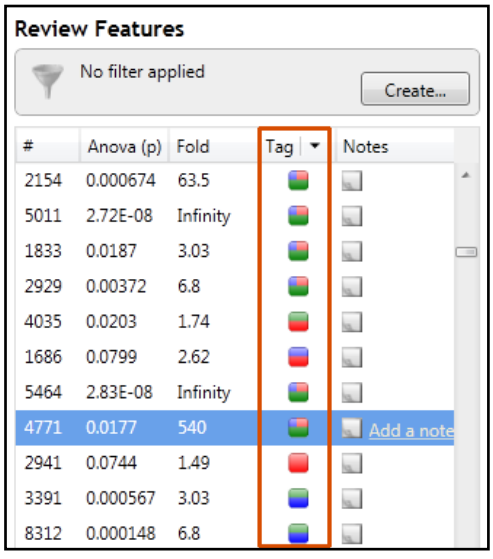
Now right click on any peptide ion in the table and select **Quick Tags** this will offer you a number of standard tag options. Select **Anova p-value....** Then set the threshold as required and adjust default name as required and click **Create Tag**.





Once this tag appears against peptide ions in the table right click on the table again and create another Quick Tag, this time for peptide ions with a **Max fold change ≥ 2**



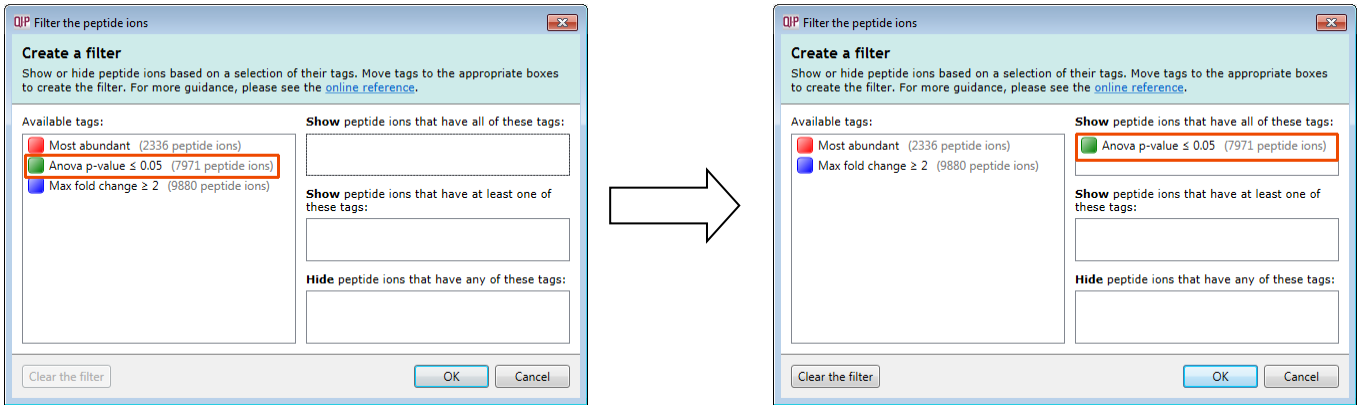


The table now displays peptide ions with multiple tags.

The tags can be used to quickly focus the table on those peptide ions that display similar properties.

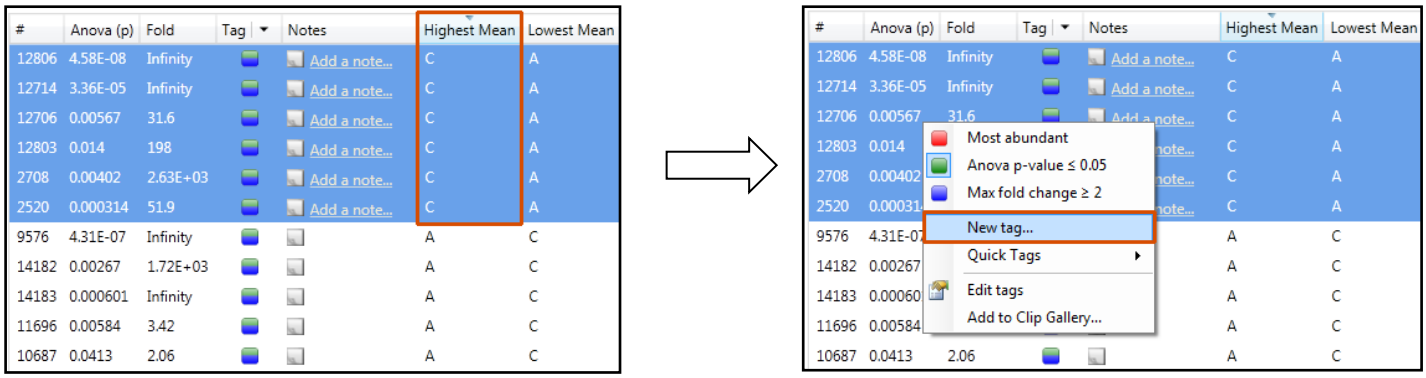
For example: to focus the table on displaying those peptide ions that have an **Anova p-value ≤ 0.05** click on **Create** on the filter panel above the table.

Drag the tag on to the panel **Show peptide ions that have all of these tags** and press OK.



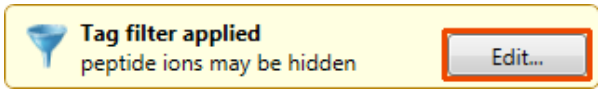
Now order the current peptide ions in the table by the **Highest mean** so that all the peptide ions showing the highest mean for **Condition C** are at the top of the list.

Then highlight all the peptide ions with the highest mean for **condition C** and create a new Tag for them.

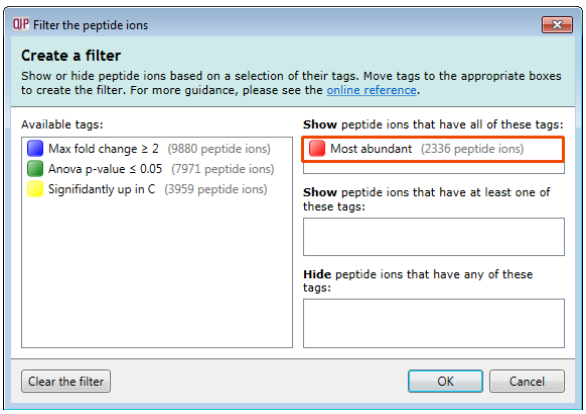


Create a tag for them called **Significantly up in C,** tagging 3959 peptide ions.

Finally view the tags you have just created by clicking on Edit in the Tag filter panel, above the table.



Make sure that only the tag for the **Most abundant** peptide ions is shown and press **OK.**



To move to the next stage in the workflow, Peptide Statistics, click **Section Complete.**

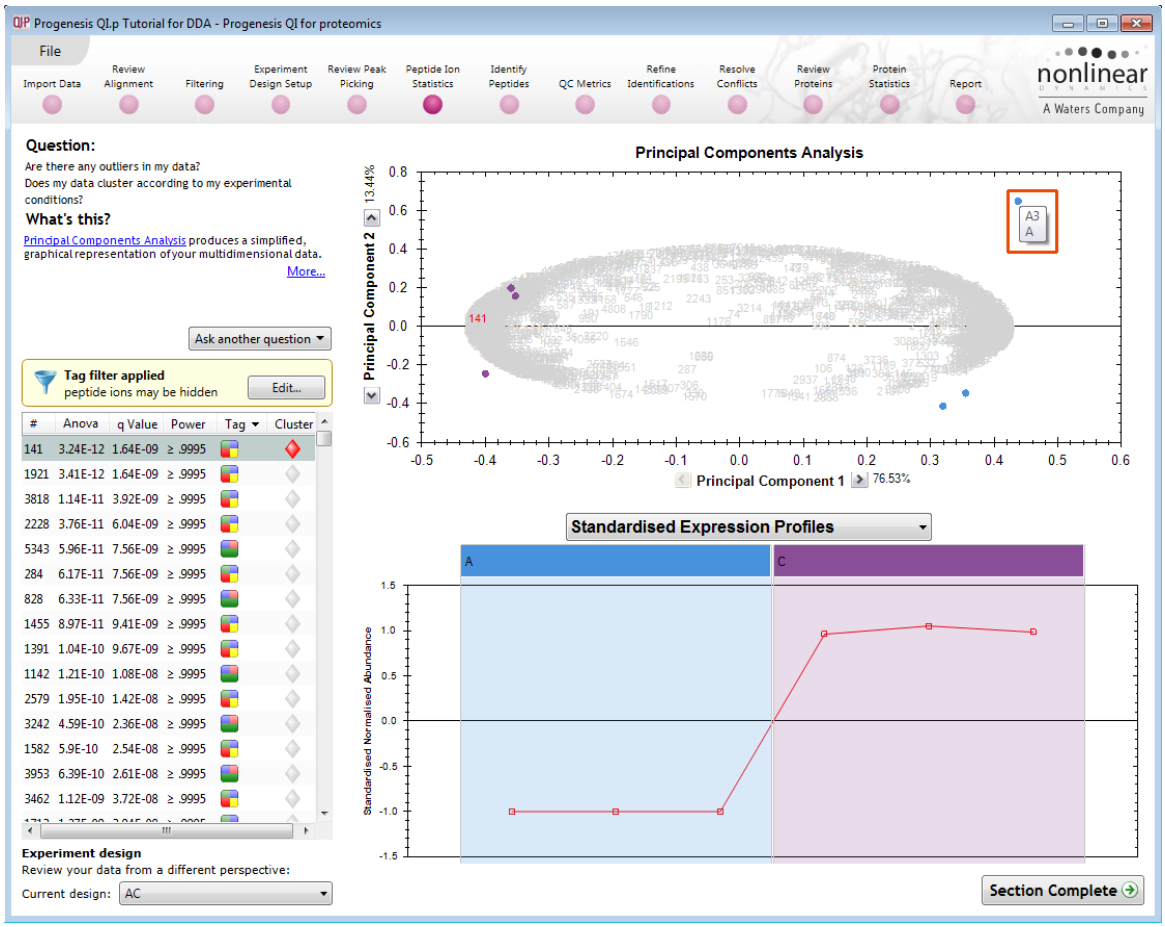
Stage 8: Peptide Ion Statistics on selected peptide ions

The user guide now describes the functionality of the Multivariate Statistics.

Peptide Statistics opens calculating the Principal Components Analysis (PCA) for the active 'tag' in this case the **Most abundant** peptide ions.

As an example we will start by examining the behaviour of the **Most abundant** peptide ions from the previous stage, **Review Peak Picking**.

The statistical analysis of the selected data is presented to you in the form of interactive graphical representation of answers to questions asked of the analysed data.



Note: the LC-MS runs (samples) are displayed as solid coloured circles on the plot. To identify the runs, a tooltip is displayed when the cursor is held over each circle.

Principal Component Analysis (PCA)

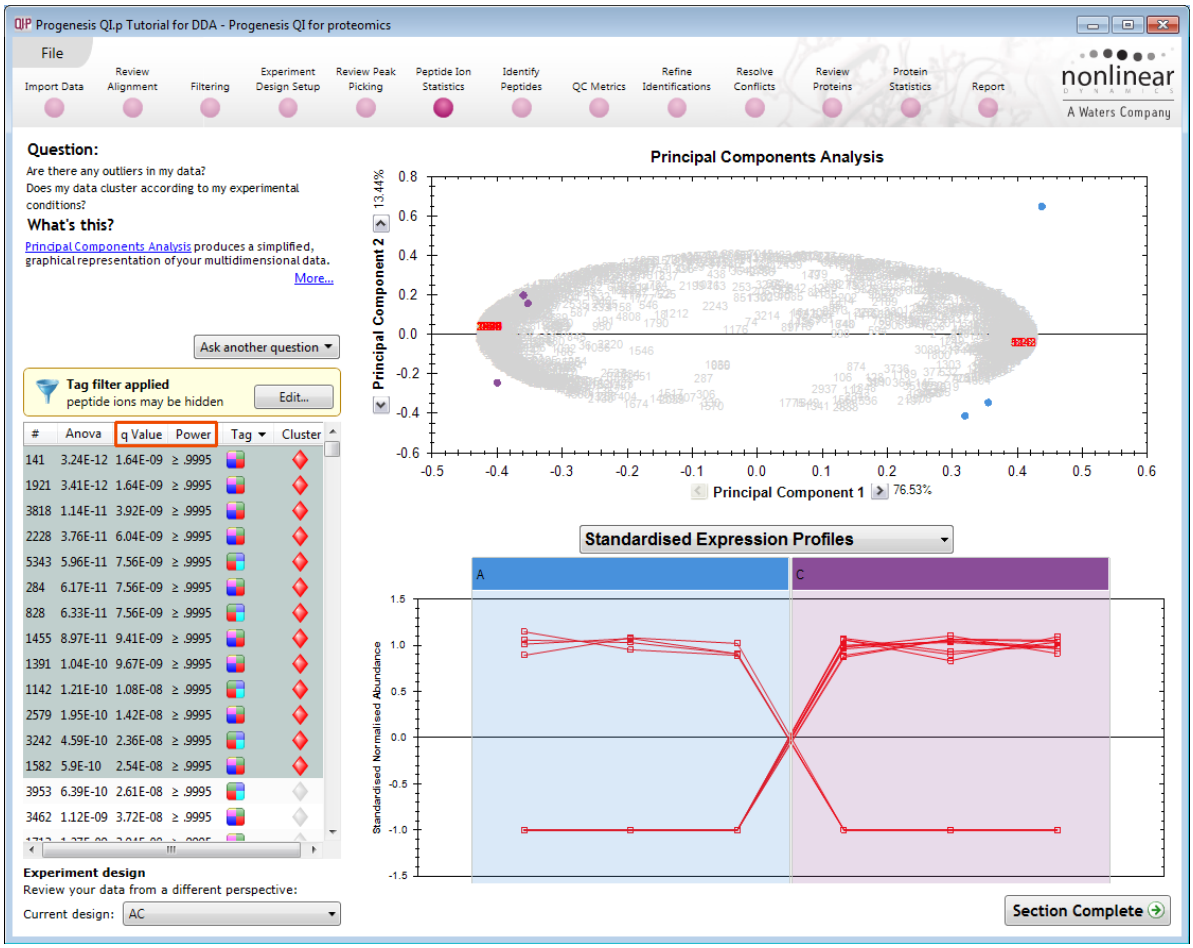
In **Peptide Stats** the first statistically based question asked of the data takes the form of a Quality Control assessment:

Are there any outliers in my data? And does my data cluster according to my experimental conditions?

It answers this question by:

'Using Principal Components Analysis (PCA) to produce a simplified graphical representation of your multidimensional data'.

PCA can be used to determine whether there are any outliers in the data and also look at how well the samples group. The groupings that can be observed on the 2D PCA plot can be compared to your experimental conditions and conclusions can be drawn regarding possible outliers in your data. Selecting peptide ions in the table will highlight the peptide ions on the 'Biplot' and their expression profiles will appear in the lower panel.



Note: the Table in the Stats view contains additional columns:

q value: tells us the expected proportion of false positives if that peptide ion's p-value is chosen as the significance threshold.

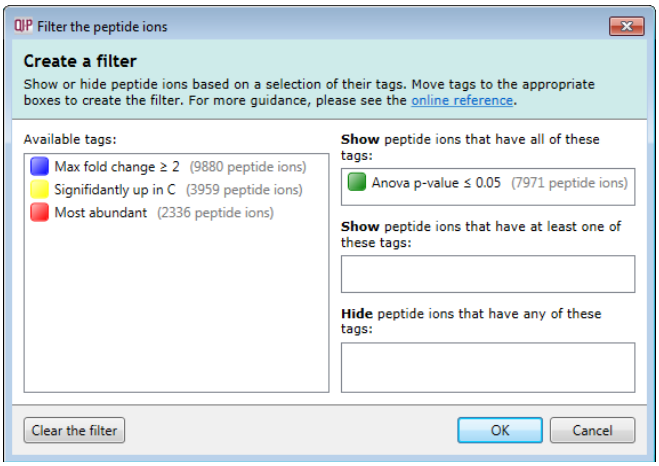
Power: can be defined as the probability of finding a real difference if it exists. 80% or 0.8 is considered an acceptable value for power. The Power Analysis is performed independently for each peptide ion, using the expression variance, sample size and difference between the means.

Also, for a given power of 80% we can determine how many samples are required to ensure we find a difference if it actually exists.

Note: Power analysis is discussed in Appendix 7 (page 83)

Correlation Analysis

Use the tags, created in Review Peak Picking, to filter the peptide ions in the table.

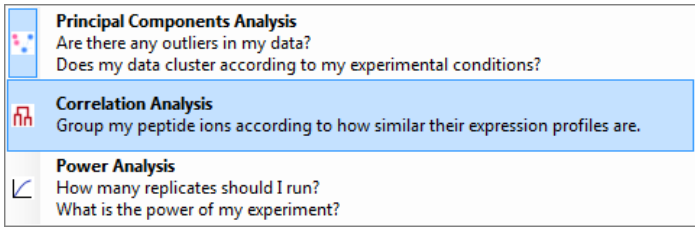


We are going to explore the Correlation Analysis for all the peptide ions that were tagged at the view results stage with an **Anova p-value≤0.05**.

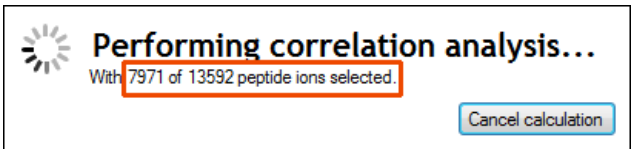
On pressing OK the PCA will recalculate using these peptide ions, you can (to save time) stop this calculation by pressing **Cancel calculation** and then set up Correlation Analysis for the 7971 peptide ions.

To set up the **Correlation Analysis** using this filtered data set click on **Ask another question** (above the table)

A selection of 3 tools will appear in the form of questions.



Select the second option to explore ‘peptide ion correlation based on similarity of expression profiles’



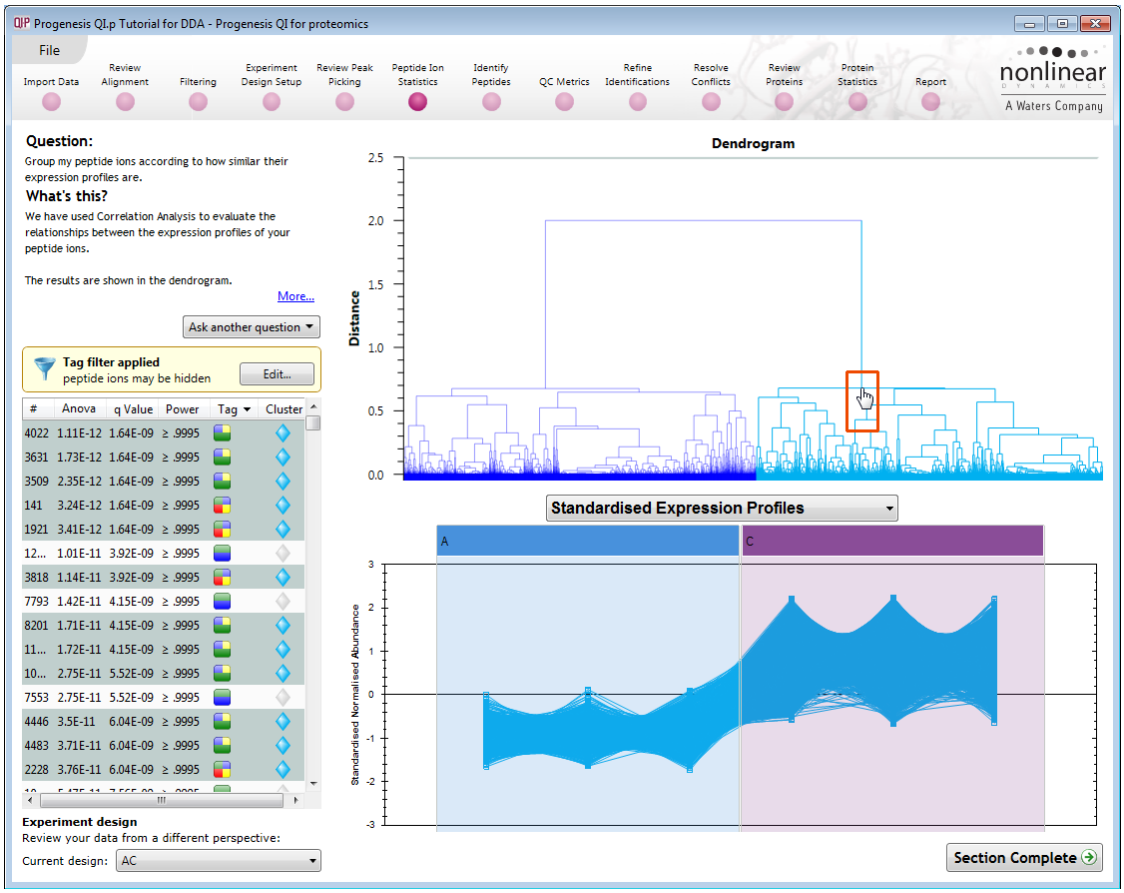
This time the statistically based question(s) being asked is:

‘Group my (selected) peptide ions according to how similar their expression profiles are’

The question is answered by:

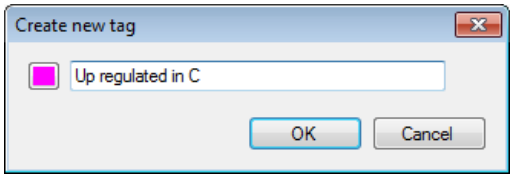
‘Using Correlation analysis to evaluate the relationships between the (selected) peptide ions’ expression profiles’.

The answer is displayed graphically in the form of an interactive dendrogram where the vertical distance, between each peptide ion, can be taken as indicative of how similar the expression profiles of each cluster of peptide ions are to each other.

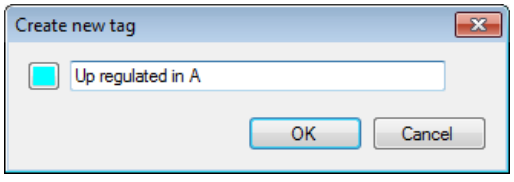


Correlation Analysis enables the grouping of peptide ions together according to how similar their expression profiles are.

To highlight all the peptide ions demonstrating **Increased expression in the C** group click on a 'node' for a branch of the Dendrogram (as shown above). As before, right click on the highlighted peptide ions in the table and create a Tag for these peptide ions (Up regulated in C).

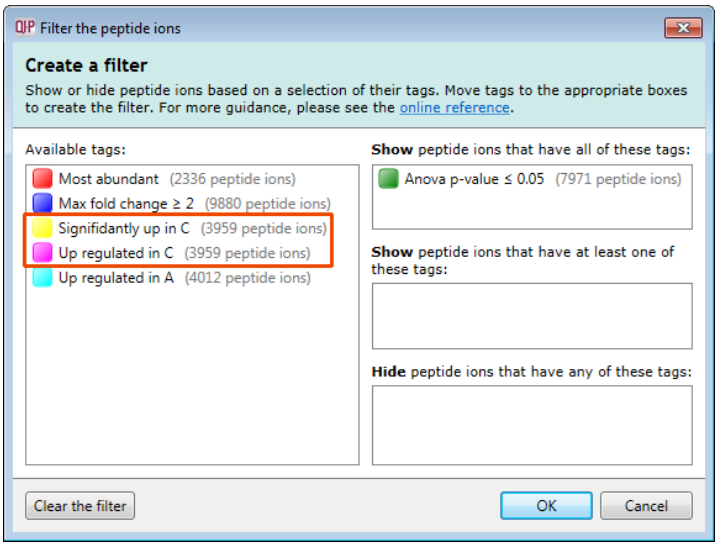


Also create a tag for those peptide ions showing **Increased expression in A** by first clicking on the other 'main' node then right click on the highlighted peptide ions in the table and creating the New tag (Up regulated in A).



Comment: When you review the tags using **Edit** you can see that the Magenta and Yellow tags have been assigned to the same number of peptide ions. This shows how tabulated information about peptide ions can be used alongside interactive graphical plots of multivariate statistical analysis to explore your data.

Note: two groups is a special case, for more groups this will not be the case, and additional tagging will be required.

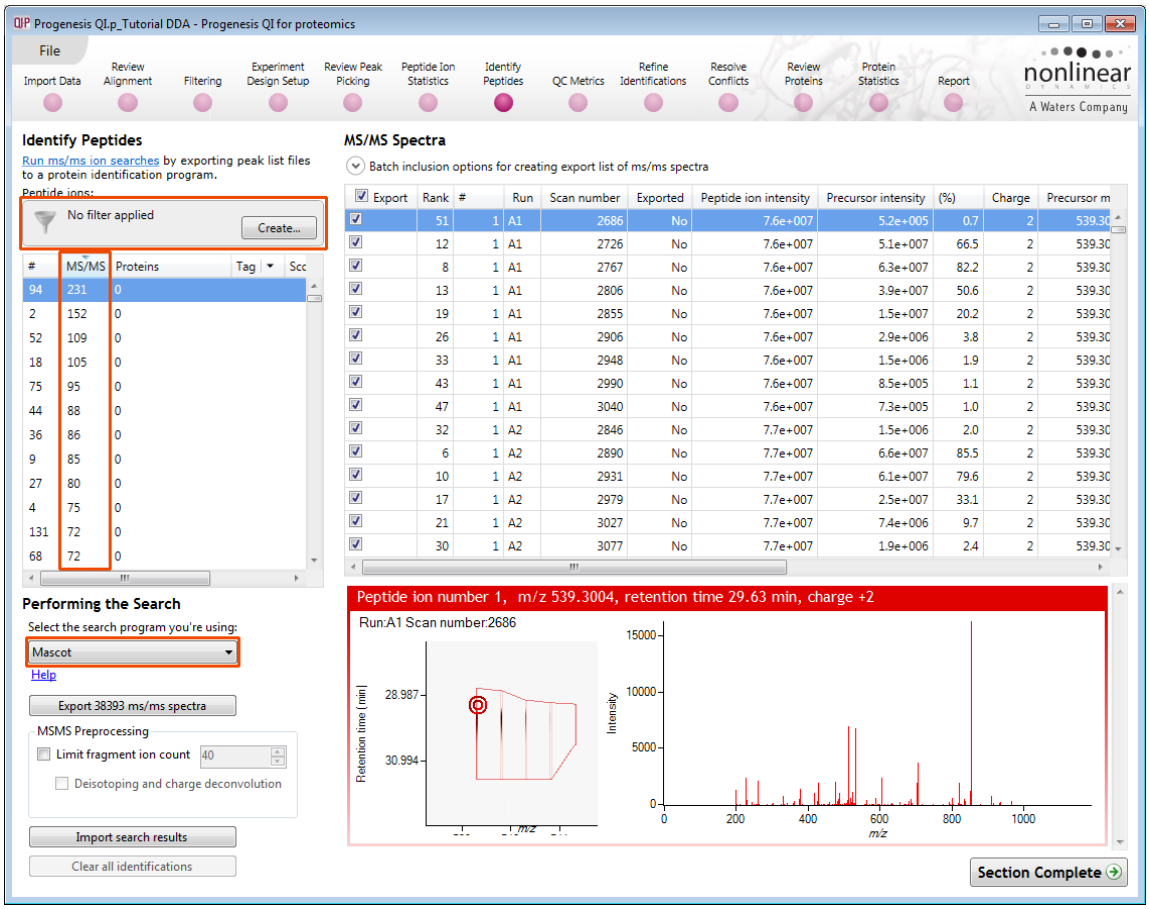


Tip: when reviewing the tags (see above) if you are not applying a new filter then use the **Cancel** button to return to the main view, this prevents unnecessary recalculation of your data.

To move to the next stage in the workflow, **Identify Peptides**, click **Section Complete.**

Stage 9: Identify peptides

Progenesis QI for proteomics does not perform peptide identifications itself for DDA data. Instead it supports identifications by allowing you to export MS/MS peak lists in formats which can be used to perform peptide searches by various search engines. The resulting identifications can then be imported back into Progenesis QI for proteomics, using a number of different file types, and matched to your detected peptide ions.



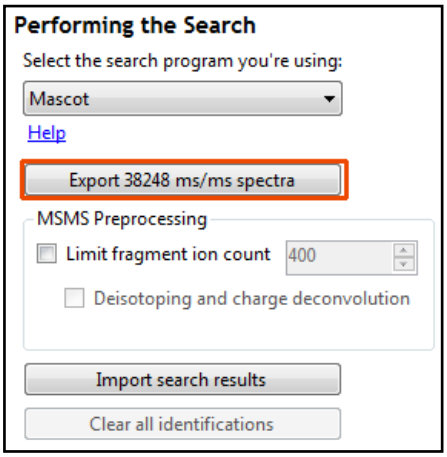
Note: ensure that **NO** tag filters are applied and that Mascot is the selected search engine.

Determining protein identification is dependent on the availability of MS/MS data for the LC-MS runs. This data may be available but limited if the LC-MS was performed in a data dependant MS/MS detection mode due to under sampling. Under these conditions MS/MS data acquisition is dependent on thresholds and parameters set prior to the acquisition of the LC-MS run.

For this example we are using LC-MS runs containing MS/MS data where the data was acquired in a data dependant mode (DDA).

The Peptide Search page shows the number MS/MS that have been matched to each peptide ion in the Peptide ion list (see above). MS/MS scans are matched to a peptide ion if their precursor *m/z* and aligned retention time fall within the area of one of the isotopes of the peptide ion. The MS/MS scans which are matched to the displayed peptide ions are shown in the MS/MS spectra list on the right.

The first step is to decide which MS/MS scans you wish to export to be identified. By default this is all the available spectra for the peptide ions displayed in the Peptide ions list. This number is visible on the Export button.



The set can be targeted using the tags and also refined with respect to quantity and quality of the spectra being sent to the search engine.

