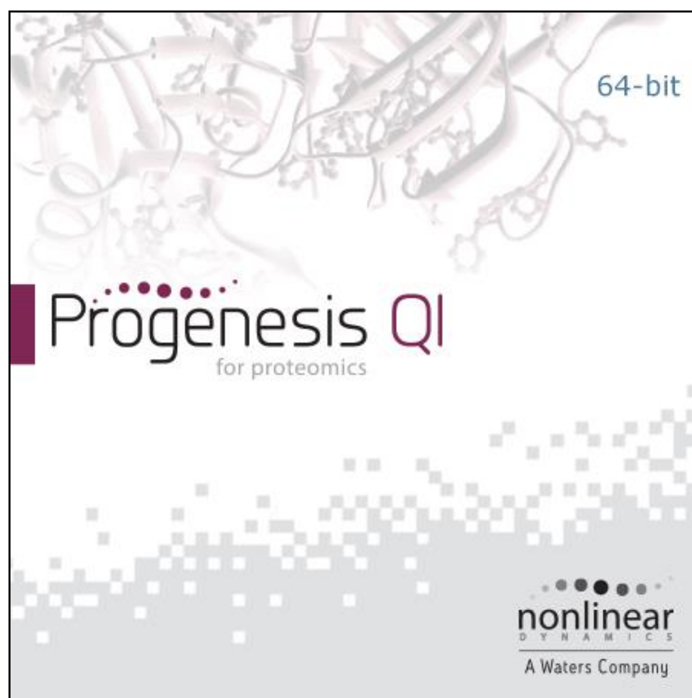


Progenesis QI for proteomics
Spectral Library
User Guide

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
Tutorial Data 4
Restoring the Tutorial 5
Stage 1: Perform Automatic Analysis of the Library Samples 6
Stage 2: Automatic Processing of your data 7
Stage 3: Filtering 12
Stage 4: Experiment Design Setup for Analysed Runs 13
Stage 5: Identify peptides 14
Stage 6: Refine Identifications 15
Stage 7 Resolving Conflicts 16
Stage 8: Creating a Spectral Library 17
Stage 9: Identifying proteins using a spectral library 22
Additional Tools for Spectral Library Construction 29
Appendix 1: Stage 1 Data Import and QC review of LC-MS data set 34
Appendix 2: Stage 2 Processing failures 39
Appendix 3: Licensing runs 41
Appendix 4: Manual assistance of Alignment 42
Appendix 5: Within-subject Design 47
Appendix 6: Resolve Conflicts 49
Protein Quantitation options 51
Appendix 7: Using Clip Gallery to Save and Export Pictures and Data 53

Introduction

This user guide takes you through how to construct a Spectral Library from the analysis of a set of 6 LC-MS runs with 2 groups (3 replicate runs per group) using the unique Progenesis QI for Proteomics workflow. It describes the initial analysis of the data followed by the creation and use of a spectral library to identify proteins in 2 sets of additional samples. More detailed descriptions of each step in the analysis workflow is described in the main User guides.

To allow ease of use the tutorial is designed to start with the restoration of Archived experiments 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. More details of the samples and the proteins present in them are available on page 4.

If you are using your own data files please refer to Appendix 1 (page 34) 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^E parameters were set to 500:30 instead of the default settings as defined in Appendix 1 (page 34). This was to done to reduce the time taken to demo the data analysis.

Tutorial Data

The data files for this tutorial were kindly provided by Catalin Doneanu at Waters Corporation, Milford MA USA . They were originally generated to examine the relative and absolute quantitation of proteins spike at different amounts into samples of a monoclonal antibody digest (Infliximab). The idea being to simulate differential contamination of the mAb preparation and therefore to monitor Host Cell Proteins with high accuracy using mass spectrometry.

There are 2 spiking amounts employed (fmol/fmol), Refered to as Sample A and Sample B:

Spiked protein	Sample A (fmol)	Sample B (fmol)	Fold change
Alcohol Dehydrogenase (ADH)	100	200	2
Bovine Serum Albumin (BSA)	200	20	10
Cytochrome C (Cyt C)	50	10	5
Enolase (ENL)	250	250	1
Gycogen Phosphorylase b (PHO)	50	500	10

These digested protein standards were added, in the amounts outlined above, to the same amount of the Infliximab digest and label free LC-MS was performed using a 30min gadient on a CSH C18 (M-class ACQUITY UPLC) attached to a Synapt G2-Si.

Six of the samples (3 each of Samples A and B) can be processed to generate spectral libraries for the Infliximab product only and also for the spiked proteins, Infliximab and any remaining contaminants of Infliximab (Host cell proteins). The remaining 4 samples are treated as example Batches of the Infliximab purified product with different levels of contamination.

The idea of the data set is to allow you to explore the processes involved in making spectral libraries and then their use in the identification and quanification of Host Cell Proteins using known protein standards to simulate different levels of contamination.

The raw data has already been loaded into Progenesis QI for proteomics 4.0 in the form of 2 experiments and licenced archives of these experiments have been created:

Spectral Library_create.ProgenesisQIPArchive

Batches.ProgenesisQIPArchive.

Note: these will open at the Import data stage when restored in Progenesis QI for proteomics

A fasta file, **Library_mouse.fas,** is provided for identification of the proteins; containing sequences for mouse proteins, Infliximab and the spiked proteins.

Note: there are also 2 experiment design setup .csv files and 2 example Spectral Libraries .msp available.

The files described above are all available in the Spectral Library Tutorial.zip downloaded from nonlinear.com.

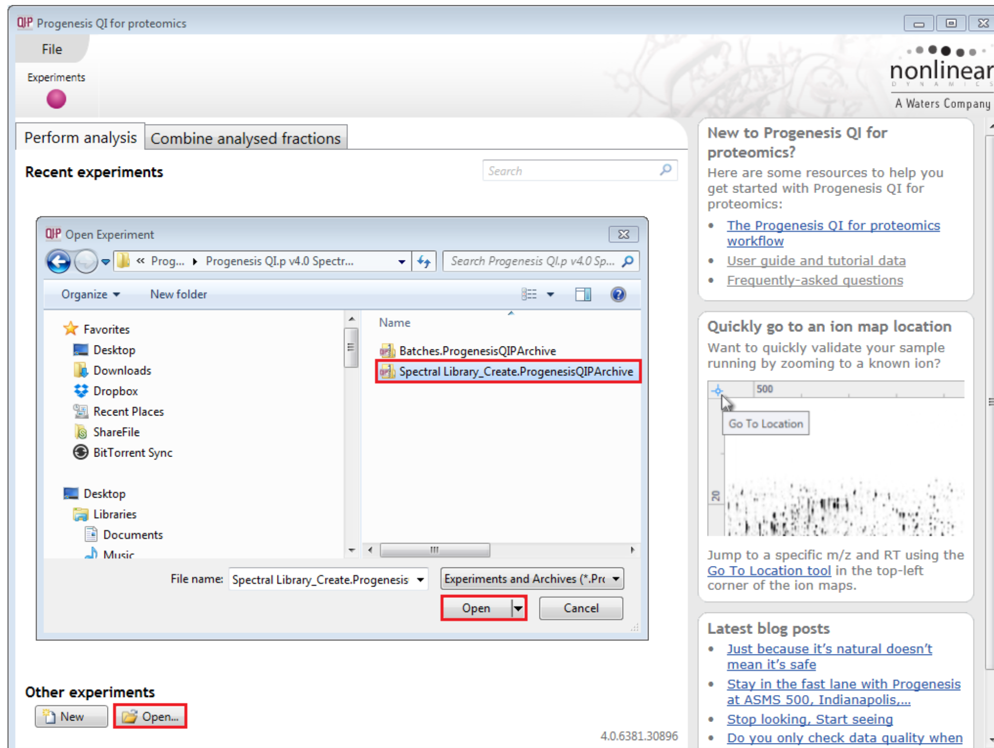
The following sections explain how to restore and process the data in Progenesis QI for proteomics v4.0.

Restoring the Tutorial

Open Progenesis QI for proteomics and download the Compressed (.zip), Spectral Library Tutorial 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.

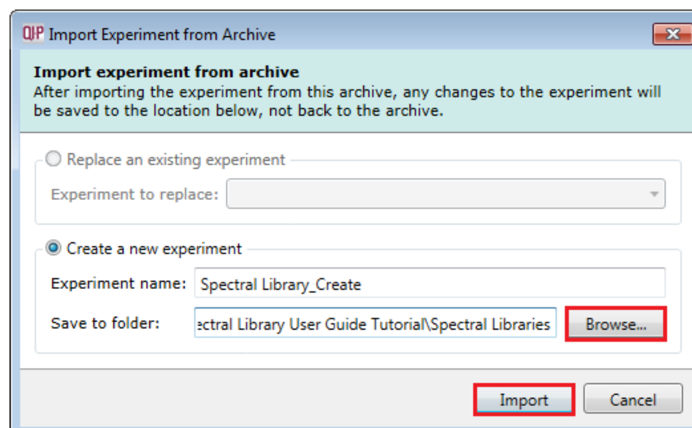
There are 2 archive files, 2 Grouping files, a fasta file and example Library files.

Now restore the 2 archive files (Spectral Library_Create and Batches). To do this, locate the '**name'.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.



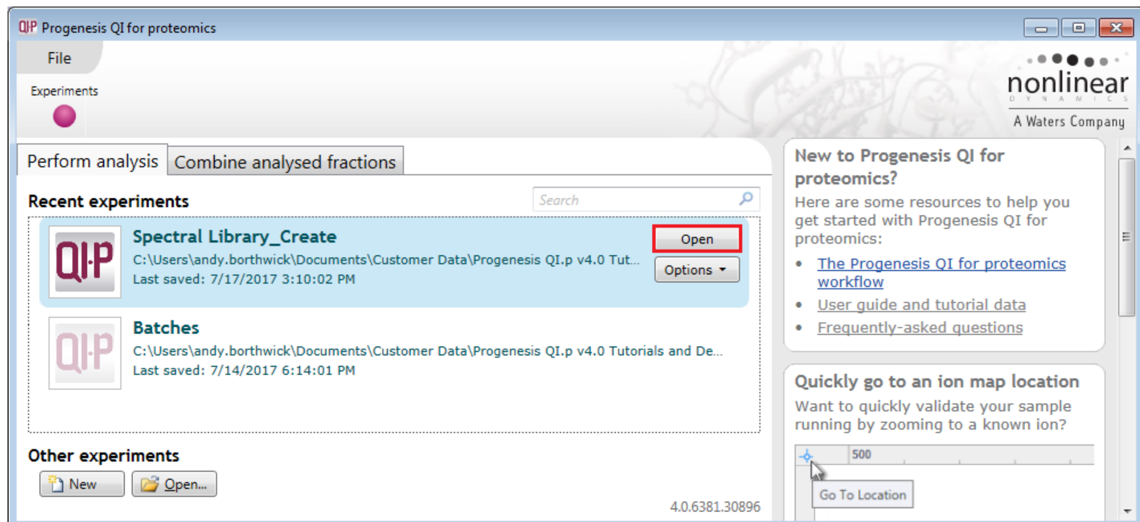
Then press **Import.** After the import is complete select **Close** from the **File menu** and then repeat the process for the other archive (Batches), restoring them into the same folder.

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

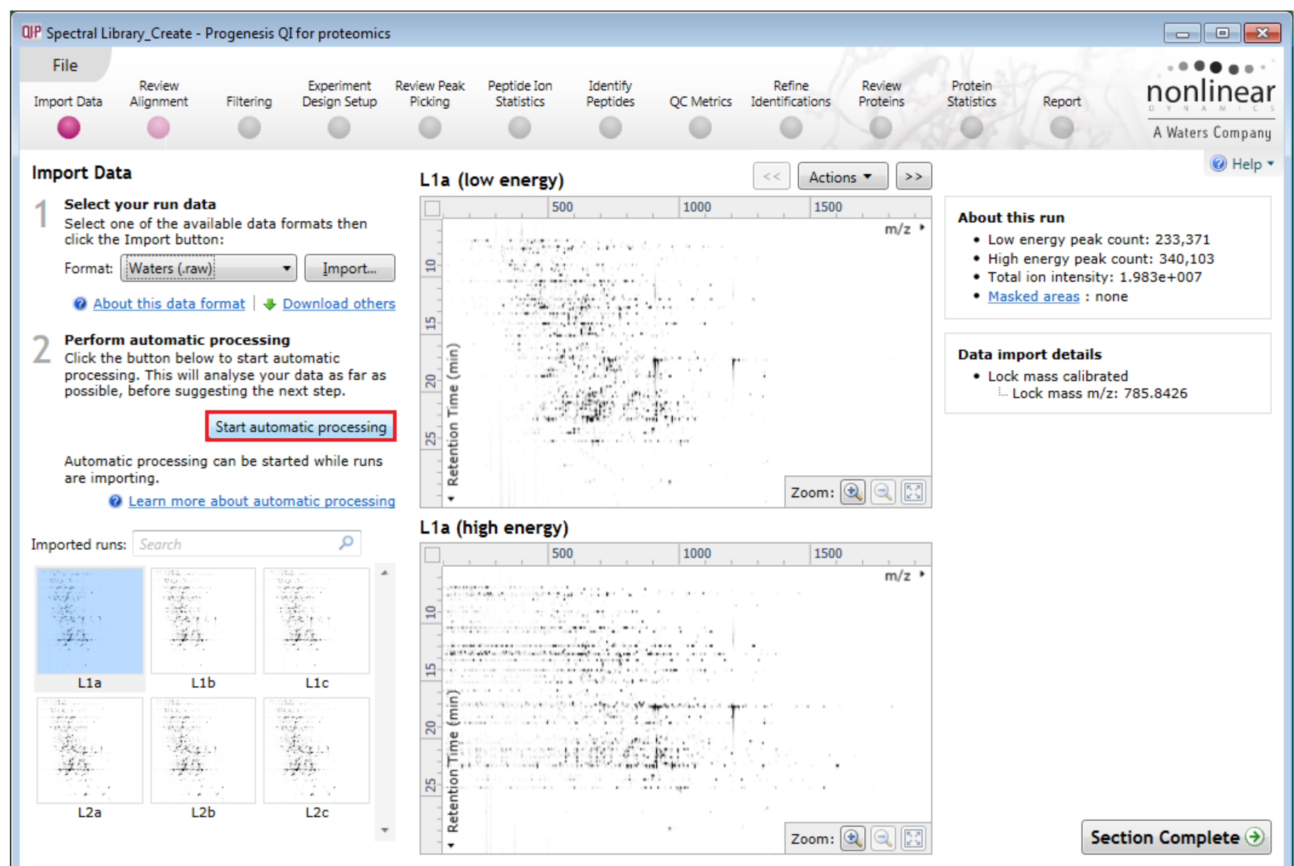
Tip: at each stage in the software there are links to more information and help on the website.

Stage 1: Perform Automatic Analysis of the Library Samples

Having opened the 2 Progenesis archive files the experiments will now be displayed when you open Progenesis QI for proteomics. Select the Spectral Library_Create experiment and click **Open**.



To start the Auto processing of the six runs in this experiment click **Start automatic processing**



The Auto processing wizard opens and a series of dialogs allow you to select how you want the runs to be analysed in Progenesis.

The following pages describe how to proceed with the analysis of this data. Additional details on using the wizard and fuller description of all the options is available in Appendix 1 (page 34) and from the main HDMSe User guide.

Stage 2: Automatic Processing of your data

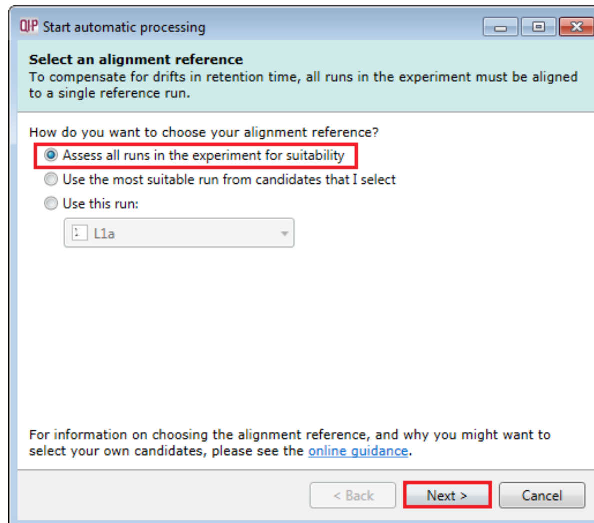
The Automatic Processing of your data can be set up and started before the import of your data has been completed by clicking on **Start automatic processing.**

2 **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.
Start automatic processing
Automatic processing can be started while runs are importing.
Learn more about automatic processing

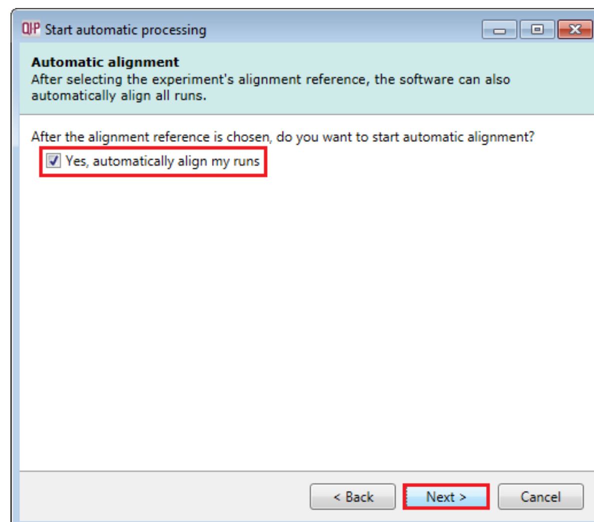
Note: for this tutorial the data has been imported already.

For the processing of the 6 files in the Spectral library_create experiment make the follow selections:

Step 1: Select the first option, **Assess all runs in the experiment for suitability**. This tells QIp to manage the selection of the Alignment reference for alignment.

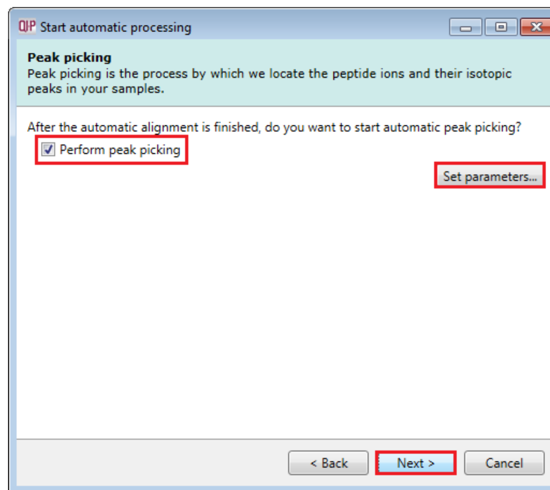


Click **Next**. The option to perform Automatic alignment will be selected by default.



