

Progenesis Q1 for proteomics User Guide

**Analysis workflow guidelines for HDMse and MSe
data**

Waters

THE SCIENCE OF WHAT'S POSSIBLE.™

Contents

Introduction	3
How to use this document	3
How can I analyse my own runs using Progenesis QI for proteomics?	3
LC-MS Data used in this user guide.....	3
Workflow approach to LC-MS run analysis	4
Restoring the Tutorial	5
Stage 1: Import Data and QC review of LC-MS data set	6
Stage 2A: Automatic Processing of your data.....	7
Stage 2B: After Automatic Processing	14
Stage 3: Licensing	15
Stage 4: Review Alignment	16
Reviewing Quality of Alignment.....	20
Stage 5A: Filtering	23
Stage 5B: Reviewing Normalisation	26
Stage 6: Experiment Design Setup for Analysed Runs.....	29
Stage 7: Review Peak Picking and editing of results	32
Stage 8: Peptide Ion Statistics on Selected Peptide ions	45
Stage 9: Identify peptides	49
Stage 10: QC Metrics	52
Stage 11: Refine Identifications.....	54
Resolving Conflicts	56
Protein Grouping	56
Protein Quantitation options	58
Stage 12: Review Proteins	61
Stage 13: Exporting Protein Data	67
Exporting Protein Data to Pathways Tool(s)	69
Stage 13: Exporting identifications for submission to PRIDE.....	71
Stage 14: Protein Statistics	73
Stage 15: Reporting.....	74
Appendix 1: Stage 1 Data Import and QC review of LC-MS data set	76
Appendix 2: Stage 1 Processing failures.....	81
Appendix 3: Licensing runs (Stage 3)	83
Appendix 4: Manual assistance of Alignment.....	84
Appendix 5: Within-subject Design.....	89
Appendix 6: Power Analysis (Peptide Ion Stats)	91
Appendix 7: Resolve Conflicts.....	92
Appendix 8: Using Clip Gallery to Save and Export Pictures and Data	98
Appendix 9: Exporting protein identities to IPA	100
Appendix 10: Exporting protein identities to Metacore	103
Appendix 11: Waters Machine Specification	105
References	105

Introduction

This user guide takes you through a complete analysis of 9 LC-MS runs with 3 groups (3 replicate runs per group) using the unique Progenesis QI for Proteomics workflow. It starts with LC-MS data loading then Alignment, followed by Peak Detection 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 Protein Stats.

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 76) then start at page 6.

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. 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?

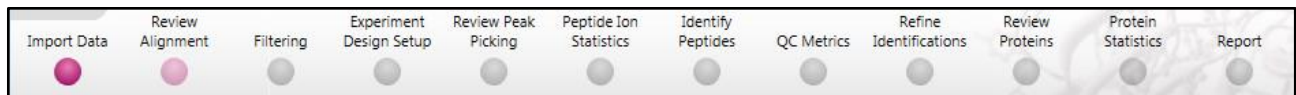
You can freely explore the quality of your LC-MS data using Data Import and then licence your own LC-MS runs using this evaluation copy of Progenesis QI for proteomics. Instructions on how to do this are included in a section at the end of the user guide document.

LC-MS Data used in this user guide

For the purposes of this data set the MS^F parameters were set to 250:125 instead of the default settings as defined in Appendix 1 (page 76). This was done to reduce the time taken to demo the data analysis.

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.

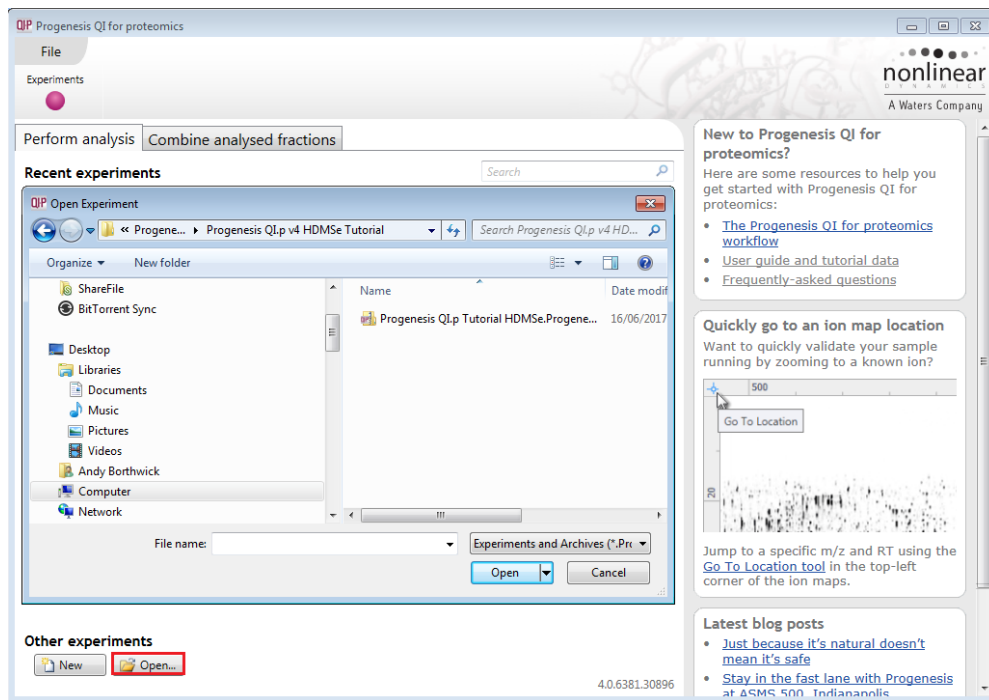


Stage	Description	Page
	LC-MS Import Data: Selection and review of data files for analysis	6
	Automatic Processing: setting up steps for automatic processing	7
	After Automatic Processing: how to work with auto analysed data	14
	Licensing: allows licensing of individual data files when there is no dongle attached (Appendix 3)	15
	Review Alignment: review of automatic and manual LC-MS run alignment	16
	Filtering: defining filters for peaks based on Retention Time, m/z, Charge State and Number of Isotopes.	23
	Review Normalisation: exploring LC-MS normalisation	26
	Experiment Design Setup: defining one or more group set ups for the analysed aligned runs	29
	Review Peak Picking: review and validate results, edit peak detection, tag groups of peaks and select peaks for further analysis	32
	Peptide Ion Statistics: performing multivariate statistical analysis on tagged and selected groups of peptide ions	45
	Identify Peptides: managing export of MS/MS spectra to, and import of peptide ids from Peptide Search engines	49
	QC metrics: quality control charts for experimental/analysed data	52
	Refine Identifications: manage filtering of peptide ids	54
	Resolve Conflicts: validation and resolution of peptide id conflicts for data entered from Database Search engines	56
	Review proteins: review protein and peptide identity and data export	61
	Protein Statistics: multivariate statistical analysis on proteins	73
	Reporting: generate a report for proteins and/or peptides	74

Restoring the Tutorial

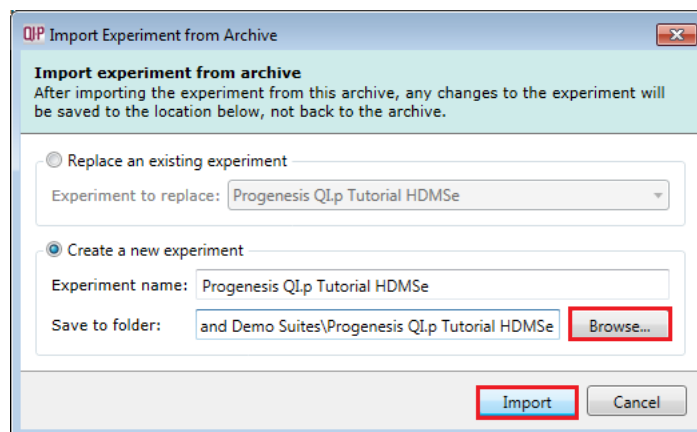
Open Progenesis QI for proteomics and download the Compressed (.zip) Tutorial Archive file from the 'User guide and tutorial' link shown below, placing it in a **new folder** on your desktop. Before restoring the tutorial in the software **you must** first right click on the (.zip) file and extract it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

Now restore the uncompressed tutorial archive file. To do this, first locate the **Progenesis QI.p Tutorial HDMSe.Progenesis QIP Archive** file using the **Open** button and press Open.

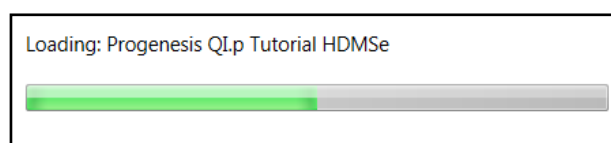


This opens the 'Import from archive' dialog.

Select the **Create a new experiment** option and select the folder in which you placed the archive, using Browse.



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).

Import Data

- Select your run data**
Select one of the available data formats then click the Import button:
Format: Waters (.raw)
[About this data format](#) | [Download others](#)
- Perform automatic processing**
Click the button below to start automatic processing. This will analyse your data as far as possible, before suggesting the next step.

Automatic processing can be started while runs are importing.
[Learn more about automatic processing](#)

Imported runs:

A_01 (low energy)

A_01 (high energy)

About this run

- Low energy peak count: 4,413,012
- High energy peak count: 10,307,607
- Total ion intensity: 6.087e+008
- Masked areas : none
- Filter strength: x2

Data import details

- Lock mass calibrated
- Lock mass m/z: 785.8426

Section Complete

Each data file appears as a 2D representation of the run.

Note: the **Experiment Properties** are available from the File menu. These were selected when the experiment was created (see Appendix 1, page 76). The tutorial data is profile data.

QIP Experiment Properties

Runs in this experiment: 9
Machine resolution: High resolution
Peak processing: Profile data

Tip: the **'Mask areas for peak picking'** facility, accessed by right clicking on the run thumbnail) 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. This is not required for this data set.

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

Now start the automatic processing.

Mask areas for peak picking...

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.

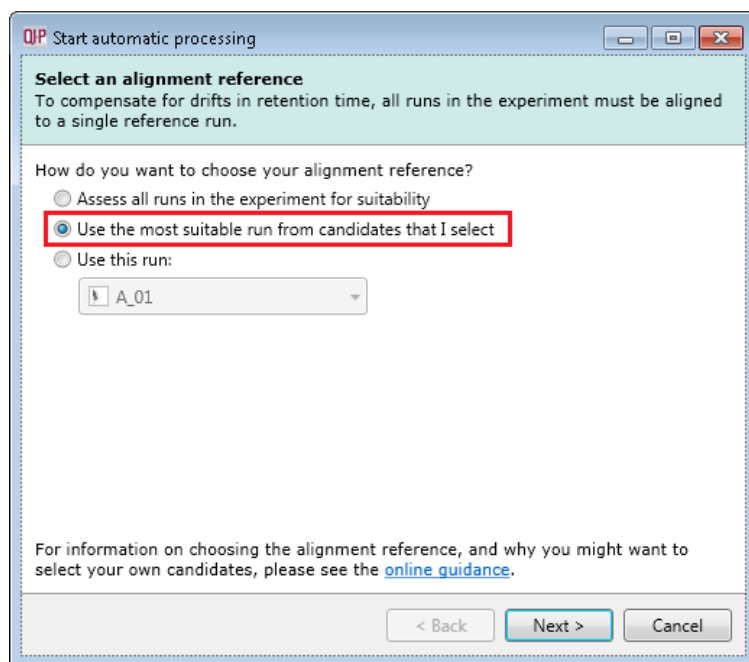


Setup of processing steps in the Analysis Workflow, up to and including Identify Peptides, can be performed in the Automatic Processing wizard, these include:

- Select an alignment reference
- Automatic alignment of all runs to a reference run
- Automatic peak picking for peptide ion detection
- Define an Experiment design
- Perform peptide and protein identification (for MS^E data only, as DDA data will require external database searching)
- Choosing a Quantitation method to be applied for abundance analysis

In this tutorial example you have 9 HDMS^e LC-MS runs, so the automatic identification of peptides is available as part of 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 selects 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 when all the candidates are from a condition that displays the largest set of common peptide ions.

3. Use this run

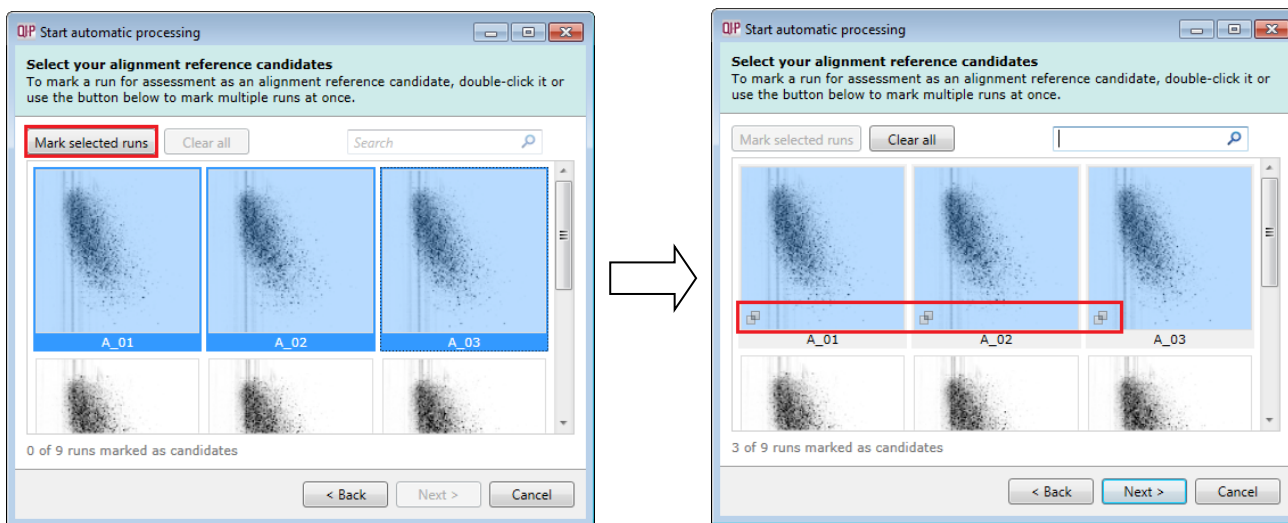
This method allows you to manually choose the reference run.

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).

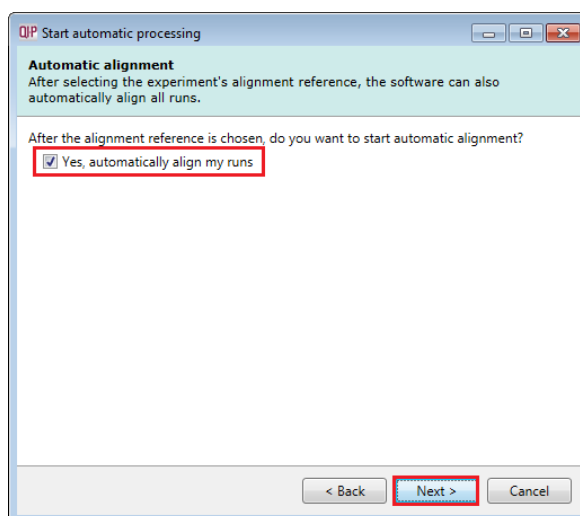
For this tutorial we will select the second option.

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.

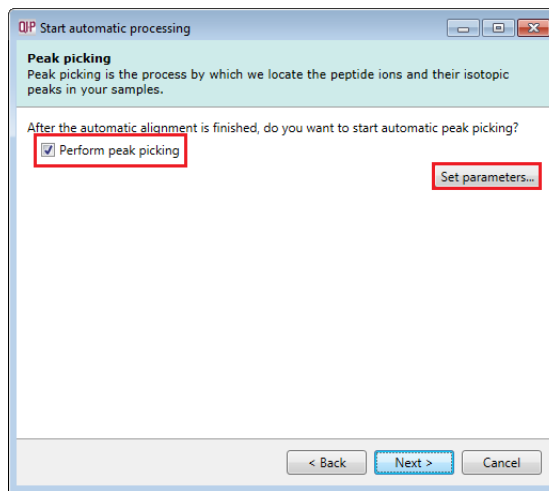


On clicking **Next** you will be asked if you want to align your runs automatically.

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



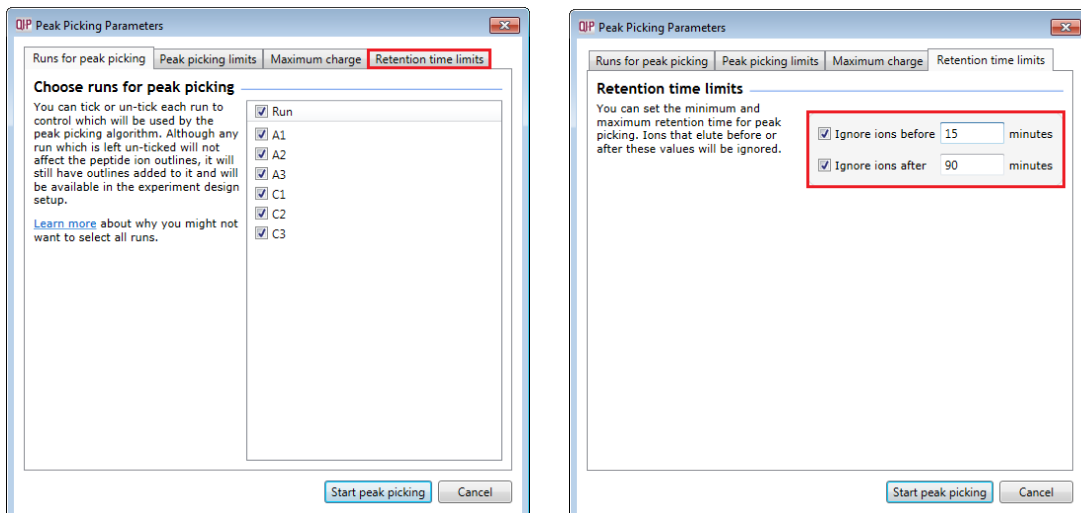
The next page of the processing wizard will ask you if you want to **Perform peak picking** and allow you to set appropriate parameters.



For the purposes of this User guide we will use the default settings for peak picking except in the case of defining Retention Time limits where we will limit the peak picking to between 15 and 90 min.

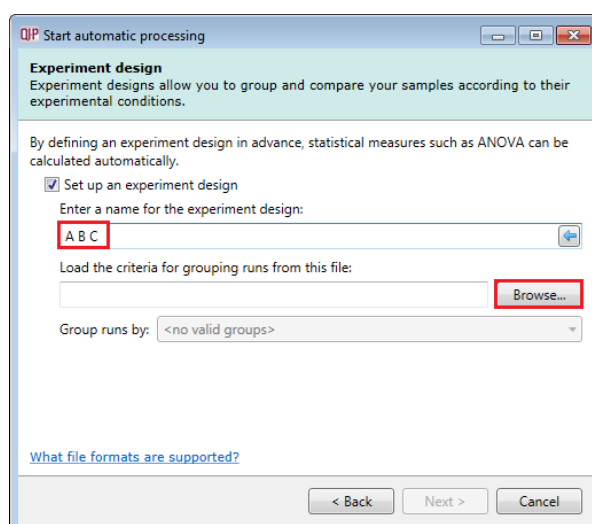
Note: for more details on setting Peak Picking parameters refer to the section on Filtering (page 23)

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 15 and 90 min and tick the boxes as shown above.

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

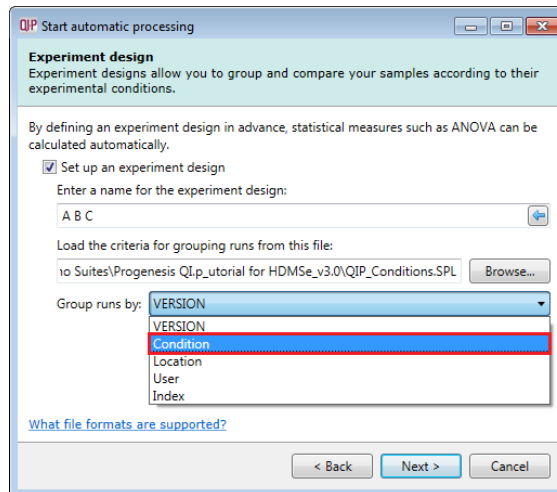


To handle the grouping of your run data you can make use of **sample tracking information** that has been stored in a spreadsheet at the time of sample collection and/or preparation.

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 **QIP_Conditions.spl** file available with the Experiment Archive you restored at the beginning of this tutorial exercise.

Give the experiment design a name (i.e A B C) and then use the **Browse** function to locate the **QIP_Conditions.spl** file.



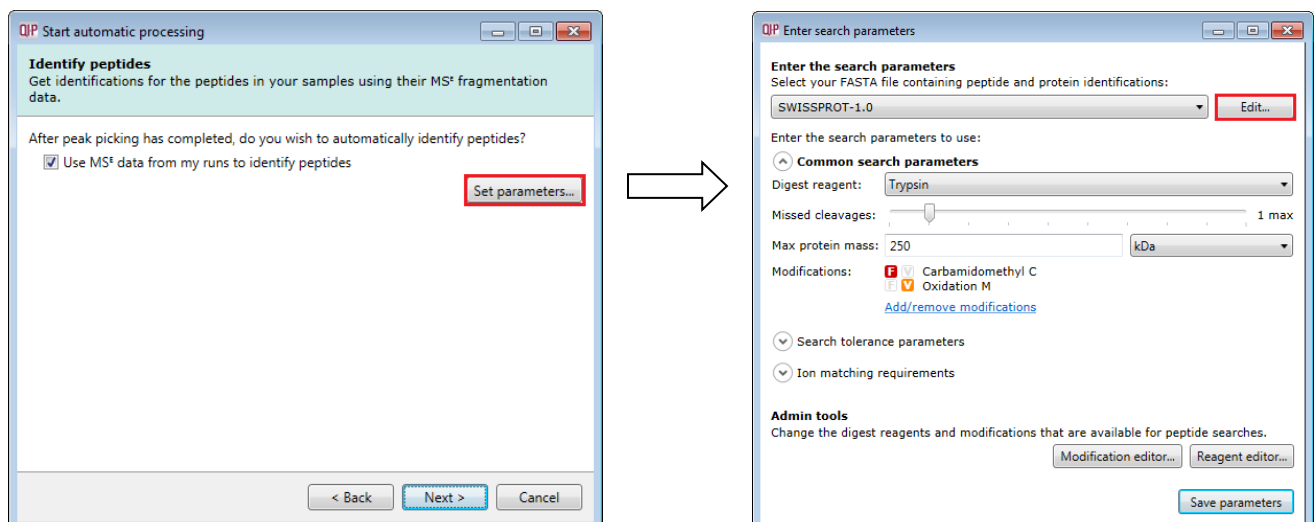
To use this approach select the **Import design from file** option from the New Experiment Design dialog. Then locate the QIP_Conditions file and select what to **Group runs by**, for example: **Condition**.

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

Select Conditions and then click **Next**.

For MS^E fragmentation data you can set up the peptide identification to be performed automatically.

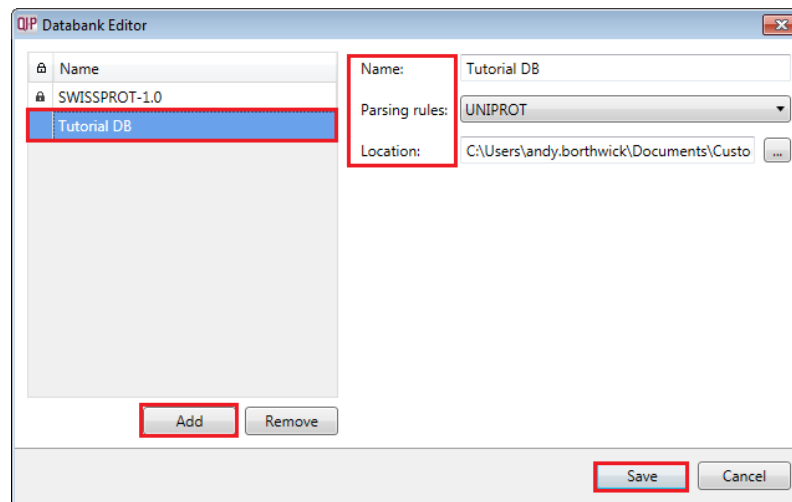
Note: if the software has detected MS^E data then this option will be ticked by default.



Click on **Set Parameters**. The default Databank is for Swissprot-1.0 (which is a locked example).

To create a new Databank from a Fasta file click on **Edit** and then create a new one using the example FASTA file that is available with the Experiment Archive you restored at the beginning of this tutorial exercise (**Tutorial_nd_DB.fasta**).

For a new Databank click **Add**. Then give it name (i.e. Tutorial DB), select the parsing rules (UNIPROT) and specify the location of the FASTA file, see the example below.



The new Data bank will appear in the left panel now click **Save** to return to the Search parameters.

If your databank is not already displayed then select it from the drop down list.

Check the **Common search parameters**

The default settings are displayed:

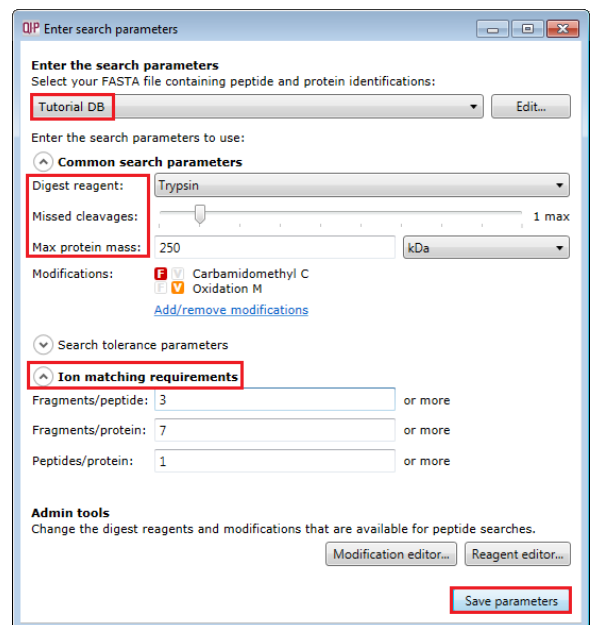
Digest reagent: is set as Trypsin. Alternative Digest reagents are available from the list and additional ones can be added to the list using the **Reagent editor...**

Missed cleavages: is set as 1.

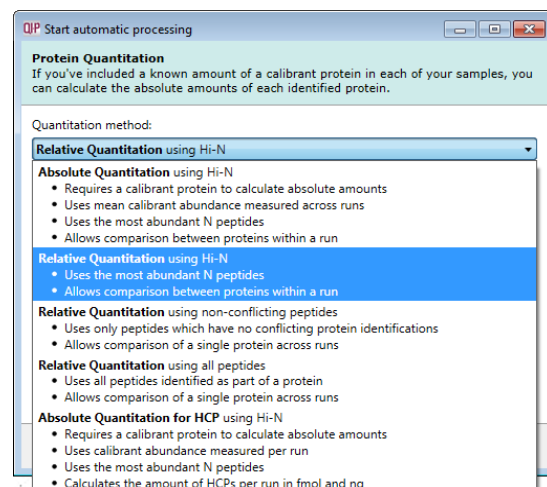
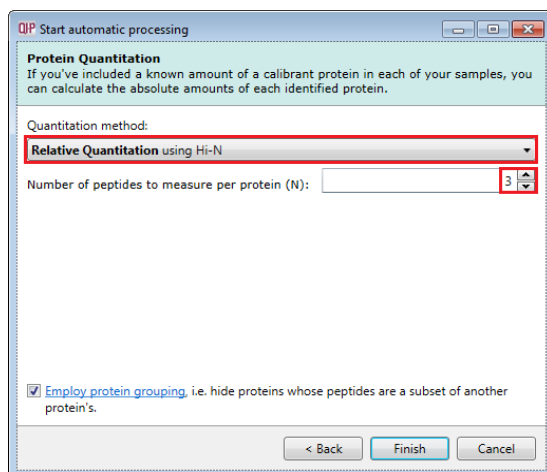
Maximum protein mass: is set at 250kDa

Modifications: are set Carbamidomethyl C (Fixed) and Oxidation M (Variable). More modifications are available from the list and additional ones can be added to the list using the **Modification editor...**

Ion matching requirements: are set at Fragments/peptide: 3, Fragments/protein: 7 and Peptides/protein: 1



Click **Save parameters** and then **Next**.

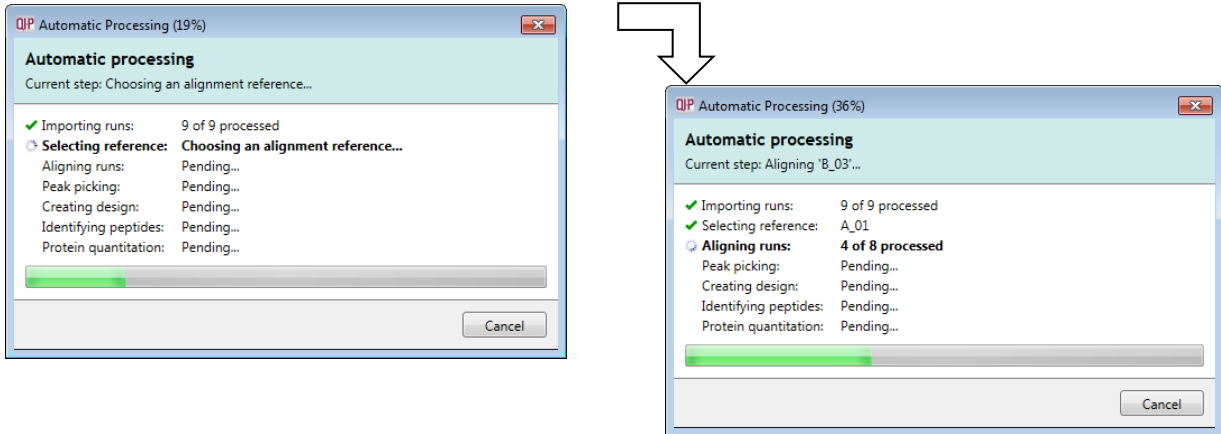


The **Protein Quantitation** dialog opens displaying the default method, **Relative Quantitation using Hi-N** which uses up to 3 peptides per protein to compute the relative amount of each 'identified' protein.

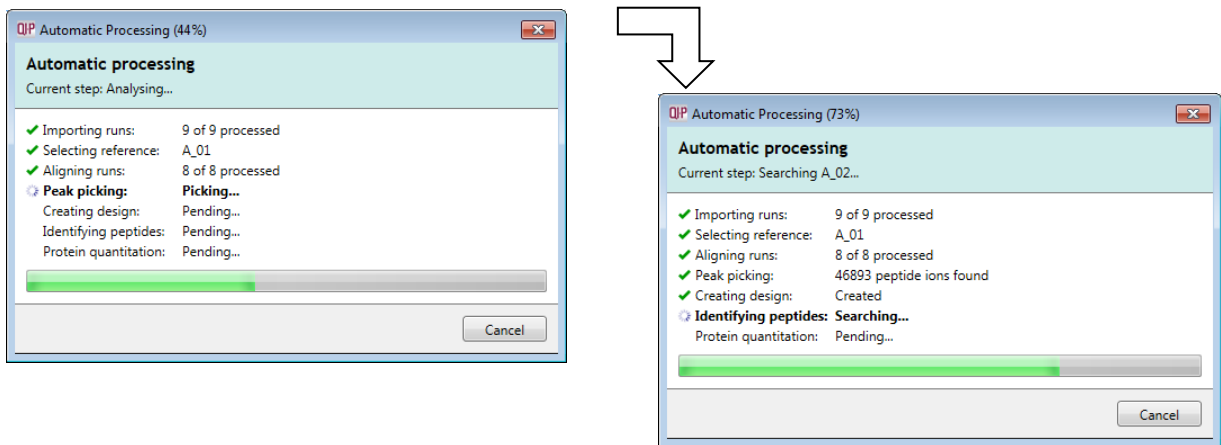
Use the drop down to reveal the alternative methods for protein quantitation

Select the Default option and click **Finish**.

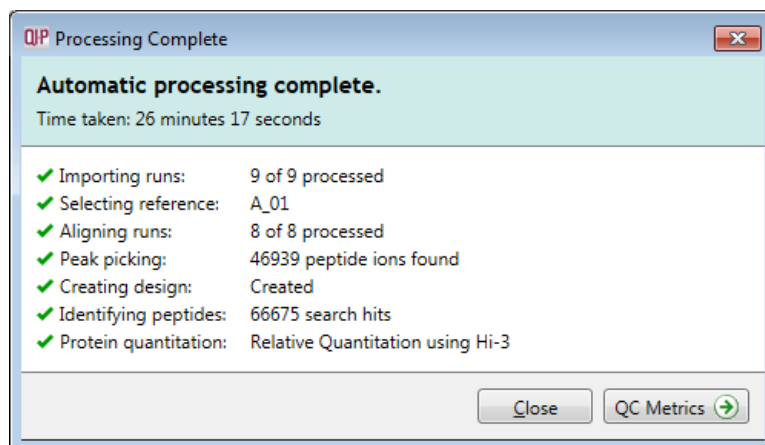
The Alignment process starts with the automatic selection of A_01 as the alignment reference



Once Alignment completes Peak Picking commences followed by Design and Identification.



Finally the Automatic Processing completes with the Quantitation being performed

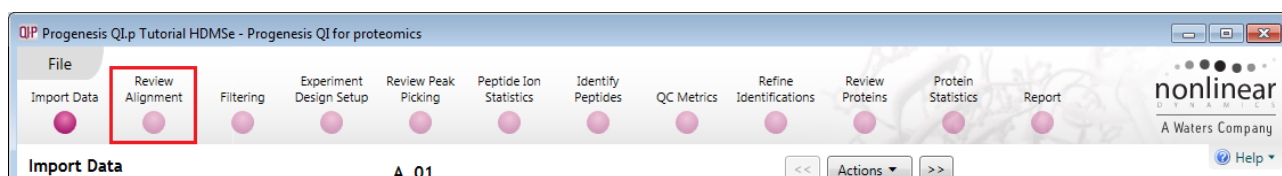


As the whole process proceeds you get information on what stage has been performed and also the % of the process that has been completed.

When Processing completes, depending on what stages you selected to perform, the Wizard displays what stage the workflow will open at. In this example it will open at QC Metrics.

You can either:

- Continue with the analysis, as the Processing dialog is not displaying any warnings, and review QC Metrics. In which case you can go to page 52
- Open the analysis at QC Metrics and immediately return to the Review Alignment stage by clicking on it in the Workflow to review alignment quality (page 16)
- Alternatively you can **Close** the dialog. This will not move you to a later stage in the workflow but instead allow you to navigate 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 81)

Please Note: the time taken to automatically process this data through all the steps including Peptide Identification takes approximately 45 minutes (as shown on the previous page). This is dependent on the specification of your PC.

Note: this does not include the time to load the data

You can explore and re-perform the steps, sequentially and/or as part of the automatic processing as described in this guide.

For this dataset the alignment and detection from the automatic processing was of good enough quality to not require any further amendments. However, in the course of exploring the data you may choose to edit and re-perform the alignment and or the peak detection as described in this document.

You will be warned that you are about to lose the analysis as you 'unlock' a step and reanalyse.

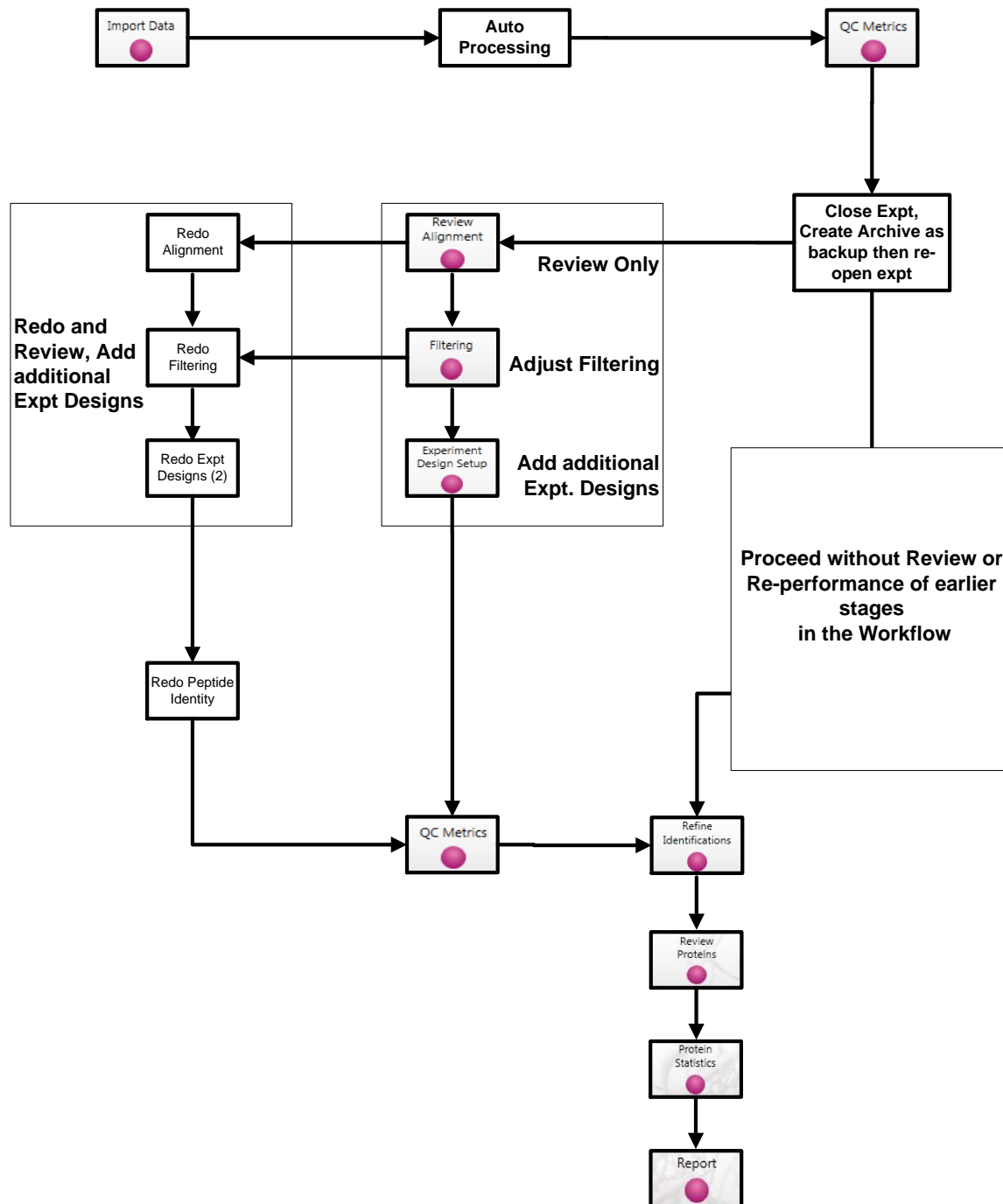
An example of the time it takes to perform each step is outlined in Appendix 10 (page 103).

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 (on the right) 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 HDMS_e (Waters) it is going to open at QC Metrics if you chose to perform all the steps in the Automatic Processing Wizard so you can either:

- Continue with the analysis, as the Processing dialog is not displaying any warnings, review the analysis at QC Metrics before proceeding with Refine Identifications. In which case you can go to page 52.
- If warnings are displayed or on Review of the Quality Metrics you require to review/redo earlier stages in the analysis then proceed to pages 16 and Appendices 1 (page 76) and 2 (page 81).

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, if you close Progenesis QI for proteomics you will be warned that the analysis will be lost.

The screenshot shows the Progenesis QI software interface. The 'License Runs' dialog box is open, displaying a table of runs and their license status. An error message dialog box is also visible, stating 'Unable to save experiment' and 'You cannot save the analysis without a valid license. If you close now your analysis will be lost.'

Run name	License state	License this run
C:\Users\andy.borthwick\Documents\Customer Data\Progenesis QI.p v3.0 Tutorials and Demo Suites\Pr...	Unlicensed	<input checked="" type="checkbox"/>
C:\Users\andy.borthwick\Documents\Customer Data\Progenesis QI.p v3.0 Tutorials and Demo Suites\Pr...	Unlicensed	<input checked="" type="checkbox"/>
C:\Users\andy.borthwick\Documents\Customer Data\Progenesis QI.p v3.0 Tutorials and Demo Suites\Pr...	Unlicensed	<input checked="" type="checkbox"/>
C:\Users\andy.borthwick\Documents\Customer Data\Progenesis QI.p v3.0 Tutorials and Demo Suites\Pr...	Unlicensed	<input checked="" type="checkbox"/>
C:\Users\andy.borthwick\Documents\Customer Data\Progenesis QI.p v3.0 Tutorials and Demo Suites\Pr...	Unlicensed	<input checked="" type="checkbox"/>
C:\Users\andy.borthwick\Documents\Customer Data\Progenesis QI.p v3.0 Tutorials and Demo Suites\Pr...	Unlicensed	<input checked="" type="checkbox"/>
C:\Users\andy.borthwick\Documents\Customer Data\Progenesis QI.p v3.0 Tutorials and Demo Suites\Pr...	Unlicensed	<input checked="" type="checkbox"/>
C:\Users\andy.borthwick\Documents\Customer Data\Progenesis QI.p v3.0 Tutorials and Demo Suites\Pr...	Unlicensed	<input checked="" type="checkbox"/>
C:\Users\andy.borthwick\Documents\Customer Data\Progenesis QI.p v3.0 Tutorials and Demo Suites\Pr...	Unlicensed	<input checked="" type="checkbox"/>
C:\Users\andy.borthwick\Documents\Customer Data\Progenesis QI.p v3.0 Tutorials and Demo Suites\Pr...	Unlicensed	<input checked="" type="checkbox"/>

Run licence code: _____

Section Complete

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 3 (page 83)

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 **Review Alignment** opens displaying the alignment of the runs to the Reference run (A_01).

Having performed the analysis automatically there will be icons next to each run in the vectors table indicating that the run is protected from editing.

Run	Includes?	Vectors	Score
A_01	✓	Ref	
A_02	✓	1963	97.6%
A_03	✓	2074	97.4%
B_01	✓	1982	96.7%
B_02	✓	1901	96.1%
B_03	✓	1937	96.4%
C_01	✓	1980	95.5%
C_02	✓	2031	96.3%
C_03	✓	1950	94.4%

In the course of reviewing the quality of alignment, you may decide that the alignment requires editing.

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 4 (page 84)

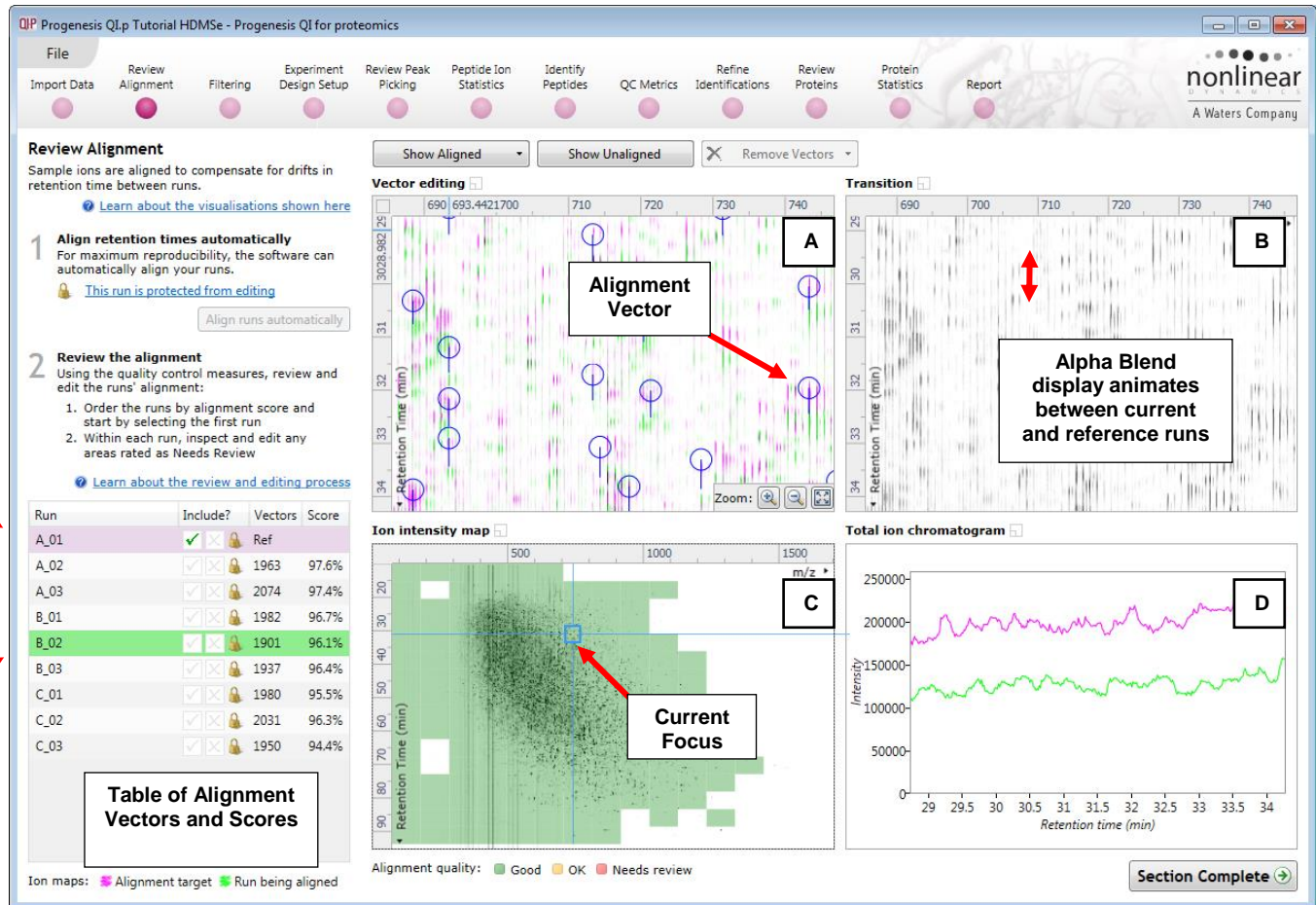
Layout of Alignment

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

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

- In the Run table click on Run B_02 to make it current. You will now be looking at the alignment of B_02 to A_01 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.