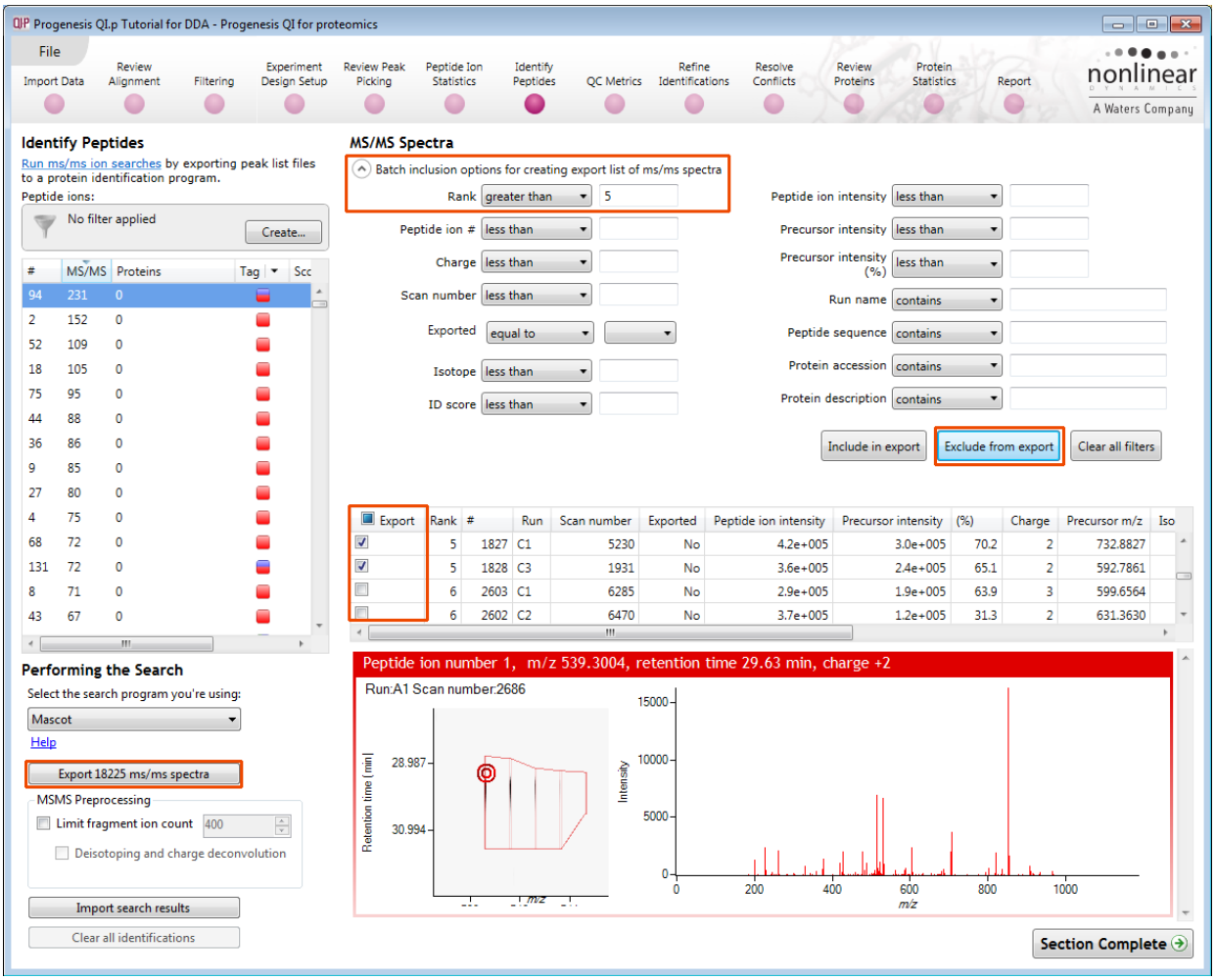
Note: by default the table is ordered on the number of MS/MS spectra available for each peptide ion.

The total number of spectra included in this set is **38248** as shown on the Export button.

Before exporting the spectra, the set can be further refined.

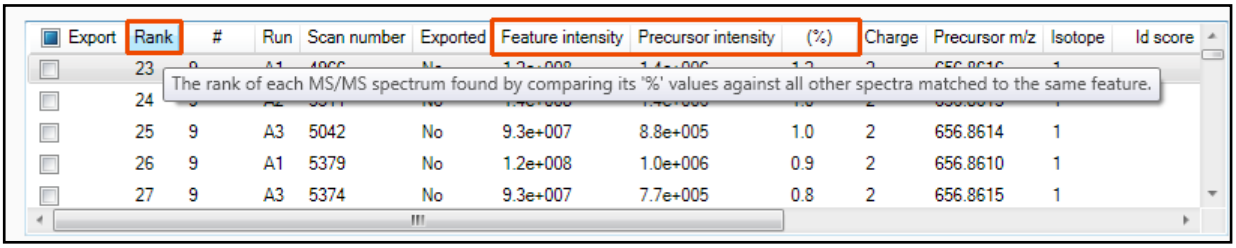
Note: many of the abundant peptide ions have a large number of spectra associated with them.

To control the number of spectra for each peptide ion, expand the **Batch inclusion options**.



For example: We will make use of the 'Rank' value to reduce the number of Spectra being used for each peptide ion to a maximum of 5.

The 'Rank' of each MS/MS spectra is determined by comparing its **% value against all other spectra matched to the same peptide ion.**

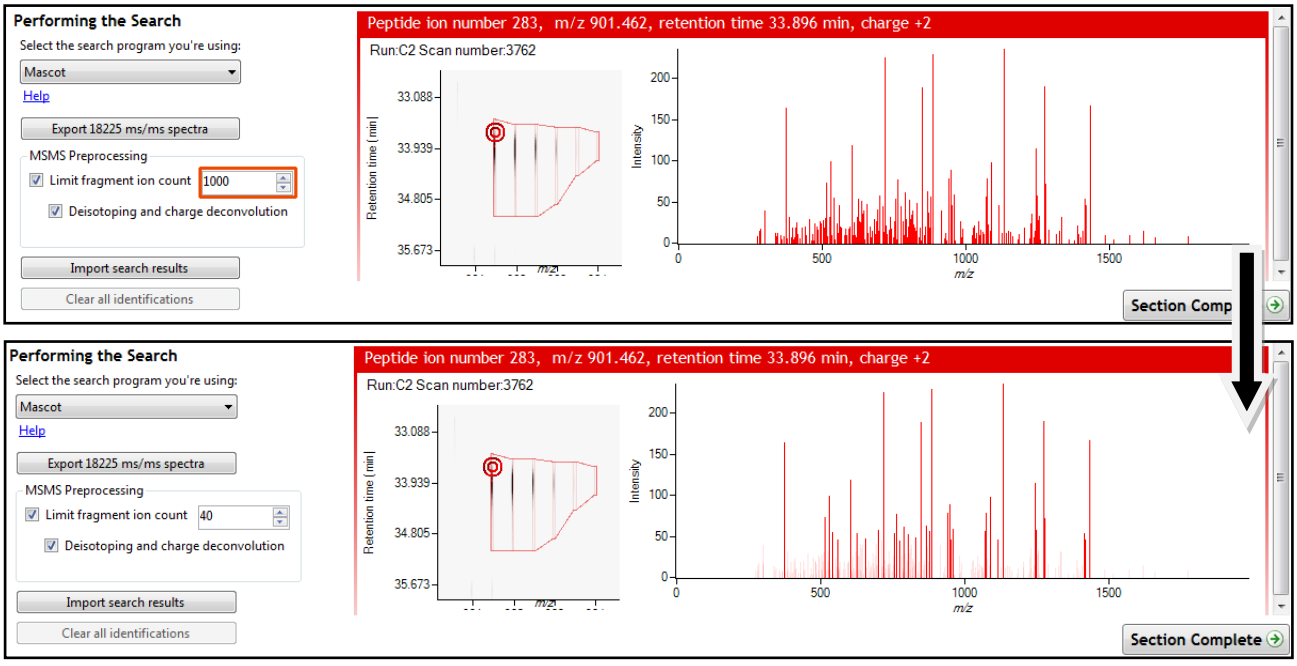


Note: the **% value** for each spectra is the **Precursor intensity as a percentage of the Peptide ion intensity**

Set the Rank filter to 'greater than' 5 and click **Exclude from export** this reduces the number to spectra to export to **18225**

Limiting the 'fragment ion count' (FIC) for the spectra being exported can improve the quality of the spectral data being used in the search by removing noisy peaks.

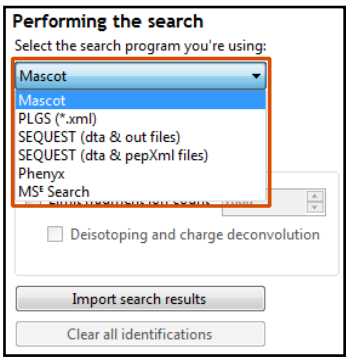
For example for the current spectra, reduce the FIC from 1000 to 40.



Note: the effect this has on the number of peaks in the spectra. This 'limitation' is applied to all the spectra being exported; hence the export file size will be reduced.

For this example we will **not** limit the fragment count, so leave it un-ticked (the default setting).

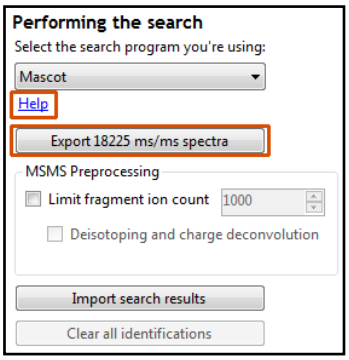
Performing an MS/MS Ion Search



Having chosen 18225 spectra to export, as described above:

1. Select appropriate search engine i.e. Mascot
2. Click 'Export current query set' to save search as file
3. Perform search on appropriate search engine and save results file
4. Click 'Import search results', locate results file and open

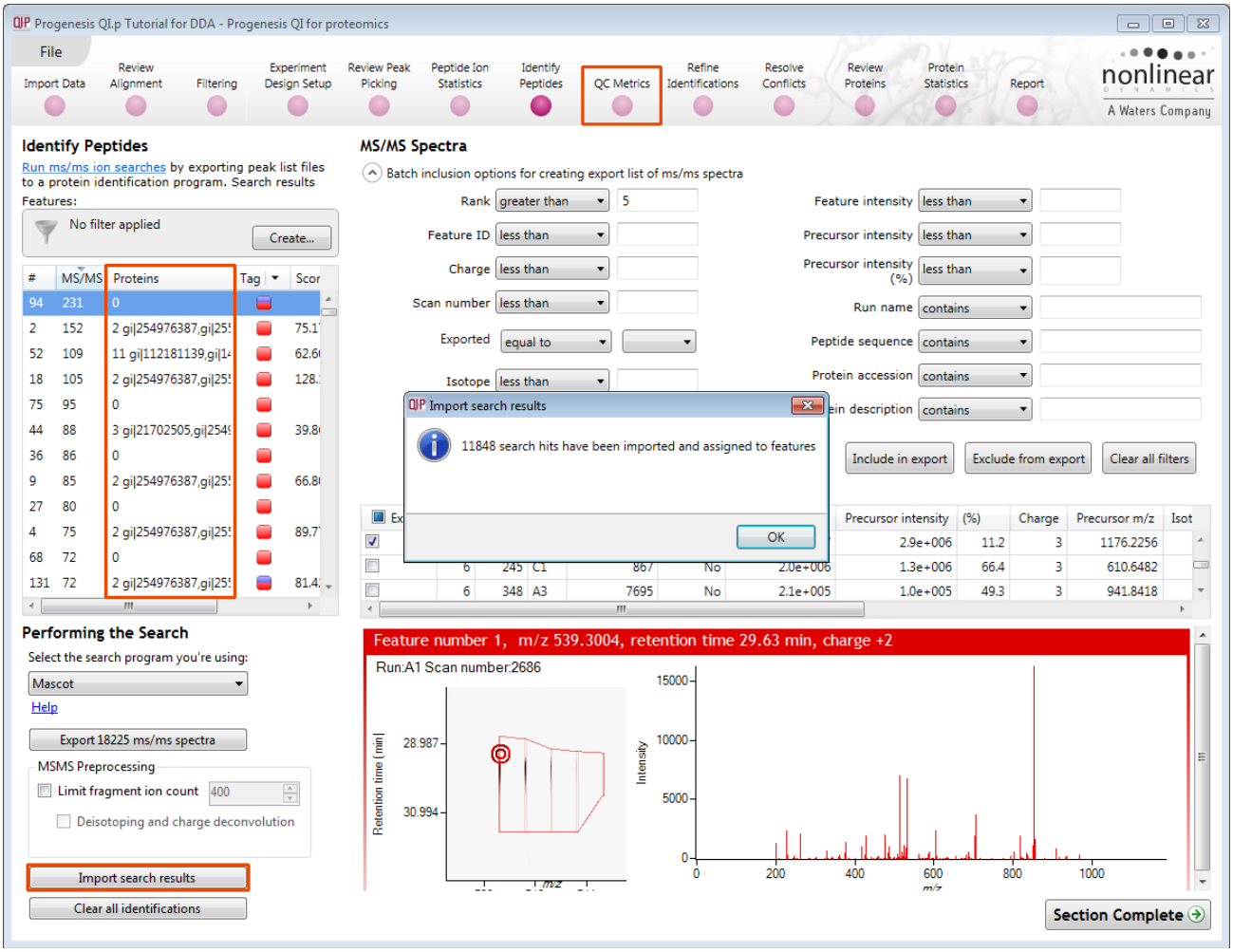
Please refer to Appendix 9 (page 86) for details of the 'Search Engine' parameters used in this example



Note: the blue link provides you with details on the appropriate formats for exporting search results and access to additional formats

Note: an example Search Results file, from a MS/MS Ion search, is available in the folder you restored the Archive to (Search Results.xml). Select the 'Mascot' method and import this file to see results like those below.

On importing the Search results the Peptide ions table updates to reflect the identified proteins and the relevant score for each searched peptide ion.

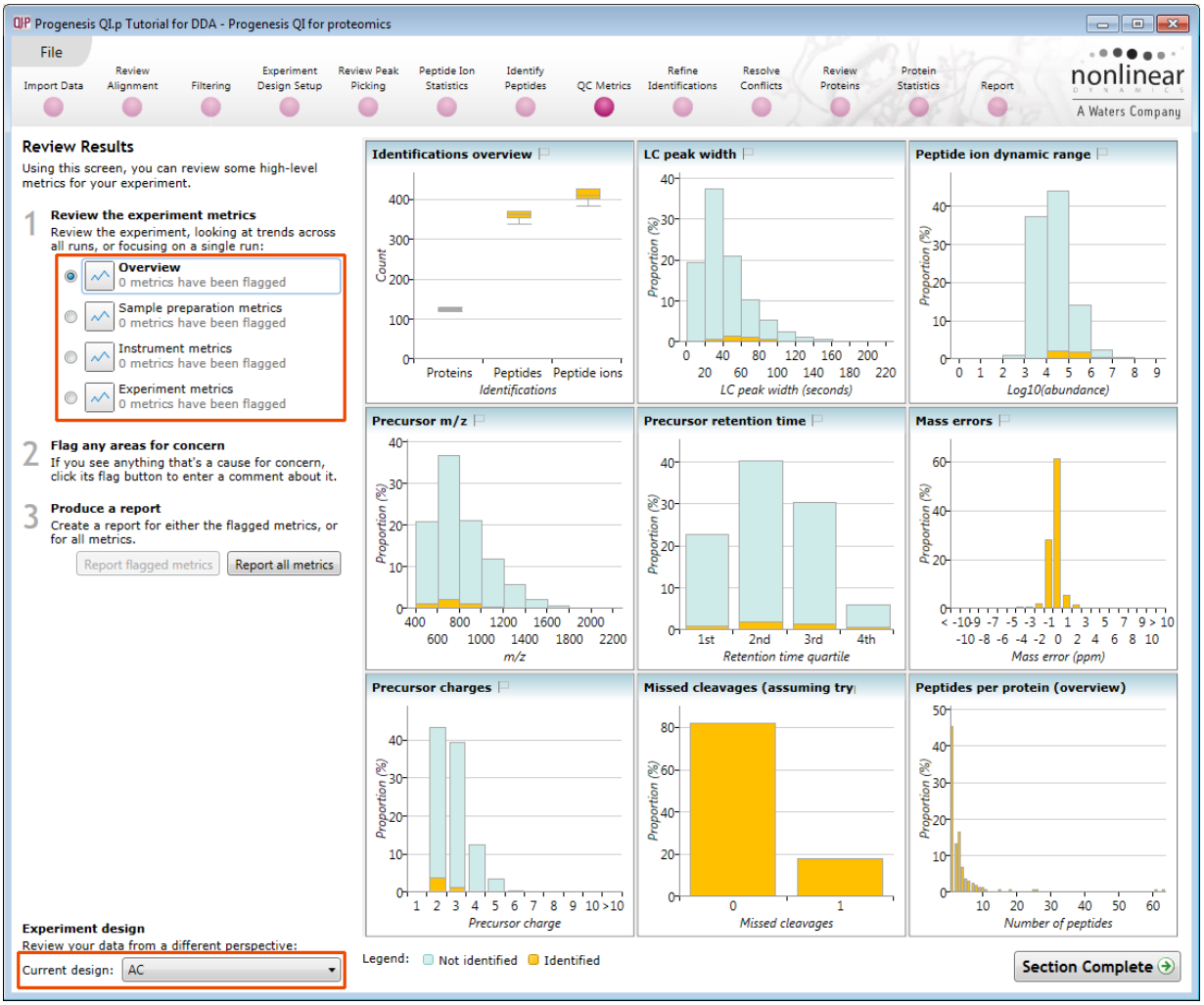


In order to review, and refine the quality of the **Peptide Search** results click on the next stage in the workflow, **Refine Identifications**.

Stage 10: QC Metrics

Progenesis QI for proteomics includes a number of batch-based QC metrics which you can examine to increase confidence in, or identify issues with, your data. These views are presented at the “QC Metrics” page. By setting up experimental groupings that correspond to your batch metadata, you are able to investigate whether there are any systematic effects associated with your processing. This gives you confidence that your conclusions are not affected by technical biases within the course of the experiment.

The charts are updated with changes in your active experimental design, allowing you to examine the effects of all stages of your processing if you wish through using different batch divisions.



Metrics

There are two sets of metrics. Firstly, the overview page provides nine at-a-glance summary measurements which cover the experiment as a whole. Then, the remainder of the pages provide more detailed information and bring in the batch-by-batch detail, some of which builds on the overview charts:

- Sample preparation metrics – highlighting issues or problems with the preparation of your samples:
  Missed cleavages, modifications and abundance dynamic range
- Instrument metrics – highlighting whether your chromatography column and mass spectrometer are configured and performing correctly:
  Mass accuracy, abundance dynamic range (again), precursor charges, MS1 scan rates
- Experiment metrics – concerning the identified proteins and peptides in your experiment, allowing you to pick out any outlying runs or conditions:
  Proteins, peptides, peptides per protein, % of peptide ions identified, proteins per condition

More detailed information on the QC metrics is available on the FAQ pages.

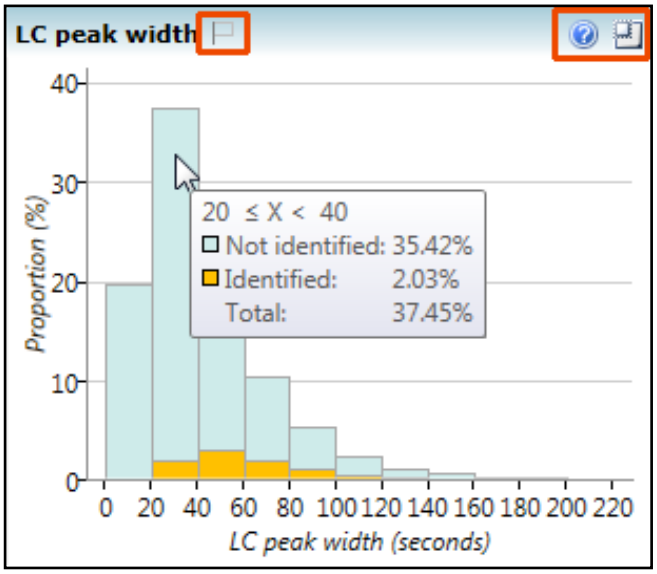
Interpretation and use

It's worth noting that there is no simple 'right' or 'wrong' answer as to whether your data are of high enough quality.

Note: QC measurements are designed to flag up potential issues for investigation, hence there must be an assignable cause to the variation observed to render QC practical.

Furthermore, the QC metrics will be rebuilt if you delete and re-do any stages leading up to them (for example, clearing identifications and re-searching with changed identification parameters, or altering the alignment). Hence, if you archive your experiment before and after your changes, or export the charts before and after, you can see the effects of your changes upon the quality measurements – a very useful method for assessing your interventions.

Note: that the overview metrics can each be expanded by clicking on the top-right icon in the sub-window. Additionally, hovering over a column will bring up a tooltip containing quantitative information on the results.



You can also tag metrics of interest or concern with a comment for your records, which is saved with the experiment. To do this, click the empty flag icon in the tab header for the given metric (or in the tile header on the overview screen):

This will pop up a dialog allowing you to enter a comment describing why this metric has been flagged.

Reporting

You can export a report for your own records, or to pass on to another member of your team:

Report all metrics:

Generates a report containing all metrics shown in the application (overview, sample preparation metrics, instrument metrics and experiment metrics). This may be useful for documentation purposes, or to verify the quality of your experiment.

Report flagged metrics:

Generates a report of only metrics you have flagged, along with the message you provided. This may be useful for giving to a technician or other team member, to highlight areas of the experiment that need improvement.

In order to review, and refine the quality of the **Peptide Search** results click on the next stage in the workflow, **Refine Identifications**.

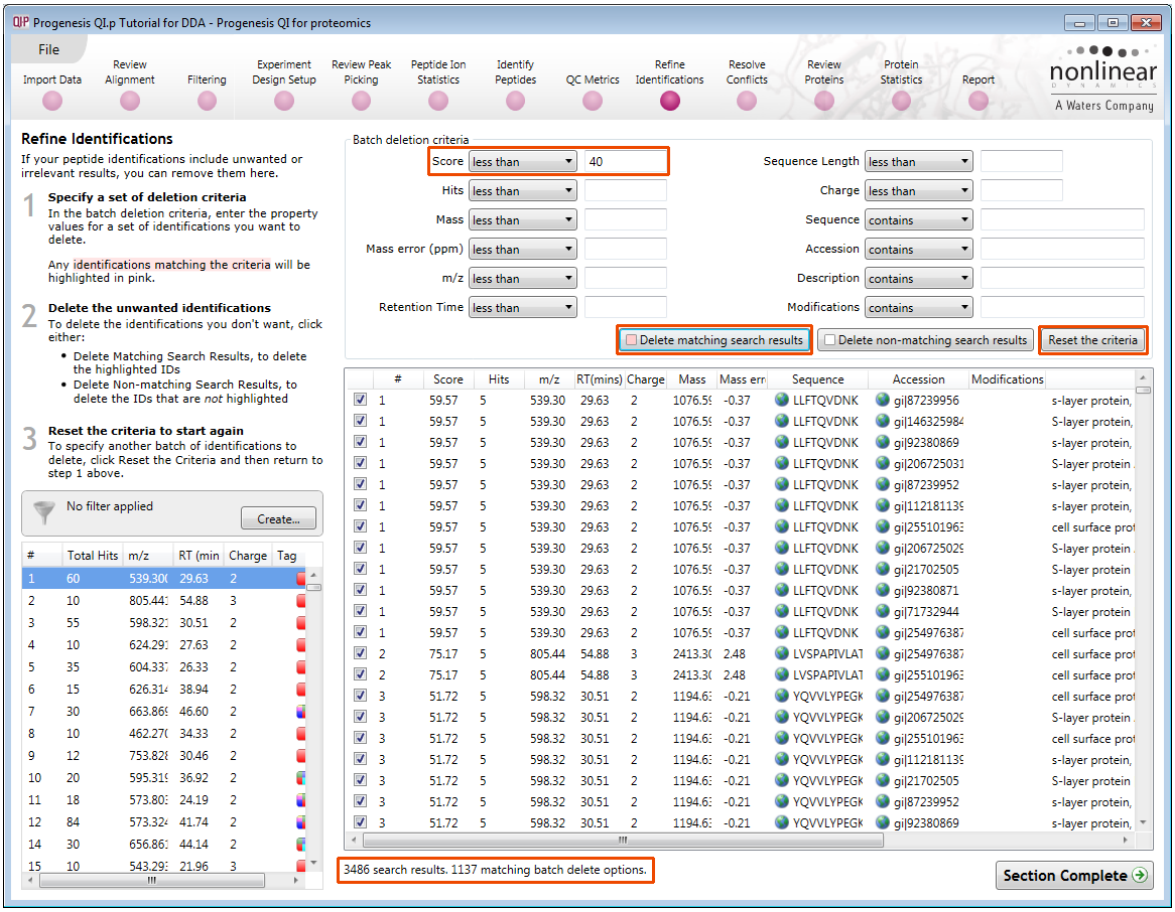
Stage 11: Refine Identifications

In this example the organism under study is *Clostridium difficile*

Note: before removing any identifications, make sure there are **no** tag filters applied at the Identify peptides stage.

As an **example** the following section describes how sequential filtering of the Peptide results can be performed using the following thresholds described below:

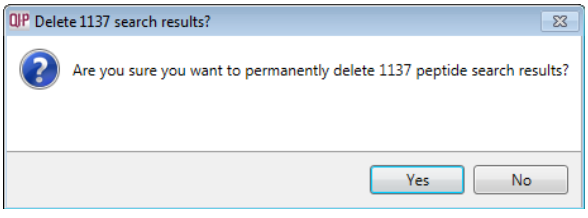
- Remove identifications with a Score less than 40
- Remove identifications where less than 2 hits were returned
- Remove all identifications where the Protein Description **Contains** 'hypothetical'
- Remove all identifications where the Protein Description **Contains** the following: 'Like', 'Putative', 'Probable', 'Potential', 'Predicted' and 'Partial'
- Remove all identifications where the Protein Description **Doesn't contain** 'Clostridium difficile'



On the Batch detection options panel, set the Score to less than 40, then **Delete matching search results**.

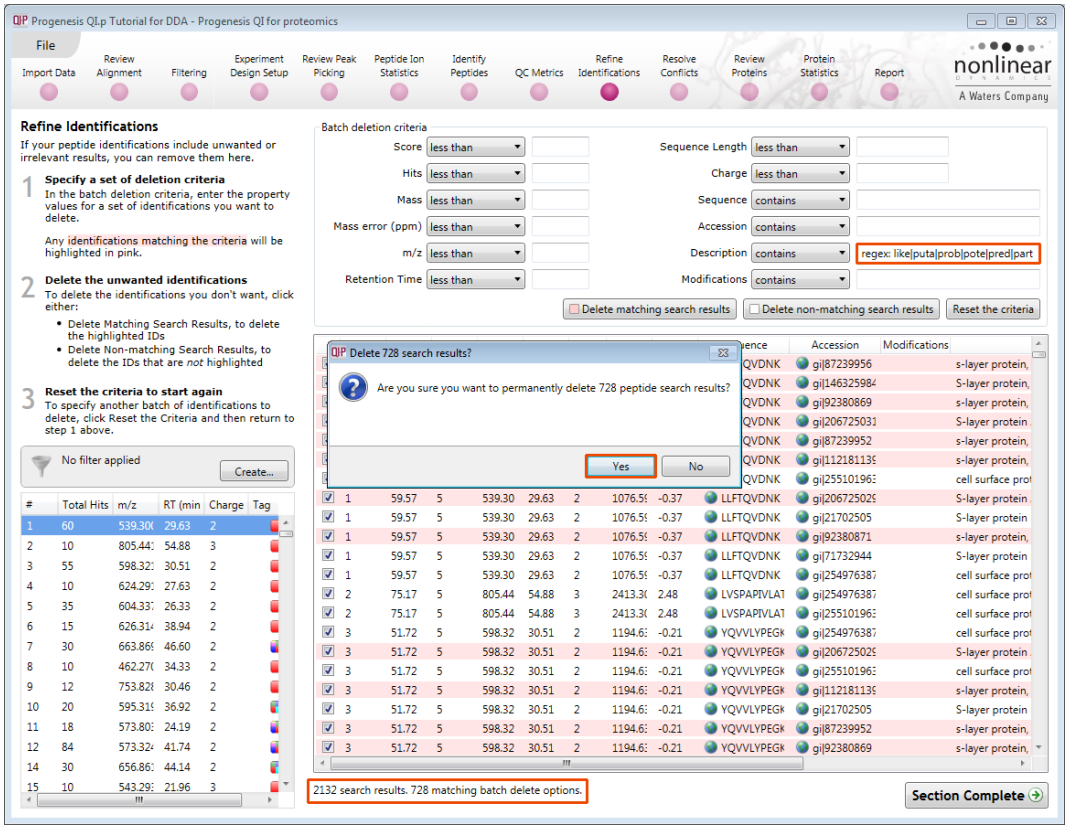
Note: the search results matching the filter criteria turn pink and the total is displayed at the bottom of the table (1137 matching out of 3486)

Note: a dialog warns you of what you are about to delete

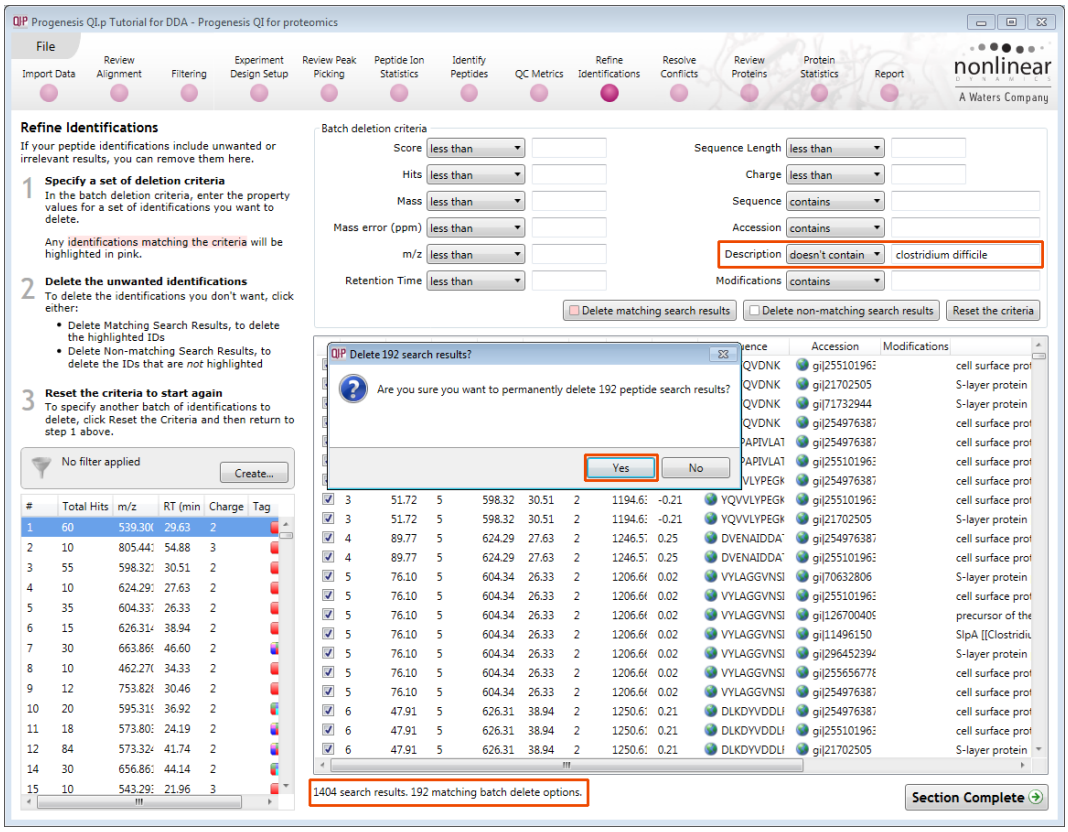


Now click **Reset the criteria** to clear the previous threshold then apply the next filter 'Hits: less than 2' .