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

Step 2: For the purposes of this User guide we will use the default settings for peak picking which you can review by clicking on Set parameters.

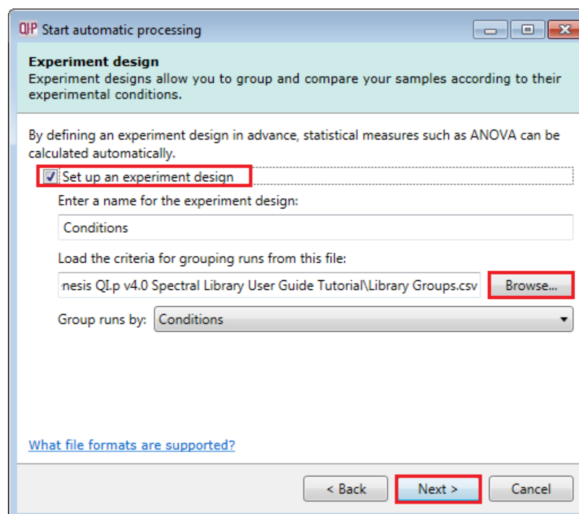


Note: for more details on setting Peak Picking parameters refer to the section on Filtering (page 12)

Click **Next**.

Step 3: To apply the experiment design in advance Click **Set up an experiment design** then locate the appropriate file (**Library Groups.csv** available in the folder you extracted the .zip file) using Browse.

Note: if you do not enter a name for the experiment design it will adopt the name of the first column in the Library Groups.csv, in this case Conditions, this can be changed as required in the main workflow.



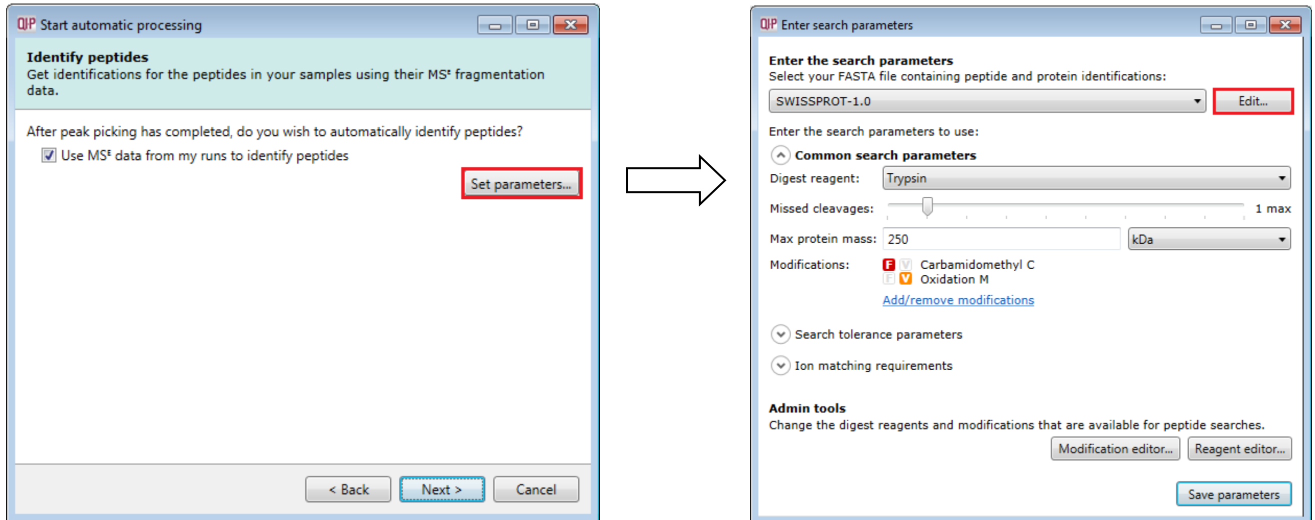
Note: you can also create additional experiment designs following the completion of automatic processing.

Select Conditions and then click **Next**.

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

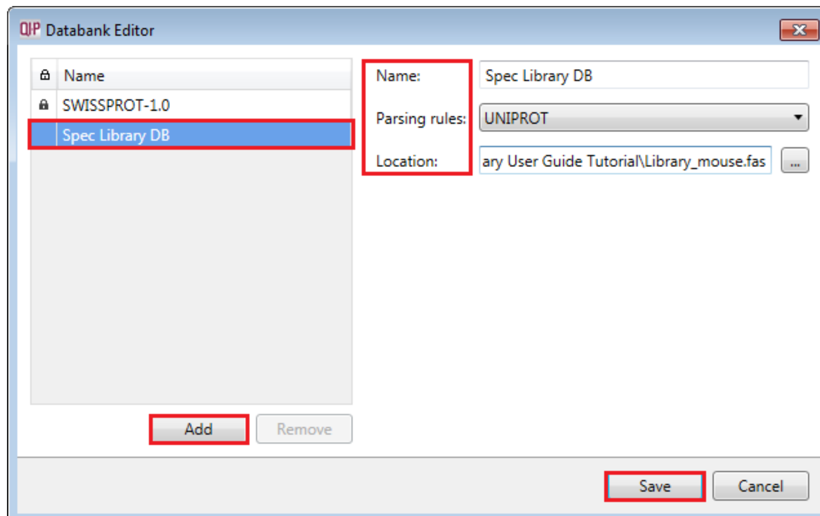
Step 4: To automatically process the identification of peptides using there MS^E fragmentation data in the tutorial data click on **Set parameters**. The default Databank is for Swissprot-1.0 (which is a locked example).

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

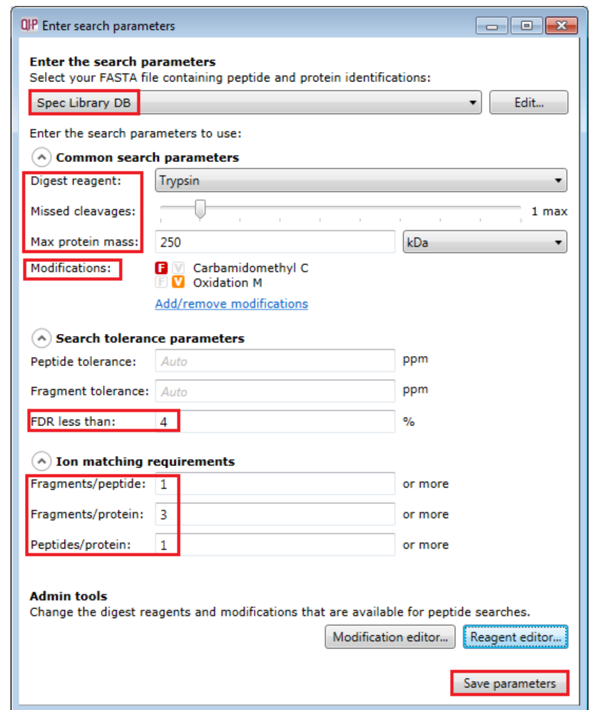


To create a new Databank from a Fasta file click on **Edit** and then create a new one using the example FASTA file that is available with the Experiment Archive you restored at the beginning of this tutorial exercise (**Library_mouse.fas**).

For a new Databank click **Add.** Then give it name (i.e. Spec Library 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.

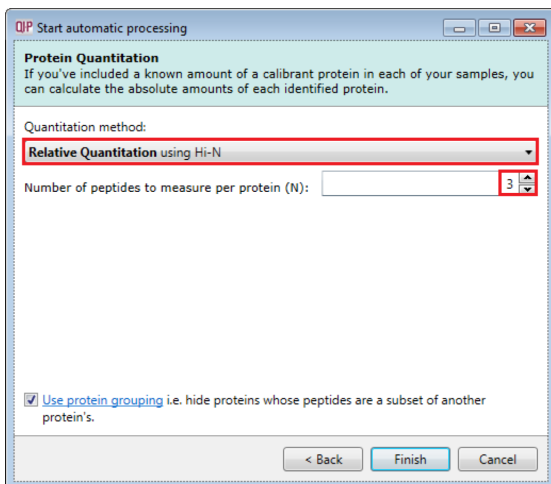
Maximum protein mass: is set at 250kDa

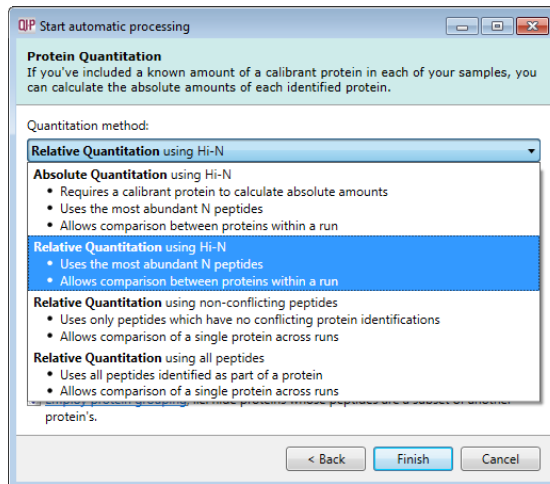
Modifications: are set Carbamidomethyl C (Fixed) and Oxidation M (Variable). More modifications are available from the list and additional ones can be added to the list using the **Modification editor…**

Search Tolerance parameters:

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

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

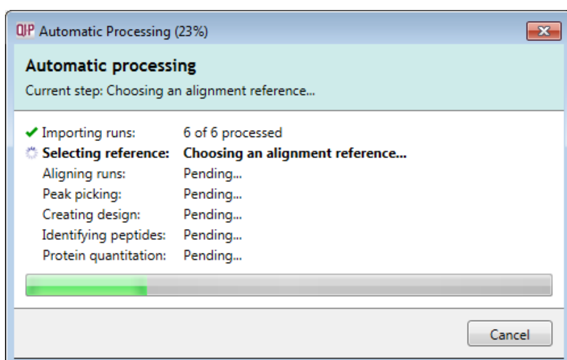


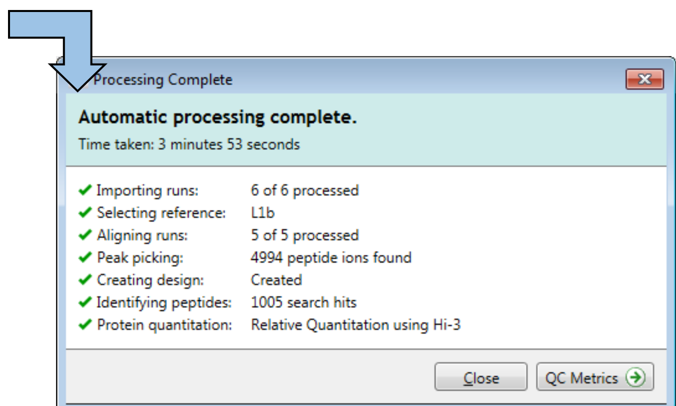


The **Protein Quantitation** dialog opens displaying the default method, **Relative Quantitation using Hi-N** which uses up to 3 peptides per protein to compute the relative amount of each 'identified' protein. Use the drop down to reveal the alternative methods for protein quantitation

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

The process starts with the selection of an alignment reference and completes with Protein quantitation.



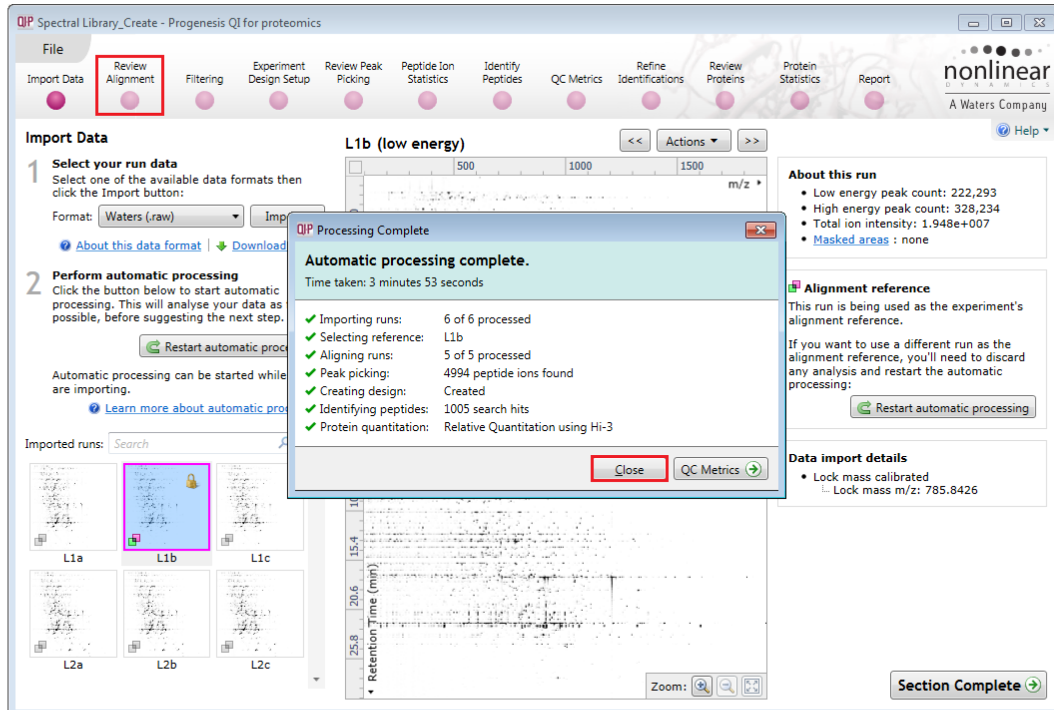


As the whole process proceeds you get information on what stage has been performed and also the % of the process that has been completed.

Note: if processing fails to complete successfully there are a number of suggested strategies you can use to proceed with your analysis. These are out lined in Appendix 2 (page 39)

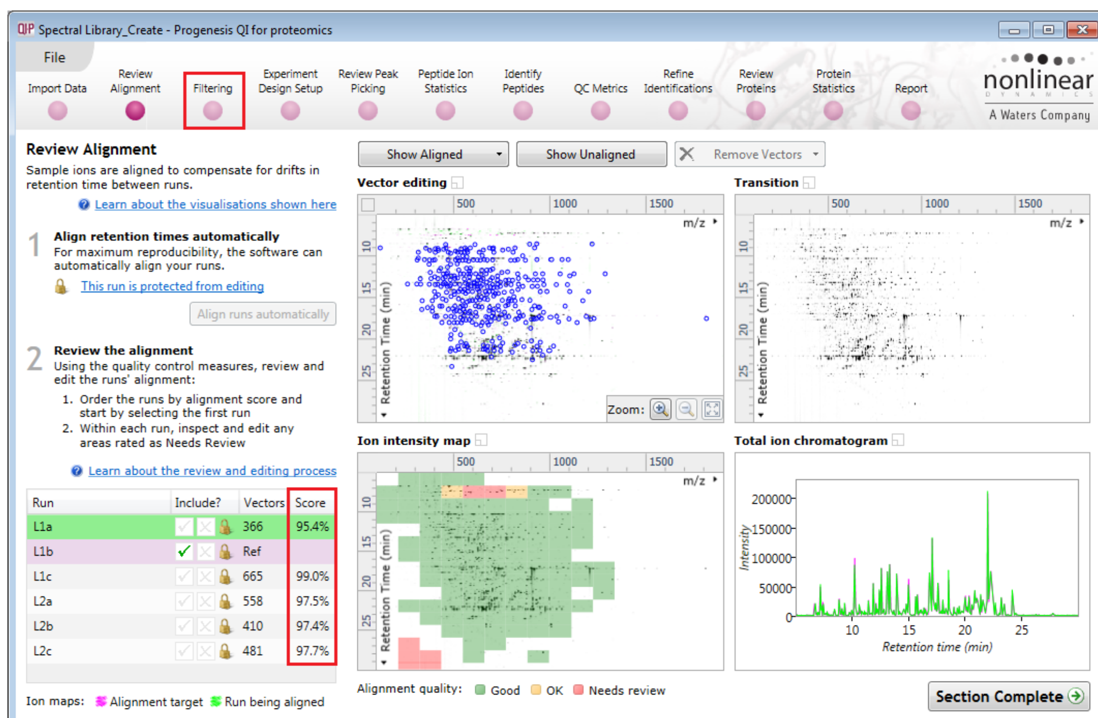
Note: this does not include the time to load the data

When Processing completes click on **Close** and the on **Review Alignment** in the Workflow



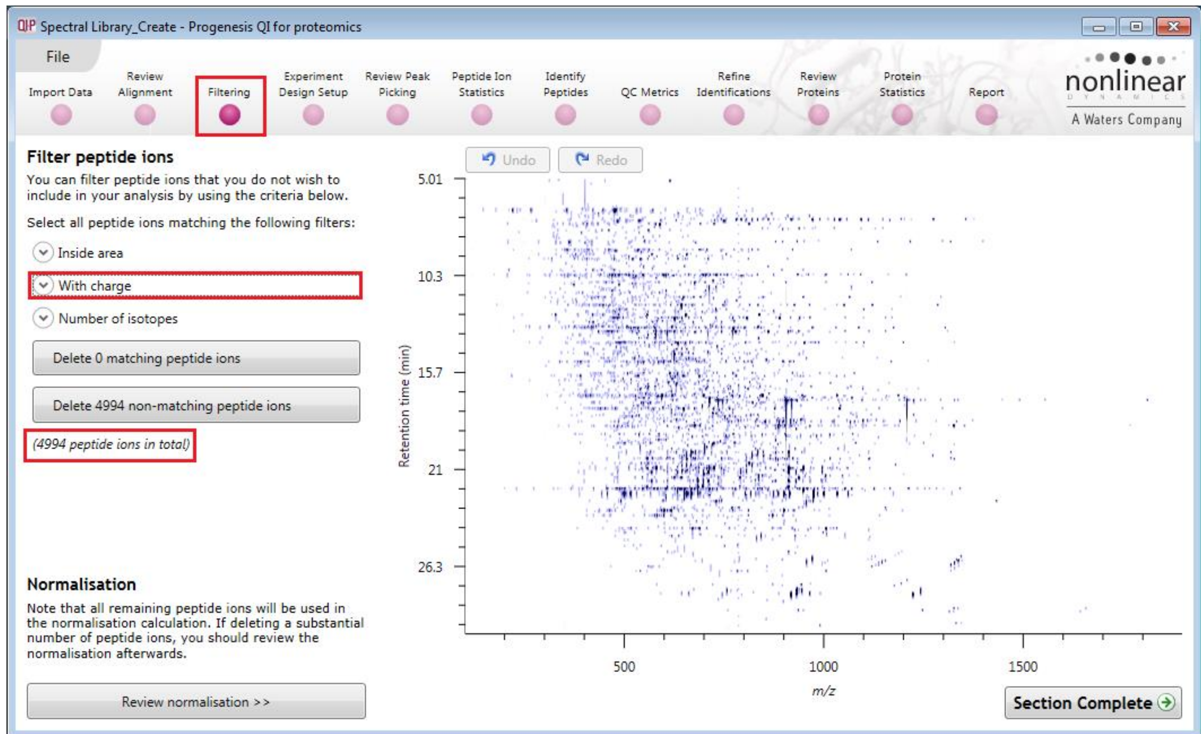
You can explore and re-perform the steps, sequentially and/or as part of the automatic processing as described in this guide.