Run	Include?	Vectors	Score
A_01	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Ref	
A_02	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 1963	97.6%
A_03	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 2074	97.4%
B_01	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 1982	96.7%
B_02	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 1901	96.1%
B_03	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 1937	96.4%
C_01	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 1980	95.5%
C_02	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 2031	96.3%
C_03	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 1950	94.4%



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.

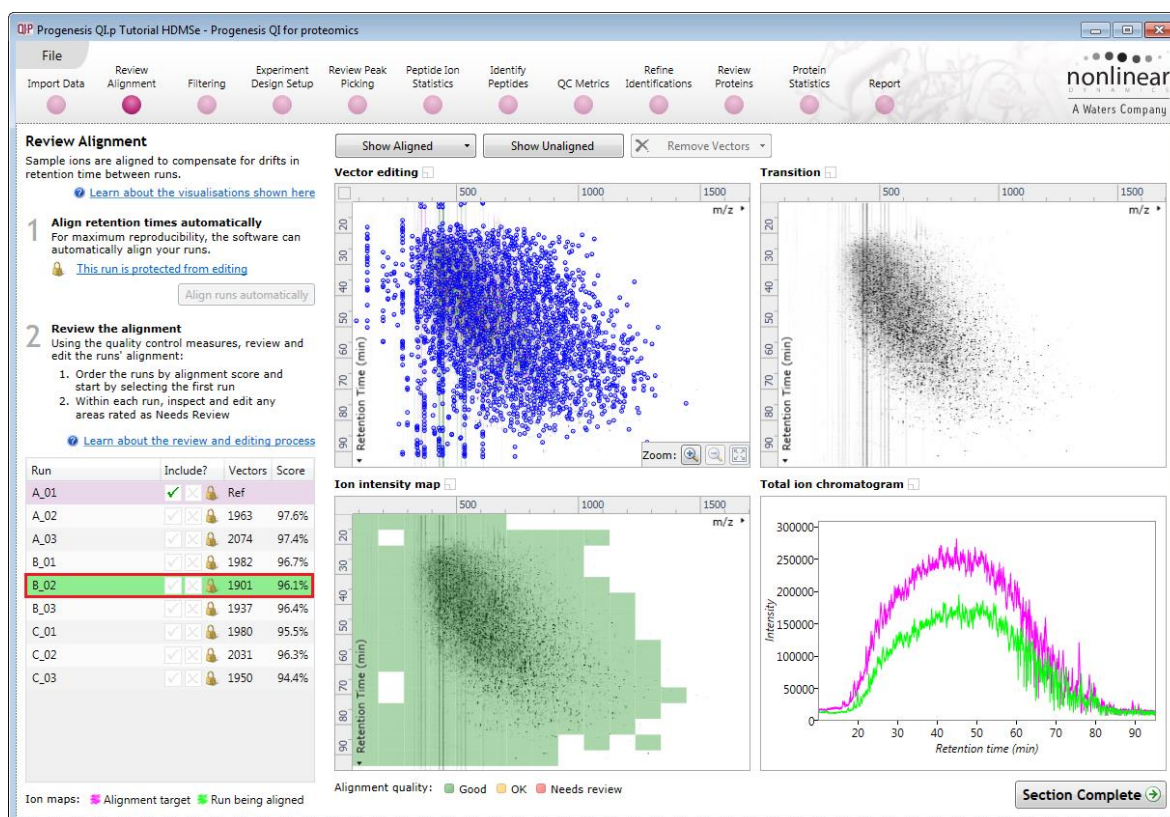
Ion Intensity Map (Window C): shows the **focus** for the other windows. When you click on the view the blue rectangle will move to the selected area. The focus can be moved systematically across the view using the cursor keys. The focus area size can be altered by clicking and dragging out a new area 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 view.

Where the alignment has worked well, the alignment views will look as below with the Ion Intensity Map showing green indicating good quality alignment and the Transition view showing peptide ions pulsing slightly but not moving up and down.



To simulate poor alignment following automatic alignment, place a single manual vector on the Vector editing view (Window A).

Placing an 'incorrect vector' to simulate miss alignment in Retention Time

This section of the user guide is provided as a **demonstration of the alignment process**, it is **not** required for the normal analysis workflow of this data.

Note: if you do explore the alignment process the final steps in this section show you how to return the alignment to the state following automatic alignment

To edit the alignment of your data you must first unlock the analysis. To do this click on the link **This run is protected from editing** and then click **Delete analysis and allow editing**

Review Alignment

Sample ions are aligned to compensate for drifts in retention time between runs.

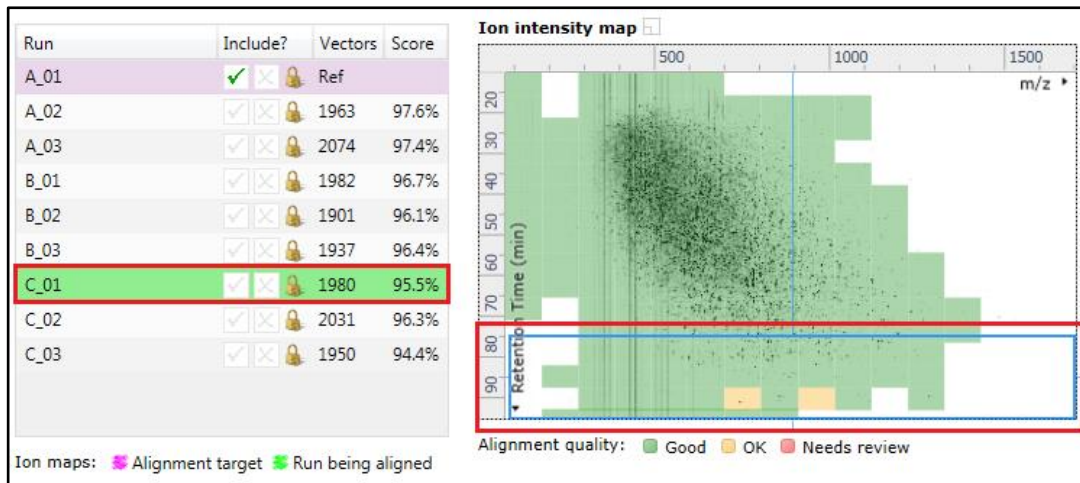
[Learn about the visualisations shown here](#)

1 **Align retention times automatically**
For maximum reproducibility, the software can automatically align your runs.

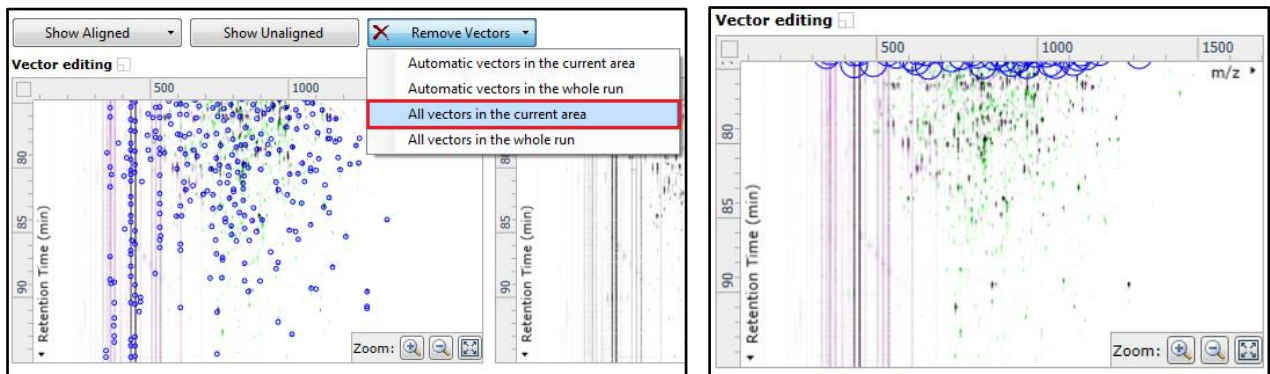
🔒 This run is protected from editing

[Align runs automatically](#)

First click on run **C_01** in the Runs table to make it current. Then remove the automatic vectors from 75 min onwards. To do this you must click on the Ion intensity map and drag out an area as shown.



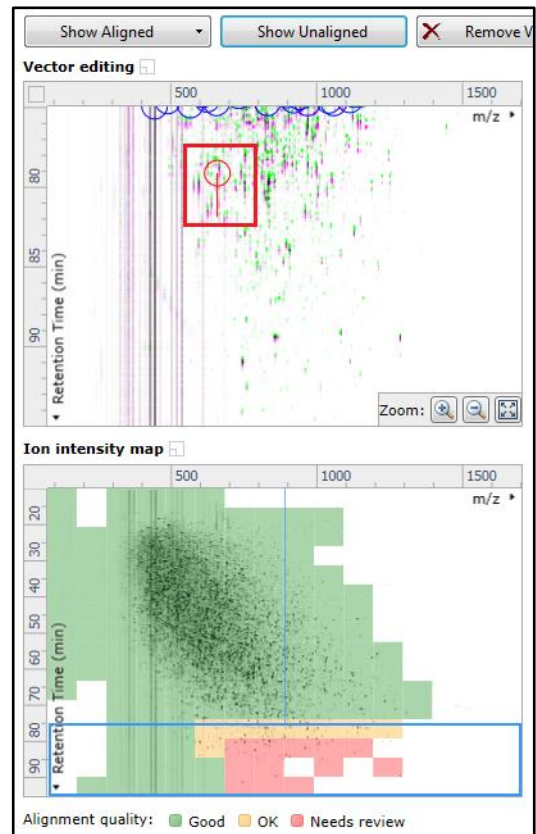
This resets all the other views. Now click **Remove Vectors** and select **All vectors in the current area** all the vectors in the Vector editing view will be removed.



Now click on the Vector editing view and drag the view slightly upwards causing the two runs to go out of synch. Then release the mouse button.

By doing this a single manual vector will appear with a length corresponding to the 'drag'.

Note: the manual vector is **red**, to distinguish it from the automatic vectors (blue)



The effect of adding this incorrect manual vector is to reduce the Alignment score and also cause a significant proportion of the Alignment quality squares to turn red on the Ion Intensity Map.

Finally click on **Show Aligned**.

Using the **Simulated** miss-aligned example, you can explore the review process for alignment. The alignment looks as below with a region of poor alignment (highlighted in red).

Review Alignment

Sample ions are aligned to compensate for drifts in retention time between runs.

1 **Align retention times automatically**
For maximum reproducibility, the software can automatically align your runs.

2 **Review the alignment**
Using the quality control measures, review and edit the runs' alignment:
1. Order the runs by alignment score and start by selecting the first run
2. Within each run, inspect and edit any areas rated as Needs Review

Run	Include?	Vectors	Score
A_01	<input checked="" type="checkbox"/>	Ref	
A_02	<input checked="" type="checkbox"/>	1963	97.6%
A_03	<input checked="" type="checkbox"/>	2074	97.4%
B_01	<input checked="" type="checkbox"/>	1982	96.7%
B_02	<input checked="" type="checkbox"/>	1901	96.1%
B_03	<input checked="" type="checkbox"/>	1937	96.4%
C_01	<input checked="" type="checkbox"/>	1716	91.6%
C_02	<input checked="" type="checkbox"/>	2031	96.3%
C_03	<input checked="" type="checkbox"/>	1951	94.4%

Ion maps: ■ Alignment target ■ Run being aligned

Alignment quality: ■ Good ■ OK ■ Needs review

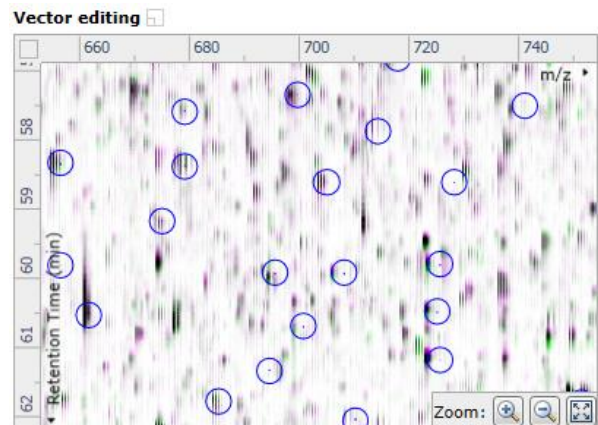
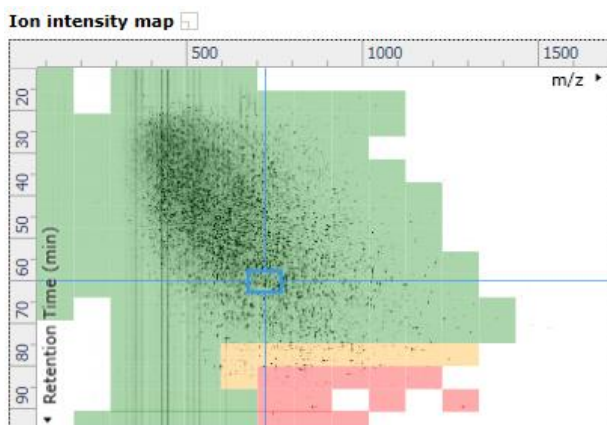
Section Complete

Note: now a smaller area in the 'green' well aligned region

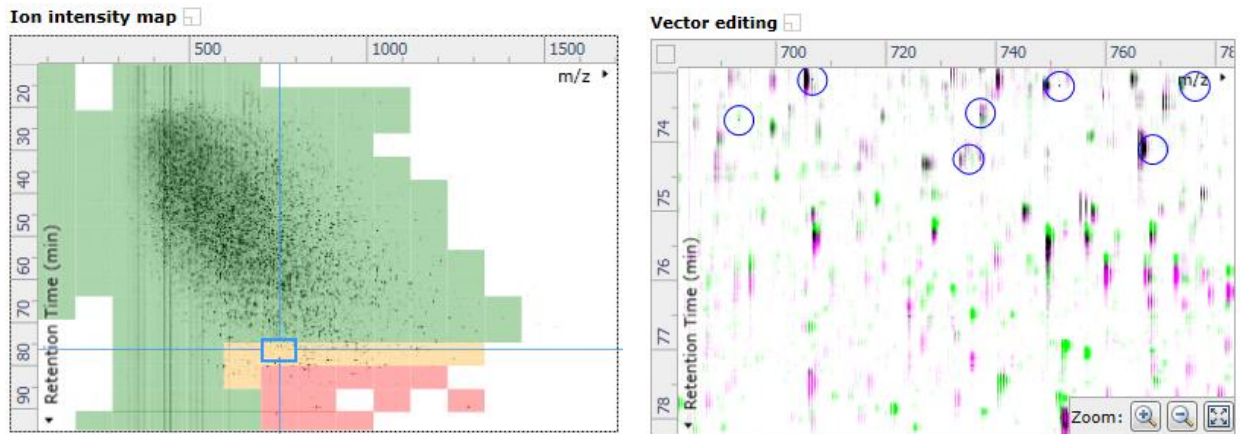
Reviewing Quality of Alignment

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). Drag out a 'Focus' area that corresponds to one of the coloured squares. Three example squares are examined here.

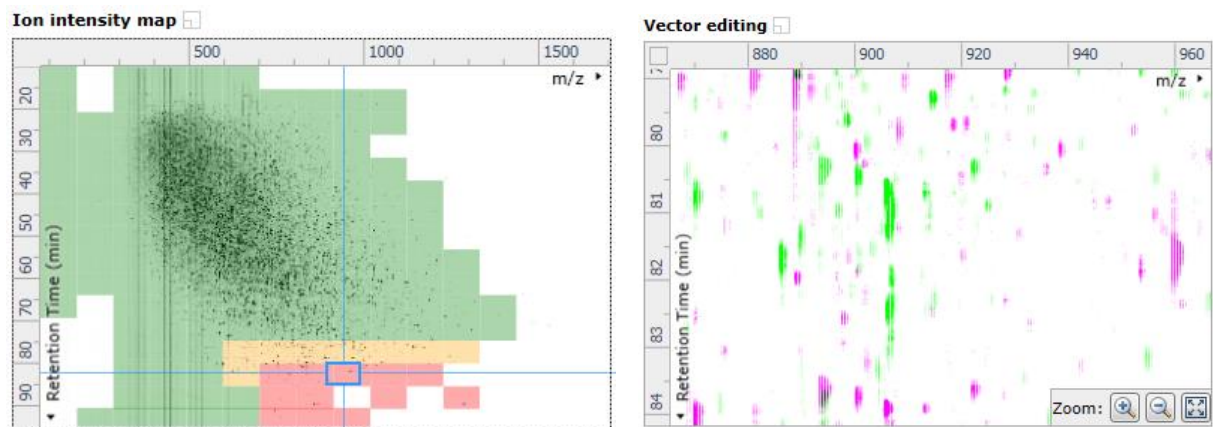
For a 'green' square, the majority of the data appears overlapped (black) indicating good alignment. When viewed in the Transition view the data appears to pulse.



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 if any 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 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's alignment.

Run	Include?	Vectors	Score
A_01	<input checked="" type="checkbox"/> <input type="checkbox"/>	Ref	
A_02	<input type="checkbox"/> <input type="checkbox"/>	1963	97.6%
A_03	<input type="checkbox"/> <input type="checkbox"/>	2074	97.4%
B_01	<input type="checkbox"/> <input type="checkbox"/>	1982	96.7%
B_02	<input type="checkbox"/> <input type="checkbox"/>	1901	96.1%
B_03	<input type="checkbox"/> <input type="checkbox"/>	1937	96.4%
C_01	<input checked="" type="checkbox"/> <input type="checkbox"/>	1047	80.4%
C_02	<input type="checkbox"/> <input type="checkbox"/>	2031	96.3%
C_03	<input type="checkbox"/> <input type="checkbox"/>	1951	94.4%

Ion maps: ■ Alignment target ■ Run being aligned

Alignment quality: ■ Good ■ OK ■ Needs review

Note: a marked red area combined with a low score clearly indicates a 'misalignment' and may require some manual intervention (see Appendix 4, page 84).

The alignment quality of this tutorial data set does not require any manual intervention so before going to the next section make sure you have removed all manual vectors and re-performed the Automatic alignment.

To do this for C_01 first select 'All vectors in the whole run' from the Remove Vectors and then click Align runs automatically.

Review Alignment
Sample ions are aligned to compensate for drifts in retention time between runs.

1 **Align retention times automatically**
For maximum reproducibility, the software can automatically align your runs.
Align runs automatically

2 **Review the alignment**
Using the quality control measures, review and edit the runs' alignment:
1. Order the runs by alignment score and start by selecting the first run
2. Within each run, inspect and edit any areas rated as Needs Review

Run	Include?	Vectors	Score
A_01	<input checked="" type="checkbox"/>	Ref	
A_02	<input checked="" type="checkbox"/>	1963	97.6%
A_03	<input checked="" type="checkbox"/>	2074	97.4%
B_01	<input checked="" type="checkbox"/>	1982	96.7%
B_02	<input checked="" type="checkbox"/>	1901	96.1%
B_03	<input checked="" type="checkbox"/>	1937	96.4%
C_01	<input checked="" type="checkbox"/>	1980	95.5%
C_02	<input checked="" type="checkbox"/>	2031	96.3%
C_03	<input checked="" type="checkbox"/>	1950	94.4%

Remove Vectors dropdown options:
 Automatic vectors in the current area
 Automatic vectors in the whole run
All vectors in the whole run

Ion maps: ■ Alignment target ■ Run being aligned

Alignment quality: ■ Good ■ OK ■ Needs review

Section Complete

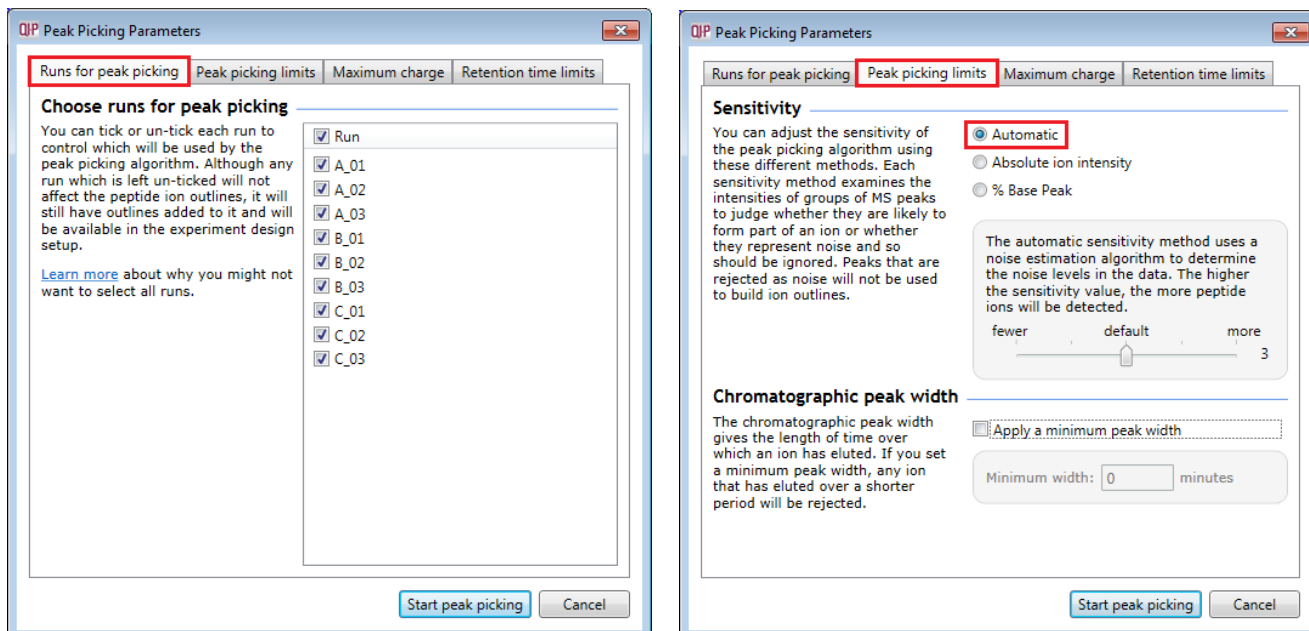
Having aligned the runs automatically, click **Section Complete** to move to Filtering.

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.



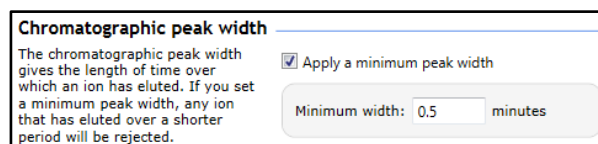
Tip: It may be appropriate **only** to pick peaks that are present in a limited number of your runs. In which case un-tick the runs that you do **NOT** want to contribute to the peptide ion detection pattern. This may be important when one or more of the runs appear noisy due to non-optimal chromatography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

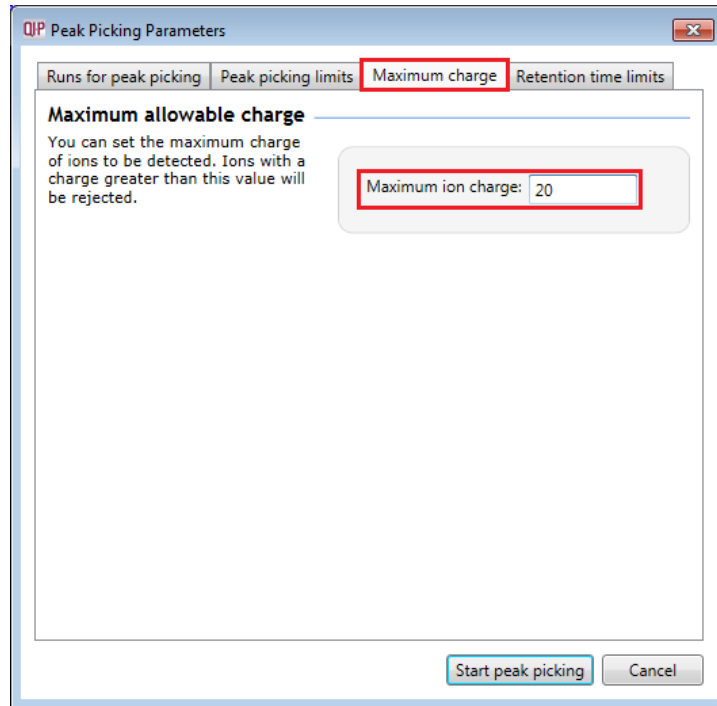
The sensitivity of the detection can be controlled by adjusting settings under the **Peak picking limits** tab.

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.

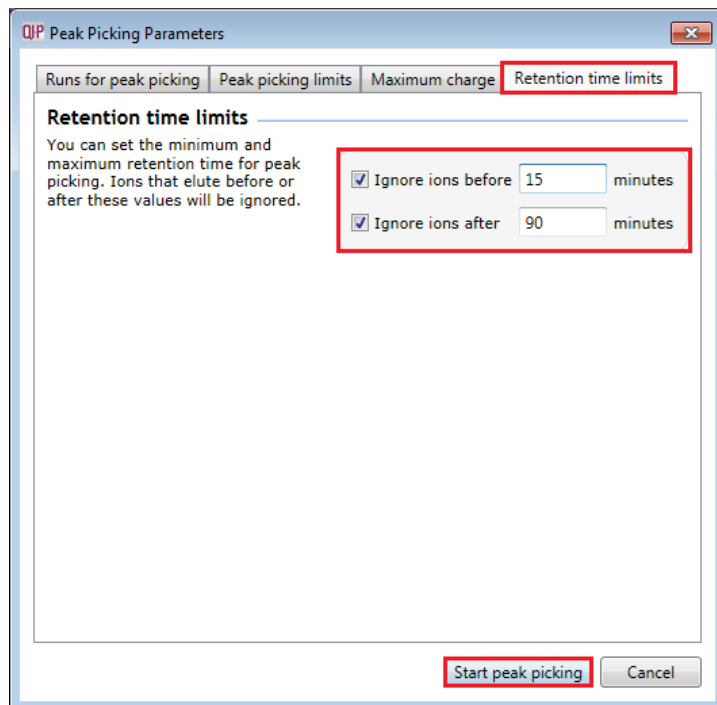


For the runs in this user guide we will use the default settings for the Automatic method and NOT apply a minimum peak width.

The third tab allows you to set the **Maximum charge** of the ions, which will be detected. The default setting is a charge state of 20.



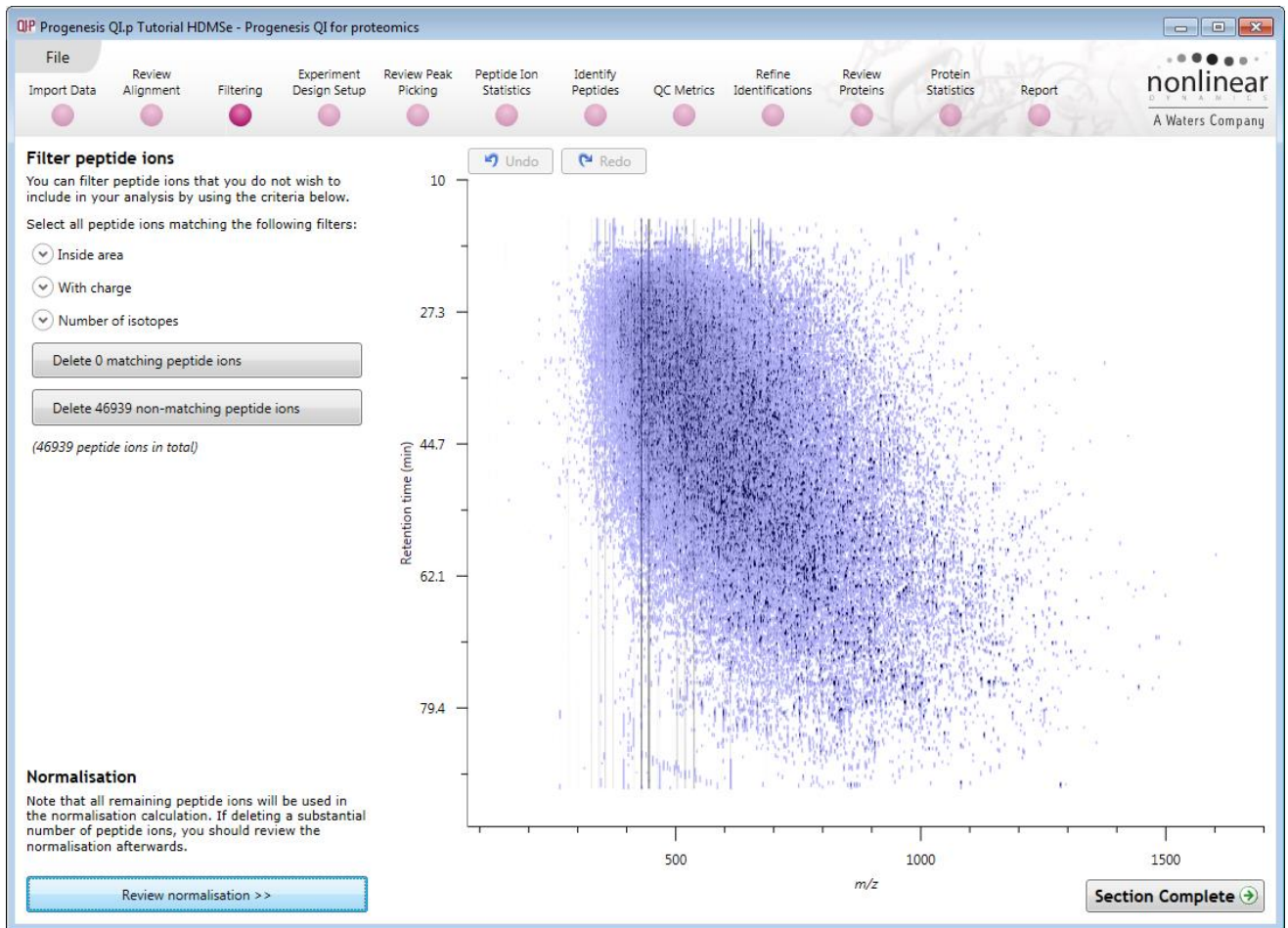
Finally, you can set **Retention time limits** for the detection. Current limits are displayed in the unticked boxes and detection will use these values whenever the box(s) are unticked.



For this dataset we will use the RT settings as shown above 15 and 90 minutes.

Press **Start peak picking** to start the detection process.

On completion of analysis, the Filtering stage will open displaying the number of peptide ions detected, in this example 46939.



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 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.

Tip: when filtering on one property of the peptide ions i.e. with charge, make sure you have 'collapsed' the other filters as expanded filters are applied concurrently.

Filter peptide ions

You can filter peptide ions that you do not wish to include in your analysis by using the criteria below.

Select all peptide ions matching the following filters:

- Inside area
- With charge
 - Charge 1 (10691 peptide ions)
 - Charge 2 (22712 peptide ions)
 - Charge 3 (11233 peptide ions)
 - Charge 4 (1240 peptide ions)
 - Charge 5 (199 peptide ions)
 - Charge 6 (298 peptide ions)
 - Charge 7 (127 peptide ions)
 - Charge 8 (83 peptide ions)
 - Charge 9 (45 peptide ions)
 - Charge 10 (18 peptide ions)
 - Charge 11 (17 peptide ions)

Filter peptide ions

You can filter peptide ions that you do not wish to include in your analysis by using the criteria below.

Select all peptide ions matching the following filters:

- Inside area
- With charge
- Number of isotopes

Filter peptide ions

You can filter peptide ions that you do not wish to include in your analysis by using the criteria below.

Select all peptide ions matching the following filters:

- Inside area
- With charge
 - Charge 1 (10691 peptide ions)
 - Charge 2 (22712 peptide ions)
 - Charge 3 (11233 peptide ions)

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 11130 Non Matching Peptide ions**.

The screenshot displays the Progenesis QI software interface during the filtering stage. The top menu bar includes options like File, Import Data, Review Alignment, Filtering, Experiment Design Setup, Review Peak Picking, Peptide Ion Statistics, Identify Peptides, QC Metrics, Refine Identifications, Review Proteins, Protein Statistics, and Report. The left sidebar shows a list of charge states from 5 to 20, each with a checkbox and the number of peptide ions. The 'Delete 11130 non-matching peptide ions' button is highlighted with a red box. Below this, a 'Normalisation' section provides a note about the calculation and a 'Review normalisation >>' button. The main plot area shows a 2D scatter plot of Retention time (min) on the y-axis (ranging from 10 to 79.4) and m/z on the x-axis (ranging from 500 to 1500). The plot contains a dense cluster of points, with many points colored red. The 'Undo' button is also highlighted with a red box. A 'Section Complete' button is visible in the bottom right corner.

You can use the **Undo** button to bring back deleted peptide ions. For this tutorial we will proceed with the 35809 matching peptide ions.

Before moving on from filtering, you can review the normalisation of the experiment.

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

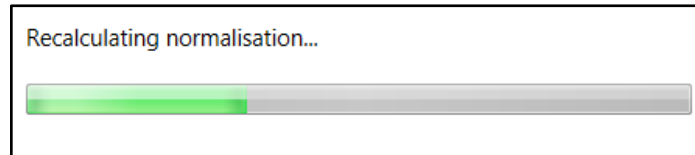
Normalisation review is accessed from the button at the bottom left corner of the filtering page.

Normalisation

Note that all remaining peptide ions will be used in the normalisation calculation. If deleting a substantial number of peptide ions, you should review the normalisation afterwards.

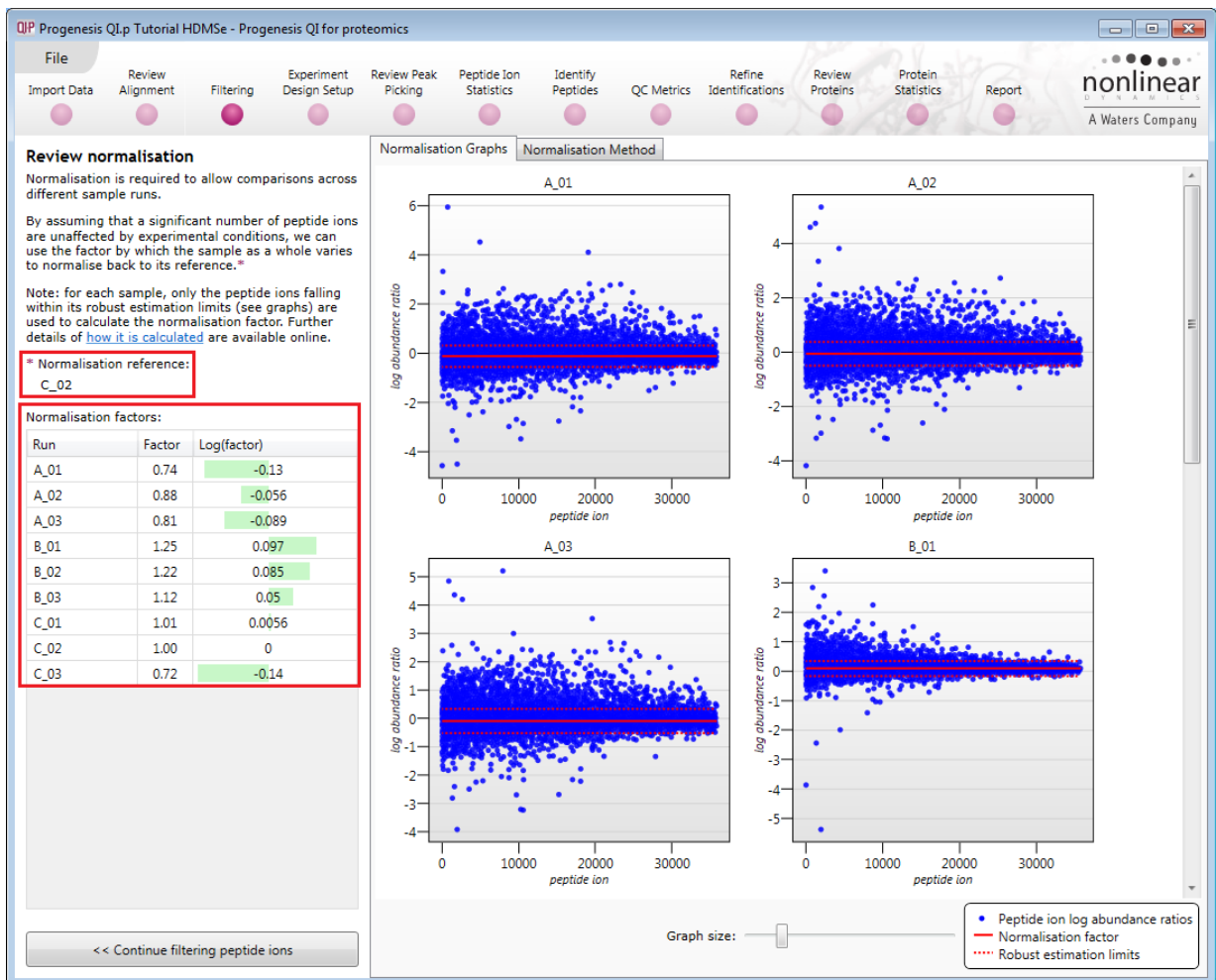
[Review normalisation >>](#)

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.



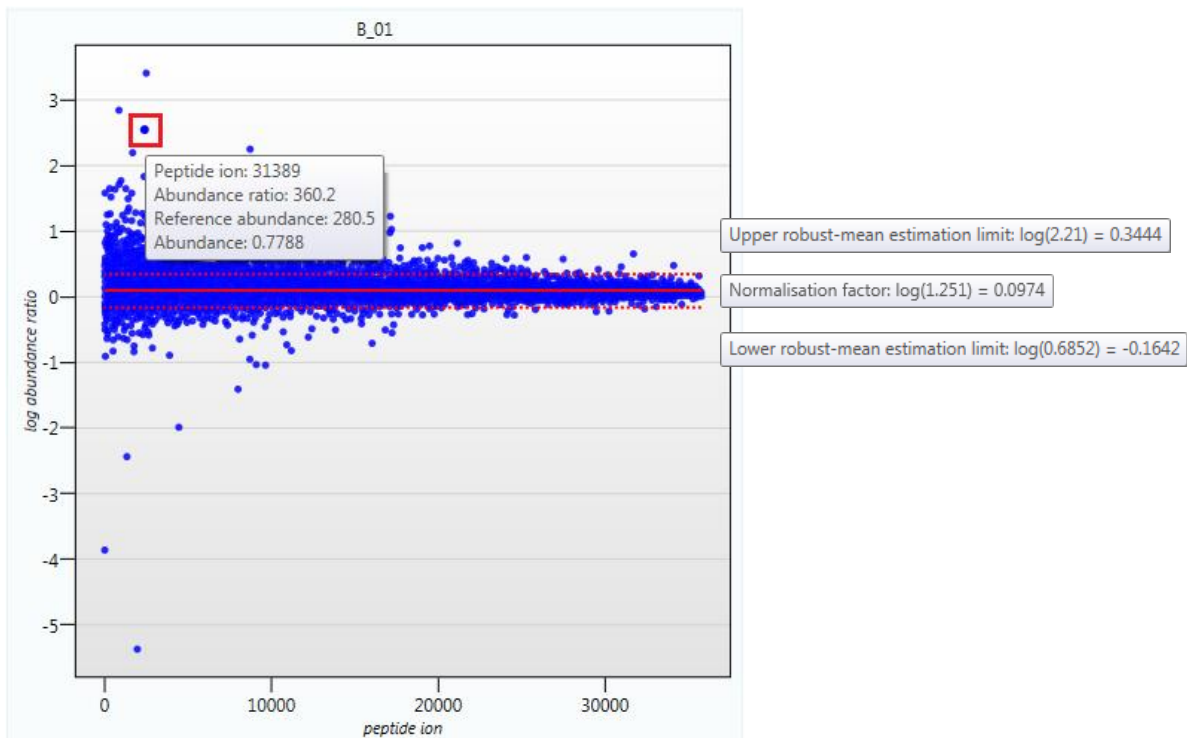
Normalisation factors are reported in the table to the left of the plots.

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 and then set this to be the 'Normalising reference'. The run used, is shown above the table of Normalisation factors (in this example C_02).

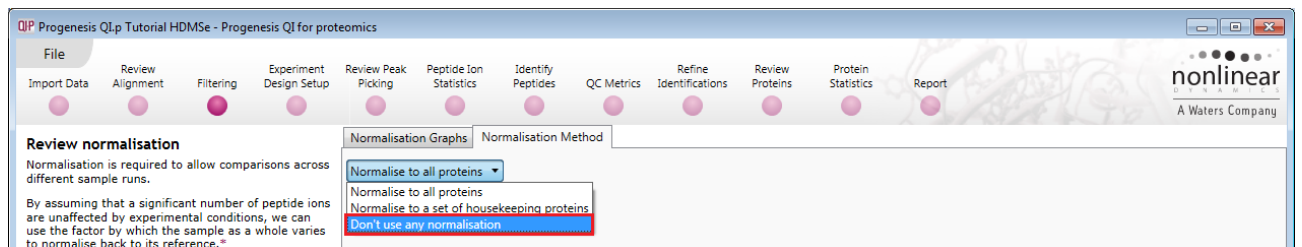
For each sample run, each blue dot shows the log of the abundance ratio for a different peptide ion (normalisation target abundance/run abundance).

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.

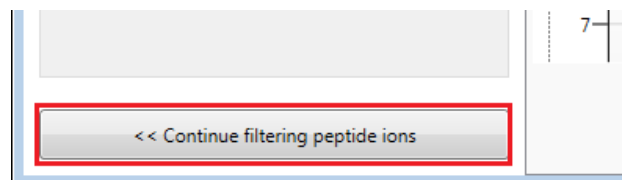
Finally, if you do **not** wish to work with normalised data then you can **use the raw abundances** by switching off the normalisation.



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

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

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



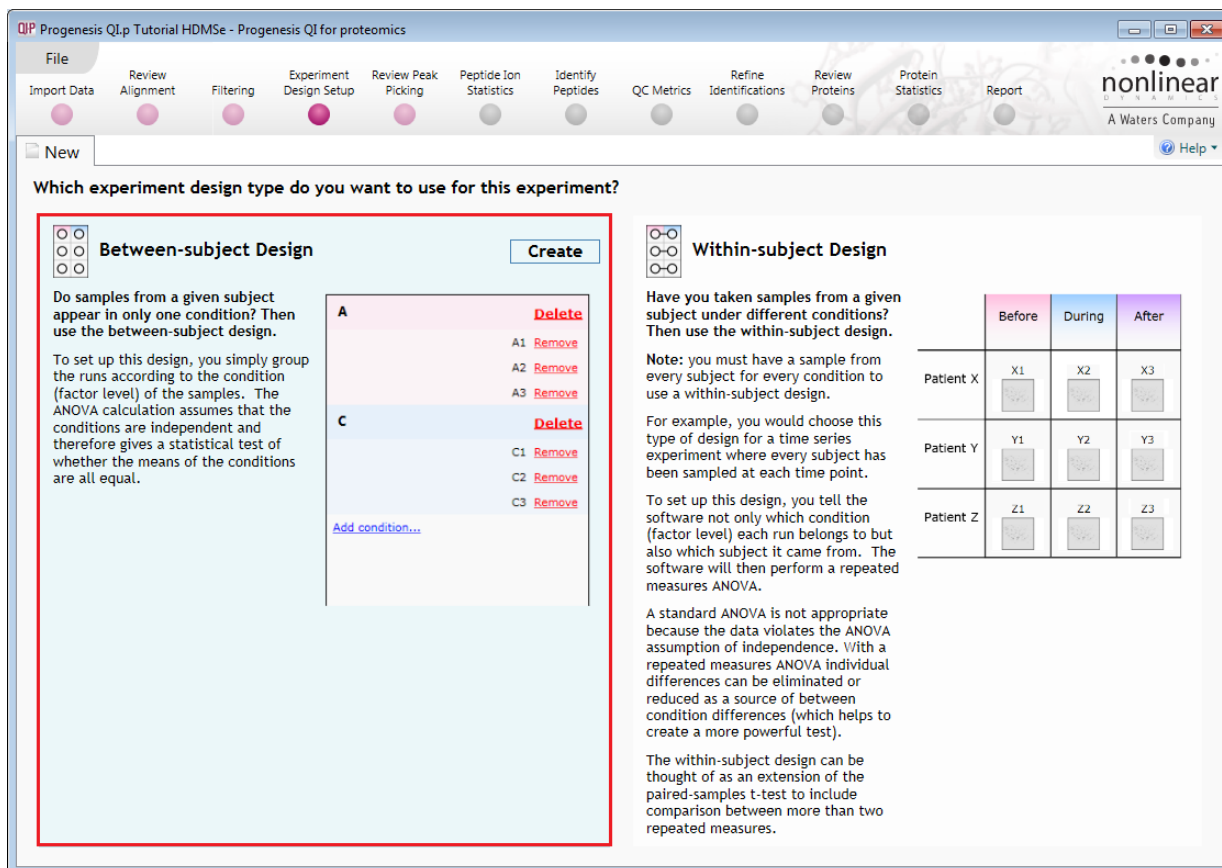
For this example, we **DO NOT** do any additional Filtering so click on **Section complete**.

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.



Which experiment design type do you want to use for this experiment?

Between-subject Design Create

Do samples from a given subject appear in only one condition? Then use the between-subject design.

To set up this design, you simply group the runs according to the condition (factor level) of the samples. The ANOVA calculation assumes that the conditions are independent and therefore gives a statistical test of whether the means of the conditions are all equal.

A Delete

A1 Remove

A2 Remove

A3 Remove

C Delete

C1 Remove

C2 Remove

C3 Remove

[Add condition...](#)

Within-subject Design

Have you taken samples from a given subject under different conditions? Then use the within-subject design.

