

# **Progenesis Q1 for proteomics HCP Spectral Library User Guide**

**Analysis workflow guidelines for MS<sup>E</sup> 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 processing of the Create Library samples.....	6
Stage 2: Automatic Processing of your data .....	9
Stage 3A: Filtering .....	14
Stage 3B: Normalisation.....	15
Stage 4: Experiment Design Setup for Analysed Runs.....	16
Stage 5: Identify peptides .....	17
Stage 6: Refine Identifications.....	18
Stage 7 Select Peptide ions for Spectral Library at Resolve Conflicts.....	19
Stage 8 Review fragmentation quality of candidate spectra .....	23
Stage 9: Creating a Spectral Library .....	26
Stage 10: Autoprocessing of the Batch Samples .....	28
Stage 11: Identifying and measuring HCP proteins using a spectral library .....	32
Appendix 1: Stage 1 Data Import and QC review of LC-MS data set .....	38
Appendix 2: Stage 2 Processing failures.....	43
Appendix 3: Licensing runs .....	45
Appendix 4: Manual assistance of Alignment.....	46
Appendix 5: Within-subject Design.....	51
Appendix 6: Resolve Conflicts.....	53
Protein Quantitation options .....	56
Appendix 7: Using Clip Gallery to Save and Export Pictures and Data .....	59
References .....	60

## Introduction

This user guide takes you through how to construct a Spectral Library from the analysis of a set of 12 LC-MS runs with 4 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 5 Batches of additional samples. More detailed descriptions of each step in the analysis workflow is described in the DDA and HDMS<sub>e</sub> 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 38) then start at page 6.

**Note:** the images used in Appendices 1 and 4 refer to the HDMS<sub>e</sub> User guide data set.

## 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](mailto: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<sup>E</sup> parameters were set to 250:150 the default settings as defined in Appendix 1 (page 38). This was to done to reduce the time taken to demo the data analysis.

## Tutorial Data

The samples were originally generated to examine the use of Spectral Libraries to monitor the presence of Host Cell Proteins (HCPs) in batches of the NIST mAb product. The idea being to simulate contamination of the mAb preparation with Chinese Hamster Ovary (CHO) cell proteins and therefore to monitor Host Cell Proteins with high accuracy using mass spectrometry.

Label free LC-MS was performed using a 30min gradient on a CSH C18 (M-class ACQUITY UPLC) attached to a Xevo-G2XS

There are 2 data sets: HCP\_Spectral Library\_ **Created** and HCP\_mAb\_ **Batches** containing the samples as indicated below, run in triplicate.

CHO proteins ppm	Create	Batches
50		Batch_5
100	CreateA	
500		Batch_4
1000	CreateB	
5000		Batch_3
10000	CreateC	
25000		Batch_2
50000		Batch_1
100000	CreateD	

The idea of the data sets is to allow you to explore the processes involved in making spectral libraries and then their use in the identification and quantification of Host Cell Proteins using different Batches where the HCPs have been simulated by introducing different levels of CHO proteins.

The raw data has already been loaded into Progenesis QI for proteomics 4.2 in the form of 2 experiments and licenced archives of these experiments have been created:

HCP\_Spectral Library\_Created.ProgenesisQIPArchive

HCP mAb\_Batches.ProgenesisQIPArchive.

**Note:** these will open at the Identify Peptides screen when restored in Progenesis QI for proteomics

A fasta file, **NIST AB\_chaperone\_CHO.fasta**, is provided for identification of the proteins in the HCP\_Spectral Library Created experiment. The FASTA contains sequences for CHO cell proteins, the NIST mAb product and the internal spiked E.coli Chaperone protein ClpB which was added to all the samples (200 fmol/sample).

All samples contain 41 pmol of NIST mAb product protein.

**Note:** there are also 2 experiment design setup (Create.spl and mAb\_Batches.spl) files and 6 example Spectral Libraries .msp available.

The files described above are all available in the 2 Spectral Library Tutorial.zip files downloaded from <http://www.nonlinear.com/progenesis/qi-for-proteomics/v4.2/user-guide/>.

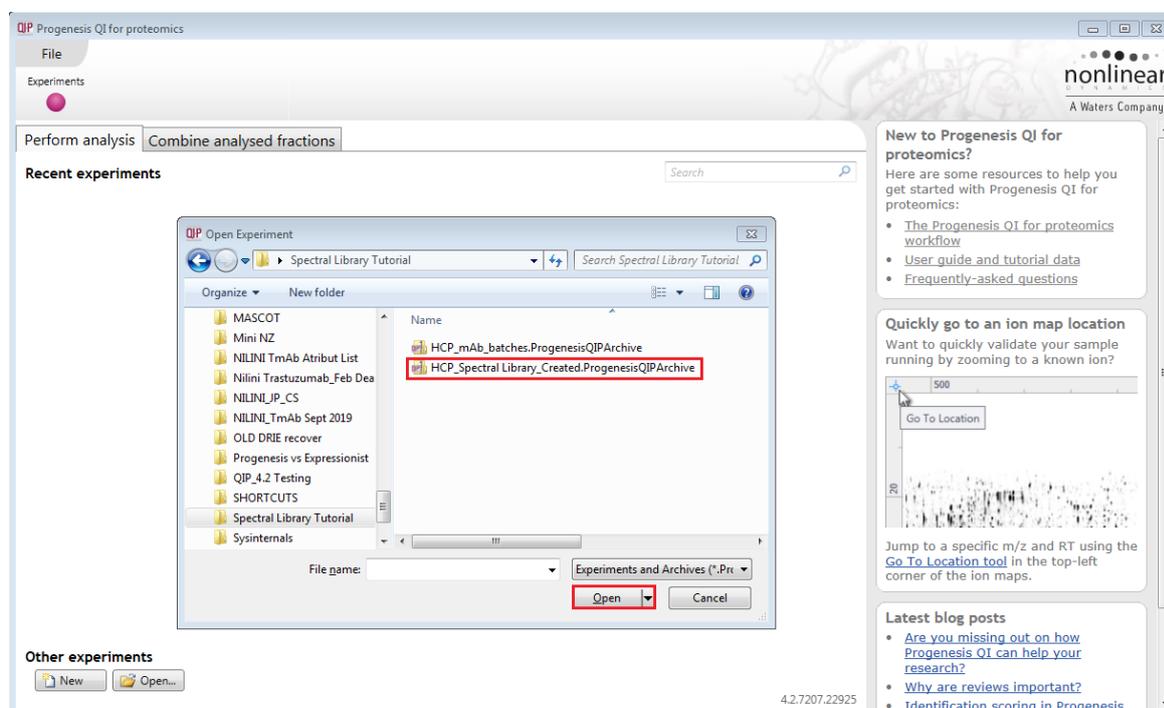
The following sections explain how to restore and process the data in Progenesis QI for proteomics v4.2.

## Restoring the Tutorial

Open Progenesis QI for proteomics and download the Compressed files (.zip), **Library\_creation\_data\_set.zip** and **Library\_search\_data\_set.zip** from the 'User guide and tutorial' link shown below, placing them in a **new folder** on your desktop (i.e. Spectral Library Tutorial). Before restoring the tutorial in the software **you must** first right click on the (.zip) files and extract them to the same folder.

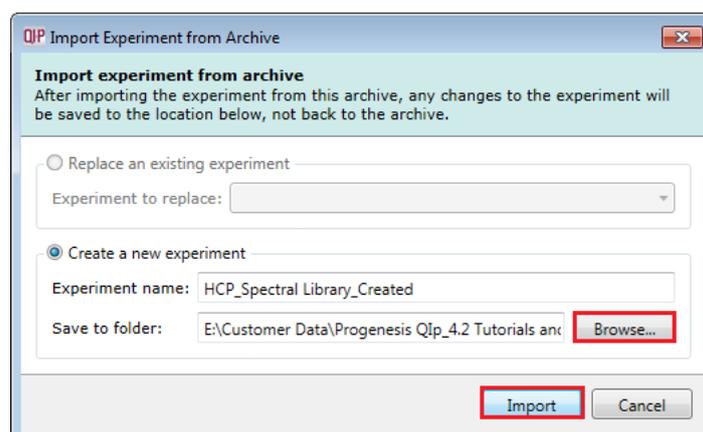
There are 2 archive files, 2 Grouping files, a fasta file and example Library files.

Now restore the 2 archive files (HCP\_Spectral Library\_Created and HCP\_mAb\_Batches). To do this, locate the '**name**'.**Progenesis QIP Archive** file using the **Open** button and press Open.



This opens the "Import Experiment from Archive" dialog.

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



Then click **Import**.

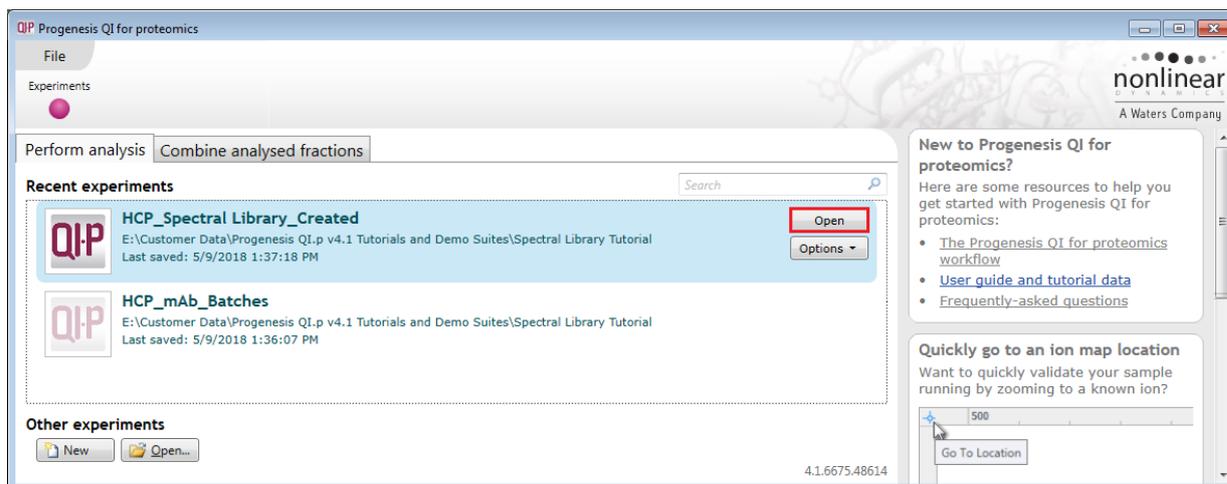
After the import is complete select **Close** from the **File** menu and then repeat the process for the other archive (Batches), restoring them into the same folder.

**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 processing of the Create Library samples

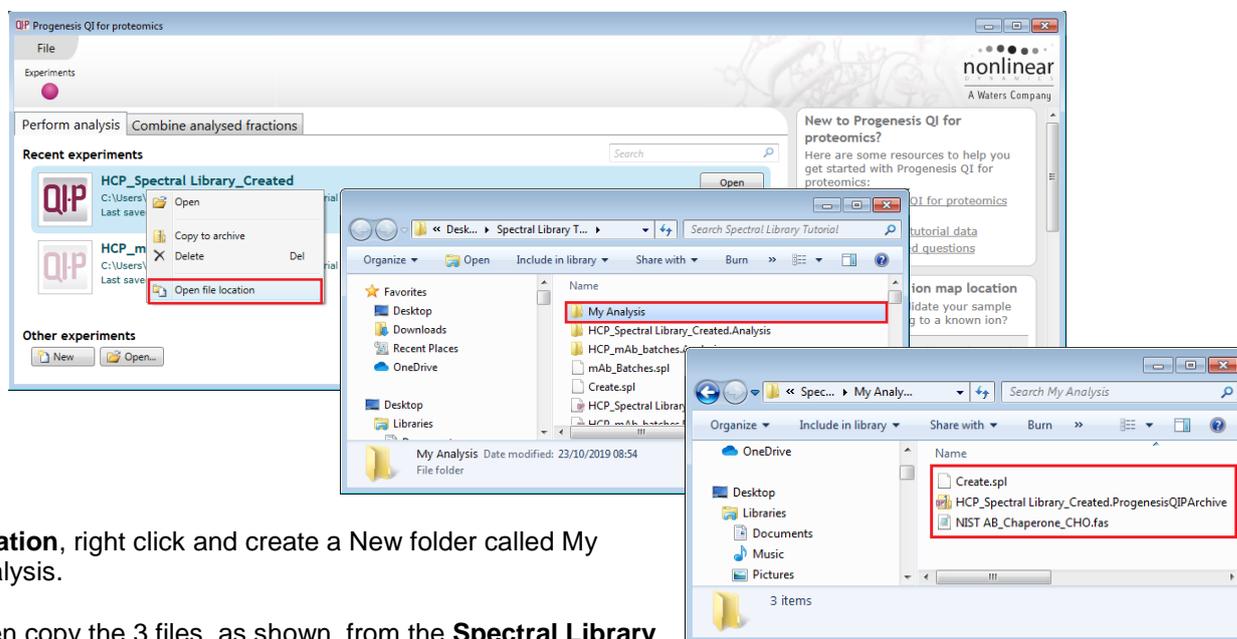
Having opened the Progenesis archive files the experiments will now be displayed when you open Progenesis QI for proteomics. Select the **HCP\_Spectral Library\_Created** experiment and click **Open**.