For this data set, the quality of the alignment (as indicated by the high scores) and detection following automatic processing does not require to be re-performed.

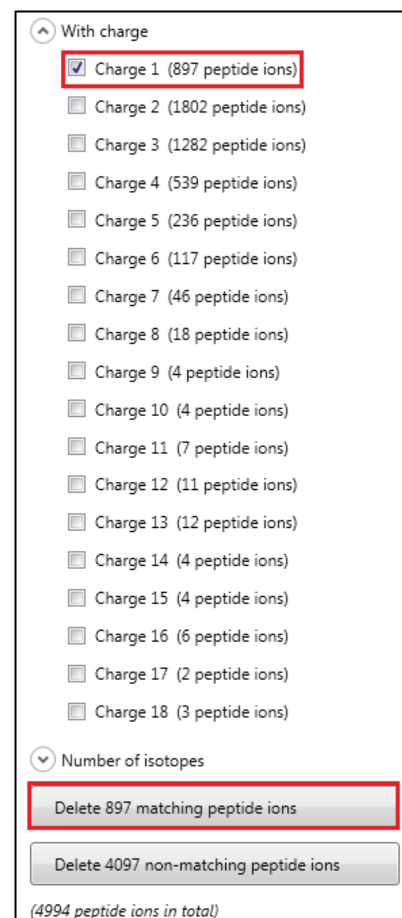


Stage 3: Filtering

At Filtering you can review the total number of peptide ions detected (4994) and choose to keep all or those which meet certain filtering criteria.



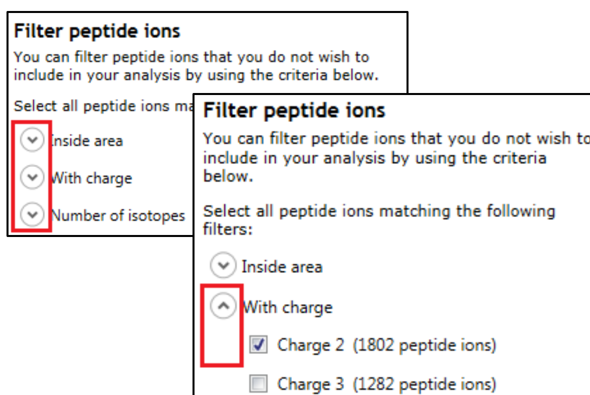
For this particular data set we will reduce the peptide ions to having a 'charge state' of 2 and above by filtering the Charge state 1 peptide ions.



Select **With charge** and tick the Peptide ions you wish to remove from the analysis.

Following deletion of the peptide ions that do not meet the criteria this leaves 4097 peptide ions with a range of charge states 2 to 20.

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



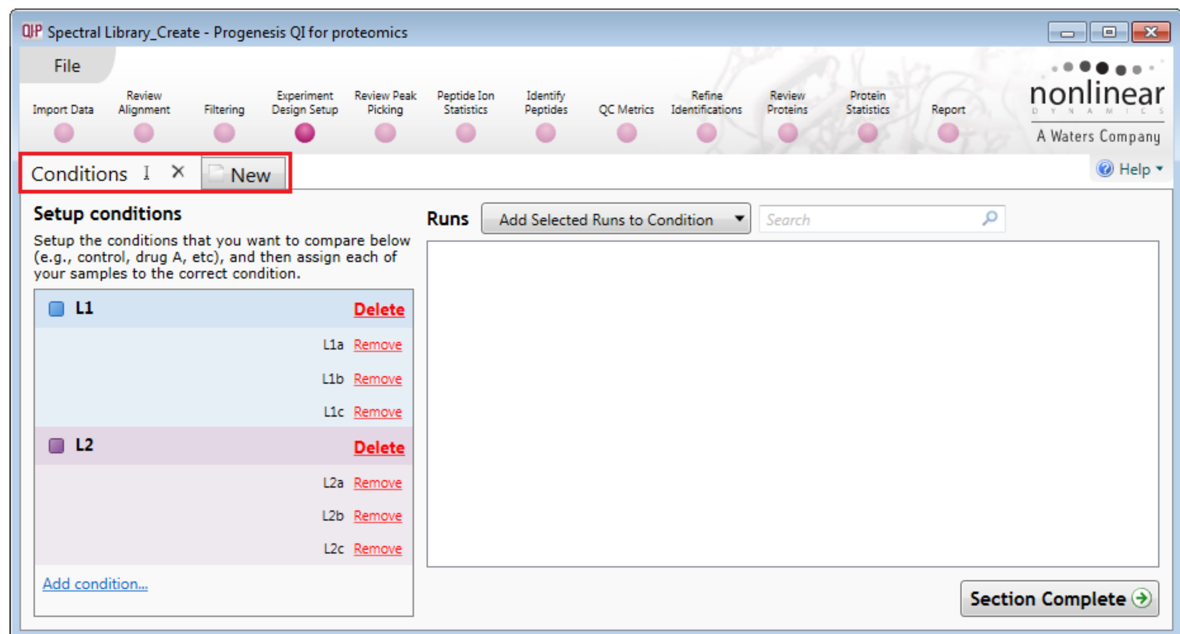
Now move on to **Experiment design**

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Stage 4: Experiment Design Setup for Analysed Runs

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

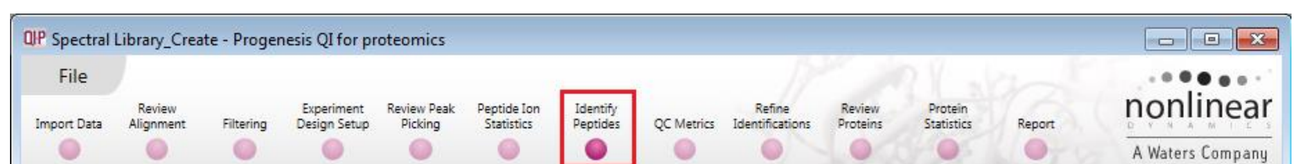
Currently the Between subject experiment design has been set during the Automatic processing of the data, with 2 conditions: L1 and L2.



Note: you can use the New tab to create another Experiment design as required .

Details on the use of the next 2 steps in the Workflow: **Review Peak Picking** and **Peptide Ion Statistics** are available in the main HDMSe User guide.

For the purposes of this Tutorial we will move directly to the **Identify Peptides** stage by clicking on the workflow



Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Stage 5: Identify peptides

Progenesis QI for proteomics is designed to perform peptide identifications either directly or by allowing you to export MS/MS spectra 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.

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

Note: Following the automatic processing, described in Stage 2 (page 7) of this guide, the Identify Peptides page currently displays the full list of the detected peptide ions in your experiment.

The left hand panel will display the Fasta File used (Spec Library DB) in the search and the parameters and settings used to control the search.

Stage 6: Refine Identifications

Before attempting to create a Spectral library from this data one should refine the quality of the search results by filtering out peptide identifications base on the score and number of hits.

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

QIP Spectral Library_Create - 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 DYNAMICS A Waters Company

Refine Identifications

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

1 **Specify a set of deletion criteria**
In the batch deletion criteria, enter the property values for a set of identifications you want to delete.

2 **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

3 **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.

No filter applied — Create...

#	Total Hits	m/z	RT (min	Charge	Tag
1	---	1207.2	17.11	3	
3	2	884.08	18.52	3	
7	3	899.46	23.21	2	
17	3	1325.6	18.53	2	
18	3	973.53	22.29	2	
23	4	937.47	21.38	2	
24	2	649.36	22.26	3	
33	2	938.47	12.60	2	
36	2	1005.4	24.17	2	
42	3	701.38	19.75	3	
47	3	632.69	22.02	3	

Resolve conflicts >>

Batch deletion criteria

Score	less than	4	Sequence Length	less than	
Hits	less than		Charge	less than	
Mass	less than		Sequence	contains	
Absolute mass error (ppm)	less than		Accession	contains	
m/z	less than		Description	contains	
Retention Time	less than		Modifications	contains	

Delete matching search results | Delete non-matching search results | Reset the criteria

#	Score	Hits	m/z	RT(mins)	Charge	Mass	Mass err	Sequence	Accession
2261	0.00	1	812.92	17.26	2	1623.8	4.88	+QSPEKGLEWVAEIR	123456
2287	0.00	1	598.80	23.19	2	1195.5	4.17	CLLNNFYPR	123456
2294	5.28	1	646.31	9.39	2	1290.6	3.72	ECCDKPLLEK	P02769
2296	6.09	2	367.22	10.73	2	732.43	12.22	NVPLYK	P00924
2308	5.14	1	633.31	17.10	4	2529.2	5.78	VSFSCRASQFVGSSIHWYQQR	123456
2316	5.28	1	431.21	9.40	3	1290.6	13.19	ECCDKPLLEK	P02769
2339	7.03	1	517.60	19.46	3	1549.7	7.37	IGEEYISDLDQLR	P00489
2357	6.25	1	348.70	6.74	2	695.39	20.09	DHLVGR	P00489
2375	5.99	2	546.82	15.75	2	1091.6	7.45	EDLIAYLKK	P62894
2382	7.12	1	435.92	14.02	3	1304.7	16.81	HLVDEPQNLIK	P02769
2383	0.00	2	495.26	10.40	2	988.50	6.68	SEVYHNLK	P00924
2402	6.60	3	503.75	8.22	2	1005.4	6.89	FGCRDPVR	P00489
2414	0.00	1	826.88	24.18	2	1651.7	3.85	GITWGEETLMEYLE	P62894
2418	6.24	2	512.02	17.84	4	2044.0	9.82	RHPYFYAPELLYYANK	P02769
2445	4.86	1	625.98	12.23	3	1874.9	5.38	VYACEVTHQGLSSPVTK	123456
2511	5.28	3	673.70	17.27	3	2018.0	5.38	LPLVGGHEGAGVVVGMGENVK	P00330
2535	---	---	418.21	8.86	2	834.40	10.08	FTISRDD	123456
2561	5.13	1	557.95	21.39	3	1670.8	15.30	TKPREEQYNSTYR	123456

439 search results. 105 matching batch delete options.

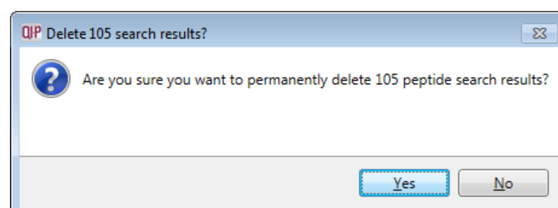
Section Complete

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

Note: the search results matching the filter criteria turn pink and the total is displayed at the bottom of the table (in this example: 105 matching out of 439)

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

Click **Yes**.



Having applied all the filters there will be **334** search results remaining

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

Stage 7 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 can be 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.
- At the resolve conflicts stage, sort the conflicts by descending order to view the list of conflicts

In this example there is a small number of conflicts to resolve as shown in the Conflicts column of the Proteins table.

Resolve the conflicts so as to maintain most of the contaminating proteins. As an example, favouring the Mouse 'Activator of 90kDa.. protein' resolve the 3 conflicts with Glycogen Phosphorylase by un-ticking them in the top right panel

Waters
THE SCIENCE OF WHAT'S POSSIBLE.

Details on performing the remaining conflict resolution are available in Appendix 6 (page 49)

After resolution of the Conflicts there will be no Conflicts remaining in the Proteins table.

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

Stage 8: Creating a Spectral Library

The **Review Proteins** stage opens displaying details for all proteins. Before creating a Spectral library from the identified proteins you can explore the correlation of the peptides/peptide ions with the behaviour of the protein with regards to the protein behaviour between the same conditions used in the experiment.

To view all the peptides assigned to a protein either double click on the protein or on **View peptide measurements.** For a protein you can order the peptides based on the correlation score and then select all of them, the example below shows a high correlation for all the identified peptides of **bovine cytochrome C**

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

You can create a spectral library to one or more proteins in Progenesis QI for proteomics by selecting the **peptide ions** identified as corresponding to a particular protein. One way of doing is to tag the **peptide ions** for a protein(s) and then simply export them to create a spectral library.

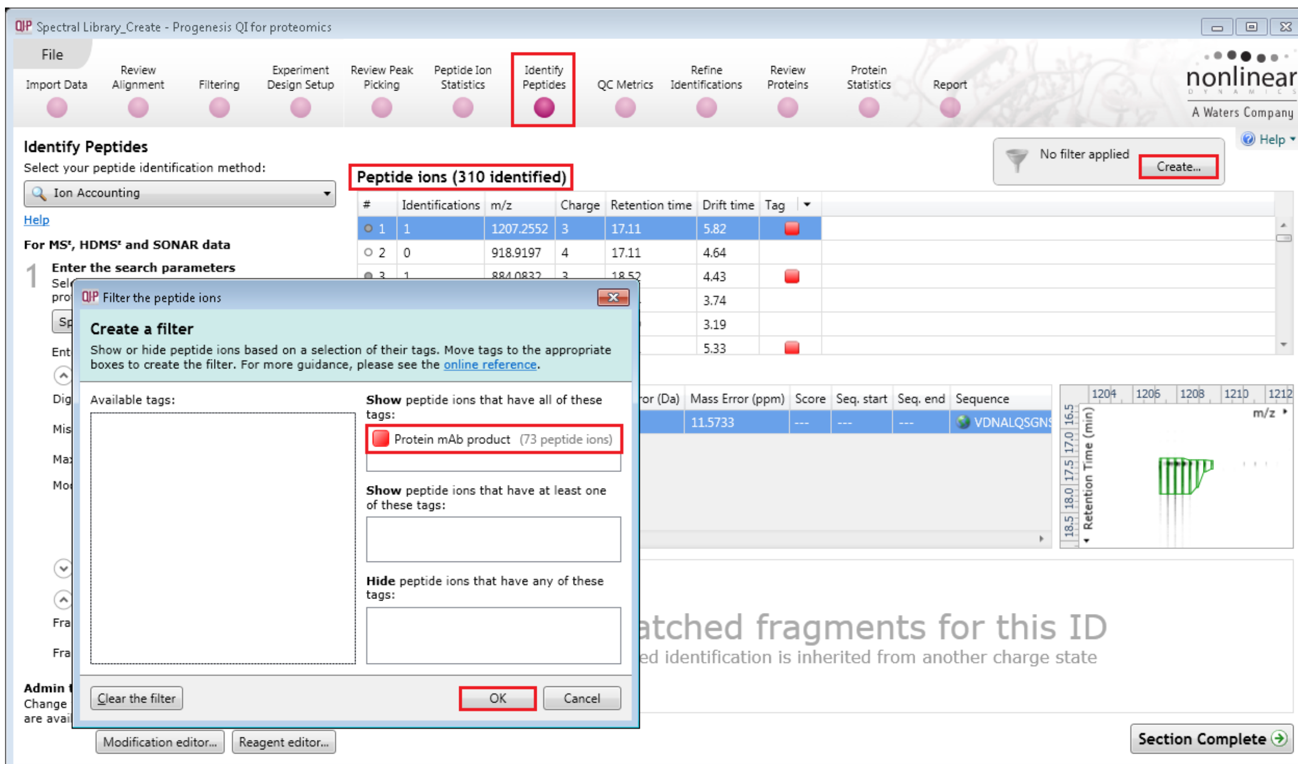
First example library to create is for the antibody protein 123456 where the table indicates that there are 44 unique peptides for this protein.

Double click on the protein or click on **View peptide measurements**

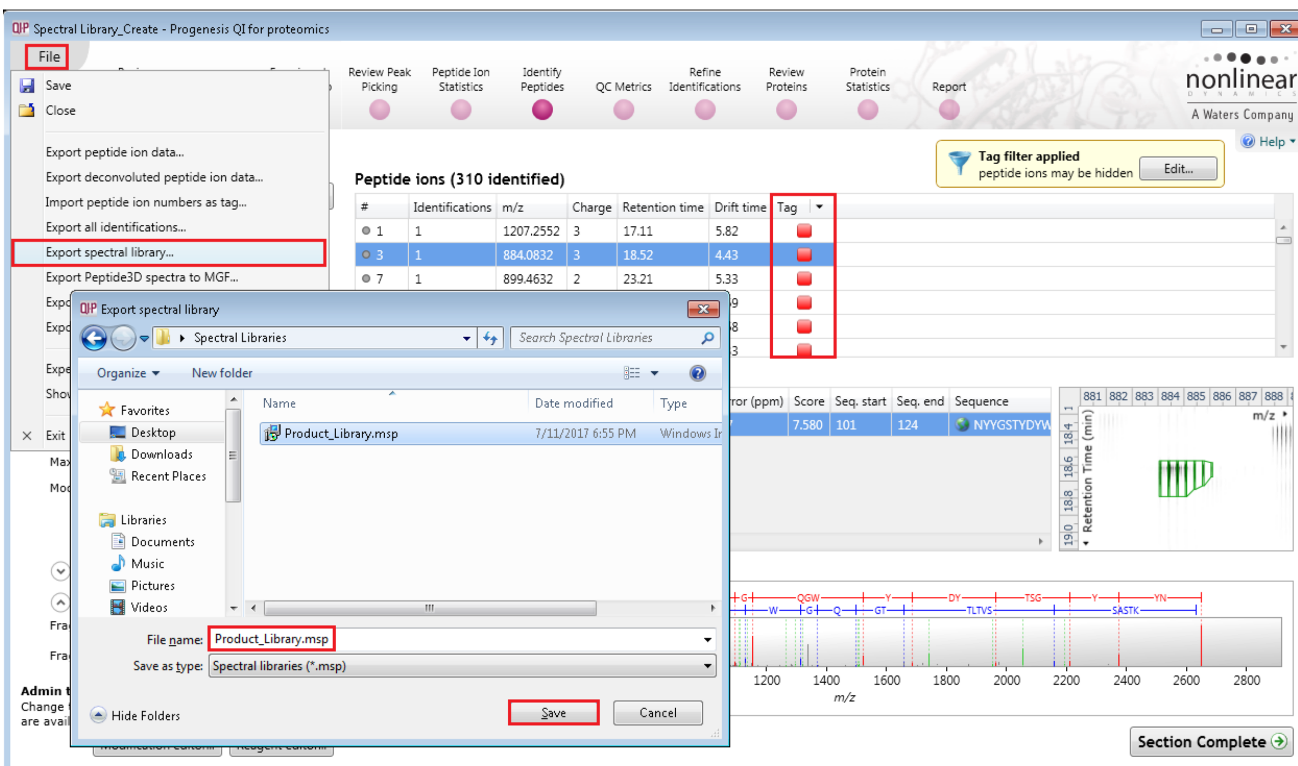
Then click **Select all peptide ions** to highlight all the rows in the table. Right click on the table and create a new tag (Protein mAb product) for all the highlighted ions.