Note: you must have a sample from every subject for every condition to use a within-subject design.

For example, you would choose this type of design for a time series experiment where every subject has been sampled at each time point.

To set up this design, you tell the software not only which condition (factor level) each run belongs to but also which subject it came from. The software will then perform a repeated measures ANOVA.

A standard ANOVA is not appropriate because the data violates the ANOVA assumption of independence. With a repeated measures ANOVA individual differences can be eliminated or reduced as a source of between condition differences (which helps to create a more powerful test).

The within-subject design can be thought of as an extension of the paired-samples t-test to include comparison between more than two repeated measures.

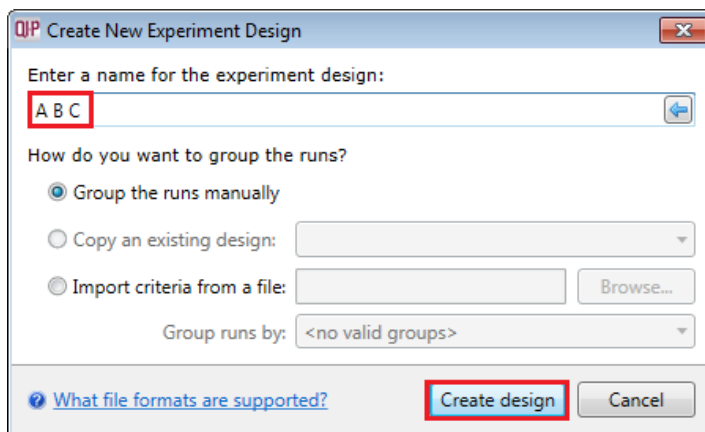
	Before	During	After
Patient X	X1	X2	X3
Patient Y	Y1	Y2	Y3
Patient Z	Z1	Z2	Z3

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 independence. 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 an extension of the paired samples t-test, including comparison between more than two repeated measures.

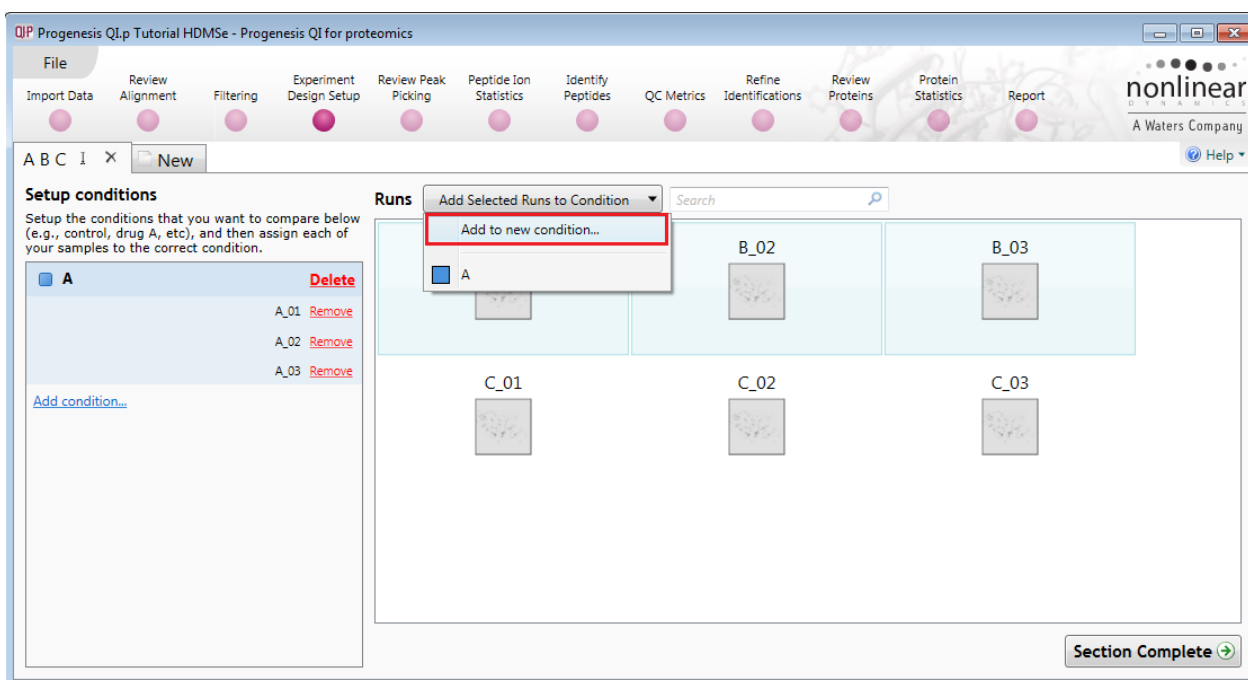
Additional information on how to apply the Within-subject Design is in Appendix 5 page 89

This experiment contains 2 conditions: Control and Treated 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** hold the cursor over this option and click to open the dialog.



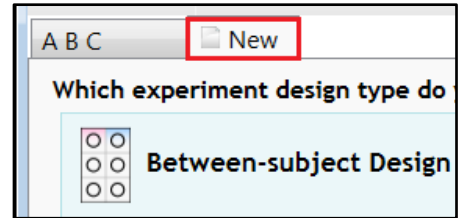
Give the new experiment design a name and then click **Create design**.



To create a new condition

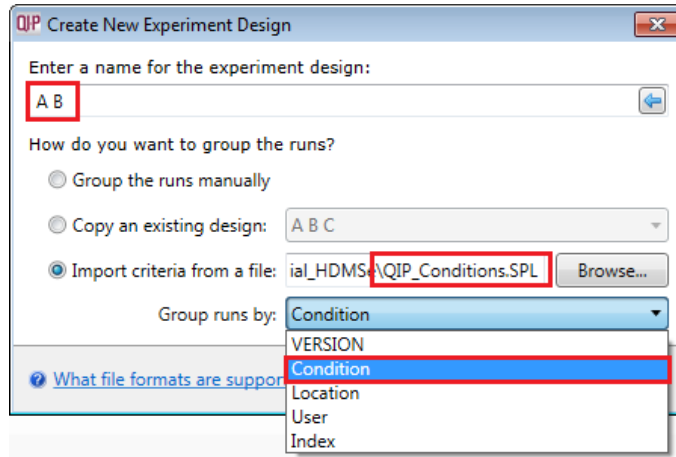
1. Select the runs for the condition by clicking on the required icon in the **Runs** panel, as shown.
2. Press the 'black triangle' next to the **Add Selected Runs to Condition** button on the main toolbar.
3. Select **Add to new condition...** from the drop down menu.
4. A new condition will appear 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 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 spreadsheet at the time of sample collection and/or preparation.



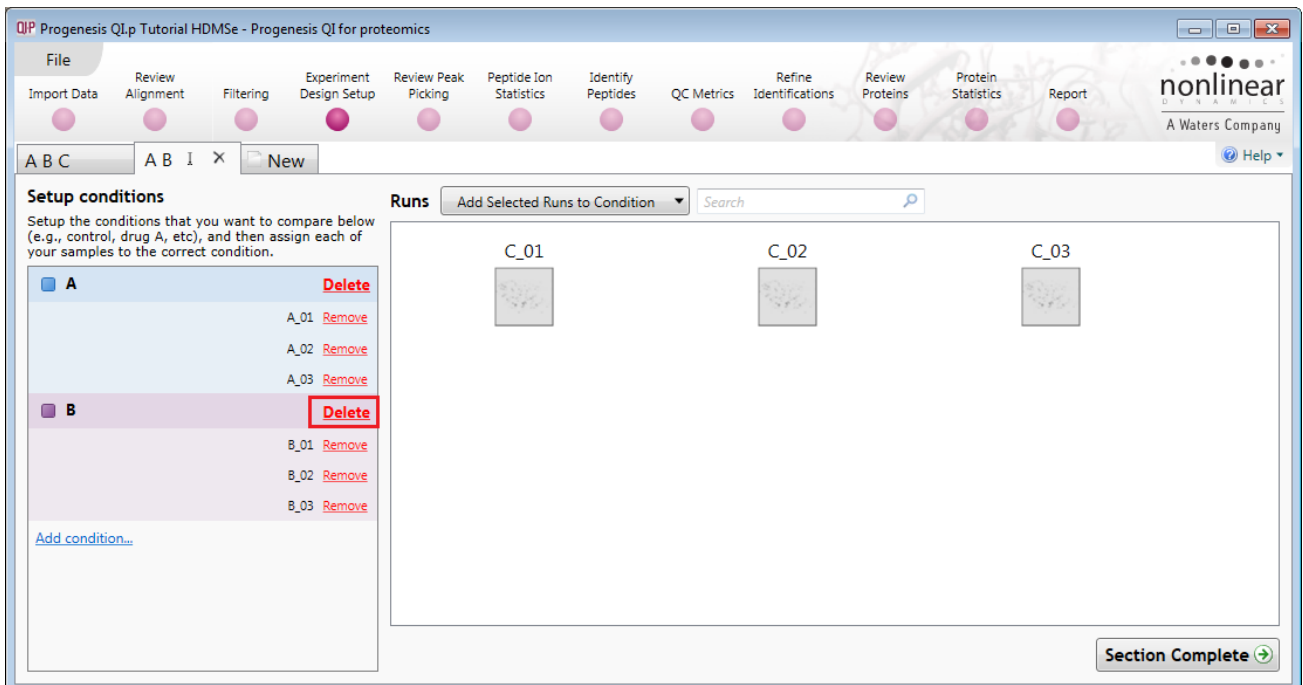
For this example there is a **QIP_Conditions.spl** file available in the Experiment Archive you restored at the beginning of this tutorial exercise.

To use this approach select the **Import criteria from file** option from the New Experiment Design dialog. Then locate the QIP_Conditions file and select what to **Group by**, for example: **Condition**.



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

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**.

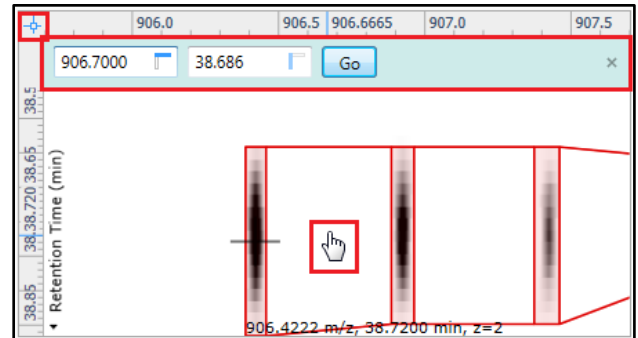
Stage 7: Review Peak Picking and editing of results

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 5 display modes: 1D, 2D, 3D, Drift Time 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

To set up the display as shown below, click on the Go To Location on the top right of the zoomed run view (right). Enter 906.700 m/z and 38.686 min and click **Go**. Then click on/or inside the peptide ion boundary to re-focus the view and table.



Window A: shows the list of peptide ions ranked by the p value for the one way **Anova** using the current Experiment Design (A B C).

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

Review Peak Picking

No filter applied

#	Anova (p)	q Value	Fold	Tag	Notes
40179	1.43E-10	7.94E-08	Infinity		
40794	2.98E-10	1.45E-07	Infinity		
4417	4.18E-10	1.8E-07	1.29		
46534	6.37E-10	2.47E-07	Infinity		
33422	7.31E-10	2.58E-07	Infinity		
25125	9.93E-10	3.21E-07	7.79		Ad
37967	1.1E-09	3.28E-07	Infinity		
15048	1.49E-09	4.1E-07	3.35		
17182	1.59E-09	4.1E-07	2.16		

Mass spectrum (38.72 min) Chromatogram (m/z = 906.4239)

1D Display 2D Display 3D Display Drift Time

Run: Expression profile

Aggregate

Select Edit

Split Merge

Add Delete

Undo Redo

Charge state: 1 2 3 4 >=5

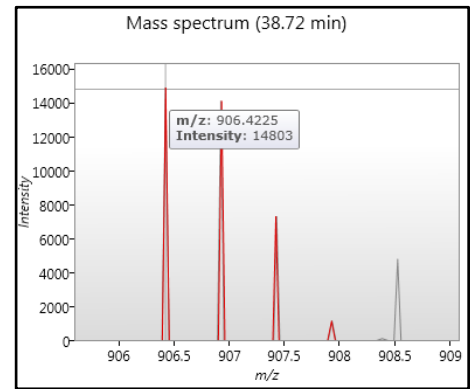
Specific drift time: Drift time: 5.11 ms

Experiment design: A B C

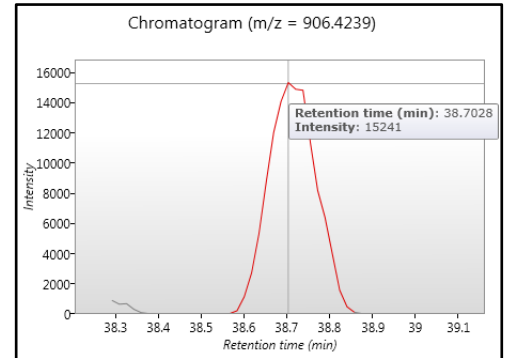
Section Complete

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

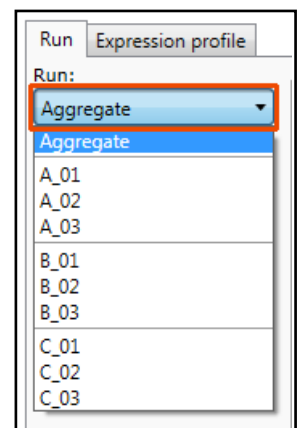


Window C: displays the Chromatogram for the current peptide ion on the selected Run (in window D). Hold the cursor over the peak to indicate Retention time and intensity.



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



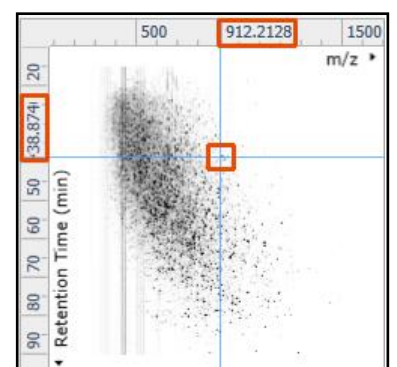
The peptide ion editing tools are located in this window (see page 41 for functional explanation).

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.



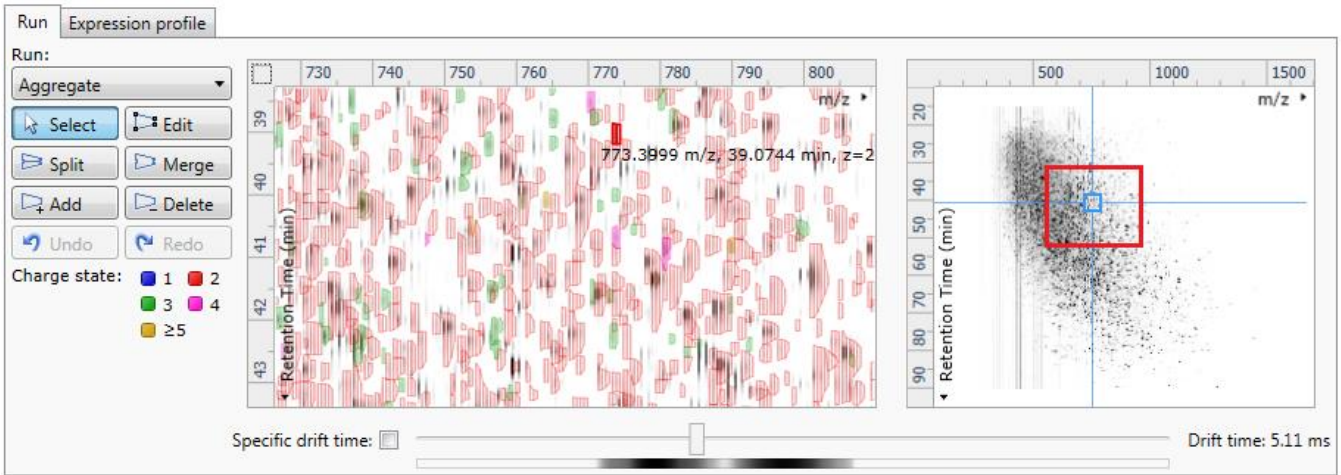
Window E: shows where the current peptide ion is located on the LC-MS run by means of the cross hairs.

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.

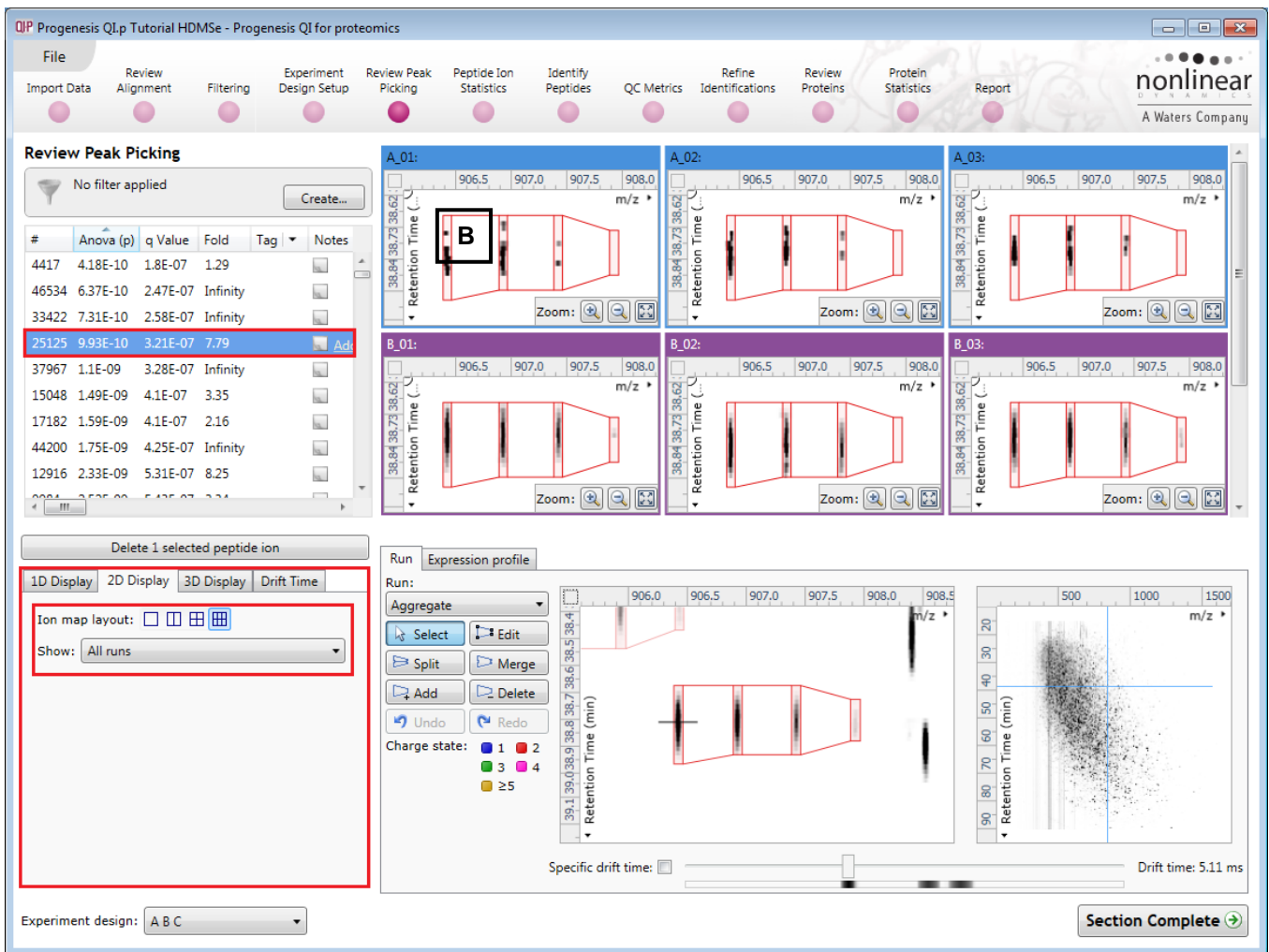
You can also drag out an area (blue square) on this view that will re-focus the other windows.



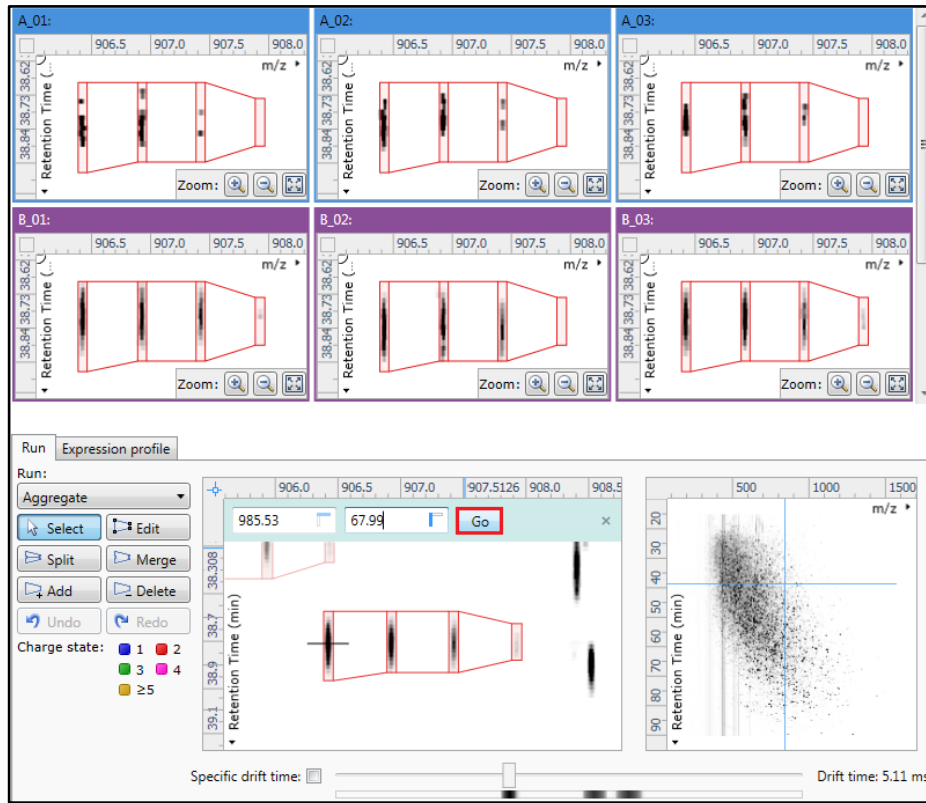
The 2D Display

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

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



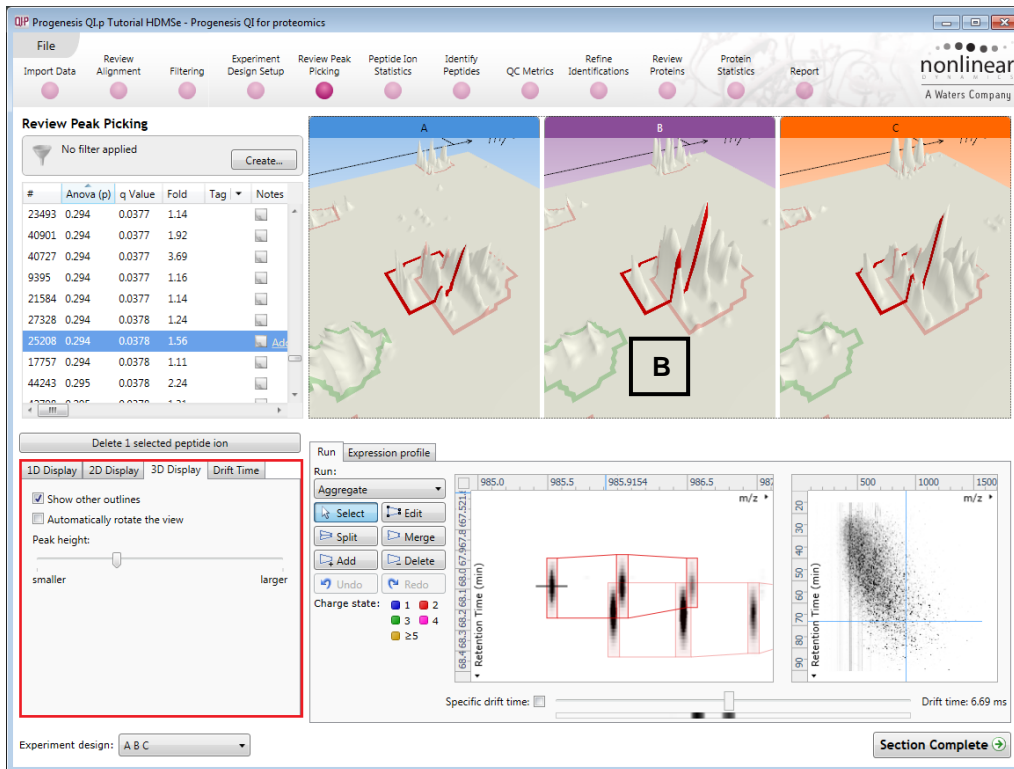
The appearance of the Montage (window B) is controlled by the panel on the bottom left of the display.



Using the 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 below. (i.e Go To Location 985.53m/z and 67.99min)

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 **Automatically Rotate** or you can rotate them manually by clicking and dragging them with the mouse.

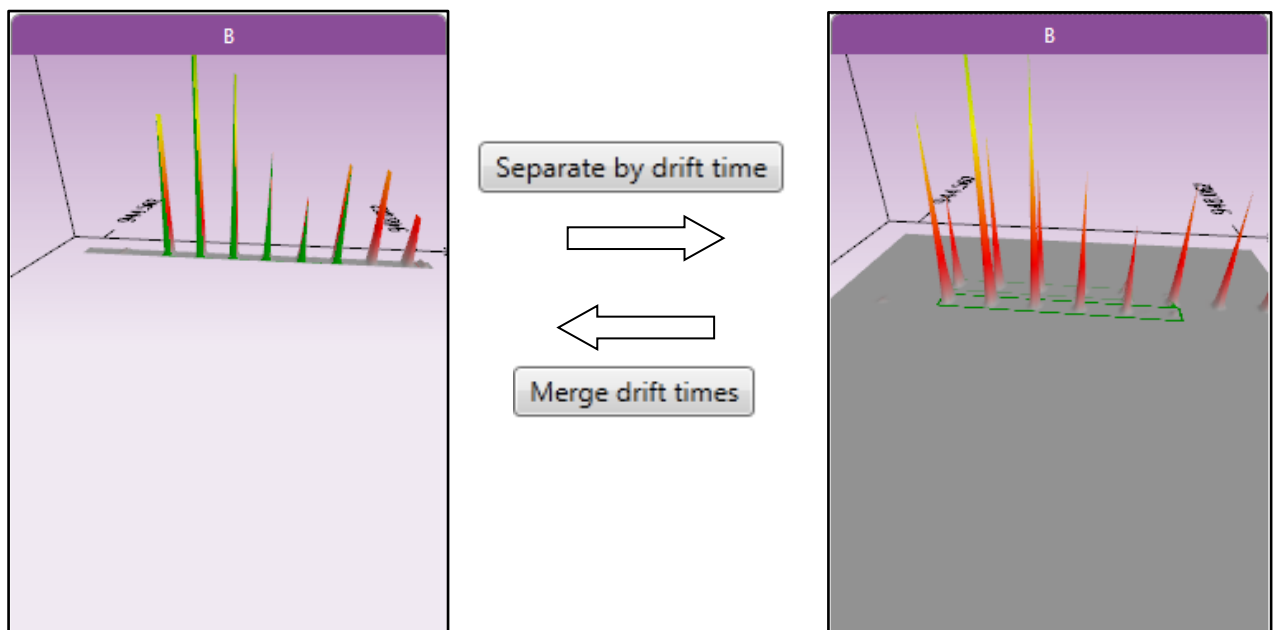
Drift Time Display

When the Drift Time tab (F) is selected in the bottom left of the display Window B changes into a 3D view displaying the Drift Time dimension (to examine this example order table on # and select 5285.

The screenshot shows the Progenesis QI software interface. The 'Review Peak Picking' window is active, displaying a table of peptide ions. The table has columns for #, Anova (p), q Value, Fold, Tag, and Notes. The row for peptide ion #5285 is highlighted in red. Below the table, there are options for '1D Display', '2D Display', '3D Display', and 'Drift Time'. The 'Drift Time' tab is selected, and the '3D montage: m/z vs. drift time' option is chosen. The 'Separate by drift time' and 'Merge drift times' buttons are visible. The 3D views (A, B, and C) show the peptide ions as peaks in a 3D space, with the Drift Time dimension being expanded.

#	Anova (p)	q Value	Fold	Tag	Notes
5280	1.39E-05	6.84E-05	1.42		
5282	0.0205	0.0049	1.21		
5283	3.26E-06	3.32E-05	1.63		
5284	0.0801	0.0134	1.07		
5285	0.0383	0.00771	2.02		
5286	0.00981	0.00294	1.16		
5288	0.00426	0.00168	1.43		
5289	0.000301	0.000341	1.66		
5290	0.213	0.029	1.08		

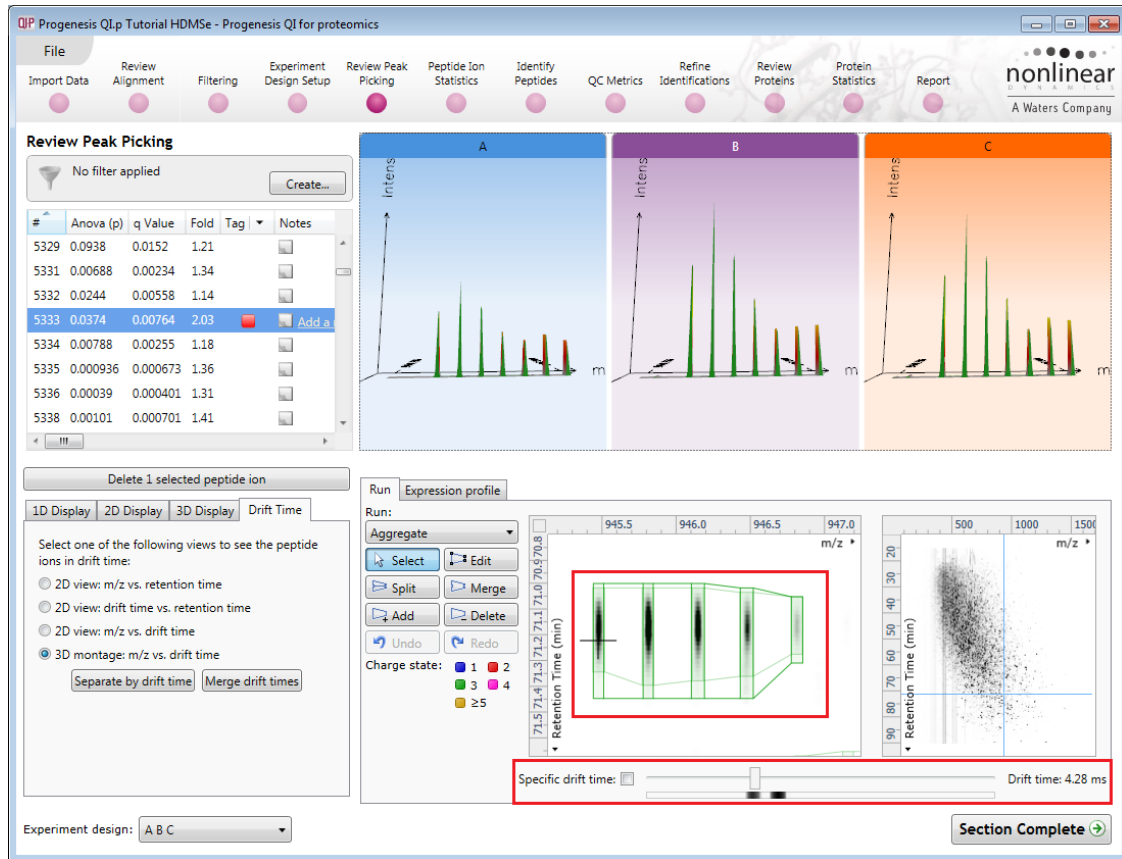
To view the drift time dimension for the current peptide ion, click on **Separate by drift time** on the Drift Time tab. This will expand the view showing the drift time separation between the detected peptide ions (below right)



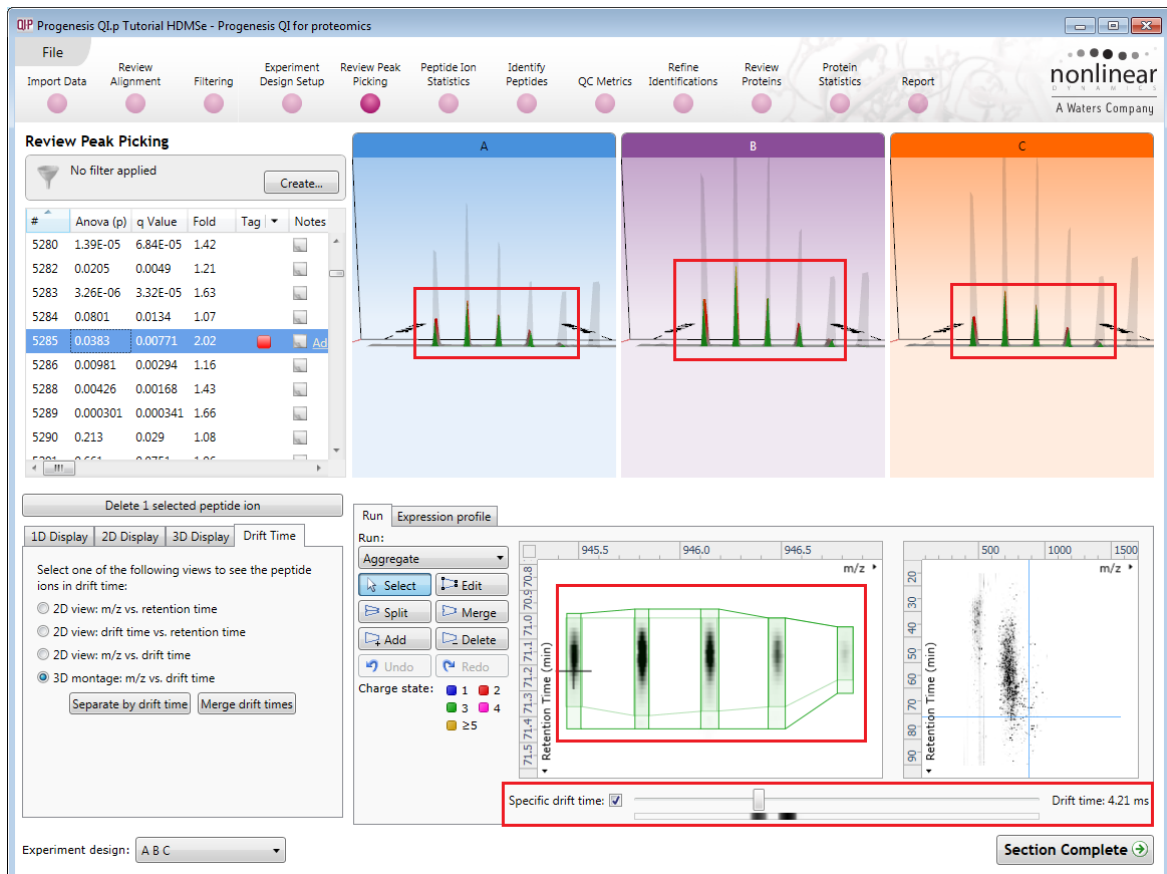
The left hand view displays the merged Drift Times.

Note: you can manipulate the orientation of the views by clicking on them and dragging the display to the required orientation. You can also zoom in and out of the panels by using the 'scroll' wheel on your mouse

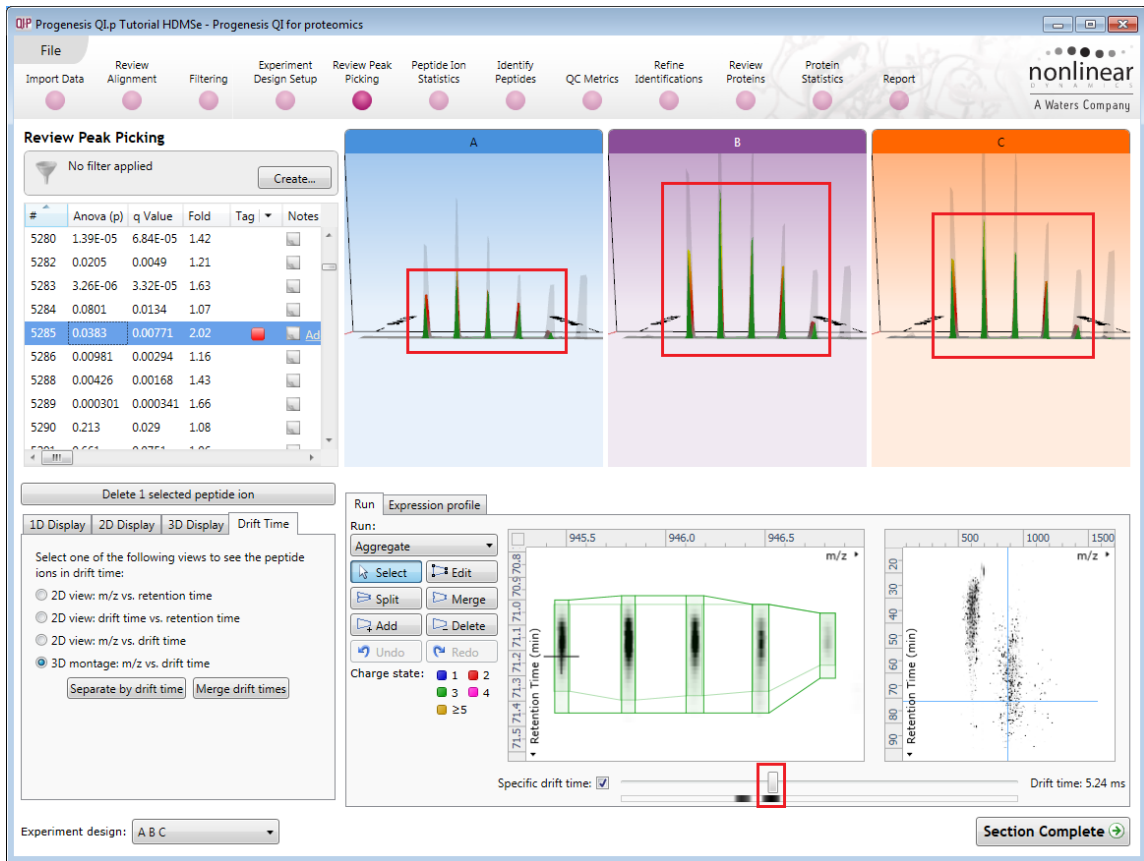
Note: you can step through the specific **Drift Times** (measured in milli seconds) for the current peptide ion by clicking on the **Specific drift time** tick box at the bottom of the display.



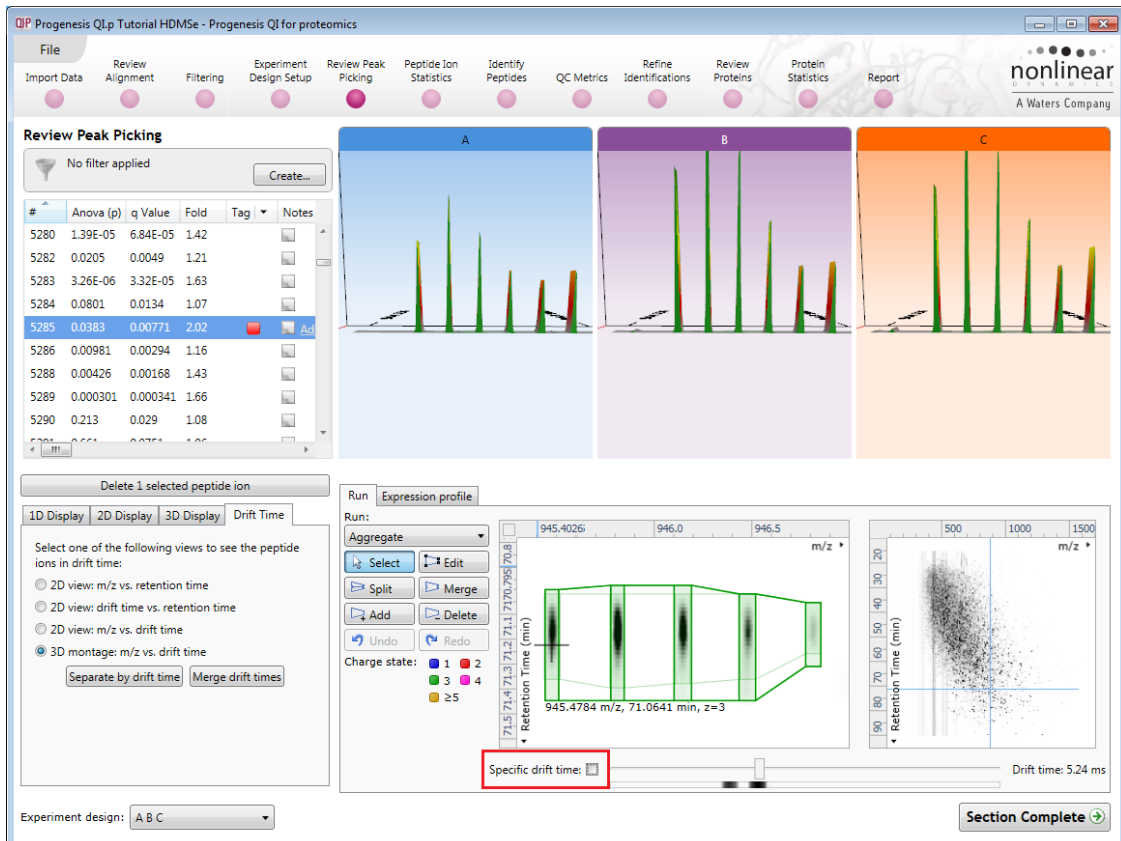
The 'crosshairs' on the peptide ion in the **Run** view identifies it as the current peptide ion in the table.



As you move the slider over the intense areas, indicated below, all the views update to the corresponding drift time.



Note: the crosshairs will remain on the original peptide ion in the table as you explore the Specific Drift times. When you un-tick the 'Specific drift time' tick box, the 3D views will return to showing the Merged Views for the current peptide ion in the table.

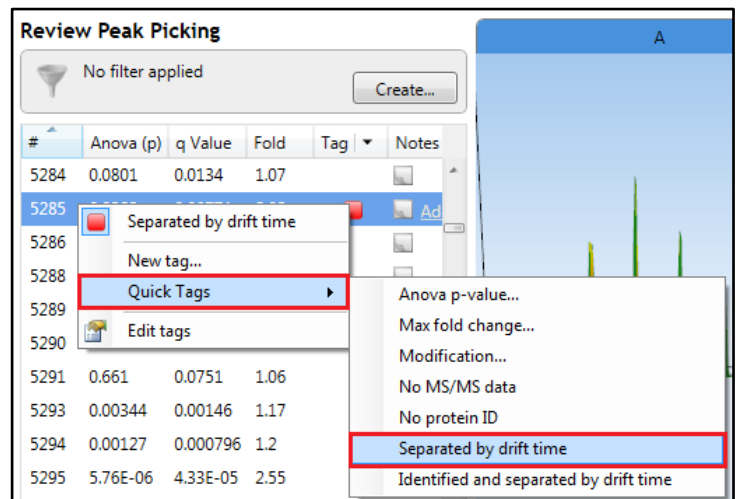
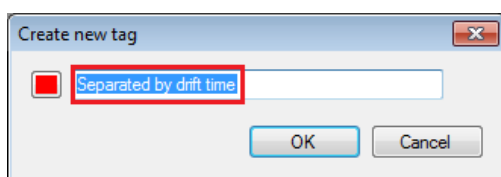


Using Quick Tags to locate examples of Drift Time

In the previous section, describing how to view Drift Time, you may have noticed the presence of a red 'Tag' in the table next to the peptide ion that we examined. Progenesis QI for proteomics allows you to assign tags based on the properties of detected peptide ions either through the manual sorting of the table or making use of the 'Quick Tags'. These tags can be used to filter the list of displayed data in order to aid exploration of the data.

To create a Quick tag for all peptide ions demonstrating separation by Drift time, right click on the table. Select **Quick Tags** then **Separated by drift time**.

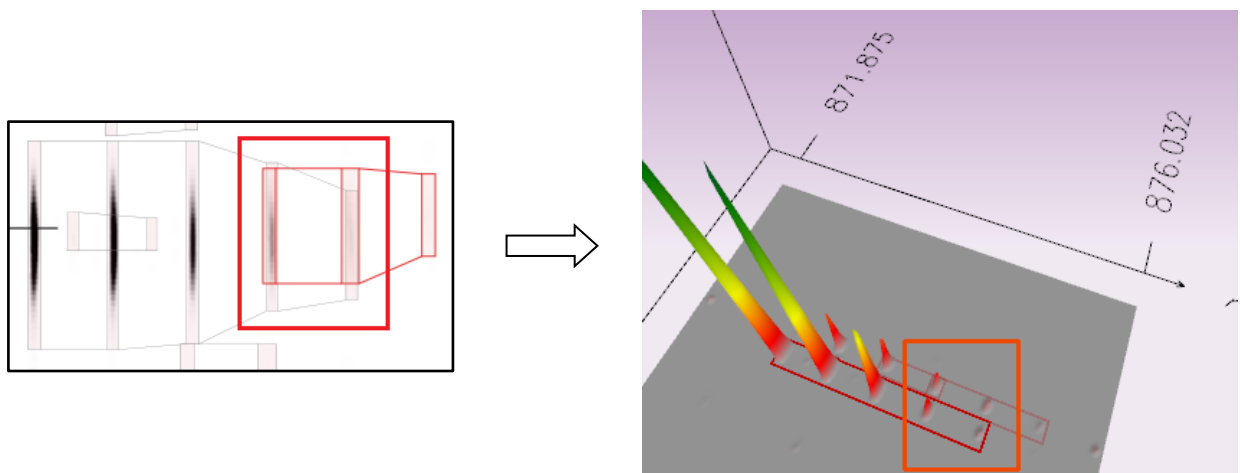
In the new tag dialog either accept or overwrite the tag name.