**Note:** having restored the **HCP\_Spectral Library\_Created** experiment, go to Stage 9 (page 26), where you can proceed with the creation of the various spectral libraries from the analysed experiment.

Alternatively, you can use the first 7 stages of this guide to perform the Automatic processing of the Created data set and then go through the processes involved in the selection and review of Peptide ions to add to the Spectral libraries. To do this you can generate a third experiment using the same **HCP\_Spectral Library\_Created** archive then remove the analysis by clicking **Restart automatic processing** as described below.

First you must create an additional folder called **My Analysis** in the same folder you restored the original archives, to do this right click on the **HCP\_Spectral Library\_Created** experiment and select **Open file**

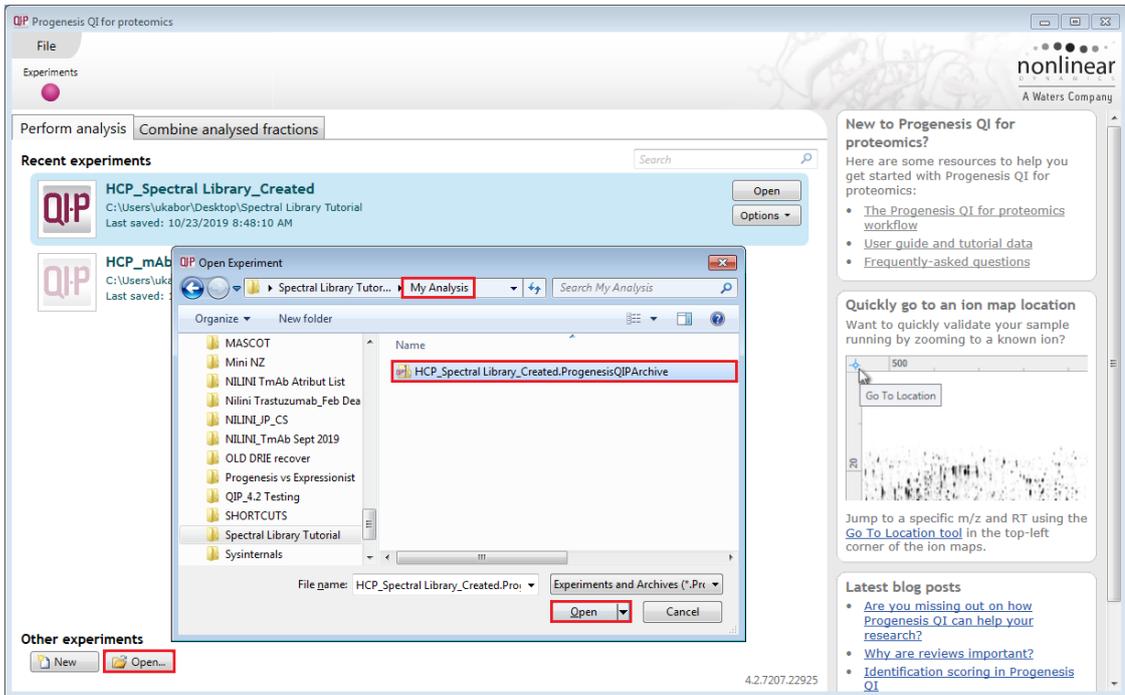


**location**, right click and create a New folder called My Analysis.

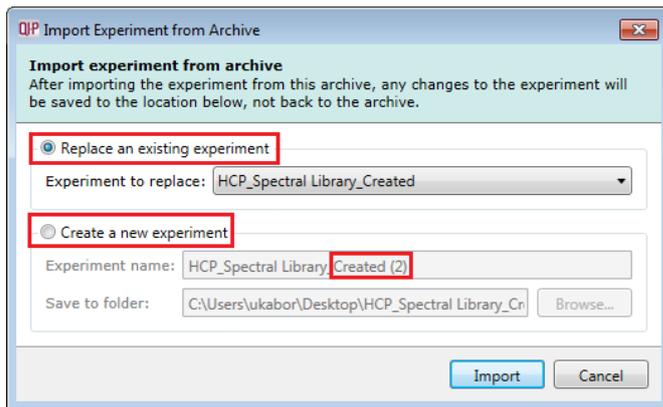
Then copy the 3 files, as shown, from the **Spectral Library Tutorial** folder to the **My Analysis** folder.

Now select **Open** on the Progenesis Experiments page and Navigate to the **My Analysis** folder.

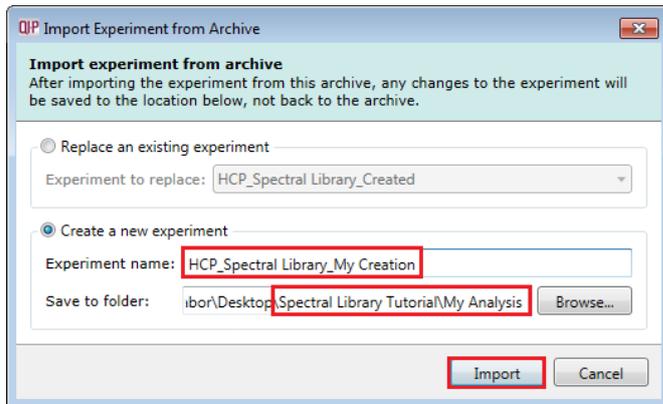
Select the **HCP\_Spectral Library\_Created.Progenesis QIPArchive** and click Open



The following dialog will open asking you to **Replace an existing experiment**, do not do this.



Select **Create a new experiment** and amend the experiment name to **HCP\_Spectral Library\_My Creation**



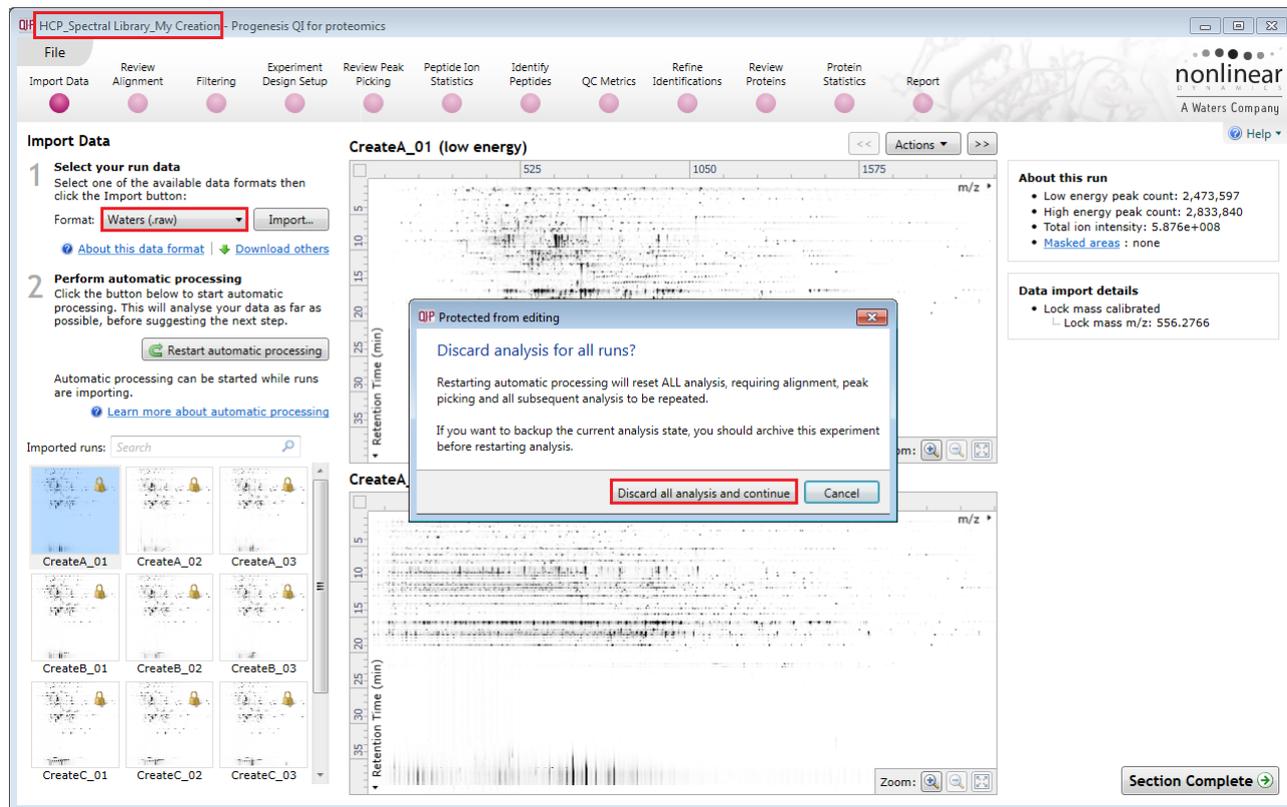
Then make sure that Save to folder is set to **My Analysis** using the **Browse** facility.

Then click **Import**

The newly created **HCP\_Spectral Library\_My Creation** experiment will open at the **Identify Peptides** stage on the workflow.

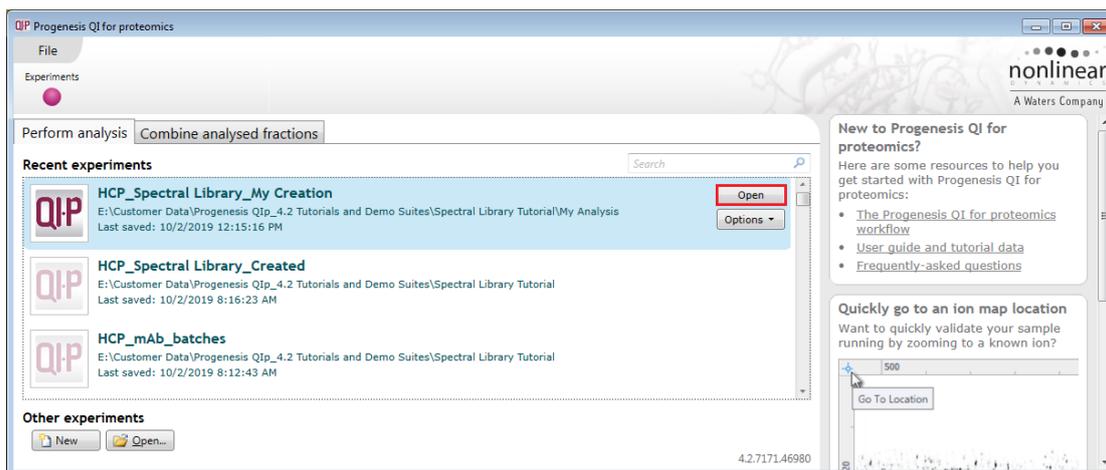
Click on **Import Data** and then click **Restart automatic processing**. This will bring up a warning dialog.

Select **Discard all analysis and continue** this will remove all existing analysis of this data set.



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

Click **Cancel** and the Close from the File menu (Top left of the Screen)



**Note:** the appearance of the new experiment in the list

The following pages describe how to proceed with the full analysis of this data.

Click **Open** and move to the next stage.

## 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**.

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

**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](#)

For the processing of the 12 files in this experiment make the following selections:

**Step 1:** Select the third option, **Use this run**. This tells QI to set the Alignment reference for to your chosen image. (In this example CreateD\_01)

QIP Start automatic processing

**Select an alignment reference**  
To compensate for drifts in retention time, all runs in the experiment must be aligned to a single reference run.

How do you want to choose your alignment reference?

- Assess all runs in the experiment for suitability
- Use the most suitable run from candidates that I select
- Use this run:  
CreateD\_01

For information on choosing the alignment reference, and why you might want to select your own candidates, please see the [online guidance](#).

< Back   **Next >**   Cancel

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

QIP Start automatic processing

**Automatic alignment**  
After selecting the experiment's alignment reference, the software can also automatically align all runs.

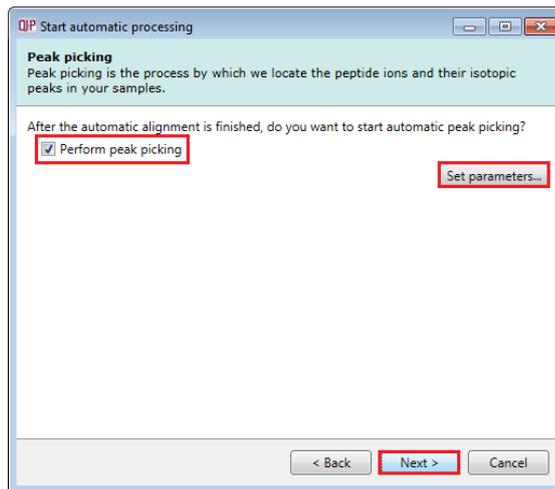
After the alignment reference is chosen, do you want to start automatic alignment?