To create the library from these tagged peptide ions click on **Identify peptides** on the workflow

Then create a filter to reduce the table to only those ions identified by the tag (Protein mAb Product). To create the filter drag the tag set on to the panel '**Show peptide ions that have all of these tags**', to restrict the table to showing only these peptide ions.



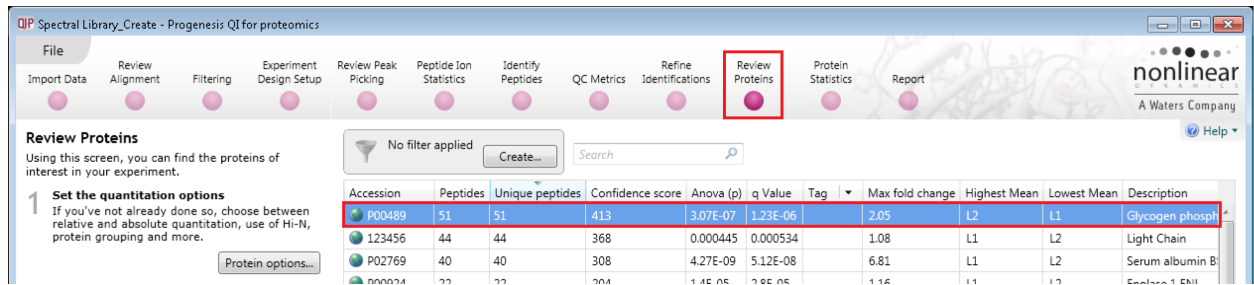
Select **Export spectral library...** from the **File** menu, name and save the library, Product _Library.msp.



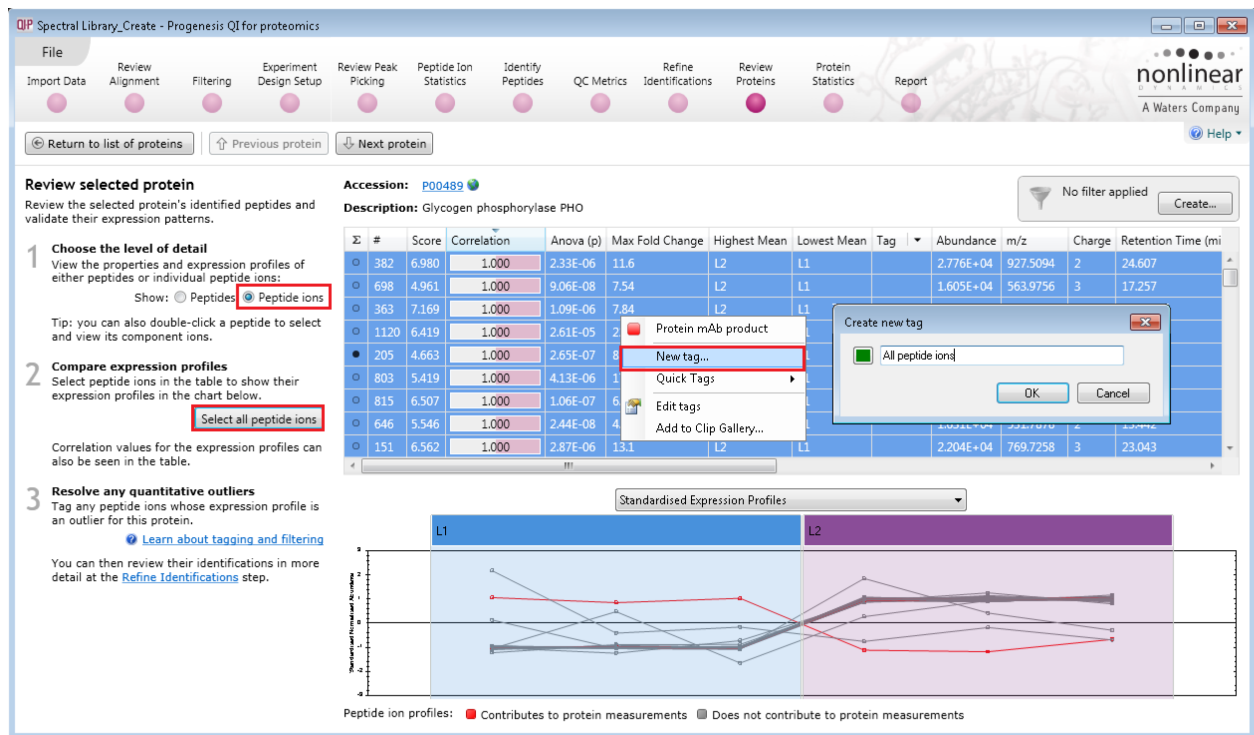
Having exported the library clear the current Tag filter by clicking on **Edit** and the **Clear the filter**

To create a second example spectral library from **all the peptide ions** for all the identified proteins in this analysis, first return to the **Review proteins** stage using the Workflow icon.

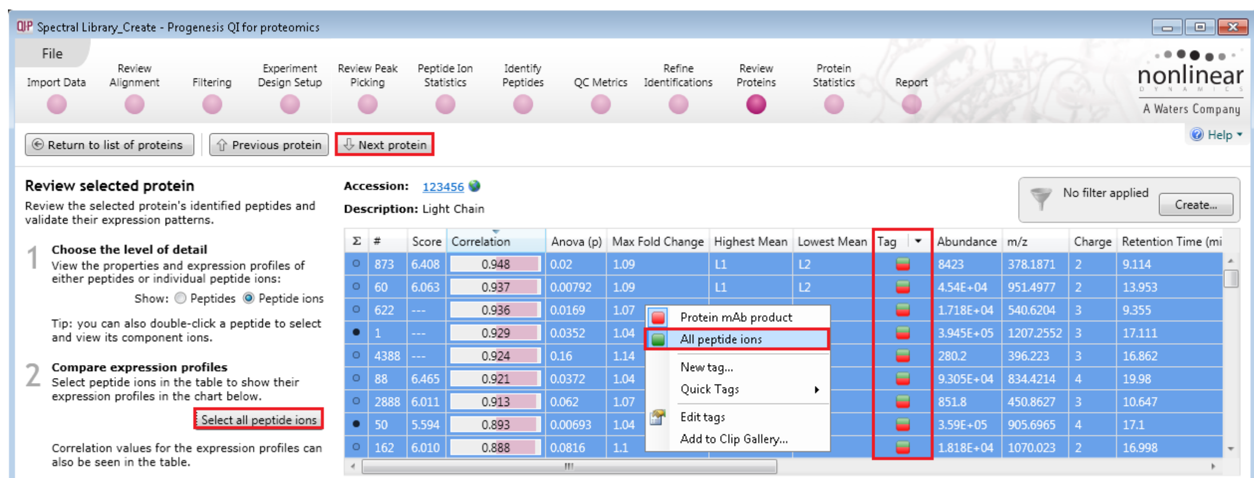
Then select the first protein, double click on it and ensure that you are looking at the peptide ions.



The click **Select all peptide ions** for Glycogen phosphorylase

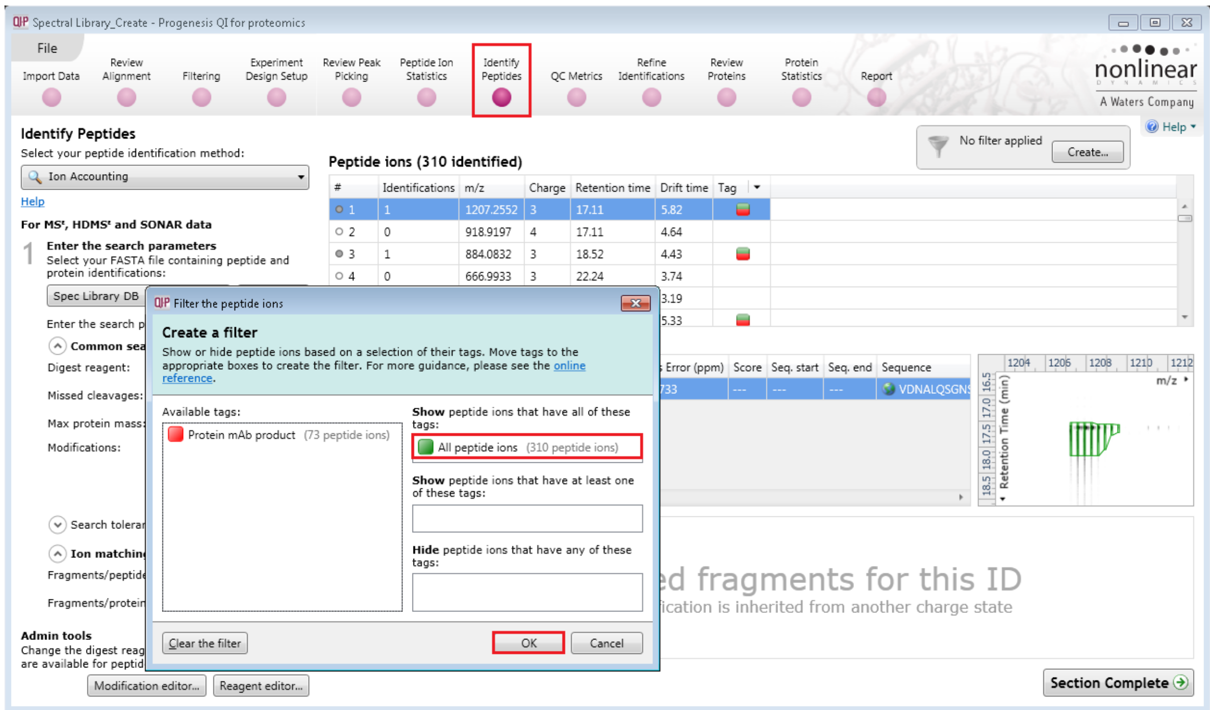


Right click on the highlighted peptide ions and select **New tag…** Create a new tag (All peptide ions) click OK to add the tag to all the peptide ions.

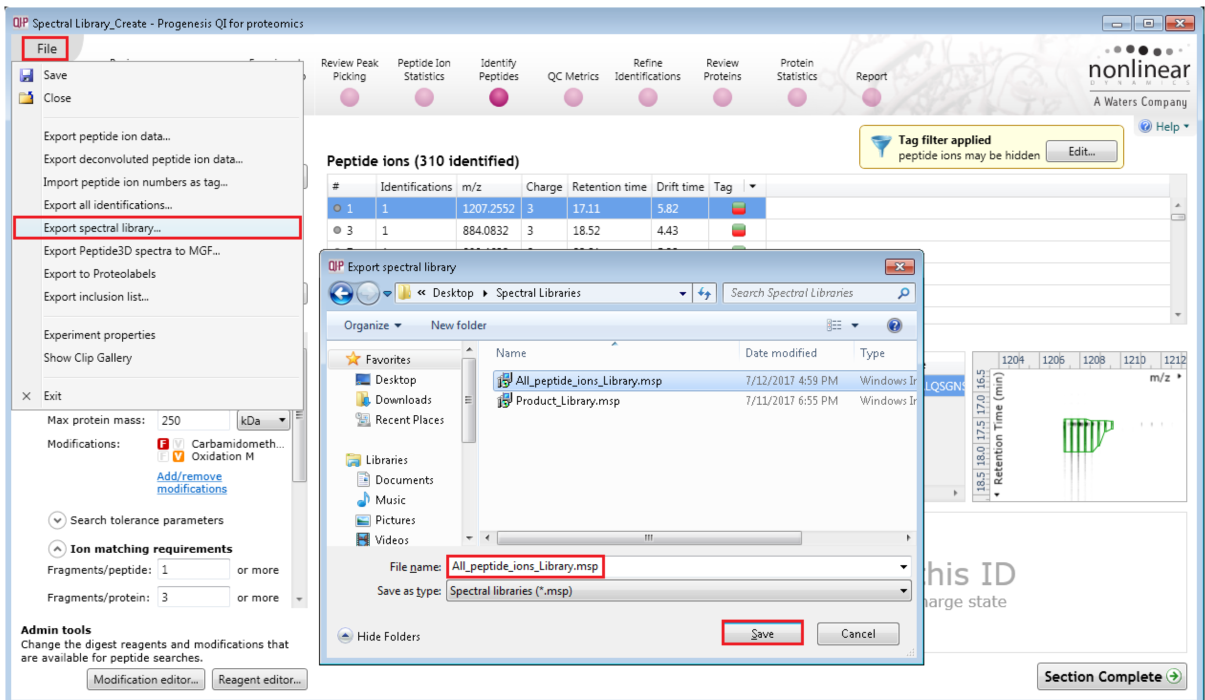


Use the **Next protein** button to step to the next protein and click **Select all peptide ions** then right click on the highlighted peptide ions and click on the **All peptide ions** tag to add the tag to all the highlighted peptide ions. Repeat this process using the **Next protein** button to step down your list. Having tagged all the

peptide ions return to **Identify Peptides**, using the workflow, and create a tag filter for the **All peptide ions** tag

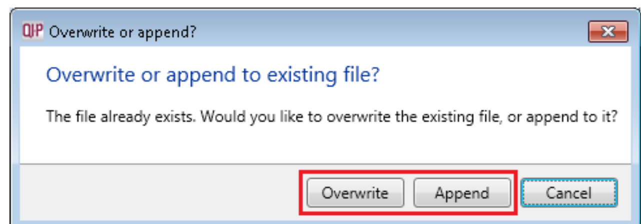


Select **Export spectral library…** from the **File** menu, and then name and save the library as,



All_peptide_ions_Library.msp

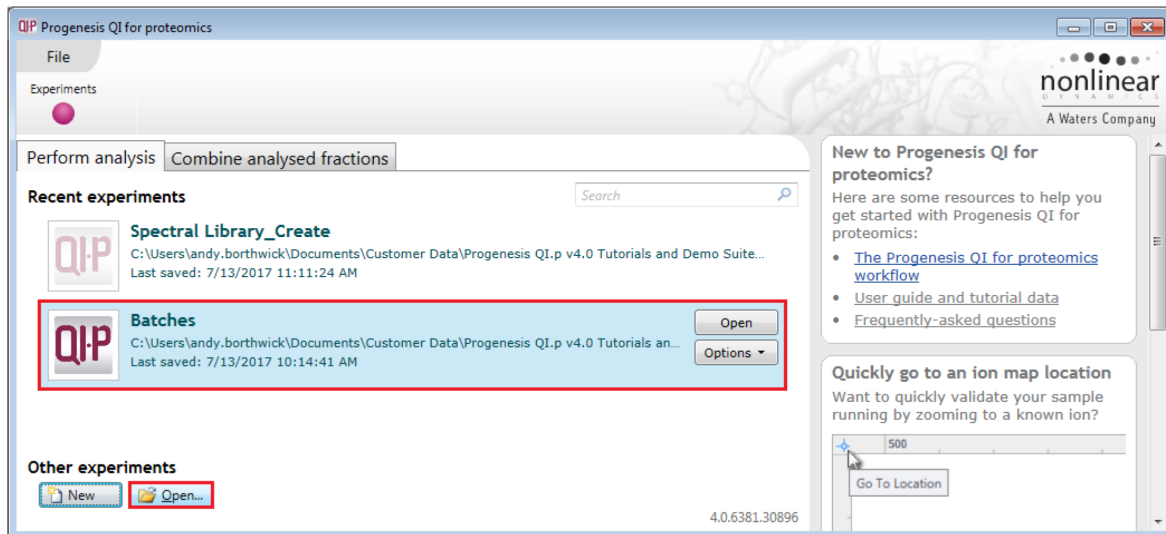
Note: if a library already exists you will get the options to Overwrite or Append the content.



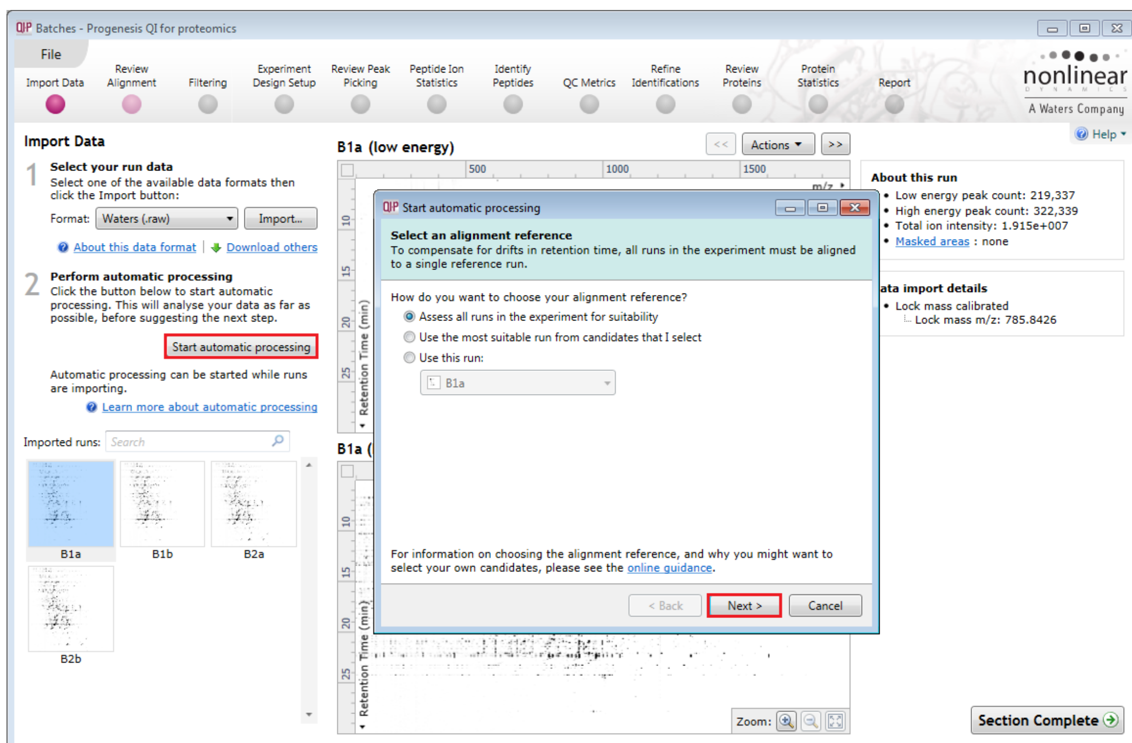
Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Stage 9: Identifying proteins using a spectral library

As an example of using the spectal library you have just created use it to identify and measure the proteins present in the **Batches** experiment downloaded with this tutorial.