When the tag is created it will appear against those peptide ions that meet the criteria for the creation of the tag, in this case:

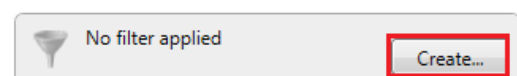
It tags peptide ions that overlap in both m/z and retention time but do not show an overlap in the drift time dimension i.e. those peptide ions that drift time has separated

For example the peptide ions below is overlapping at the same m/z and RT but are **separated in drift time**



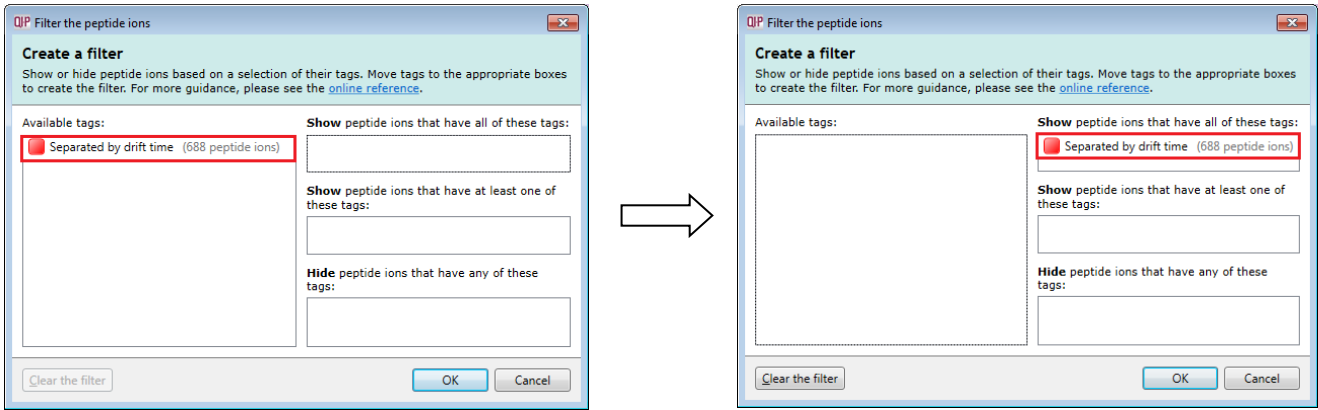
Now filter the table so that it currently only displays a list of peptide ions containing the **separated by drift time** tag.

Click on **Create** on the filter panel above the table.



This will open a Tag filter dialog, in this example, displaying that you have created assigned the **Separated by drift time** tag to 683 peptide ions in your experiment.

To display only those peptide ions containing this tag drag the **Separated by drift time** tag on to the **Show** panel and click OK.



When you apply the tag filter the table will now only display the peptide ions with the appropriate tag(s).

#	Anova (p)	q Value	Fold	Tag	Notes
390	0.00519	0.00192	1.49		
448	1.74E-05	7.69E-05	1.39		
483	0.000121	0.000211	1.75		
716	0.0539	0.00992	1.22		
783	0.011	0.00318	1.51	Ag	
797	0.00207	0.00107	1.24		
919	0.168	0.0239	1.25		
957	0.117	0.0179	1.17		

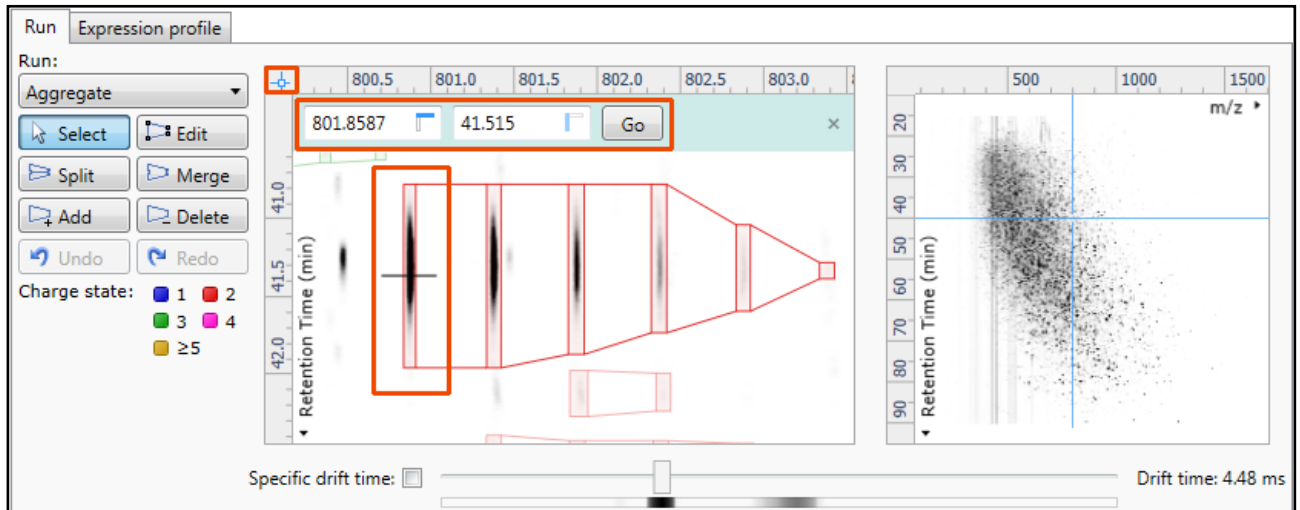
Note: with this **Tag filter applied** you can easily review the effect of Drift time separation for the peptide ions.

To remove the filter click on **Edit**, above the table, and **Clear the filter** followed by **OK**.

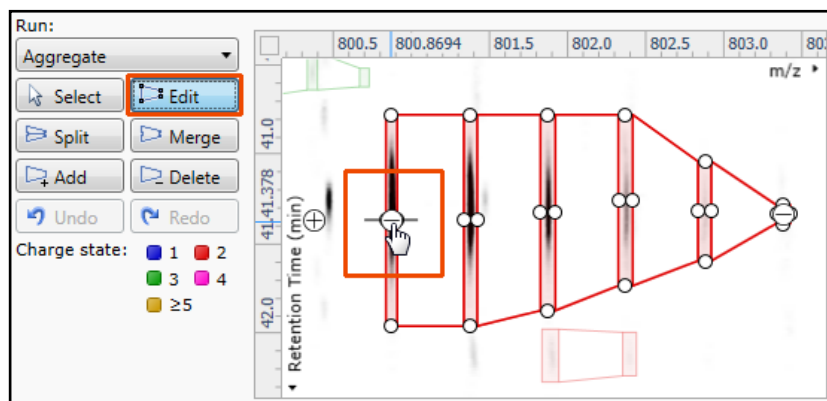
Editing of peptide ions in the View Results stage

As an example of using the editing tools which are located on the left of the LC-MS 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.

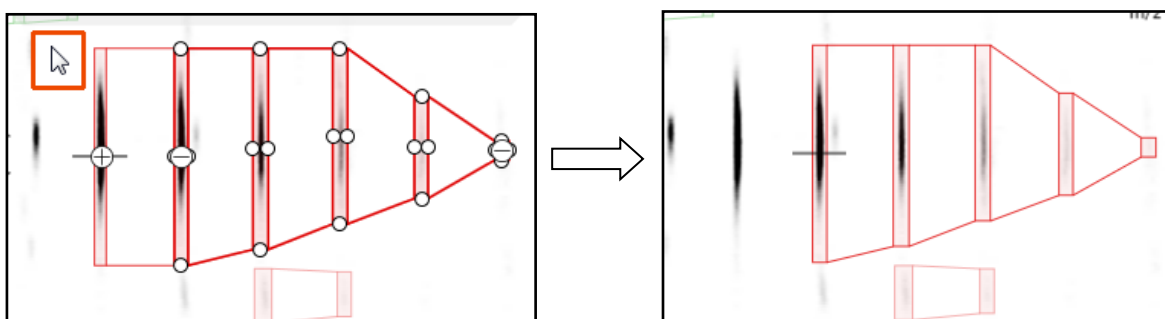
1. Locate the peptide ion at approx 801.86 m/z and 41.5 min using the **Go To Location** tool (top left of zoomed ion map), right click and zoom out and click the cross hairs on the monoisotopic peak to set the zoom.



2. Select the **Edit** tool and click on the peptide ion (in the Run view) to reveal the 'edit handles'. You can zoom in more by dragging an area around the peptide ion of interest.

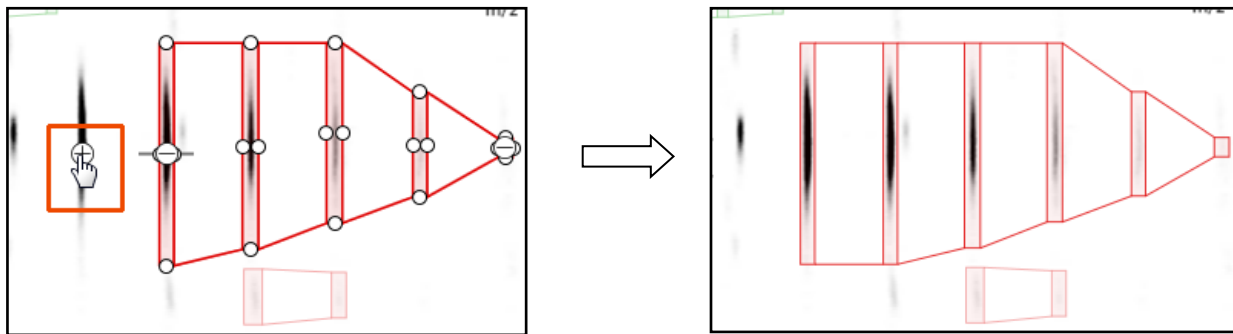


3. Click on the 'minus' handle over the monoisotopic peak to remove it.



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.

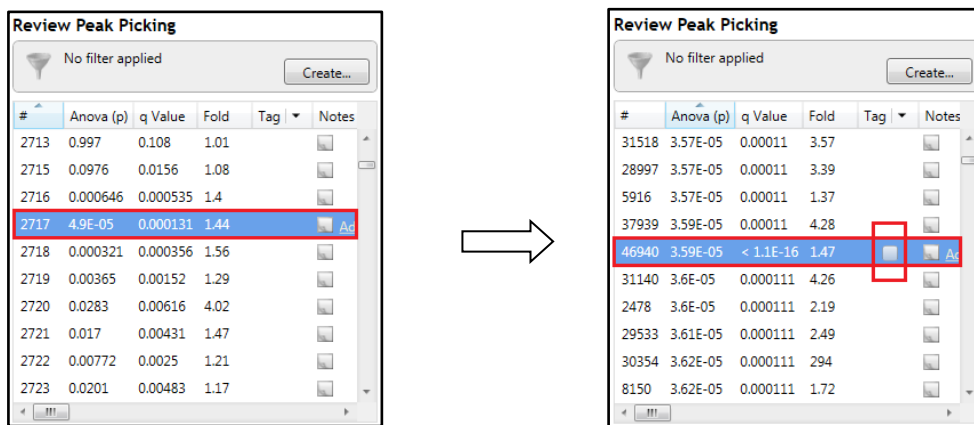
6. Click on the 'plus' handle on the peak to add it.



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

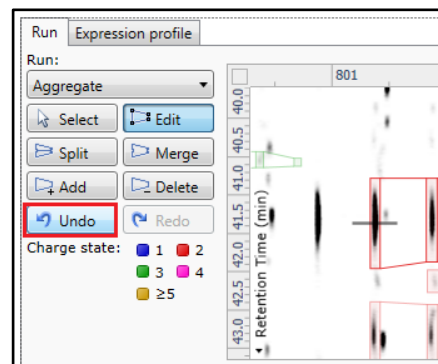
8. **Note:** If you are not satisfied with the editing use the **Undo** button and retry.

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.

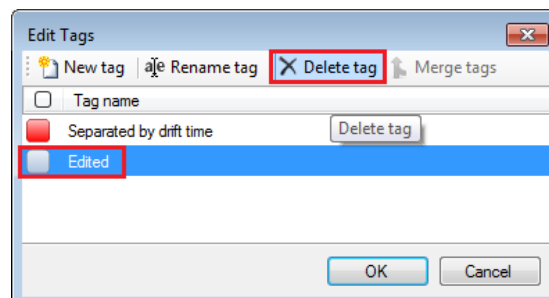


The other tools: **split**, **merge**, **add** and **delete** behave in a similar fashion and their use can be combined to achieve the desired results.

Note: if you have been exploring the editing of peptide ions then a quick way of reversing the edit is to repeatedly press **Undo**.



Then right click on the table and select **Edit tags**. Then select the automatically added **Edited** tag and press **Delete tag**.



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 29). An example is described below.

First expand the 'Peptide ions' table to show all the details by double 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 > 2x10⁴.

#	Anova (p)	q Value	Fold	Tag	Notes
18180	5.19E-05	0.000137	1.35		
21698	5.2E-05	0.000137	1.54		
35011	5.2E-05	0.000137	6.99		
28122	5.2E-05	0.000137	2.58		
2716	5.21E-05	0.000137	1.44		
10738	5.21E-05	0.000137	1.59		
26113	5.22E-05	0.000137	10.2		
15956	5.22E-05	0.000137	1.91		
12382	5.24E-05	0.000137	2.98		
34456	5.24E-05	0.000137	2.68		

Right click on the highlighted peptide ions and select 'New Tag'. Give the Tag a name. i.e. '**Most Abundant**'.

The screenshot shows the Progenesis QI software interface. The 'Review Peak Picking' window is open, displaying a table of peptide ions. A right-click context menu is open over the highlighted row (peptide ion 2716), with 'New tag...' selected. Below the table, there are chromatograms and a 'Section Complete' button.

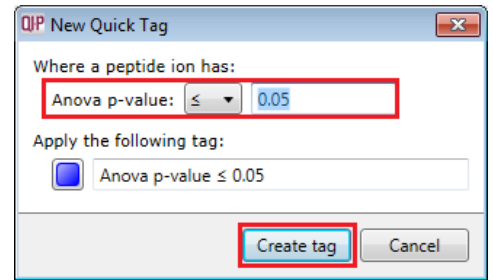
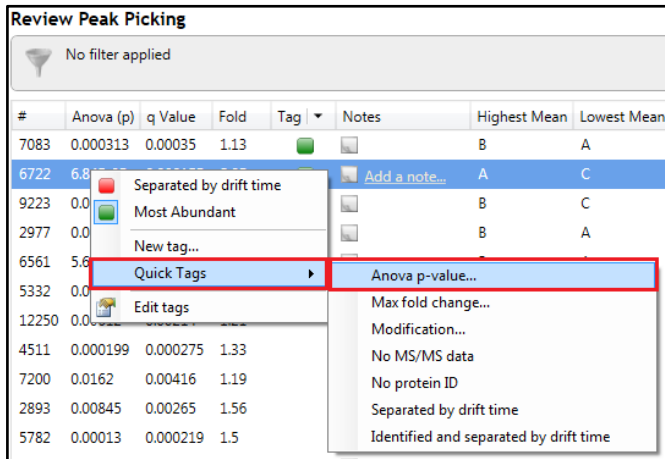
Note: there is already a red tag present that was assigned to those peptide ions that are **Separated by drift time**, which you created in the previous section.

The 'Create new tag' dialog box is shown. The text 'Most Abundant' is entered in the input field. The 'OK' button is highlighted.

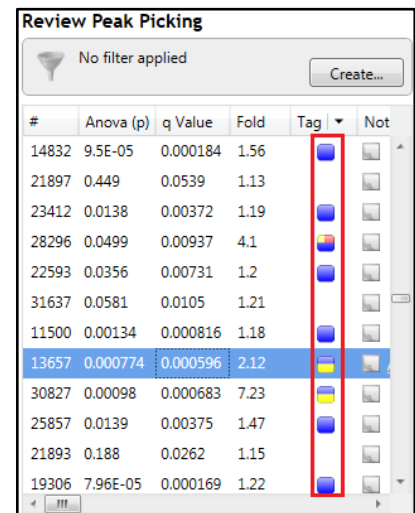
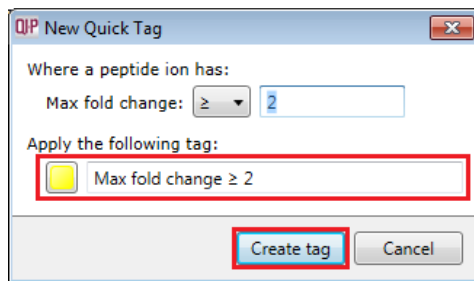
On clicking **OK** the Tag is added to the peptide ions highlighted in the table (signified by a coloured square, green in this example).

#	Anova (p)	q Value	Fold	Tag	Notes
7083	0.000313	0.00035	1.13		
6722	6.84E-05	0.000155	2.05		
9223	0.0263	0.00585	1.11		
2977	0.000161	0.000248	1.47		
6561	5.64E-05	0.000142	1.63		
5332	0.0159	0.0041	1.26		
12250	0.00612	0.00214	1.21		
4511	0.000199	0.000275	1.33		
7200	0.0162	0.00416	1.19		
2893	0.00845	0.00265	1.56		

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**.

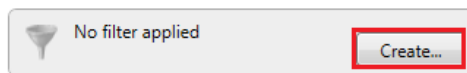


Once this tag appears against peptide ions in the table right click on the table again and create another Quick Tag, this time for peptide ions with a **Max fold change ≥ 2**



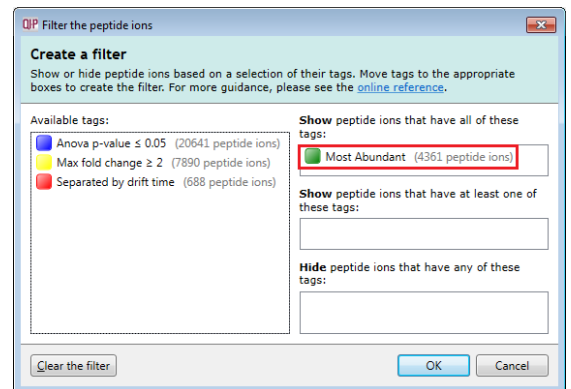
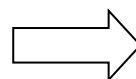
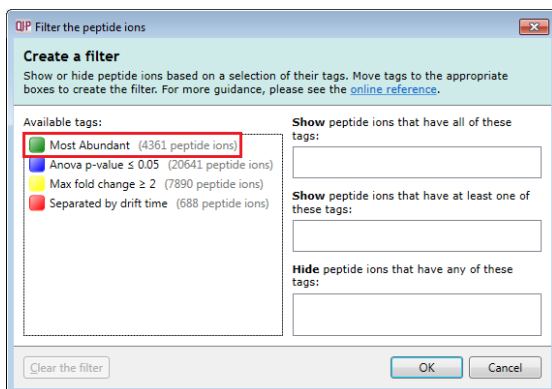
The table now displays peptide ions with multiple tags.

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 are **Most Abundant** 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.

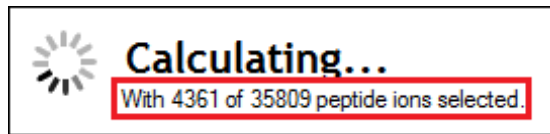


To move to the next stage in the workflow, Peptide Ion Statistics, click **Section Complete**.

Stage 8: Peptide Ion Statistics on Selected Peptide ions

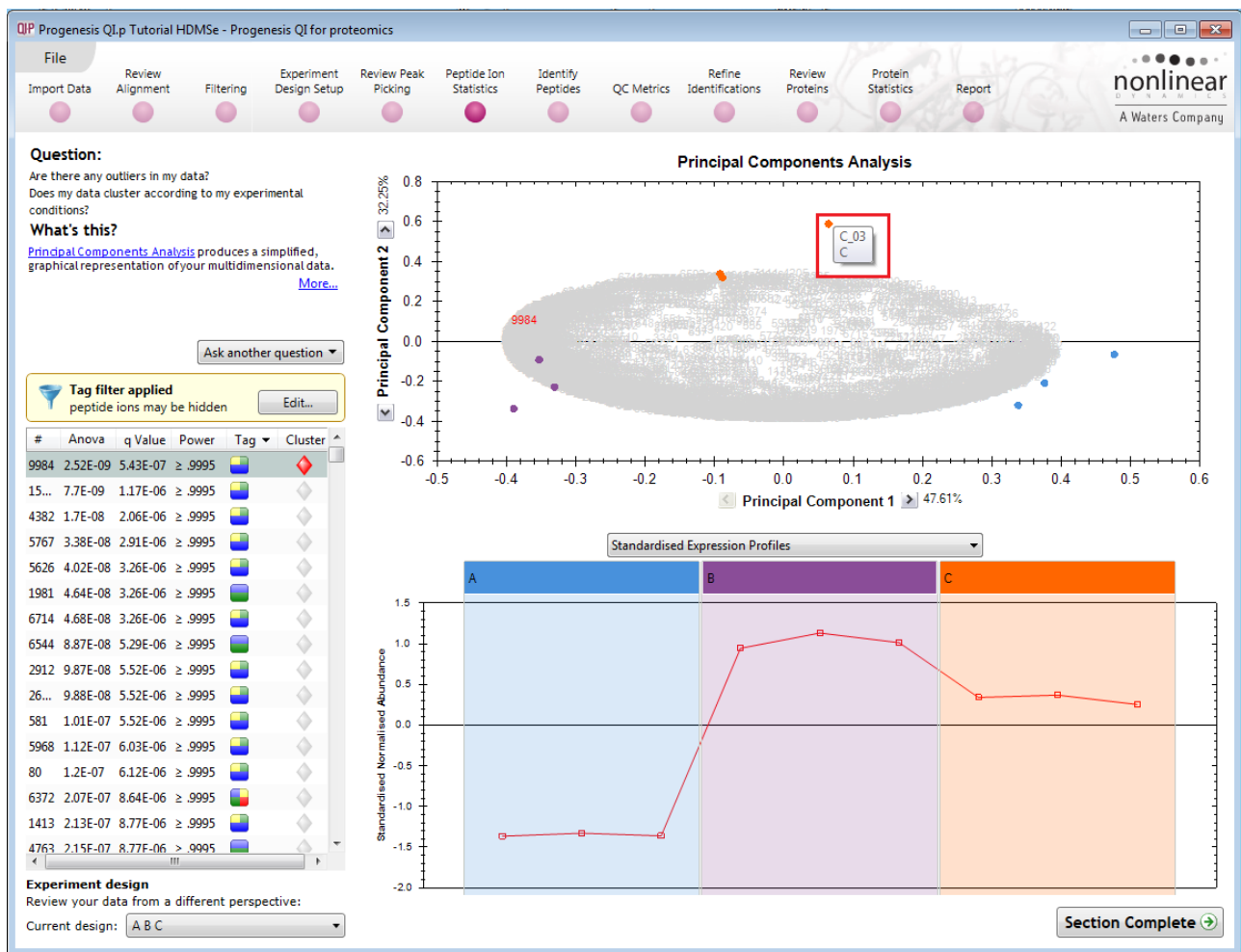
The user guide now describes the functionality of the Multivariate Statistics.

Peptide Ion 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.

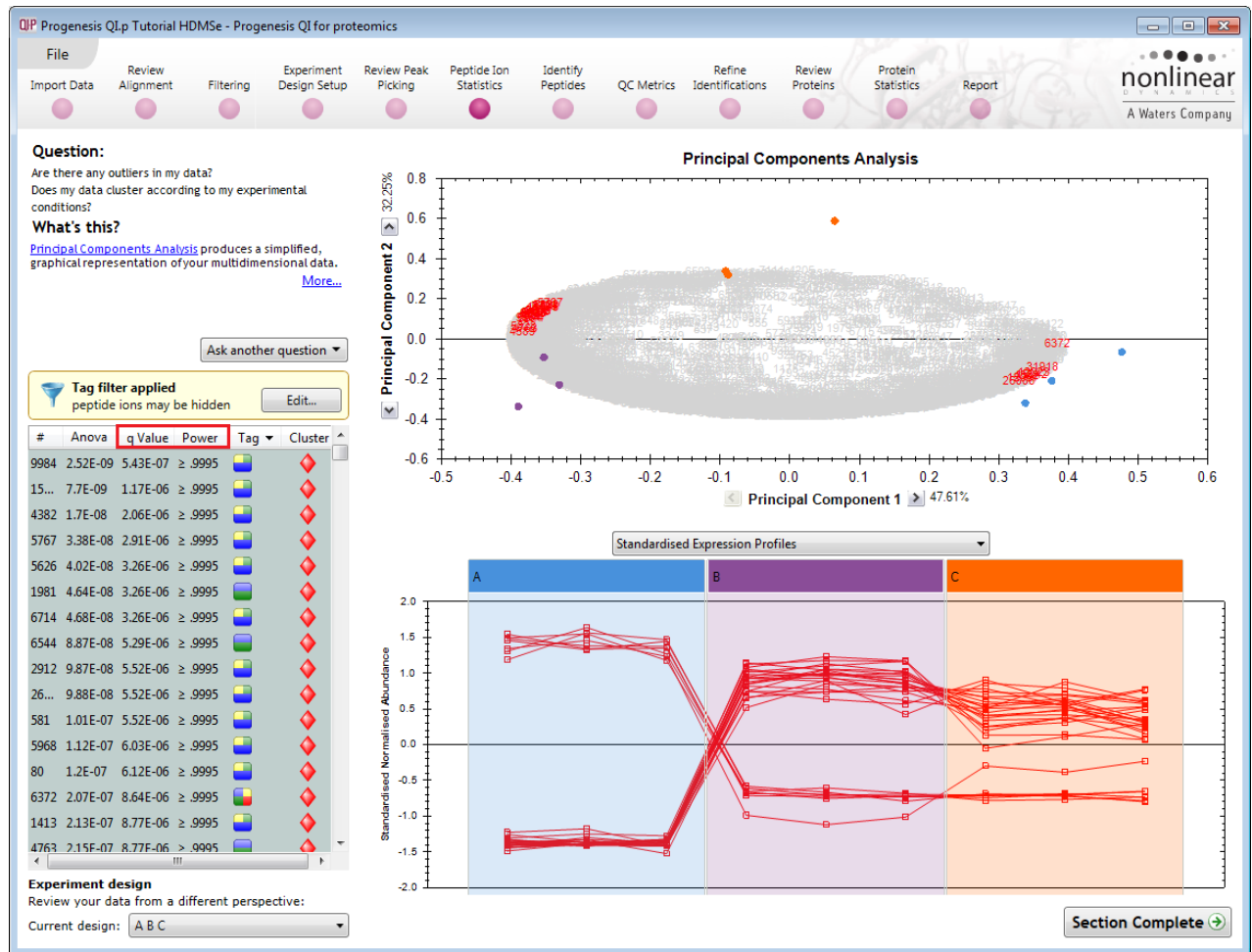
Principal Component Analysis (PCA)

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.

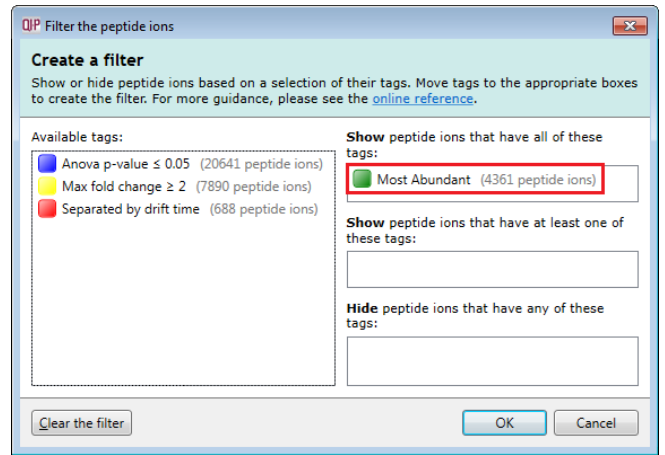
Also, for a given power of 80% we can determine how many samples are required to ensure we find a difference if it actually exists.

Note: Power analysis is discussed in Appendix 6 (page 91)

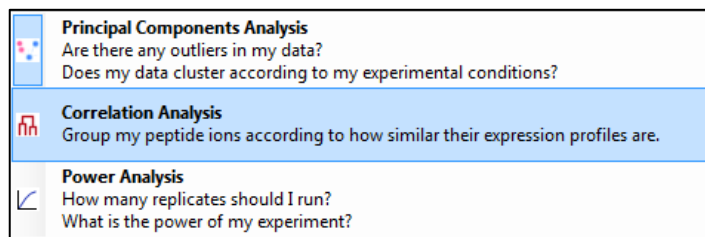
Correlation Analysis

With the tag filter still set to display only the top 4361 **Most Abundant** peptide ions, we are going to explore the Correlation Analysis of these peptide ions.

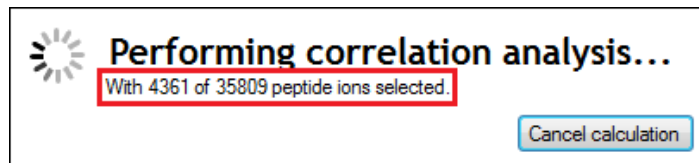
To set up the **Correlation Analysis** using this filtered data set click on **Ask another question** (above the table)



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'



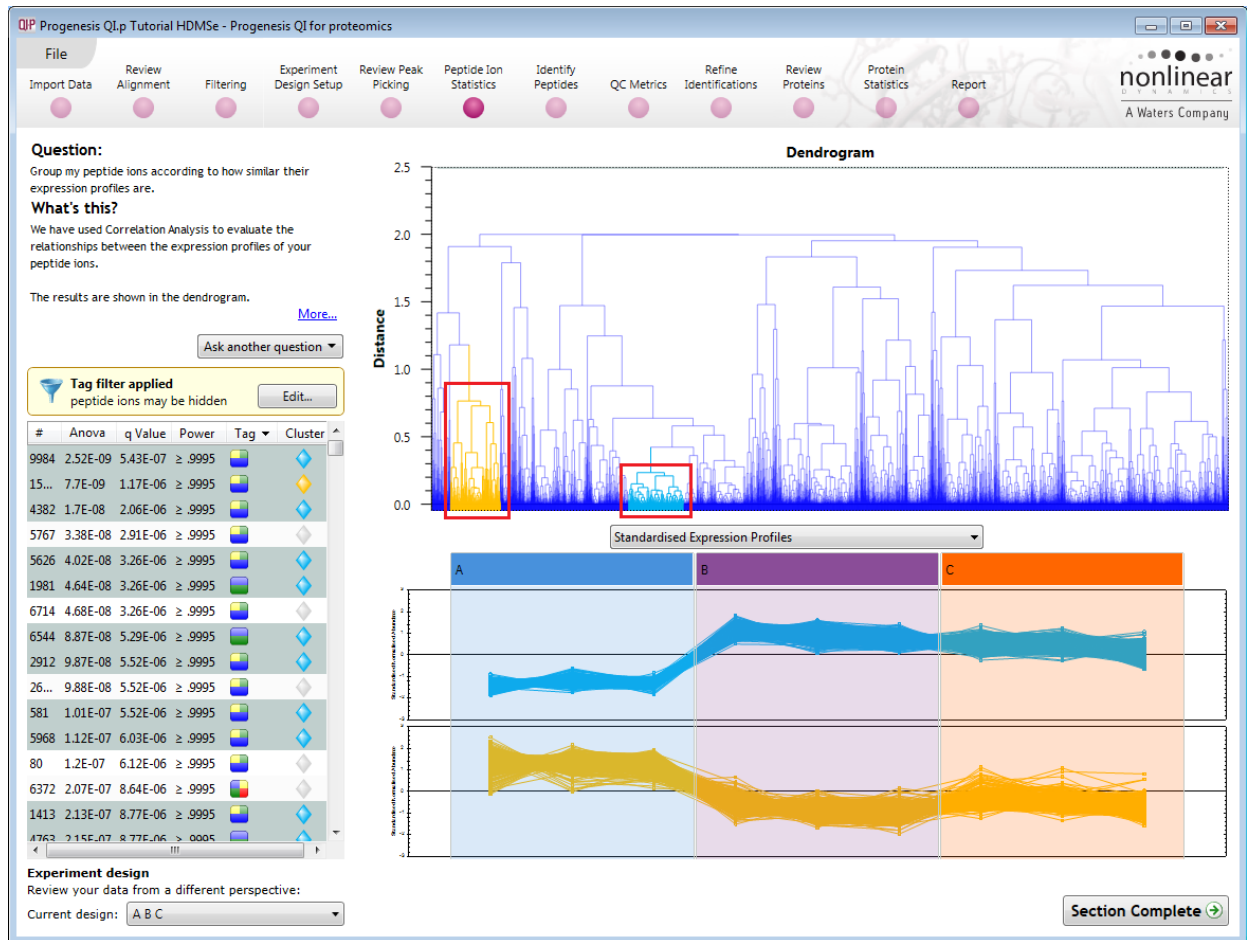
This time the statistically based question(s) being asked is:

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

The question is answered by:

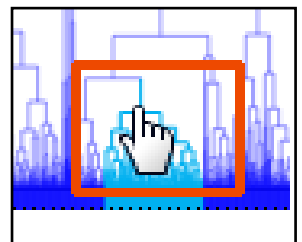
'Using Correlation analysis to evaluate the relationships between the (selected) peptide ions' expression profiles'

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.

For example: To highlight groups of peptide ions demonstrating **different expression profiles** click on a 'node' for a branch of the Dendrogram (as shown above) while holding the mouse button, hold down the **Ctrl** key and then click on another node as shown.



If you have selected 2 nodes then there will be two expression profile graphs

Note: by highlighting a group of peptide ions with similar expression the peptide ions are identified with the same colour of cluster flag in the table. This allows the table to be sorted on cluster and tagged accordingly

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.

Before moving to the Identify Peptides stage in the Workflow, first return to the PCA display and clear all tag filters **Clear all Tag filters**.

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

Stage 9: Identify peptides

Progenesis QI for proteomics is designed to perform peptide identifications either directly or 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.

QIP Progenesis QI.p Tutorial HDMSe - Progenesis QI for proteomics

File Import Data Review Alignment Filtering Experiment Design Setup Review Peak Picking Peptide Ion Statistics Identify Peptides QC Metrics Refine Identifications Review Proteins Protein Statistics Report

nonlinear A Waters Company

Identify Peptides

Select your peptide identification method:

Ion Accounting

Mascot Version: 1.0
PLGS (*.xml) Version: 1.0
SEQUEST (dta & out files) Version: 1.0
SEQUEST (dta & pepxml files) Version: 1.0
Phenyx Version: 4.2.7145.32390
Spectral Library Search Version: 4.2.7145.32390
Ion Accounting Version: 4.2.7145.32390
Max protein mass: 250 kDa

Modifications: Carbamidomethyl C Oxidation M
Add/remove modifications

Search tolerance parameters
Ion matching requirements

2 Search for identifications
Identifications will be assigned to the relevant features automatically.
9 / 9 runs ready for searching

Change Apex3D parameters
Search for identifications
Import existing results...
Clear identifications

Admin tools
Change the digest reagents and modifications that are available for peptide searches.
Modification editor... Reagent editor...

Features

#	Identifications	m/z	Charge	Retention time	Drift time	Tag
1	0	828.4192	2	58.67	4.90	
3	0	822.4427	2	69.31	4.76	
4	0	962.7957	3	76.54	4.97	
5	0	894.4745	3	67.50	4.48	
6	0	964.1703	3	63.92	5.45	
7	0	823.4391	2	75.96	5.04	
8	0	763.7624	3	69.69	4.07	
9	0	913.5255	2	71.77	6.28	

Identifications for peptide ion 1

Peak mass Peptide mass Protein mass Mass Error (Da) Mass Error (ppm) Score Seq. start Seq. end Seq

Fragment matches for:

Intensity (counts) vs m/z

Section Complete

For this example we are using the direct method **Ion Accounting**.

Note: Following the full automatic processing, described in Stage 2 of this guide, the Identify Peptides page currently displays the full list of the detected peptide ions in your experiment and some of their attributes, including the number of **Identifications** (as shown on page 51). If search results exist these can be cleared by clicking **Clear all identifications**, this will allow you to re-perform the search.

Entering Search Parameters

Firstly you need to select the FASTA file containing peptide and protein identifications.

SWISSPROT-1 is provided with the installation of the software.

To add new Databanks in the form of FASTA files click on **Edit...** to open the Databank editor

Note: the SWISSPROT-1.0 is locked

Identify Peptides

Select your peptide identification method:

Ion Accounting

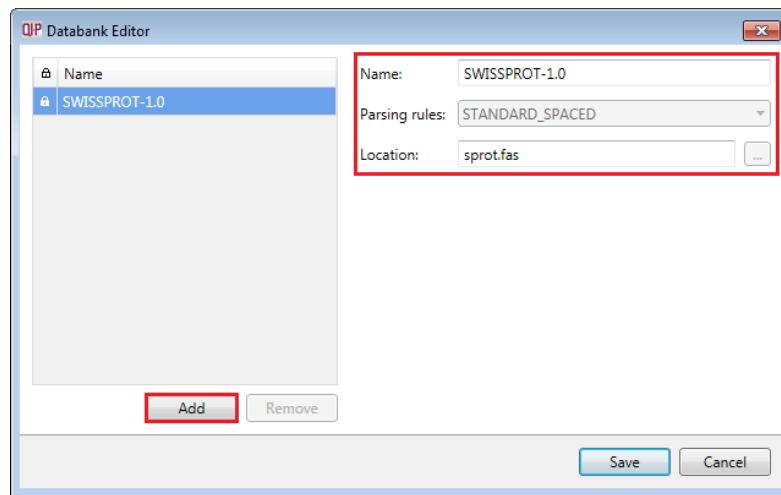
Help

For MS^s, HDMS^s and SONAR data

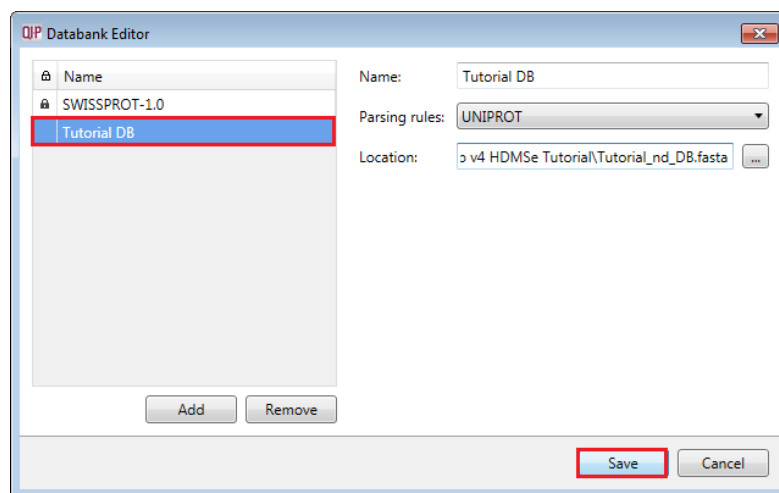
1 Enter the search parameters
Select your FASTA file containing peptide and protein identifications:

SWISSPROT-1.0 Edit...
SWISSPROT-1.0 STANDARD_SPACED
e:

Common search parameters
Search tolerance parameters
Ion matching requirements



For a new Databank you need to give it name, select the parsing rules and specify the location of the FASTA file, see the example below.



The new Data bank will appear in the left panel now click **Save** to return to the Search parameters.

If your databank is not already displayed then select it from the drop down list.

Expand the **Common search parameters**

The default settings are displayed:

Digest reagent: is set as Trypsin. Alternative Digest reagents are available from the list and additional ones can be added to the list using the **Reagent editor...**

Missed cleavages: is set as 1.

Maximum protein mass: is set at 250kDa

Modifications: are set Carbamidomethyl C (Fixed) and Oxidation M (variable). More modifications are available from the list and additional ones can be added to the list using the **Modification editor...**

Identify Peptides

Select your peptide identification method:

Ion Accounting

[Help](#)

For MS¹, HDMS² and SONAR data

1 Enter the search parameters
Select your FASTA file containing peptide and protein identifications:

Tutorial DB Edit...

Enter the search parameters to use:

Common search parameters

Search tolerance parameters

Ion matching requirements

Common search parameters

Digest reagent: Trypsin

Missed cleavages: 1 max

Max protein mass: 250 kDa

Modifications: Carbamidomethyl C Oxidation M

[Add/remove modifications](#)

Having selected the Databank then set the Search tolerance parameters and ion matching parameters as shown to the right.

When all the runs are ready for searching, click **Search for identifications**.

Depending on the search parameters and the number of runs the **Ion Accounting** search can take some time.

Once the Ion Accounting is complete, peptide ions with identifications are identified with a solid grey symbol and the number of identifications appears in the next column.

Search tolerance parameters

Peptide tolerance: ppm

Fragment tolerance: ppm

FDR less than: %

Ion matching requirements

Fragments/peptide: or more

Fragments/protein: or more

Peptides/protein: or more

2 Search for identifications
 Identifications will be assigned to the relevant features automatically.

9 / 9 runs ready for searching

The screenshot shows the Progenesis QI software interface. The 'Identify Peptides' section is active, with 'Ion Accounting' selected as the method. The 'Search for identifications' step is completed, showing '9 / 9 runs ready for searching'. A table of 'Peptide ions (10855 identified)' is displayed, with the following data:

#	Identifications	m/z	Charge	Retention time	Drift time	Tag
1	1	828.4192	2	58.67	4.90	
3	3	822.4427	2	69.31	4.76	
4	1	962.7957	3	76.54	4.97	
5	1	894.4745	3	67.50	4.48	
6	1	964.1703	3	63.92	5.45	
7	1	823.4391	2	75.96	5.04	
8	2	763.7624	3	69.69	4.07	
9	0					

An 'Import search results' dialog box is open, indicating that 54814 spectrum matches have been imported and assigned to peptide ions. The 'Identifications' table shows a peak mass of 1654.824. A 'Fragment matches for: DLSEASVYAEYALPK' plot is shown at the bottom, displaying intensity (counts) versus m/z. The plot shows a series of peaks corresponding to the peptide sequence, with the sequence DLSEASVYAEYALPK displayed above the peaks. The x-axis ranges from 0 to 1800 m/z, and the y-axis ranges from 0 to 100 intensity (counts). A 'Section Complete' indicator is visible in the bottom right corner.

Details for the current peptide ion identifications are displayed in the table below and the Fragment ions for the current identification are displayed in the bottom panel.

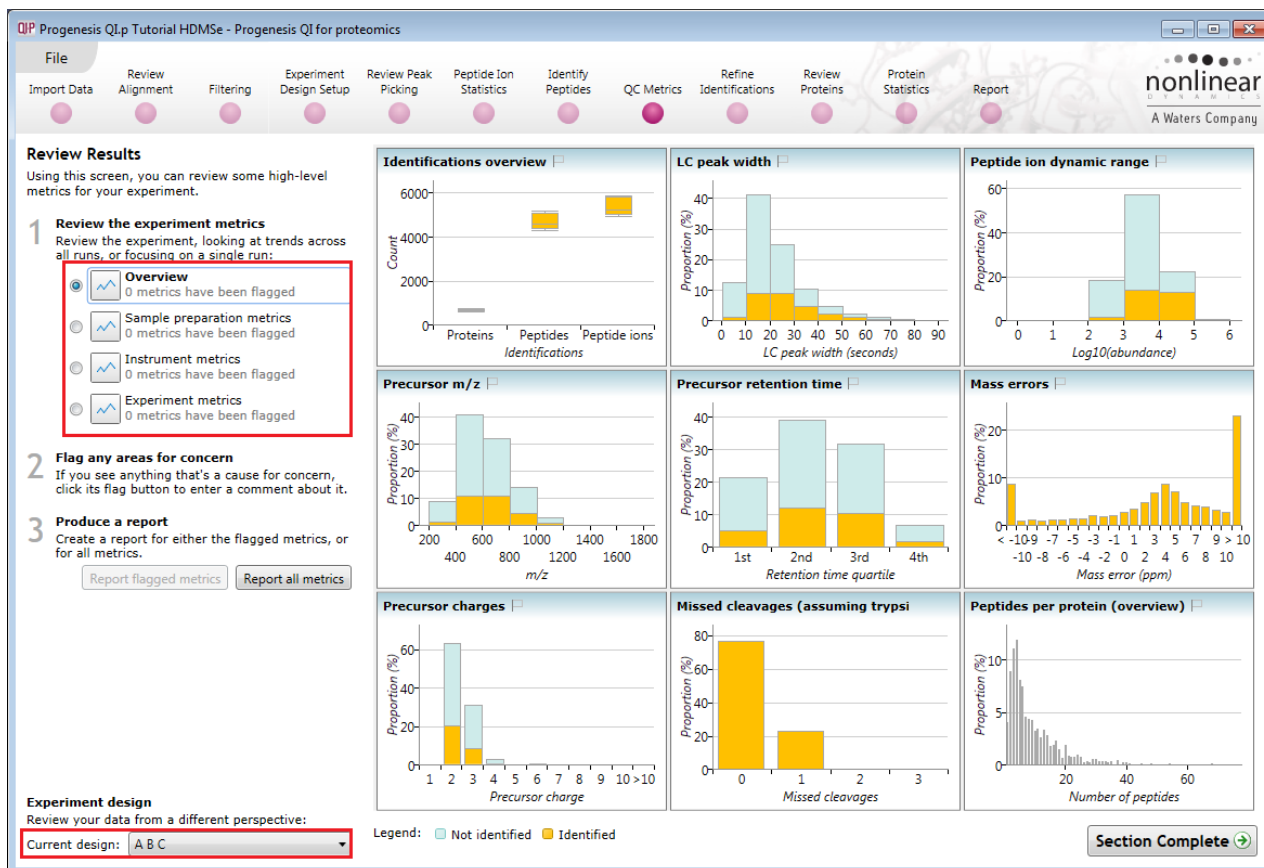
Note: if you want to perform the search with a new set of parameters then first select **Clear all identifications**

Having performed the process of peptide identification you can review the overall quality of your analysis by using the various batch-based quality metrics in the next section, by clicking on QC Metrics on the workflow.