Yes, automatically align my runs

< Back   **Next >**   Cancel

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

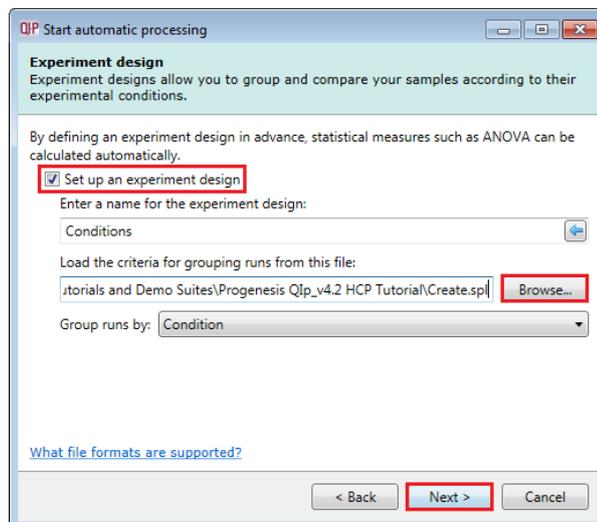
**Step 2:** For the purposes of this User guide we will use the default settings for peak picking which you can review by clicking on Set parameters.



Click **Next**.

**Step 3:** To apply the experiment design in advance Click **Set up an experiment design** then locate the appropriate file (**Create.spl** 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 Create.spl, in this case Conditions, this can be changed as required in the main workflow.



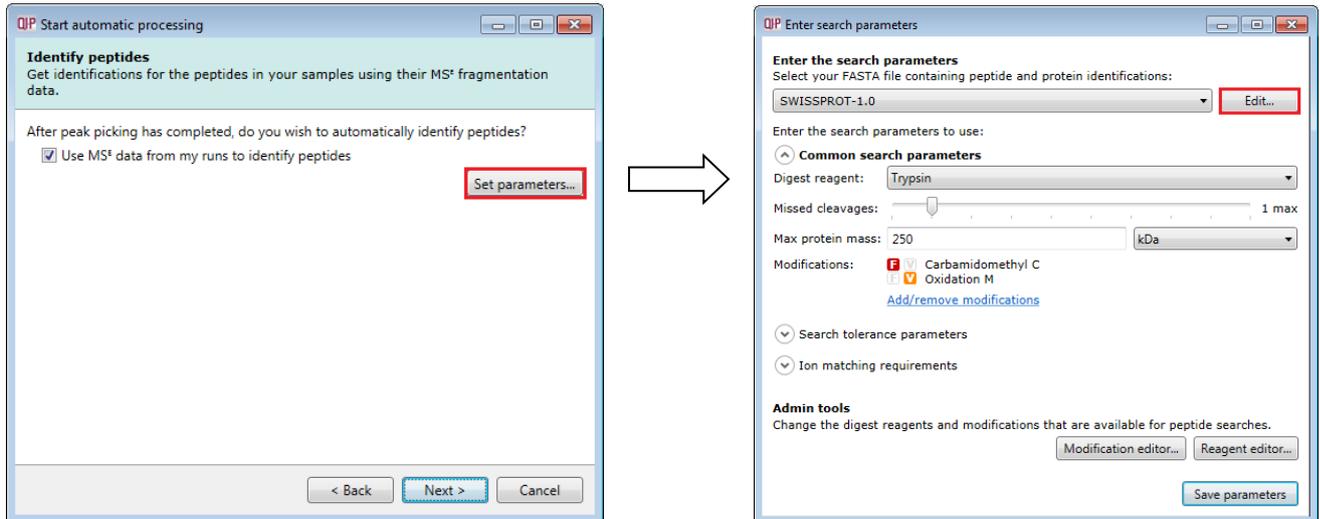
**Note:** you can also create additional experiment designs following the completion of automatic processing.

Select Conditions and then click **Next**.

For MS<sup>E</sup> fragmentation data you can set up the peptide identification to be performed automatically.

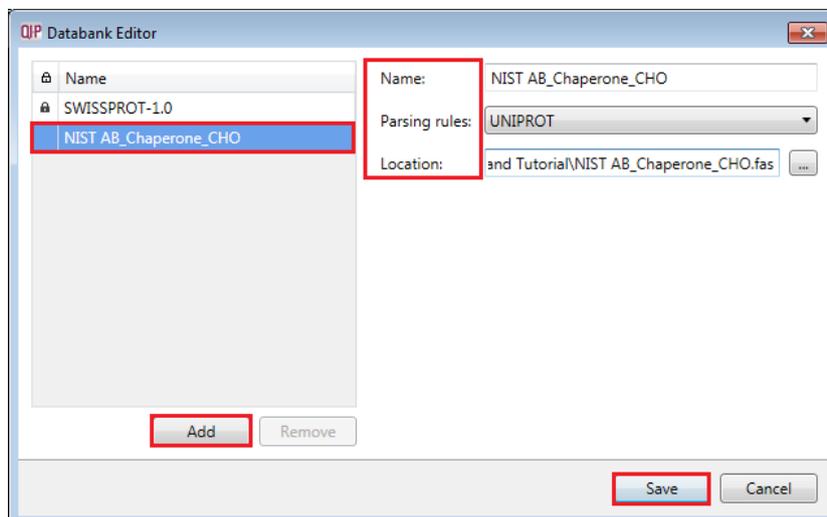
**Step 4:** To automatically process the identification of peptides using their MS<sup>E</sup> fragmentation data in the tutorial data click on **Set parameters**. The default Databank is for Swissprot-1.0 (which is a locked example).

**Note:** if the software has detected the presence of MS<sup>E</sup> 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 (**NIST AB\_Chaperone\_CHO.fas**).

For a new Databank click **Add**. Then give it name (i.e. NIST AB\_Chaperone\_CHO), 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...**

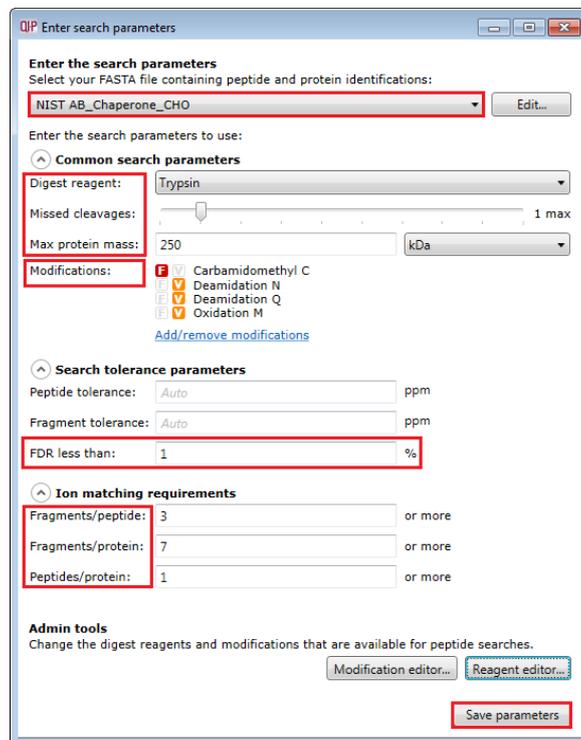
**Missed cleavages:** is set as 1.

**Maximum protein mass:** is set at 250kDa

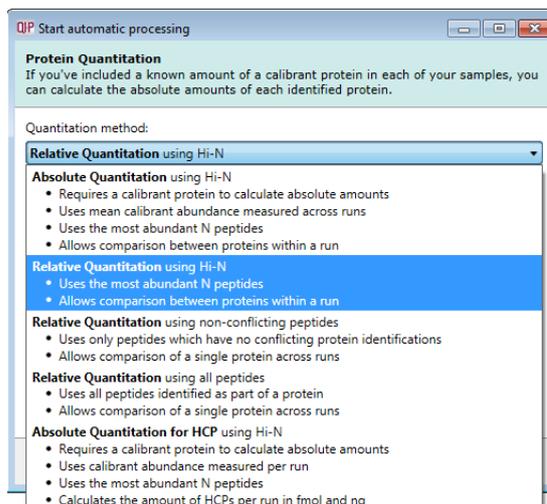
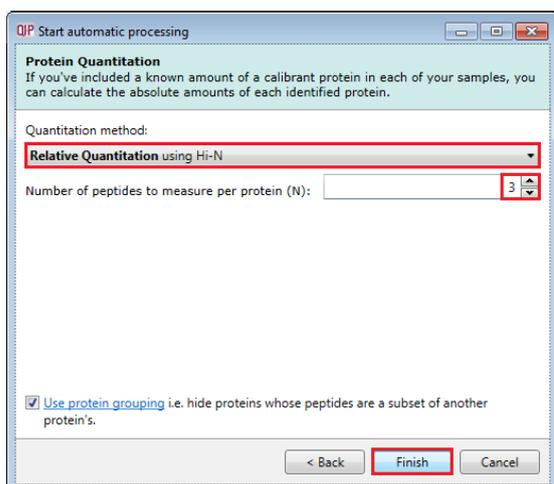
**Modifications:** are set Carbamidomethyl C (Fixed) and Oxidation M (Variable). More modifications are available from the list and additional ones can be added to the list using the **Modification editor...** In this example Deamidation on N and Q have been added

**Search Tolerance parameters:** The false discovery rate (FDR) has been set to 1%

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



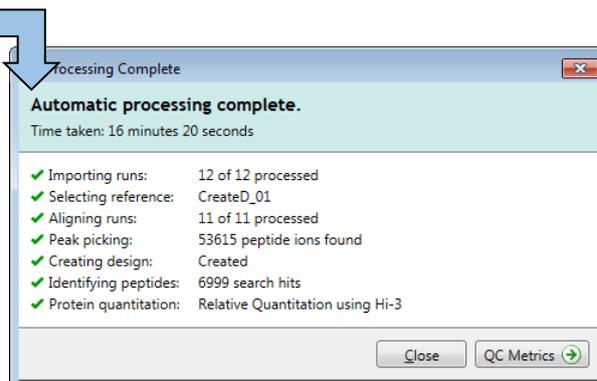
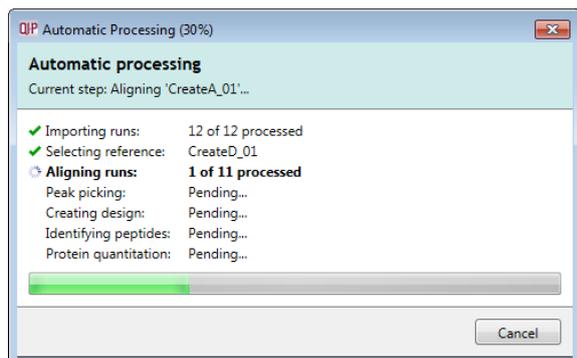
Click **Save parameters** and then **Next**.



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**.

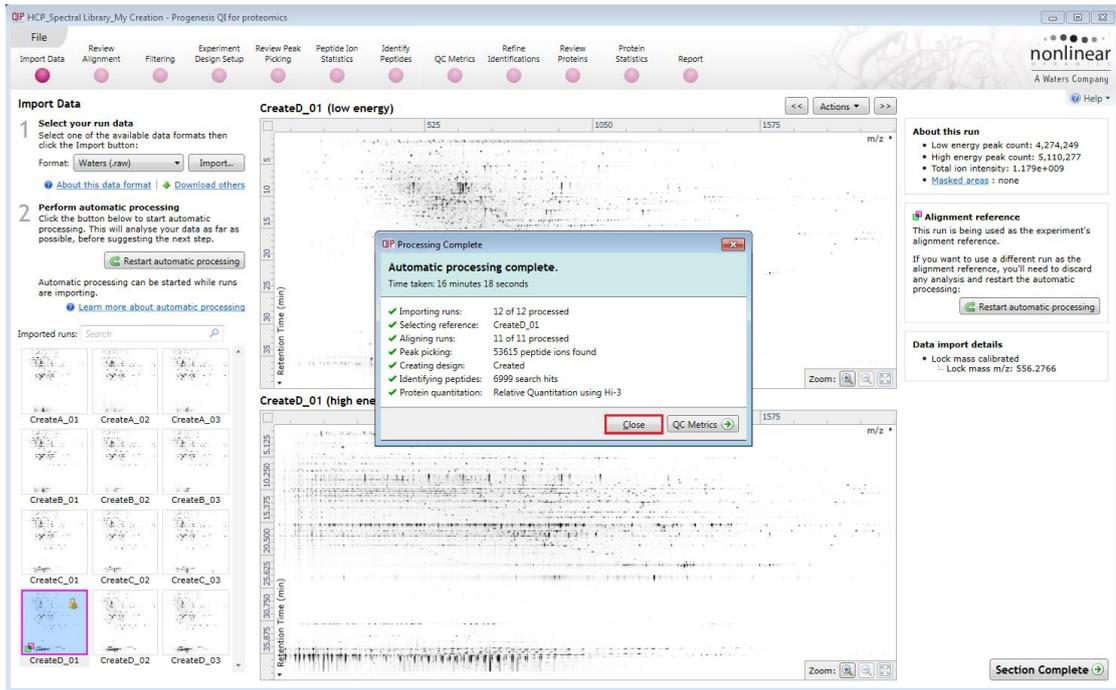
The process starts with the selection of an alignment reference and completes with Protein quantitation.