Now in the 'Description' field enter '**hypothetical**' and delete matching search results. Then also in Description field enter the 'regular expression': **regex: like|puta|prob|pote|pred|part** and delete matching search results.



Finally alter the Description to 'doesn't contain Clostridium difficile' and delete the matching search results.



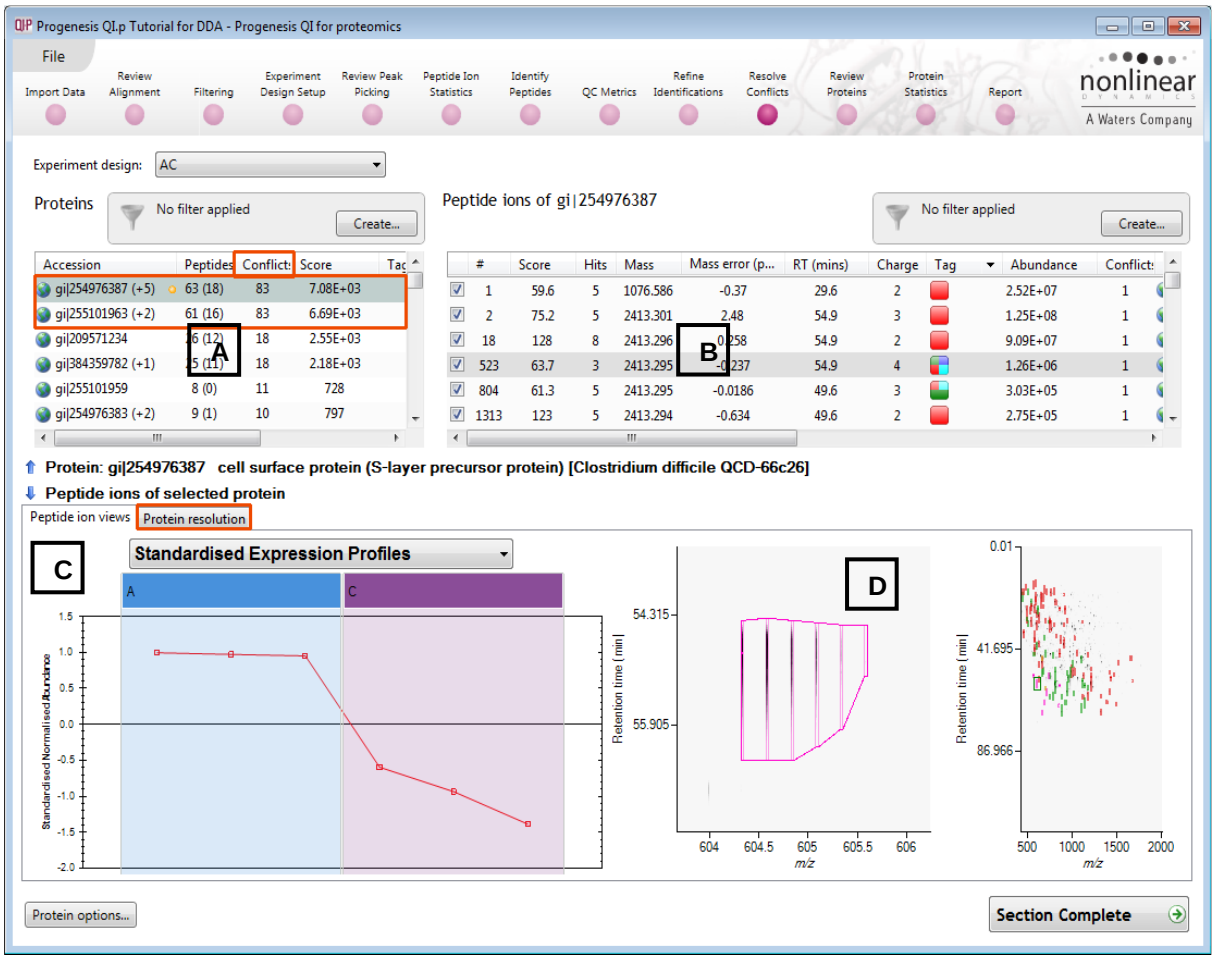
Having applied all the filters there will be **1212** search results remaining

To validate the Peptide search results at the protein level select the next stage in the workflow by clicking on **Resolve Conflicts**.

Stage 12: Resolve Conflicts

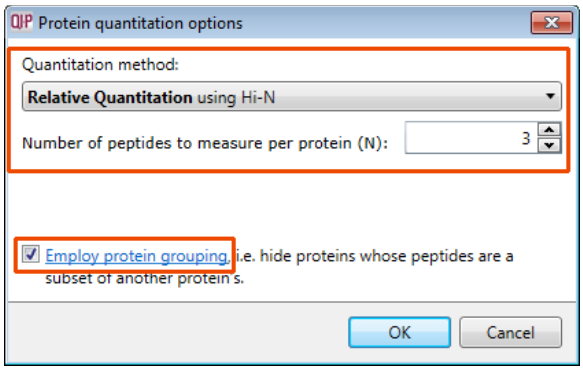
This stage allows you examine the behaviour of the identified peptides and resolve any conflicts for the various peptide assignments at the protein level.

The **Resolve Conflicts** stage provides a number of interrelated graphical and tabular views to assist you in the validation of the peptides that have been assigned to proteins and also to review the relevance of the data returned from the search.



The number of conflicts you have to resolve will depend on the scope and stringency of the filters you apply at the **Refine Identifications** stage.

Note: the default **Protein options** for protein grouping and Protein quantitation are set as shown



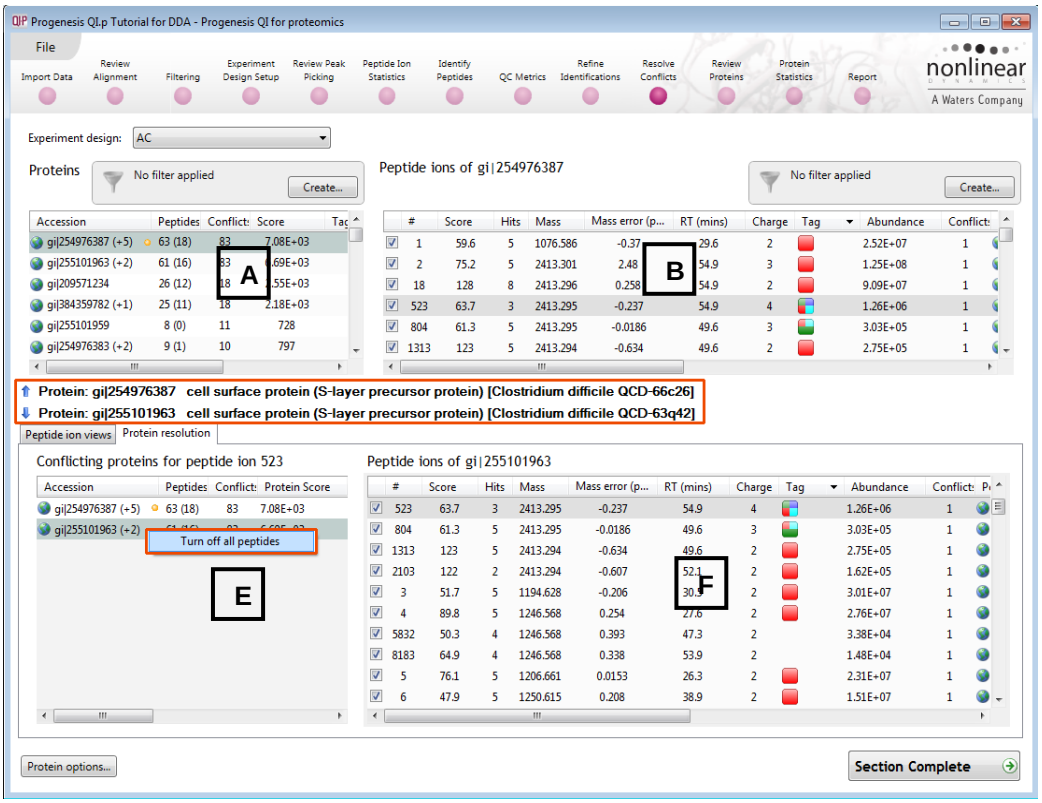
This means that if you choose **not** to resolve the conflicts then proteins, to be considered for quantitation, require at least one unique peptide (number in brackets after peptide count).

For more details on **Protein Grouping** page 53 and **Protein Quantitation** go to page 54

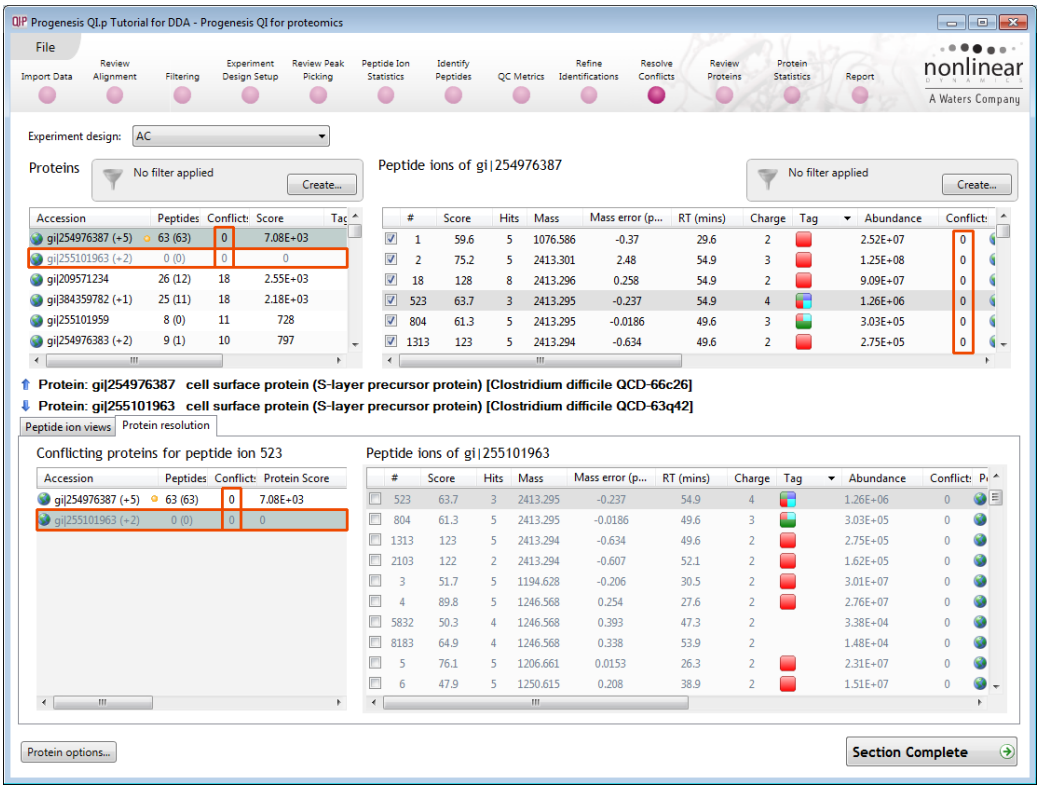
Open **Resolve Conflicts** and order the data in the Proteins table (A) on the basis of **Conflicts**.

Note: the look of the tables (with regards to ordering) in the following section may vary slightly.

Depending on the ordering, make '**cell surface protein (S-layer precursor etc '** the current protein by clicking on it in Window A (a circular orange symbol indicates current protein). This protein has 63 peptides assigned (window B) which have a total of 83 conflicts. To view the conflicting assignments click on the **Protein Resolution** tab (window C).



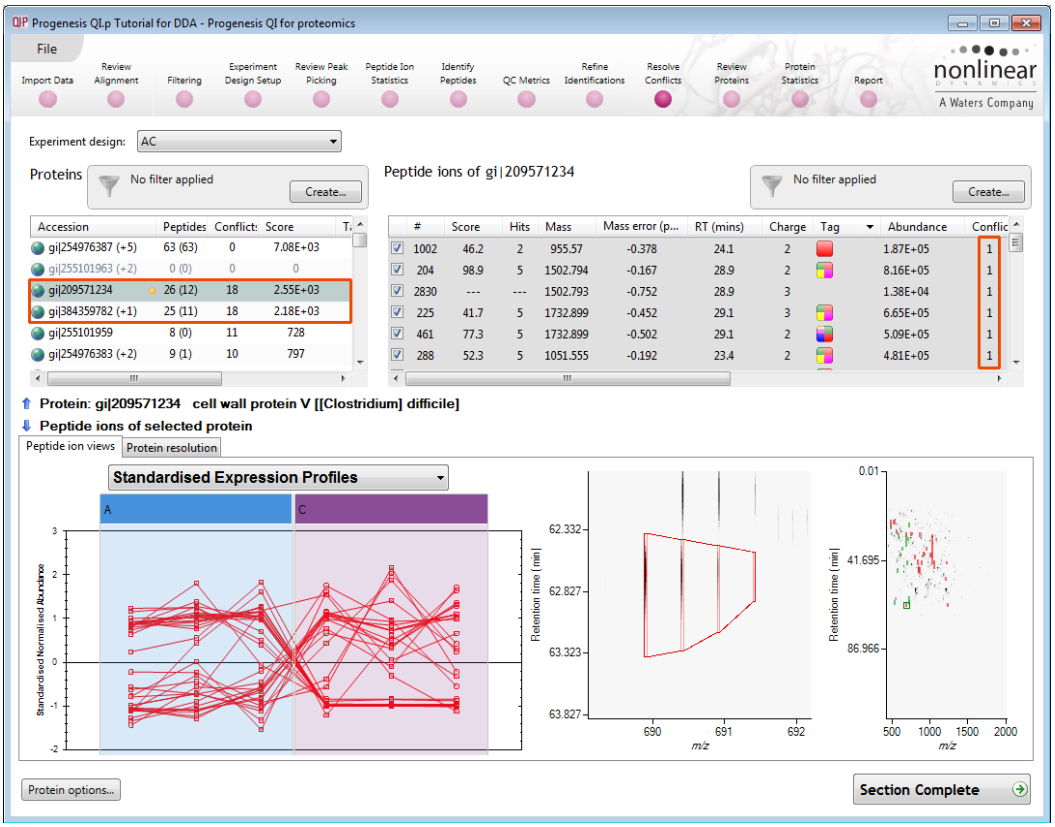
In this case the conflicting peptide assignments are with '**The same protein'** (from a different strain) which also contains 83 conflicts. A simple resolution to these conflicts is to right click on the conflicting protein and turn off all its peptides (based on lower number of peptides and score).



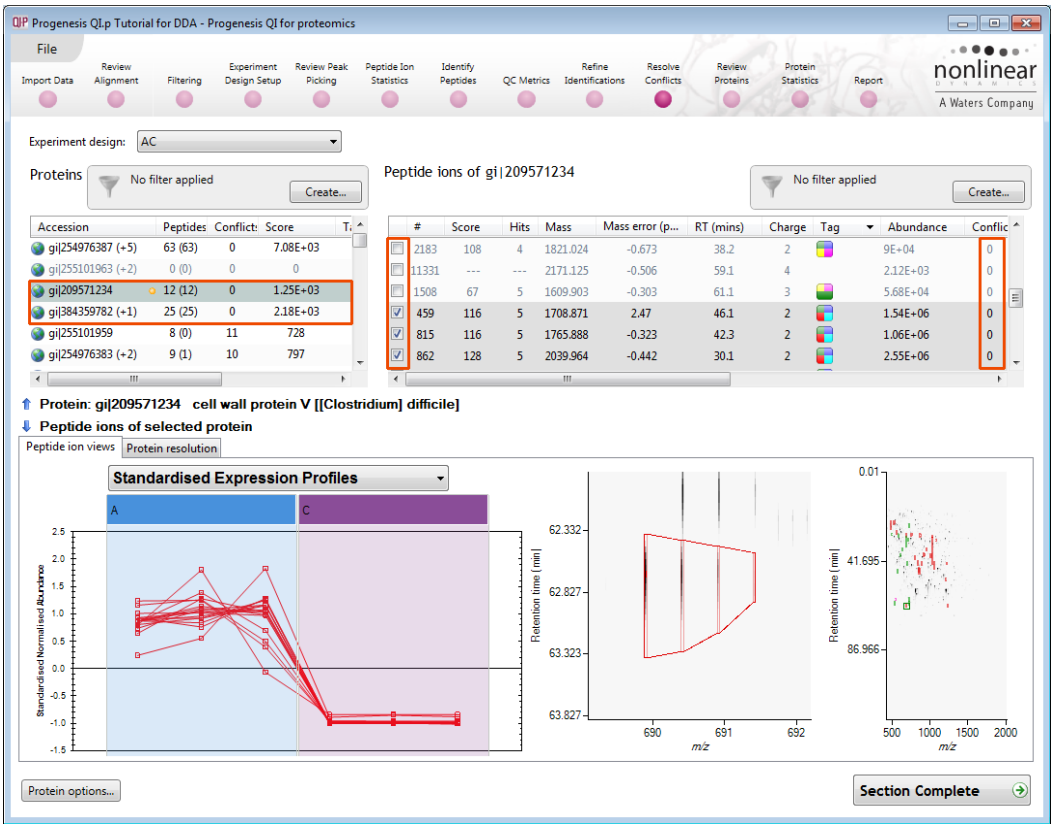
Note: as you un-assign the peptides the number of conflicts update 'on the fly' in all the windows.

In this case the conflicting peptides are unassigned from the 'precursor' protein.

In some cases you can resolve the conflicts between 2 proteins on the basis of consistent peptide expression. In the example below the proteins share 18 conflicts. For the protein showing 2 clear patterns of expression you can un-tick all the peptides with conflicts in the corresponding peptides table

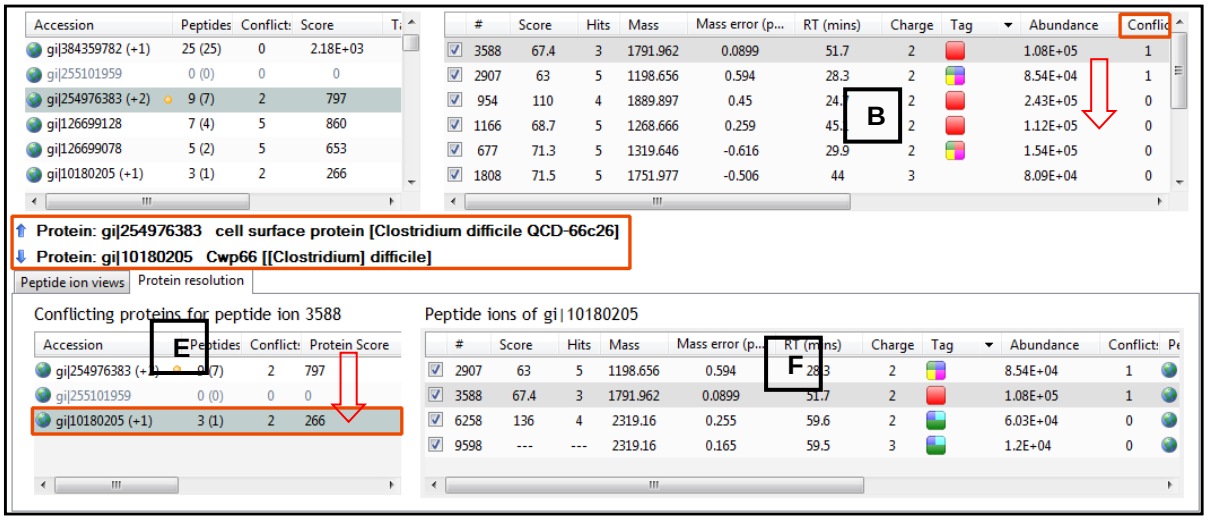


This leaves the peptides with the same expression pattern assigned with the appropriate protein thus resolving the conflicts.

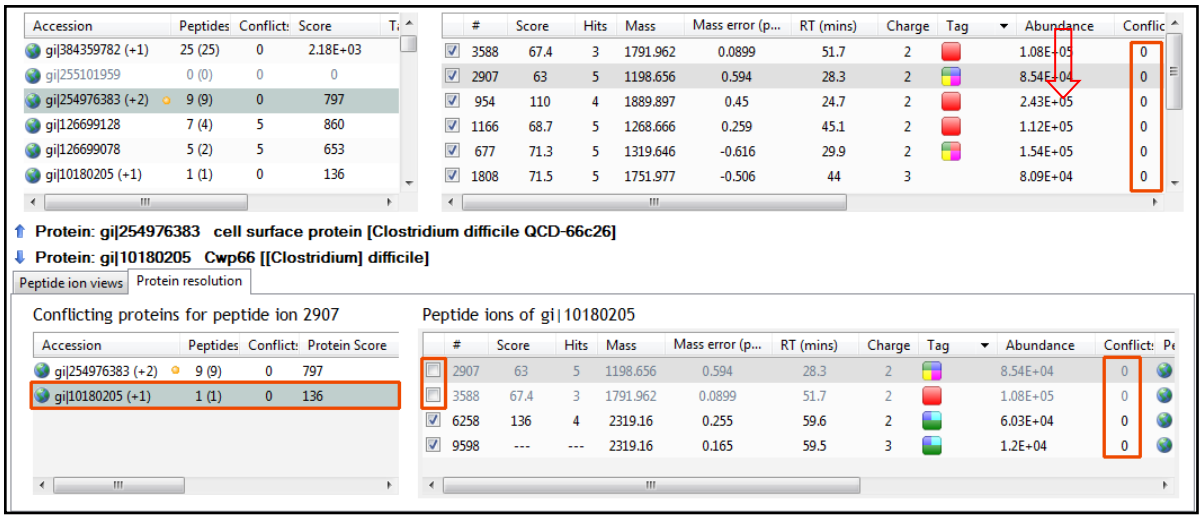


Note: the number of conflicts you have to resolve will depend on the scope and stringency of the filters you apply at the **Refine Identifications** stage.

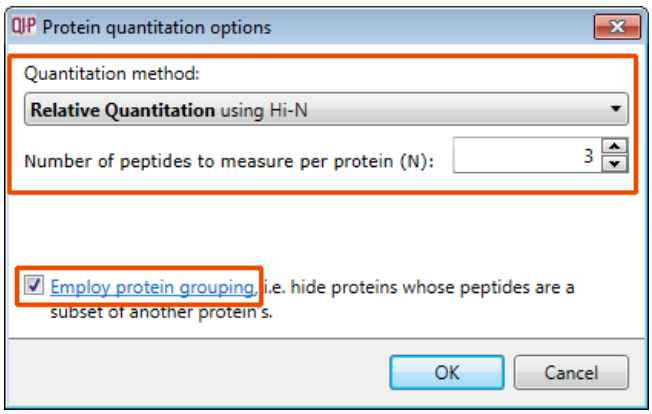
To resolve the remaining conflicts in the example shown below first order the conflicts in panel B and select the top one (which may still be selected) Panel B will display the peptides for this protein and the number of conflicts for each peptide. Panel E will also update to show the conflicting protein.



Favouring the protein with the higher score, resolve the conflict by switching off (or un-assigning) the peptide in panel F for the protein with the lower score. By doing this, the other 3 panels update to show the change in conflicts.



Note: the number of conflicts you have to resolve will depend on the scope and stringency of the filters you apply at the **Refine Identifications** stage.

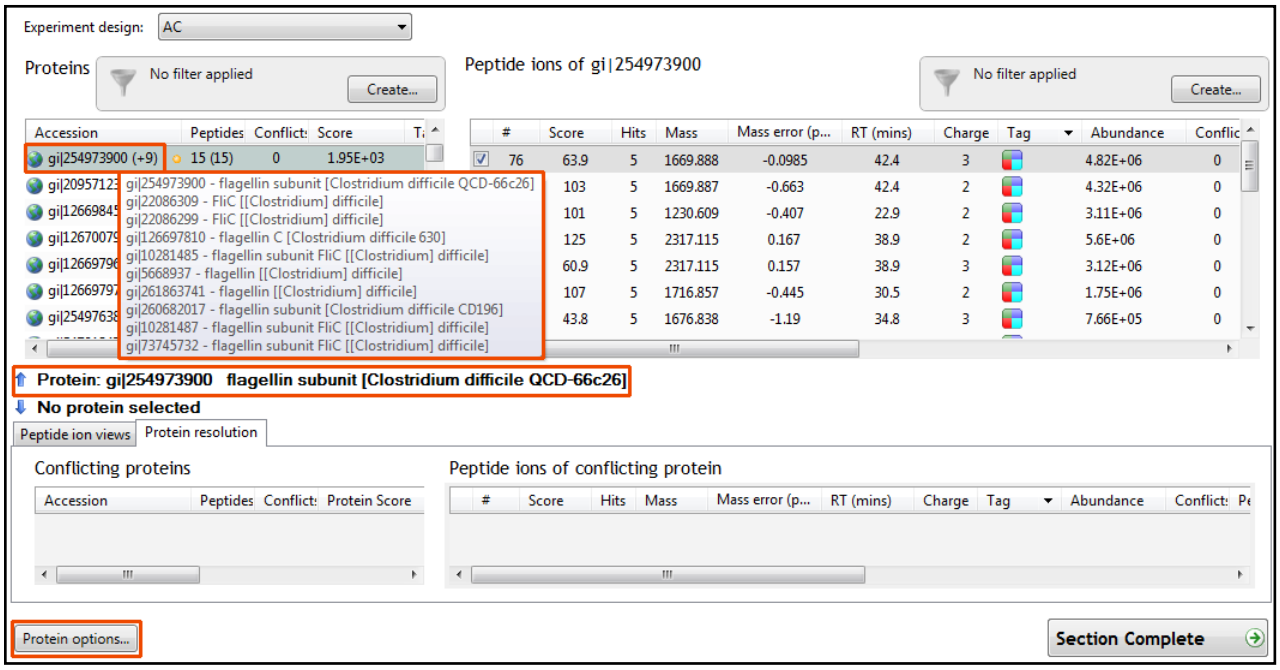


Note: the default **Protein options** for protein quantitation and protein grouping are set as shown

Protein Grouping

Where proteins are identified containing the same peptides then they are effectively indistinguishable aside from the score. Also where one protein contains only peptides that represent a subset of another protein's peptides, the protein with fewer peptides can be subsumed into that with the greater number.

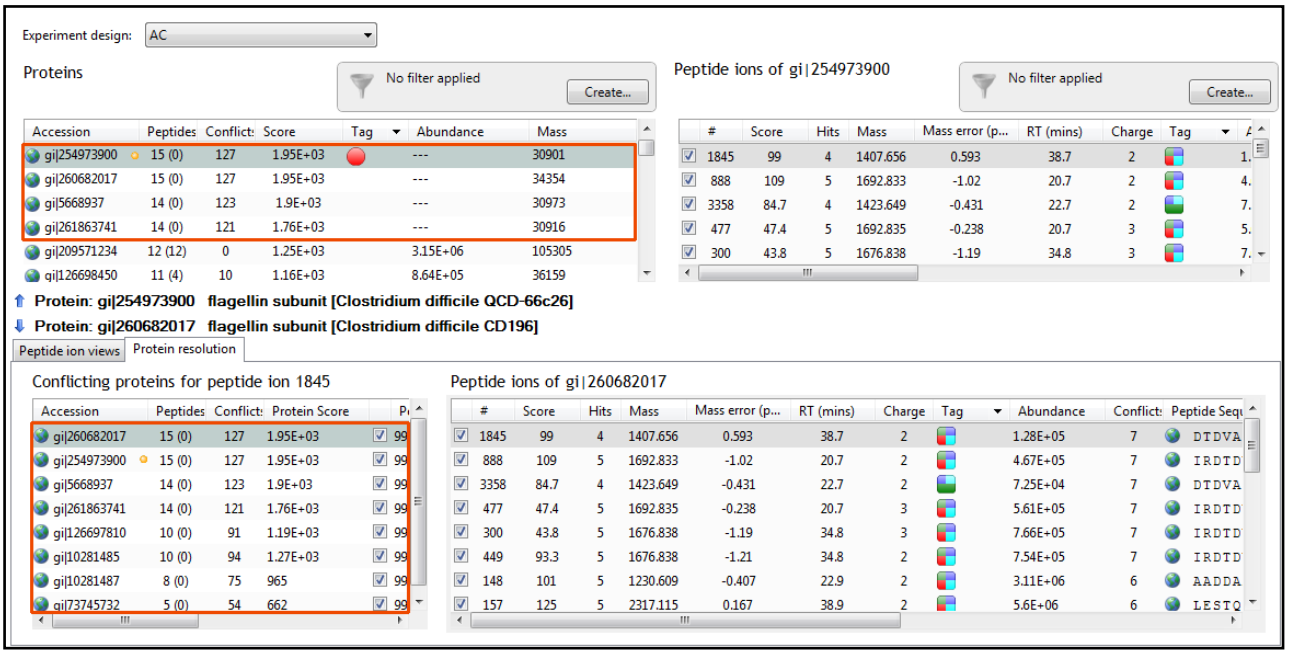
With protein grouping switched on (default setting) protein groups and the additional members are indicated by a bracketed number located after the Accession number. Taking **flagellin subunit** as an example, when the cursor is *held over the accession number the group members (9 in total) appear in a tool tip.*



Having performed the conflict resolution with **Employ Protein Grouping** now switch off the protein grouping to examine the effect of the process.