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

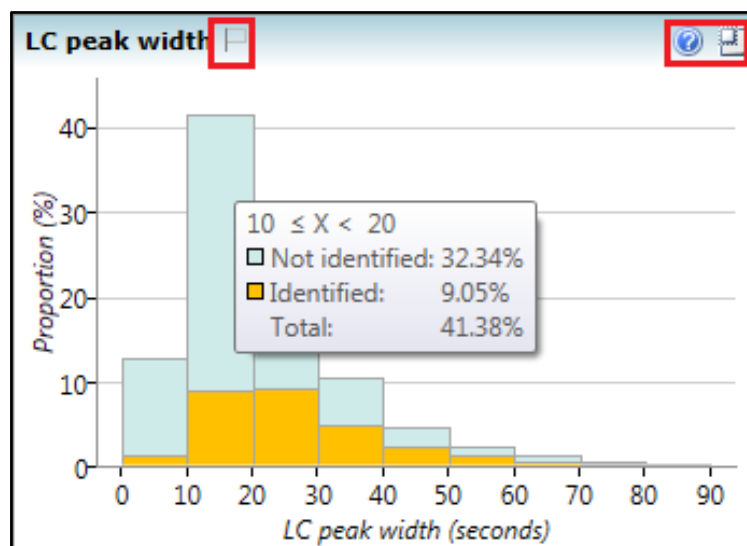
More detailed information on the QC metrics is available on the FAQ pages.

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 we are going to apply a number of filters to 'refine' the quality of the Databank search.

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 4
- Remove identifications where less than 2 hits were returned
- Remove all identifications where the Protein Description **Contains** the following: 'Putative', 'Probable', 'Like', 'Potential' and 'Predicted'

To perform these filters, on the Batch detection options panel, set the Score to less than 4, then **Delete matching search results**.

Refine Identifications

If your peptide identifications include unwanted or irrelevant results, you can remove them here.

- Specify a set of deletion criteria**
In the batch deletion criteria, enter the property values for a set of identifications you want to delete.
- Delete the unwanted identifications**
To delete the identifications you don't want, click either:
 - Delete Matching Search Results, to delete the highlighted IDs
 - Delete Non-matching Search Results, to delete the IDs that are not highlighted
- Reset the criteria to start again**
To specify another batch of identifications to delete, click Reset the Criteria and then return to step 1 above.

Batch deletion criteria

Score: less than 4

Hits: less than

Mass: less than

Absolute mass error (ppm): less than

Retention Time: less than

Sequence Length: less than

Charge: less than

Sequence: contains

Accession: contains

Description: contains

Modifications: contains

Delete matching search results Delete non-matching search results

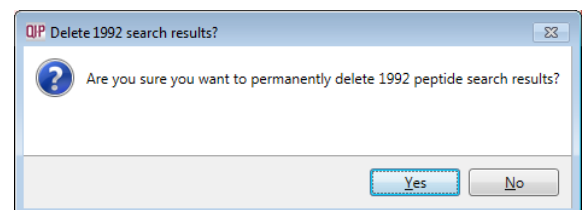
#	Score	Hits	m/z	RT(min)	Charge	Mass	Mass err	Sequence	Accession	Modifications
148	3.67	1	665.86	51.25	2	1329.7	2.25	ILKEMPSE...	C4V191	Putative uncharacterized pr
154	4.34	1	658.36	49.10	2	1314.7	-27.97	NLVSGSTV...	C4YK7	Putative uncharacterized pr
154	4.56	5	658.36	49.10	2	1314.7	-11.92	ILGPGNQF...	O74712	Histidine biosynthesis trifun
154	8.32	2	658.36	49.10	2	1314.7	8.24	EQLSQVLV...	C4YQP9	60S ribosomal protein L35 C
155	---	---	830.46	56.33	2	1658.9	15.68	VINEPTAA...	C4YK39	Heat shock protein SSC1_r
155	5.02	1	830.46	56.33	2	1658.9	8.93	SNPIFTKN...	C4YS46	Putative uncharacterized pr
155	8.48	8	830.46	56.33	2	1658.9	15.68	IINEPTAAA...	P46587	Heat shock protein SSA2 Gf
155	8.47	5	830.46	56.33	2	1658.9	15.68	IINEPTAAA...	P10591	Heat shock protein SSA1 O5
155	5.02	1	830.46	56.33	2	1658.9	8.93	SNPIFTKN...	Q5AJY6	Likely 26S proteasome regu
155	8.48	8	830.46	56.33	2	1658.9	15.68	IINEPTAAA...	P41797	Heat shock protein SSA1 Gf
158	3.69	1	543.98	44.89	3	1628.9	21.88	QALSETVD...	C4YK19	Putative uncharacterized pr
158	3.73	1	543.98	44.89	3	1628.9	8.86	TTLTEINGP...	Q5A9Y0	Putative uncharacterized pr
161	0.00	1	618.84	50.83	2	1235.6	42.52	GLSVSKDE...	C4YK5	Putative uncharacterized pr
161	5.71	1	618.84	50.83	2	1235.6	-8.39	FDNGLRFIVR	Q5A202	Putative uncharacterized pr
161	8.59	4	618.84	50.83	2	1235.6	13.06	TLASDGLA...	Q5A516	ADP_ATP carrier protein GN
165	8.11	9	697.37	56.25	2	1392.7	7.96	ALDADVVS...	P82610	5-methyltetrahydropteroylt
166	5.32	1	965.48	51.10	2	1928.9	8.09	AVSSGMV...	C4YE86	Glycyl-tRNA synthetase 1 Gl
166	8.10	8	965.48	51.10	2	1928.9	11.92	VWLDPNE...	C4YQK4	60S ribosomal protein L19 C
167	4.56	3	821.96	50.91	2	1641.9	11.73	LASESTLPV...	C4YSQ4	Putative uncharacterized pr

14747 search results. 1992 matching batch delete options.

Section Complete

Note: the search results matching the filter criteria turn pink and the total is displayed at the bottom of the table (in this example: 1992 matching out of 14747)

Note: a dialog warns you of what you are about to delete
Click **Yes**.



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

Tip: always click **Reset the Criteria** after each deletion.

Refine Identifications

If your peptide identifications include unwanted or irrelevant results, you can remove them here.

- Specify a set of deletion criteria**
In the batch deletion criteria, enter the property values for a set of identifications you want to delete.
- Delete the unwanted identifications**
To delete the identifications you don't want, click either:
 - Delete Matching Search Results, to delete the highlighted IDs
 - Delete Non-matching Search Results, to delete the IDs that are not highlighted
- Reset the criteria to start again**
To specify another batch of identifications to delete, click Reset the Criteria and then return to step 1 above.

Batch deletion criteria:

- Score: less than
- Hits: less than 2
- Mass: less than
- Absolute mass error (ppm): less than
- m/z: less than
- Retention Time: less than
- Sequence Length: less than
- Charge: less than
- Sequence: contains
- Accession: contains
- Description: contains
- Modifications: contains

Buttons: Delete matching search results Delete non-matching search results

Confirmation dialog: **QIP Delete 4968 search results?**
Are you sure you want to permanently delete 4968 peptide search results?
Buttons:

Main interface status: 12755 search results. 4968 matching batch delete options.

Now in the Description first enter 'Like' and delete matching search results. Then enter the 'regular expression': **regex: Puta|Prob|Pote|Pred** and delete matching search results.

Refine Identifications

If your peptide identifications include unwanted or irrelevant results, you can remove them here.

- Specify a set of deletion criteria**
In the batch deletion criteria, enter the property values for a set of identifications you want to delete.
- Delete the unwanted identifications**
To delete the identifications you don't want, click either:
 - Delete Matching Search Results, to delete the highlighted IDs
 - Delete Non-matching Search Results, to delete the IDs that are not highlighted
- Reset the criteria to start again**
To specify another batch of identifications to delete, click Reset the Criteria and then return to step 1 above.

Batch deletion criteria:

- Score: less than
- Hits: less than
- Mass: less than
- Absolute mass error (ppm): less than
- m/z: less than
- Retention Time: less than
- Sequence Length: less than
- Charge: less than
- Sequence: contains
- Accession: contains
- Description: contains regex: Puta|Prob|Pote|Pred
- Modifications: contains

Buttons: Delete matching search results Delete non-matching search results

Confirmation dialog: **QIP Delete 2158 search results?**
Are you sure you want to permanently delete 2158 peptide search results?
Buttons:

Main interface status: 7454 search results. 2158 matching batch delete options.

Having applied all the filters there will be **5296** search results remaining

To validate the Peptide search results at the protein level select **Resolve Conflicts (bottom left)**.

Resolving Conflicts

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

The **Resolve Conflicts** stage is now accessed the bottom left of the Refine Identifications stage.

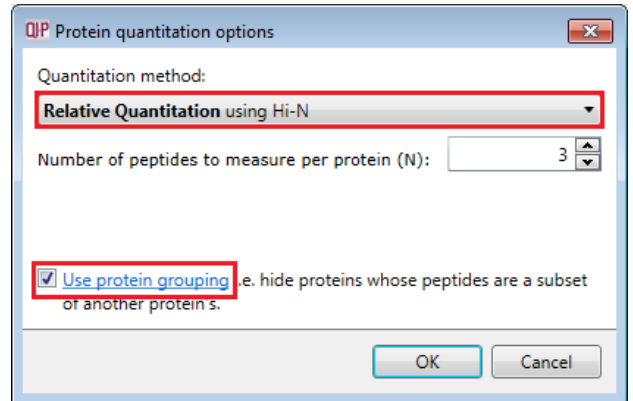
The number of conflicts you have to resolve will depend on the scope and stringency of the filters you apply at the **Refine Identifications** stage.

Details on performing conflict resolution are available in Appendix 7 (page 92)

For this guide: we will NOT resolve the conflicts.

The following pages describe the handling of Protein grouping and Protein Quantitation. Options for these are available at both Resolve Conflicts and Review Proteins

Note: the default settings are to **Use Protein Grouping** and **Relative Quantitation using Hi-N**



Protein Grouping

Where proteins are identified containing the same peptides then they are effectively indistinguishable aside from 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 **Glutathione reductase (C4YR40)** as an example, we can look at the other group members by changing the protein options.

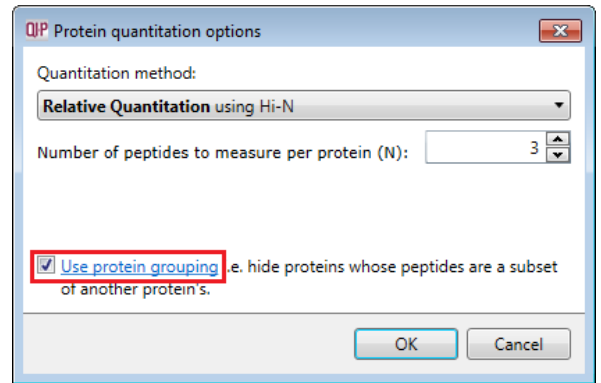
Accession	Peptides	Unique	Conflict	Score	Tag
C4YR40 (+2)	12	12	0	77.4	
C42766	12	12	0	77.4	
C4YDN4	12	11	1	83.5	

#	Σ	⊞	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict
1006	3		6.41	7	0.292	1100.643	-1.19	42.3	2		6.57E-04	0
1286	2		5.73	2	0.338	1092.59	8.18	42.2	2		5.74E-04	0
6170	0		6.72	9	0.986	1195.694	0.952	50.1	2		1.76E-04	0
6651	1		6.51	7	0.835	1393.662	9.94	41.5	2		1.15E-04	0

Although Conflict resolution has not been performed this protein has no conflicting peptides.

Protein grouping is switched on by default. To switch it off click on **Protein options...** (bottom left) and untick the option.

As grouping is switched off the grouped proteins appear with conflicts to the other group members.

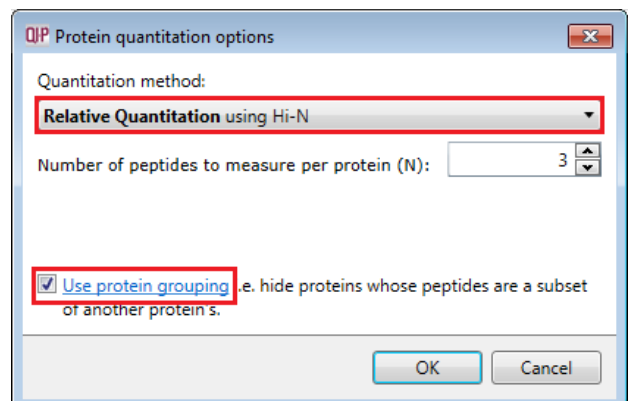


Accession	Peptides	Unique	Conflict	Score	Tag	AI
C4YE07	12	11	3	62		3.2
C4YR40	12	0	21	77.4		---
Q5A2T2	12	12	0	95.3		8.0
Q5A017	12	11	2	132		1.1

Accession	Peptides	Unique	Conflict	Protein Score	Pept
C4YR40	12	0	21	77.4	6.41
Q5W322	12	0	21	77.4	6.41
Q9UV69	9	0	18	58.1	6.41

Note: when the grouping is switched off the other group members appear in panel E showing conflicts with the other group members and **no** Unique peptides

Before moving to **Review Proteins** it is worth understanding the available options to handling Protein Quantitation in Progenesis QI for proteomics.



The next section in this guide describes the use of alternatives to the default method shown to the right.

Note: If you are using the default method, you can go directly to the **Review Protein** stage page 61 using the workflow.

Protein Quantitation options

There are 5 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.

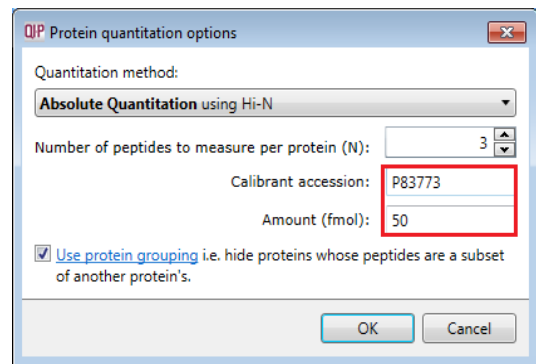
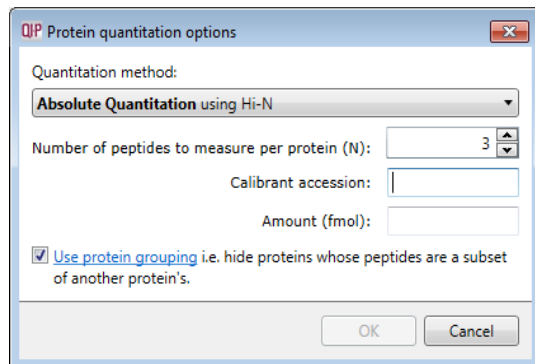
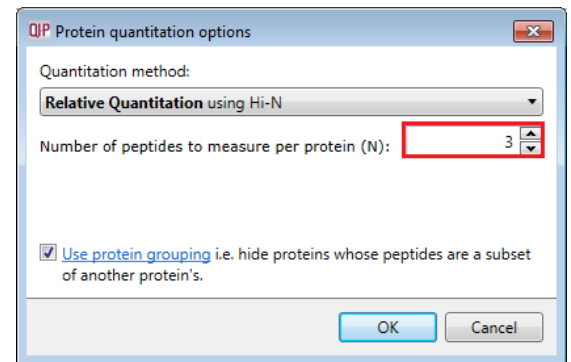
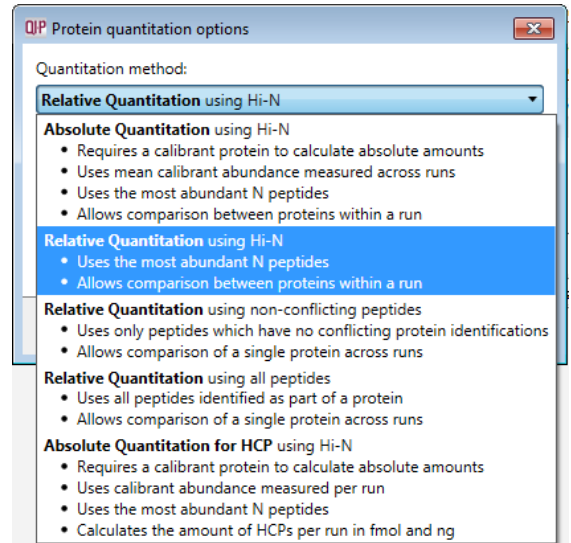
Hi-N in Progenesis QI for proteomics is an implementation of Hi-3 as described by Silva *et al.* [References]. After peptide and protein identification, the abundance of each peptide is calculated from all its constituent peptide ions.

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.

Review Proteins
Using this screen, you can find the proteins of interest in your experiment.

No filter applied Create... Search

Description	Amount (fmol) - A	Amount (fmol) - B	Amount (fmol) - C
Glyceraldehyde 3-phosphate dehydrogenase GN=TDH3 PE=3 SV=1	477	800	852
Glycerol-3-phosphate dehydrogenase GN=GPD2 PE=3 SV=1	79	140	121
Glycerol-3-phosphate dehydrogenase_mitochondrial GN=GUT2 PE=3 SV=1	30.3	33.5	33.1
Glycine cleavage system H protein GN=CAWG_00084 PE=4 SV=1	27.1	37.4	31.8

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.

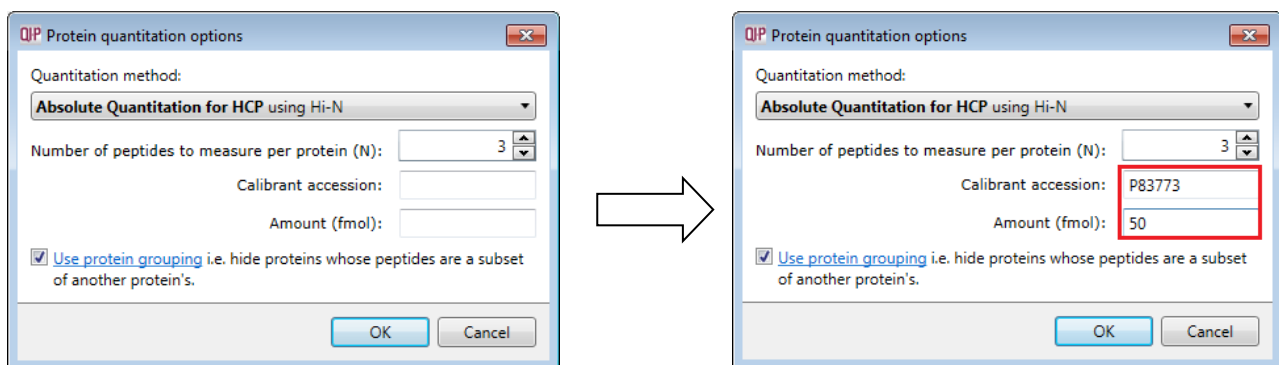
To generate values for **Absolute Quantitation for HCP** (Host Cell Proteins) select the 5th option on the drop down at the beginning of this section.

Note: there are 2 differences between this HCP-specific option and the Absolute Quantitation option:

- The mass of protein present is also reported in ng in addition to the amount in fmol
- The calculation of protein amounts and masses present is based on measurements of the calibrant protein present in the same run, and not a pooled measurement over all the runs. It does still assume a specified amount of a calibrant in every run. However, it does not pool information across the runs to derive its relationship between the calibrant and observed abundance. The amount of a contaminant is instead calculated directly by relation to the observed amount of the calibrant in the same run.

Description	Amount (fmol) - A	Amount (fmol) - B	Amount (fmol) - C	Amount (ng) - A	Amount (ng) - B	Amount (ng) - C
Glyceraldehyde 3-phosphate dehydrogenase GN=TDH3 PE=3 SV=1	471	724	971	16.9	26	34.9
Glycerol-3-phosphate dehydrogenase GN=GPD2 PE=3 SV=1	78.1	127	138	3.22	5.22	5.68
Glycerol-3-phosphate dehydrogenase_mitochondrial GN=GUT2 PE=3 SV=1	29.9	30.3	37.7	2.18	2.2	2.75
Glycine cleavage system H protein GN=CAWG_00084 PE=4 SV=1	26.8	33.8	36.1	0.548	0.692	0.739

As for **Absolute Quantitation using Hi-N**, 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.

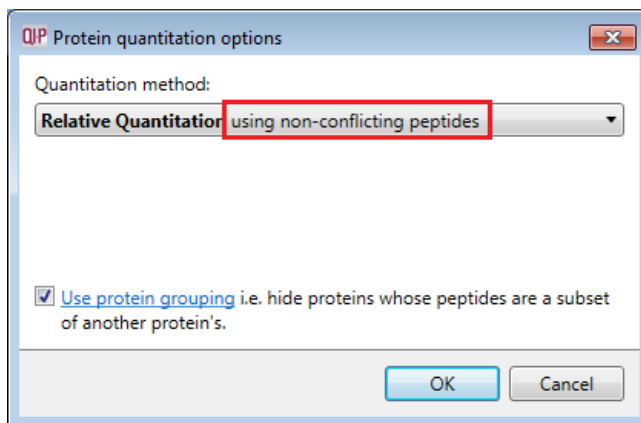
How the values for the calibrant behave across runs comparing Absolute Quant with Absolute Quant for HCP using the same calibrant (Acetyl-CoA hydrolase)

Description	Amount (fmol) - A	Amount (fmol) - B	Amount (fmol) - C
Acetolactate synthase small subunit_mitochondrial GN=CAWG_03647 PE=4 SV=1	23.1	26	30.7
Acetyl-CoA hydrolase GN=ACH1 PE=1 SV=2	50.6	55.3	44
Acetyl-coenzyme A synthetase 1 GN=ACS1 PE=3 SV=1	6.66	8.89	7.95

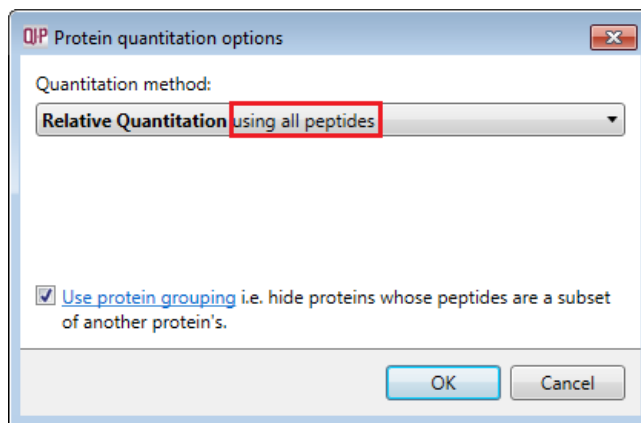
Description	Amount (fmol) - A	Amount (fmol) - B	Amount (fmol) - C	Amount (ng) - A	Amount (ng) - B	Amount (ng) - C
Acetolactate synthase small subunit_mitochondrial GN=CAWG_03647 PE=4 SV=1	22.9	23.5	35	0.764	0.785	1.17
Acetyl-CoA hydrolase GN=ACH1 PE=1 SV=2	50	50	50	2.91	2.91	2.91
Acetyl-coenzyme A synthetase 1 GN=ACS1 PE=3 SV=1	6.59	8.03	9.03	0.496	0.605	0.68

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.

Make sure that the Protein quantitation method is set to **Relative Quantitation** using Hi-N (where N=3) and Use protein grouping is ticked.

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

Stage 12: Review Proteins

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

Review Proteins

Using this screen, you can find the proteins of interest in your experiment.

- Set the quantitation options**
If you've not already done so, choose between relative and absolute quantitation, use of Hi-N, protein grouping and more.
- Create a shortlist to review**
In the table, sort and filter the proteins based on their measurements, to generate a shortlist for further review.
- Review the proteins**
For each protein of interest, review its peptide measurements and correlations:
- Export data for further processing**
By exporting your data to external tools, there's no limit to your analysis.

Accession	Peptides	Unique peptides	Confidence score	Anova (p)	q Value	Tag	Max fold change	Highest Mean	Lowest Mean	Description
Q0P457	4	4	31	1.36E-06	8.73E-06		1.95	B	A	Cu-binding metallothi...
P87219	3	3	22.9	2.13E-06	1.04E-05		4.1	C	A	Sorbose reductase SO...
Q59W63	2	1	11.1	2.65E-06	1.04E-05		1.35	A	C	Ribosomal RNA assem...
P43066	3	3	24.8	2.84E-06	1.04E-05		1.78	B	A	D-arabinitol 2-dehydr...
Q9P8Q7	7	7	62.2	4.14E-06	1.26E-05		2.01	B	A	Isocitrate lyase GN=IC...
Q7Z8E8	3	3	25.4	1.26E-05	2.14E-05		3.28	C	A	Cell surface hydropho...
C4YNC2 (+1)	9	9	25.4	1.26E-05	2.14E-05		1.51	B	A	Phosphoenolpyruvate...
C4YS49	3	3	25.4	1.26E-05	2.14E-05		1.51	B	A	Dihydroxy-acid dehyd...
Q5AE88	4	3	25.4	1.26E-05	2.14E-05		1.51	B	A	Proteasome compone...
C4YFV8	2	2	25.4	1.26E-05	2.14E-05		1.51	B	A	Glycine cleavage syste...
C4YNI1	4	2	25.4	1.26E-05	2.14E-05		1.51	B	A	1,3-beta-glucan synth...
C4YHP3	5	5	25.4	1.26E-05	2.14E-05		1.51	B	A	Carnitine O-acetyltran...
P42800	4	4	26	1.49E-05	2.14E-05		2.7	B	A	Inositol-3-phosphate ...
Q59LQ6	1	1	8.06	1.59E-05	2.14E-05		1.68	A	B	Protein translation fac...
O94039	25	25	260	1.71E-05	2.14E-05		1.77	B	A	Transketolase 1 GN=T...

Selected protein: Isocitrate lyase GN=ICL1 PE=3 SV=1

View peptide measurements

Quantifiable proteins displayed: 516

As an example, let us explore Isocitrate Lyase.

First filter the table to show only proteins showing a significant 2 fold or greater change.

Review Proteins

Using this screen, you can find the proteins of interest in your experiment.

- Set the quantitation options**
If you've not already done so, choose between relative and absolute quantitation, use of Hi-N, protein grouping and more.
- Create a shortlist to review**
In the table, sort and filter the proteins based on their measurements, to generate a shortlist for further review.
- Review the proteins**
For each protein of interest, review its peptide measurements and correlations:
- Export data for further processing**
By exporting your data to external tools, there's no limit to your analysis.

Tag filter applied
proteins may be hidden

Accession	Peptides	Unique peptides	Confidence score	Anova (p)	q Value	Tag	Max fold change	Highest Mean	Lowest Mean	Description
Q8UVL1	3	3	26.5	5.87E-07	6.56E-06		2.44	A	C	Non-histone chrom...
C4VL44	2	2	12.8	7.7E-07	6.56E-06		2.19	B	A	Extracellular matrix
P87219	3	3	22.9	2.13E-06	1.04E-05		4.1	C	A	Sorbose reductase S...
Q9P8Q7	7	7	62.2	4.14E-06	1.26E-05		2.01	B	A	Isocitrate lyase GN=IC...
Q7Z8E8	3	3	25.4	4.42E-06	1.26E-05		3.28	C	A	Cell surface hydroph...
P42800	4	4	26	1.49E-05	2.14E-05		2.7	B	A	Inositol-3-phosphat...
C4YF25	1	1	6.57	0.000162	7.27E-05		2.26	A	B	Bud site selection p...
C4YRH4	6	6	35.4	0.0011	0.000218		2.11	B	A	NAD(P)H-depender...
C4YE92	2	2	13.4	0.00125	0.000234		2.65	C	A	6,7-dimethyl-8-nibi...
Q5ALX8	5	5	51.3	0.00205	0.000314		2.05	C	A	Adenine phosphoni...
Q59UB3	1	1	5.68	0.00278	0.000395		4.06	C	A	Ribonucleoside-dip...
Q9HFQ6	1	1	14.6	0.00406	0.000492		2.22	C	A	60S acidic ribosom...
Q59NN8	1	1	6.72	0.0122	0.00108		3.46	A	B	Hsp70 nucleotide e...
Q59WG0	1	1	7	0.0146	0.00124		3.24	C	A	Hit family protein 1...

Selected protein: Isocitrate lyase GN=ICL1 PE=3 SV=1

View peptide measurements

Quantifiable proteins displayed: 14

The table indicates that this protein is most highly expressed in Condition B, 2.01 fold over the lowest condition (A).

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** beside table and ensure that Show **Peptides** is selected.

Review selected protein
 Accession: Q9P8Q7
 Description: Isocitrate lyase GN=ICL1 PE=3 SV=1

Σ	Identifier	Ions	Score	Correlation	Anova (p)	Max Fold Change	Highest Mean	Lowest Mean	Tag	Abundance	Neutral Mass	Retention Time
●	53.57_1729.8822n	1	7.423	0.957	0.000368	1.52	B	A	■	2.817E+04	1729.8822	53.566
●	64.49_1503.8215n	1	7.388	0.934	2.92E-05	2.72	B	A	■	1.807E+04	1503.8215	64.486
●	63.34_1322.6874n	1	7.247	0.971	2.61E-06	2.92	B	A	■	1.379E+04	1322.6874	63.343
○	53.57_1729.8549n	1	7.423	0.991	0.000386	7.4	B	A	■	3626	1729.8549	53.566
○	38.07_1302.7206n	1	6.440	0.989	5.47E-06	2.08	B	A	■	5533	1302.7206	38.068
○	38.09_1302.7213n	1	6.440	0.878	0.0814	2.08	B	A	■	634.5	1302.7213	38.086
○	38.29_1302.6459n	1	6.440	0.943	0.00224	1.32	B	A	■	4070	1302.6459	38.291
○	47.52_1312.7092n	1	6.960	0.970	1.06E-05	2.9	B	A	■	7300	1312.7092	47.517
○	34.10_1489.7484n	1	7.208	0.991	3.44E-05	3.21	B	A	■	8931	1489.7484	34.097
○	46.27_1696.8563n	1	5.650	0.016	0.0126	1.26	A	C	■	3183	1696.8563	46.27

Standardised Expression Profiles

Peptide profiles: ● Contributes to protein measurements ■ Does not contribute to protein measurements

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

Note: a correlation score (between -1 and 1) for each peptide is available to facilitate the validation of peptide expression.

You can control what peptides are used in Protein Quantitation by using the **Protein options...** at the protein level of **Review Proteins**

Review Proteins
 Using this screen, you can find the proteins of interest in your experiment.

1 **Set the quantitation options**
 If you've not already done so, choose between relative and absolute quantitation, use of Hi-N, protein grouping and more.

2 **Create a shortlist to review**
 In the table, sort and filter the proteins based on their measurements, to generate a shortlist for further review.

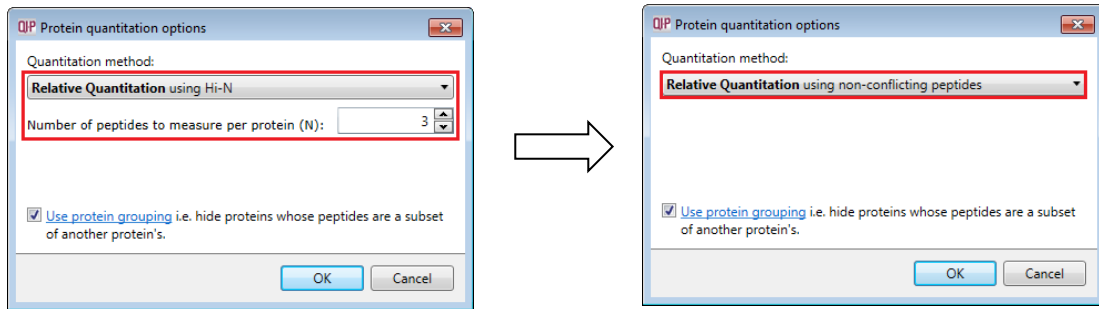
Tag filter applied: proteins may be hidden

Accession	Peptides	Unique peptides
C4YL44	2	2
P87219	3	3
Q9P8Q7	7	7
Q7Z8E8	3	3
P42800	4	4
C4YF25	1	1
C4YRH4	6	6

Protein options...

Note: the default quantitation method is based on the Relative Quantitation using Hi-N, in this case n=3 so the top 3 most abundant peptides are used to determine the relative abundance for each protein. Where there are multiple charge states the combined abundance of the charge states is used.

Changing to **Relative Quantitation using non conflicting peptides** will utilise all non conflicting peptides in the abundance calculation. (as shown below)



Note: doing this will update the peptide table to show open symbols against conflicting peptides which will not contribute to the quantitation of the protein.

Review selected protein
 Accession: [Q59W33](#)
 Description: Glycerol-3-phosphate dehydrogenase GN=GPD2 PE=3 SV=1

1 **Choose the level of detail**
 View the properties and expression profiles of either peptides or individual peptide ions:
 Show: Peptides Peptide ions

Σ	Identifier	Ions	Score	Correlation	Anova (p)	Max Fold Change	Highest Mean	Lowest Mean	Tag	Abundance	Neutral Mass	Retention T
●	62.70_2049.9990n	1	7.950	0.970	2.24E-06	2.44	C	A		4236	2049.999	62.697
●	56.02_2070.9708n	2	8.888	0.978	4.53E-07	4.08	B	A		4.183E+04	2070.9708	56.024
●	62.66_2049.9663n	1	7.950	0.983	6.77E-08	11.5	C	A		4518	2049.9663	62.663
●	59.63_959.5408n	1	8.230	0.984	5.76E-06	2.55	B	A		1.917E+04	959.5408	59.628
●	62.61_1423.6541n	1	8.390	0.991	4.03E-06	4.81	C	A		1.996E+04	1423.6541	62.611
○	60.94_1440.8230n	2	8.549	0.992	6.93E-08	2.9	B	A		8.757E+04	1440.823	60.937
○	None of the ions contribute to protein measurements				1.65E-06	2.57	B	A		3.169E+04	1787.9145	41.223
●	59.75_1970.9109n	1	8.734	0.993	2.51E-07	4.17	C	A		2.901E+04	1970.9109	59.748
●	54.42_1777.9120n	1	8.589	0.993	3.9E-06	2.58	B	A		3.193E+04	1777.912	54.418
●	35.12_1666.9070n	2	7.292	0.995	5.04E-07	3.12	B	A		4.057E+04	1666.907	35.12

2 **Compare expression profiles**
 Select peptides in the table to show their expression profiles in the chart below.
 Select all peptides

3 **Resolve any quantitative outliers**
 Tag any peptide ions whose expression profile is an outlier for this protein.
 Learn about tagging and filtering

Experiment design
 Review your data from a different perspective:
 Current design: A B C

Standardised Expression Profiles

In this example if you view the data at the level of the peptide ions then the two open symbols correspond to the 2 charge states of the conflicting peptide ion

Review selected protein
 Accession: [Q59W33](#)
 Description: Glycerol-3-phosphate dehydrogenase GN=GPD2 PE=3 SV=1

1 **Choose the level of detail**
 View the properties and expression profiles of either peptides or individual peptide ions:
 Show: Peptides Peptide ions

Charge	m/z	Retention Time (mins)	Mass error (ppm)	Drift time (ms)	Peptide Sequence	Modifications
E+04	596.3515	2	50.072	10.56	LLSTSIDTLK	
E+04	691.3333	3	55.989	12.24	TFTEESAGVADLITTCGGR	[16] Carbamidomethyl C
E+04	481.2813	3	60.937	2.25	VIEDVVGASIAQALK	
E+04	721.4193	2	60.937	3.56	VIEDVVGASIAQALK	

Returning to Isocitrate Lyase and the **Peptides** view of the table

Review selected protein
 Accession: [Q9P8Q7](#)
 Description: Isocitrate lyase GN=ICL1 PE=3 SV=1

1 **Choose the level of detail**
 View the properties and expression profiles of either peptides or individual peptide ions:
 Show: Peptides Peptide ions

Σ	Identifier	Ions	Score	Correlation	Anova (p)	Max Fold Change	Highest Mean	Lowest Mean	Tag	Abundance	Neutral Mass	Retention T
●	46.27_1696.8563n	1	5.650	0.016	0.0126	1.26	A	C		3183	1696.8563	46.27
●	38.09_1302.7213n	1	6.440	0.878	0.0814	2.08	B	A		634.5	1302.7213	38.086
●	64.49_1503.8215n	1	7.388	0.934	2.92E-05	2.72	B	A		1.807E+04	1503.8215	64.486
●	38.29_1302.6459n	1	6.440	0.943	0.00224	1.32	B	A		4070	1302.6459	38.291

If you order the Peptide table using the **Correlation** column then hold down the Ctrl key and then click on the first peptide (with the lowest correlation) you can remove them from the expression profiles.

As an example: the removal of a poorly correlated peptide from a protein requires you to first create a tag (at the peptide ion level for the corresponding peptide ions of the peptide you wish to remove (see below).

Review selected protein
 Review the selected protein's identified peptides and validate their expression patterns.

1 **Choose the level of detail**
 View the properties and expression profiles of either peptides or individual peptide ions:
 Show: Peptides Peptide ions

2 **Compare expression profiles**
 Select peptide ions in the table to show their expression profiles in the chart below.
 Select all peptide ions

3 **Resolve any quantitative outliers**
 Tag any peptide ions whose expression profile is an outlier for this protein.
 Learn about tagging and filtering

Experiment design
 Review your data from a different perspective:
 Current design: A B C

Accession: Q9P8Q7
 Description: Isocitrate lyase GN=ICL1 PE=3 SV=1

Σ	#	Score	Correlation	Anova (p)	Max Fold Change	Highest Mean	Lowest Mean	Tag	Abundance	m/z	Charge	Retention Time
●	24078	5.650	0.016	0.0126	1.26	A	C		3183	566.6261	3	46.27
●	40839	6.440	0.878	0.0814	2.08	B	A		634.5	652.3679	2	38.086
●	9787	7.388	0.934	0.0126	1.26	B	A		1.807E+04	752.918	2	64.486
●	20412	6.440	0.943	0.0126	1.26	B	A		652.3302	2	38.291	
●	6734	7.423	0.957	0.0126	1.26	A	A		2.817E+04	577.6347	3	53.566
●	14321	6.960	0.970	0.0126	1.26	A	A		7300	438.577	3	47.517
●	9973	7.247	0.971	0.0126	1.26	A	A		1.379E+04	662.351	2	63.343
●	12033	6.440	0.989	0.0126	1.26	A	A		5533	435.2475	3	38.068
●	26031	7.423	0.991	0.0126	1.26	A	A		3626	865.9347	2	53.566
●	16259	7.208	0.991	0.0126	1.26	A	A		8931	745.8815	2	34.097

Standardised Expression Profiles

Peptide ion profiles: Contributes to protein measurements Does not contribute to protein measurements

Once the tag is created, return to the peptide level, to identify additional, poorly correlated peptides. For each additional peptide, double click on it and then right click on the corresponding peptide ion to select and add the **Remove peptide** tag, the tag appears in the tag column.

Review selected protein
 Review the selected protein's identified peptides and validate their expression patterns.

1 **Choose the level of detail**
 View the properties and expression profiles of either peptides or individual peptide ions:
 Show: Peptides Peptide ions

2 **Compare expression profiles**
 Select peptide ions in the table to show their expression profiles in the chart below.
 Select all peptide ions

3 **Resolve any quantitative outliers**

Accession: Q9P8Q7
 Description: Isocitrate lyase GN=ICL1 PE=3 SV=1

Σ	#	Score	Correlation	Anova (p)	Max Fold Change	Highest Mean	Lowest Mean	Tag	Abundance	m/z	Charge	Retention Time (m)
●	24078	5.650	0.016	0.0126	1.26	A	C		3183	566.6261	3	46.27
●	40839	6.440	0.878	0.0814	2.08	B	A	<input checked="" type="checkbox"/>	634.5	652.3679	2	38.086
●	9787	7.388	0.934	0.0126	1.26	A	A		1.807E+04	752.918	2	64.486
●	20412	6.440	0.943	0.0126	1.26	A	A		4070	652.3302	2	38.291
●	6734	7.423	0.957	0.0126	1.26	A	A		2.817E+04	577.6347	3	53.566
●	14321	6.960	0.970	0.0126	1.26	A	A		7300	438.577	3	47.517
●	9973	7.247	0.971	0.0126	1.26	A	A		1.379E+04	662.351	2	63.343
●	12033	6.440	0.989	0.0126	1.26	A	A		5533	435.2475	3	38.068
●	26031	7.423	0.991	0.0126	1.26	A	A		3626	865.9347	2	53.566
●	16259	7.208	0.991	0.0126	1.26	A	A		8931	745.8815	2	34.097

Now at the Peptides level, set up a Tag Filter to display only the poorly correlated peptide ions in this example 2 (with the **Remove peptide** tag).

Review selected protein
 Review the selected protein's identified peptides and validate their expression patterns.

1 **Choose the level of detail**
 View the properties and expression profiles of either peptides or individual peptide ions:
 Show: Peptides Peptide ions

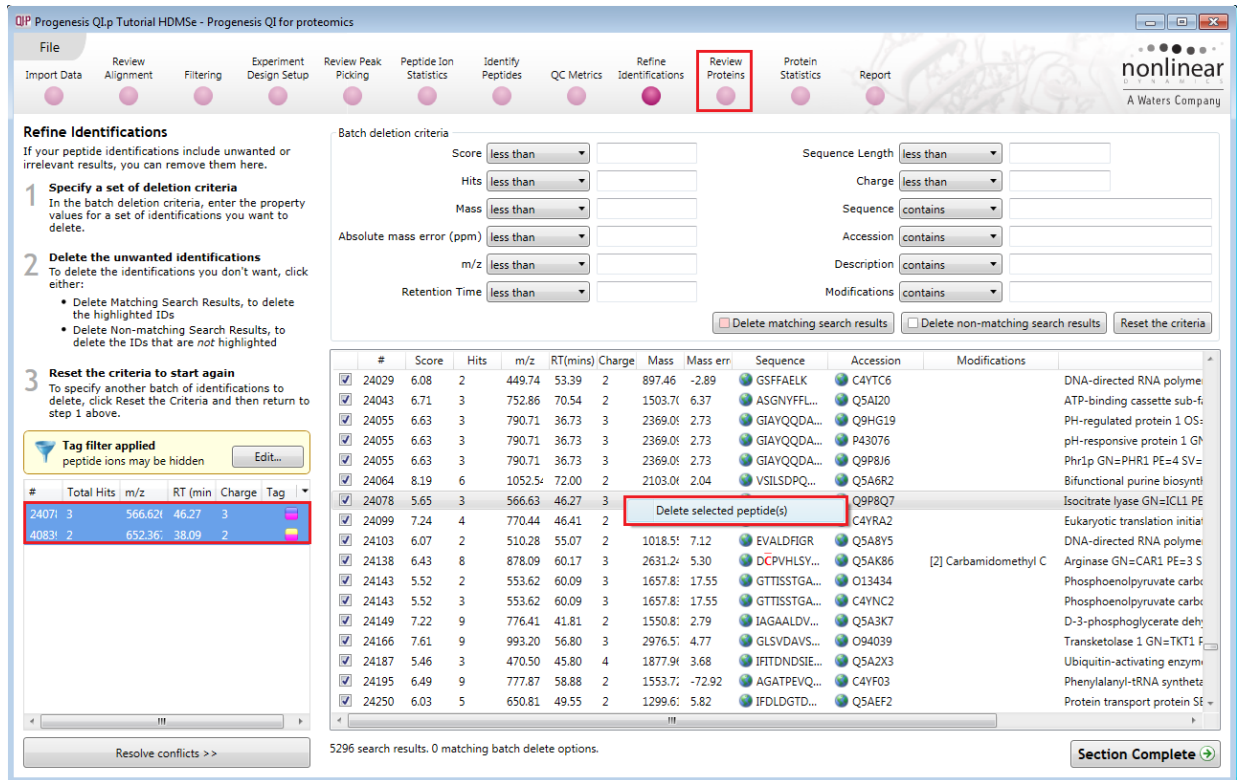
Accession: Q9P8Q7
 Description: Isocitrate lyase GN=ICL1 PE=3 SV=1

Tag filter applied
 peptides may be hidden

Σ	Identifier	Ions	Score	Correlation	Anova (p)	Max Fold Change	Highest Mean	Lowest Mean	Tag	Abundance	Neutral Mass	Retention Time
●	46.27_1696.8563n	1	5.650	0.016	0.0126	1.26	A	C	<input checked="" type="checkbox"/>	3183	1696.8563	46.27
●	38.09_1302.7213n	1	6.440	0.878	0.0814	2.08	B	A	<input checked="" type="checkbox"/>	634.5	1302.7213	38.086

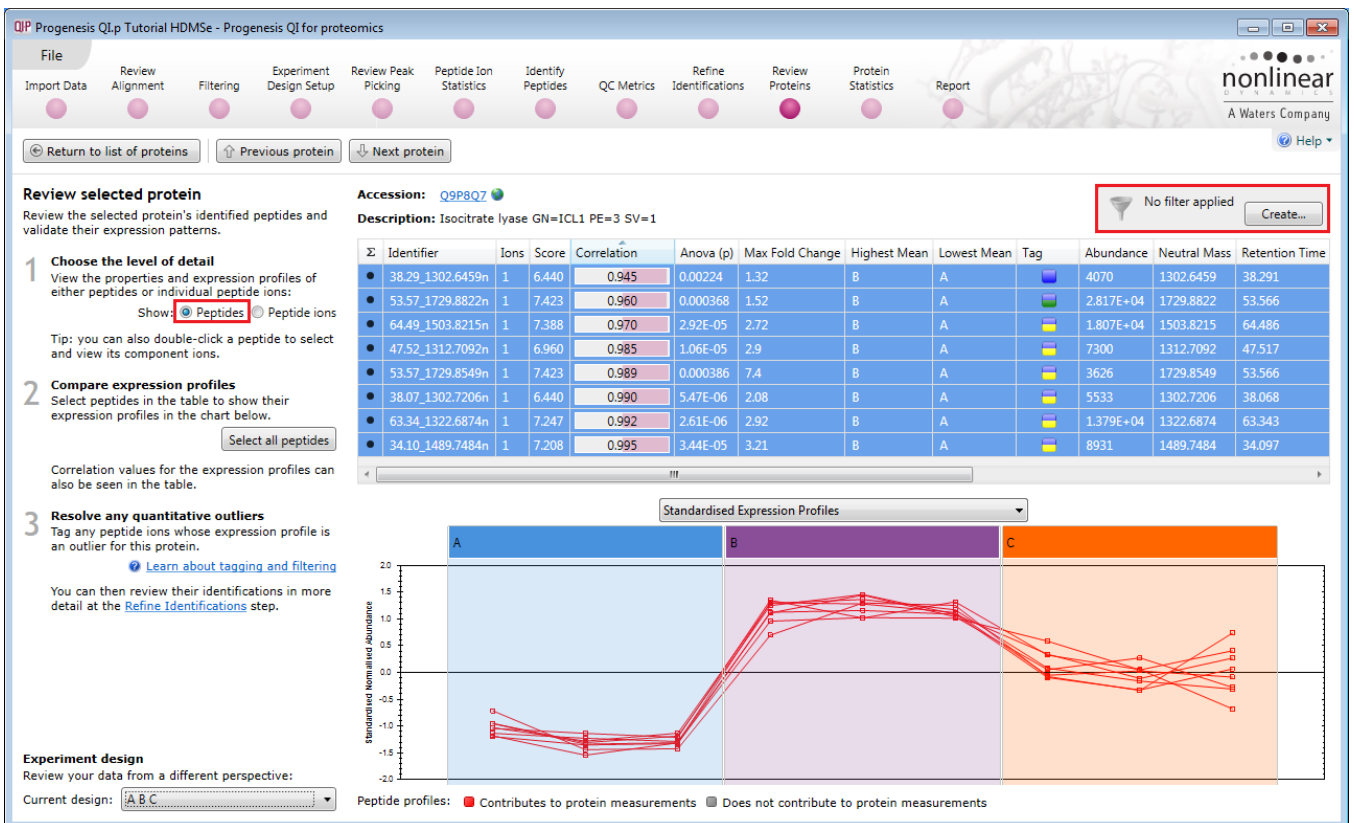
Then click on **Refine Identifications** on the workflow.