As the whole process continues 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 outlined in Appendix 2 (page 43).

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



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

**Note:** you can explore and re-perform the steps, sequentially and/or as part of the automatic processing as described in this guide

**Note:** additional details on alignment are available in Appendix 4 (page 46).

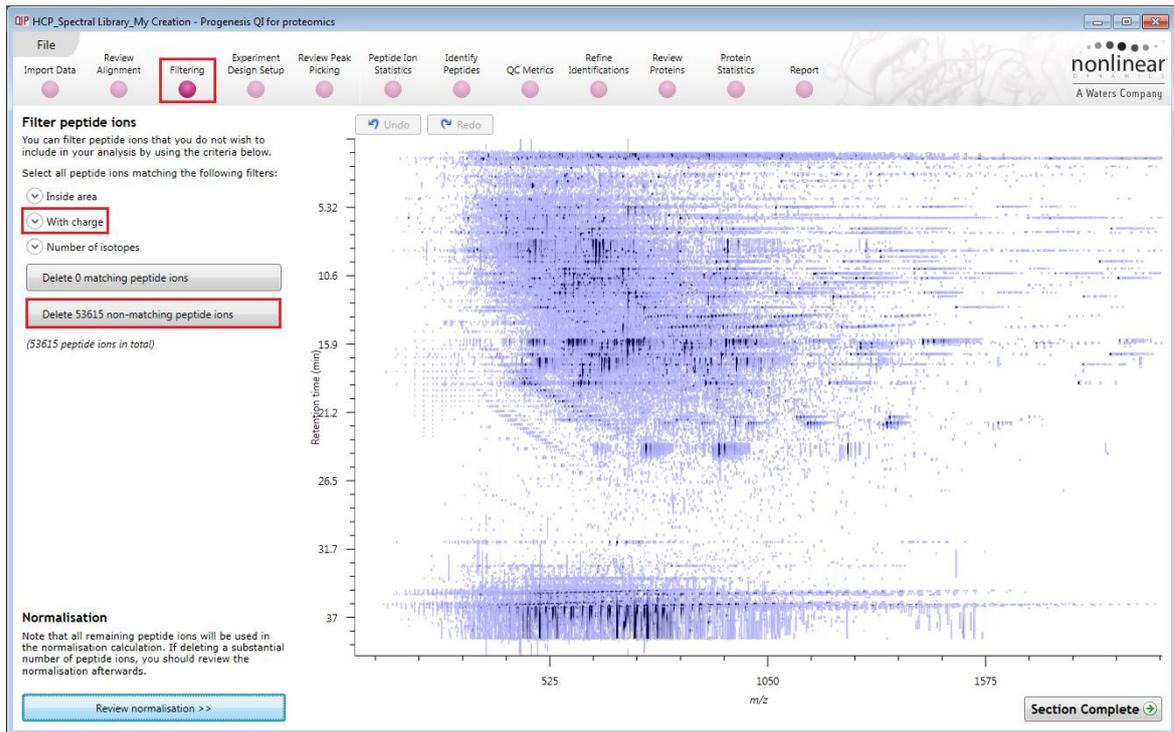
For this data set, the quality of the alignment (as indicated by the high scores) and detection following automatic processing does not require to be re-performed.



Click on **Filtering** on the workflow to move to the next section.

## Stage 3A: Filtering

At Filtering you can review the total number of peptide ions detected (53,615) and choose to keep all or those which meet certain filtering criteria.

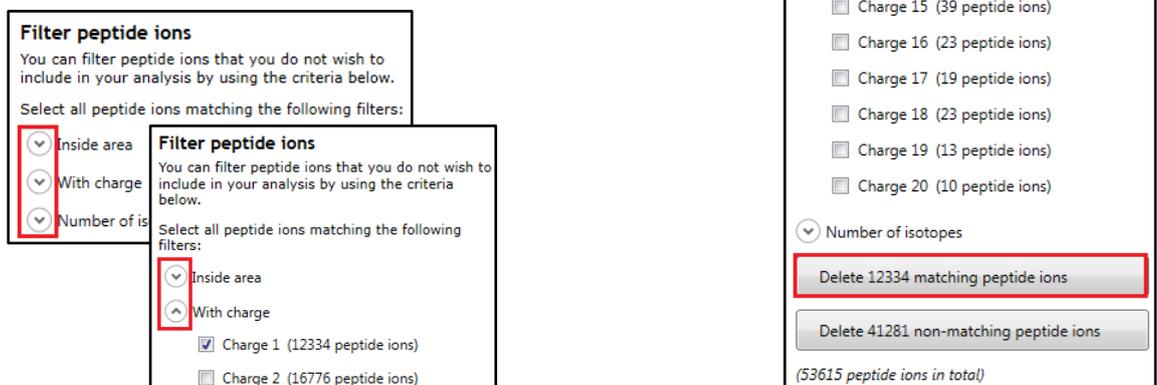


For this particular data set we will reduce the peptide ions to having a 'charge state' of 2 and above by filtering out the Charge state 1 peptide ions.

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 41,281 peptide ions with a range of charge states 2 to 20.

**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.



Now move on to **Normalisation** by clicking on 'Review normalisation', bottom left of the Filter page.

## Stage 3B: Normalisation

At this stage in the workflow you can review and change the normalisation method; by default it will be set to normalise to all proteins.

To change it click on the Normalisation Method tab and select **Normalise to a set of housekeeping proteins** from the drop down. Enter 'Light chain' in the proteins you want to normalise against. Select all of the peptide ions for this protein by selecting all the rows in the table and ticking the **Use** box.

**Review normalisation**  
Normalisation is required to allow comparisons across different sample runs.

By assuming that a significant number of peptide ions are unaffected by experimental conditions, we can use the factor by which the sample as a whole varies to normalise back to its reference.\*

Note: for each sample, only the peptide ions falling within its robust estimation limits (see graphs) are used to calculate the normalisation factor. Further details of [how it is calculated](#) are available online.

\* Normalisation reference:  
CreateC\_02

**Normalisation factors:**

Run	Factor	Log(factor)
CreateA_01	1.00	-0.001
CreateA_02	1.02	0.0072
CreateA_03	1.01	0.0037
CreateB_01	0.97	-0.012
CreateB_02	0.93	-0.032
CreateB_03	0.94	-0.025
CreateC_01	1.05	0.021
CreateC_02	1.00	0
CreateC_03	1.04	0.015
CreateD_01	0.84	-0.076
CreateD_02	0.82	-0.084
CreateD_03	0.88	-0.057

Normalisation Graphs | **Normalisation Method**

Normalise to a set of housekeeping proteins

Check the housekeeping proteins you want to normalise against: Light Chain

Use?	#	m/z	z	Mass	Retention time	Tag	Accession	Protein Description
<input checked="" type="checkbox"/>	1	711.8746	4	2843.4693	15.55		867100	Light Chain
<input checked="" type="checkbox"/>	2	937.4697	2	1872.9249	16.69		867100	Light Chain
<input checked="" type="checkbox"/>	3	946.9597	2	1891.9048	10.76		867100	Light Chain
<input checked="" type="checkbox"/>	5	948.8270	3	2843.4592	15.57		867100	Light Chain
<input checked="" type="checkbox"/>	6	631.6429	3	1891.9069	10.75		867100	Light Chain
<input checked="" type="checkbox"/>	7	701.0759	4	2800.2744	10.03		867100	Light Chain
<input checked="" type="checkbox"/>	8	625.3162	3	1872.9268	16.63		867100	Light Chain
<input checked="" type="checkbox"/>	9	561.0633	5	2800.2802	10.03		867100	Light Chain
<input checked="" type="checkbox"/>	12	934.4296	3	2800.2669	10.02		867100	Light Chain
<input checked="" type="checkbox"/>	13	904.5116	2	1807.0087	17.44		867100	Light Chain
<input checked="" type="checkbox"/>	14	924.9028	2	1847.7909	9.46		867100	Light Chain
<input checked="" type="checkbox"/>	15	848.7167	3	2543.1283	15.80		867100	Light Chain
<input checked="" type="checkbox"/>	19	1272.5750	2	2543.1355	15.80		867100	Light Chain
<input checked="" type="checkbox"/>	20	593.8331	2	1185.6517	11.04		867100	Light Chain
<input checked="" type="checkbox"/>	22	938.4740	2	1874.9334	6.92		867100	Light Chain
<input checked="" type="checkbox"/>	27	661.3477	2	1320.6808	9.10		867100	Light Chain
<input checked="" type="checkbox"/>	31	476.7766	2	951.5386	8.65		867100	Light Chain
<input checked="" type="checkbox"/>	37	557.8091	4	2227.2073	15.66		867100	Light Chain
<input checked="" type="checkbox"/>	38	712.6654	3	2134.9745	5.23		867100	Light Chain
<input checked="" type="checkbox"/>	39	603.6716	3	1807.9930	17.85		867100	Light Chain
<input checked="" type="checkbox"/>	40	559.9427	3	1676.8063	11.30		867100	Light Chain
<input checked="" type="checkbox"/>	42	1068.4944	2	2134.9742	5.23		867100	Light Chain
<input checked="" type="checkbox"/>	43	899.4560	2	1796.8975	18.87		867100	Light Chain

<< Continue filtering peptide ions

**Note:** the Normalisation factors will update, reflecting the use of the selected protein for normalisation.

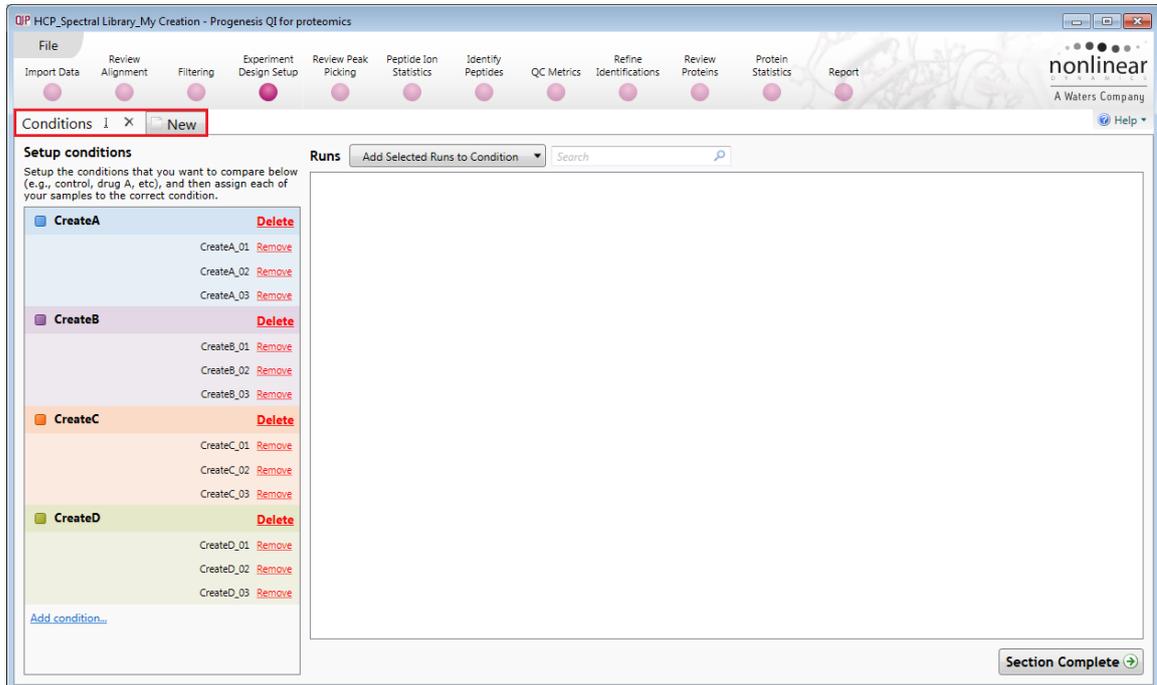
**Note:** As the same amount of mAb protein (41 pmol) is present in each sample the normalisation was performed against the peptide ions for this protein.

Now move on to **Experiment design setup** by clicking on the icon on the workflow.

## 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 4 conditions: CreateA to D (CHO proteins spiked @ 100 ppm to 100000ppm).



**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.



## 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.

The data described in this tutorial was acquired on a Waters Xevo G2XS it is MS<sup>E</sup> and therefore the peptide identification method used was **Ion accounting** as shown below.