As grouping is switched off the grouped proteins appear with conflicts to the other group members and the number of unique peptides that are used for quantitation appear in brackets after the peptide number.

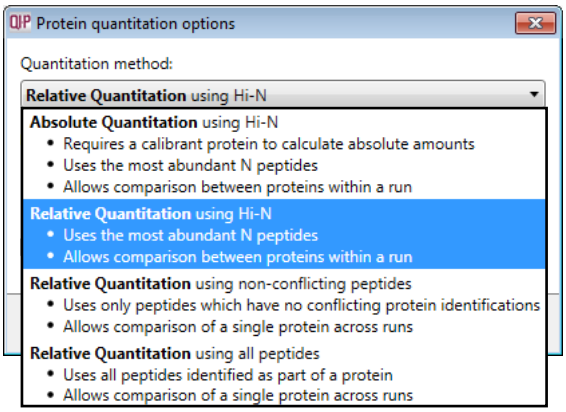
Note: flagellin has **no unique** peptides (brackets after the peptides field) as they are all present in flagellin subunit protein hence the reason for grouping. As a result all the conflicts are internal to the group.



Now set the Protein Options back to **Employ protein grouping**

Protein Quantitation options

There are 4 options with which to control how the Protein Quantification is performed by Progenesis QI for proteomics.



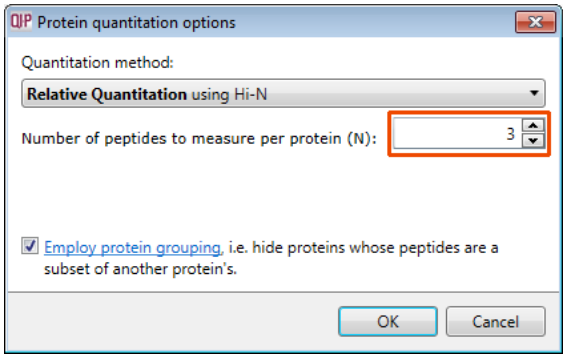
The default option that will be applied is **Relative Quantitation using Hi-N**.

If you have selected one of the other options during the setup of the Auto Processing of your data (Stage 1) then this option will be applied.

Hi-N in Progenesis QI for proteomics is an implementation of Hi-3 as described by Silva *et al.* [References]. After peptide and protein identification, the abundance of each peptide is calculated from all its constituent peptide ions.

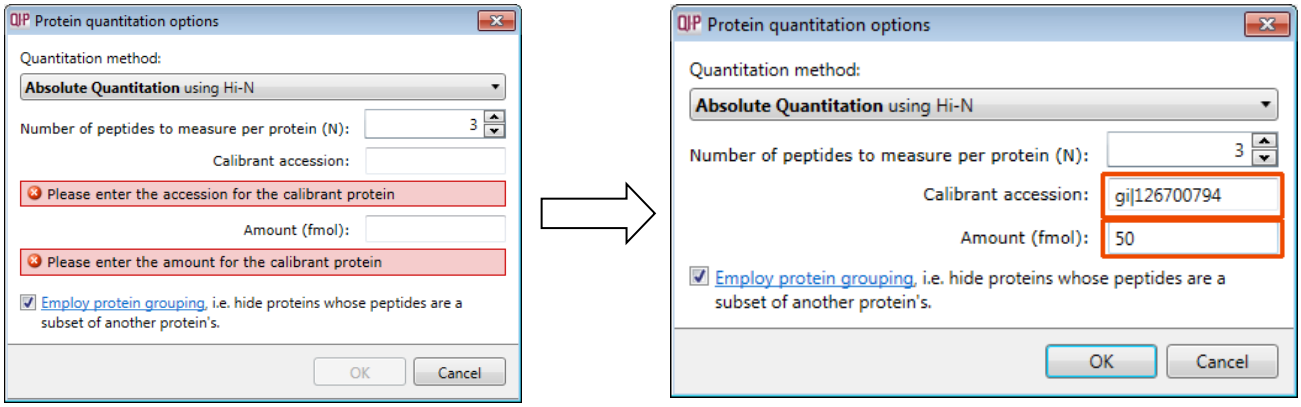
For each protein, the N most abundant peptides (N being set according to the user selection) have their **abundances averaged** to provide a reading for the protein signal.

The ranking of peptide abundance is based on the integrated value across all the runs, allowed by the accurate alignment and lack of missing values. This gives added confidence in the peptide selection, taking all runs into account to make the ranking robust.

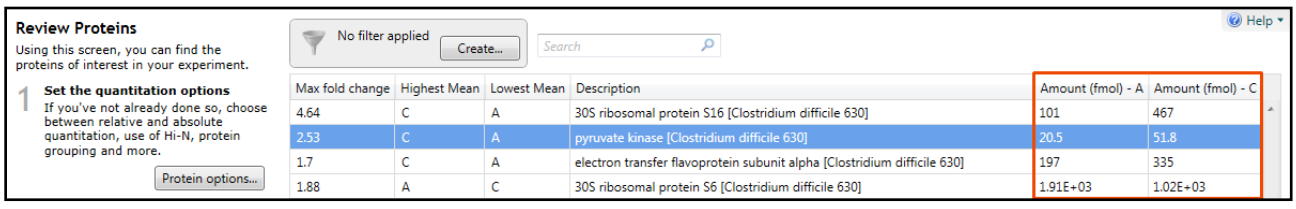


The averaged abundance readings not only make possible the **relative quantitation** of the same protein across all runs to be determined but also allow, with the inclusion of a known amount of a calibrant protein in each run, this to be converted to an **absolute** reading for protein amount.

To generate values for **absolute quantitation** enter the accession number and amount for the calibrant.



The absolute amounts, based on the calibrant used, are reported at the Review Proteins stage as additional columns (one for each condition) following the protein description.



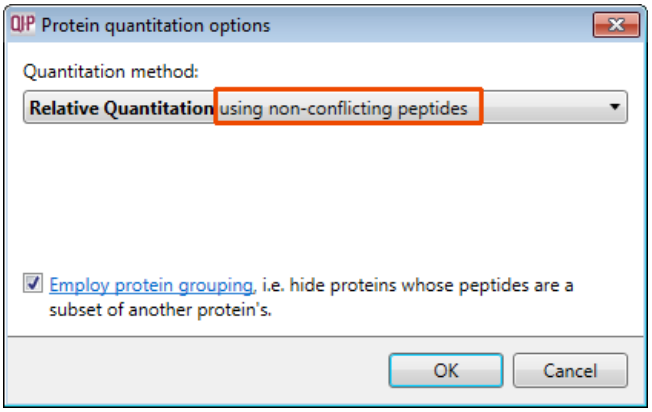
The ranking of peptide abundance is based on the integrated value across all the runs, allowed by the accurate alignment and lack of missing values. This gives added confidence in the peptide selection, taking all runs into account to make the ranking robust. These Hi-N methods allow the relative and absolute comparison of proteins within the same run.

Note: When there are **peptide conflicts** (a peptide is shared between two proteins, for example) it is important to assign the signal correctly for absolute quantitation. To do this, Progenesis QI for proteomics carries out a two-step process. Firstly, Hi-N is carried out only on the N most abundant **unique** (non-conflicting) peptides of the proteins concerned. This provides a ratio estimate for the two proteins based only on unshared peptides. The abundance of any shared peptides is then divided and allotted in this ratio between the two proteins, and the full Hi-N calculation is then applied using the divided values for conflicted peptides.

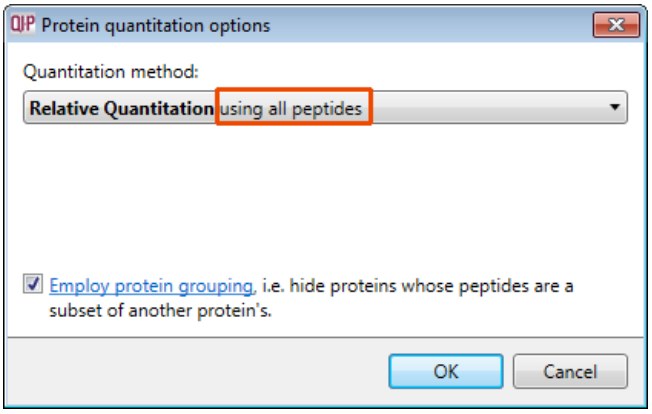
Naturally, if the conflicted peptides are not among the N most abundant in either protein initially, then this will not cause any difference in the result.

Relative Quantitation can also be performed comparing a single protein across all the runs using only the unique or non-conflicting peptides. Select the third method from the drop down

Using non conflicting peptides



The relative Quantitation can also be performed using all peptides.

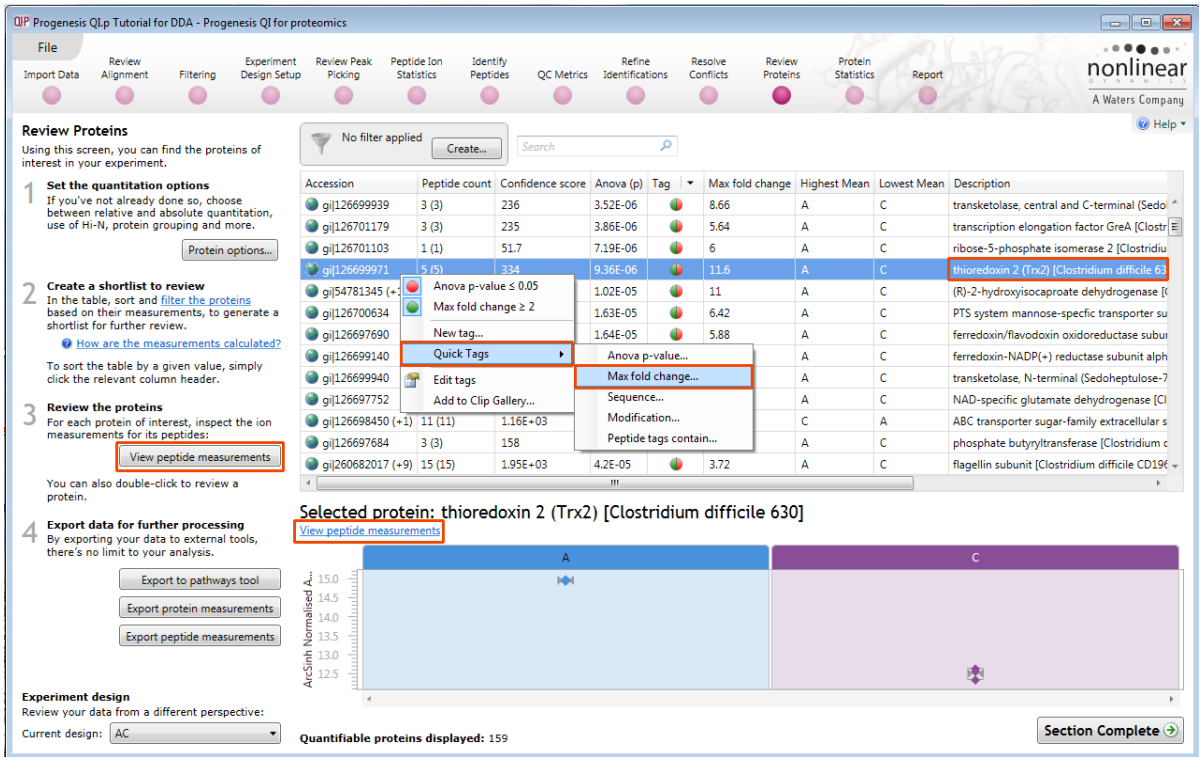


Note: if you have performed conflict resolution then there will be no difference between these methods.

Now move to the Review Proteins section by clicking on **Review Proteins** icon on the workflow at the top of the screen.

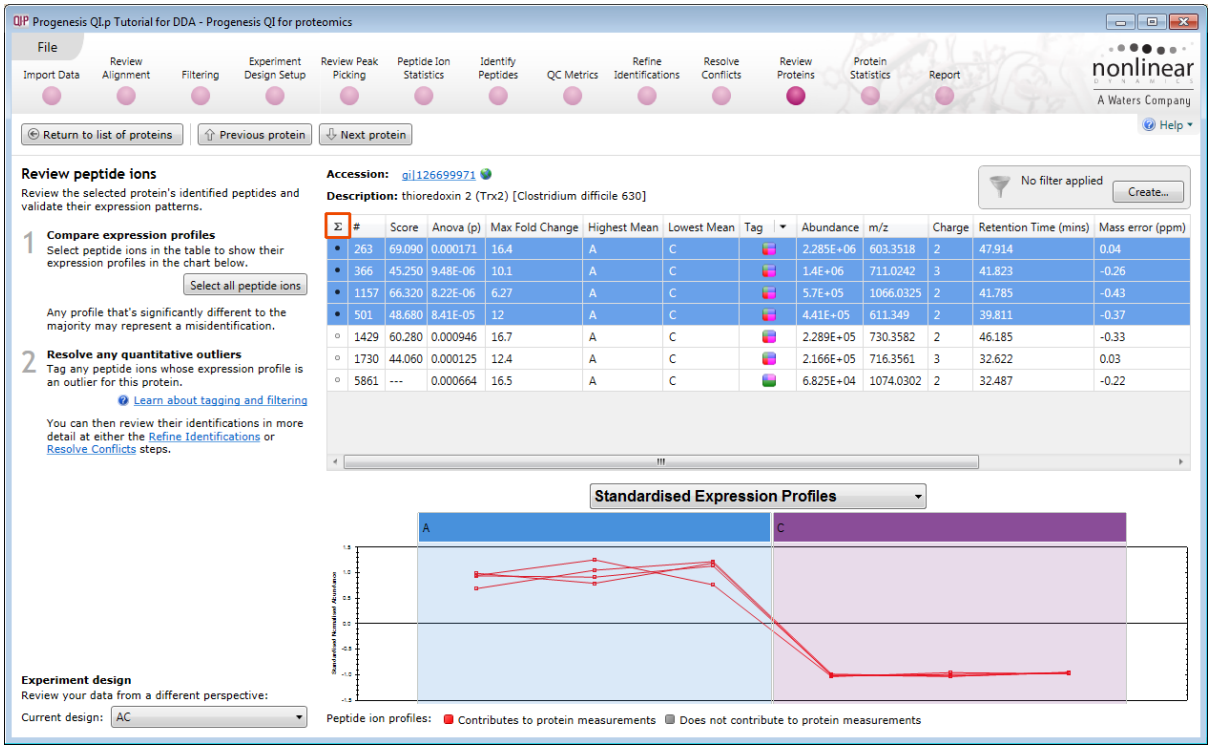
Stage 13: Review Proteins

The **Review Proteins** stage opens displaying details for all the proteins. You can now create tags at the protein level for proteins with an Anova p value ≤ 0.05 and Max Fold change ≥ 2 by right clicking on the table and selecting **Quick Tags…**



As an example let us explore Thioredoxin 2.

The table indicates that this protein is most highly expressed in Condition A by 11.6 fold over the lowest condition (C).

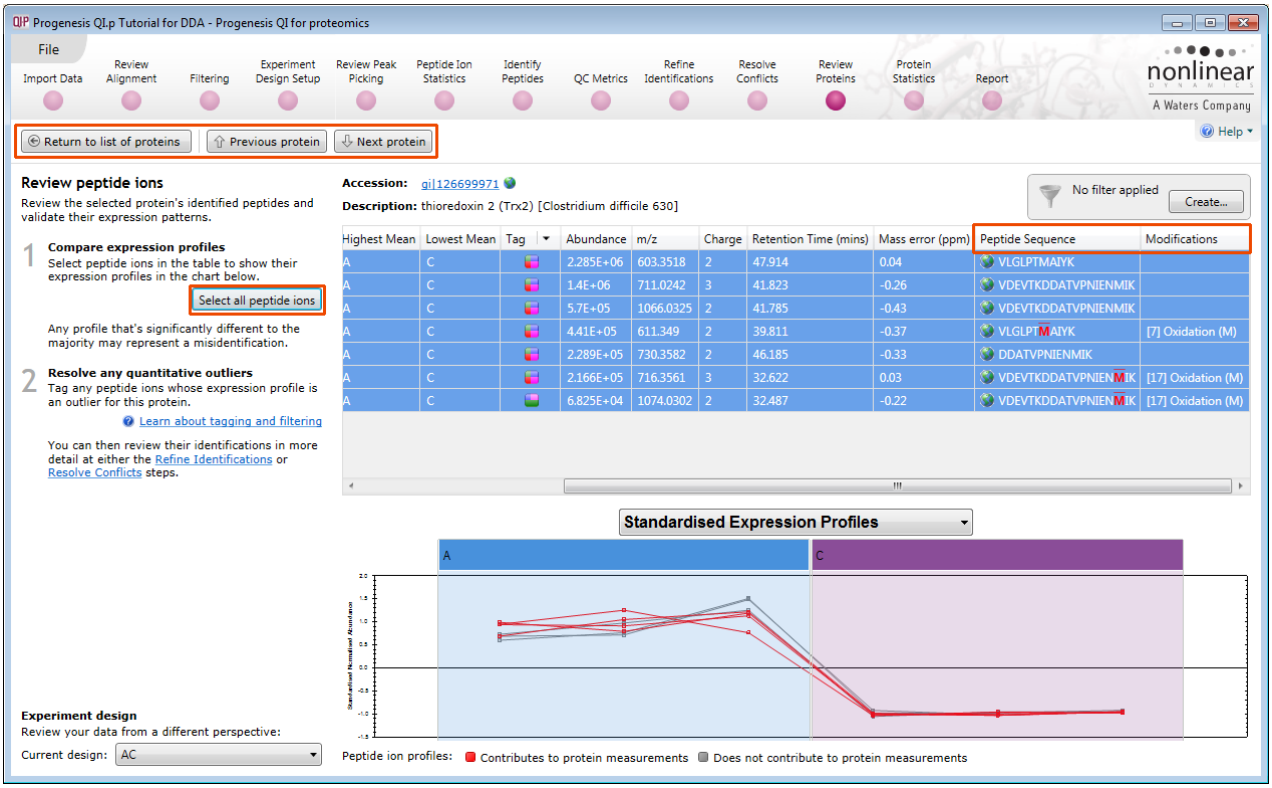


To view the corresponding peptide measurements for the current protein either double click on the protein in the table or use the **View peptide measurements** below the table.

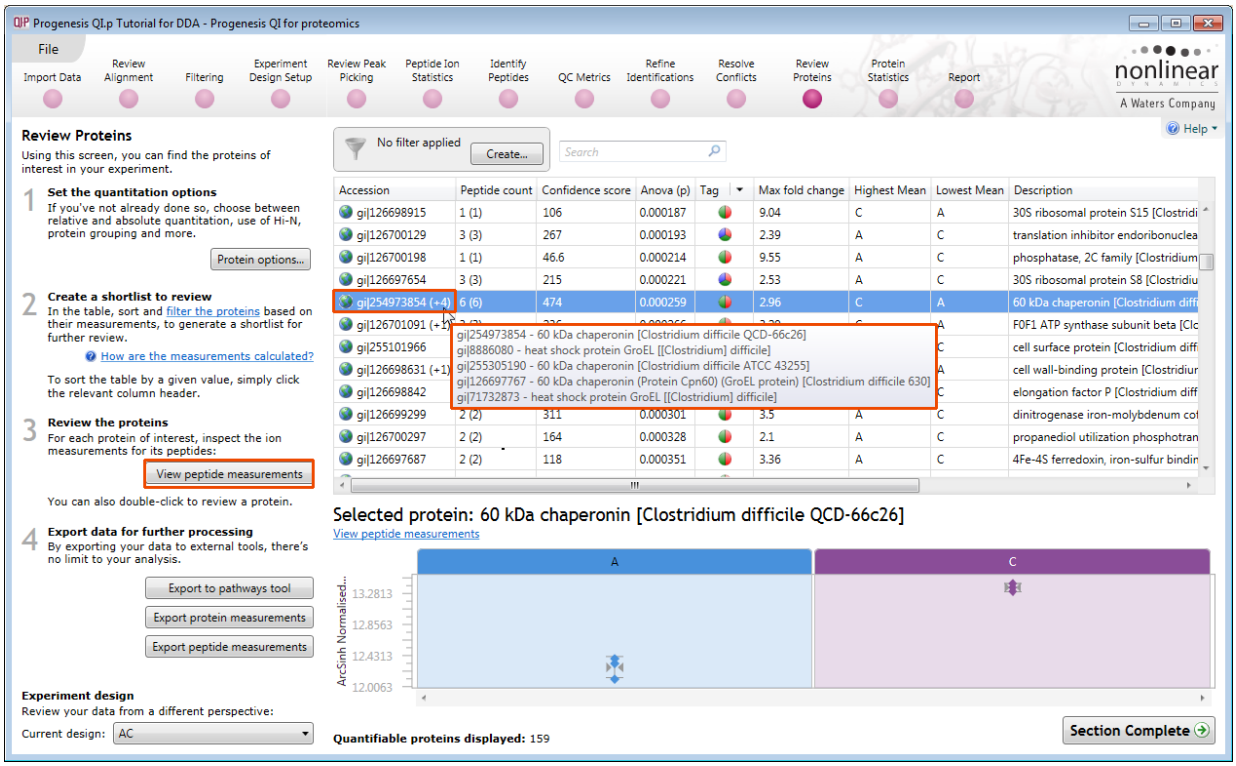
The solid icon in **Σ** column indicates that the peptide contributes to protein measurements.

By selecting all the peptide ions you can compare the pattern of expression across all the samples allowing you to identify any 'atypical' behaviour assigned to the current protein.

Scroll to the right on the table for the sequence and modification details. Use the navigation buttons on the top left either to review each protein's peptides by stepping through the list or return to the protein list and select a specific protein to review.

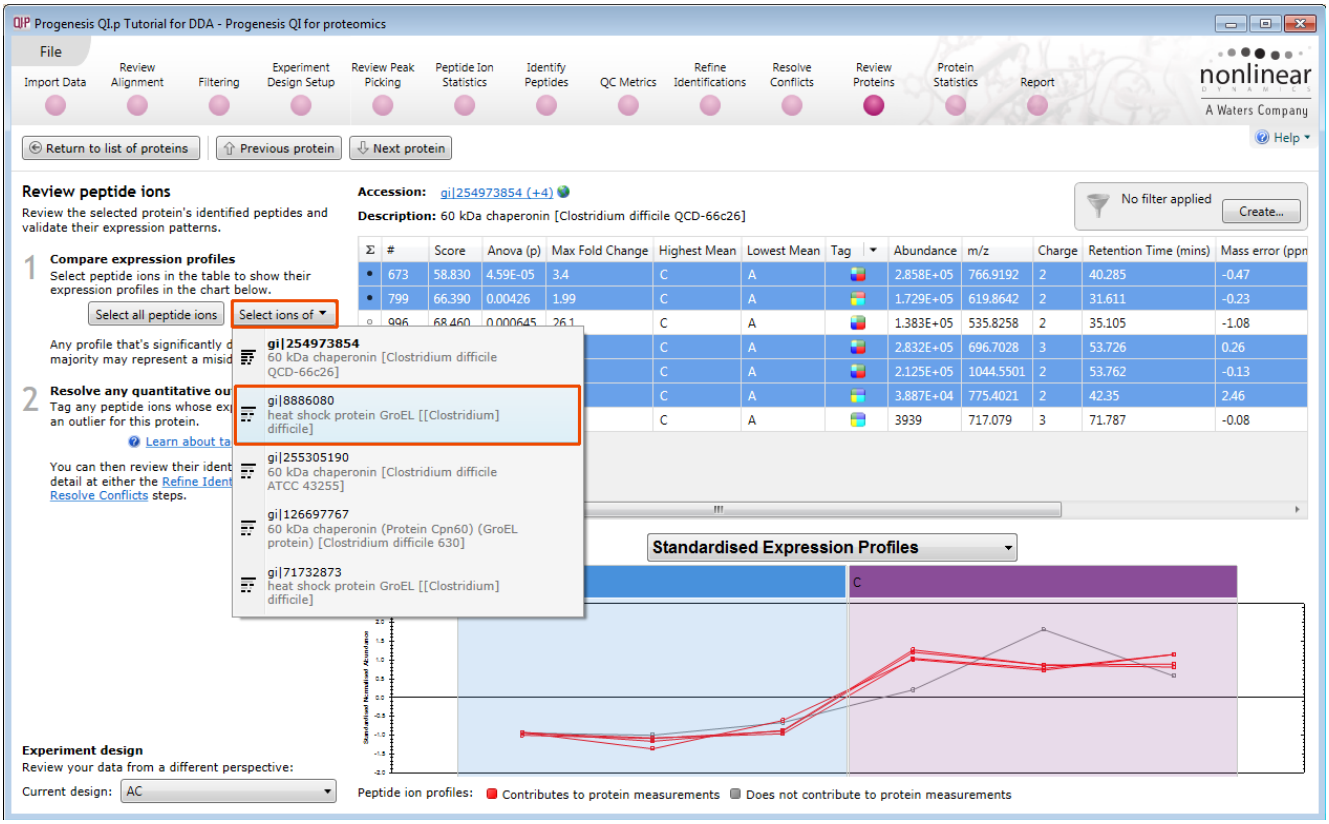


With protein grouping switched on (default setting) protein groups and the additional members are indicated by a bracketed number located after the Accession number. The additional members can be seen by holding the cursor over the accession number.

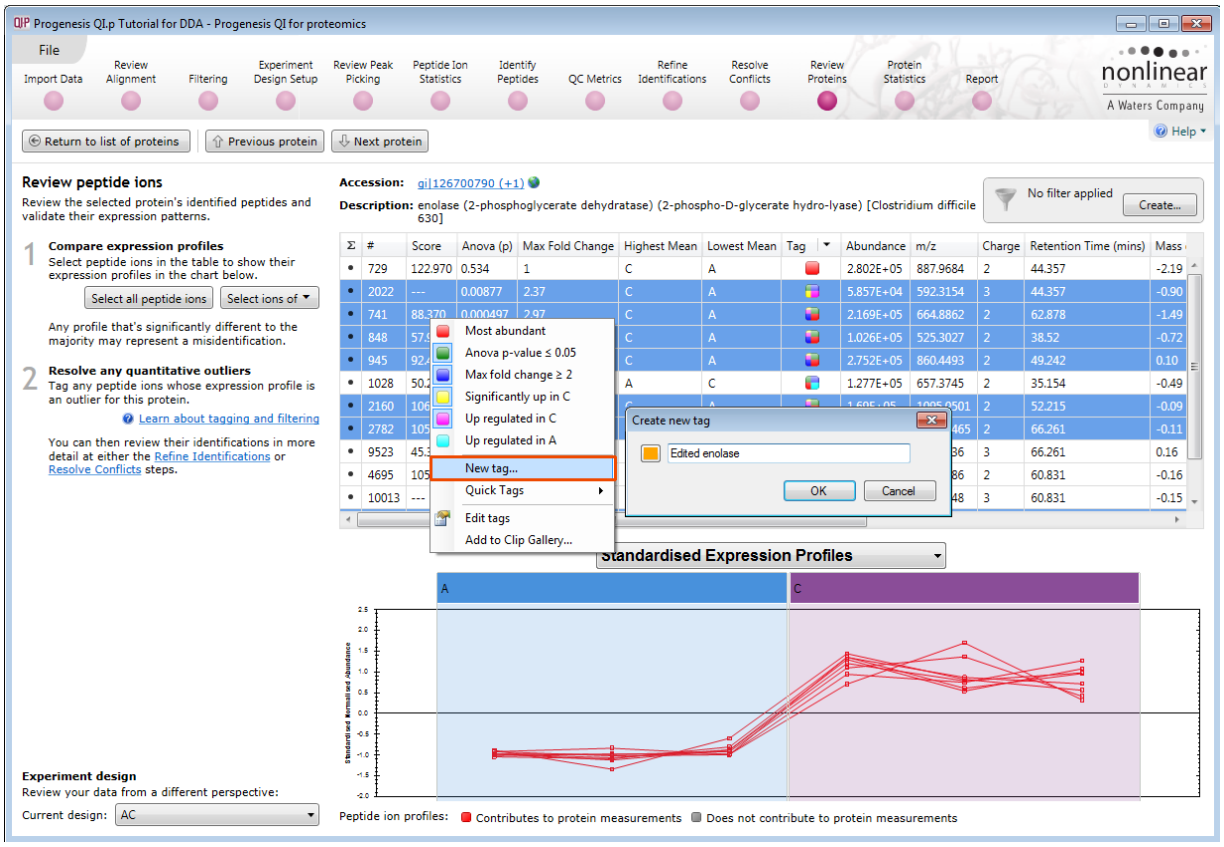


To view members peptide ions click on **View peptide measurements**.

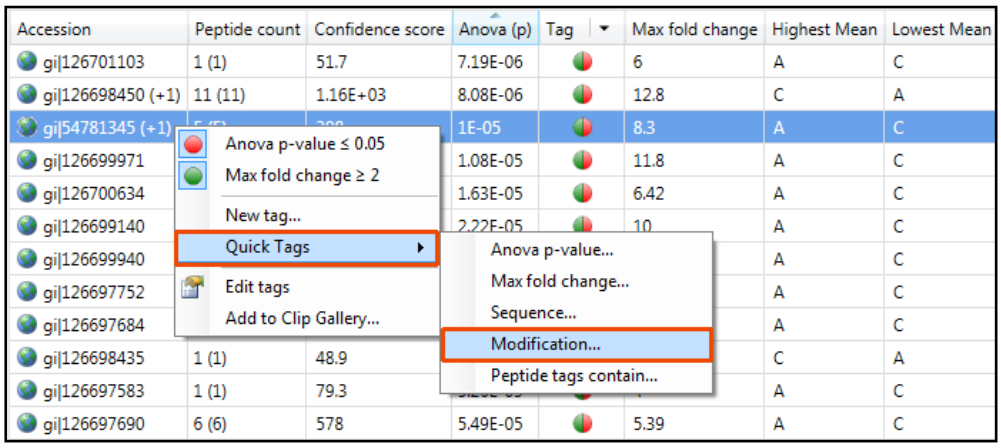
Click on **Select ions of** to show the list of additional group members. Then as you click on one of the member proteins its peptide ions are highlighted in the table.