Select the peptides showing in the left hand table and then right click on one of the highlighted rows in the right hand table and click **Delete selected peptide(s)**.



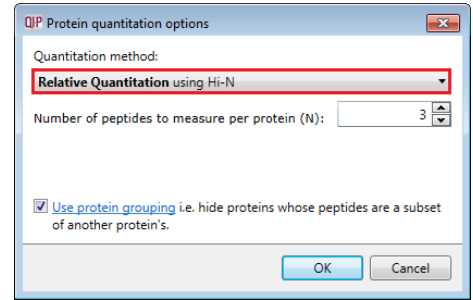
Now return to **Review Proteins**, double click on Isocitrate Lyase (Q9P8Q7) and go to the **Peptides** view. Clear the filter, to view the remaining peptides, and then click **Select all peptides** for Isocitrate Lyase.

Note: the poorly correlated peptides are no longer present.

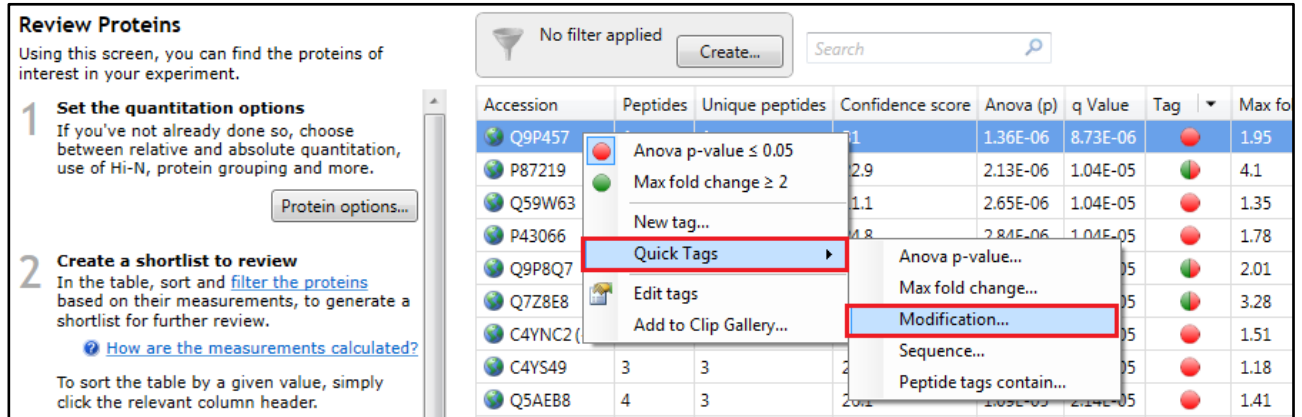


Now return to the protein level by clicking on Return to list of proteins (top left).

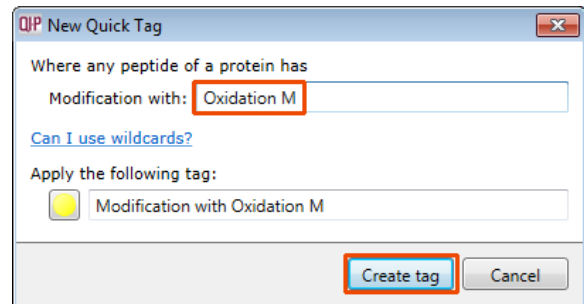
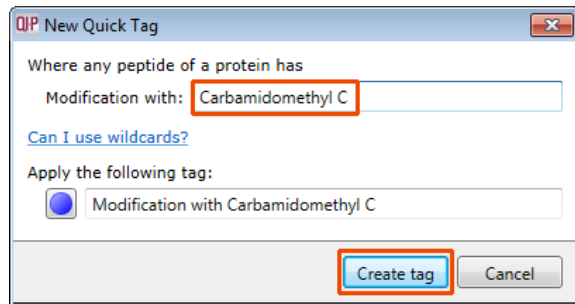
Before creating tags for proteins containing modified peptides, reset the Quantitation method to **Relative Quantitation using Hi-N** using the **Protein options...**



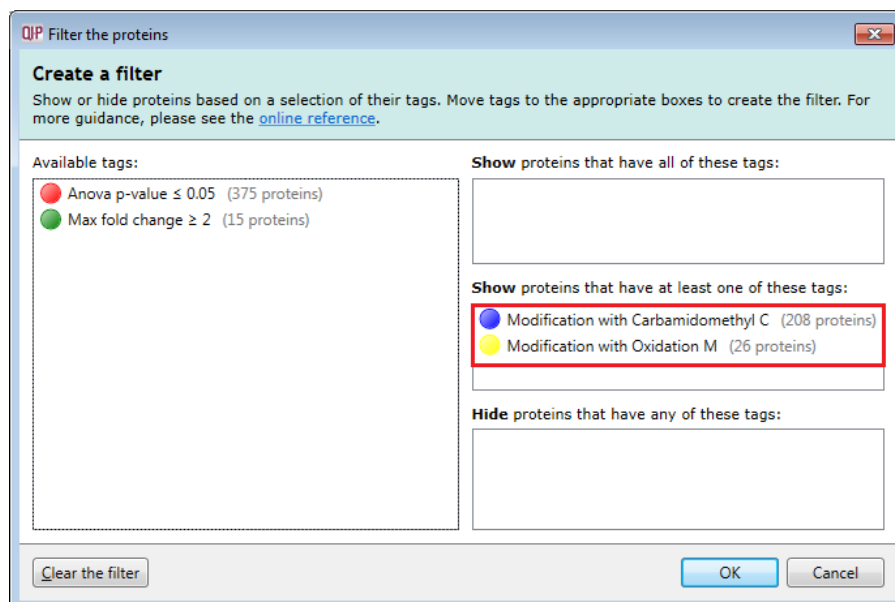
Now right click on the Protein table and select **Modification** from the **Quick tags**.



To find those proteins containing peptides with Carbamidomethylated cysteine and Oxidated methionine residues create Quick tags for each modification as shown below.



To reduce the table to displaying only these proteins with modified peptides (on cysteines and or methionines) use a tag filter to focus on these proteins by placing the appropriate tags in the **Show proteins that have at least one of these tags:**



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

Review Proteins
Using this screen, you can find the proteins of interest in your experiment.

1 Set the quantitation options
If you've not already done so, choose between relative and absolute quantitation, use of Hi-N, protein grouping and more.

2 Create a shortlist to review
In the table, sort and filter the proteins based on their measurements, to generate a shortlist for further review.

3 Review the proteins
For each protein of interest, review its peptide measurements and correlations:

4 Export data for further processing
By exporting your data to external tools, there's no limit to your analysis.

Anova (p)	q Value	Tag	Max fold change	Highest Mean	Lowest Mean	Description
0.000465	0.000134		1.26	A	C	Ribosome biogenesis protein YTM1 GN=YTM1 PE=3 SV=1
0.000481	0.000137		1.24	B	A	UTP-glucose-1-phosphate uridylyltransferase GN=UGP1 PE=4 SV=1
0.000501	0.000137		1.76	C	A	Pyruvate decarboxylase GN=PDC11 PE=3 SV=1
0.000503	0.000137					Stt1 GN=STI1 PE=4 SV=1
0.000521	0.000139					Threonine aldolase GN=GLY1 PE=3 SV=1
0.000523	0.000139					Fatty acid synthase alpha subunit reductase GN=CAWG_02796 PE=4 SV=1
0.000541	0.000143		1.42	A	C	60S ribosomal protein L10 GN=RPL10 PE=4 SV=1
0.000559	0.000146		1.4	B	A	Aldehyde dehydrogenase_mitochondrial GN=ALD5 PE=3 SV=1
0.000591	0.000153		1.19	B	A	Tyrosyl-tRNA synthetase GN=TY51 PE=4 SV=1
0.000719	0.000174		1.31	A	C	Inorganic phosphate transporter PHO84 GN=CAWG_00289 PE=4 SV=1
0.000724	0.000174		1.15	B	A	Adenylyl-sulfate kinase GN=MET14 PE=3 SV=1
0.000738	0.000174		1.45	B	A	Peroxisome oxidin TSA1 GN=TSA1 PE=2 SV=1
0.000749	0.000174		1.4	B	A	Heat shock protein SSA2 GN=SSA2 PE=1 SV=3

Selected protein: Pyruvate decarboxylase GN=PDC11 PE=3 SV=1

Quantifiable proteins displayed: 217

Note: hold the cursor over the tags for a description of the protein's current tags.

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

Stage 13: Exporting Protein Data

Data can be exported in a csv file format. You can either export the **protein, peptide or peptide ion 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 for only proteins that have Oxidised Methionine residues.

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

4 Export data for further processing
By exporting your data to external tools, there's no limit to your analysis.

Export to pathways tool

Export protein measurements

Export peptide measurements

Export peptide ion measurements

Filter the proteins

Create a filter
Show or hide proteins based on a selection of their tags. Move tags to the appropriate boxes to create the filter. For more guidance, please see the [online reference](#).

Available tags:

- Anova p-value ≤ 0.05 (377 proteins)
- Modification with Carbamidomethyl C (207 proteins)
- Max fold change ≥ 2 (15 proteins)

Show proteins that have all of these tags:

- Modification with Oxidation M (25 proteins)

Show proteins that have at least one of these tags:

Hide proteins that have any of these tags:

Clear the filter

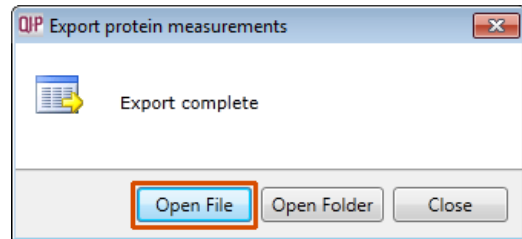
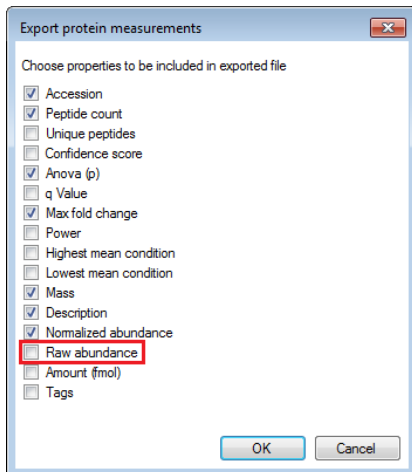
OK Cancel

File

- Save
- Close
- Export peptide ion measurements...
- Export peptide measurements...
- Export protein measurements...
- Export to pathways tool...
- Import additional protein data...
- Import protein accessions as tag...
- Export mzIdentML for PRIDE submission...
- Experiment properties
- Show Clip Gallery
- Exit

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.

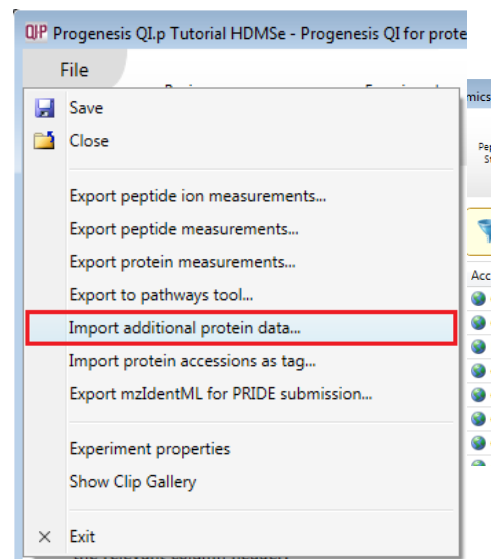


Excel will open displaying the exported protein measurements

Accession	Peptide count	Anova (p)	Max fold change	Mass	Description	Normalized abundance								
						A			B			C		
						A_01	A_02	A_03	B_01	B_02	B_03	C_01	C_02	C_03
Q59WG3	19	0.000286	1.393265562	91276.19	Cell division control protein 48 GN=CDC48 PE=3 SV=1	6.04165	6.104069	6.382303	7.810237	8.469195	8.124259	9.235389	8.411273	8.167793
Q5A1E2	19	0.000501	1.757985787	62783.45	Pyruvate decarboxylase GN=PDC11 PE=3 SV=1	30.47532	34.81616	35.18093	61.06671	57.59173	54.2401	64.75432	61.17964	50.6951
P46587	34	0.00075	1.395119219	70241.81	Heat shock protein SSA2 GN=SSA2 PE=1 SV=3	26.34676	27.96146	28.2539	36.62435	40.75456	37.80508	33.07268	33.9658	30.17681
Q9Y725	5	0.002606	1.242929018	39469.74	Mannose-1-phosphate guanyltransferase 1 OS=Candid	5.923814	6.143487	6.088348	5.100299	4.794784	4.712066	5.066129	5.133183	4.493118
Q59Z65	5	0.003284	1.239495422	31564.35	Proteasome component PRE2 GN=PRE2 PE=3 SV=1	2.14134	2.00107	1.959733	2.456694	2.641729	2.465155	2.293807	2.248603	2.465606
Q5ADM7	20	0.004259	1.785445695	35947.04	Glyceraldehyde 3-phosphate dehydrogenase GN=TDH	28.27132	37.47768	35.05756	66.21397	52.08768	50.68558	65.07962	63.1759	51.72912
Q59LW3	14	0.007171	1.929316718	187412.7	DNA-directed RNA polymerase GN=RPA190 PE=3 SV=1	4.219746	3.771511	3.794856	6.538796	4.287708	4.245464	8.246794	7.903809	6.588543
C4YTC6	10	0.00903	1.22209187	132058.3	DNA-directed RNA polymerase I 135 kDa polypeptide	9.059609	10.47197	9.842404	10.69975	11.48251	11.08544	9.237917	9.320345	8.663667
Q59N00	15	0.009063	1.182776974	27285.6	40S ribosomal protein S3 GN=RPS3 PE=4 SV=1	30.95634	32.97579	31.92331	30.09772	32.01223	30.41159	27.94132	28.06409	25.03729
C4YSV1	6	0.020115	1.197574763	15842.55	60S ribosomal protein L25 GN=CAWG_05170 PE=4 SV=1	19.09519	21.06827	20.28896	21.3714	20.04349	20.17272	17.93076	18.02852	15.46766
C4YQM1	4	0.022335	1.124601067	15995.81	40S ribosomal protein S15 GN=CAWG_02780 PE=4 SV=1	7.402242	7.74925	7.425526	7.501278	7.528884	7.452533	6.944223	6.957132	6.174225
P30575	18	0.026576	1.713187284	47231.58	Enolase 1 GN=ENO1 PE=1 SV=1	21.86692	29.96505	28.59235	51.95777	35.8221	35.52757	50.63594	49.372	37.77399
C4YHD6	7	0.031059	1.083204829	76870.01	NADPH-cytochrome P450 reductase GN=CAWG_03482	4.676369	5.02495	4.740551	5.020557	4.936881	4.881926	4.561702	4.682273	4.455524
C4YTC4	11	0.035844	1.174491751	91821.43	Glutaminyl-tRNA synthetase GN=CAWG_05411 PE=4 SV=1	9.448774	10.68459	9.762518	9.3306	8.400595	8.364039	8.782928	8.725951	7.945433
C4YRA2	9	0.041867	1.150593676	59119.25	Eukaryotic translation initiation factor 2 gamma subun	4.030048	4.073439	4.109665	5.053444	4.451311	4.54762	4.49272	4.689095	4.194535
P41797	27	0.056709	1.244772849	70495.2	Heat shock protein SSA1 GN=SSA1 PE=1 SV=2	15.05905	11.85971	11.42823	15.36704	16.33804	16.0282	13.4975	13.78449	12.17764
Q5A516	12	0.059216	1.179345683	32926.06	ADP_ATP carrier protein GN=PET9 PE=3 SV=1	30.36237	34.1933	32.87809	34.50778	35.19443	33.92876	30.93623	30.89512	26.04023
Q5ABS1	7	0.06238	1.069201884	14423.79	Ubiquinol-cytochrome c reductase complex 14 kDa prc	7.27818	7.431786	7.889224	8.234258	7.998214	7.930625	7.727017	7.876603	7.862713
Q5A900	9	0.078281	1.248516734	26987.34	40S ribosomal protein S2 GN=RPS21 PE=3 SV=1	15.07801	16.72155	15.13405	14.1835	12.10492	11.30308	15.55814	15.47995	12.76609
Q59M82	19	0.08783	1.124167619	190630.7	Clathrin heavy chain GN=CHC1 PE=4 SV=1	9.715146	10.26357	10.03633	10.34802	10.91702	9.966933	9.749938	9.565339	8.467036
C4YR46	38	0.099847	1.121176792	117623.1	Elongation factor 3 GN=CAWG_04545 PE=4 SV=1	23.97338	26.44841	25.42621	27.59592	29.83916	27.60993	26.99834	26.37865	23.40255
Q5ACQ0	11	0.137185	1.118347026	77833.41	Long-chain-fatty-acid-CoA ligase 4 GN=FAA4 PE=4 SV=1	9.58484	10.05479	9.555411	9.596871	8.595828	8.515559	8.989106	9.185083	7.931336
P16017	19	0.234967	1.194311094	50468.03	Elongation factor 1-alpha GN=TEF1 PE=1 SV=1	68.59949	67.33289	65.09681	70.54589	49.78468	48.19559	69.49089	73.01285	58.76892
Q59ZX4	14	0.314524	1.114106362	39269.36	60S ribosomal protein L4-B GN=RPL4B PE=4 SV=1	37.76197	40.11539	38.34233	39.72134	36.06694	34.58901	37.5628	36.55015	30.20356

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 metadata 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 QI 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.

MetaCore: MetaCore is an integrated software suite for functional analysis of experimental data. MetaCore is based on a curated database of human protein-protein, protein-DNA interactions, transcription factors, signaling and metabolic pathways, disease and toxicity, and the effects of bioactive molecules.

Ingenuity Pathways Analysis (IPA): Explores experimental data by identifying relationships, mechanisms, functions, and pathways of relevance.

Note: Plugins for these tools are provided as standard.

Note: Access to IPA is dependent on having a local licence for IPA. Details of exporting data to IPA are described in Appendix 9 (page 100)

As an example of exporting to a Pathways tool first use the tag filtering to 'focus' on the set to export. Then click **Export to pathways tool**, select **Panther** and then the test to be performed.

The screenshot shows the Progenesis QI software interface. The 'Review Proteins' section is active, displaying a table of protein data. A dialog box titled 'Select a pathways tool' is open, showing a list of tools: IMPALA: Integrated Molecular Pathway Level Analysis, IMPALA: Integrated Molecular Pathway Level Analysis, IPA: Ingenuity Pathway Analysis, MetaCore, and Panther Classification System. The Panther Classification System is highlighted. Below the dialog box, a chromatogram plot is visible.

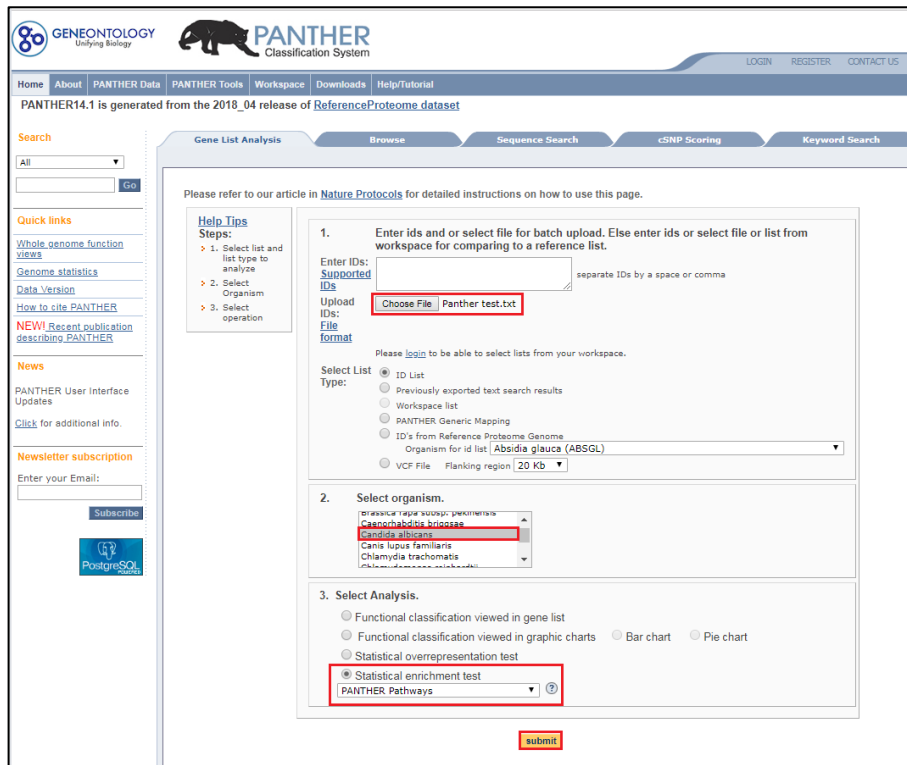
The screenshot shows a dialog box titled '4 Export data for further processing'. It contains four buttons: 'Export to pathways tool', 'Export protein measurements', 'Export peptide measurements', and 'Export peptide ion measurements'. The 'Export to pathways tool' button is highlighted with a red box.

Select either **Statistical over-representation test** or **Statistical enrichment test**.

Make sure the **Open Panther in my browser** is ticked and then click **Export proteins to file**. Save file with appropriate name.

When Panther opens locate the file, select the appropriate organism (*Candida albicans*) and tick Statistical enrichment test.

The screenshot shows a dialog box titled 'Configure your export'. It contains the following options: 'Select the type of analysis to perform:' with a dropdown menu set to 'Statistical enrichment test'; 'For statistical enrichment testing, choose two experimental conditions that you would like to compare.' with 'Baseline:' set to 'A' and 'Comparison:' set to 'B'; 'If greater, gives a negative fold change.' and 'If greater, gives a positive fold change.'; and a checked checkbox for 'Open Panther in my browser'. At the bottom, there is a button labeled 'Export proteins to file' which is highlighted with a red box.



Click **submit**.

Analysis results list is returned:

The screenshot shows the 'PANTHER PATHWAY LIST' results page. It includes a search bar, a 'Display: 30 items per page' dropdown, and a 'Refine Search' link. Below this is a table with 12 rows of results. Each row contains a checkbox, a pathway ID, a list of mapped IDs, a pathway name, and counts for components, subfamilies, and associated sequences.

clr	all	Pathway Accession	Mapped IDs	Pathway Name	Components	Subfamilies	Associated Sequence
<input type="checkbox"/>		P00005	CANAL CGD=CAL0000192659 UniProtKB=O42825	Angiogenesis	77	246	1254
<input type="checkbox"/>		P00052	CANAL EnsemblGenome=CaO19.1760 UniProtKB=Q59XU5 CANAL CGD=CAL0000188340 UniProtKB=P28870 CANAL CGD=CAL0000196167 UniProtKB=Q59VR3	TGF-beta signaling pathway	20	168	830
<input type="checkbox"/>		P02738	CANAL CGD=CAL0000188396 UniProtKB=Q5A0L0 CANAL CGD=CAL0000191707 UniProtKB=Q5AG68 CANAL CGD=CAL0000179678 UniProtKB=Q5AFP2 CANAL CGD=CAL0000191899 UniProtKB=Q5A6R2	De novo purine biosynthesis	23	81	575
<input type="checkbox"/>		P00059	CANAL CGD=CAL0000191263 UniProtKB=P43063	p53 pathway	70	107	591
<input type="checkbox"/>		P00048	CANAL EnsemblGenome=CaO19.1760 UniProtKB=Q59XU5	PI3 kinase pathway	35	78	425
<input type="checkbox"/>		P05734	CANAL CGD=CAL0000181188 UniProtKB=Q59YF0	Synaptic vesicle trafficking	12	32	164
<input type="checkbox"/>		P00024	CANAL CGD=CAL0000173961 UniProtKB=Q9P940 CANAL CGD=CAL0000197744 UniProtKB=Q5ADM7 CANAL CGD=CAL0000185645 UniProtKB=P30575 CANAL CGD=CAL0000197690 UniProtKB=P83780 CANAL CGD=CAL0000185566 UniProtKB=P82612 CANAL CGD=CAL0000184891 UniProtKB=P46273	Glycolysis	10	42	141
<input type="checkbox"/>		P02723	CANAL CGD=CAL0000192887 UniProtKB=Q5ALX8	Adenine and hypoxanthine salvage pathway	10	21	120
<input type="checkbox"/>		P00044	CANAL CGD=CAL0000181188 UniProtKB=Q59YF0	Nicotinic acetylcholine receptor signaling pathway	13	252	962
<input type="checkbox"/>		P02762	CANAL CGD=CAL0000194226 UniProtKB=Q5A017 CANAL CGD=CAL0000197690 UniProtKB=P83780	Pentose phosphate pathway	13	20	57
<input type="checkbox"/>		P02721	CANAL CGD=CAL0000187344 UniProtKB=P0C8K9	ATP synthesis	12	18	64
<input type="checkbox"/>		P00016	CANAL CGD=CAL0000192659 UniProtKB=O42825	Cytoskeletal regulation by Rho GTPase	20	174	824

Click to explore returned analysis.

Note: the success of a Pathways analysis is dependent on the organism under study being available to search using **Impala** and **Panther**.

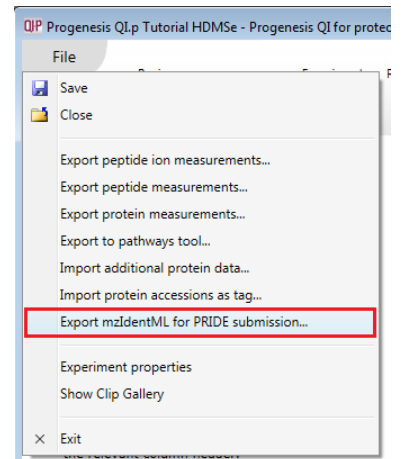
Clear any applied tag filters then move to the next section/stage.

Stage 13: Exporting identifications for submission to PRIDE

If you are submitting your raw data and results to PRIDE using the **mzIdentML** Complete submission process (described in section 2A on the PRIDE site) then it is now possible to export the identification results as mzIdentML which can be checked in **Pride Inspector tool** as part of the submission process.

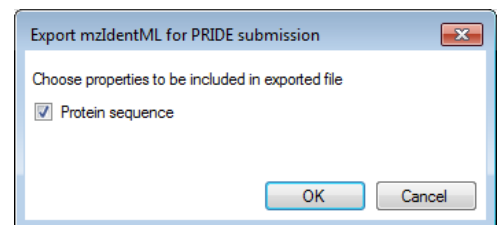
Note: this only supports identifications obtained from searches performed in Progenesis QI for proteomics version 4.0 or later, using the search methods Mascot or Ion Accounting.

Export Identifications is carried out at the Review Proteins stage of the workflow. Click on the **File** menu and select **Export mzIdentML for PRIDE submission....**

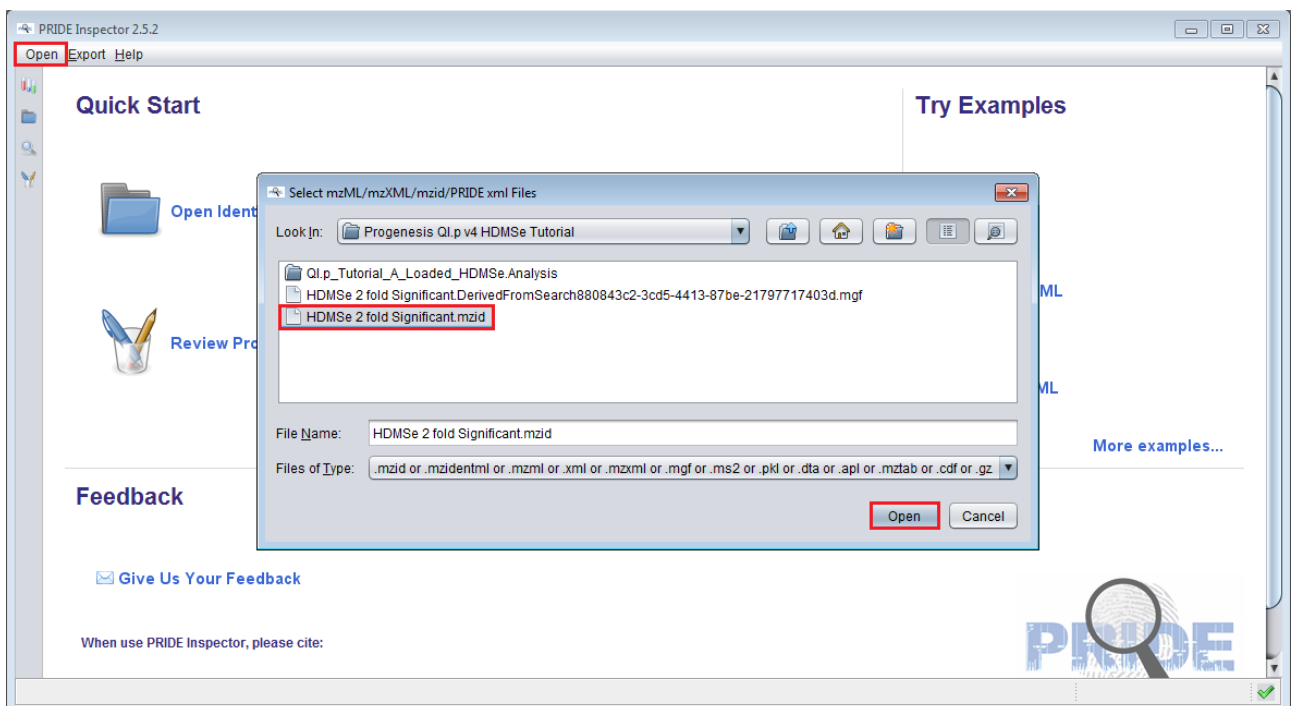


A dialog will appear allowing you to include the protein sequence or not in your export. Choosing to export the sequence will allow the display of the coverage map in PRIDE Inspector.

Enter a file name. Progenesis will save two files the mzIdentML as a .mzid file and also a corresponding .mgf file containing the spectra.



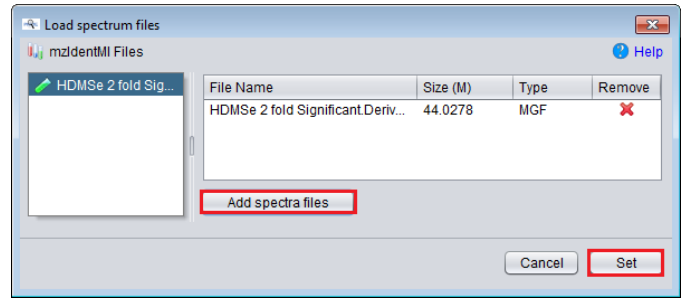
Open PRIDE Inspector and locate and select the mzid file to import.



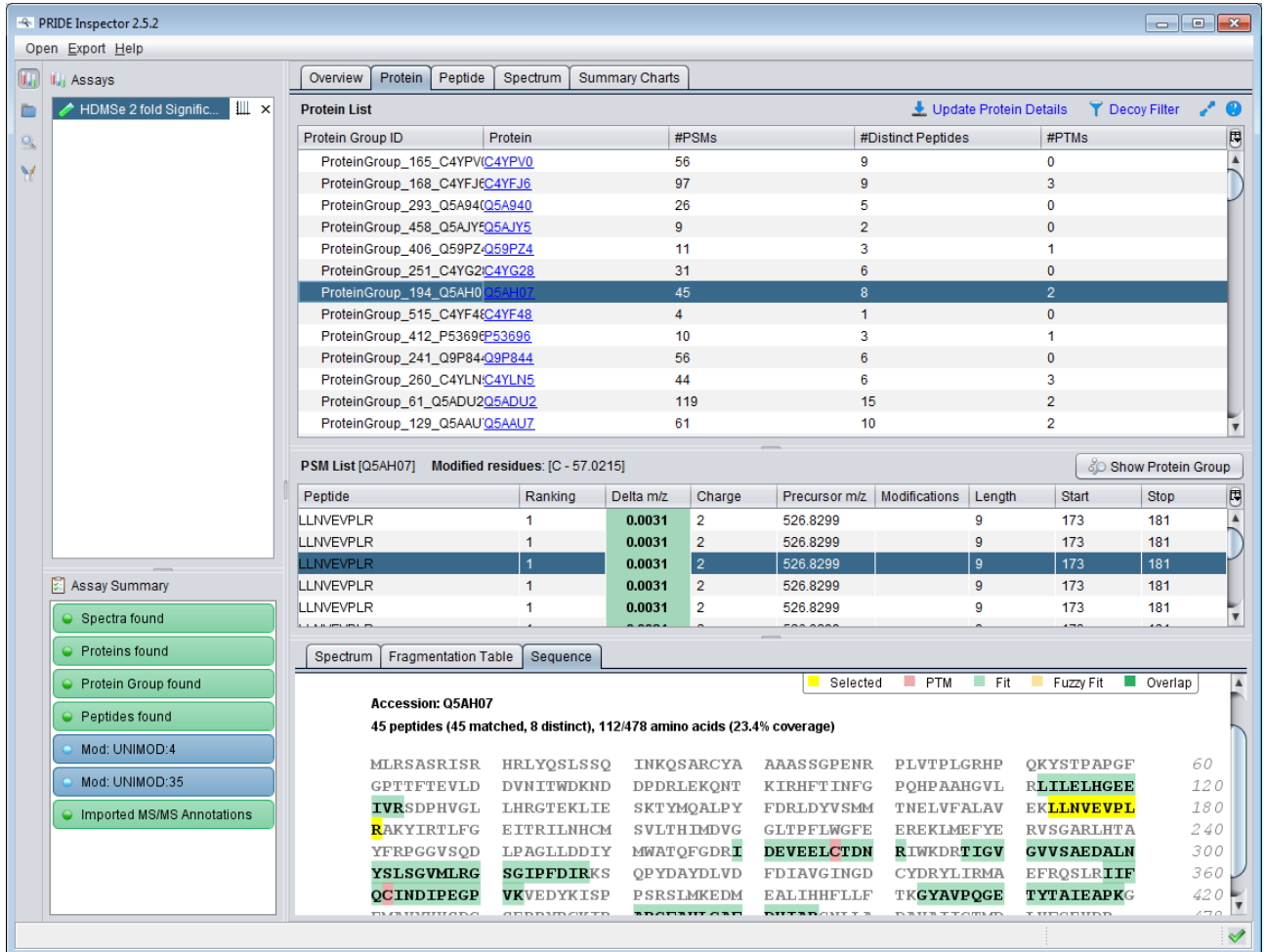
Click open.

You will then be asked if you want to load the corresponding spectrum files. Click yes and locate the .mgf file.

A dialog opens and allows you to **Add spectra files**. Add the corresponding .mgf file and click **Set**



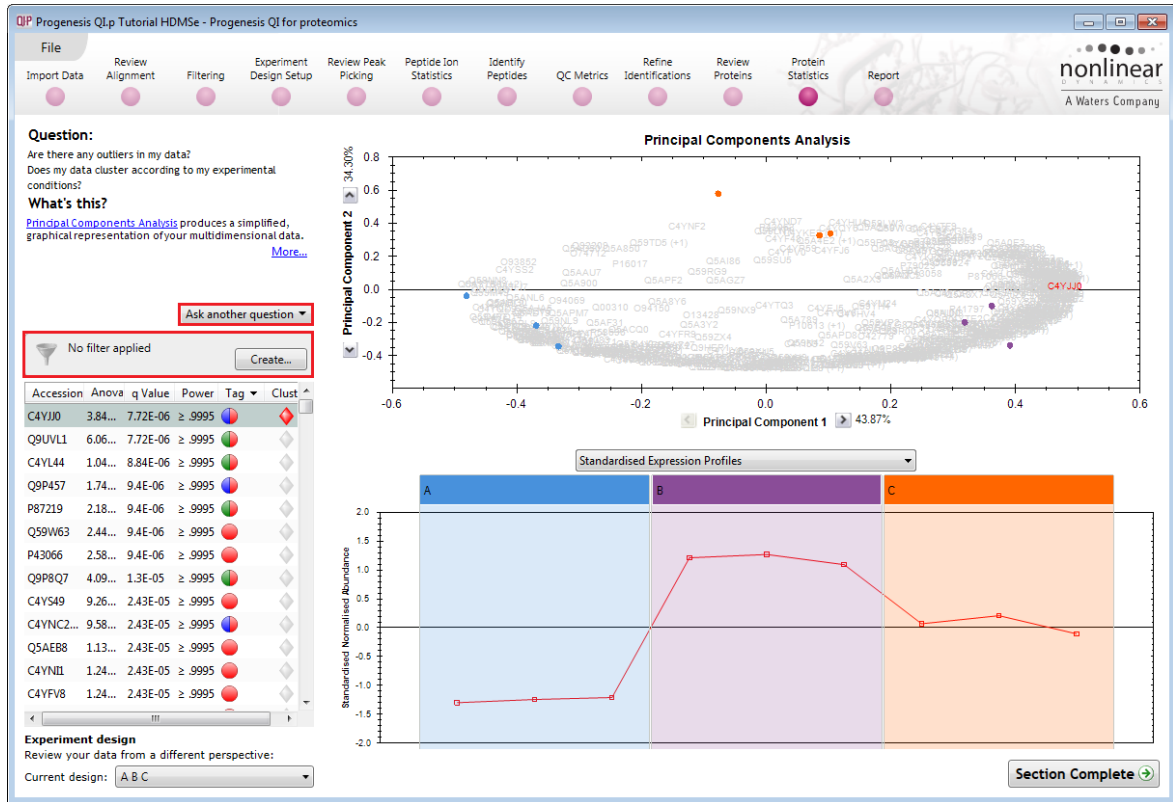
PRIDE Inspector will open allowing you to check the data.



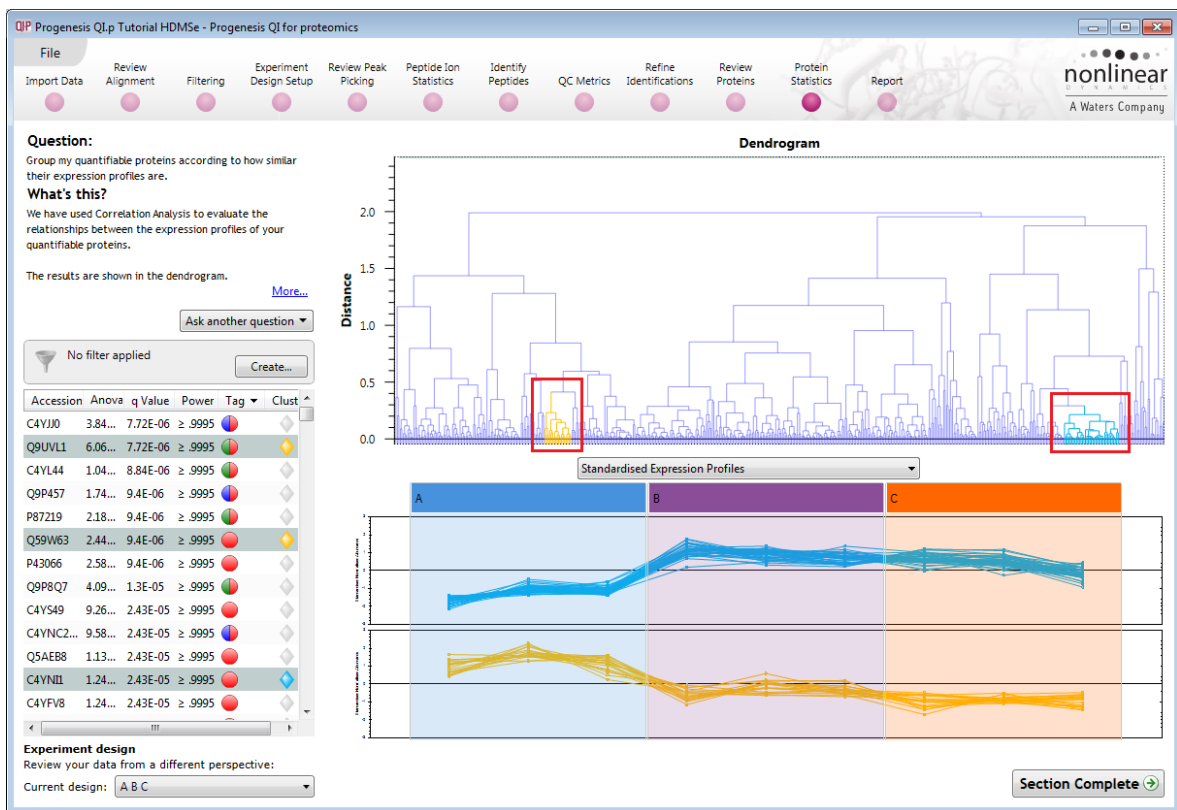
Once checked you can submit the results data to PRIDE, along with your raw data, as part of your submission process, following the guide lines provided on the PRIDE site.

Stage 14: 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. Allowing you to identify similar patterns of expression using the Correlation Analysis. Click on 2 of the branches (holding the **Ctrl** key down) to see differing patterns of expression.

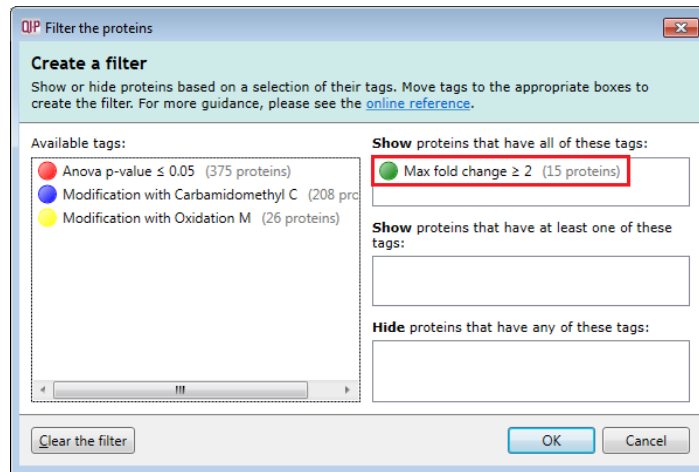


Now move to the **Report** section to report on Proteins and /or peptides.

Stage 15: 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**.

Note: this facility is used to generate Html reports on a limited selection of Proteins in your data. Creating a report on all the data in your experiment can take a long time



As an example we will create a report for **only** the proteins showing a Max Fold change of greater than 2.

1. First reduce the proteins to report on by selecting the '**Max fold change \geq 2**' tag. In this example it reduces the number of proteins in the table to 15.
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.

Protein report | Peptide report

Report on your proteins
Generate a report containing the proteins of interest in your experiment.

1 **Choose what to report upon**
Using the list below, filter your data to show only the proteins you want to include in the report.
[Learn about tagging and filtering](#)

2 **Customise the report**
Enter a title for your report and select the sections you want to include in it using the controls to the right.