**Identify Peptides**  
Select your peptide identification method:  
**Ion Accounting**

**Peptide ions (2094 identified)**

#	Identifications	m/z	Charge	Retention time	Tag
1	1	711.8746	4	15.55	
2	1	937.4697	2	16.69	
3	1	946.9597	2	10.76	
5	1	948.8270	3	15.57	
6	1	631.6429	3	10.75	
7	1	701.0759	4	10.03	
8	1	625.3162	3	16.63	
9	1	561.0633	5	10.03	

**Identifications for peptide ion 1**

Peak mass	Peptide mass	Protein mass	Mass Error (Da)	Mass Error (ppm)	Score	Seq. start	Seq. end	Sequence
2843.469	2843.450	23412.939	0.0190	6.6697	9.515	226	251	THTCPCPCAPPELLGGPSVFLFPPKPKK

**Fragment matches for: THTCPCPCAPPELLGGPSVFLFPPKPKK**

Intensity (counts) vs m/z plot showing fragmentation patterns.

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 9) of this guide, the Identify Peptides page currently displays the full list of the detected peptide ions in your experiment.

**Note:** At the Filtering stage you will have removed the charge state 1 peptide ions, so you should re-perform the search using the parameters shown on the right.

There will be 2094 identified.

The left hand panel will display the Fasta File used (NIST\_AB\_Chaperone\_CHO) in the search and the parameters and settings used to control the search.

**For MS<sup>E</sup>, HDMS<sup>E</sup> and SONAR data**

**1 Enter the search parameters**  
Select your FASTA file containing peptide and protein identifications:  
**NIST\_AB\_Chaperone\_CHO** Edit...

Enter the search parameters to use:  
**Common search parameters**

Digest reagent: **Trypsin**

Missed cleavages: 1 max

Max protein mass: 250 kDa

Modifications: Carbamidomethyl C, Deamidation N, Deamidation Q, Oxidation M

**Search tolerance parameters**

Peptide tolerance: Auto ppm

Fragment tolerance: Auto ppm

**FDR less than:** 1 %

**Ion matching requirements**

**Fragments/peptide:** 3 or more

**Fragments/protein:** 7 or more

**Peptides/protein:** 1 or more

## Stage 6: Refine Identifications

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

**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 5
- Hits less than 2
- Absolute mass error (ppm) greater than 10

To perform these filters, on the Batch detection options panel, set the Score to less than 5, 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 5

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

Modifications: contains

Delete matching search results  Delete non-matching search results

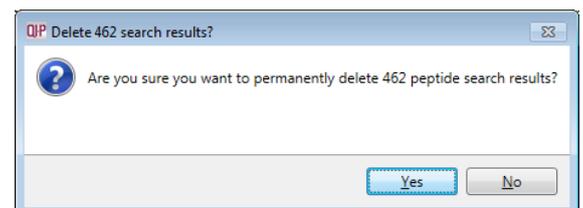
#	Score	Hits	m/z	RT(mins)	Charge	Mass	Mass err	Sequence	Accession	Modifications	
115	10.13	12	905.68	12.17	4	3618.61	-3.55	VDNALQ...	867100		Light Chain
116	0.00	1	836.42	16.63	2	1670.81	3.28	PPVLSDG...	867100		Light Chain
117	10.13	12	1207.24	12.17	3	3618.71	-0.67	VDNALQ...	867100		Light Chain
118	8.41	6	419.76	6.09	2	837.51	14.07	ALPAPIEK	867100		Light Chain
120	8.59	4	418.23	7.19	2	834.44	15.57	DTLMISR	867100		Light Chain
128	---	---	1422.74	15.57	2	2843.41	-1.11	LGGTIDD...	G3HDR3	[8] Carbamidomethyl C	T-complex prote
128	9.45	5	1422.74	15.57	2	2843.41	5.78	THTCPPCP...	867100	[4] Carbamidomethyl C[7] Carbamidomethyl C	Light Chain
130	9.19	5	811.40	5.86	2	1620.71	1.95	VGYMHWY...	867100		Light Chain
131	8.98	6	558.05	16.26	4	2228.11	-2.16	VVSVLTVL...	867100	[14] Deamidation N	Light Chain
132	8.62	8	581.81	12.13	2	1161.61	2.30	NQVSLTCL...	867100	[1] Deamidation N[7] Carbamidomethyl C	Light Chain
145	9.44	12	609.68	9.00	5	3043.31	-4.69	SRWQQGN...	867100	[11] Carbamidomethyl C	Light Chain
158	0.00	1	912.03	15.57	2	1822.04	2.83	PELLGGPS...	867100		Light Chain
159	4.76	1	807.40	6.91	2	1612.71	-2.61	SHAFTIHL...	G3H319	[10] Carbamidomethyl C	Kinesin-like prot
160	9.01	3	406.21	5.86	4	1620.81	8.85	VGYMHWY...	867100		Light Chain
175	9.07	12	743.41	15.66	3	2227.21	1.66	VVSVLTVL...	867100		Light Chain
189	---	---	501.59	13.00	3	1501.71	8.28	HTFVESDE...	G31216		Triosephosphate
189	8.96	4	501.59	13.00	3	1501.71	1.69	DSTVSLST...	867100		Light Chain
198	5.91	1	981.87	14.53	5	4904.31	5.36	IVGGYTCA...	P00761	[7] Carbamidomethyl C[25] Carbamidomethyl	Trypsin OS=Sus
198	6.17	1	981.87	14.53	5	4904.31	5.36	IVGGYTCA...	P00761	[7] Carbamidomethyl C[15] Deamidation Q[25]	Trypsin OS=Sus
199	6.17	1	818.39	14.53	6	4904.31	4.85	IVGGYTCA...	P00761	[7] Carbamidomethyl C[15] Deamidation Q[25]	Trypsin OS=Sus
199	5.91	1	818.39	14.53	6	4904.31	4.85	IVGGYTCA...	P00761	[7] Carbamidomethyl C[25] Carbamidomethyl	Trypsin OS=Sus
206	9.59	6	1273.01	16.06	2	2544.11	2.12	GFYPSDIA...	867100	[14] Deamidation N	Light Chain
206	9.34	4	1273.01	16.06	2	2544.11	2.12	GFYPSDIA...	867100	[16] Deamidation Q	Light Chain
220	8.51	7	429.73	2.16	2	857.44	1.64	HYNPSLK	867100		Light Chain

2867 search results. 462 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: 462 matching out of 2867)

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



Then click **Reset the criteria** and enter the next criteria and repeat the process

Having applied the 3 filters there will be **1320** search results remaining

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

## Stage 7 Select Peptide ions for Spectral Library at 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.

In addition to indicating the number of conflicting peptide identifications other tools have been introduced at this stage in the workflow that will facilitate the review and selection of peptide ions (Spectra) to be entered into the spectral library.

This allows you to choose spectra on the basis of:

- Unique peptide ions, no alternative conflicting peptide identifications
- No overlapping peptide ions
- Correlation of the peptide ions expression profile with other peptide ions identified for the same protein

The **Resolve Conflicts** stage can be accessed at the bottom left of the **Refine Identifications** stage.

The screenshot displays the Progenesis QI software interface. The top navigation bar includes 'File', '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 'Refine Identifications' stage is active, showing a table of peptide ions for protein G3I8R9. The table has columns for Accession, Peptides, Unique, Conflict, Score, Tag, Abundance, #, Σ, Correlation, Mass, Mass error (p...), RT (mins), Charge, Tag, Abundance, Conflict, and Peptide Sequence. The 'Protein: G3I8R9 78 kDa glucose-regulated protein OS=Cricetulus griseus GN=I79\_019946 PE=1 SV=1' is selected. Below the table, the 'Standardised Expression Profiles' and 'Peptide ions of selected protein' plots are shown. The 'Standardised Expression Profiles' plot shows a line graph of Standardised Normalised Abundance versus Retention time (min) for four conditions: CreateA, CreateB, CreateC, and CreateD. The 'Peptide ions of selected protein' plot shows a heatmap of Retention time (min) versus m/z for the selected protein.

**Note:** If you decide to resolve all the conflicts before applying the selection strategy described below then refer to Appendix 6 (page 53)

For this tutorial, Conflict resolution was **NOT** performed.

The current protein, G3I8R9 (left hand table), has a total number of 19 peptides, 17 of which are uniquely assigned to this protein, the remaining 2 have a total of 12 conflicting protein assignments.

A possible strategy for the selection of peptide ions for the current protein is to use the tools in the peptide ion table to order and select only those peptide ions that have:

- No overlapping peptide ions
- No conflicts
- Good correlation (i.e  $>0.925$ ) of the peptide ion's expression with all the other ions identified for this protein

To achieve this you will make use of tagging the 'candidate' ions to be added to the library.

In this example data set we will create 2 spectral libraries:

- A) **HCP library\_All** which will contain all the selected peptide ions for all the Host Cell proteins
- B) **NIST\_Product Library** which contains all the selected peptide ions for the mAb product

Starting with the HCP library\_All we will continue with the G318R9 protein, as described above. First order the peptide ion table on Overlapping peptide ions and highlight all of the zero entries.

The screenshot shows the Progenesis QI software interface. 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 main window is titled "HCP\_Spectral Library\_My Creation - Progenesis QI for proteomics".

Under the "Proteins" section, "G318R9" is selected. Below it, the "Peptide ions of G318R9" table is displayed. The table has columns for #, I, Score, Hits, Correlation, Mass, Mass error (p...), RT (mins), Charge, Tag, Abundance, Conflict, and Peptide Sequence. Several rows are highlighted, with the "Conflict" column showing values of 0, 2, 0, 4, 0, 0, 0, 0, 0.

Below the table, there are three visualization panels:
 

- Standardised Expression Profiles:** A line graph showing abundance across four conditions (CreateA, CreateB, CreateC, CreateD). The y-axis is "Standardised Normalised Abundance" ranging from -1.0 to 2.5.
- Peptide ions of selected protein:** A plot of Retention time (min) vs m/z. The y-axis ranges from 19.907 to 20.125, and the x-axis ranges from 1074 to 1078.
- Peptide ions of selected protein:** A mass spectrum plot of Retention time (min) vs m/z. The y-axis ranges from 10.603 to 31.745, and the x-axis ranges from 500 to 2000.

At the bottom, there are buttons for "Refine Identifications" and "Protein options...", and a "Section Complete" indicator.

**Note:** the good correlation of the expression of the highlighted peptide ions as shown in the bottom left graph as well as the **Correlation** scores highlighted in the peptide ion table.

This is a close-up view of the "Peptide ions of G318R9" table. The table lists various peptide ions with their corresponding scores, hits, and correlation values. The "Conflict" column is highlighted, showing values of 0, 2, 0, 0, 2, 0, 0, 0, 0.

#	I	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict	Peptide Sequence
5276	2	8.19	2	0.999	1396.78	-1.22	13.7	2		2.28E-04	0	ELEEIVQPIIS
3999	2	8.07	2	0.998	1814.98	-4.74	12	3		2.91E-04	0	IINEPTAAAIA
36326	2	---	---	0.997	1315.63	0.323	11.2	3		1.02E-03	0	NELESYAYSLK
13168	2	7.56	2	0.999	2147.995	2.58	20	3		1.32E-04	0	IETESFFEGED
1938	2	9.09	2	0.999	1535.792	1.09	16.7	2		8.93E-04	0	TFAPEEISAMV
4738	2	7.66	2	0.997	980.477	-4.32	5.54	2		1.25E-04	0	ETAAYLGM
2724	2	8.29	2	0.999	1524.87	-4.03	11.3	3		4.14E-04	0	KELEEIVQPII
2726	2	8.95	2	1.000	1459.751	-0.819	13.5	2		4.46E-04	0	SDIDEIVLVGG

**Note:** if you highlight a peptide ion with overlapping peptide ions, these can be displayed in the bottom right panel by selecting 'Overlapping peptide ions' from the drop down.

This is a close-up view of the visualization panels. The top panel shows a plot of Retention time (min) vs m/z. The y-axis ranges from 16.528 to 16.947, and the x-axis ranges from 768 to 772. A red box highlights a region of overlapping peptide ions. The bottom panel shows a mass spectrum plot of Retention time (min) vs m/z. The y-axis ranges from 10.603 to 31.745, and the x-axis ranges from 500 to 2000. A yellow box highlights a region of overlapping peptide ions.

Returning to the highlighted peptide ions for protein with no overlapping peptide ions, now holding down the **Ctrl** key deselect the 2 ions showing 2 and 4 conflicts.