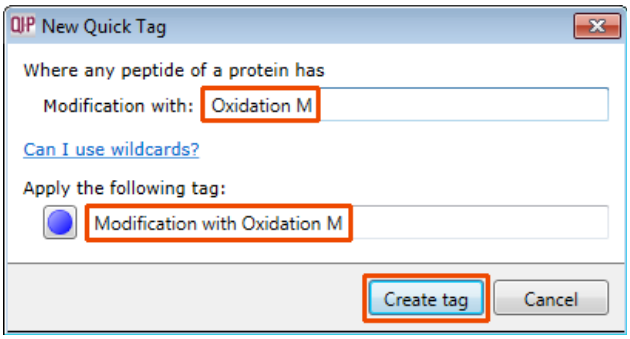
You can tag all the peptide ions for a protein or edit and tag accordingly for 'atypical behaviour'. Then remove these peptides at the **Refine identification** stage in the Workflow.



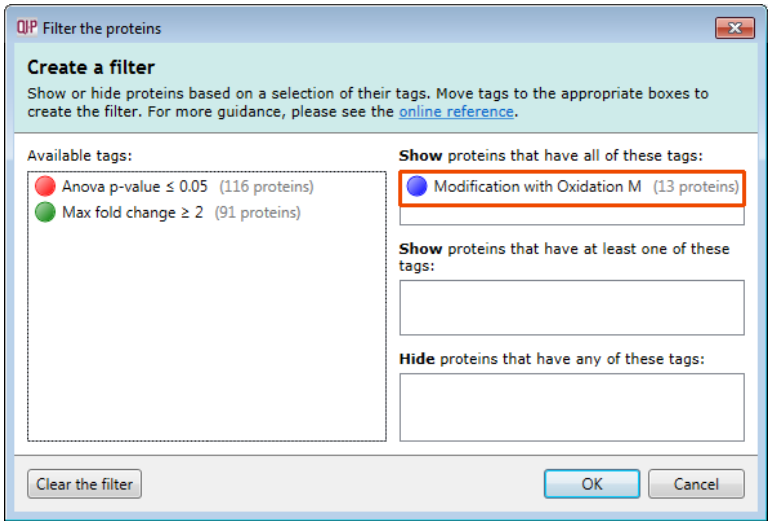
Modified proteins can be located by specifically searching for proteins containing modified peptides. Click the **Return to list of proteins** button to return to the proteins list and right click on it and select **Modification** from the list of **Quick Tags.**



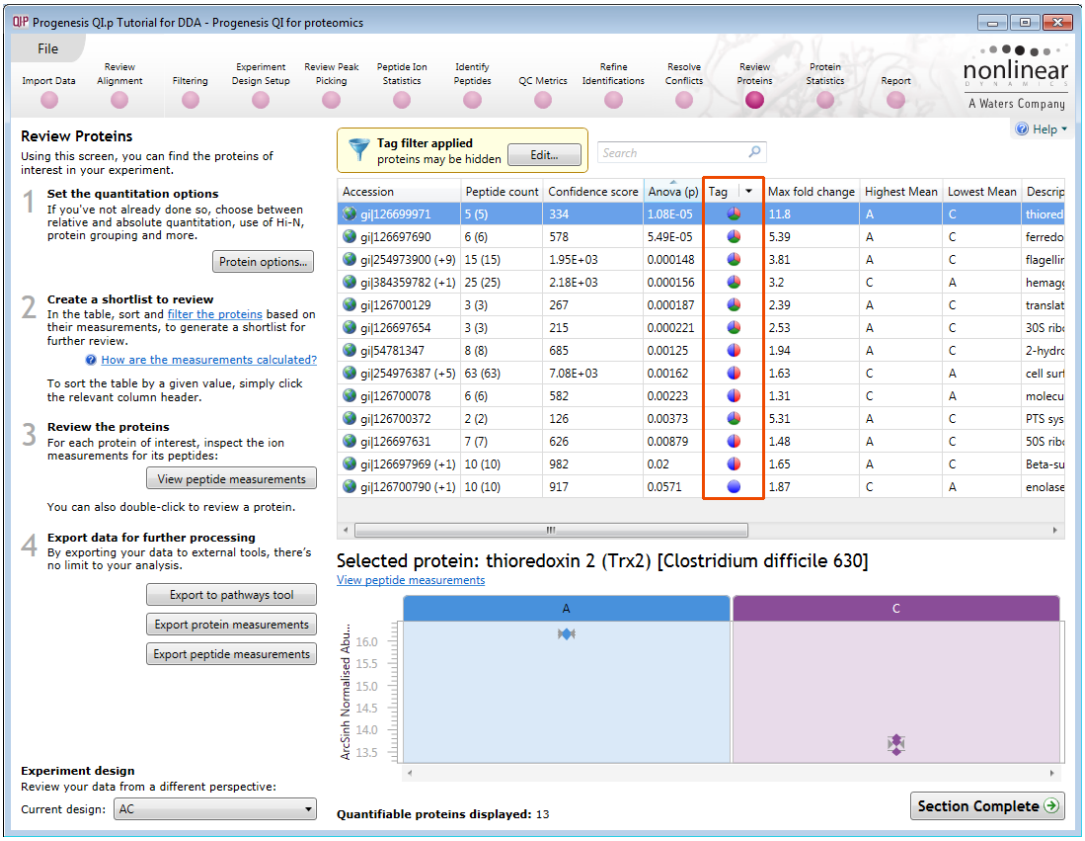
To find those proteins containing peptides with Oxidation on Methionine residues enter **Oxidation (M).** This will automatically provide a named tag when you click **Create tag**.



To reduce the table to displaying only these proteins with modified peptides (Oxidation on methionine) use the tag filter to focus on these proteins.



The proteins table will now only display those proteins containing modified peptides.

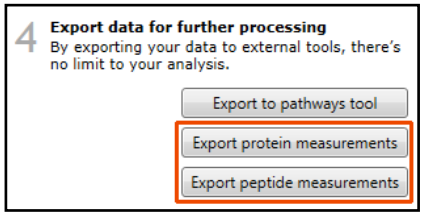


Note: the **Sequence** Quick tag can be used to locate Proteins containing peptides with specific motifs.

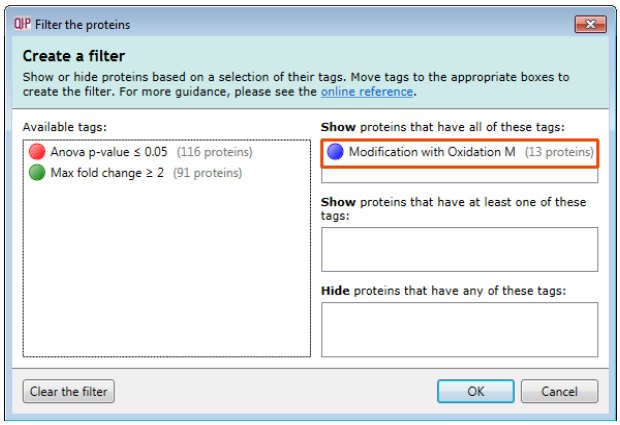
Stage 14: Exporting Protein Data

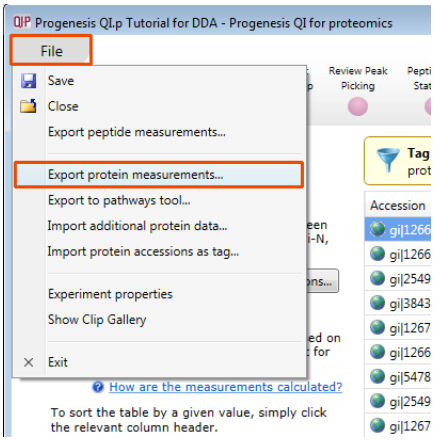
Protein data can be exported in a csv file format. You can either export the **Protein and/or peptide measurements** using the options in the File Menu or use the buttons under Step 4 both available at the **Review Proteins** stage**.**

As an example of Data export use the Tag filtered set from the previous section. Where you are only going to export measurements for those proteins that a have Oxidised Methionine residues.



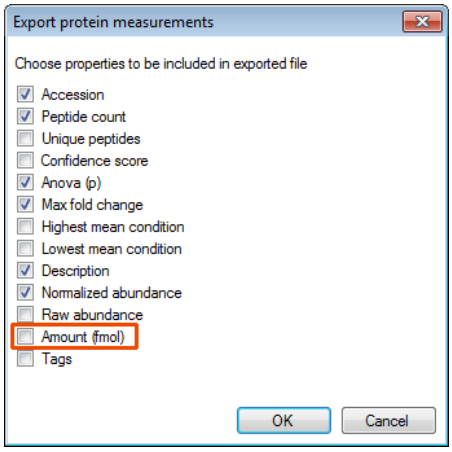
First set the tag filter as shown below. Then select **Export Protein Measurements**.

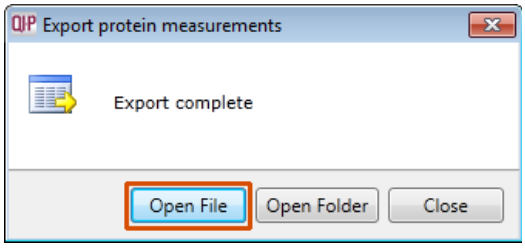




The Export Protein Measurements dialog opens. Select the required fields and click OK. **Save** the file and then open the exported data file using the dialog that opens.

Note: if you have performed **Absolute Quantification** then the 'Amount' field will be available.



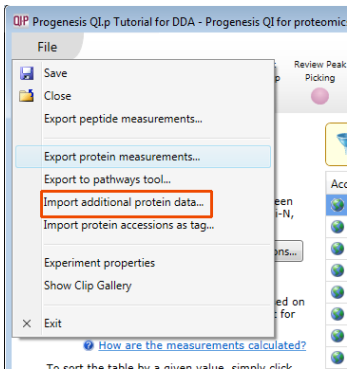


Excel will open displaying the exported protein measurements

	A	B	C	D	E	F	G	H	I	J	K
1						Normalized abundance					
2						A			C		
3	Accession	Peptide count	Anova (p)	Max fold change	Description	A1	A2	A3	C1	C2	C3
4	gi|126699971	5	9.36E-06	11.58476058	thioredoxin 2 (Trx2) [Clostridium difficile 630]	1525647	1617593	1553317	113654.1	140964.9	150789.2
5	gi|126697690	6	1.64E-05	5.879936323	ferredoxin/flavodoxin oxidoreductase subunit gamma	539615.9	622072.7	489155.7	90251.77	95359.21	95147.9
6	gi|254973900;gi|10281485;	15	4.20E-05	3.724777972	flagellin subunit [Clostridium difficile QCD-66c26]	7082586	6203247	7681870	1856199	1801158	1971892
7	gi|126700129	3	0.000192748	2.390273556	translation inhibitor endoribonuclease [Clostridium di	1868979	1833114	2077924	842205.4	722151.4	853783.7
8	gi|126697654	3	0.000220716	2.533396933	30S ribosomal protein S8 [Clostridium difficile 630]	338209.3	383095	400934.2	153782.3	156568.4	132627.1
9	gi|384359782;gi|209570719	25	0.002009624	1.884385068	hemagglutinin/adhesin [Clostridium difficile BI1]	585079.9	571830.8	706836.6	1226165	1039788	1246065
10	gi|126700372	2	0.003731629	5.312860261	PTS system HPr protein [Clostridium difficile 630]	34569.83	45956.03	34245.29	4020.826	9869.877	7711.814
11	gi|54781347	8	0.003987149	1.956788527	2-hydroxyisocaproate-CoA transferase [Clostridium dif	530107.9	427769.4	420032.8	232920.9	201308.1	269940.1
12	gi|126700078	6	0.007662556	1.358344492	molecular chaperone DnaK [Clostridium difficile 630]	169694.5	155660.9	140698	222159.5	210880.7	200020.8
13	gi|126697631	7	0.010228808	1.476686476	50S ribosomal protein L7/L12 [Clostridium difficile 630]	800237.5	686088.3	910574	559940.4	526226.6	536994
14	gi|254976387;gi|11496150;	63	0.010811893	1.316557403	cell surface protein (S-layer precursor protein) [Clostri	79867266	84107432	83973903	1.1E+08	97449336	1.19E+08
15	gi|126697969;gi|25565442	10	0.023675339	1.842240485	Beta-subunit of electron transfer flavoprotein [Clostric	2288685	2176419	3361065	1480229	1158617	1609334
16	gi|126700790;gi|29645204	10	0.148353122	1.752066307	enolase (2-phosphoglycerate dehydratase) (2-phospho	120504	70395.35	283290.2	282282	264544.4	283985.1
17											

Note: where there are multiple group members the other accession numbers are also exported.

At each stage in the Work flow there are a number of Export and Import options available from the **File** Menu. This includes the option to **Import Additional Protein Data** which can be used to increase the Protein meta data and also be used to sort the existing tabular data.



Exporting Protein Data to Pathways Tool(s)

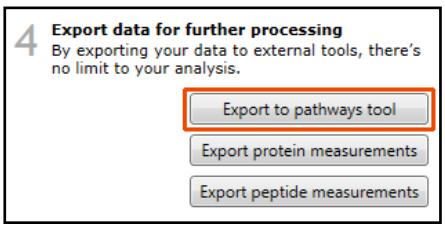
Using Progenesis, you can export protein lists to pathway analysis tools to help you understand your data in a wider biological context.

Currently Progenesis for proteomics supports the export to:

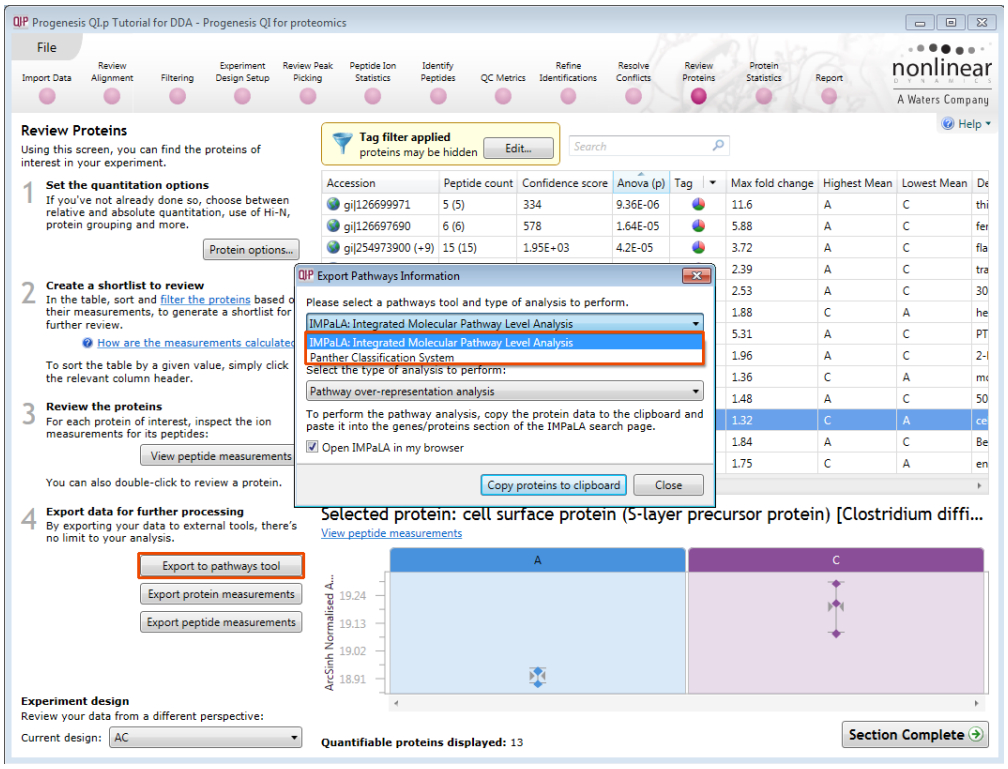
IMPaLA: which aggregates and queries many other pathway analysis tools including KEGG and Reactome)

PANTHER: classifies proteins for high-throughput analysis.

Note: Plugins for these tools are provided as standard.

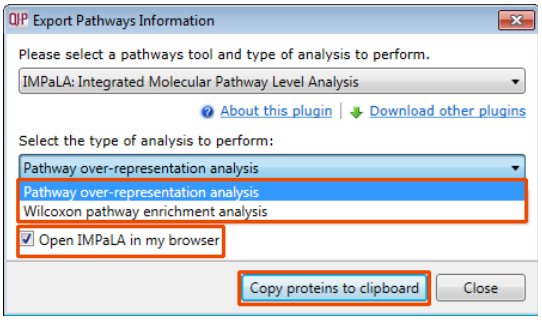


First select the protein data to export to the pathways tool using tag filtering to 'focus' the set to export. Then click **Export to pathways tool**, and select the tool followed by the test to be performed

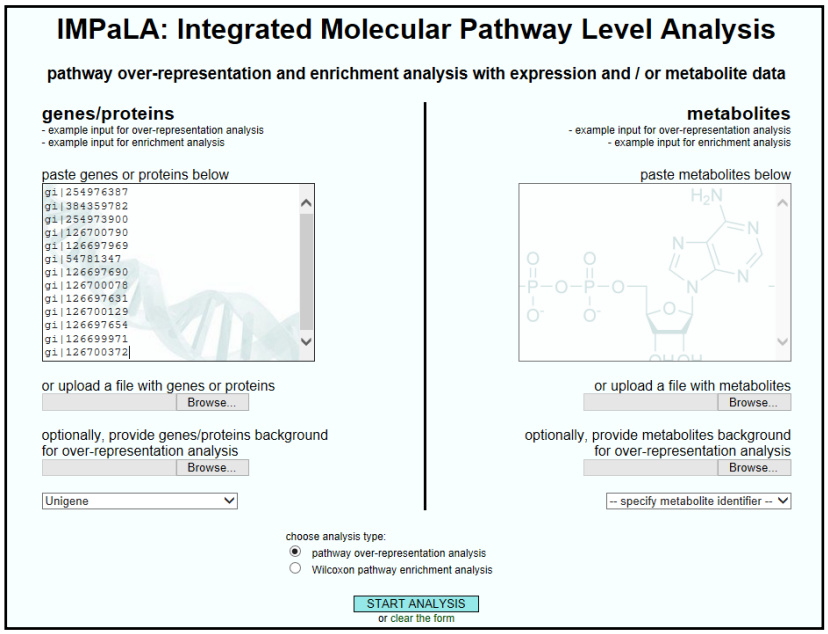


Select either **Pathway over-representation analysis** or **Wilcoxon pathway enrichment analysis.**

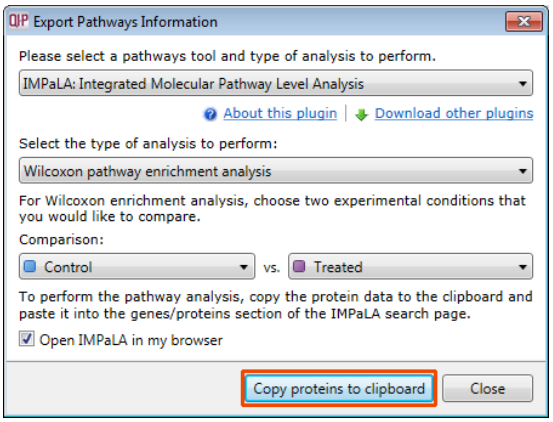
Make sure the **Open IMPaLA in my browser** is ticked and then click **Copy proteins to clipboard**



When IMPaLA opens paste in the exported values and perform the test.



Note: currently the Clostridium difficile protein set being used in this user guide does not yield any pathway 'hits' with Impala or Panther.

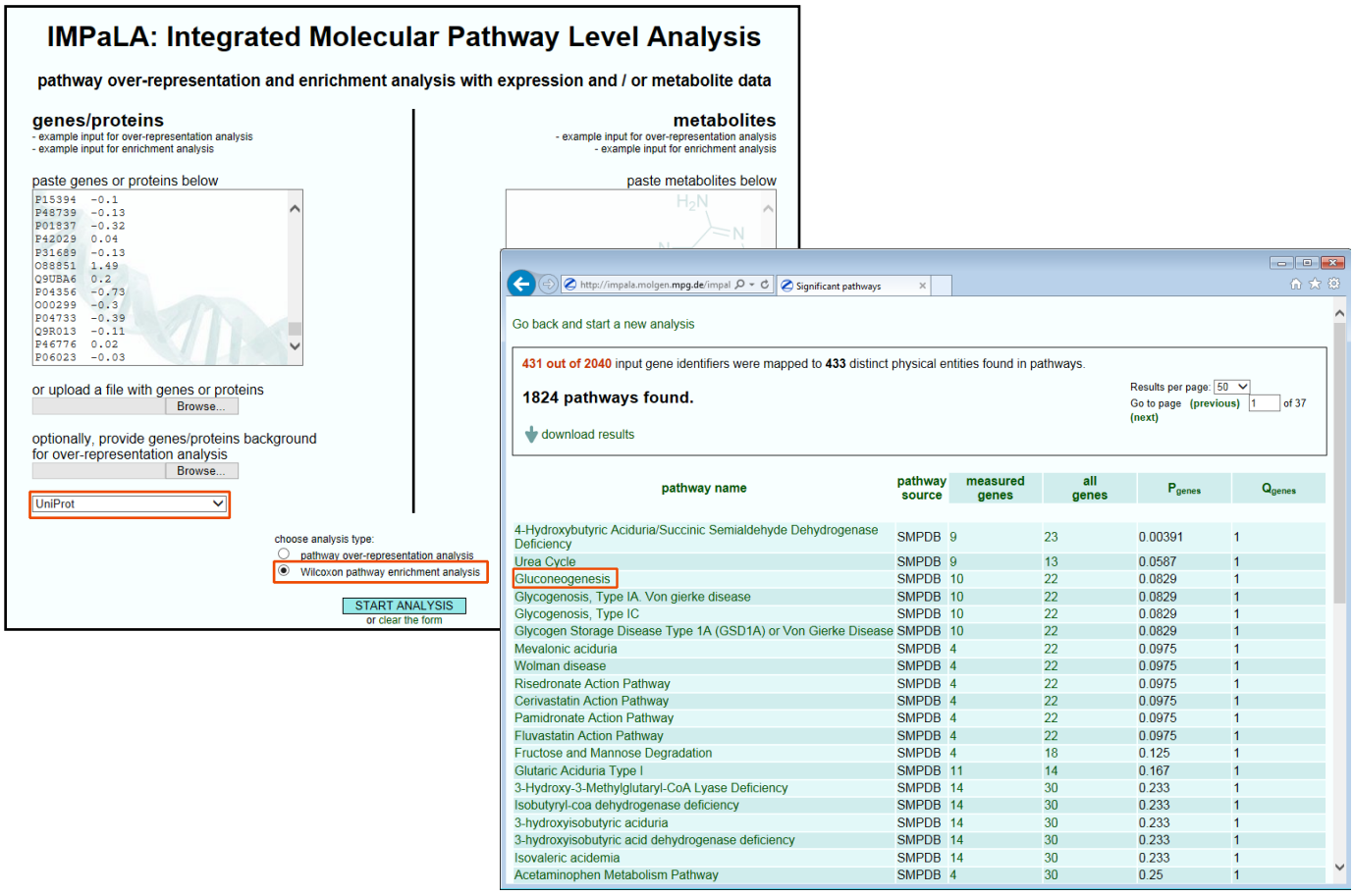


Although the previous page outlines the process of exporting data to pathway tools, a text file (**Impala_Wilcoxon_demo.txt**) has been included in the tutorial download to allow the demonstration of exporting data to Impala for pathways analysis.

In Progenesis QI for proteomics set up the **Export to pathways tool** dialog to export the data to Impala to perform a Wilcoxon pathway enrichment analysis. Click copy to clipboard to open Impala

To **simulate** the **Copy proteins to clipboard** open the text file in 'Notepad', select all and copy. In the Impala window, open in your browser, right click on the genes/protein panel and paste the contents of the file Impala_Wilcoxon_demo.txt.

Select UniProt as the identifier, tick Wilcoxon pathway enrichment analysis then Start Analysis.



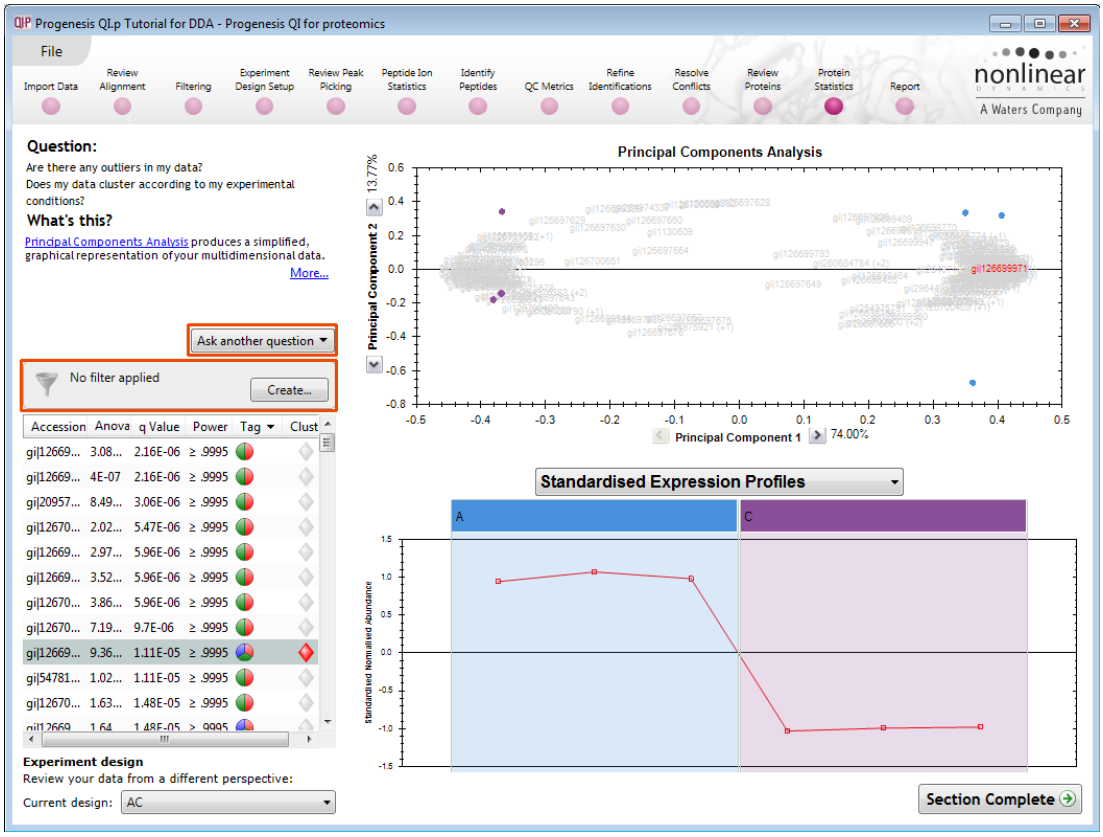
Note: if you are using your own data then pressing **Copy proteins to clipboard** will open **Impala** and allow you to paste directly into Impala, without saving to a separate file.

Note: in the case of exporting to **Panther** the normal process requires a file to be created and stored then opened in Panther to perform the pathway analysis.

Clear any applied tag filters before moving to the **Protein Statistics** section to report on Proteins and /or peptides.

Stage 15: Protein Statistics

Protein Statistics opens with a Principal Components Analysis (PCA) for all the proteins displayed.

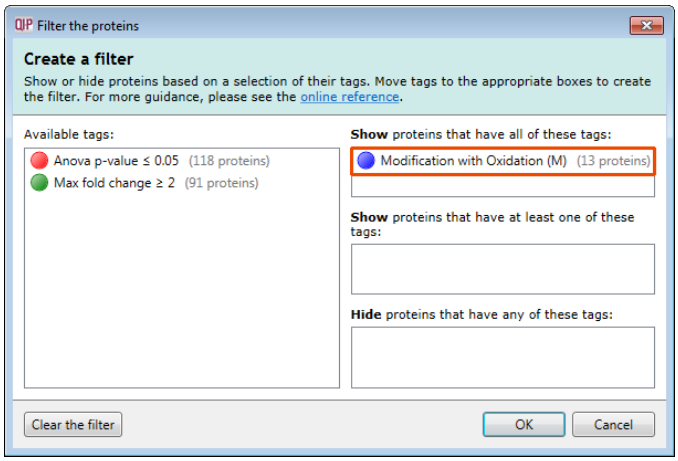


The Multivariate Stats can now be applied to all or subsets of proteins as determined by the current Tag filters (in this example none). Allowing you to identify similar paterns of expression using the Correlation Analysis.

Now move to the **Report** section to report on Proteins and /or peptides.