Tag filter applied
proteins may be hidden

Accession	Anova (p)	Fold	Tag	Description
Q9UVL1	6.06E-07	2.4	●●●	Non-histone ch
C4YL44	1.04E-06	2.2	●●●	Extracellular ma
P87219	2.18E-06	4.1	●●●	Sorbose reduct
Q9P8Q7	4.09E-06	2.0	●●●	Isocitrate lyase
P42800	2.16E-05	2.7	●●●	Inositol-3-phos
Q7Z8E8	2.39E-05	3.3	●●●	Cell surface hyd
C4YF25	3.44E-05	2.3	●●●	Bud site selecti
C4YR44	0.00108	2.1	●●●	NAD(P)H-deper
C4YE92	0.0018	2.6	●●●	6,7-dimethyl-8-
Q5ALX8	0.00206	2.1	●●●	Adenine phosph
Q59U83	0.00315	4.1	●●●	Ribonucleoside
Q59NN8	0.00544	3.5	●●●	Hsp70 nucleotic
Q9HFQ6	0.00581	2.2	●●●	60S acidic ribos
Q59WGO	0.0125	3.2	●●●	Hit family prote
Q59Z58	0.188	10.6	●●●	Protoplast secr

Experiment design: ABC

Structure the report
Enter a report title:
Max Fold Increased Proteins

Select the sections to include in the report:

- Overview run
- Data processing methods
- Experiment design
- Protein report
Includes a table describing the structure of the experiment i.e. the samples in each experimental condition
 - Protein table
 - Peptide tables
 - Protein Details
Reports the full details of every protein which matches your current filter
 - Tags
 - Expression profile
 - Peptide ion table
 - Peptide ion details

Create report

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

Max Fold Increased Proteins

Experiment: Progenesis QI.p Tutorial HDMSe
Report created: 02/09/2019 13:31:52

Proteins

Protein building options
 Protein grouping **Group similar proteins**
 Protein quantitation **Absolute Quantitation using Hi-3**

Accession	Peptides	Score	Anova (p)*	Fold	Tags	Description	Average Normalised Abundances		
							A	B	C
Q9P8Q7	6 (6)	56.53	4.09e-006	2.01		Isocitrate lyase GN=ICL1 PE=3 SV=1	2.07	4.16	3.31

Q9P8Q7

Isocitrate lyase GN=ICL1 PE=3 SV=1
6 peptides

Sequence	Peptide Ion	Score	Hits	Mass	Charge	Tags	Conflicts	Modifications	In quantitation	Drift time (ms)	Average Normalised Abundances		
											A	B	C
ASADIFGSNLLAVAR	9787	7.39	9	1503.8215	2		0		yes	4.28	6636.02	1.81e+004	1.49e-004

Accession Q9P8Q7

Description Isocitrate lyase GN=ICL1 PE=3 SV=1
Peptides 6 (6)
Score 56.53
Anova 4.09e-006
Fold 2.01

- Anova p-value ≤ 0.05
- Max fold change ≥ 2

Condition	Mean Abundance
A	~1.5
B	~2.1
C	~1.9

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.

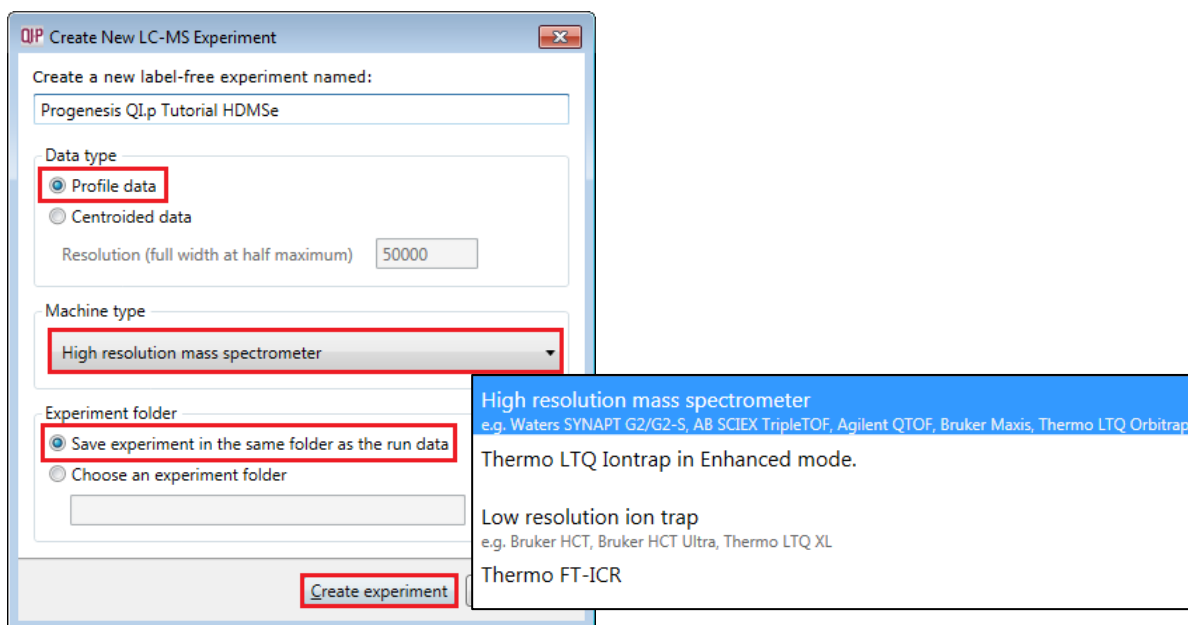
Note: there are separate panels for reporting on Proteins and Peptides.

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, Thermo, Bruker, SCIEX and Agilent) or, for other Vendors, convert them to mzXML or mzML format first.

To create a new experiment with your (Waters) 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'.

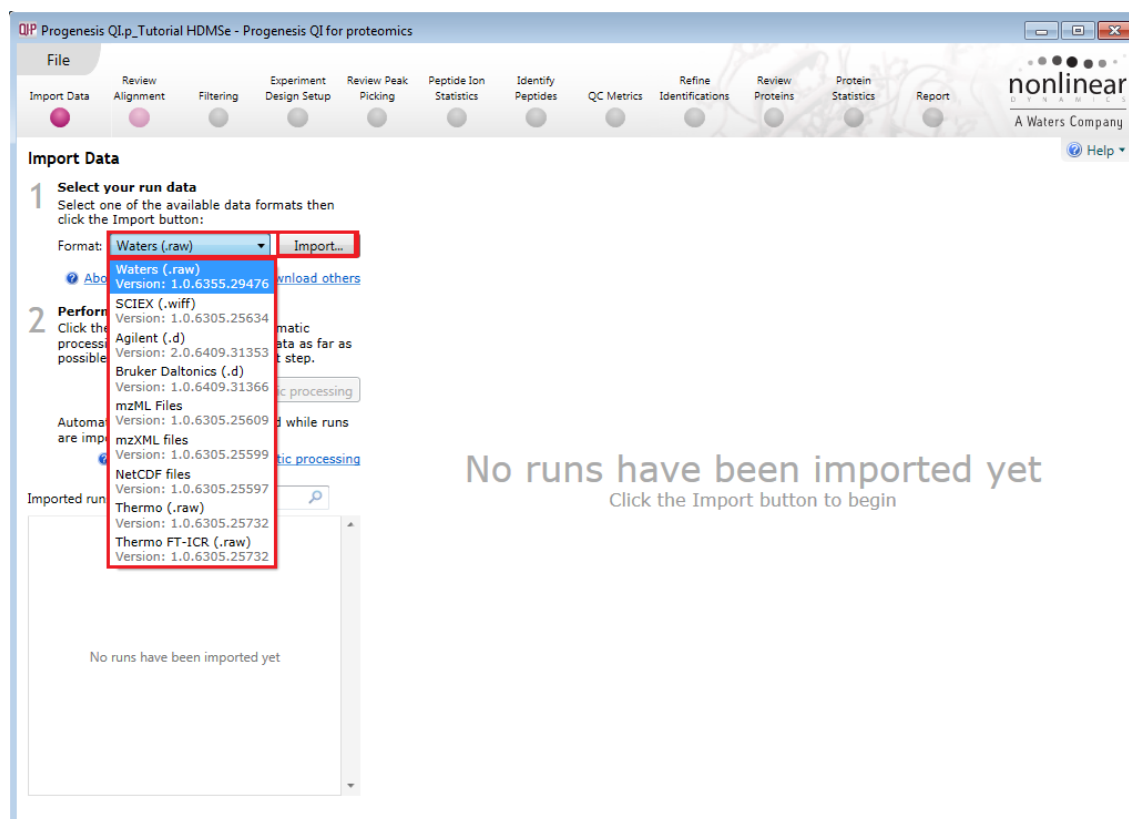
Note: if you have converted or captured the data as centroided then select Centroided data and enter the Resolution for the MS machine used.



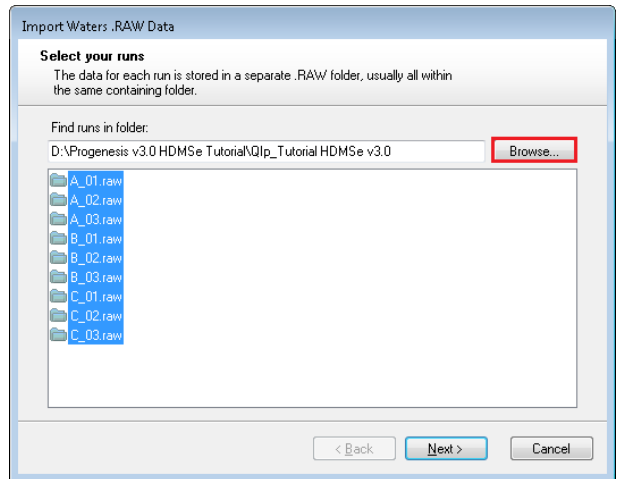
Click **Create experiment** to open the LC-MS Data Import stage of the workflow.

Select the 'Import Data file format', in this example they are Waters/SYNAPT data

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

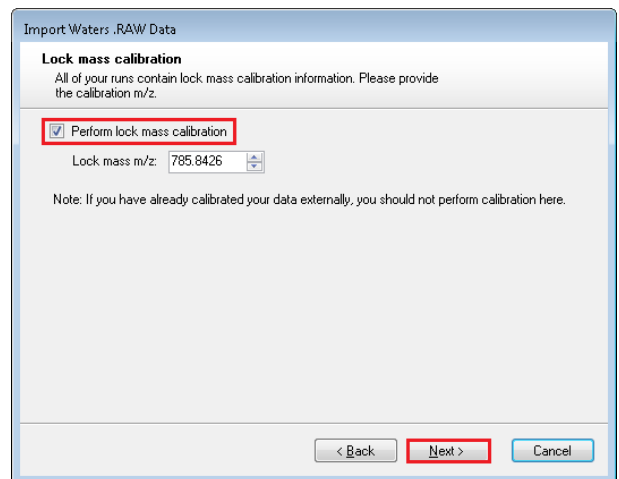


Locate and select all the .RAW folders (A_01 to C_03).



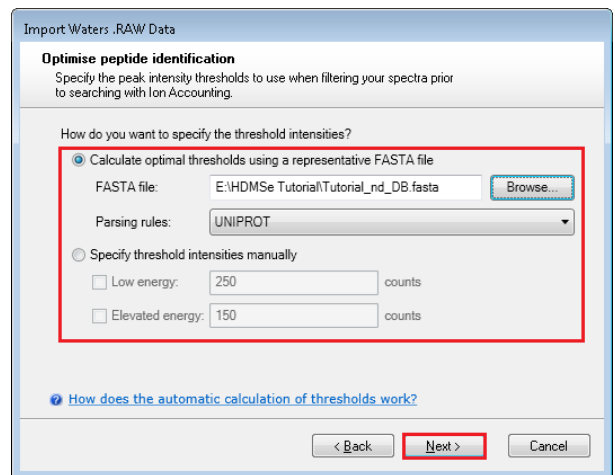
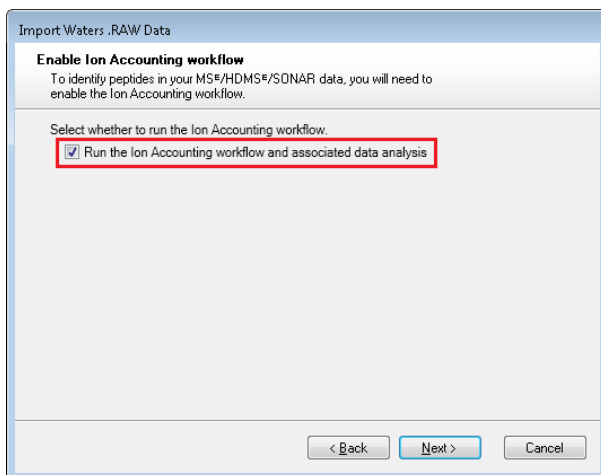
On importing, the lock mass calibration is read and presented on this dialog

You can, if required, alter the lock mass calibration at this step.



For MSe, HDMSe and SONAR data formats the Ion accounting workflow is selected as default if your computer has a GPU.

Click **Next**. You can either choose to calculate optimal thresholds using an appropriate FASTA file for your data set or set the Thresholds manually.



If you choose to determine the thresholds automatically then Progenesis determines appropriate thresholds by sampling each run and finding the thresholds that yield the most protein identifications in the sample area. For each run, Progenesis performs the following steps:

- It finds the 5-minute retention time window that contains the highest total intensity.
- It extracts the ions within this window and performs multiple Ion Accounting searches, each one using a different set of threshold values.

- Finally, it selects the thresholds that resulted in the largest number of protein identifications and applies those to the whole run. If more than one set of thresholds results in similar numbers of identifications (within 10% of the maximum), it will choose the highest thresholds as a way of optimising system performance.

Note: for HDMSe the default settings are 150 and 30 and for MSe the default settings are 250 and 150 for the Low and Elevated energies respectively.

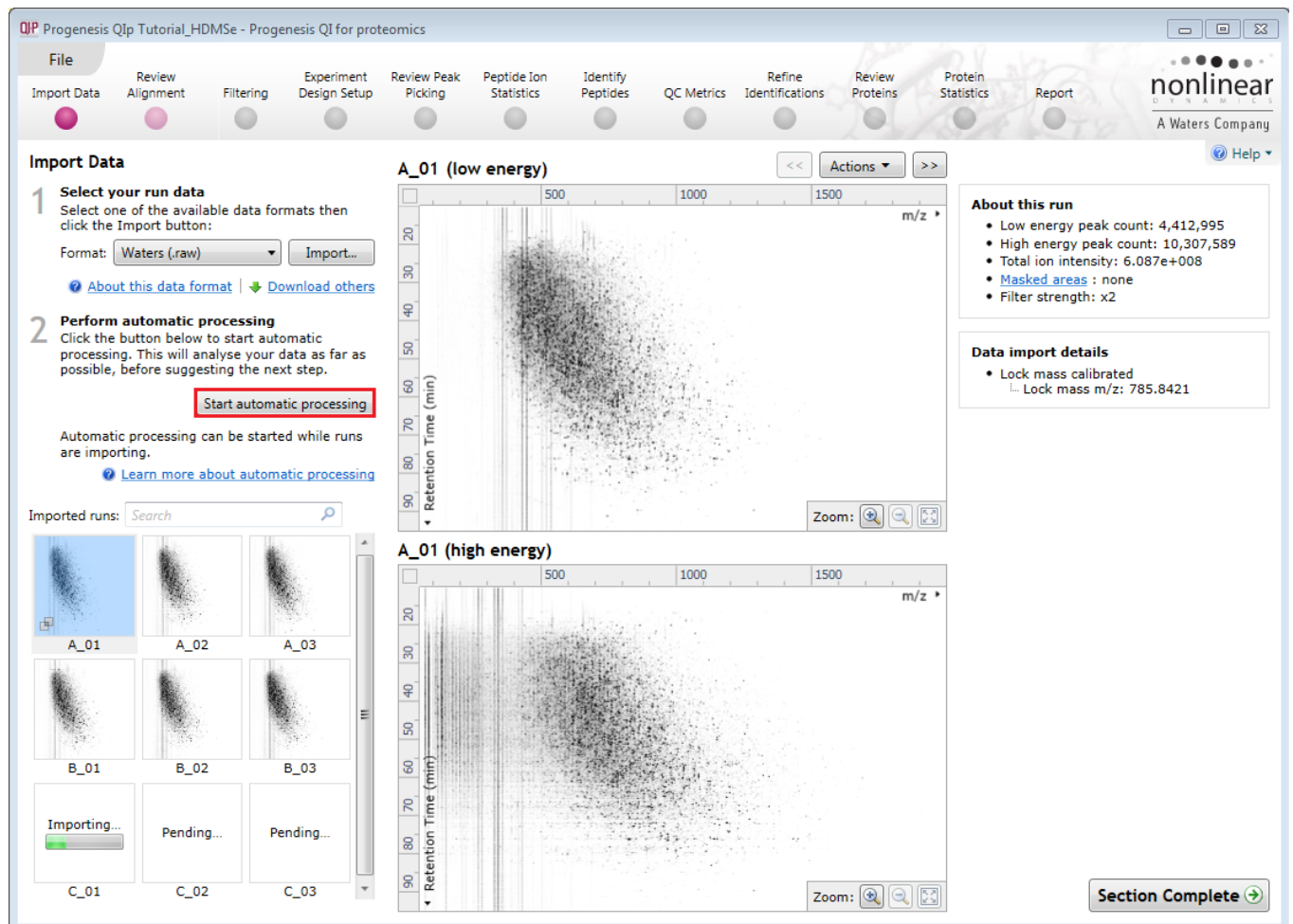
Having selected how to handle the thresholds you will get the option to specify your own elution limits, the default is Start and End of the run, accept or make changes as necessary.

A summary of the loading parameters is provided before you click Import

On loading the selected runs your data set will be automatically examined and the size of each file will be reduced by a 'data modelling 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: you can start the automatic processing before the loading has completed.



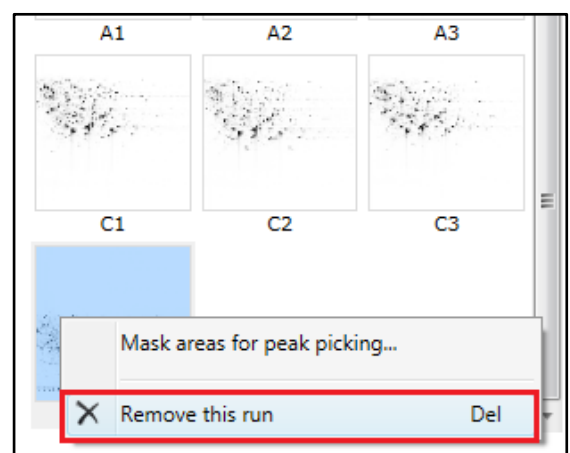
For details of setting up the steps in the automatic processing wizard return to Stage 2A page 7.

Review Chromatography

Each data file appears as a 2D representation of the run. If you created a **profile** experiment, at this stage you will be warned if any of the data files have been 'centroided' during the data acquisition and conversion process, as files must be of one format or the other.

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

Note: you can also multi-select runs to remove by holding down the Ctrl key.



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

The screenshot shows the Progenesis QI software interface during the 'Import Data' stage. The main window displays a 2D plot of 'A_01 (low energy)' and 'A_01 (high energy)' with axes for 'Retention Time (min)' and 'm/z'. The plot shows a dense cluster of peaks. On the right side, there is a sidebar with information about the current run, including peak counts and ion intensity. A 'Section Complete' button is visible at the bottom right.

Import Data

- Select your run data**
Select one of the available data formats then click the Import button:
Format: Waters (.raw) Import...
[About this data format](#) [Download others](#)
- Perform automatic processing**
Click the button below to start automatic processing. This will analyse your data as far as possible, before suggesting the next step.
[Restart automatic processing](#)
Automatic processing can be started while runs are importing.
[Learn more about automatic processing](#)

Imported runs: Search

A_01 A_02 A_03
B_01 B_02 B_03
C_01 C_02 C_03

A_01 (low energy)

Retention Time (min) m/z

A_01 (high energy)

Retention Time (min) m/z

About this run

- Low energy peak count: 4,413,012
- High energy peak count: 10,307,607
- Total ion intensity: 6.087e+008
- Masked areas : none
- Filter strength: x2

Alignment reference

This run is being used as the experiment's alignment reference.

If you want to use a different run as the alignment reference, you'll need to discard any analysis and restart the automatic processing:

[Restart automatic processing](#)

Data import details

- Lock mass calibrated
Lock mass m/z: 785.8426

Section Complete

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** on the workflow or **Section Complete** to move forward to the Review Alignment Stage.

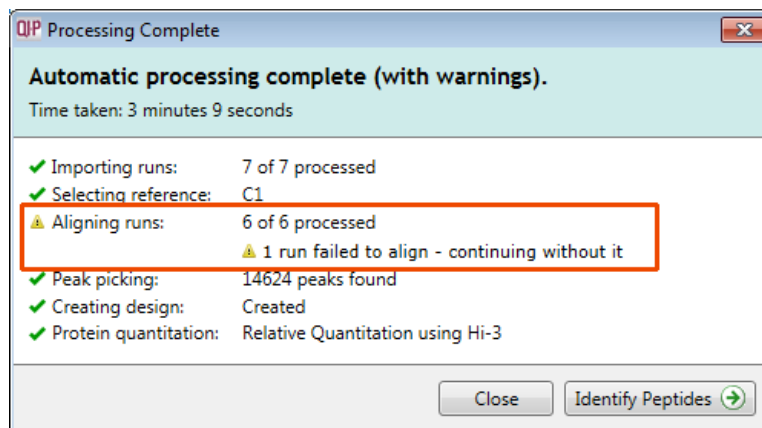
Note: you will be offered the automatic alignment if you have not performed it automatically already.

Now move to the next stage in the workflow (page 7 in this user guide).

Appendix 2: Stage 1 Processing failures

During automatic processing if a stage fails to complete successfully or only partially completes, the automatic processing dialog will warn you of the problem. Depending on the type of failure 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; however, the automatic processing dialog will prompt you to 'drop-off' at the **Review Alignment** stage on completion to investigate the problem.



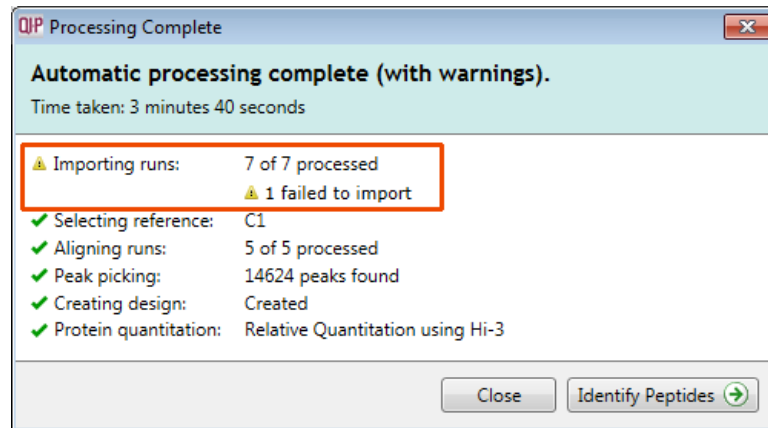
Note: in this example the run that failed to align will not contribute to the peak picking and will be excluded at the alignment stage (a cross appears in the include column).

Run	Include?	Vectors	Score
A1	<input checked="" type="checkbox"/>	221	84.3%
A2	<input checked="" type="checkbox"/>	240	82.2%
A3	<input checked="" type="checkbox"/>	244	83.9%
C1	<input checked="" type="checkbox"/>	Ref	
C2	<input checked="" type="checkbox"/>	362	97.8%
C3	<input checked="" type="checkbox"/>	422	99.0%
D1	<input checked="" type="checkbox"/>	0	2.0%

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.

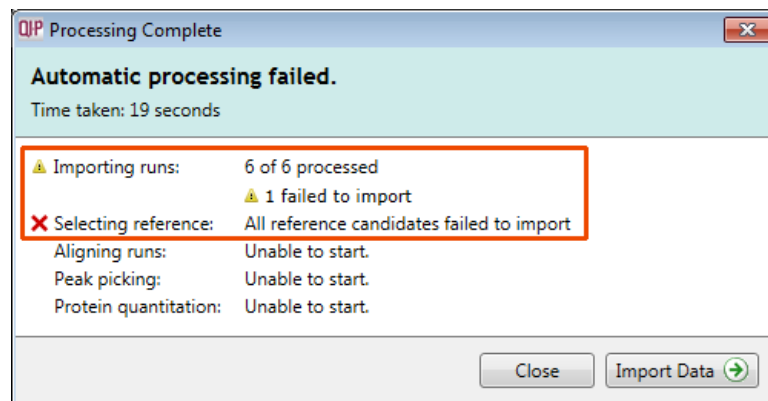
If some runs in a data set fail to import (but not all), the automatic processing will continue informing you that one or more runs have failed to import.



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: while importing, you specified the selection of the alignment reference to be made 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: 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.

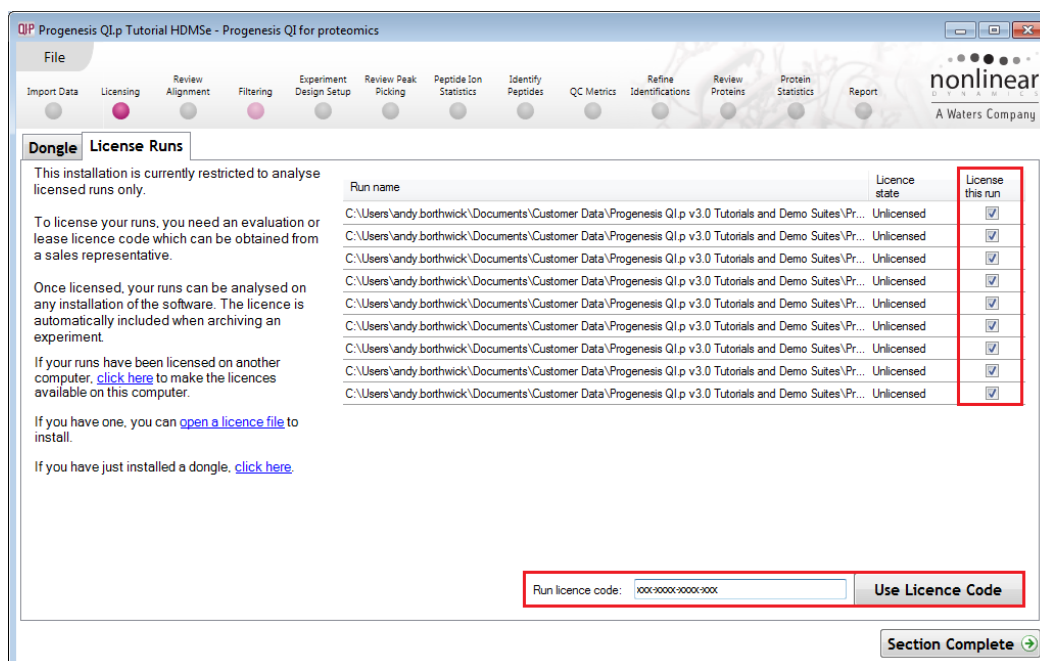
To use this page to License your Runs you must first either obtain an **'Evaluation' Licence Code from a Sales Person or purchase a licence code directly**.

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

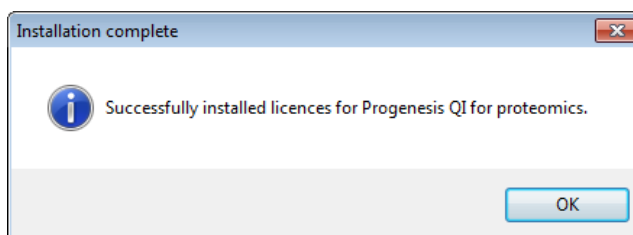
The runs in your experiment will be listed as shown below.

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 4: Manual assistance of Alignment

Approach to alignment

To place manual alignment vectors on a run (B_02 in this example):

Review Alignment
Sample ions are aligned to compensate for drifts in retention time between runs.

1 **Align retention times automatically**
For maximum reproducibility, the software can automatically align your runs.

2 **Review the alignment**
Using the quality control measures, review and edit the runs' alignment:

- Order the runs by alignment score and start by selecting the first run
- Within each run, inspect and edit any areas rated as Needs Review

Run	Include?	Vectors	Score
A_01	<input checked="" type="checkbox"/>	Ref	
A_02	<input checked="" type="checkbox"/>	0	56.79
A_03	<input checked="" type="checkbox"/>	0	81.79
B_01	<input checked="" type="checkbox"/>	0	66.39
B_02	<input checked="" type="checkbox"/>	0	40.29
B_03	<input checked="" type="checkbox"/>	0	48.09
C_01	<input checked="" type="checkbox"/>	0	35.79
C_02	<input checked="" type="checkbox"/>	0	87.29
C_03	<input checked="" type="checkbox"/>	0	63.19

Ion maps: ■ Alignment target ■ Run being aligned

Alignment quality: ■ Good ■ OK ■ Needs review

Section Complete

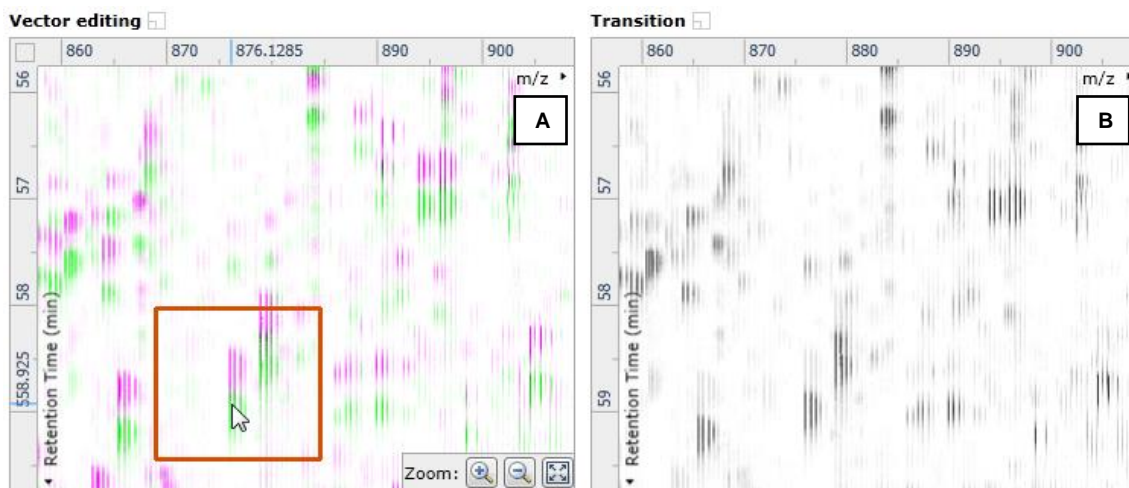
- Click on Run B_02 in the **Runs** panel, this will be highlighted in green and the reference run (A_01) will be highlighted in magenta.
- You will need to place approximately 5 - 10 **alignment vectors** evenly distributed from top to bottom of the whole run (RT range).
- First drag out an area on 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.

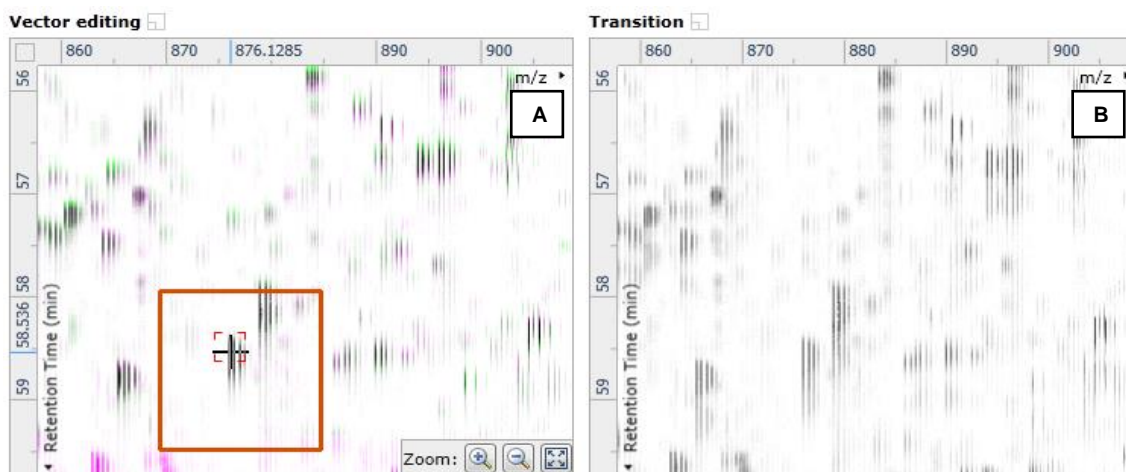
Note: the Ion Intensity Map gives you a colour metric, visually scoring the current alignment and an overall score is placed next to the Vectors column in the table. With each additional vector added this score will update to reflect the 'changing' overall quality of the alignment. The colour coding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

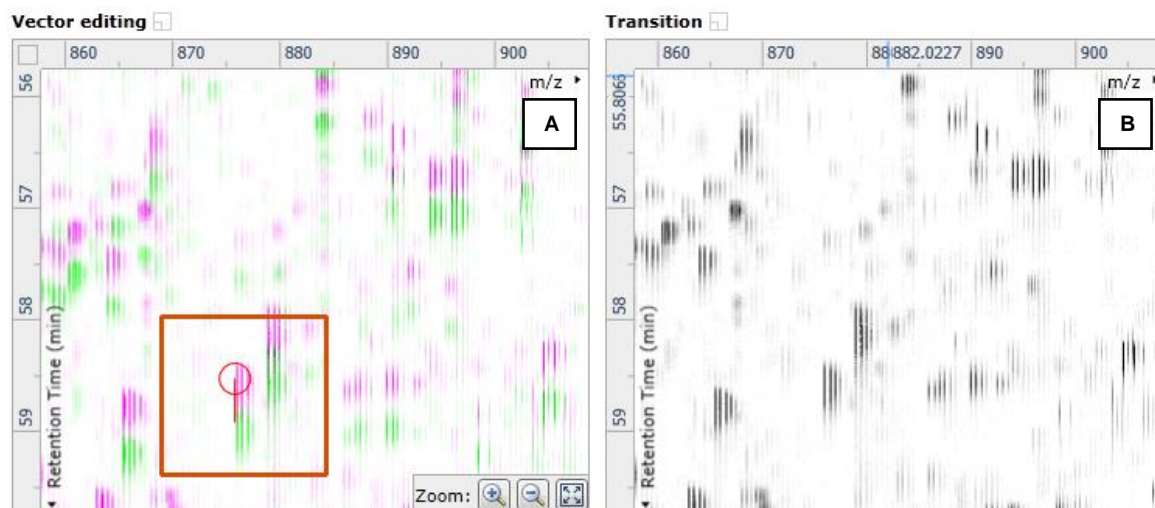


5. As you are holding down the left mouse button (depending on the severity of the misalignment), the alignment vector will automatically find the correct lock. If not, drag the green peptide ion over the corresponding magenta peptide ion of the reference run. The red box will appear as shown below indicating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



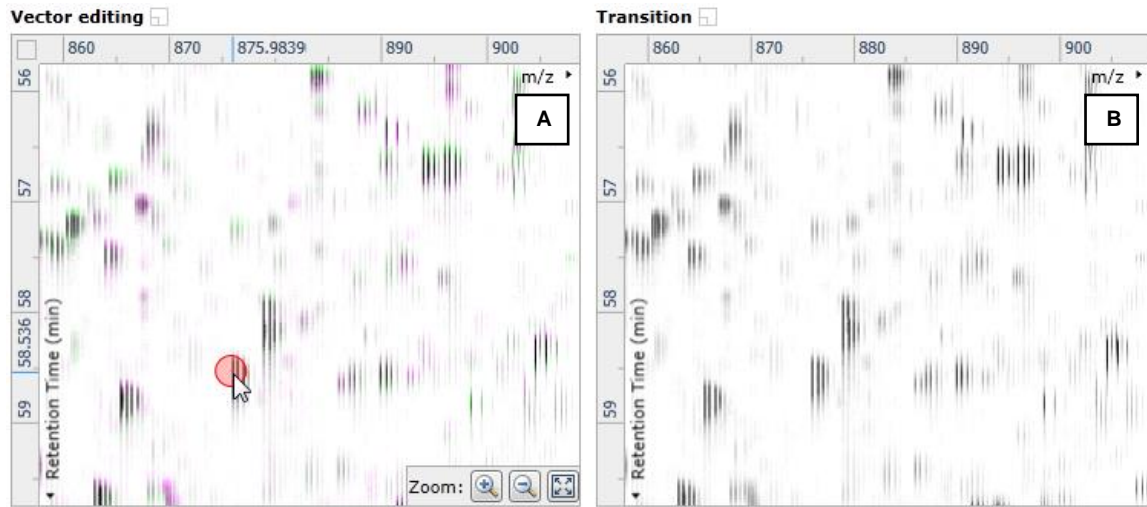
Tip: while holding down the mouse button hold down the **Alt** key. This will allow smooth movement of the cursor as the **Alt** key allows you to override the 'automatic alignment' performed as you depress the mouse button.

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.

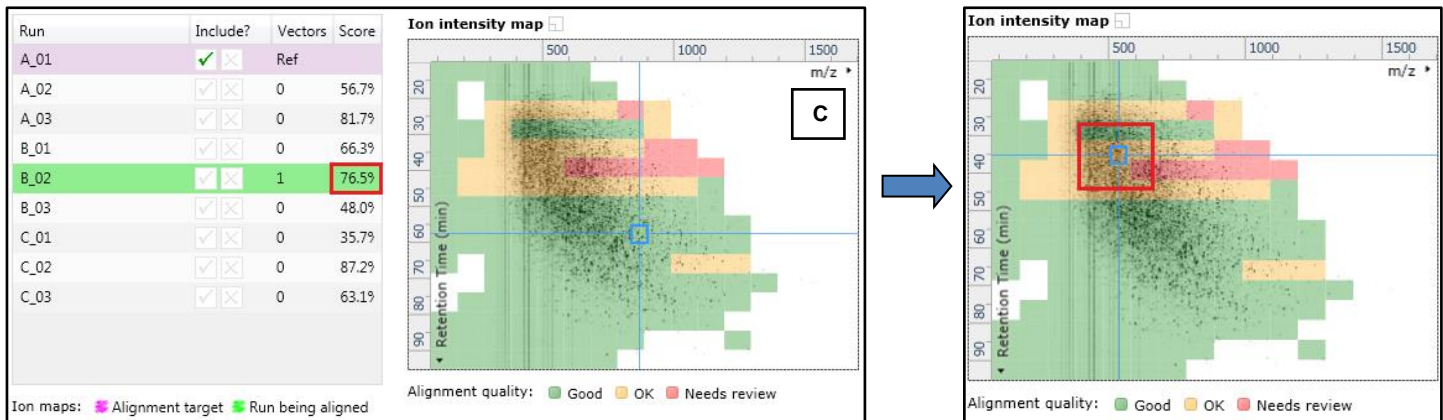


Note: an incorrectly placed vector is removed by right clicking on it in the **Vector Editing** window and selecting delete vector.

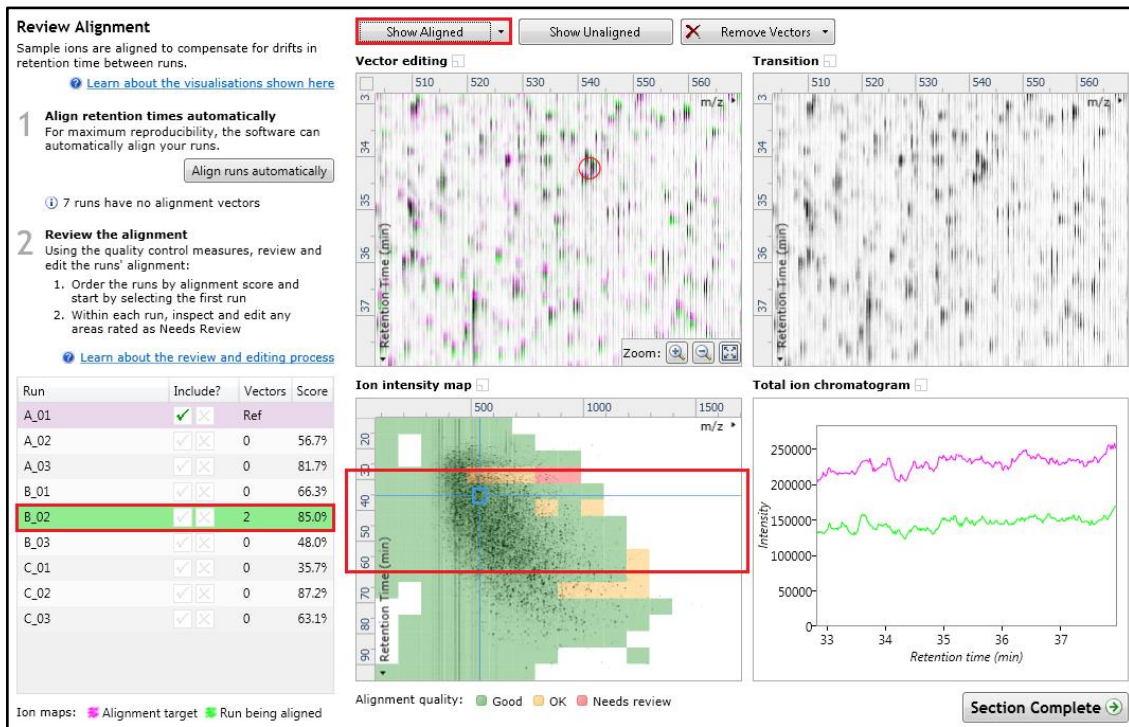
- Now click **Show Aligned** on the top tool bar to see the effect of adding a single vector.



- 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.

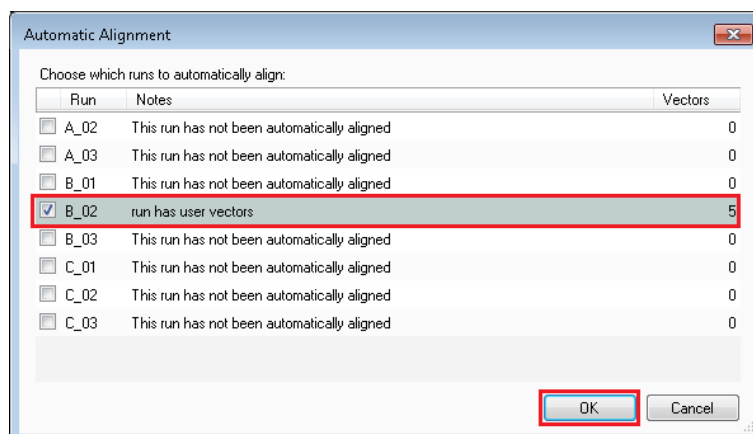


10. The shift in the Retention Time (RT) is as a result of incorrect running of the chromatography. In many of these cases if the Automatic Alignment fails to generate a good alignment then removing all the alignment vectors for this run and placing a single manual vector to act as a 'seed' for the Automatic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

11. In the case of the example shown above placing a small number of vectors from the top to the bottom of the run is sufficient to markedly improve the alignment.



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



13. On pressing OK the Automatic Alignment will run for the selected run. On completion the table and views will update to display the automatically generated vectors (shown in blue).

Review Alignment
Sample ions are aligned to compensate for drifts in retention time between runs.

[Learn about the visualisations shown here](#)

- Align retention times automatically**
For maximum reproducibility, the software can automatically align your runs.
- Review the alignment**
Using the quality control measures, review and edit the runs' alignment:
 - Order the runs by alignment score and start by selecting the first run
 - Within each run, inspect and edit any areas rated as Needs Review

[Learn about the review and editing process](#)

Run	Include?	Vectors	Score
A_01	<input checked="" type="checkbox"/>	Ref	
A_02	<input checked="" type="checkbox"/>	0	56.79
A_03	<input checked="" type="checkbox"/>	0	81.79
B_01	<input checked="" type="checkbox"/>	0	66.39
B_02	<input checked="" type="checkbox"/>	1842	96.09
B_03	<input checked="" type="checkbox"/>	0	48.09
C_01	<input checked="" type="checkbox"/>	0	35.79
C_02	<input checked="" type="checkbox"/>	0	87.29
C_03	<input checked="" type="checkbox"/>	0	63.19

Ion maps: ■ Alignment target ■ Run being aligned

Alignment quality: ■ Good ■ OK ■ Needs review

Section Complete

Vector editing **Transition**

Ion intensity map **Total ion chromatogram**

Buttons: Show Aligned, Show Unaligned, Remove Vectors

14. Repeat this process for all the runs to be aligned.

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.

Also the 'ease' of addition of vectors is dependent on the actual differences between the LC-MS runs being aligned.

To review the vectors, automatic and manual, return to page 18

Appendix 5: 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

QIP Progenesis QI.p Tutorial HDMSe - Progenesis QI for proteomics

File Import Data Review Alignment Filtering Experiment Design Setup Review Peak Picking Peptide Ion Statistics Identify Peptides QC Metrics Refine Identifications Review Proteins Protein Statistics Report

nonlinear
A Waters Company

Which experiment design type do you want to use for this experiment?

Between-subject Design

Do samples from a given subject appear in only one condition? Then use the between-subject design.

To set up this design, you simply group the runs according to the condition (factor level) of the samples. The ANOVA calculation assumes that the conditions are independent and therefore gives a statistical test of whether the means of the conditions are all equal.

Within-subject Design

Have you taken samples from a given subject under different conditions? Then use the within-subject design.

A standard ANOVA is not appropriate because the data violates the ANOVA assumption of independence. With a repeated measures ANOVA individual differences can be eliminated or reduced as a source of between condition differences (which helps to create a more powerful test).

The within-subject design can be thought of as an extension of the paired-samples t-test to include comparison between more than two repeated measures.

QIP Create New Experiment Design

Enter a name for the experiment design:
Before During and After Treatment

How do you want to group the runs?