Peptide ions of G318R9

No filter applied

#	Σ	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict:	Peptide Sequence
42732	0	7.49	2	0.994	2443.227	-1.59	16.4	2		2.41E+03	0	KEDVGTVVGID
48016	0	---	---	0.995	1216.612	-9.46	12.3	3		333	0	DAGTIAGLNVN
20984	0	7.38	2	0.995	1973.903	1.22	13	2		5.28E+03	0	IEWLESHQDAD
7586	0	7.53	2	0.998	1045.563	-2.21	2.92	2		5.53E+03	0	VLESDLKK
2658	0	9.37	2	1.000	1658.882	-3.52	13.8	3		4.12E+04	2	IINEPTAAATA
15842	0	7.56	2	0.998	2147.993	1.39	20	2		9.5E+03	0	IEIESFFEGED
29426	0	---	---	0.998	1227.616	-3.64	5.21	3		1.07E+03	4	VEI IANDQGNR
13632	0	8.95	2	0.999	1459.749	-1.82	13.5	3		4.97E+03	0	SDIDEIVLVGG
1692	0	8.79	2	1.000	1216.619	-3.38	12.3	2		5.7E+04	0	DAGTIAGLNVN
4119	1	8.86	2	1.000	1676.804	2.29	9.59	2		3.93E+04	0	NQLTSNPENTV

This leaves 7 remaining peptide ions, with no conflicts and no overlapping peptide ions all showing good correlation of their expression profiles.

Now create a new tag for these peptide ions by right clicking on the highlighted ions, select **New tag** and call it HCP\_library\_All

Peptide ions of G318R9

No filter applied

#	Σ	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict:	Peptide Sequence
42732	0	7.49	2	0.994	2443.227	-1.59	16.4	2		2.41E+03	0	KEDVGTVVGID
48016	0	---	---	0.995	1216.612	-9.46	12.3	3		333	0	DAGTIAGLNVN
20984	0	7.38	2	0.995	1973.903	1.22	13	2		5.28E+03	0	IEWLESHQDAD
7586	0	7.53	2	0.998	1045.563	-2.21	2.92	2		5.53E+03	0	VLESDLKK
2658	0	9.37	2	1.000	1658.882	-3.52	13.8	3		4.12E+04	2	IINEPTAAATA
15842	0	7.56	2	0.998	2147.993	1.39	20	2		9.5E+03	0	IEIESFFEGED
29426	0	---	---	0.998	1227.616	-3.64	5.21	3		1.07E+03	4	VEI IANDQGNR
13632	0	8.95	2	0.999	1459.749	-1.82	13.5	3		4.97E+03	0	SDIDEIVLVGG
1692	0	8.79	2	1.000	1216.619	-3.38	12.3	2		5.7E+04	0	DAGTIAGLNVN
4119	1	8.86	2	1.000	1676.804	2.29	9.59	2		3.93E+04	0	NQLTSNPENTV

Right-click context menu options: No tags to assign, **New tag...**, Quick Tags, Edit tags

Create new tag dialog:  HCP\_library\_All

To keep track of which proteins you have reviewed, create a protein tag (HCP\_library) for the left hand table.

Accession	Peptides	Unique	Conflict:	Score	Tag	Abundance
867100	59	59	28	888		2.38E+07
P02769	38	38	0	503		3.74E+05
G3HL13 (+1)	24	24	0	301		1.03E+05
G3GV00 (+6)	19	8	13	252		3.53E+05
G3BR9	19	17	12	215		7.65E+04
G3H0U6	0	186	0	186		5.52E+04
G3HQM6	0	159	0	159		5.74E+04
G3HQP8 (+5)	13	165	0	165		9.1E+04
G3I06 (+2)	34	148	0	148		5.05E+04
G3IAQ0 (+5)	13	13	0	130		4.22E+04

Right-click context menu options: **HCP\_library**, New tag..., Quick Tags, Edit tags

#	Σ	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict:	Peptide Sequence
42732	0	7.49	2	0.994	2443.227	-1.59	16.4	2		2.41E+03	0	KEDVGTVVGID
48016	0	---	---	0.995	1216.612	-9.46	12.3	3		333	0	DAGTIAGLNVN
20984	0	7.38	2	0.995	1973.903	1.22	13	2		5.28E+03	0	IEWLESHQDAD
7586	0	7.53	2	0.998	1045.563	-2.21	2.92	2		5.53E+03	0	VLESDLKK
2658	0	9.37	2	1.000	1658.882	-3.52	13.8	3		4.12E+04	2	IINEPTAAATA
15842	0	7.56	2	0.998	2147.993	1.39	20	2		9.5E+03	0	IEIESFFEGED
29426	0	---	---	0.998	1227.616	-3.64	5.21	3		1.07E+03	4	VEI IANDQGNR
13632	0	8.95	2	0.999	1459.749	-1.82	13.5	3		4.97E+03	0	SDIDEIVLVGG
1692	0	8.79	2	1.000	1216.619	-3.38	12.3	2		5.7E+04	0	DAGTIAGLNVN
4119	1	8.86	2	1.000	1676.804	2.29	9.59	2		3.93E+04	0	NQLTSNPENTV

Now work down the list of proteins (ordered on Peptides) in the Left hand table, highlight in turn all the 'non overlapping' peptide ions (in the peptide ion table). Then while holding down the **Ctrl** key (to retain the selection) click on the peptide ions with conflicts to un-highlight them. Finally **right** click on the remaining highlighted peptide ions and select the **HCP\_library\_All** tag.

Accession	Peptides	Unique	Conflict:	Score	Tag	Abundance
867100	59	59	28	888		2.38E+07
P02769	38	38	0	503		3.74E+05
G3HL13 (+1)	24	24	0	301		1.03E+05
G3GV00 (+6)	19	8	13	252		3.53E+05
G3BR9	19	17	12	215		7.65E+04
G3H0U6	17	17	0	186		5.52E+04
G3HQM6	16	16	0	159		5.74E+04
G3HQP8 (+5)	14	6	13	165		9.1E+04

Right-click context menu options: **HCP\_library\_All**, New tag..., Quick Tags, Edit tags

#	Σ	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict:	Peptide Sequence
8561	0	8.94	2	0.979	1926.789	-1.29	6.47	2		1.68E+04	0	CCAADREARFF
8900	0	---	---	0.999	1141.702	-4.76	9.46	3		4.28E+03	0	KQTALVELLK
1560	0	8.94	2	1.36	7.63E+04	-1.36	4.84	2		7.63E+04	0	YIYDNDQTISS
13269	0	8.94	2	3.73	3.88E+03	-3.73	4.84	3		3.88E+03	0	YIYDNDQTISS
13677	1	8.69	2	-2.17	1.59E+04	-2.17	10.5	2		1.59E+04	0	RFPSALTPE
1198	0	1.929	2	-0.9	1.04E+05	-0.9	8.16	2		1.04E+05	0	HLVDEPQNLIK
1202	1	8.82	2	0.995	788.482	-2.76	5.77	2		4.7E+04	0	LVDTLTK
1062	1	8.95	2	0.999	1013.615	2.44	13	2		9.23E+04	0	QTALVELLK

Repeat this process for all the HCP proteins with un-conflicted peptide ions.

Now highlight the NIST mAb protein, in this case the one with the greatest number of peptides, right click and tag it (NIST\_Ab\_Product). Then in the Peptide Ions table order on overlapping peptide ions and highlight all those with no overlaps. Un-highlight those with Conflicts. Then right click on the remaining highlighted peptide ions and create the tag **NIST\_product library**.

**Note:** at this stage, in addition to the tagged groups described above you **can** also create a very conservative library by only tagging the 3 most abundant peptide ions, assigned to different sequences that have no overlapping peptides, no conflicts and good correlation of their expression profiles.

To do this first apply a Filter so that only the Peptide ions that are tagged **HCP\_library\_All** are shown in the Peptide Ions table

For example returning to the G318R9 protein, order on Abundance and highlight the 3 most abundant peptide ions with different sequences

Now, before actually creating the Spectral Libraries from the tagged groups of peptide ions you can review the Quality of fragmentation for each 'candidate' spectra being entered into the library and add additional tags to indicate quality of fragmentation. This then allows to easily control the creation and addition of spectra to a library through application of tag filters.

## Stage 8 Review fragmentation quality of candidate spectra

Once you have tagged your candidate spectra at the Resolve Conflicts stage you can now return to the Identify Peptides stage and review the quality of fragmentation for each peptide ion, before adding it to the spectral library.

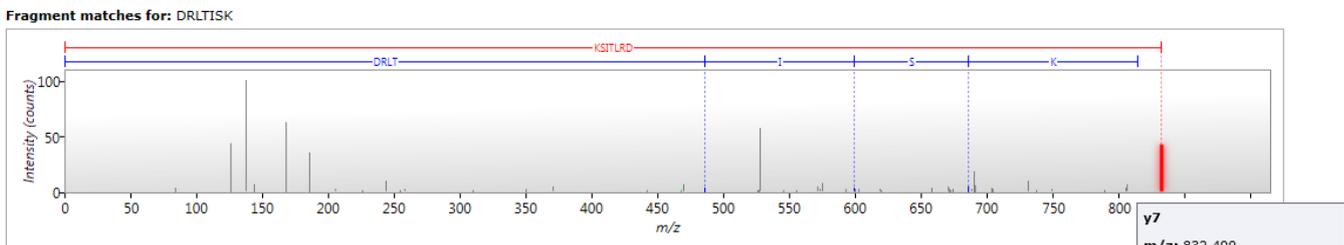
**Note:** the following section describes one ‘possible strategy’ on how to decide which spectra to add to the spectral library on the basis of the number and quality of fragment matches. It outlines the processes involved **but** it should be clear that it is by no means definitive as there will always be an element of subjectivity, quality is not an ‘exact metric’.

First select the group of candidate spectra to review i.e. start with the NIST\_product library

The screenshot shows the 'Identify Peptides' stage in Progenesis QI. A 'Create a filter' dialog box is open, allowing the user to filter peptide ions based on tags. The 'Available tags' list includes 'HCP\_HI\_3\_Library (119 peptide ions)', 'NIST\_product library (55 peptide ions)', and 'HCP\_library\_All (253 peptide ions)'. The 'NIST\_product library' tag is selected. Below the dialog, a fragmentation plot for the peptide WQQGNVFSQVMHEALHNHYTQK is displayed, showing intensity (counts) versus m/z. The plot shows several peaks corresponding to the peptide sequence, with a prominent peak at m/z 867.100.

Before going through the filtered list of tagged peptide ions for the NIST Ab product you should first create a new tag to attach to all the spectra that are judged poor in terms of fragmentation or no matches.

For example: for the fragmentation of DRLTISK, shown below, there is only a single match for the y (in this case y 7 representing the un-fragmented precursor) and only 4 b ions matched.



The screenshot shows the 'Peptide ions (825 identified)' table in Progenesis QI. A 'Create new tag' dialog box is open, showing a new tag 'NIST\_product library\_poor frag' being created. The table lists peptide ions with columns for #, Identifications, m/z, Charge, Retention time, and Tag. The ion at m/z 416.7468 is highlighted.

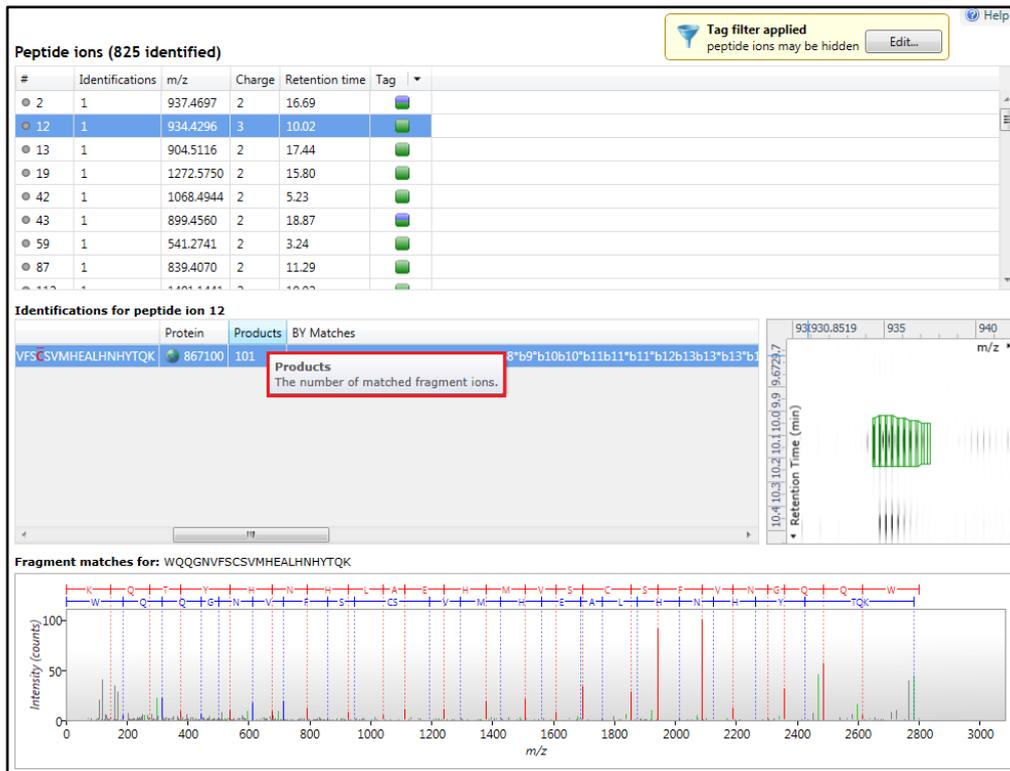
#	Identifications	m/z	Charge	Retention time	Tag
618	1	639.9958	3	4.66	
959	1	416.7468	2	2.69	
1007	1			2.17	
1021	1			2.83	
1234	1			2.95	
1267	1			2.51	
1285	1			3.00	
1352	1			3.91	