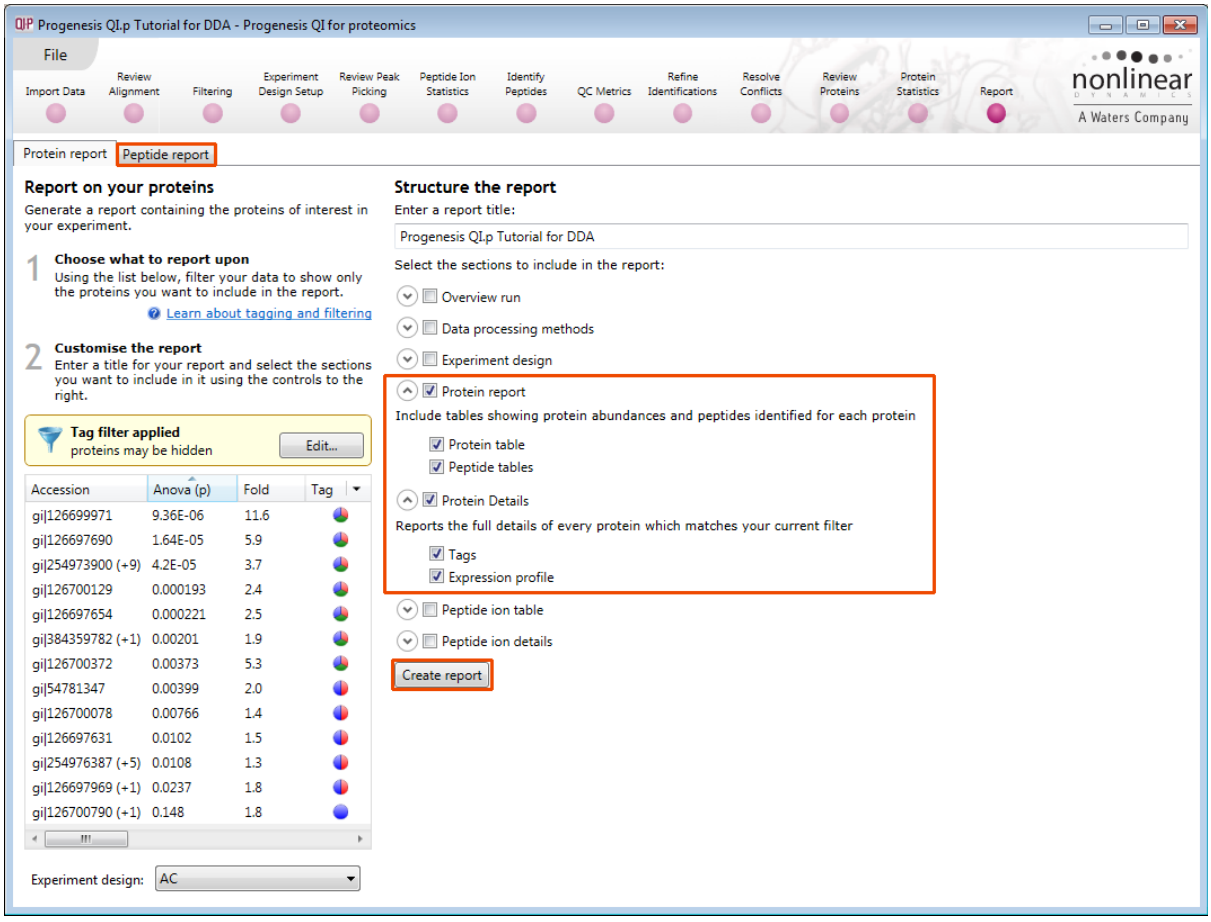
Stage 16: Reporting

The **Report Design** stage allows you to select what views you want to include in a report based on the list of **currently selected proteins.**



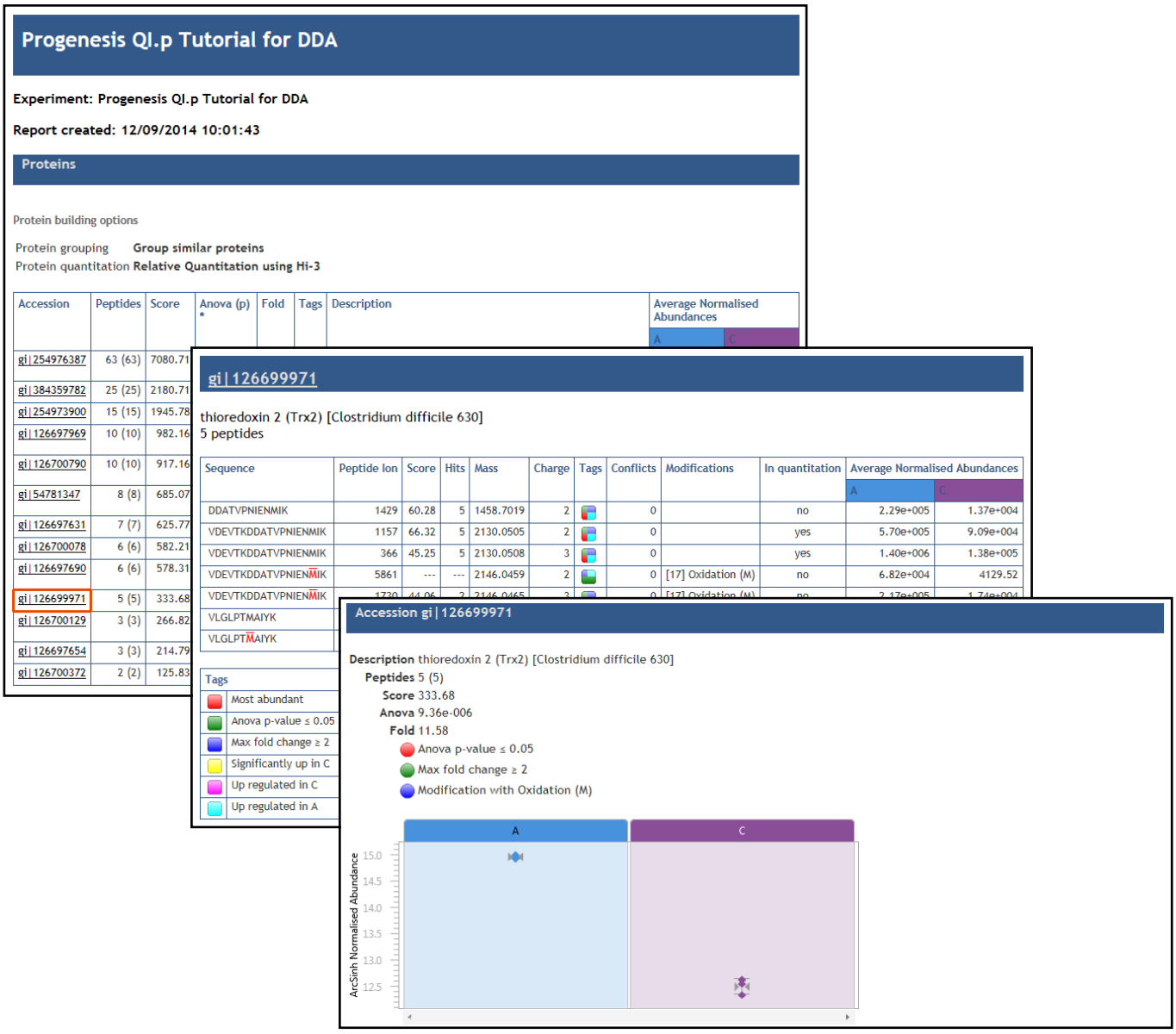
As an example we will create a report for **only** the proteins containing peptides with Oxidation of Methionine residues enter **Oxidation (M).**

1. First reduce the proteins to report on by selecting the '**Modification with Oxidation (M)**' tag. In this example it reduces the number of proteins in the table to 13.
2. Expand the various Report Design options (by default they are all selected)
3. Un-tick as shown below
4. Click **Create Report**



This opens a dialog to allow you to save the report, after which it will be opened in the form of a web page.

Click on the **Accession No**. in the proteins section of the Report and this will take you to the Assigned peptides for this protein



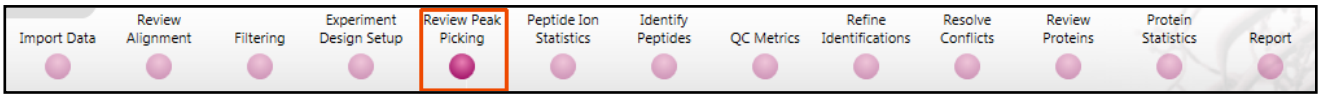
Having closed the report it can be reopened by double clicking on the saved html file.

Note: you can also copy and paste all or selected sections of the report to Excel and/or Word.

Creating an Inclusion list

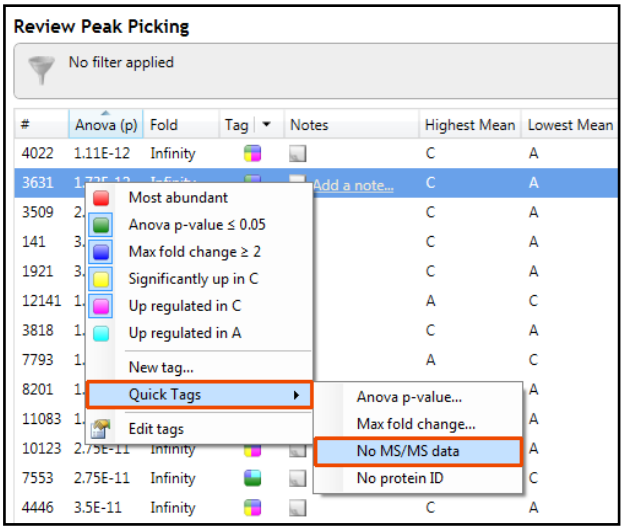
Inclusion lists can be used to try to increase the number of identified peptides you have in your experiment. They are used to control your mass spectrometer to try and concentrate the collection of MS/MS scans from specified m/z – retention time positions. Runs made using the inclusion list can then be imported into Progenesis QI for proteomics and the extra MS/MS scans added to the experiment.

As an example we are going to create an inclusion list for all the peptide ions that show a Significant difference between groups A and C (Anova p<0.05) and have **no** MS/MS spectra.

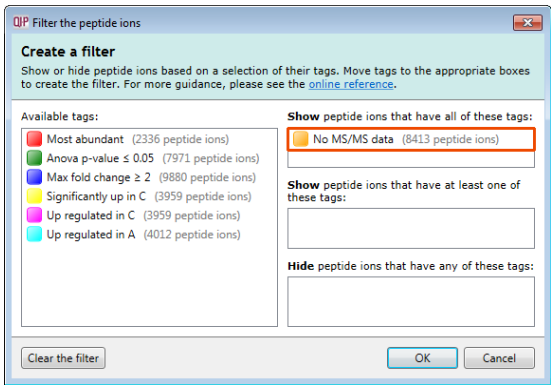


First return to **Review Peak Picking** using the Workflow icons.

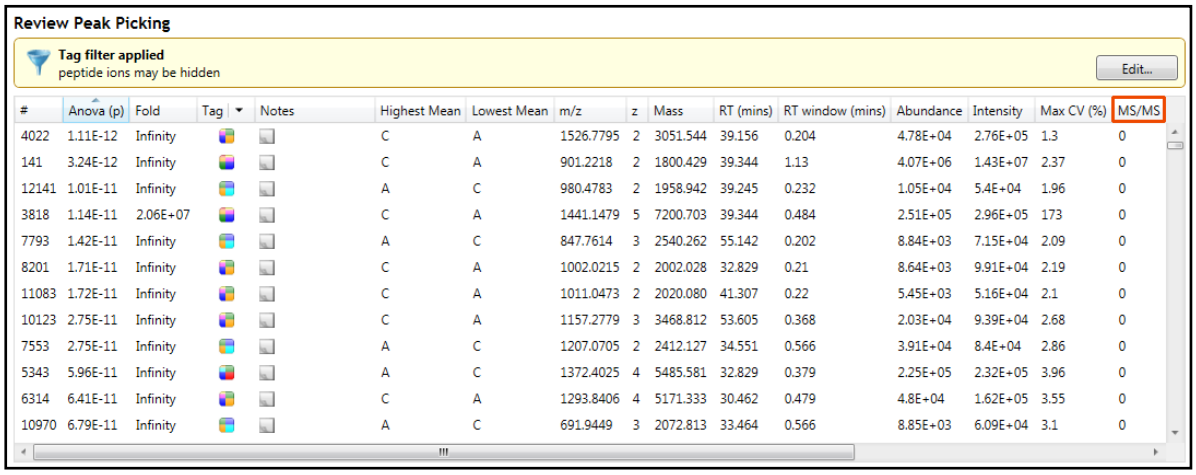
With no filters applied right click on a peptide ion in the table, select **Quick tags** and click on **No MS/MS data**.



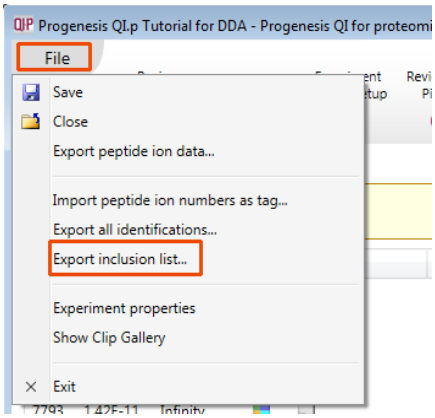
Filter the table so that it is only showing peptide ions with **No MS/MS data**. To do this click **Create** and drag the tag on to the Show peptide ions that have all these tags and click OK.



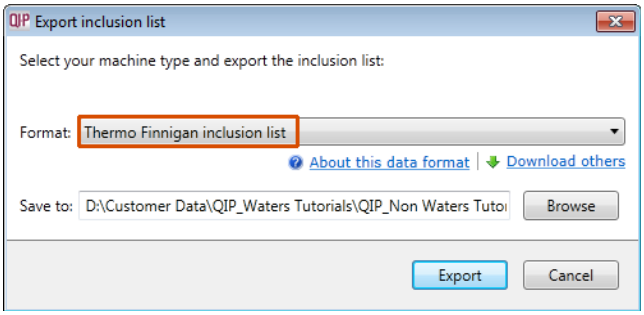
The table will now only be displaying peptide ions that have no ms/ms.



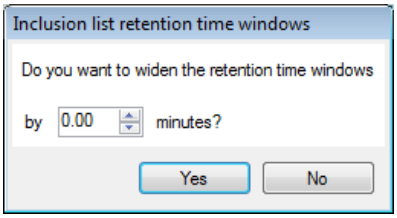
Now select **Export Inclusion List...** from the file menu



Then select the appropriate format.



Finally export the file to an appropriate location



Note: with certain MS machines it is possible to widen the retention time windows being used, this can be controlled using the following dialog.

If you require further information on the inclusion list file formats then click **Help.**

Note: The new LC-MS runs can then be added to the original experiment to increase the MS/MS coverage using the **Add files** facility at the Data Import Stage.

Congratulations!

This document has taken you through a complete analysis using Progenesis QI for proteomics, from Alignment through the analysis workflow to generate lists of proteins exhibiting expression changes across biological conditions using powerful Multivariate Statistical analysis of the data.

Hopefully our example has shown you how this unique technology can deliver significant benefits with

- Speed
- Objectivity
- Statistical Power

If you would like to see the benefits of running Progenesis QI for proteomics using your own data and explore the Progenesis QI for proteomics workflow please go to Appendix 4: Licensing Runs (page 76).

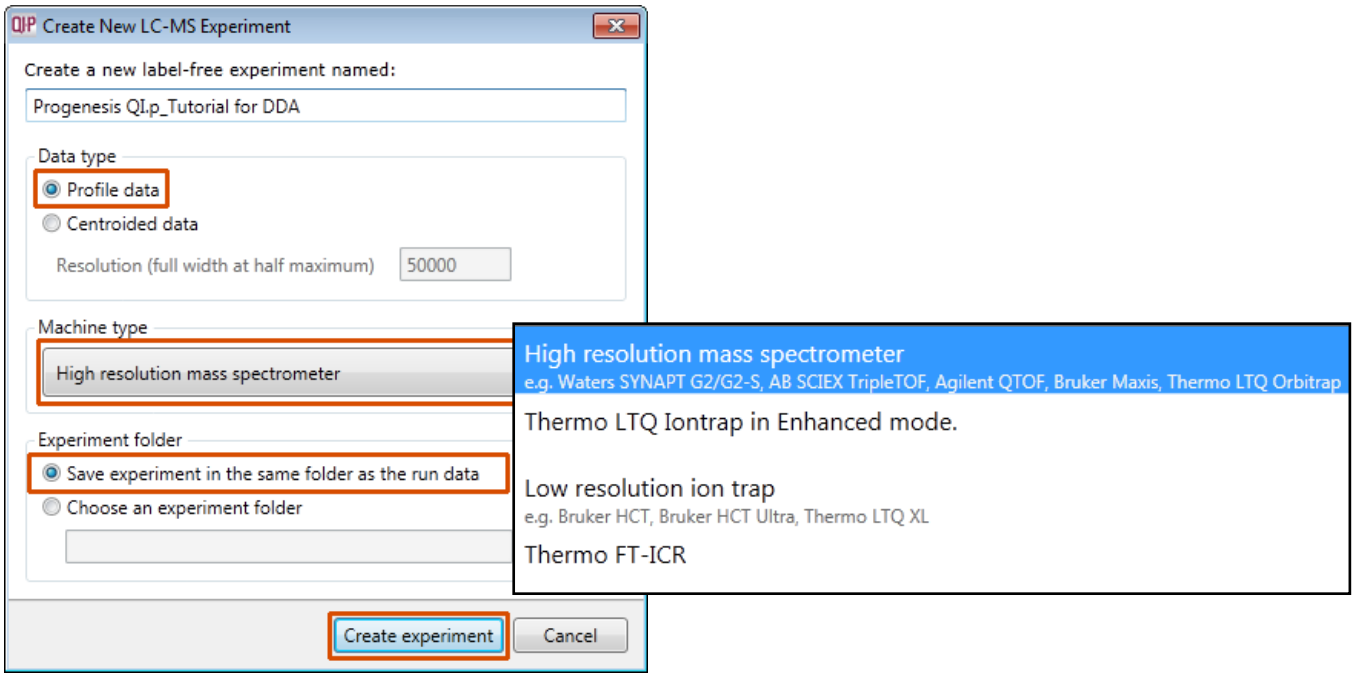
Appendix 1: Stage 1 Data Import and QC review of LC-MS data set

You can use your own data files, either by directly loading the raw files (Waters, AB SCIEX, Agilent, Bruker and Thermo) or, for other Vendors, convert them to mzXML or mzML format first.

To create a new experiment with your files: open Progenesis QI for proteomics and click **New,** bottom left of the **Experiments** page and give your experiment a name. Then select data type, the default is 'Profile data'.

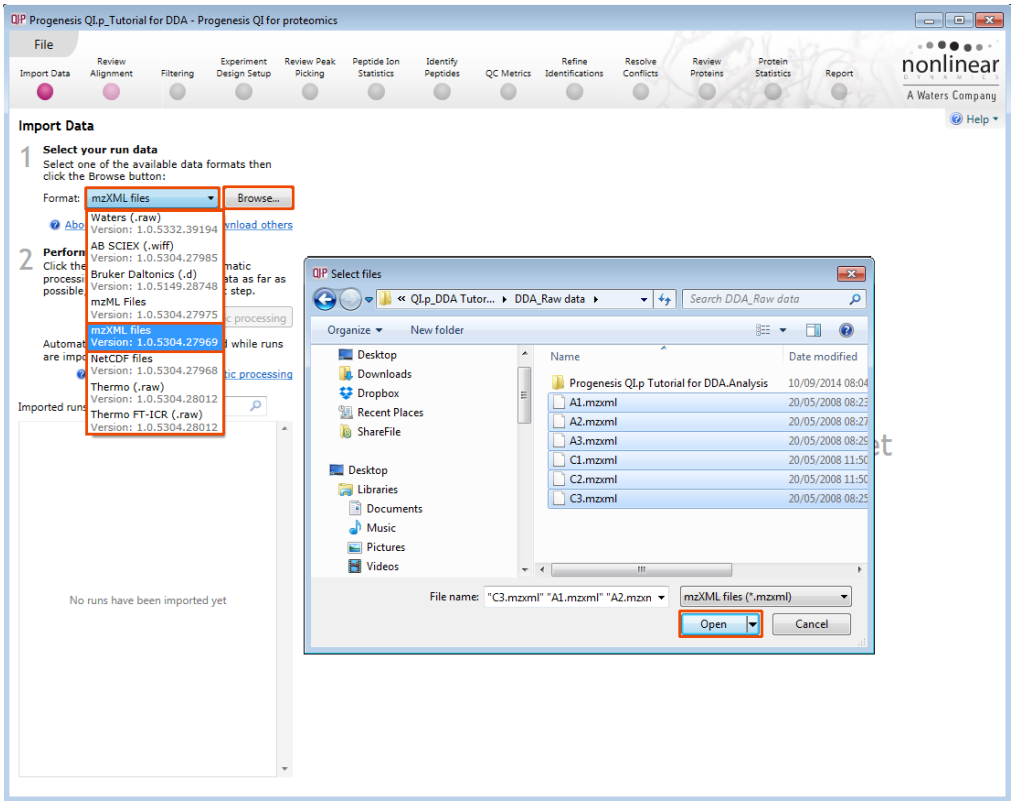
Note: if you have converted or captured the data as centroided then select Centroided data and enter the Resolution for the MS machine used.

Click **Create experiment** to open the LC-MS Data Import stage of the workflow.



Select the 'Import Data file format', in this example they are mzXML files

Then locate your data files using **Import...**

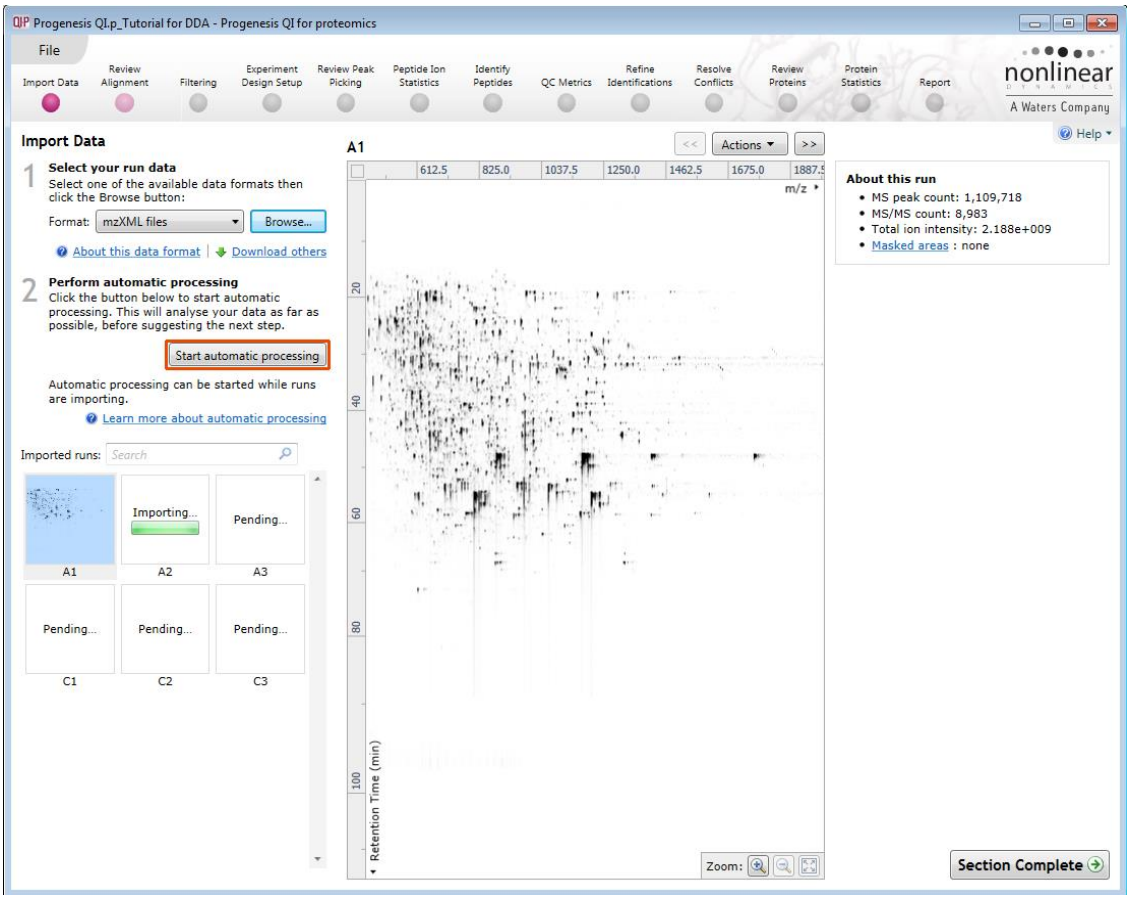


Locate and select all the Data files (A1 to C3).

On loading the selected runs your data set will be automatically examined and the size of each file will be reduced by a 'data reduction routine', which reduces the data by several orders of magnitude but still retains all the relevant quantitation and positional information.

Note: For a large number of files this may take some time.

Note: as the loading process starts you can also start the automatic processing before the loading has completed.



For details of setting up the steps in the automatic processing wizard return to Stage 2A page 7.

Additional details on Selection of Alignment Reference

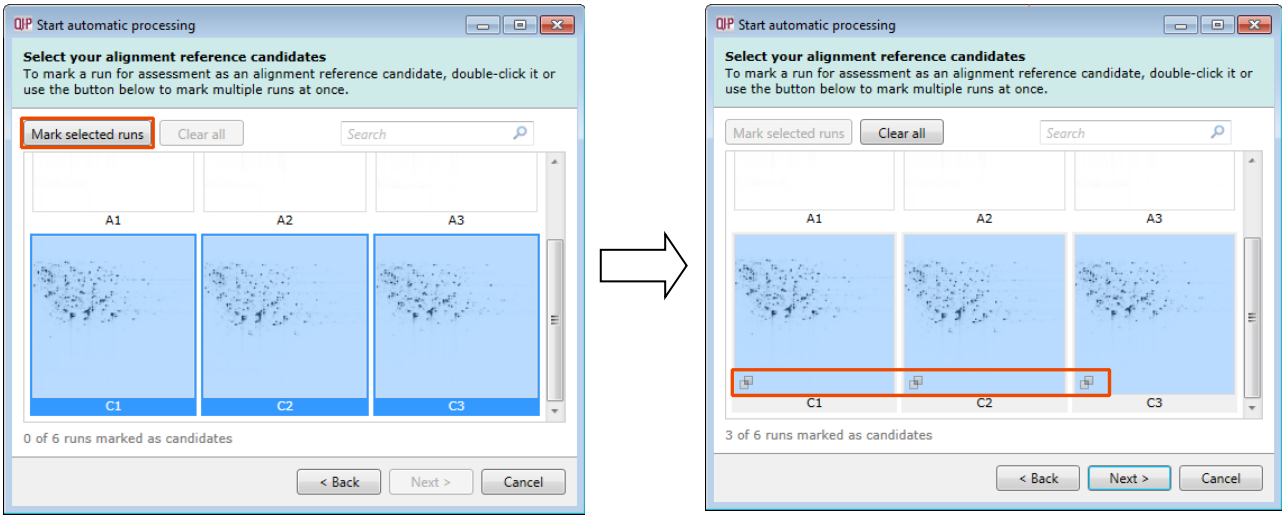
1. **Assess all runs in the experiment for suitability**

- This method compares every run in your experiment to every other run for similarity.
- The run with the greatest similarity to all other runs is chosen as the alignment reference.

If you have no prior knowledge about which of your runs would make a good reference, then this choice will normally produce a good alignment reference for you. This method can take a long time.

2. **Use the most suitable run from candidates that I select**

- This method asks you to choose a selection of reference candidates, and the automatic algorithm chooses the best reference from these runs.



Select a subset of your runs as 'alignment reference candidates' and click **Mark selected runs.** An icon appears on the bottom left of each run to indicate that it is selected as a candidate.

This method most appropriate when you have some prior knowledge of your runs suitability as references:

- runs from pooled samples
- runs for one of your experimental conditions will contain the largest set of common peptides.

Also preferred option when analysing a large number of runs

3. **Use this run**

- This method allows you to manually choose the reference run.

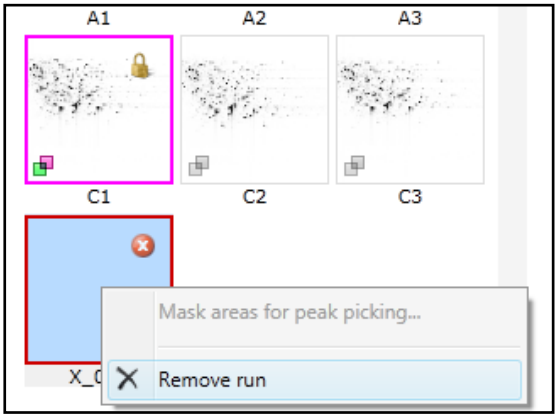
Manual selection gives you full control, but there are a couple of risks to note:

- If you choose a pending run which subsequently fails to load, alignment will not be performed.
- If you choose a run before it fully loads, and it turns out to have chromatography issues, alignment will be negatively affected (for this reason we recommend that you let your reference run fully load and assess it's chromatography before loading further runs).

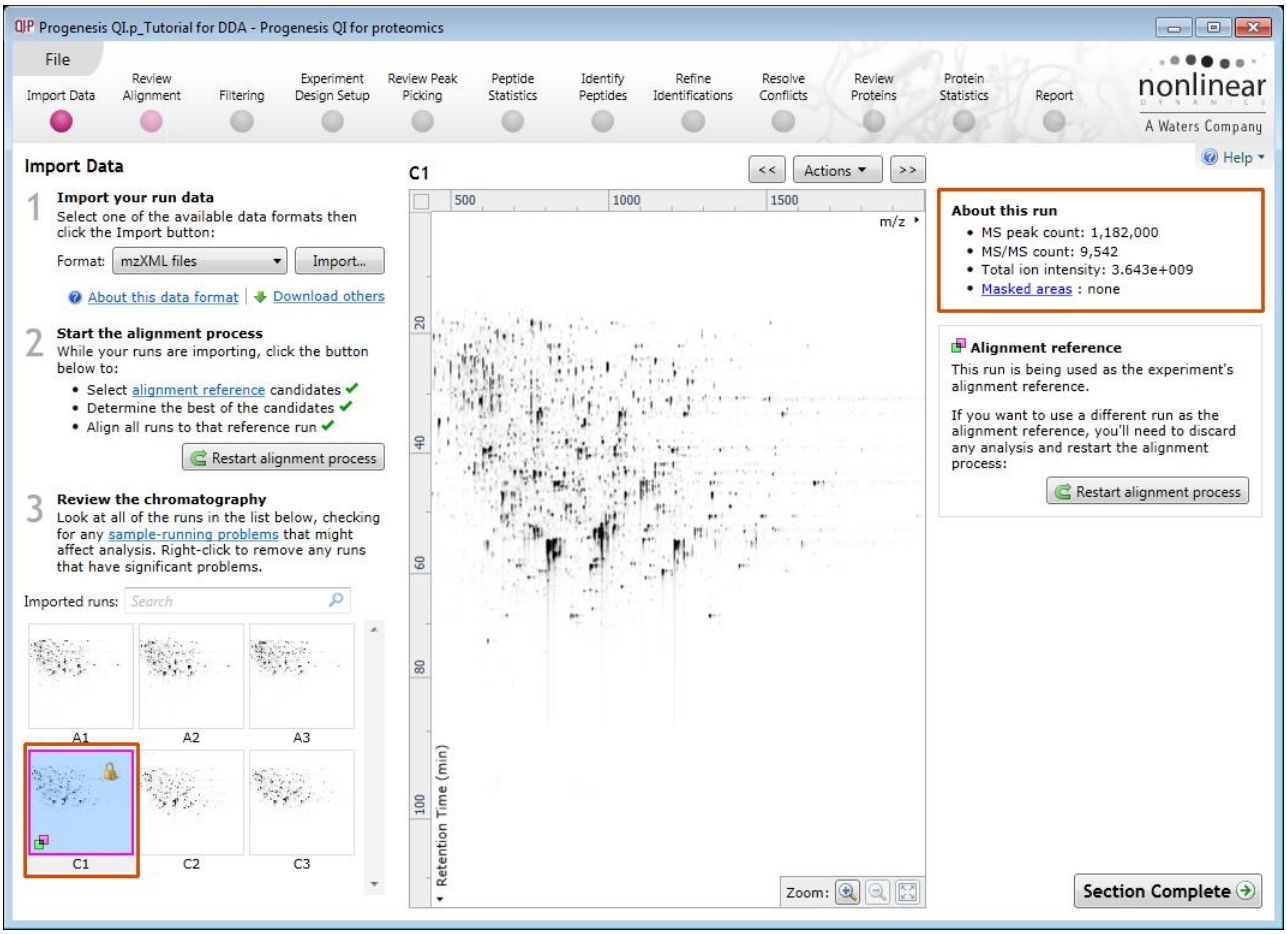
Review Chromatography

Each data file appears as a 2D representation of the run. At this stage you will be warned if any of the data files have been 'centroided' during the data acquisition and conversion process.