Open and automatically process the **Batches** experiment as follows



Setting up the steps for the auto processing allow:

Default option for selection of alignment reference

Automatic alignment

Default settings at Peak Picking

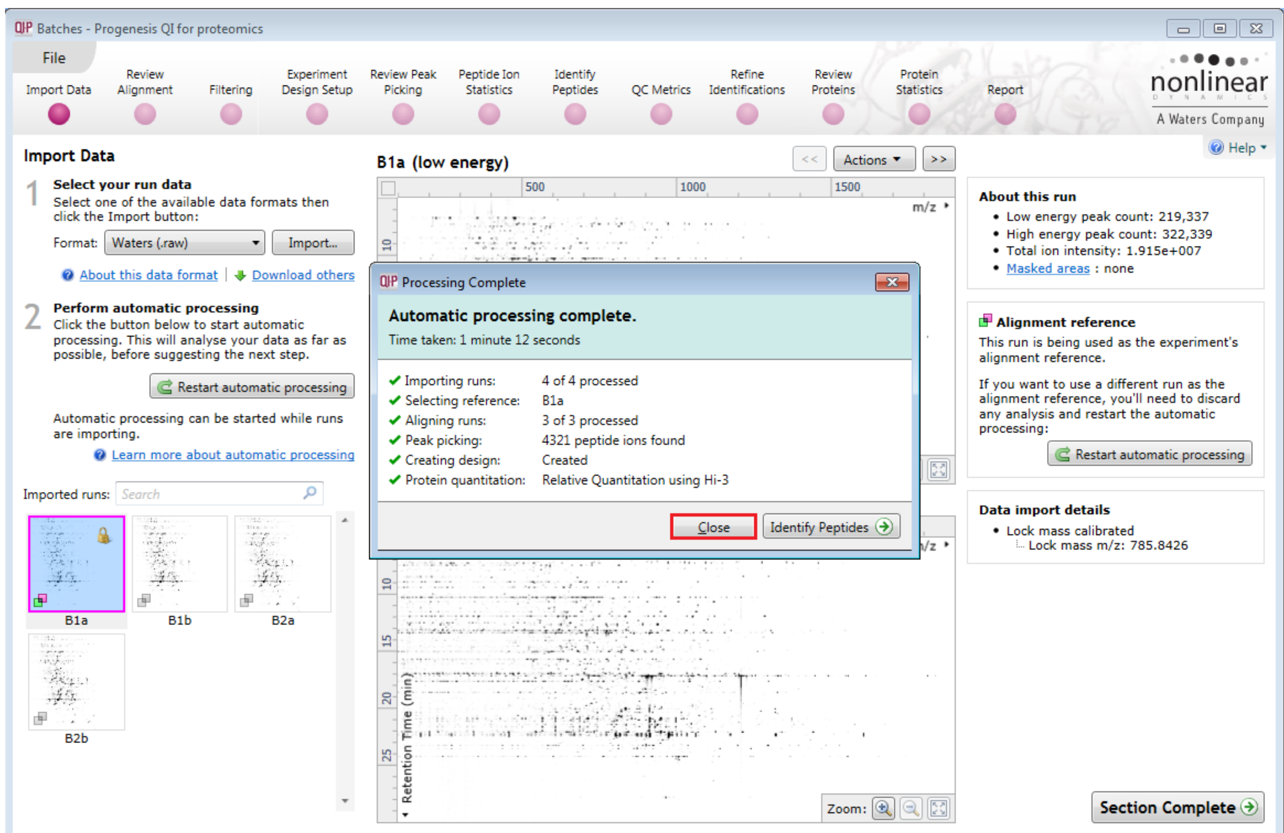
For Experiment design setup: use file Group Batches.csv

Do not perform identification by deselecting **Use MS^{E} data........ untick this option**

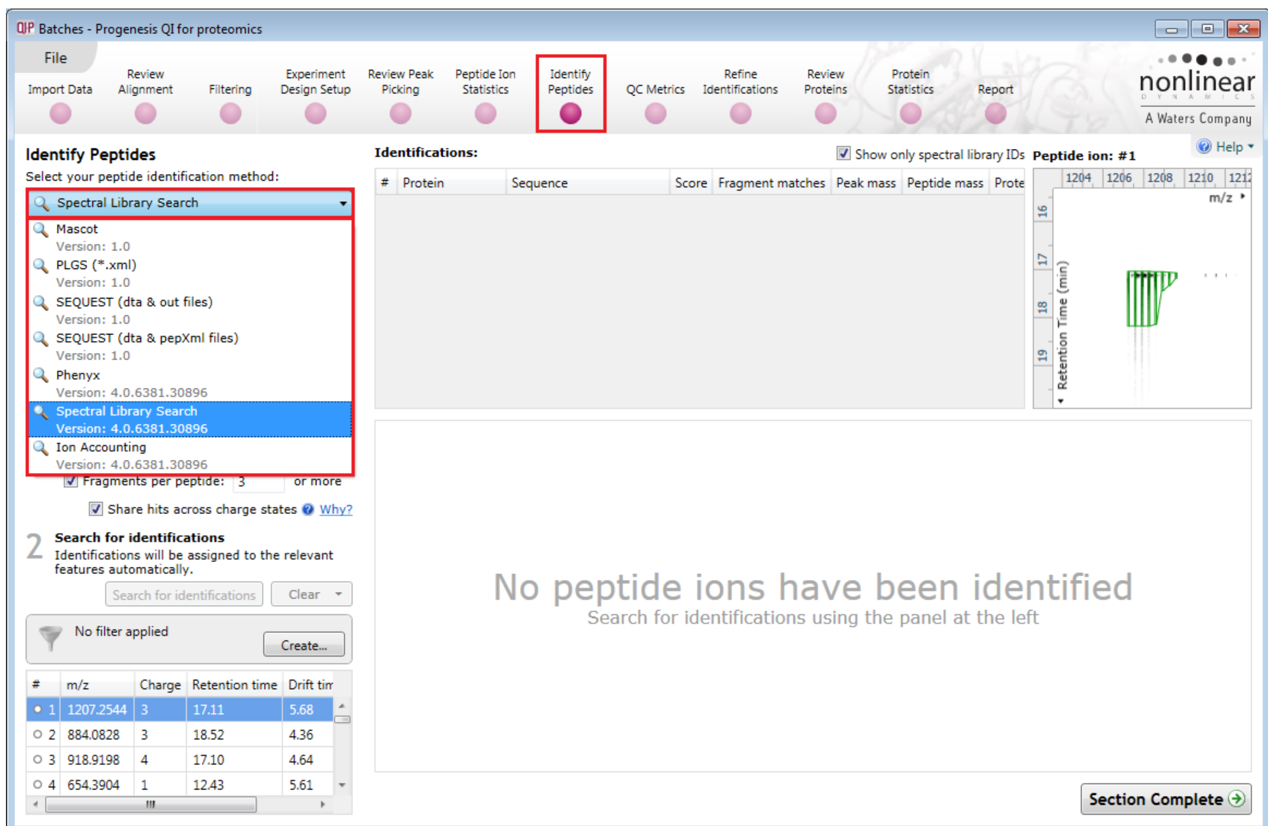
Use default settings for Quantitation and Use protein grouping

Click **Finish** to start the analysis.

Waters
THE SCIENCE OF WHAT'S POSSIBLE.



When the analysis completes click on **Close** then on **Identify Peptides** using the workflow icon.



Select **Spectral Library Search** as the **Peptide identification method**

Locate and select the Product_library (msp file created from the Spectral Library analysis)

Set the **Precursor** and **Fragment** tolerances to 25ppm and ensure that **Fragments per peptide to 3 or more** is ticked and the **Retention time** and **Share hits** options are un-ticked.

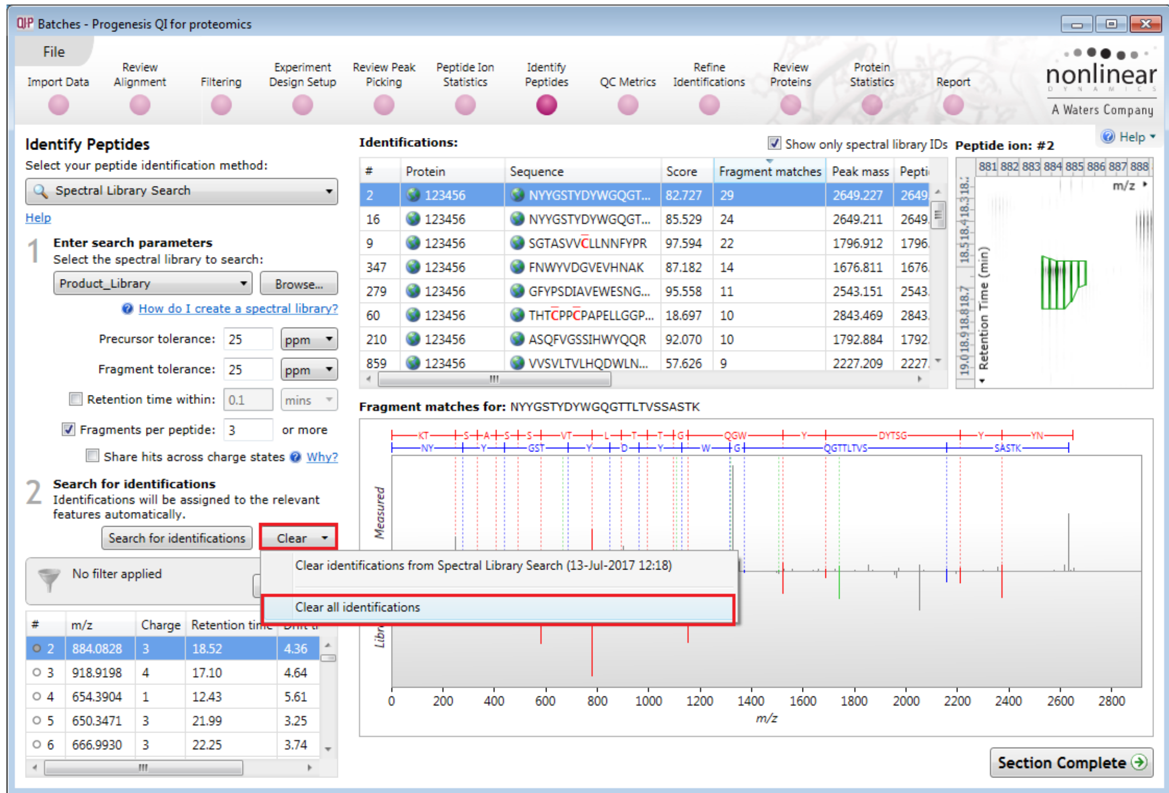
Then click **Search for identifications** as shown above.

When the search completes order the table on Fragment matches, and select the first example in the list. The 'mirror' plot will now display the fragment matches for the current peptide sequence showing the measured and Library b and y ions.

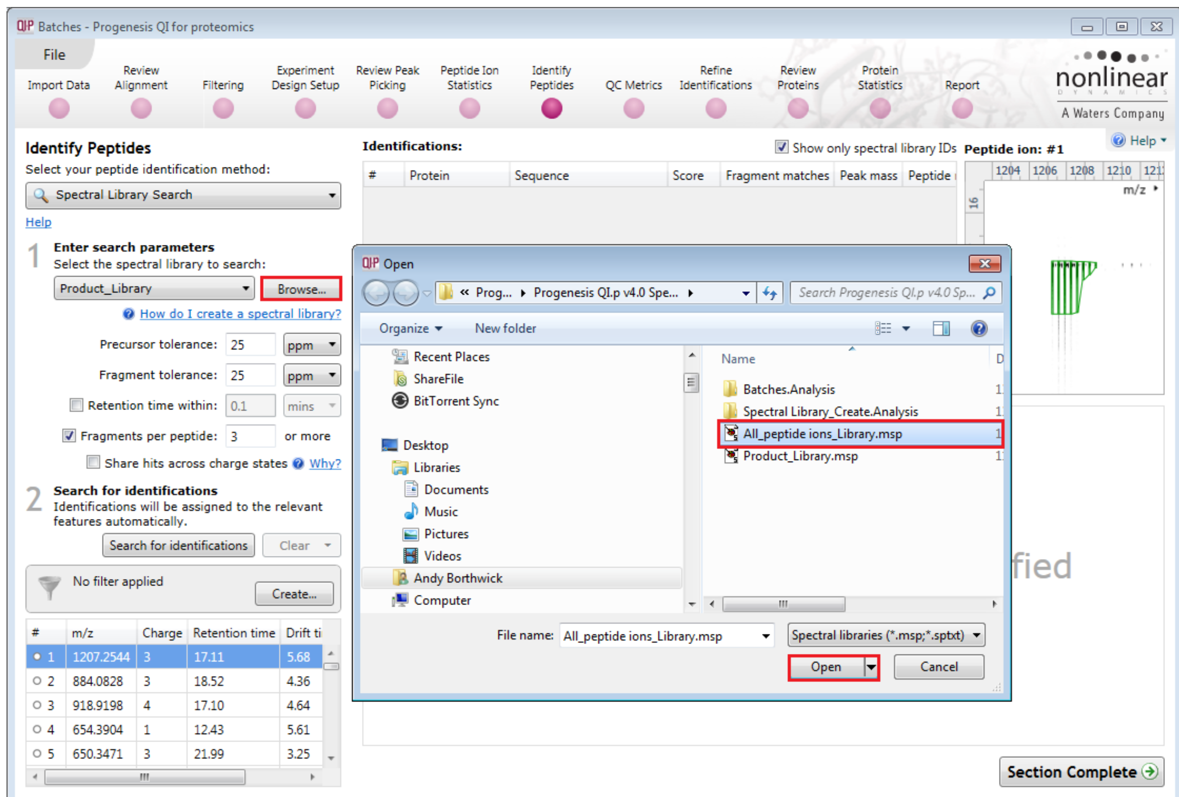
Now click on **Review proteins** on the workflow to examine the quantified proteins. As the Product library only contained the spectra for one protein (123456) then only one protein is displayed.

As the protein Quantitation is Relative based on the 3 most abundant peptides (Hi3) B2 (Batch2) shows a moderate increase of 12% over B1 (Batch 1)

You can also use the second library that you created in the previous section to explore the relative contamination of the 2 batches as a result of differential spiking of the product with 5 known proteins as well as well as low level contamination from the mouse host cell proteins. Enolase was spiked at 250fmols per sample and can be used to determine Absolute Quantitation of the proteins identified as present in the batches using the second spectral library to focus the search.



Return to **Identify peptides** and firstly clear the existing search results.



Using the same parameters click **Search for identifications**

Waters
THE SCIENCE OF WHAT'S POSSIBLE.

There are spectra from 10 proteins contained in the All_peptide_ions_Library

Viewing the proteins identified by Spectral Library matching, shows the presence of 10 proteins where 3 of them have been identified by the matching of a single unique peptide

As 250fmol of Enolase has been been spiked into Batch 1 and Batch 2 samples, Absolute Quantitation can be performed. Click on **Protein options** and select **Absolute Quantitation using HiN.**

Tip: as you start to type the Accession number for the Calibrant it will offer you the list of possible accessions from the Protein table, click on the required one as it becomes visible in the selection and then enter the amount of fmols added and click **OK**.

This will not affect the observed fold change between the 2 batches but it will generate additional columns in the table for each Batch in (fmol).

Description	Amount (fmol) - B1	Amount (fmol) - B2
Activator of 90 kDa heat shock protein ATPase homolog 2 OS=Mus musculus GN=Ahsa2 PE=2 SV=2	6.56	45.6
Serum albumin OS=Mus musculus GN=Alb PE=1 SV=3	109	18.3
bovine cytochrome C	357	119
Glycogen phosphorylase PHO	43.5	277
Serum albumin BSA	174	26.9
Hemoglobin subunit beta-1 OS=Mus musculus GN=Hbb-b1 PE=1 SV=2	5.27	27.3
Light Chain	1.79E+03	2E+03
Alcohol dehydrogenase 1 ADH	81.9	154
Enolase 1 ENL	266	234
Trypsin OS=Sus scrofa PE=1 SV=1	74.2	68.2

The **Absolute quantitation** values are included as an option in the **Export protein measurement** dialog.

Note: you can also capture the table by right clicking on it and Adding it to the Clip gallery as described in Appendix 7 (page 53)

Additional Tools for Spectral Library Construction

The description and methods employed to construct and use a Spectral Library only address some of the strategies that can be used in Progenesis QI for proteomics. Other tools have been added or made available at other stages in the workflow that can and will help in the review process of the quality of the spectra added to the Spectral Library.

For example: the capacity to look at the correlation of the peptides and peptides ions has been available since Progenesis QI for proteomics v3.1. This is useful not only in confirming patterns of peptide ion behaviour that match the changes in abundance of the protein between experimental conditions but also in identifying atypical behaviour of assigned peptides when there is a clear difference between the actual abundance of a protein.

However, the use of correlation *per se* must be viewed with some caution when there is no, or very little actual difference in the protein between the conditions.

For example: Compare the behaviour of peptide ion correlation for Cytochrome C and Enolase in the Create Spectral Library experiment described in this user guide.

Waters
THE SCIENCE OF WHAT'S POSSIBLE.

At **Review Proteins** go to the peptide ion level, order on Correlation and **Select all peptide ions**

Viewing all the Standardised expression profiles for the Cytochrome C (spiked protein) where we are expecting a 5 fold decrease in the expression, the pattern for the correlation of all peptide ions shows this

Whereas for Enolase, where **No** difference between the conditions is expected, the Standardised plot indicates approximately half of the peptide ions exhibiting increased expression in L2 compared with L1 and the other half showing a decrease.

With Enolase peptide ions selected, the Standardised view presents what looks like two populations of Peptide ion behaviour. When you look at the actual relative fold change these reflect a small % change about 1.0 which is accentuated by how the Standardised view is derived.

With the table ordered on correlation, scroll to the region where the correlation drops below 0.9.

Note: these peptide ions display p values >0.05, highlight these rows in the table and create a tag filter for them so that they can be hidden and then specifically deleted them as required.

Create a tag filter for the peptide ions.

When this is applied, reselect all the remaining peptide ions in the table (see below).

Having removed the poorly correlated peptide ions the plot of all the peptide ions now looks like this:

The ability to do correlation based filtering has been included at Conflict resolution (see below) where you can order the table by **Correlation** and tag the peptide ions with the poorest correlation.

Hide them by applying the tag filter and then select all the remaining peptide ions in the table.

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Finally you can also assess the possible interference of other peptide ions that overlap with the peptide ion of interest. To do this order the peptide ion data on the column related to peptide ion overlap as shown below.

The actual overlapping ions can also be viewed by setting the current ion's focus to show the **Overlapping peptide ions** to assist in the process of evaluating which peptide ions to include in the Spectral Library.

Note: the table also displays (by a solid icon) which peptide ions currently contribute to the protein's quantitation.