To tag this ‘poorly fragmented peptide ion’ right click on the ion in the table and create a new tag

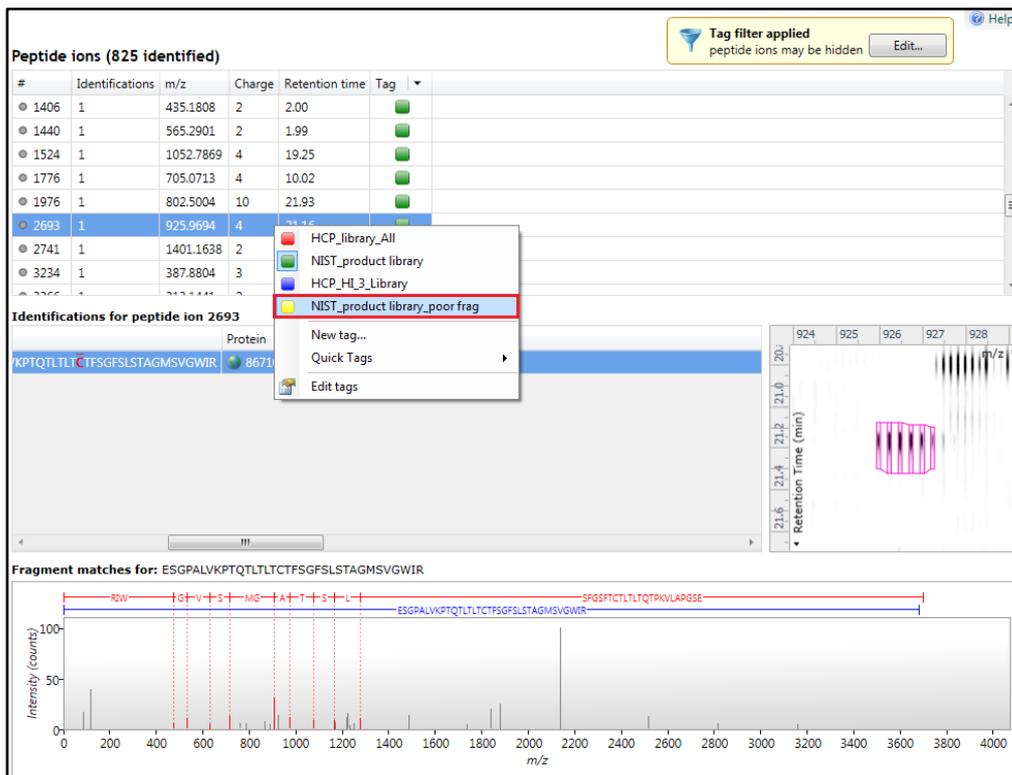
NIST\_product library\_poor frag

Return to the top of the filtered list of Peptide ions, then step through the list tagging the 'poorly fragmented ions where there is a low number of Fragment matches, poor coverage of the peptide ion sequence or no matched fragments (due to inherited identification from another charge state).

Example of good fragmentation matches, leave untagged.

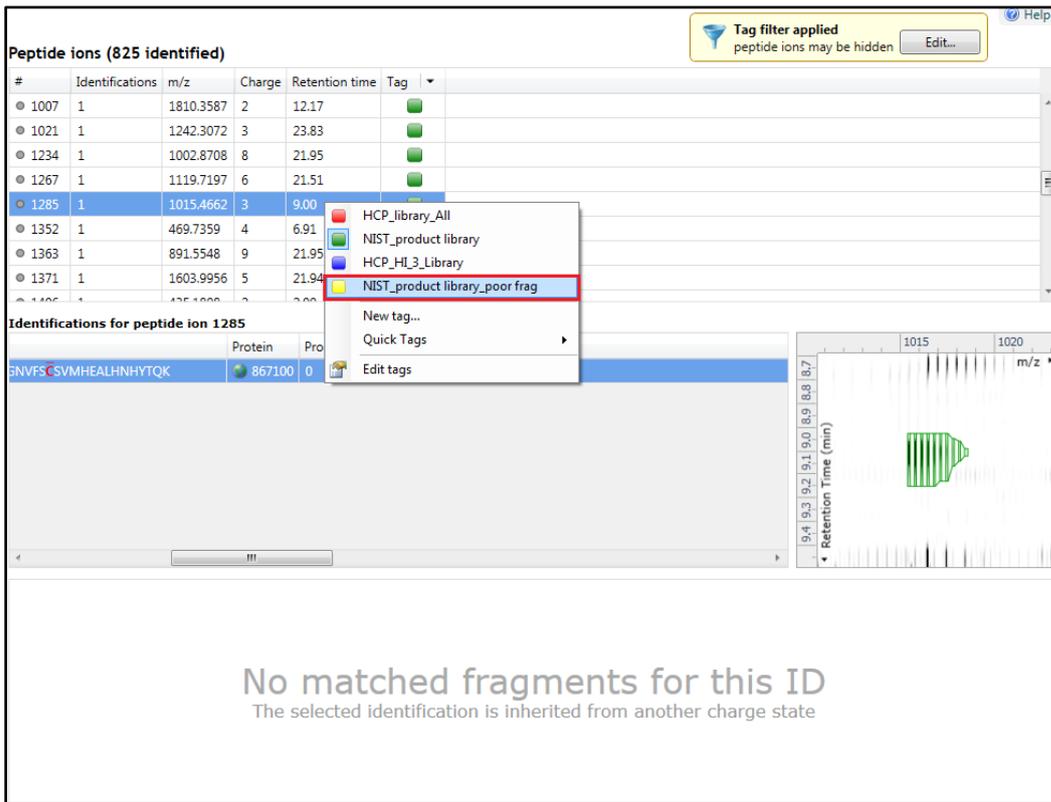


Example of poor fragmentation matches, right click on the peptide ion in the table and select the NIST\_product library\_poor frag



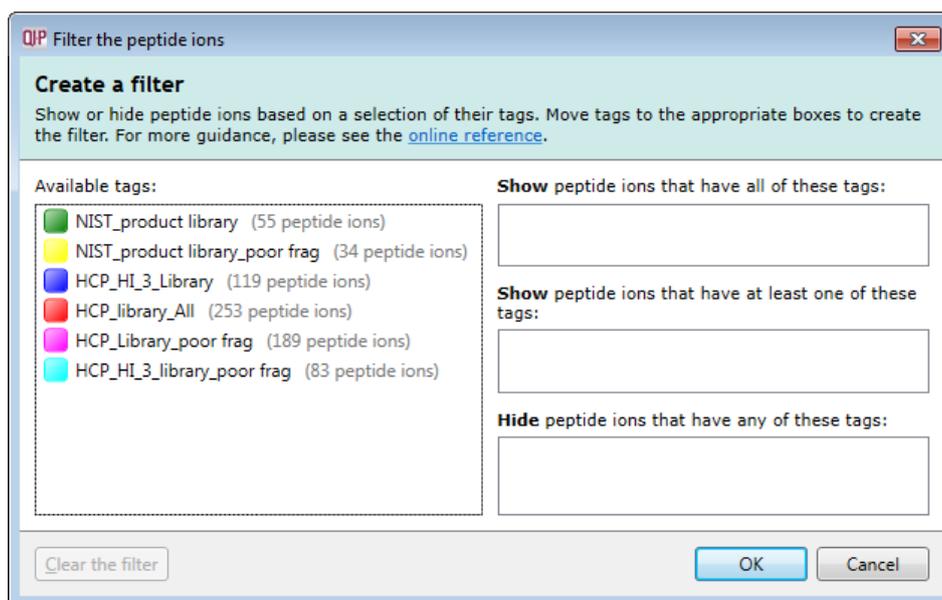
Here there is not only limited Fragment matches for y ions but also only the matching of the precursor ion.

Finally where you have initially tagged peptide ions, where the identity has been inherited from another charge state, this will show no Fragment matches. At this point, the option would be not to have tagged it in the first place, when you were creating the original tagged list as described in the previous section.



Once you have completed the review process of the NIST\_product library then repeat the process for the remaining 2 candidate libraries (HCP\_library\_All and as required HCP\_HI\_3\_Library) creating additional tags for the corresponding poor fragmentation (i.e. HCP\_library\_poor frag).

Having done this you will have 6 tagged groups of peptide ions as shown below.



**Note:** the actual numbers will vary depending on how the spectra were reviewed.

Now move forward to the next section to see how these tags can be used to create Spectral Libraries.

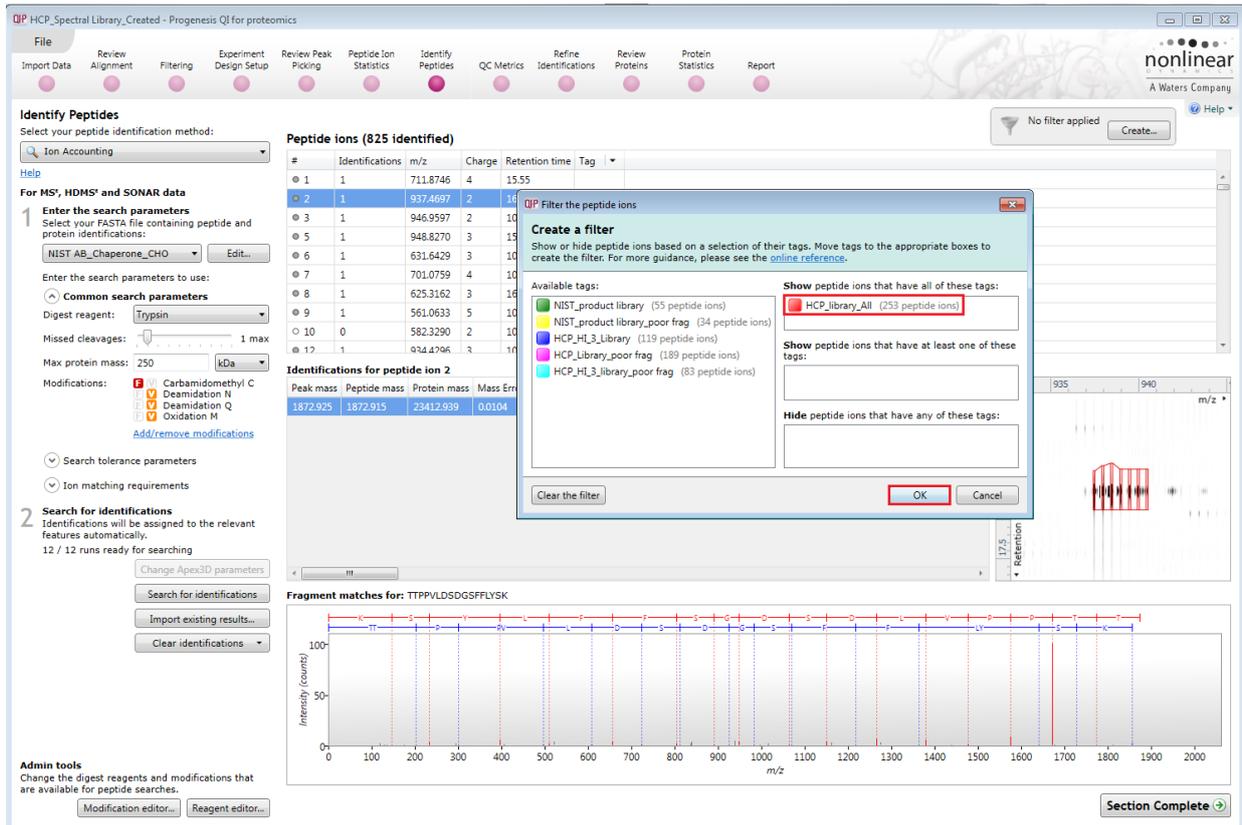
## Stage 9: Creating a Spectral Library

Having tagged and reviewed the ‘candidate’ lists of peptide ions in the previous section you can now create the actual libraries by first filtering to the required list of peptide ions and then exporting them to a Spectral library.