You can delete run(s) by left clicking on the run in the list.



At the Import Data stage you can examine the quality of the imported runs using the 2D representation of the runs



Note: details of the current run appear on the top right of the view.

Once you have reviewed the imported runs click on Review Alignment or Section Complete to move forward to the Review Alignment Stage.

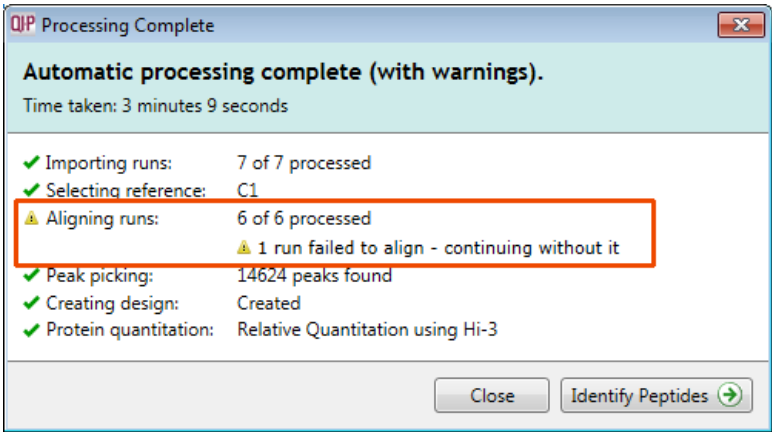
Note: you will be offered the automatic alignment if you have not performed it automatically already.

Now move to the next stage in the workflow (page 14 in this user guide) by clicking **Section Complete.**

Appendix 2: Stage 1 Processing failures

If a stage fails to complete successfully or only partially completes, the automatic processing dialog will warn you of the problem. This may or may not allow the automatic processing to complete.

For example: a run that fails to automatically align will trigger a warning, although analysis will continue; the automatic processing dialog will prompt you to 'drop-off' at the **Review Alignment** stage on completion to investigate the problem.

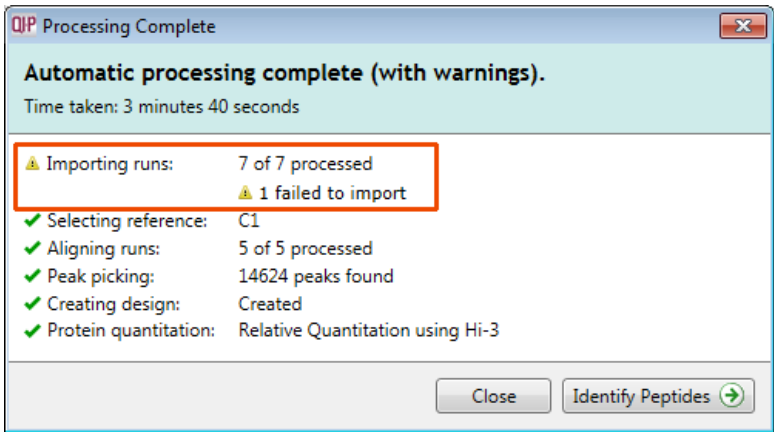


Note: in this example the run that failed to align will not contribute to the peak picking and will be excluded at the alignment stage (a cross appears in the include column).

You can either remove the run from the experiment at the **Import Data** or add it back in at the **Review Alignment** stage once the alignment of the run has been corrected.

As another example, runs that import successfully but with warnings at the **Import Data** stage will cause a flag in the readout to notify you of the potential quality issue.

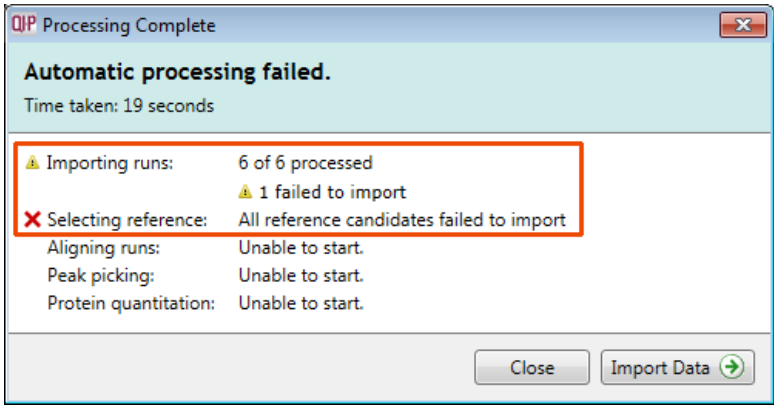
If some runs fail to import (but not all), the automatic processing will continue, informing you that one or more runs have failed to import.



In this case you can remove the runs at Import Data and if appropriate replace them with additional runs.

Note: adding additional runs will then be aligned and peak picking should be re-done to include data from the added runs in the generation of the aggregate

An example of a problem that would halt the automatic processing would be the failure to successfully import all the potential reference candidates, (for example, where in importing, you specified selection of the alignment reference from several runs before they were fully imported and set the processing underway, and they later failed to import owing to problems with the runs).



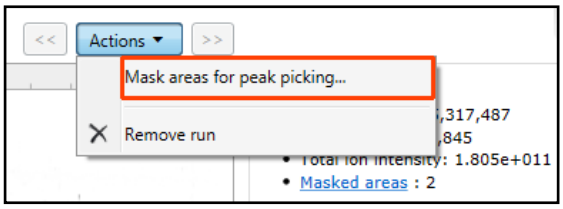
In this case, the processing dialog would halt and prompt you to select another reference.

Appendix 3: Stage 1 Data QC review and addition of exclusion areas

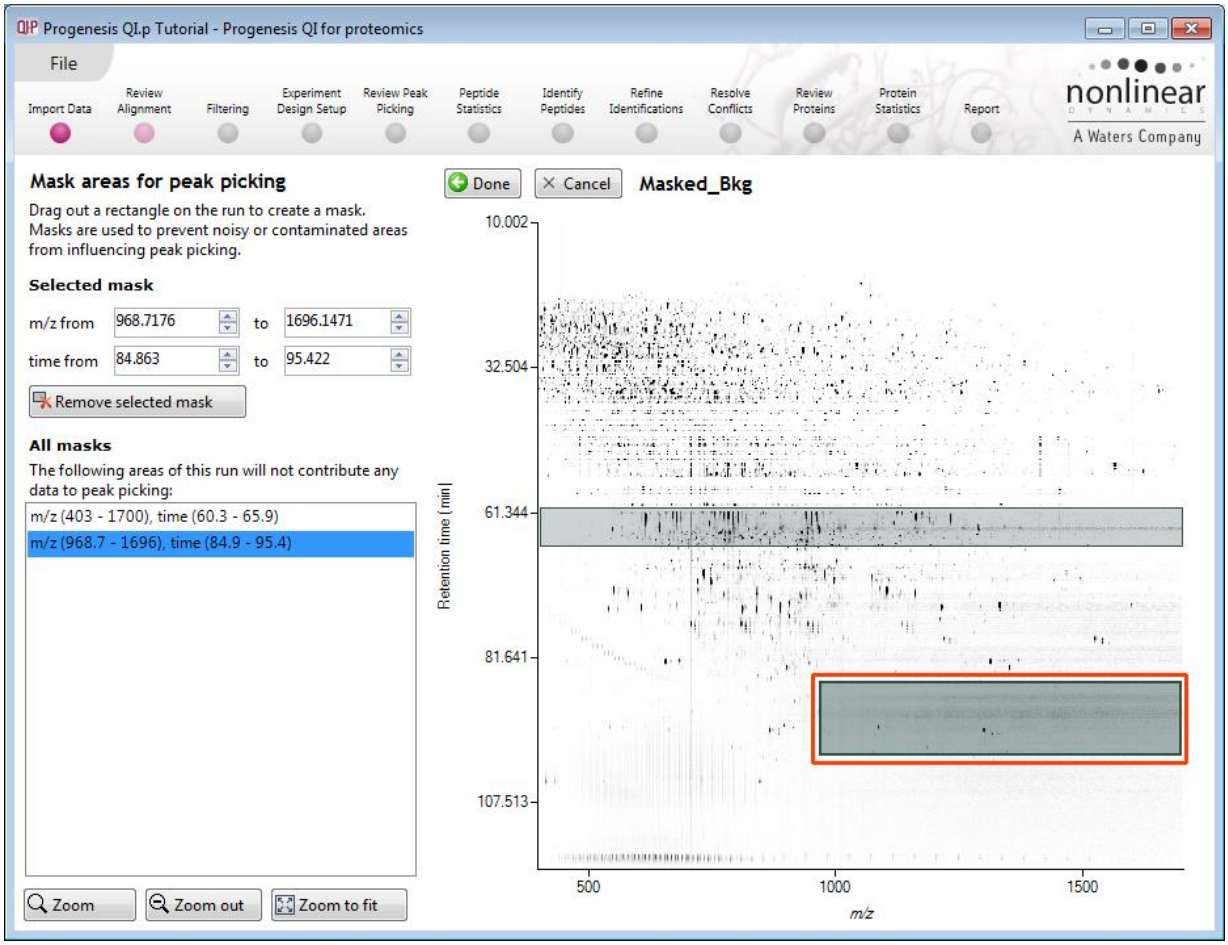
During the process of Data QC you may identify areas of the raw data for a particular run that appear 'noisy' yet still have identifiable 'isotopic patterns'.

For example if the run is part of a 'replicate set' of runs it is possible to exclude such areas on the noisy run by applying a mask to the area. By doing so this area is excluded during the initial part of the detection process in order that it does not 'interfere' with the detection of the peptide ions in the replicate group.

To do this select Masked areas from selected run on the bottom left of the screen.



Drag out an area over the noisy part of the run to create the mask.



Note: Click **Done** to return to the **Import Data** view where you can zoom into the masked where you will see the isotopic peptide ions in the noise.

Note: if the level of noise is high and affecting many of your runs a preferred approach would be to re-optimise the chromatography to improve the levels of noise in your data.

Appendix 4: Licensing runs (Stage 3)

When setting up a **New experiment** if you are evaluating Progenesis QI for proteomics with unlicensed runs then the licensing page will open after **Import Data section**.

If you already have a programmed dongle attached to your machine then the License Runs page will not appear.

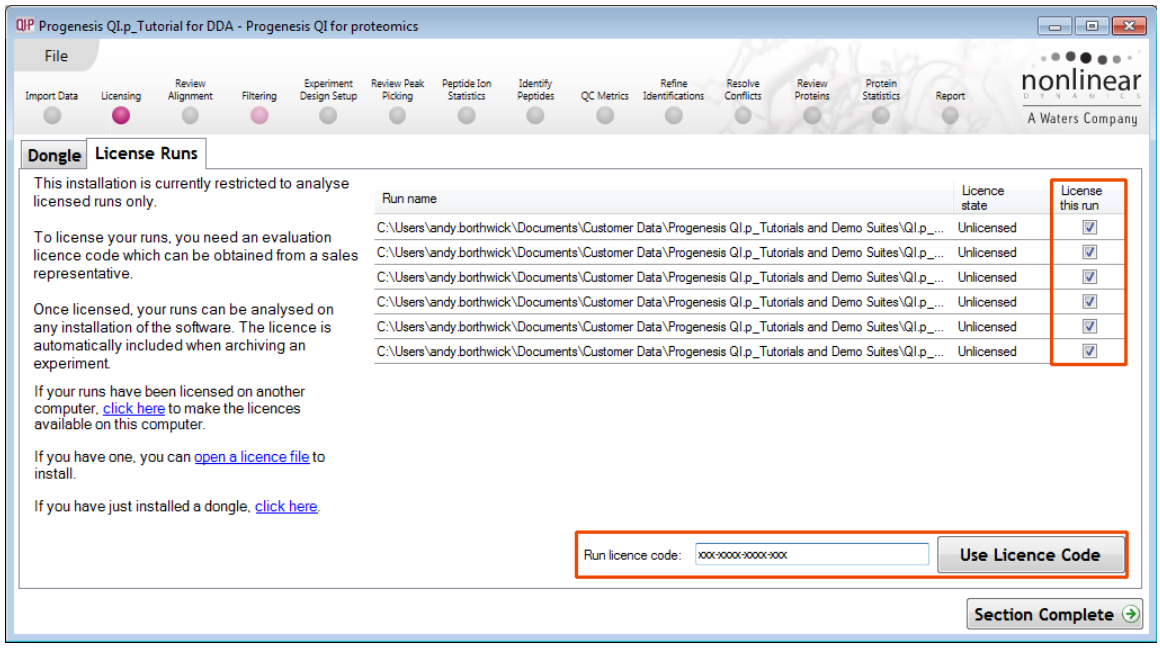
To use this page to License your Runs you must first either obtain an **'Evaluation' Licence Code from a Sales Person or purchase a licence code directly**.

Each code will allow you to license a set number of runs.

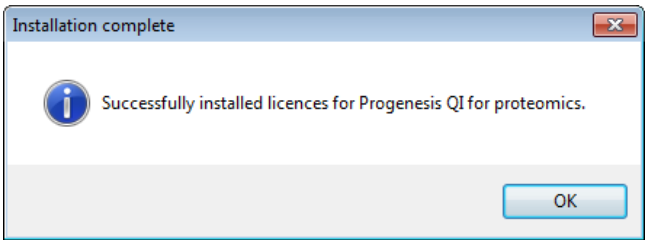
The runs in your experiment will be listed as shown below.

To activate license(s) for the selected runs enter the code in the space provided and click **Use Licence code**.

Note: you will need an internet connection to use this method.



A message confirming successful installation of your licences will appear.



Click **OK**, the view will update and Alignment, the next stage in the workflow, will open with the licensed files.

Appendix 5: Manual assistance of Alignment

Approach to alignment

To place manual alignment vectors on a run (C2 in this example):

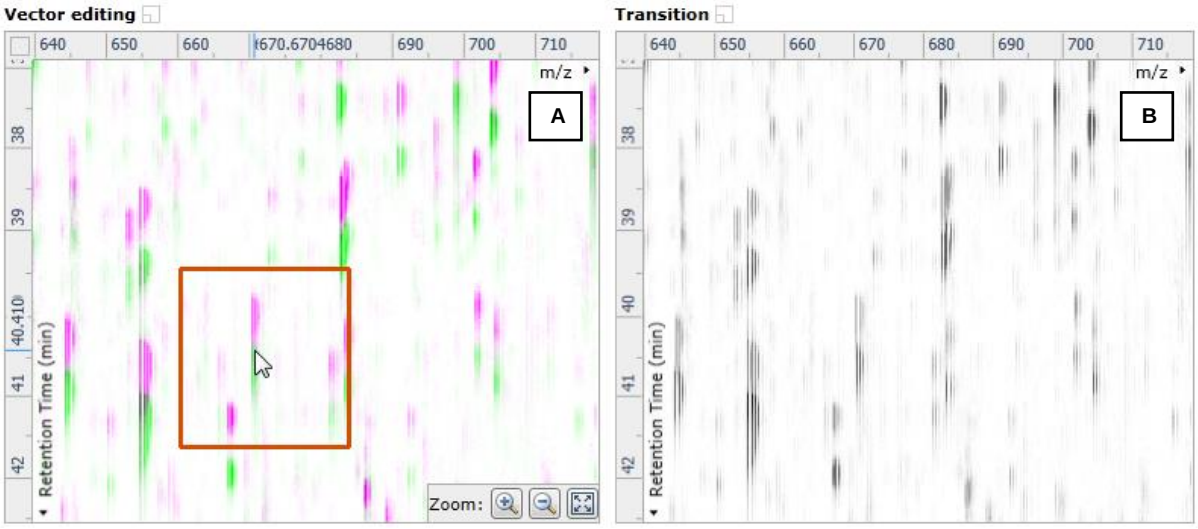
1. Click on Run C2 in the **Runs** panel, this will be highlighted in green and the reference run (C1) will be highlighted in magenta.
2. You will need approximately 5 **alignment vectors** evenly distributed from top to bottom of the whole run.
3. First drag out an area on the **Ion Intensity Map** (C), this will reset the other 3 windows to display the same 'zoomed' area

Note: the peptide ions moving back and forwards between the 2 runs in the **Transition** window (B) indicates the misalignment of the two runs.

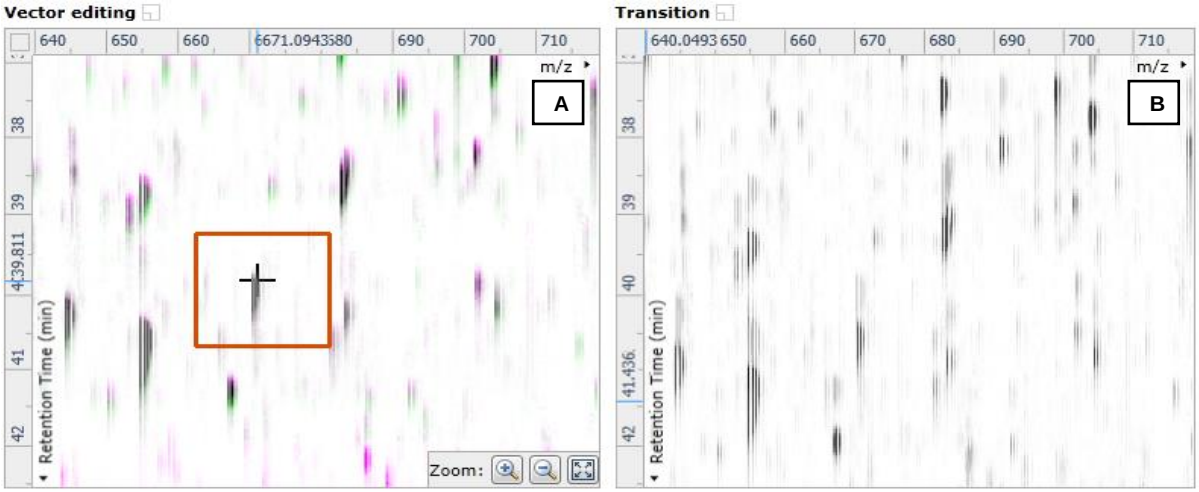
Note: the Ion Intensity Map gives you a colour metric, visually scoring the current alignment. The overall score is placed next to the Vectors column in the table. With each additional vector added this score will update to reflect the 'changing' overall quality of the alignment. The colour coding on the Ion intensity Map will also update with each additional vector.

Note: The **Total Ion Chromatograms** window (D) also reflects the misalignment of the 2 runs for the current Retention Time range (vertical dimension of the current Focus grid in the **Ion Intensity Map** window).

4. Click and hold on a green peptide ion in Window A as shown below

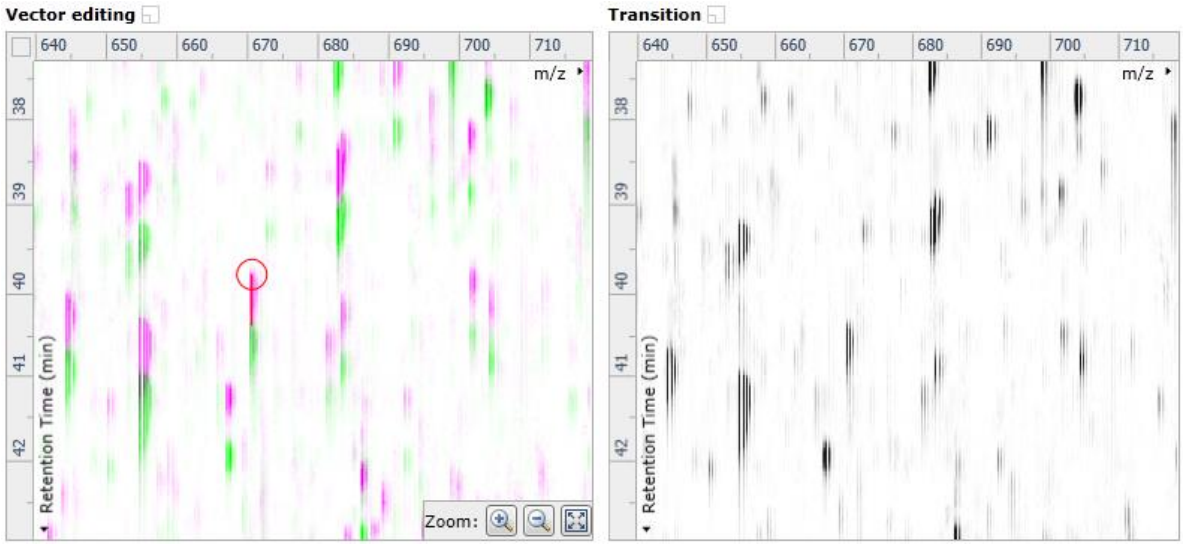


5. As you are holding down the left mouse button (depending on the severity of the misalignment), the alignment vector will automatically find the correct lock. If not, drag the green peptide ion over the corresponding magenta peptide ion of the reference run. The red box will appear as shown below indicating that a positional lock has been found for the overlapping peptide ions.



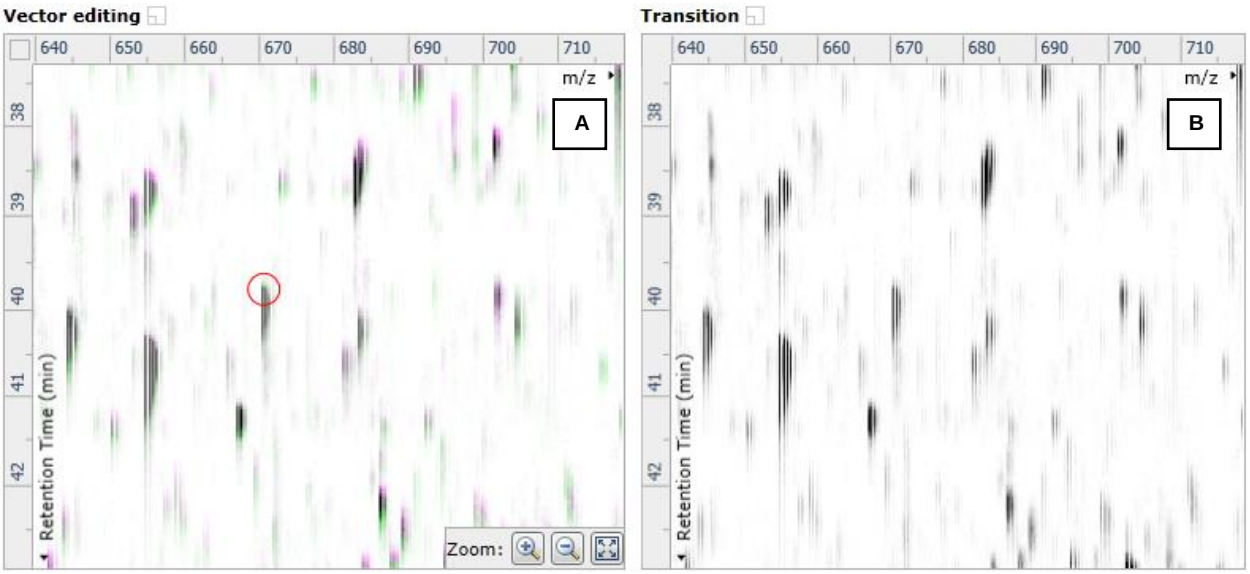
***Tip*:** while holding down the mouse button hold down the **Alt** key. This will allow smooth movement of the cursor as the **Alt** key allows you to override the 'automatic alignment' performed as you depress the mouse button.

6. On releasing the left mouse button the view will 'bounce' back and a red vector, starting in the green peptide ion and finishing in the magenta peptide ion will appear.

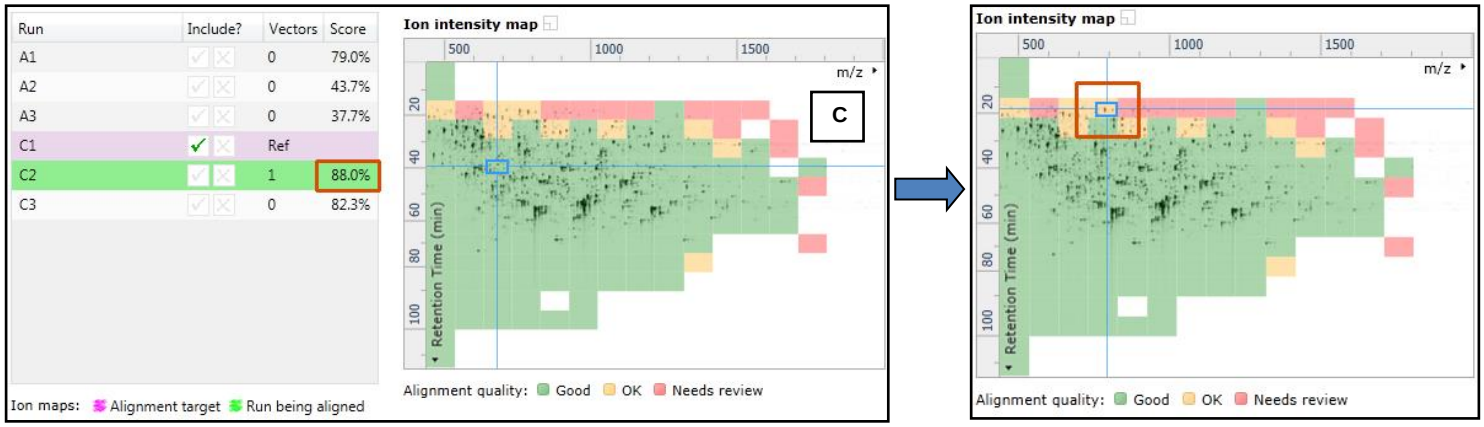


Note: an incorrectly placed vector is removed by right clicking on it in the **Vector Editing** window

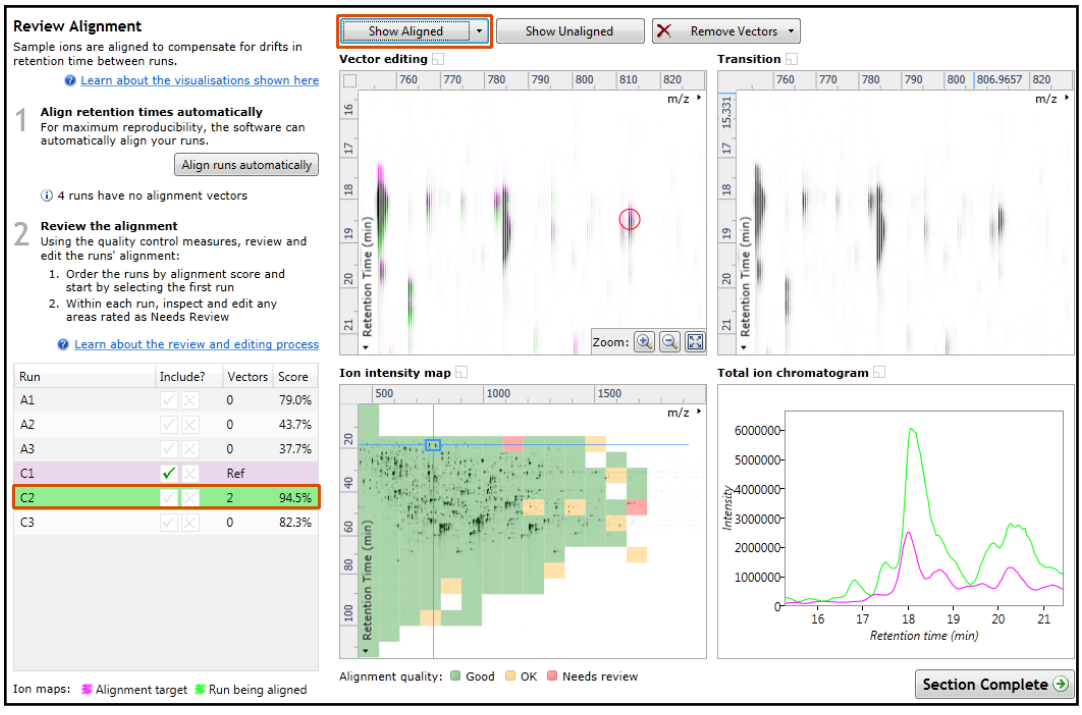
7. Now click **Show Aligned** on the top tool bar to see the effect of adding a single vector.



8. With the placement of a single manual vector the increase in the proportion of the **Ion Intensity Map** (C) showing green is reflected in the improved alignment score in the table. Now click in the Ion Intensity Map to relocate the focus in order to place the next manual vector.