Group the runs manually

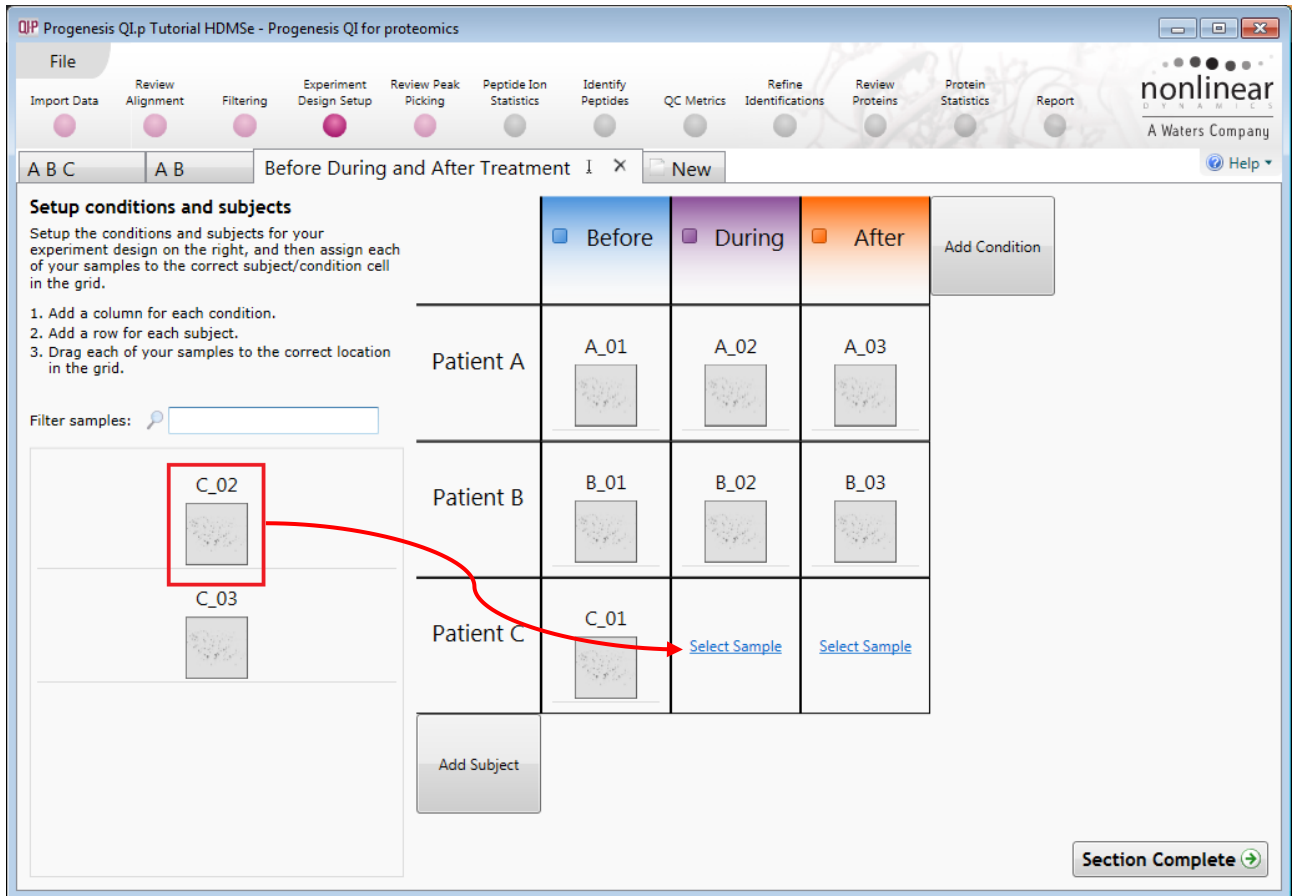
Copy an existing design: _____

Create design Cancel

	Before	During	After
Patient X	X1	X2	X3
Patient Y	Y1	Y2	Y3
Patient Z	Z1	Z2	Z3

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 the later stages of the workflow with the exception of **Identify Peptides**, and **Refine Identifications** (including Resolve Conflicts).

Appendix 6: Power Analysis (Peptide Ion 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 Peptide Ion 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 Peptide Ion Stats screen. A selection of 3 tools will appear in the form of questions.

Principal Components Analysis
Are there any outliers in my data?
Does my data cluster according to my experimental conditions?

Correlation Analysis
Group my quantifiable proteins according to how similar their expression profiles are.

Power Analysis
How many replicates should I run?
What is the power of my experiment?

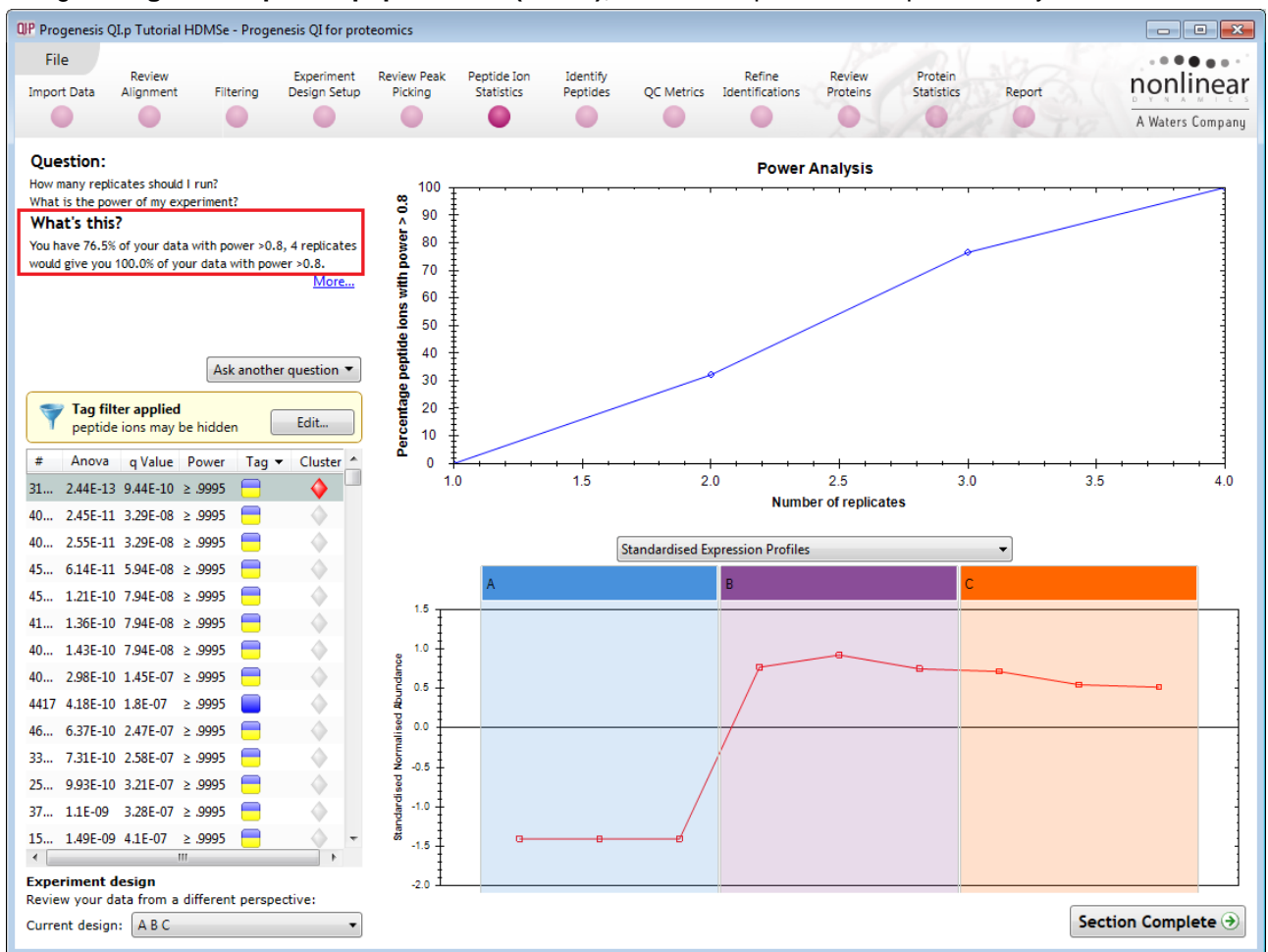
Select the option

'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 (20641)**, as an example, view the power analysis.



This is displayed graphically showing that 76.5% of the 20641 peptide ions have a power of 80% or that 4 replicates would give you 100% 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%.

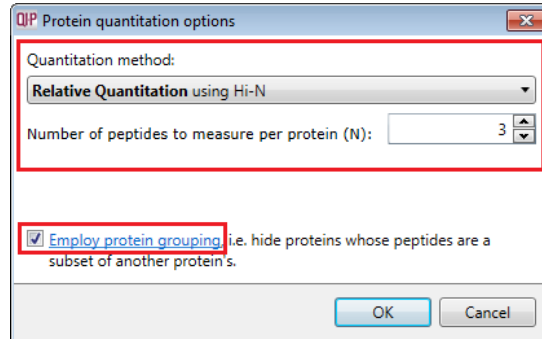
Appendix 7: Resolve Conflicts

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

The **Resolve Conflicts** stage is now accessed at the bottom left of the **Refine Identifications** stage.

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



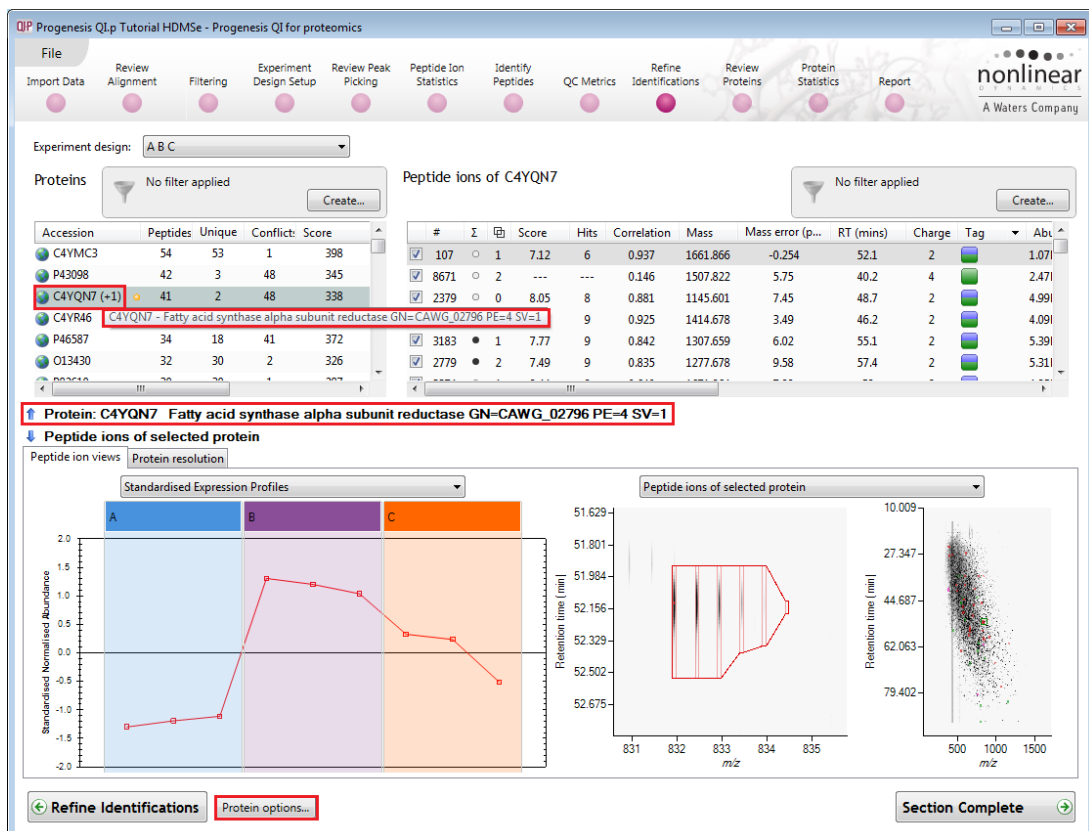
This means that if you choose **not** to resolve the conflicts then proteins, to be considered for quantitation, require at least one unique peptide (see Unique column).

For this guide: we will not resolve the conflicts but click on Section Complete to go to Review Proteins (page 61).

(For more details on Protein Grouping page 56 and Protein Quantitation options go to page 58).

If you wish/require to resolve conflicts then the remainder of this appendix provides some guidance and explanation of performing this process in Progenesis QI for proteomics.

With **Group similar proteins** selected the additional members are indicated by a bracketed number located after the Accession number.



The number of the **Peptides** used for quantitation is displayed in the **Unique** column.

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.

With **Resolve Conflicts** open 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 pages may vary slightly.

The screenshot shows the Progenesis QI software interface. At the top, there is a menu bar with options like File, Review Alignment, Experiment Design Setup, Review Peak Picking, Peptide Ion Statistics, Identify Peptides, QC Metrics, Refine Identifications, Review Proteins, Protein Statistics, and Report. Below the menu, the 'Experiment design' is set to 'A B C'. The 'Proteins' section shows a list of proteins with columns for Accession, Peptides, Unique, Conflict, and Score. The first protein, P34731, is highlighted with a red box and labeled 'A'. The 'Peptide ions of P34731' section shows a list of peptide ions with columns for #, I, Score, Hits, Correlation, Mass, Mass error (p...), RT (mins), Charge, Tag, and Abun. The first peptide ion, 742, is highlighted with a red box and labeled 'B'. Below these sections, there are two main panels. Panel 'C' is titled 'Standardised Expression Profiles' and shows a line graph with three distinct peaks labeled A, B, and C. Panel 'D' is titled 'Peptide ions of selected protein' and shows two plots: a retention time vs. m/z plot and a retention time vs. m/z scatter plot. The software interface also includes buttons for 'Refine Identifications', 'Protein options...', and 'Section Complete'.

Select the first protein in list A (in this case it has 70 conflicts) the panel to the right B lists the peptides for this protein and the conflicting protein for each peptide.

This screenshot shows a different view of the Progenesis QI software interface. The 'Proteins' section shows a list of proteins with columns for Accession, Peptides, Unique, Conflict, and Score. The first protein, P34731, is highlighted with a red box and labeled 'A'. The 'Peptide ions of P34731' section shows a list of peptide ions with columns for #, I, Score, Hits, Correlation, Mass, Mass error (p...), RT (mins), Charge, Tag, and Abun. The first peptide ion, 742, is highlighted with a red box and labeled 'B'. Below these sections, there are two main panels. Panel 'E' is titled 'Conflicting proteins for peptide ion 742' and shows a table with columns for Accession, Peptides, Unique, Conflict, and Score. The first protein, P34731, is highlighted with a red box and labeled 'A'. The second protein, C4YQR7, is highlighted with a red box and labeled 'E'. Panel 'F' is titled 'Peptide ions of C4YQR7' and shows a list of peptide ions with columns for #, I, Score, Hits, Correlation, Mass, Mass error (p...), RT (mins), Charge, Tag, and Abun. The first peptide ion, 742, is highlighted with a red box and labeled 'F'. The software interface also includes buttons for 'Refine Identifications', 'Protein options...', and 'Section Complete'.

Panel C shows the expression profile(s) for the peptide(s) selected in list B

Panel D shows the details for the selected peptide.

Now click on the **Protein Resolution** in Panel C to display the proteins that are conflicting.

The lower left panel E displays the Conflicting proteins for the peptide ion highlighted in panel B this includes the current protein in panel A as indicated by the orange ball to the right of the accession.

The Accession and description for the 2 proteins highlighted in Panels A and E are shown in the middle margin. As most of the peptide ions are conflicting between the 2 closely related proteins one simple way to resolve these conflicts is to favour the protein with the higher score and greater number of non-conflicting peptides.

One way to do this is to right click on the lower scoring protein in panel E which only has one unique peptide and turn off all its peptides

Accession	Peptides	Unique	Conflict	Protein Score
P34731	58	4	70	529
C4YQR7	55	1	70	500

Turn off all peptides

All the peptides are now switched off in panel B and all the entries for the lower scoring protein are set to zero. The higher scoring protein now has 56 non-conflicting peptides and only 4 remaining conflicts

Panel A: Protein Summary

Accession	Peptides	Unique	Conflict	Score	Tag
P34731	58	56	3	529	
C4YQR7	0	0	0	0	

Panel B: Peptide ions of C4YQR7

#	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict
742	8.11	9	---	1478.765	9.73	45.3	2		6.74E+04	0

Panel E: Conflicting proteins for peptide ion 742

Accession	Peptides	Unique	Conflict	Protein Score	Pept
P34731	58	56	3	529	8.11
C4YQR7	0	0	0	0	8.11

Panel F: Peptide ions of C4YQR7

#	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict
742	8.11	9	---	1478.765	9.73	45.3	2		6.74E+04	0

To resolve the remaining conflict(s) 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.

Panel A: Protein Summary

Accession	Peptides	Unique	Conflict	Score	Tag
P34731	58	56	3	529	
C4YQP9	3	1	3	25.2	

Panel B: Peptide ions of C4YQP9

#	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict
2183	5.9	2	0.718	1529.861	3.14	45	3		5.05E+04	1

Panel E: Conflicting proteins for peptide ion 2183

Accession	Peptides	Unique	Conflict	Protein Score	Pept
P34731	58	56	3	529	5.9
C4YQP9	3	1	3	25.2	5.9

Panel F: Peptide ions of C4YQP9

#	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict
2183	5.9	2	0.718	1529.861	3.14	45	3		5.05E+04	1

Favouring the protein with the higher score, but this time 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.

Panel A: Protein Summary

Accession	Peptides	Unique	Conflict	Score	Tag
P34731	58	57	2	529	
C4YQP9	3	1	2	25.2	

Panel B: Peptide ions of C4YQP9

#	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict
2183	5.9	2	0.718	1529.861	3.14	45	3		5.05E+04	0

Panel E: Conflicting proteins for peptide ion 2183

Accession	Peptides	Unique	Conflict	Protein Score	Pept
P34731	58	57	2	529	5.9
C4YQP9	3	1	2	25.2	5.9

Panel F: Peptide ions of C4YQP9

#	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict
2183	5.9	2	0.718	1529.861	3.14	45	3		5.05E+04	0

Repeat this process until there are no conflicts remaining for the current protein in Panel A.

Now repeat using a similar approach for the next protein in Panel A, here the situation is similar.

Protein resolution results for peptide ion 107:

Accession	Peptides	Unique	Conflict	Protein Score	Peptide Score
P43098	42	3	48	345	7.12
C4YQN7 (+1)	41	2	48	338	7.12
Q5A017	12	1	132	8.69	8.69

Peptide ions of C4YQN7:

#	Σ	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict
107	1	7.12	6	0.932	1661.866	-0.254	52.1	2	1.07E+05	1	
8671	2	0.093	1507.822	5.75	40.2	4	2.47E+04	1	

Resolution of conflicts for this protein

Protein resolution results for peptide ion 107:

Accession	Peptides	Unique	Conflict	Protein Score	Peptide Score
P43098	42	40	2	345	7.12
C4YQN7 (+1)	0	0	0	0	7.12
Q5A017	12	11	1	132	8.69

Peptide ions of Q5A017:

#	Σ	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict
107	1	8.69	9	0.979	1661.866	8.95	52.1	2	1.07E+05	1	
520	0	8.45	8	0.957	1345.749	29.6	67.6	2	1.62E+05	0	

Finally move to the remaining conflict in panel B, and favour the higher scoring protein

Protein resolution results for peptide ion 8671:

Accession	Peptides	Unique	Conflict	Protein Score	Peptide Score
P43098	42	41	1	345	...
Q5A5Q8	13	5	11	116	...
C4YQN7 (+1)	0	0	0	0	...

Peptide ions of Q5A5Q8:

#	Σ	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict
8671	2	0.316	1507.822	29	40.2	4	2.47E+04	1	
4804	1	7.95	2	0.896	831.466	2.55	33.9	2	1.76E+04	1	

Adopting a similar approach to the next protein favouring the protein with the highest score as each conflict is examined.

Protein resolution results for peptide ion 2054:

Accession	Peptides	Unique	Conflict	Protein Score	Peptide Score
P41797	27	11	43	314	8.1
P46587	34	18	41	372	8.1
P10591	15	4	37	153	8.1

Peptide ions of P46587:

#	Σ	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict
2054	0	8.44	6	0.946	1674.74	10.1	33.2	2	4.21E+04	4	
12391	1	7.96	9	0.863	1525.747	8.3	57.5	3	1.36E+04	3	

In this case the first peptide for protein (P41797) has 4 conflicting proteins in panel E. Resolve the conflict in favour of the protein with the higher score (P46587) by unticking the peptide ion in panel B then move on to the next conflicting peptide ion in Panel B (which has 3 conflicts) and resolve the conflict in favour of the protein with the higher score.

Protein: P41797 Heat shock protein SSA1 GN=SSA1 PE=1 SV=2
Protein: P10592 Heat shock protein SSA2 OS=Saccharomyces cerevisiae GN=SSA2 PE=1 SV=3

Peptide ion views Protein resolution

Conflicting proteins for peptide ion 2054

Accession	Peptides	Unique	Conflict:	Protein Score	P
P46587	34	18	26	364	8.4
P41797	27	12	27	314	8.4
P10591	0	0	0	0	8.3
Q5A3Y2	8	7	7	63.6	5.6
P10592	7	7	7	63.6	8.4

Peptide ions of P10592

#	I	Score	Hits	Correlation	Mass	Mass error (p...	RT (mins)	Charge	Tag	Abundance	Conflict:
2054	0	8.18	3	0.940	1674.74	10.1	33.2	2	4.21E+04	1	
8166	4	7.21	2	0.896	960.52	7.04	21.2	2	1.18E+04	1	
34780	0	---	---	0.928	960.535	23.3	21.2	3	1.67E+03	0	
18915	0	6.32	3	0.621	1423.668	7.06	18.2	3	4.16E+03	0	
44755	0	6.32	3	0.570	1423.667	6.58	18.2	4	365	0	

Alternatively you can review the other conflicting proteins. In this case P10591 and P10592 are from a similarly named protein but a different species so you can right click and turn off all the peptides.

Then reviewing the remaining conflicts between P41797 and P46587 they have 19 remaining conflicts. To resolve these in favour of P46587 you can highlight all the remaining conflicted peptides in panel B and click on the tick box to untick them, this will remove them from P41797.

Protein: P41797 Heat shock protein SSA1 GN=SSA1 PE=1 SV=2
Protein: P46587 Heat shock protein SSA2 GN=SSA2 PE=1 SV=3

Peptide ion views Protein resolution

Conflicting proteins for peptide ion 297

Accession	Peptides	Unique	Conflict:	Protein Score	Pept
P46587	34	19	20	373	8.06
P41797	27	12	19	314	8.06

Peptide ions of P46587

#	I	Score	Hits	Correlation	Mass	Mass error (p...	RT (mins)	Charge	Tag	Abundance	Conflict:
297	1	8.06	4	0.961	1680.838	16	52.8	2	1.61E+05	1	
30364	0	8.63	8	0.703	2166.023	2.92	59.7	2	3.02E+03	1	
1632	0	8.59	7	0.921	1590.767	13	45.8	2	7.77E+04	1	
1419	1	8.64	8	0.925	1787.006	12.9	52.3	3	6.21E+04	1	
17676	0	8.58	5	0.546	1786.989	3.48	52.3	2	4.37E+03	1	

The display will update to show one remaining conflict for P46587. Click on this protein in panel A.

Protein: P46587 Heat shock protein SSA2 GN=SSA2 PE=1 SV=3
Protein: Q5ADM7 Glyceraldehyde 3-phosphate dehydrogenase GN=TDH3 PE=3 SV=1

Peptide ion views Protein resolution

Conflicting proteins for peptide ion 186

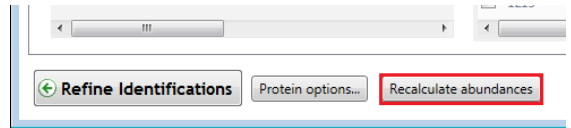
Accession	Peptides	Unique	Conflict:	Protein Score	Pept
P46587	34	33	1	363	6.58
Q5ADM7	20	19	1	247	8.45

Peptide ions of Q5ADM7

#	I	Score	Hits	Correlation	Mass	Mass error (p...	RT (mins)	Charge	Tag	Abundance	Conflict:
186	7	8.45	5	0.884	1292.682	1.38	49.9	2	3.01E+05	1	
3200	0	9.12	9	0.873	2885.364	5.15	76.6	2	4.07E+04	0	
6	0	8.02	8	0.969	2889.489	1.97	63.9	3	2.4E+05	0	
18769	0	8.02	7	0.757	2889.489	2.08	63.9	4	8.77E+03	0	
39	1	8.73	8	0.979	1765.824	16.8	54.3	2	3.06E+05	0	

The strategies that can be used to resolve conflicts can include differences in Mass error as well. However, if there is no difference for a given peptide then either resolution in favour of the protein with the highest protein score or unticking the peptide for both conflicting proteins is an option.

Note: the abundances will need to be recalculated as a result of performing Conflict resolution. This is achieved by clicking on the Recalculate abundances, which appears during Conflict resolution

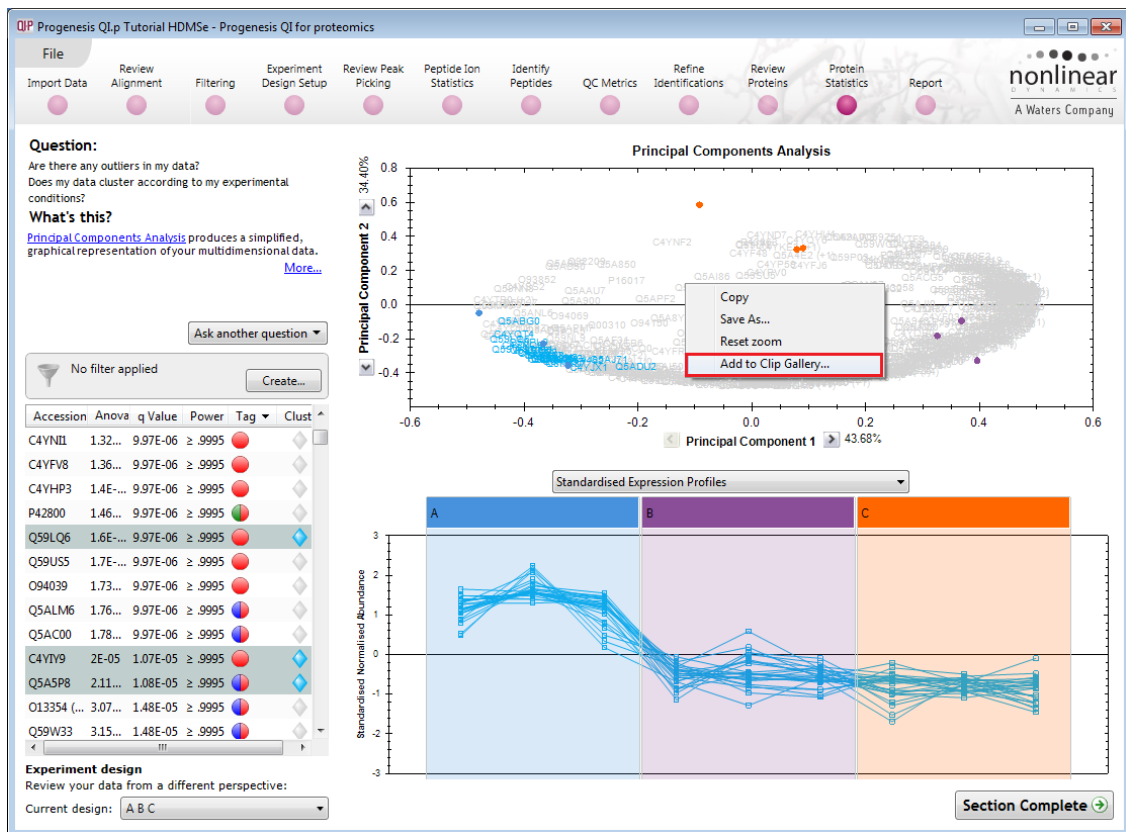


Appendix 8: Using Clip Gallery to Save and Export Pictures and Data

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

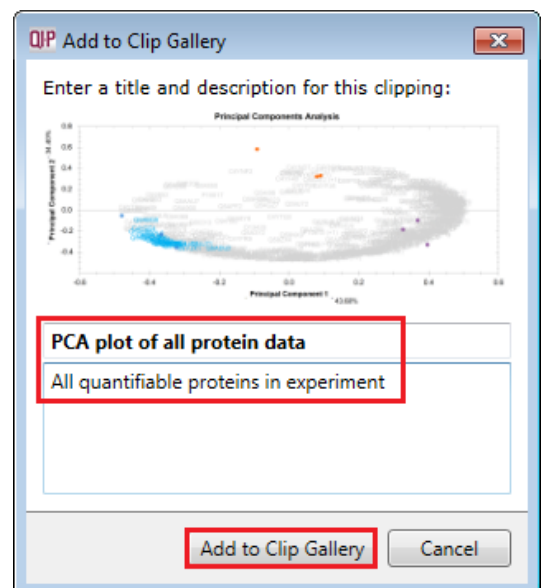
The saved images of the Views and the tables are retained as part of the experiment and are stored accordingly. This facility allows you to capture (high resolution) images that can be used in the development of specific reports and/or used as part of the process of publishing your experimental findings.

As an example of using the Clip Gallery, at the **Protein Statistics** view while displaying the PCA plot right click on the **Biplot View** and select **Add to Clip Gallery...**

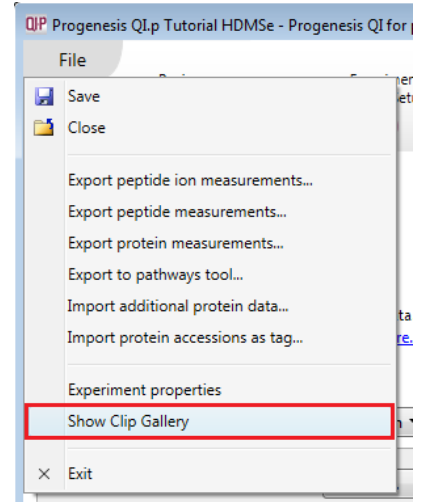


This will open a dialog displaying what is to be saved and allows you to alter the title and provides a description of the image for later reference.

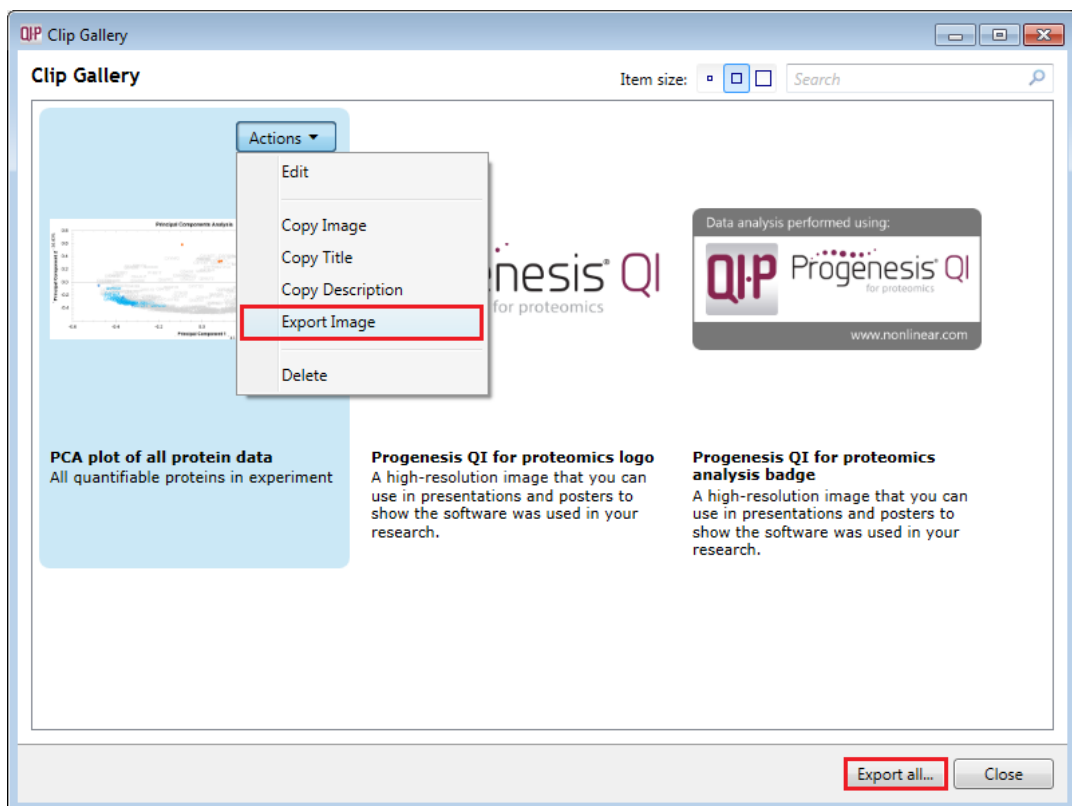
Enter details as required and click **Add to Clip Gallery**



To view, edit and/or export from the clip gallery 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: Exporting protein identities to IPA

IPA (Ingenuity Pathway Analysis): is a commercial web-based software application for the analysis, integration and interpretation of data derived from any experiments that generate gene and chemical lists with expression data, such as metabolomics and proteomics.

Before exporting data to IPA, make sure that you can launch IPA using the default browser on your system. The plug-in relies on the browser to upload the Progenesis-exported data to IPA.

Note: the use of IPA requires a licensed copy, for licensing details please visit the [IPA website](#).

First select the Protein data to export to the pathways tool using tag filtering to 'focus' the set to export.

Click **Export to pathways tool** and select IPA and click **Next**.

The screenshot displays the Progenesis QI software interface during the 'Review Proteins' step. A dialog box titled 'QIP Export Pathways Information' is open, prompting the user to 'Select a pathways tool'. The dropdown menu shows 'IPA: Ingenuity Pathway Analysis' selected. The background interface includes a navigation menu, a table of protein data, and a graph of ArcSinh Normalised Abundance.

QIP Export Pathways Information

Select a pathways tool
Choose a pathways tool from the list below. You can find out more or download new plugins using the links below.

Which pathways tool do you want to use?
 IPA: Ingenuity Pathway Analysis
[About this plugin](#) | [Download other plugins](#)

< Back Next > Cancel

SV=1

ArcSinh Normalised Abundance

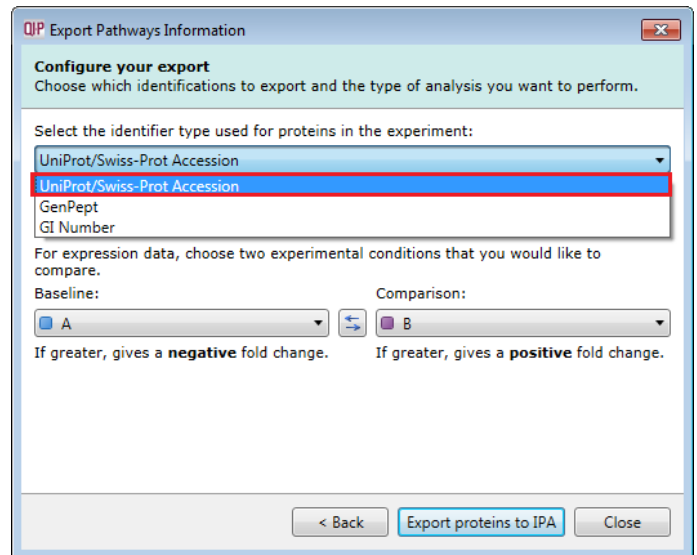
Quantifiable proteins displayed: 512

Section Complete

Select the appropriate identifier type used for proteins in the experiment.

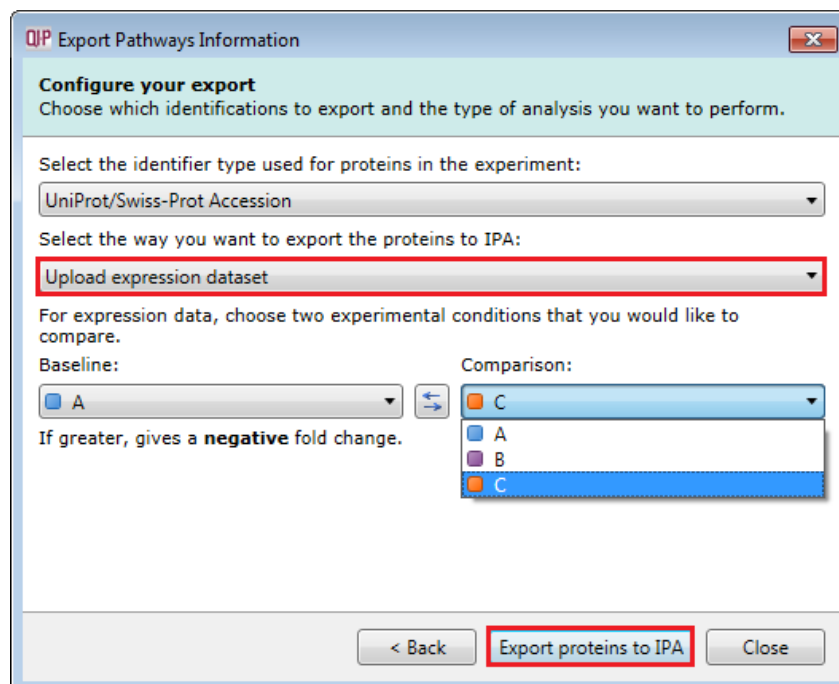
Note: for proteins with different identifier types, filter the protein list to contain only compounds with single identifier type at a time, then perform the export operation and merge the resulting lists in IPA.

Finally select the way you want to export the compounds to IPA.



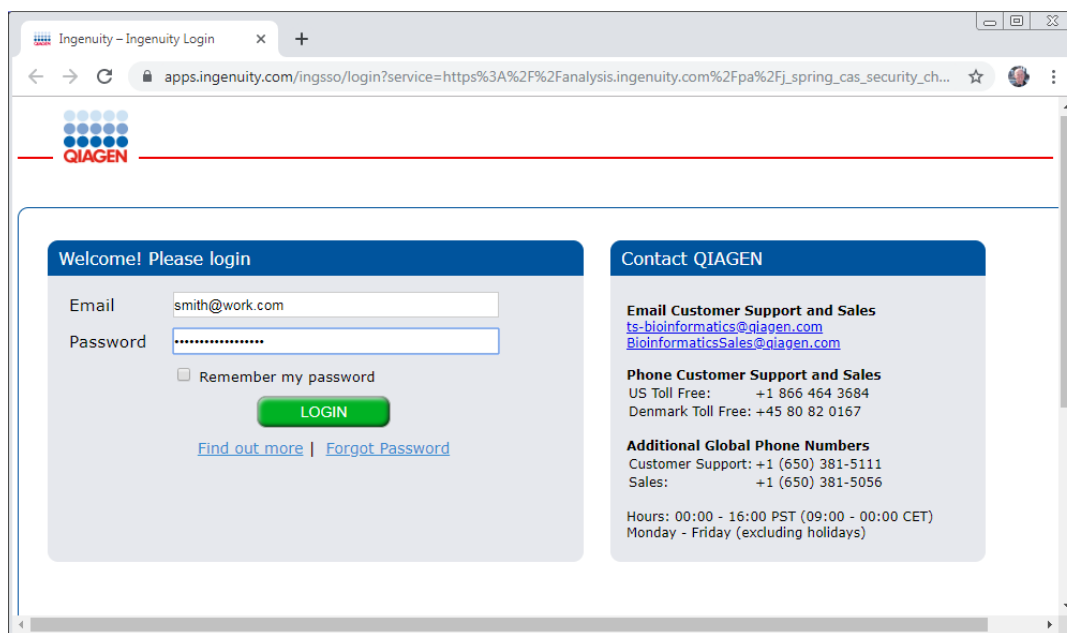
If you are performing over-representation analysis or other types of analysis that do not require the protein expression data then you can make use of the **Create a list** option to export data to IPA.

If you require to perform enrichment analysis or other types of analyses that use protein expression data, select the option to **Upload expression dataset** to IPA.



Note: You need to select the type of protein identifier used by at least one protein whose expression data you want to export. If there are proteins with different identifier types, you will be able to select additional identifier types in IPA after the dataset is uploaded.

The operation exports data from a single observation at a time - select the two experimental conditions (taken from the current **experiment design as set in Progenesis QI for proteomics** you wish to compare as that observation).

Click **Export compounds to IPA**.

If you have access to a licenced copy of IPA then log in

IPA will open displaying the imported data from Progenesis QI for Proteomics.

ID	Observation 1	Ignore	Ignore	Ignore	Ignore
1	P34731	-1.00902155389087	---	---	---
2	C4YQR7	1.03618874042727	---	---	---
3	C4YMC3	-1.1630436093987	---	---	---
4	C4YQN7	-1.14248088552437	---	---	---
5	P43098	-1.00167260838215	---	---	---
6	C4YR46	1.11144195146305	---	---	---
7	O13430	1.0005161363124	---	---	---
8	P46587	1.16697510036399	---	---	---
9	P82610	1.14143675382931	---	---	---
10	P46598	1.06004923965961	---	---	---
11	Q96V89	-1.10613636924535	---	---	---
12	P41797	1.50747573107542	---	---	---
13	P28877	-1.20624862059398	---	---	---
14	C4YK39	1.05494661189734	---	---	---
15	Q59KZ1	1.03999813845519	---	---	---
16	C4YL05	-1.0312270752554	---	---	---
17	P46273	1.31165982609058	---	---	---
18	C4YL8	1.26909493012736	---	---	---
19	O94039	1.55325257659893	---	---	---
20	O13287	1.38535524141575	---	---	---

You can now explore your protein expression data using the tools available in IPA. Instructions on how to manage and explore your imported data set are provided by IPA.

Appendix 10: Exporting protein identities to Metacore

Metacore: is a commercial web-based software application for the analysis, integration and interpretation of data derived from any experiments that generate gene and chemical lists with expression data, such as metabolomics and proteomics.

Before exporting data to MetaCore, make sure that you can launch MetaCore using the default browser on your system. The plug-in relies on the browser to upload the Progenesis-exported data to MetaCore.

First select the Protein data to export to the pathways tool using tag filtering to 'focus' the set to export.

Click **Export to pathways tool** and select MetaCore

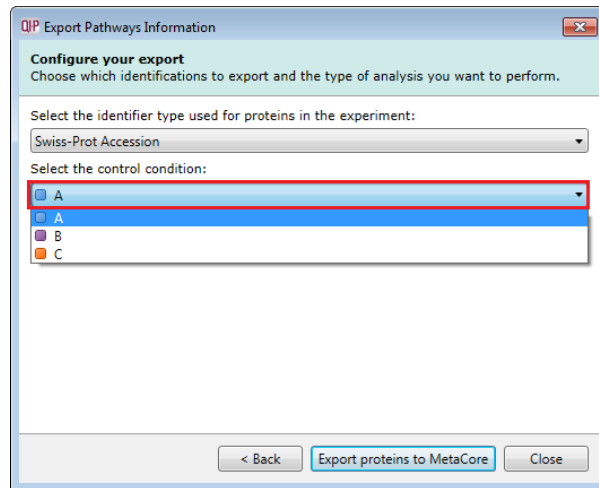
The screenshot shows the Progenesis QI software interface. The 'Review Proteins' section is active, and the 'Export to pathways tool' button is highlighted with a red box. A dialog box titled 'QIP Export Pathways Information' is open, showing a list of pathways tools with 'MetaCore' selected. The background shows a table of protein data with columns for Accession, Peptides, Unique peptides, Confidence score, Anova (p), q Value, Tag, Max fold change, Highest Mean, and Lowest Me. The 'Tag' column is set to 'C'.

Then click **Next** to Configure your export page .

Select the appropriate identifier type used for proteins in the experiment.

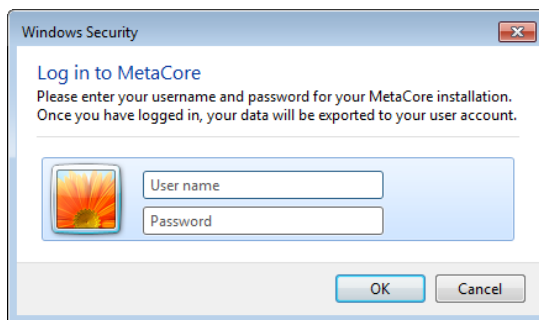
The screenshot shows the 'QIP Export Pathways Information' dialog box, specifically the 'Configure your export' section. The 'Select the identifier type used for proteins in the experiment:' dropdown menu is open, showing options: Swiss-Prot Accession, GI Number, ENSEMBL, and EMBL. 'Swiss-Prot Accession' is selected and highlighted with a red box. The 'Export proteins to MetaCore' button is highlighted with a blue box.

If you have multiple conditions in your current experiment design you can select which condition should be used as the control



Finally click **Export proteins to MetaCore**.

If you have not already done so, you will be prompted to enter your MetaCore *User name* and *Password*; once you have done this, your data will be exported to your MetaCore account, and the MetaCore user interface will open in your default web-browser.



Metacore will open displaying the imported data from Progenesis QI for Proteomics.

You can now explore your protein expression data using the tools available in MetaCore. Instructions on how to manage and explore your imported data set are provided by MetaCore.

Appendix 11: Waters Machine Specification

This appendix provides information on the approximate time(s) taken at each stage and the total time taken to analyse a set of 9 (Phase 1) HDMSe runs on a Waters Demo Spec PC.

Machine Spec: Lenovo

Processor: Intel® Xeon® CPU E5-1630 v3 @ 3.70GHz 12core K40c GPU card

RAM: 64.0 GB

System Type: 64-bit Operating System

File Folder Size: Each file folder (.RAW): 40.9 Gig

Analysis Stages:		Per file	Total	
Import Data:	Loading of Raw data per file	10min	1hr 12min	for 9 files
	Apex Background processing (re-opening at Import Data)	18min (max)	2hr 42min 20s	for 9 files
Alignment:	Automatic alignment of data (re-opening at Alignment)		3min 45s 10s	
	Peak Detection: Automatic Detection of data (re-opening at Peak Detection)		9min 25s 10s	
Identify Peptides:	Performing MS ^E Search (re-opening at Identify Peptides)		13min 15s 10s	
	Total Analysis Time: Excluding Background Apex Processing		1hr 38min	
	Including Apex processing		2hr 51min	
Restoring:	Tutorial Archive		3.5min	

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