Peptide ions of P00489

Indicates which peptide ions contribute to the protein quantitation

	#	Σ		Score	Hits	Correlation	Mass	Mass error (p...
☑	1112							
☑	246	●	1	5.5	3	1.000	2117.163	15.2
☑	752	○	1	4.51	1	1.000	1990.033	17.2
☑	361	○	1	6.69	5	1.000	1052.582	15.9
☑	442	○	1	5.42	3	0.999	1889.029	15
☑	1109	○	1	6.45	3	0.999	946.481	16.3
☑	233	○	2	7.06	5	0.999	1439.749	13.5

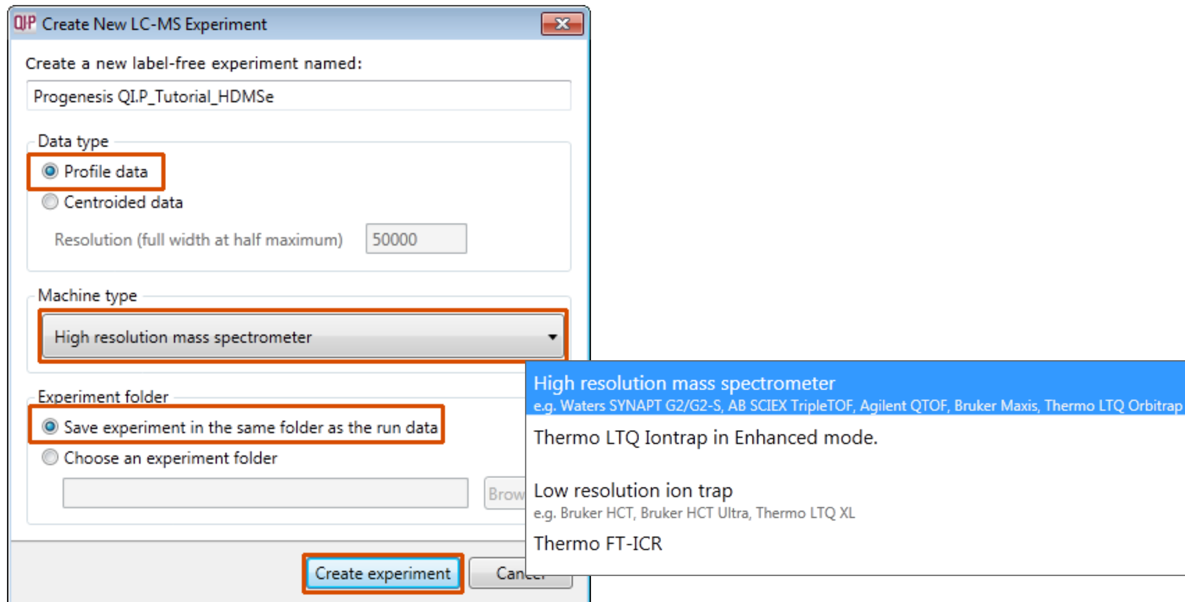
Combining the use of these tools and approaches will allow you to make informed decisions on the number and quality of the spectra that you include in the Spectral Library.

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

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

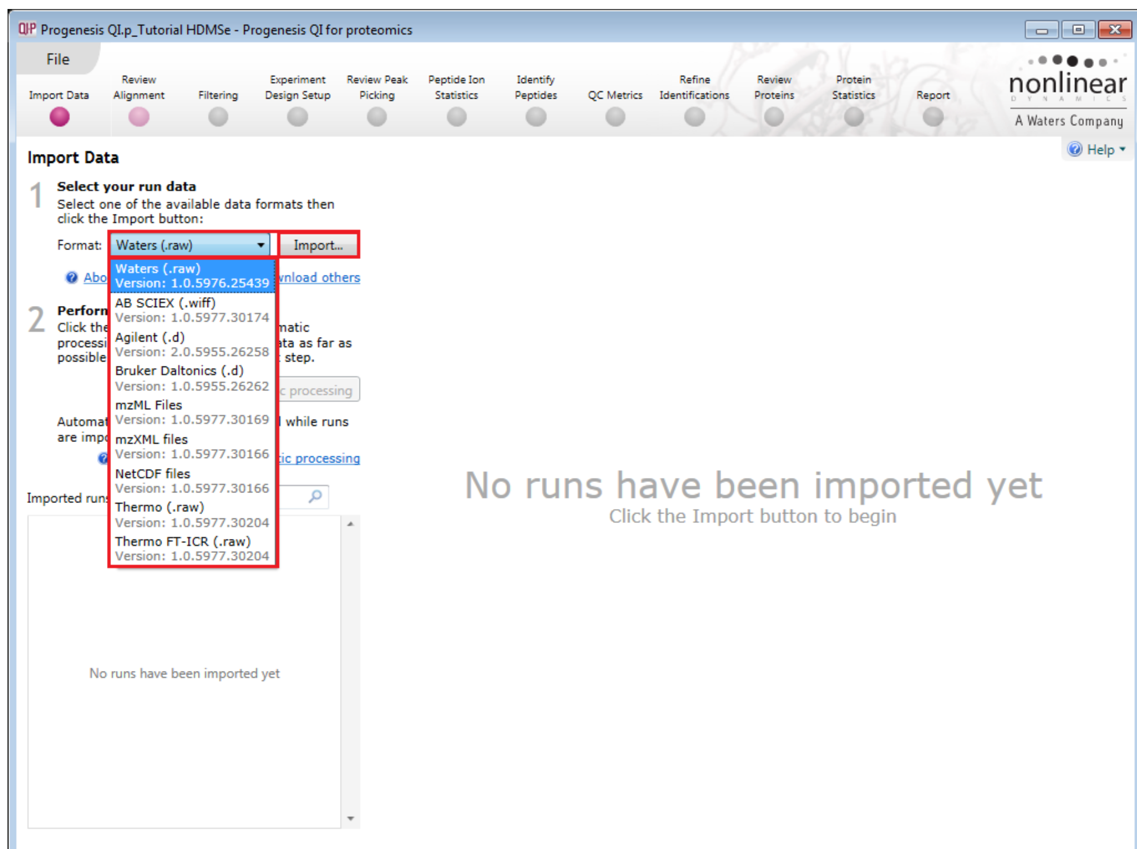
Note: if you have converted or captured the data as centroided then select Centroided data and enter the Resolution for the MS machine used.



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

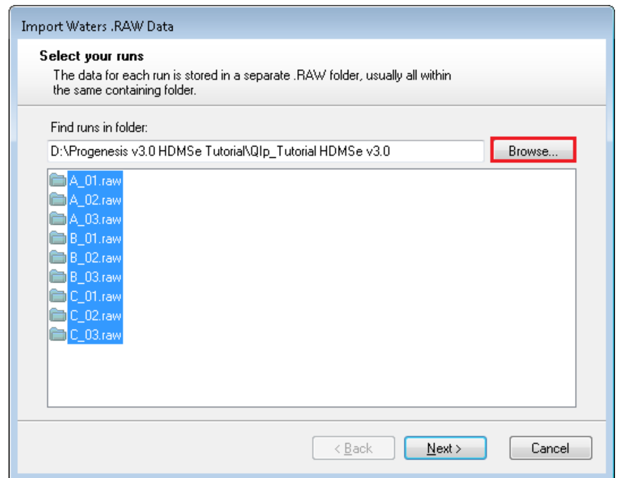
Select the 'Import Data file format', in this example they are Waters/SYNAPT data

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



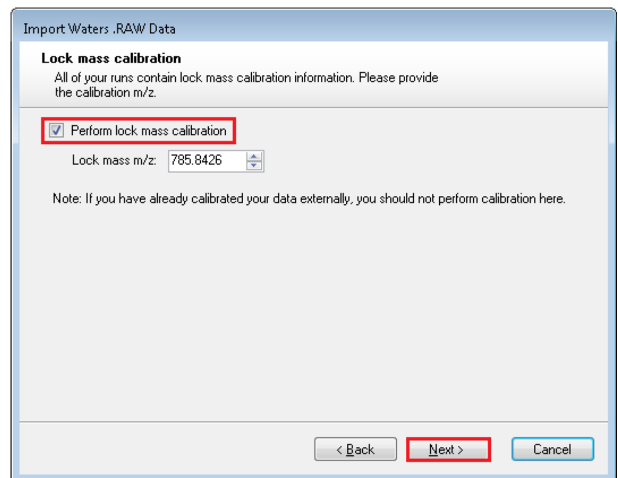
Waters
THE SCIENCE OF WHAT'S POSSIBLE.

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



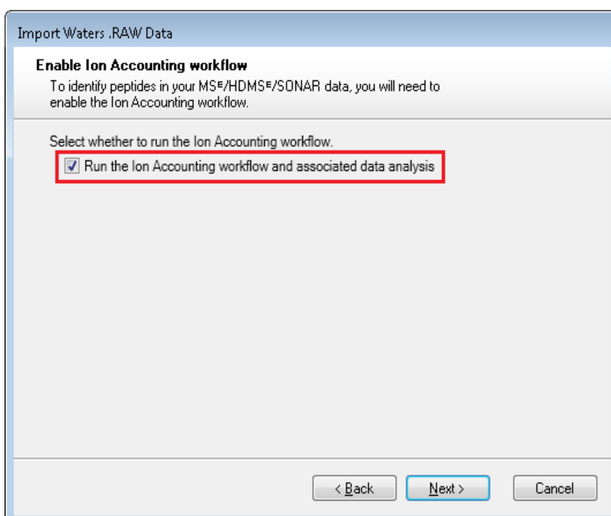
On importing, the lock mass calibration is read and presented on this dialog

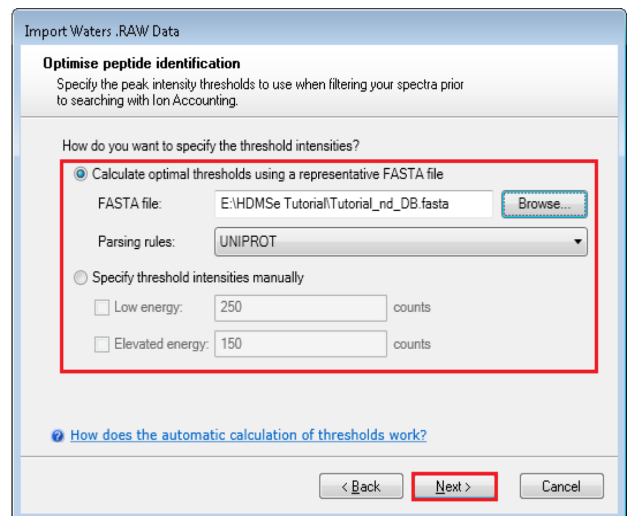
You can, if required, alter the lock mass calibration at this step.



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.





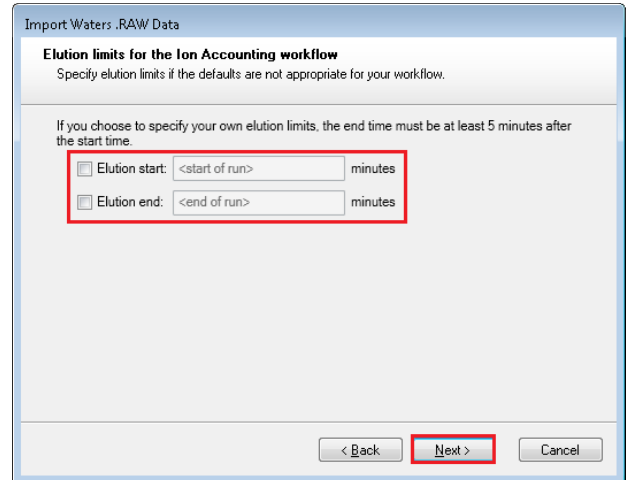
If you choose to determine the thresholds automatically then Progenesis determines appropriate thresholds by sampling each run and finding the thresholds that yield the most protein identifications in the sample area. For each run, Progenesis performs the following steps:

- 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.

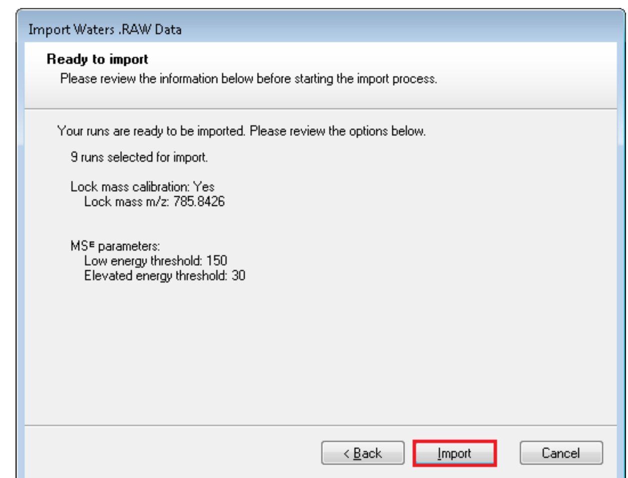
- Finally, it selects the thresholds that resulted in the largest number of protein identifications and applies those to the whole run. If more than one set of thresholds results in similar numbers of identifications (within 10% of the maximum), it will choose the highest thresholds as a way of optimising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

Note: for HDMSe the settings are 150 and 30 and for MSe the 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.

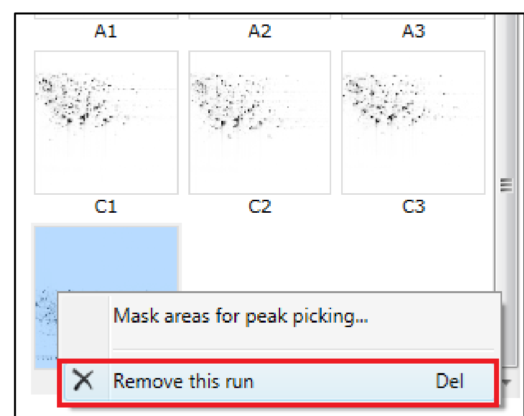
Note: For a large number of files this may take some time.

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 right 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

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.

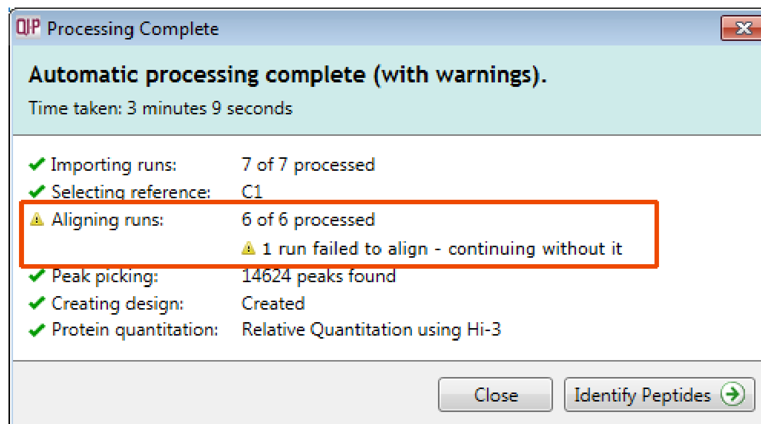
Note: you will be offered the automatic alignment if you have not performed it automatically already.

Now move to the next stage in the workflow (page 7 in this user guide) by clicking **Section Complete.**

Appendix 2: Stage 2 Processing failures

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

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

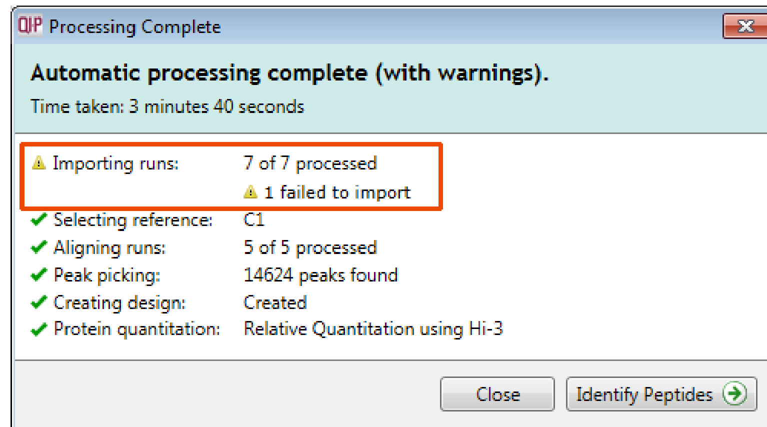


Note: in this example the run that failed to align will not contribute to the peak picking and will be excluded at the alignment stage (a cross appears in the include column).

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.

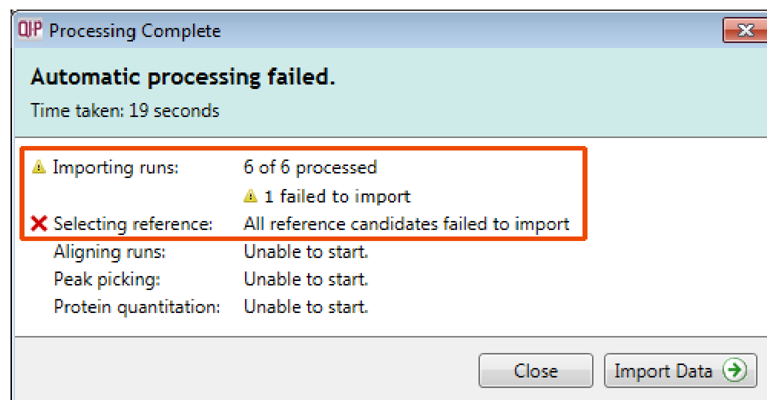
If some runs in a data set fail to import (but not all), the automatic processing will continue informing you that one or more runs have failed to import.



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.

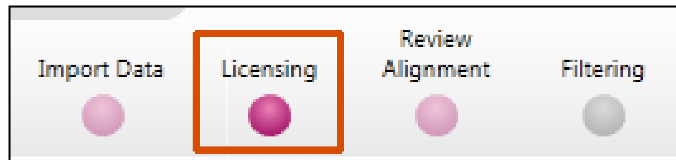
An example of a problem that would halt the automatic processing would be the failure to successfully import all the potential reference candidates, (for example: while importing, you specified the selection of the alignment reference to be made from several runs before they were fully imported and set 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

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.