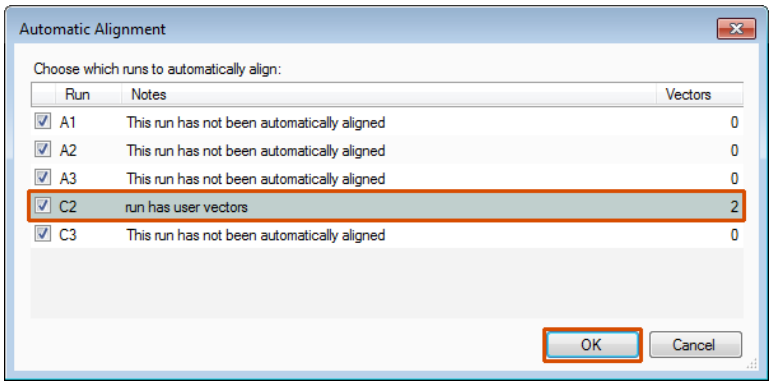
9. Adding an additional vector will improve the alignment further as shown below.



10. The shift in the Retention Time (RT) is as a result of incorrect running of the chromatography. In many of these cases if the Automatic Alignment fails to generate a good alignment then removing all the alignment vectors for this run and placing a **single** manual vector to act as a 'seed' for the Automatic Alignment algorithm maybe all that is required to generate a good alignment.

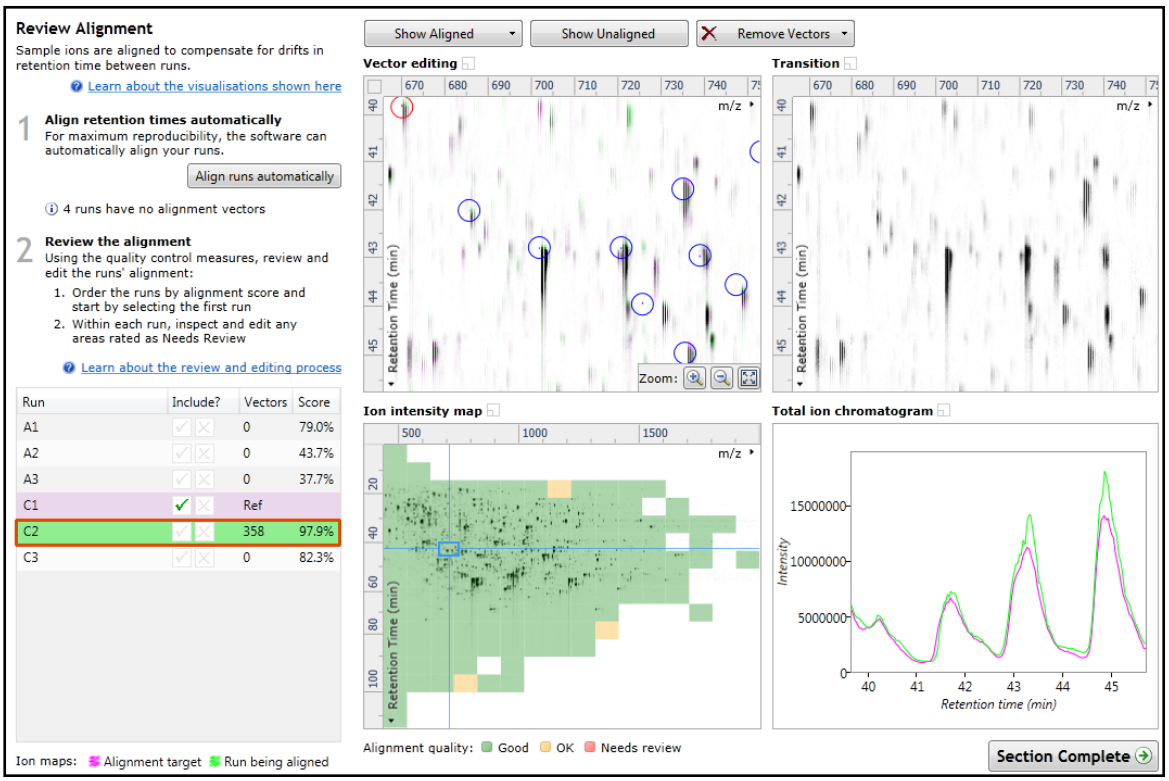
11. In the case of the example shown above placing 2 vectors from the top to the bottom of the run is sufficient to markedly improve the alignment (**Note**: the improvement in the score with the addition of only 2 vectors).

12. At this point you would redo the automatic alignment of this image by selecting **Align runs automatically**. **Note**: if you are focusing only on the alignment of one run, then un-tick the other runs in the alignment dialog.



The number of manual vectors that you add at this stage is dependent on the misalignment between the current run and the Reference run.

Note: In many cases only using the Automatic vector wizard will achieve the alignment.



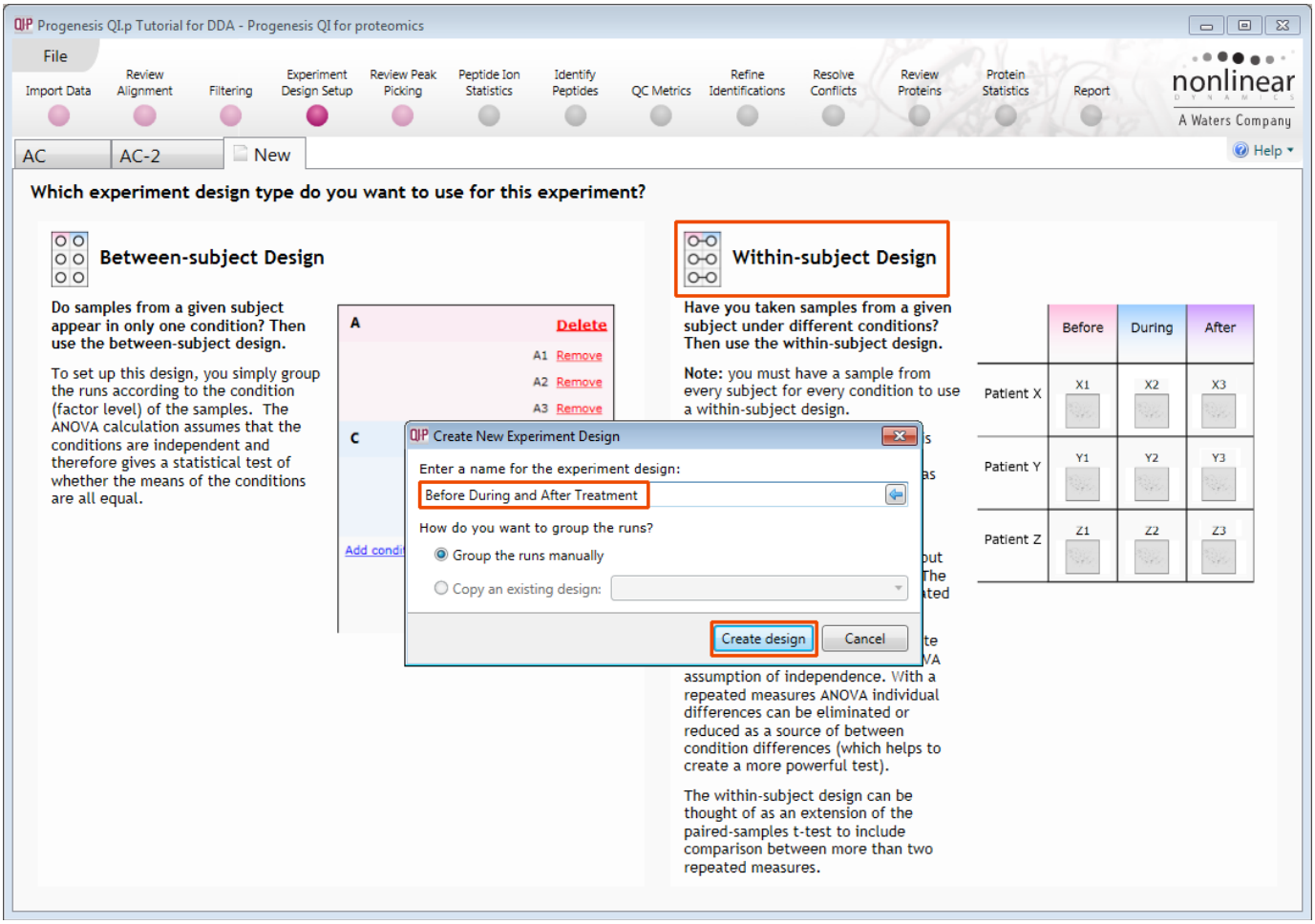
Tip: a normal alignment strategy would be: to run the automatic alignment first for all runs, then order the alignments based on score. For low scoring alignments remove all the vectors and place 1 to 5 manual vectors to increase the score then perform automatic alignment. Then review the improved alignment score.

To review the vectors automatic and manual, return to page 15.

Appendix 6: Within-subject Design

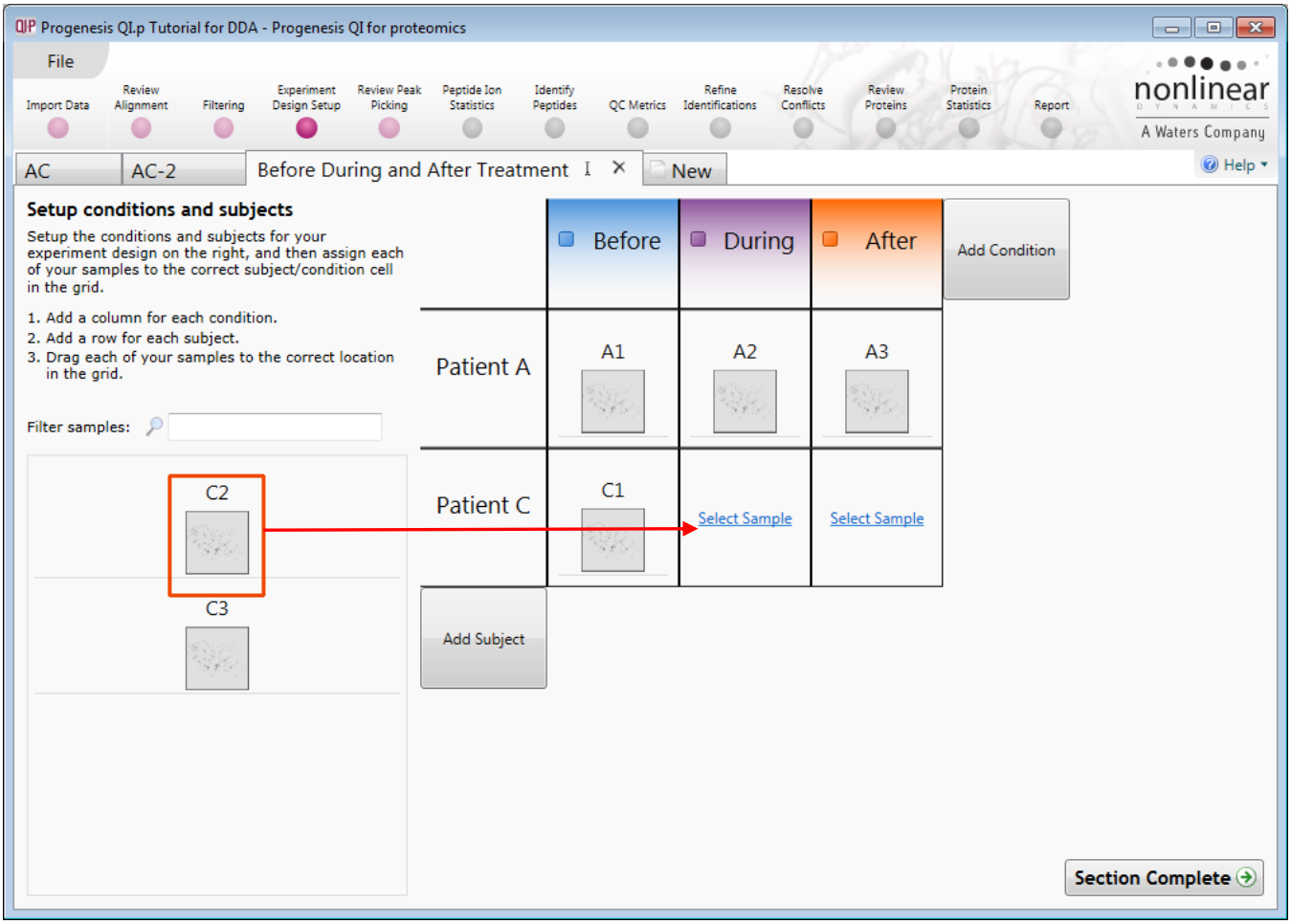
To create a **Within-subject Design** for your data set select this option on the **Experiment Design Setup** page and enter the name of the design.

In this example there are 3 Subjects (i.e. patients A, B and C) who have been individually sampled: Before(1), During (2) and After (3) treatment



When the design page opens use the **Add Subject** and **Add Condition** buttons to create the matrix that fits your experimental design, over typing the names as required.

Then Drag and drop the Samples on to the correct 'cell' of the matrix.



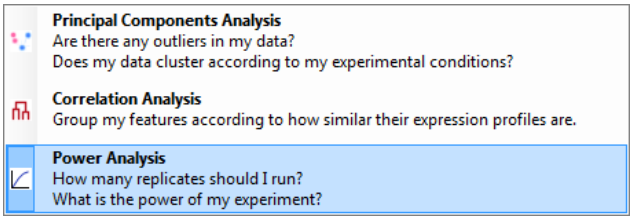
You can create additional Experimental Designs using the New tab

All of these Experimental Designs are available at all the following stages in the workflow.

Appendix 7: Power Analysis (Peptide Stats)

Power analysis is a statistical technique that is used to gauge how many replicates are needed to reliably see expression differences in your data. It is available through the Progenesis Stats section of the workflow.

To perform a power analysis of the data click on **Ask another question** at the top of the table in the Progenesis Stats screen. A selection of 3 tools will appear in the form of questions.



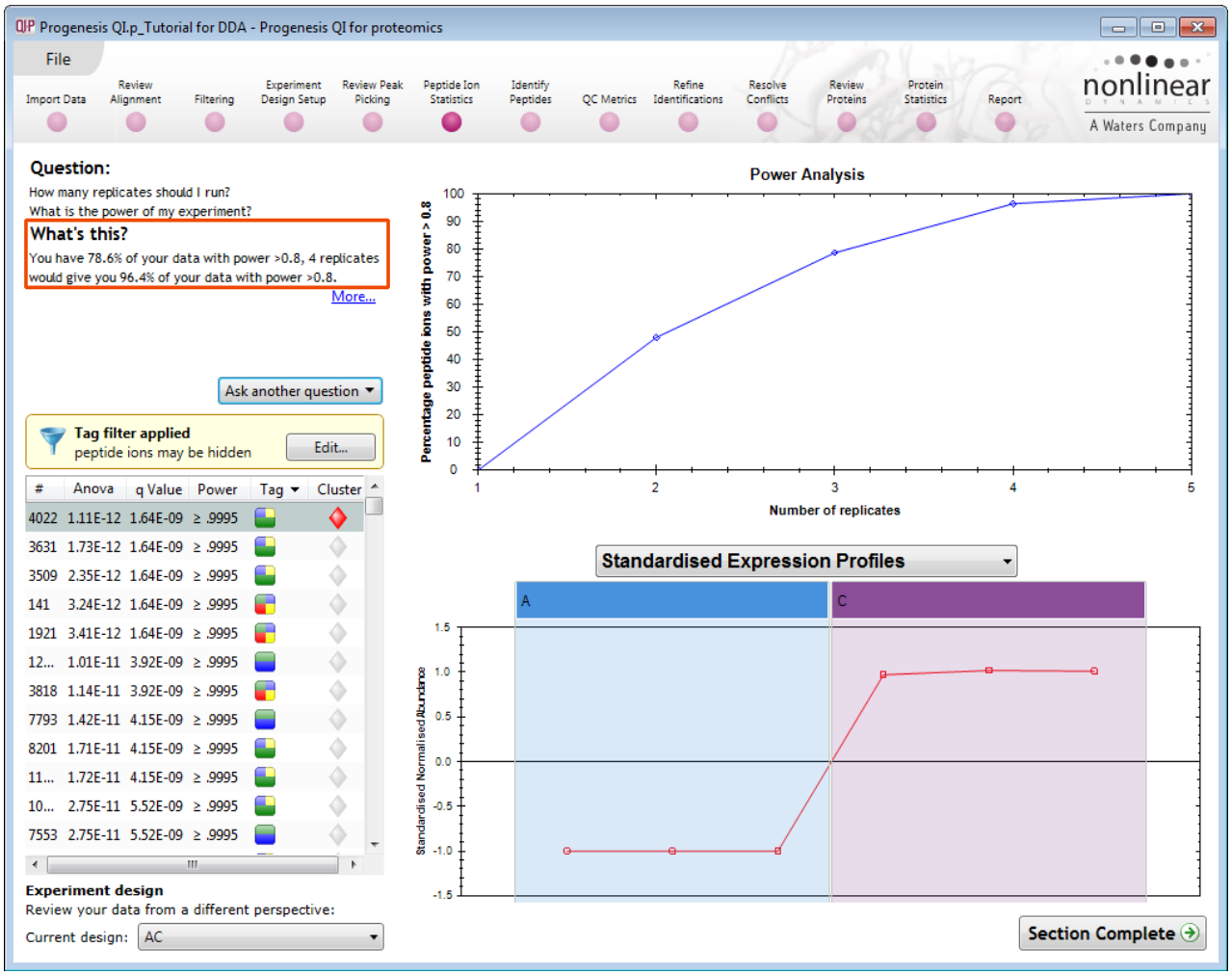
Select the option

'How many replicates should I run and what is the power of my experiment?'

It answers this question by informing you:

'How many replicates you need so that at least 80% of your peptide ions with a power >0.8'

Using the **Significant p<0.05 peptide ions (7971),** as an example, view the power analysis.



This is displayed graphically showing that 78.6% of the 7991 peptide ions have a power of 80% or that 4 replicates would give you 96.4% of your data with power > 0.8.

- **The power of a statistical test** reflects our confidence in the experimental data's ability to find the differences that do actually exist
- The power is expressed as a percentage, where 80% power is an accepted level, therefore allowing you to assess the number of sample replicates that would be required to achieve a power of 80%.

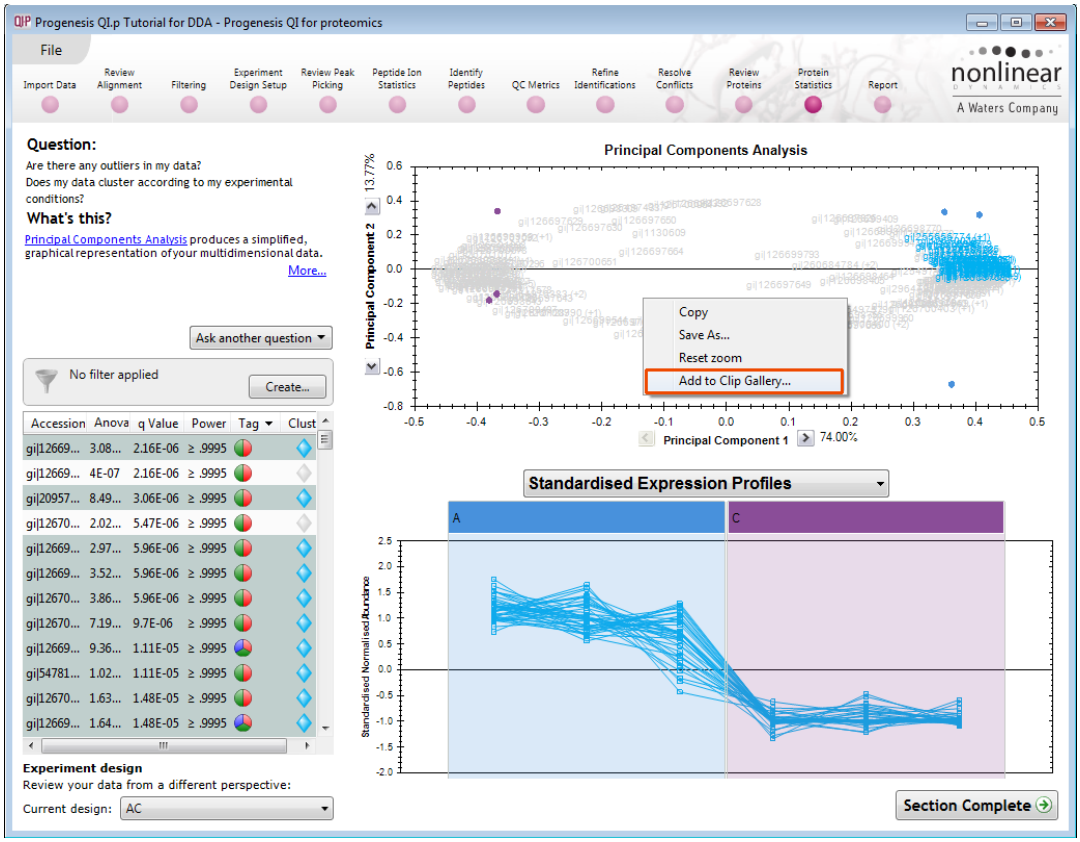
Appendix 8: Using Clip Gallery to Save and Export Pictures and Data

At every stage of the Progenesis QI for proteomics workflow the images and data tables can be added to the Clip Gallery.

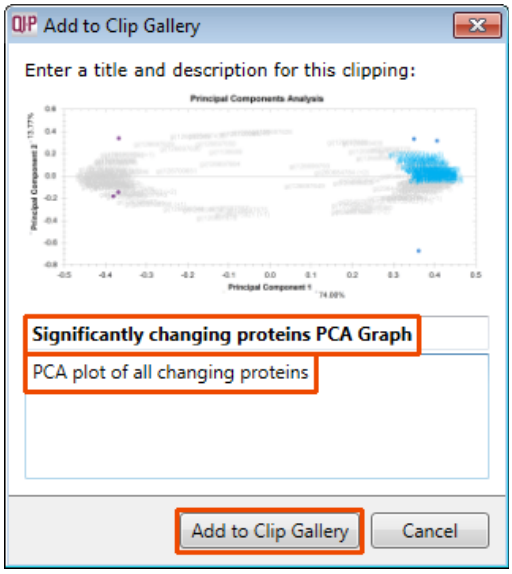
The saved images are retained as part of the experiment and are stored accordingly. This facility allows you to capture (high resolution) images that can be used in the development of specific reports and/or used as part of the process of publishing your experimental findings.

As an example of using the Clip Gallery, at the **Protein Statistics** view, displaying the PCA plot.

Right click on the **Biplot** View and select Add to clip gallery….

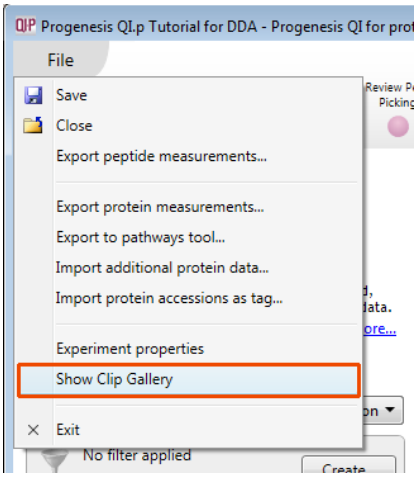


This will open a dialog displaying what is to be saved and allows you to alter the title and provides a description of the image for later reference.

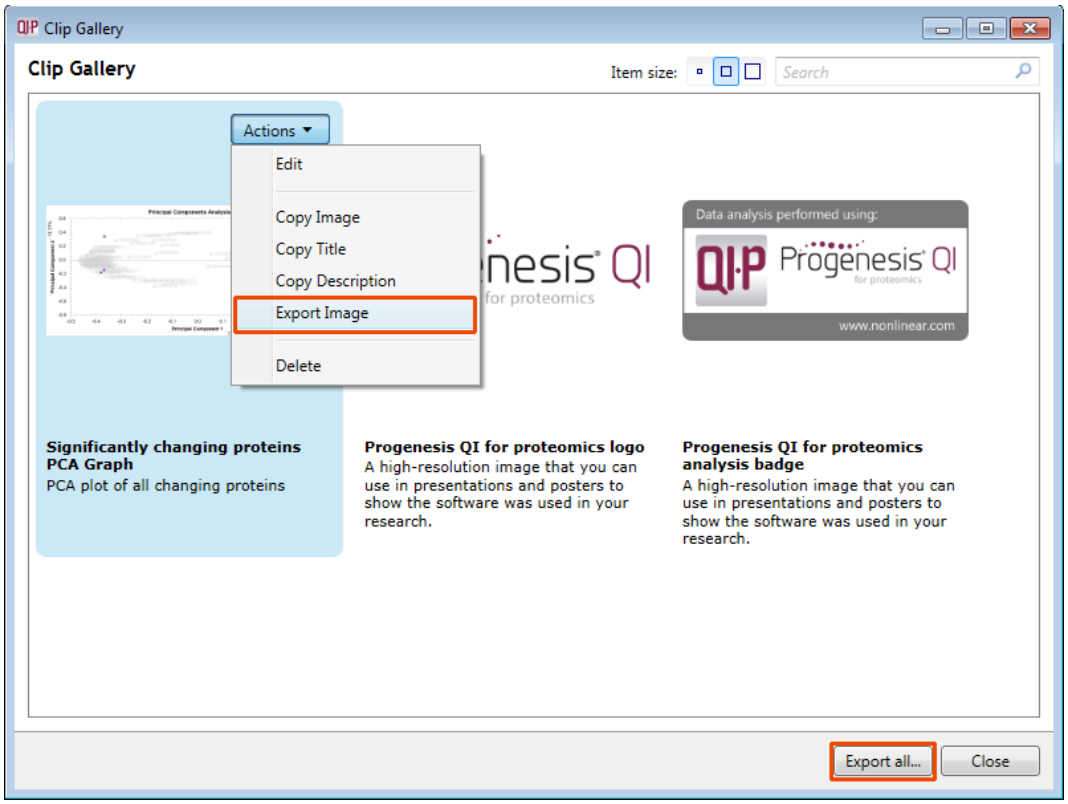


Enter details as required and click **Add to clip gallery**

To view, edit and/or export from the clip galley the gallery can be accessed from the **File** menu.



Selecting an image in the gallery makes available an **Actions** menu that allows you to manipulate the output of the image.

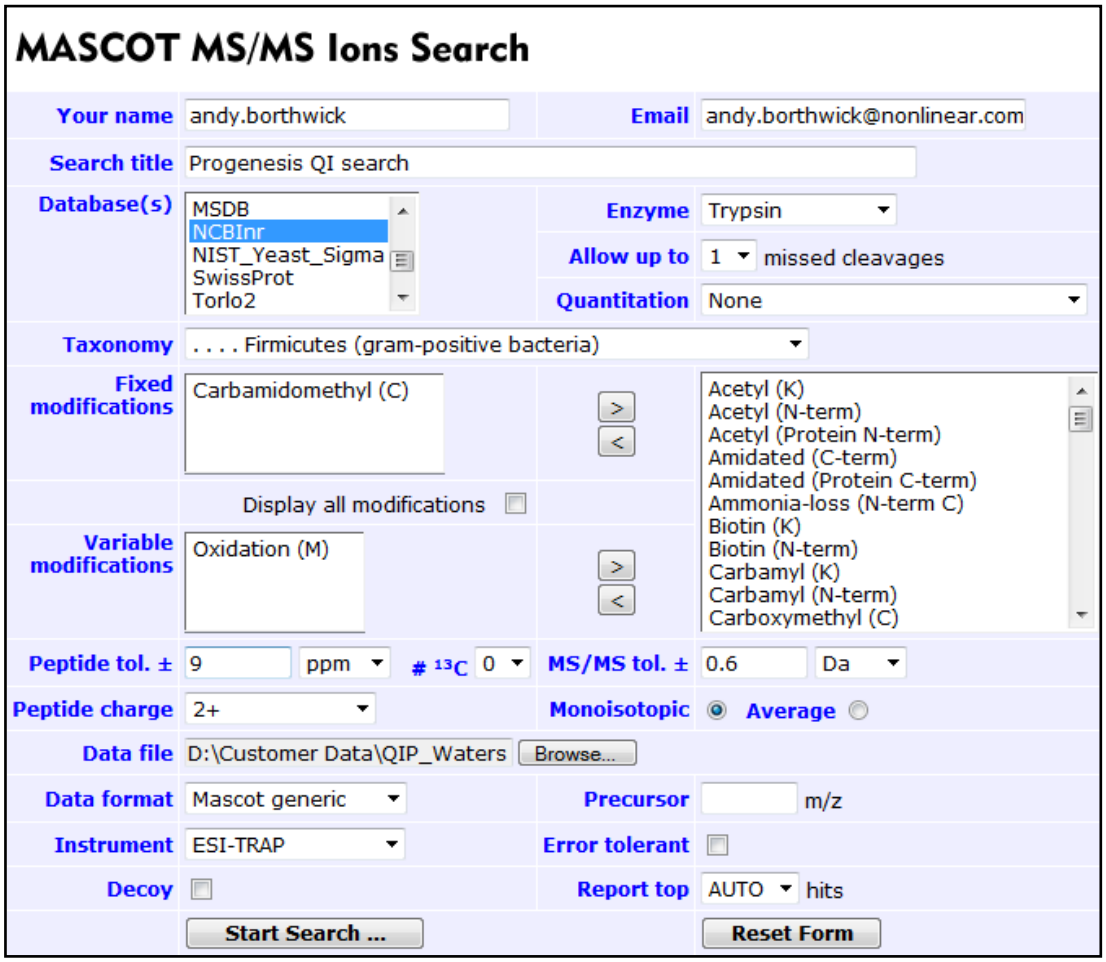


Note: there is also the capacity to **Export all…** the images in the experiments clip gallery which creates a list of files in a folder of your choice where the file name is based on the image title.

Note: right clicking on a table and adding it to the Clip Gallery allows you to export the current content to Excel.

Appendix 9 Search engine parameters (Stage 9) Mascot

The parameters applied to the Mascot search that yielded the search results used in this user guide are shown below:



Database : NCBInr (circa 01/14) was used with the Taxonomy restriction set to Fermicutes

Fixed modifications: Carbamylation(C) and variable modification Oxidation (M)

Peptide Tol: 9ppm

Instrument: ESI-Trap

References

Silva, JC, Gorenstein, MV, Li, G-Z, Vissers, JPC and Geromanos. Molecular and Cellular Proteomics (2006); 5 : 144-156 Absolute Quantification of Proteins by LCMS^{E}