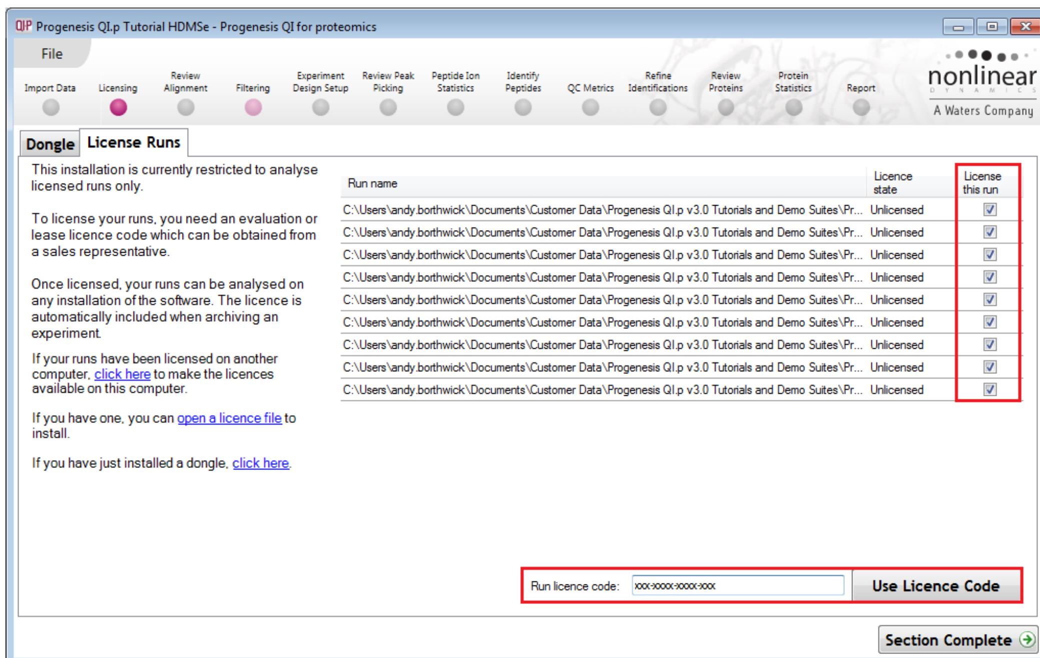
To use this page to License your Runs you must first either obtain an **'Evaluation' Licence Code from a Sales Person or purchase a licence code directly**.

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

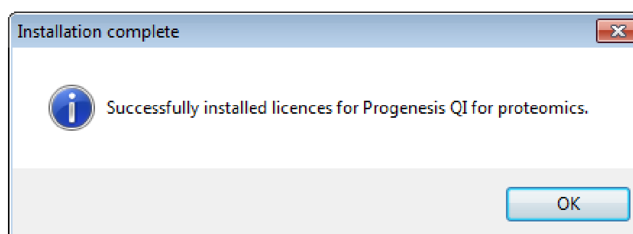
The runs in your experiment will be listed as shown below.

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 .

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Appendix 4: Manual assistance of Alignment

Approach to alignment

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

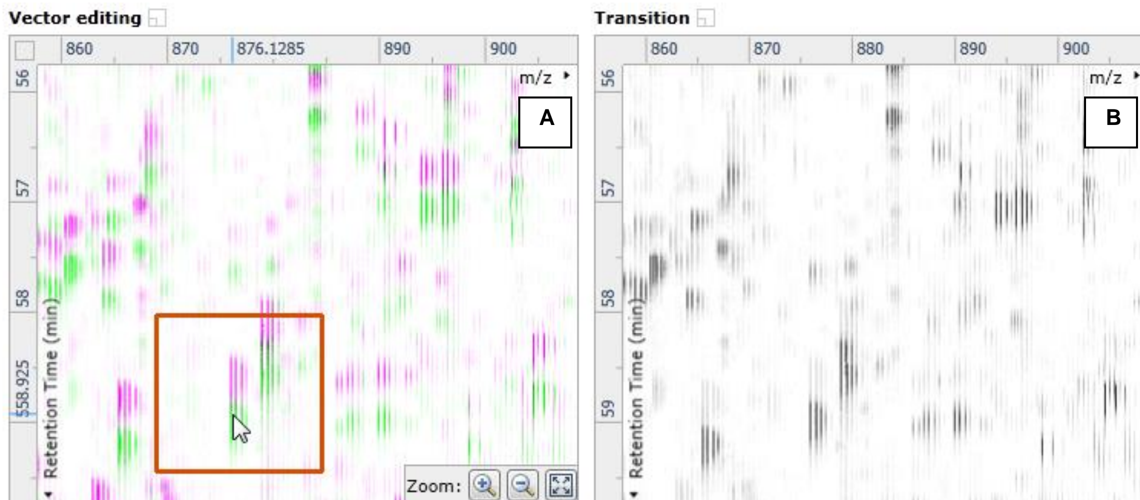
1. Click on Run B_02 in the **Runs** panel, this will be highlighted in green and the reference run (A_01) will be highlighted in magenta.
2. You will need to place approximately 5 - 10 **alignment vectors** evenly distributed from top to bottom of the whole run (RT range).
3. First drag out an area on the **Ion Intensity Map** (C), this will reset the other 3 windows to display the same 'zoomed' area

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

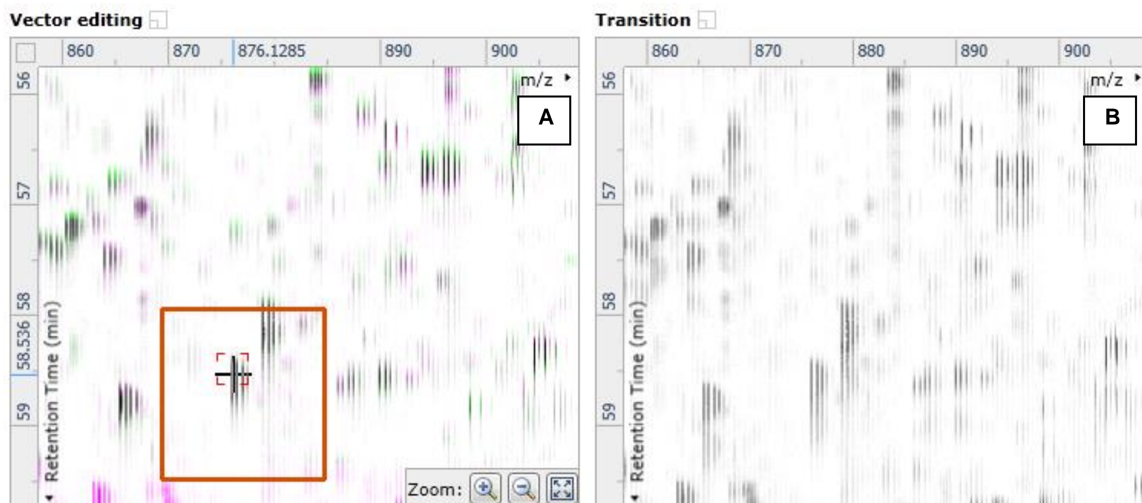
Note: the Ion Intensity Map gives you a colour metric, visually scoring the current alignment and an overall score is placed next to the Vectors column in the table. With each additional vector added this score will update to reflect the 'changing' overall quality of the alignment. The colour coding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

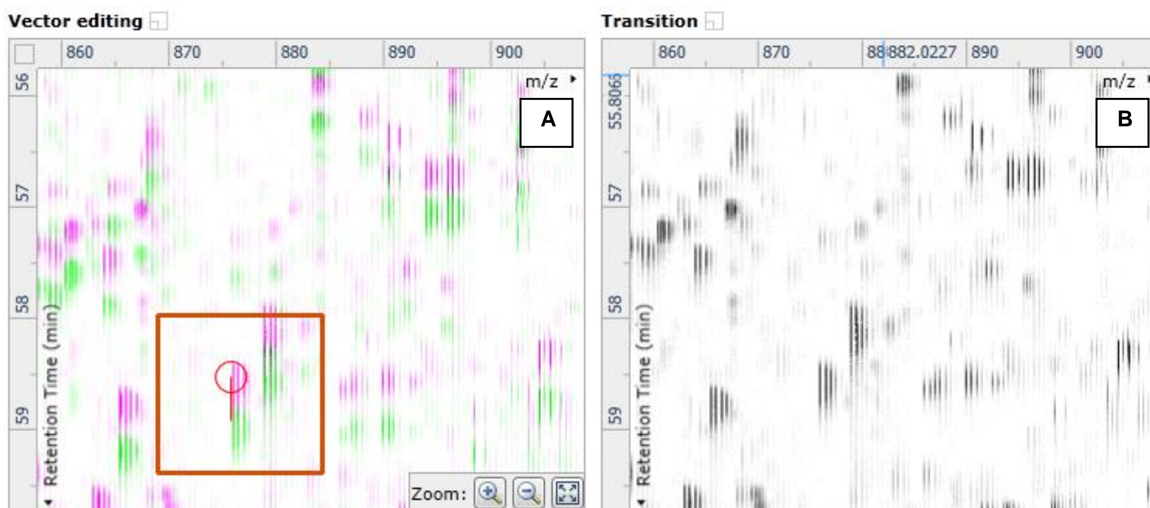


5. As you are holding down the left mouse button (depending on the severity of the misalignment), the alignment vector will automatically find the correct lock. If not, drag the green peptide ion over the corresponding magenta peptide ion of the reference run. The red box will appear as shown below indicating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



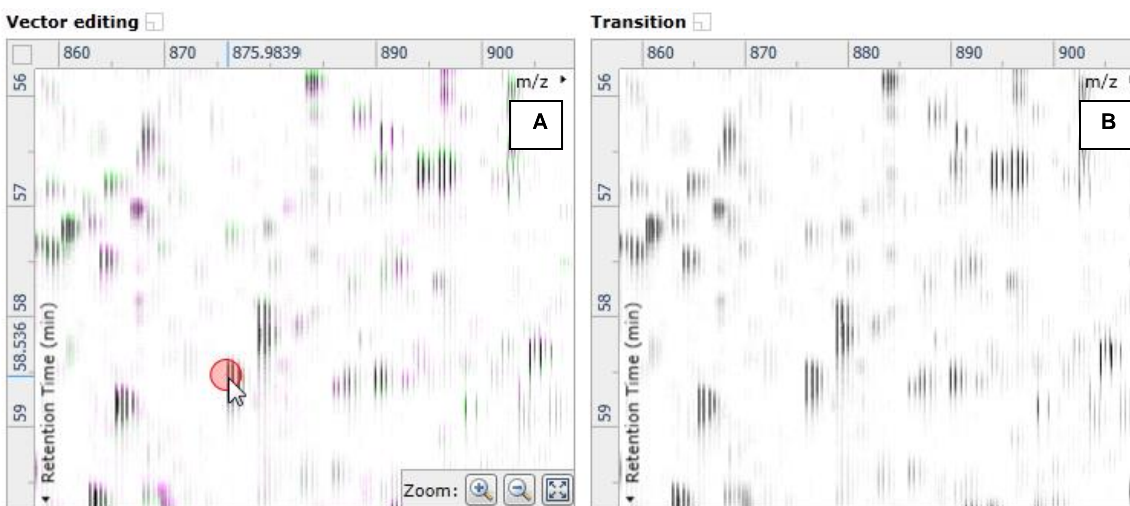
Tip: while holding down the mouse button hold down the **Alt** key. This will allow smooth movement of the cursor as the **Alt** key allows you to override the 'automatic alignment' performed as you depress the mouse button.

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.

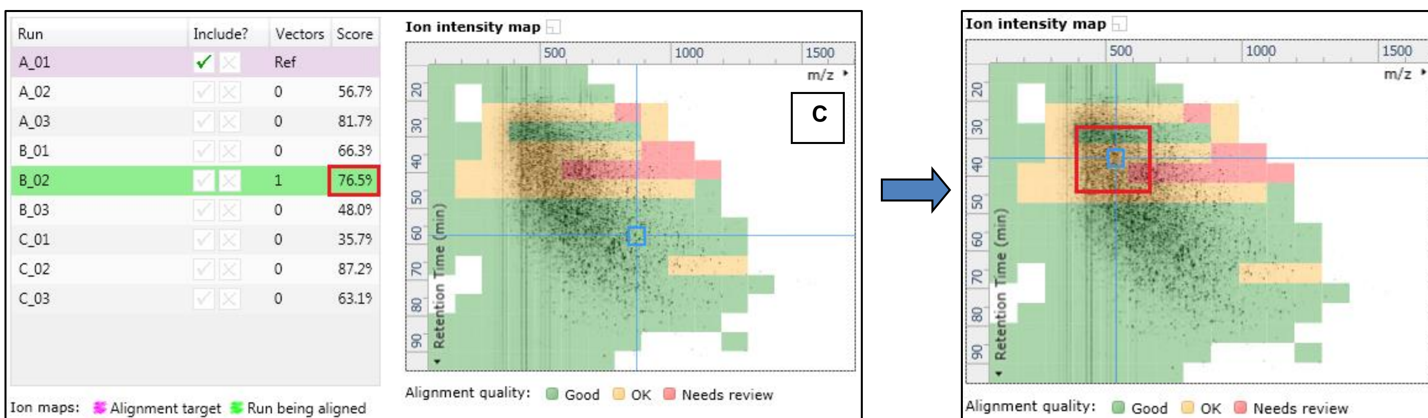


Note: an incorrectly placed vector is removed by right clicking on it in the **Vector Editing** window and selecting delete vector.

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



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



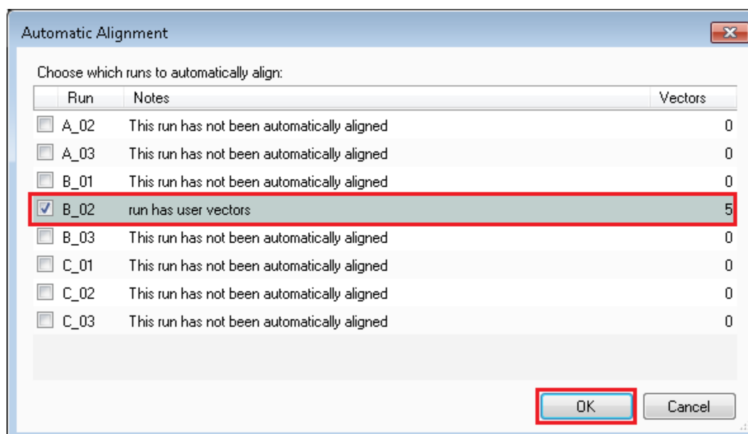
Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

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).

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 **Error! Bookmark not defined.**

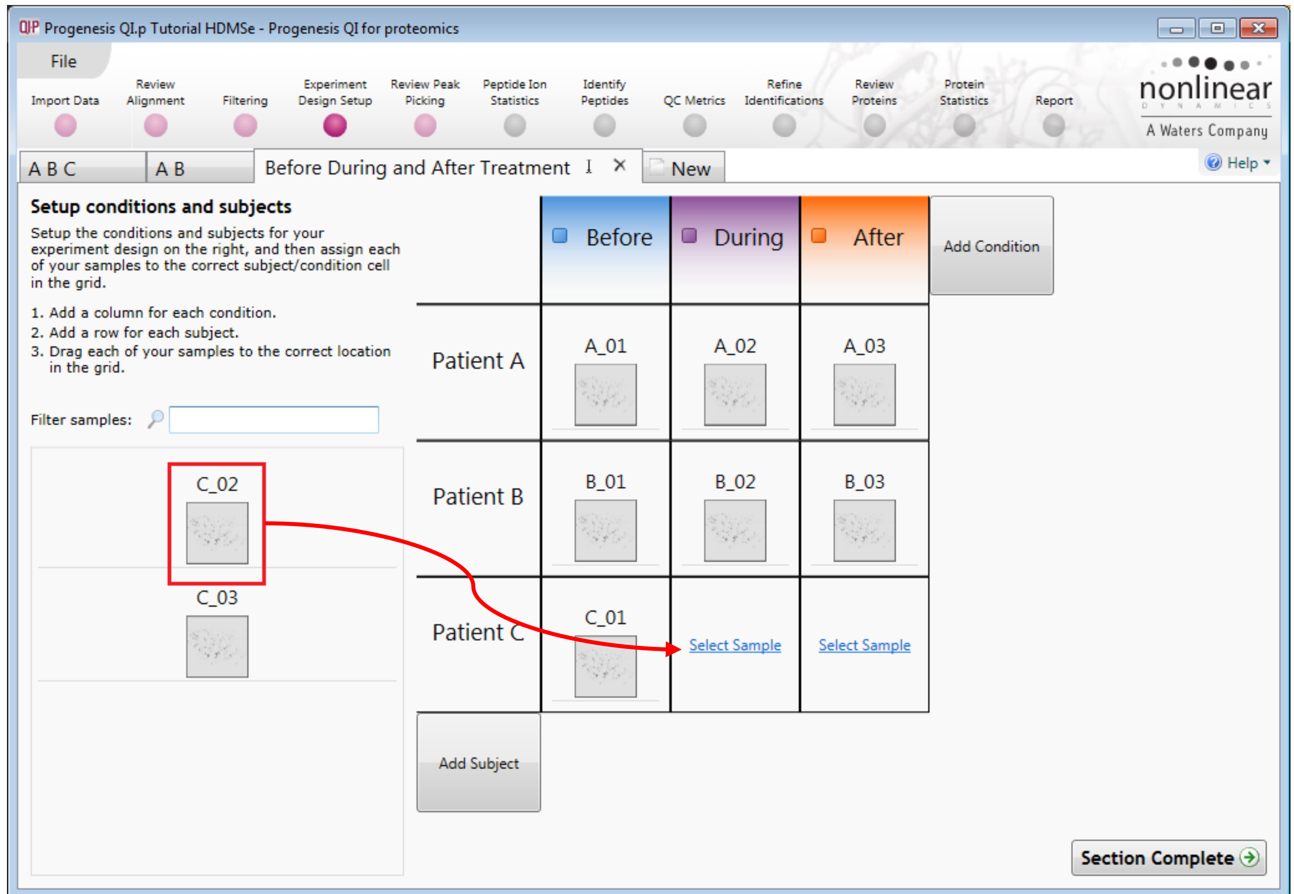
Appendix 5: Within-subject Design

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

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: 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.

To complete the resolution of conflicts started on page 16 carry on favouring the contaminating proteins.

Select the next protein in the ordered list (P00924 Enolase, also one of the spiked proteins) which like Glycogen phosphorylase has 3 conflicts .

Untick the conflicts favouring first Glyphorin

Accession	Peptides	Unique	Conflict:	Score	Tag
P00489	51	51	0	413	
Q8N9S3	3	3	0	13.1	
P00924	23	22	1	211	
Q78HU7	1	1	0	4.39	
P00330	20	20	2	144	
P07724	2	1	2	17.6	
P02769	41	40	2	322	
123456	44	43	1	368	
P62894 (+2)	11	11	0	64.9	
P02088 (+2)	2	2	0	13.6	
P00761	5	5	0	31.9	
Q8BKA3	1	1	0	0	
O70200	1	1	0	0	

	#	Σ	⧉	Score	Hits	Correlation	Mass	Mass error (p...	RT (mins)	Charge	Tag	Abundance	Conflict:
☐	511	○	1	7.1	4	---	1871.993	17.2	23	3		2.18E+04	0
☐	1739	○	0	---	---	---	1871.968	4.21	23	2		2.97E+03	0
☑	839	○	0	6.58	1	0.965	813.513	20.5	14.6	2		5.85E+03	1
☑	493	●	0	5.9	1	0.980	1839.925	5.37	15.6	2		9.25E+03	0
☑	156	●	0	7.62	3	0.987	1577.815	13.3	21.4	2		3.08E+04	0
☑	432	●	0	7.34	1	0.947	1754.971	16.8	19.3	3		1.41E+04	0
☑	123	●	1	6.26	1	0.992	1839.95	19.2	15.6	3		4.2E+04	0
☑	1387	●	0	---	---	0.996	1577.81	10.5	21.4	3		4.09E+03	0
☑	204	○	0	7.15	3	0.988	1415.738	16.7	14.5	2		2.72E+04	0
☑	3334	○	0	5.85	3	0.498	1415.723	6.24	14.5	3		570	0
☑	548	○	0	7.28	3	0.979	1577.804	6.59	21.9	2		8.35E+03	0
☑	3399	○	0	7.28	3	0.634	1577.8	4.27	21.9	3		641	0
☑	334	○	0	7.42	6	-0.975	1411.835	14	19.7	2		1.94E+04	0

Protein: P00924 Enolase 1 ENL

Protein: Q78HU7 Glycophorin-C OS=Mus musculus GN=Gypc PE=3 SV=2

Peptide ion views | Protein resolution

Conflicting proteins for peptide ion 1739

Accession	Peptides	Unique	Conflict:	Protein Score	Pe
P00924	23	22	1	211	☐ ---
Q78HU7	1	1	0	4.39	☑ ---

Peptide ions of Q78HU7

	#	Σ	⧉	Score	Hits	Correlation	Mass	Mass error (p...	RT (mins)	Charge	Tag	Abundance	Conflict:
☑	511	●	1	4.39	1	0.999	1871.993	-3.61	23	3		2.18E+04	0
☑	1739	●	0	---	---	0.999	1871.968	-16.6	23	2		2.97E+03	0

And then ADH

Accession	Peptides	Unique	Conflict:	Score	Tag
P00489	51	51	0	413	
Q8N9S3	3	3	0	13.1	
P00924	22	22	0	204	
Q78HU7	1	1	0	4.39	
P00330	20	20	1	144	
P07724	2	1	2	17.6	
P02769	41	40	2	322	
123456	44	43	1	368	
P62894 (+2)	11	11	0	64.9	
P02088 (+2)	2	2	0	13.6	
P00761	5	5	0	31.9	
Q8BKA3	1	1	0	0	
O70200	1	1	0	0	

	#	Σ	⧉	Score	Hits	Correlation	Mass	Mass error (p...	RT (mins)	Charge	Tag	Abundance	Conflict:
☐	511	○	1	7.1	4	---	1871.993	17.2	23	3		2.18E+04	0
☐	1739	○	0	---	---	---	1871.968	4.21	23	2		2.97E+03	0
☐	839	○	0	6.58	1	---	813.513	20.5	14.6	2		5.85E+03	0
☑	493	●	0	5.9	1	0.982	1839.925	5.37	15.6	2		9.25E+03	0
☑	156	●	0	7.62	3	0.987	1577.815	13.3	21.4	2		3.08E+04	0
☑	432	●	0	7.34	1	0.945	1754.971	16.8	19.3	3		1.41E+04	0
☑	123	●	1	6.26	1	0.990	1839.95	19.2	15.6	3		4.2E+04	0
☑	1387	●	0	---	---	0.997	1577.81	10.5	21.4	3		4.09E+03	0
☑	204	○	0	7.15	3	0.989	1415.738	16.7	14.5	2		2.72E+04	0
☑	3334	○	0	5.85	3	0.489	1415.723	6.24	14.5	3		570	0
☑	548	○	0	7.28	3	0.978	1577.804	6.59	21.9	2		8.35E+03	0
☑	3399	○	0	7.28	3	0.636	1577.8	4.27	21.9	3		641	0
☑	334	○	0	7.42	6	-0.972	1411.835	14	19.7	2		1.94E+04	0

Protein: P00924 Enolase 1 ENL

Protein: P00330 Alcohol dehydrogenase 1 ADH

Peptide ion views | Protein resolution

Conflicting proteins for peptide ion 839

Accession	Peptides	Unique	Conflict:	Protein Score	Pe
P00924	22	22	0	204	☐ 6.58
P00330	20	20	1	144	☑ ---

Peptide ions of P00330

	#	Σ	⧉	Score	Hits	Correlation	Mass	Mass error (p...	RT (mins)	Charge	Tag	Abundance	Conflic
☑	839	○	0	---	---	-0.955	813.513	20.5	14.6	2		5.85E+03	0
☑	820	○	0	5.37	1	-0.981	1311.661	-7.36	14.3	2		5.79E+03	1
☑	231	●	0	6.62	4	0.996	1250.685	18.2	15.8	2		2.6E+04	0
☑	1791	●	0	5.71	5	0.995	1446.807	7.21	19.5	3		2.05E+03	0

For the remaining conflict for ADH resolve in favour of the Antibody product 123456.

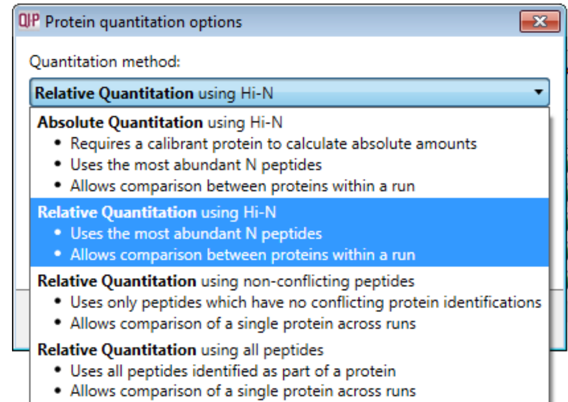
Finally resolve the 2 remaining conflicts between the Serum Albumin (spike) and the native Mouse Serum Albumin by favouring the mouse contaminant and un-ticking the conflicts for P02769.

Now there are no remaining conflicts return to page 17.

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

Protein Quantitation options

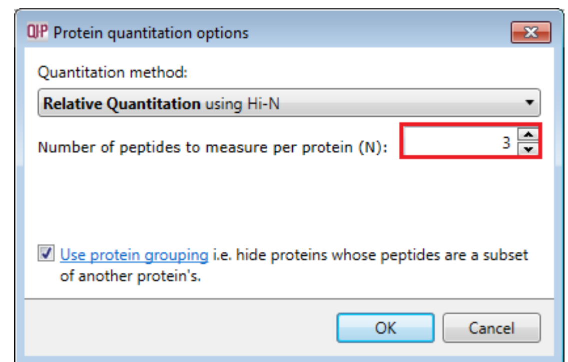
There are 4 options with which to control how the Protein Quantification is performed by Progenesis QI for proteomics.



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

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

Hi-N in Progenesis QI for proteomics is an implementation of Hi-3 as described by Silva *et al.* [**Error! eference source not found.**]. After peptide and protein identification, the abundance of each peptide is calculated from all its constituent peptide ions.

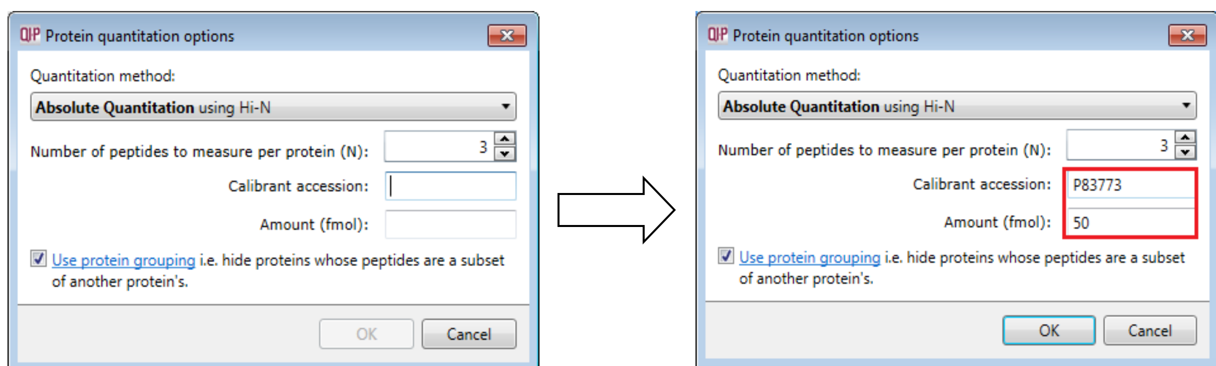


For each protein, the N most abundant peptides (N being set according to the user selection) have their **abundances averaged** to provide a reading for the protein signal.

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.

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.
Protein options...

No filter applied Create... Search Help

Description	Amount (fmol) - A	Amount (fmol) - B	Amount (fmol) - C
Glyceraldehyde 3-phosphate dehydrogenase GN=TDH3 PE=3 SV=1	576	838	794
Glycerol-3-phosphate dehydrogenase GN=GPD2 PE=3 SV=1	90	195	180
Glycerol-3-phosphate dehydrogenase_ mitochondrial GN=GUT2 PE=3 SV=1	31.2	36	35.9
Glycine cleavage system H protein GN=CAWG_00084 PE=4 SV=1	61.3	79.2	65.1

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

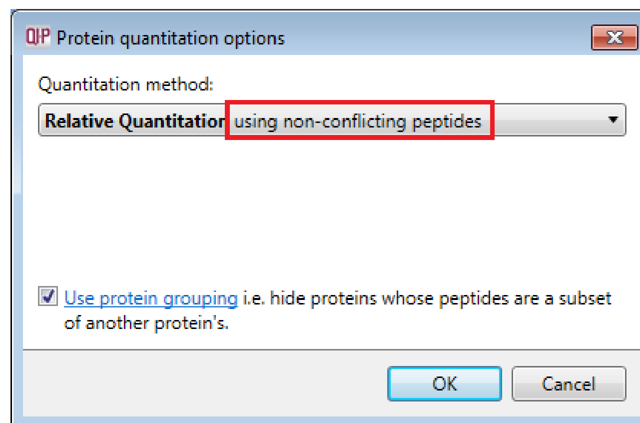
The ranking of peptide abundance is based on the integrated value across all the runs, allowed by the accurate alignment and lack of missing values. This gives added confidence in the peptide selection, taking all runs into account to make the ranking robust. These Hi-N methods allow the relative and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

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

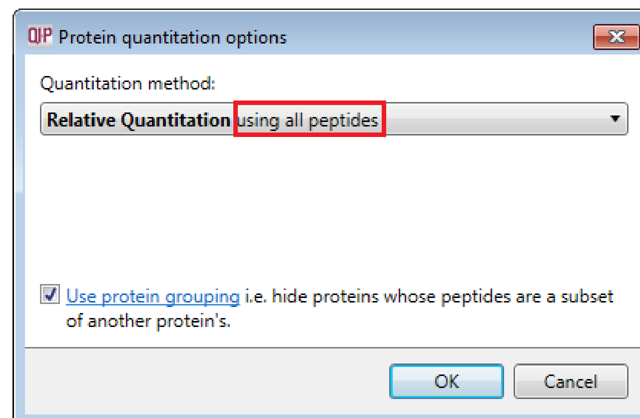
Naturally, if the conflicted peptides are not among the N most abundant in either protein initially, then this will not cause any difference in the result.

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

Using non conflicting peptides



The relative Quantitation can also be performed using all peptides.



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

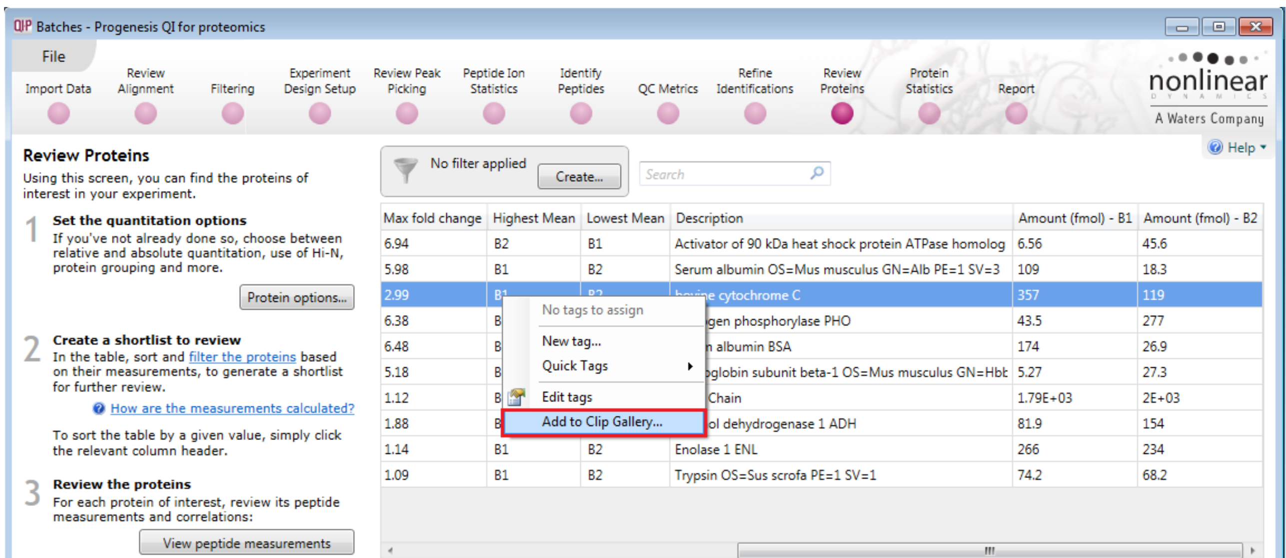
Now move to the Review Proteins section by clicking on **Review Proteins** icon on the workflow at the top of the screen.

Appendix 7: Using Clip Gallery to Save and Export Pictures and Data

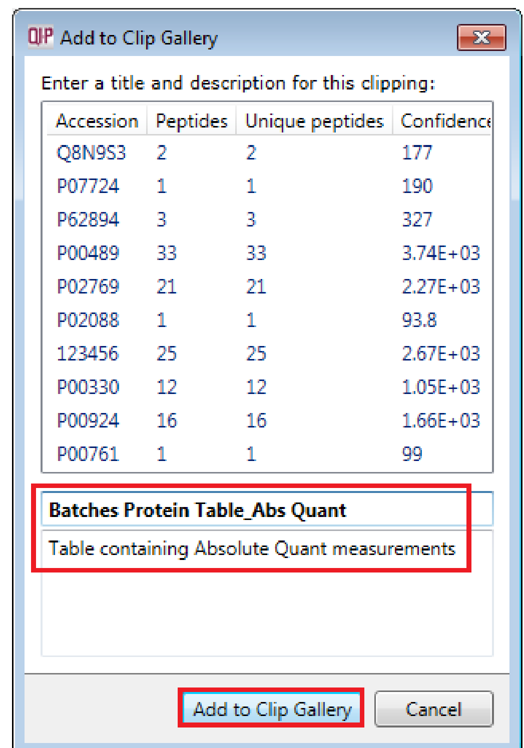
At every stage of the Progenesis QI for proteomics workflow the views and data tables can be added to the Clip Gallery.

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 **Review Proteins** stage right click on the Protein Table and select **Add to clip gallery…**



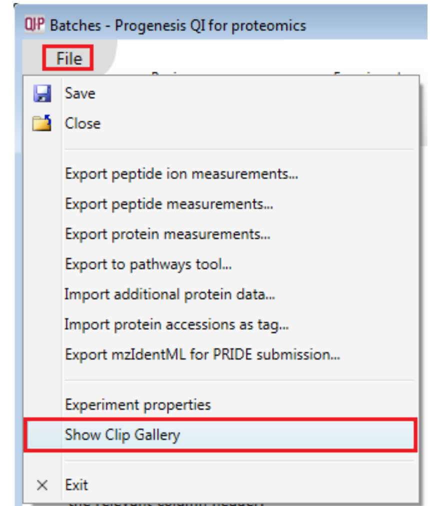
This will open a dialog displaying what is to be saved and allows you to alter the title and provide a description of the item for later reference.



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

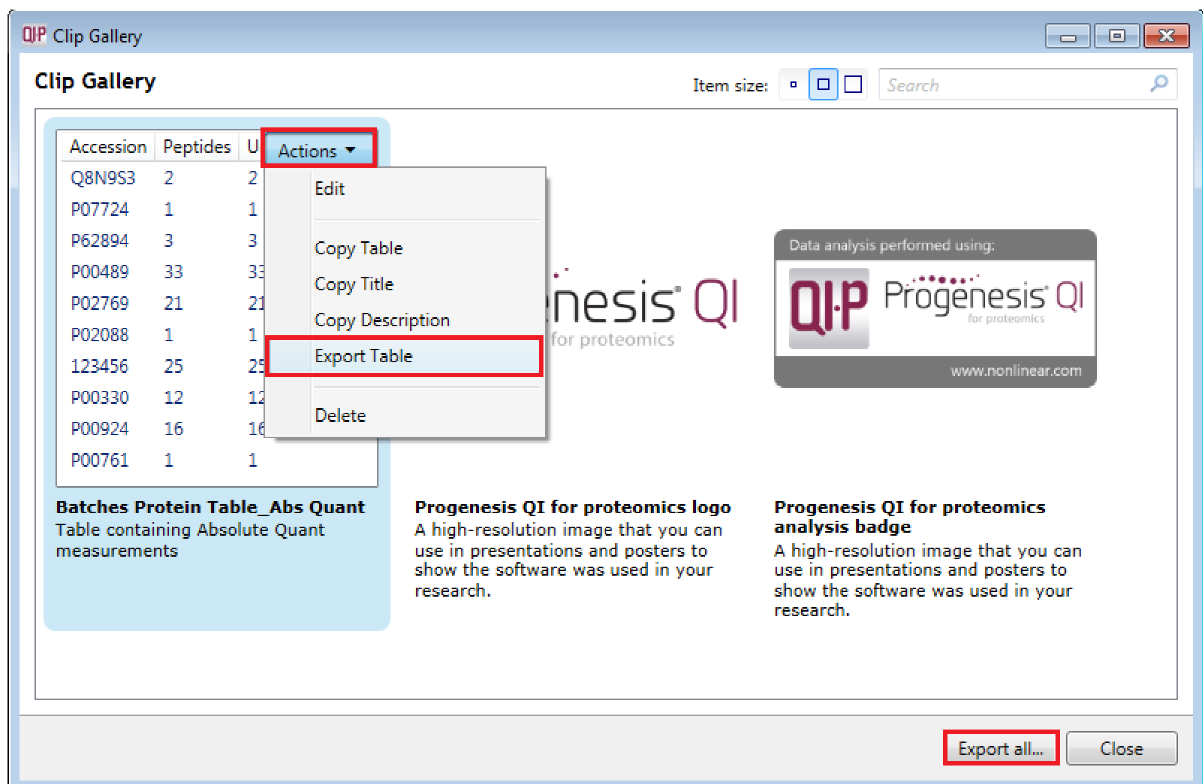
Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

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



Click **Show Clip Gallery**

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



Note: there is also the capacity to **Export all…** the items 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 item title.

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