**Note:** if you have come straight to this point in the workflow and you have already unzipped the **Library\_search\_data\_set.zip** into same folder then examples of the spectral libraries will already exist as described below

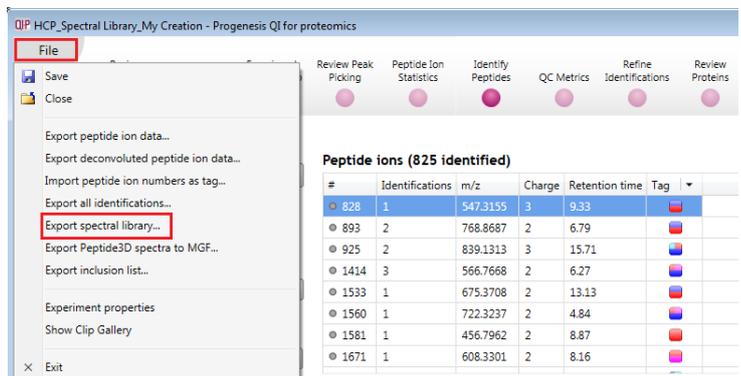
As an example, create two Spectral Libraries firstly a Spectral Library called **HCP\_library\_All\_raw** which is based on all peptide ions tagged **HCP\_library\_All** and a second library **HCP\_library\_All\_reviewed** is the same as the first library minus the spectra judged as having poor fragmentation.

Click on **Create** and drag the **HCP-library\_All** tag on to the Show panel and click OK. This will reduce the table to showing only those peptide ions which have this tag.



To create the library from the tagged filter group, select **Export spectral library...** from the file menu

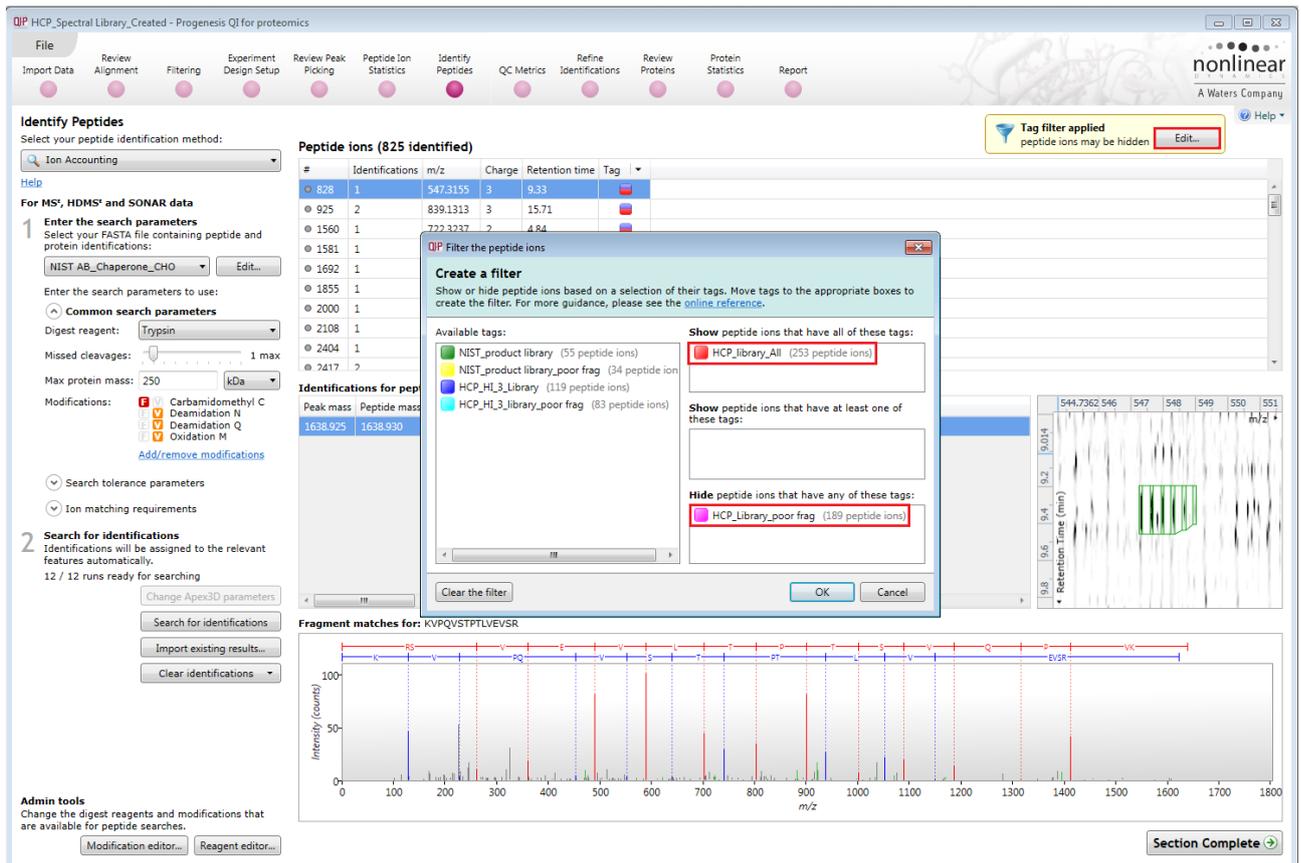
Then save the new spectral library as **HCP\_library\_All\_raw.msp**



Now create the **HCP\_library\_All\_raw\_reviewed** spectral library.

First modify the applied tag filter to hide the spectra that have been judged as poorly fragmented.

Click on **Edit** to open the existing Tag filter. Then drag the tag HCP\_Library\_poor frag on to the **Hide** panel while leaving the existing tag in the **Show** panel

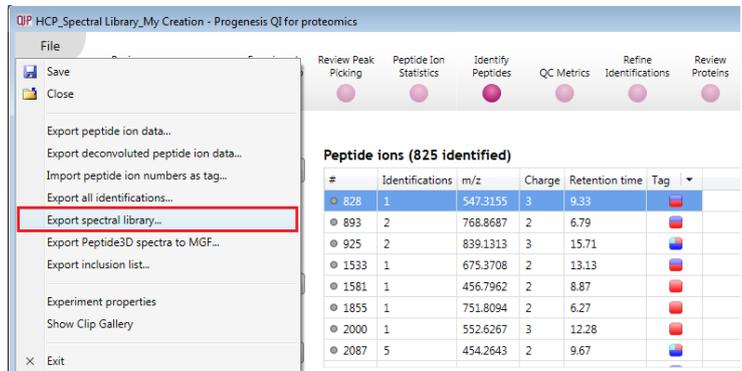


This will further reduce the number of tagged peptide ions in the table.

As before to create the library from the tagged filter group, select **Export spectral library...** from the file menu

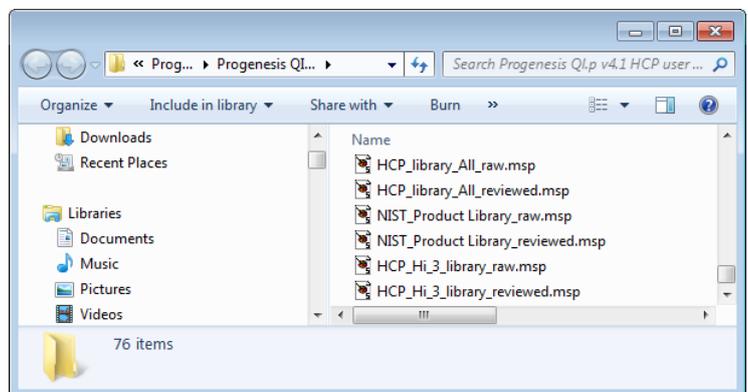
Then name and save the library, **HCP\_library\_All\_reviewed.msp**

**Tip:** if saving the library to the same folder as the example library files, use a different name if you wish to retain both libraries.



Now clear the filter by clicking on **Edit** followed by **Clear the filter**, then repeat the process for the **NIST\_product library** and the **HCP Hi\_3\_Library**.

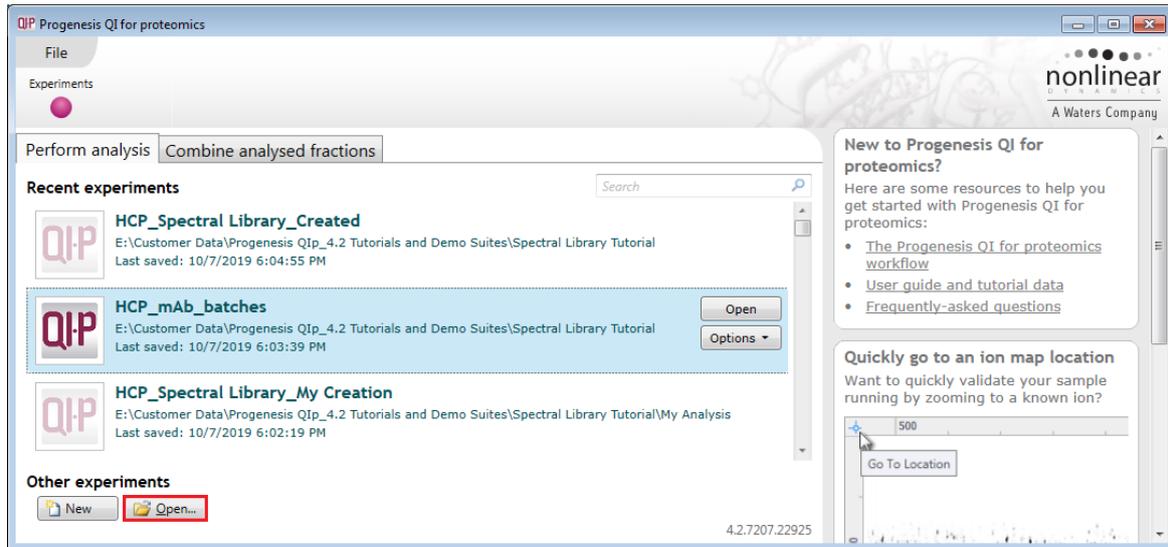
Once you have done this you will have 6 .msp files that you can use for searching the Batch samples described in the next section.



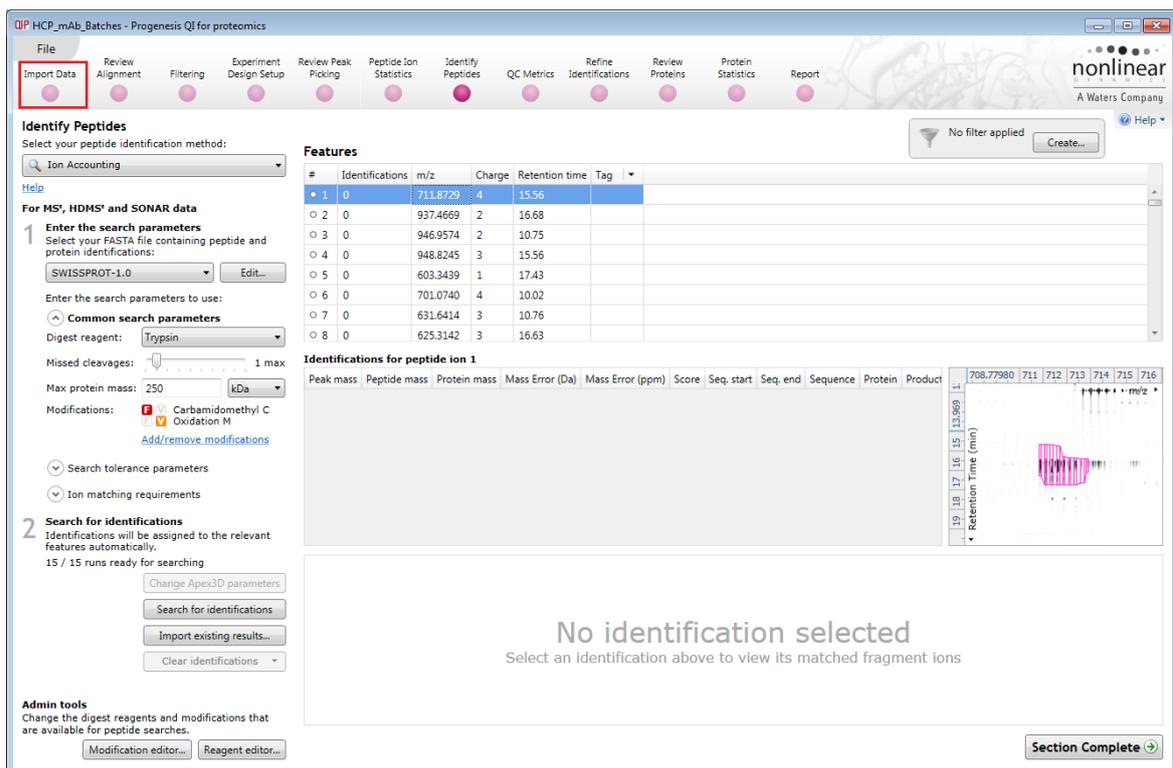
**Note:** Examples of these Spectral Libraries are provided with the Tutorial data that you downloaded to perform this tutorial.

## Stage 10: Autoprocessing of the Batch Samples

As an example of using the spectral library you have just created use it to identify and measure the Host Cell Proteins (HCP) present in the **HCP\_mAb\_Batches** of NIST Ab product. To do this use the second tutorial archive that you restored at the beginning of this tutorial.



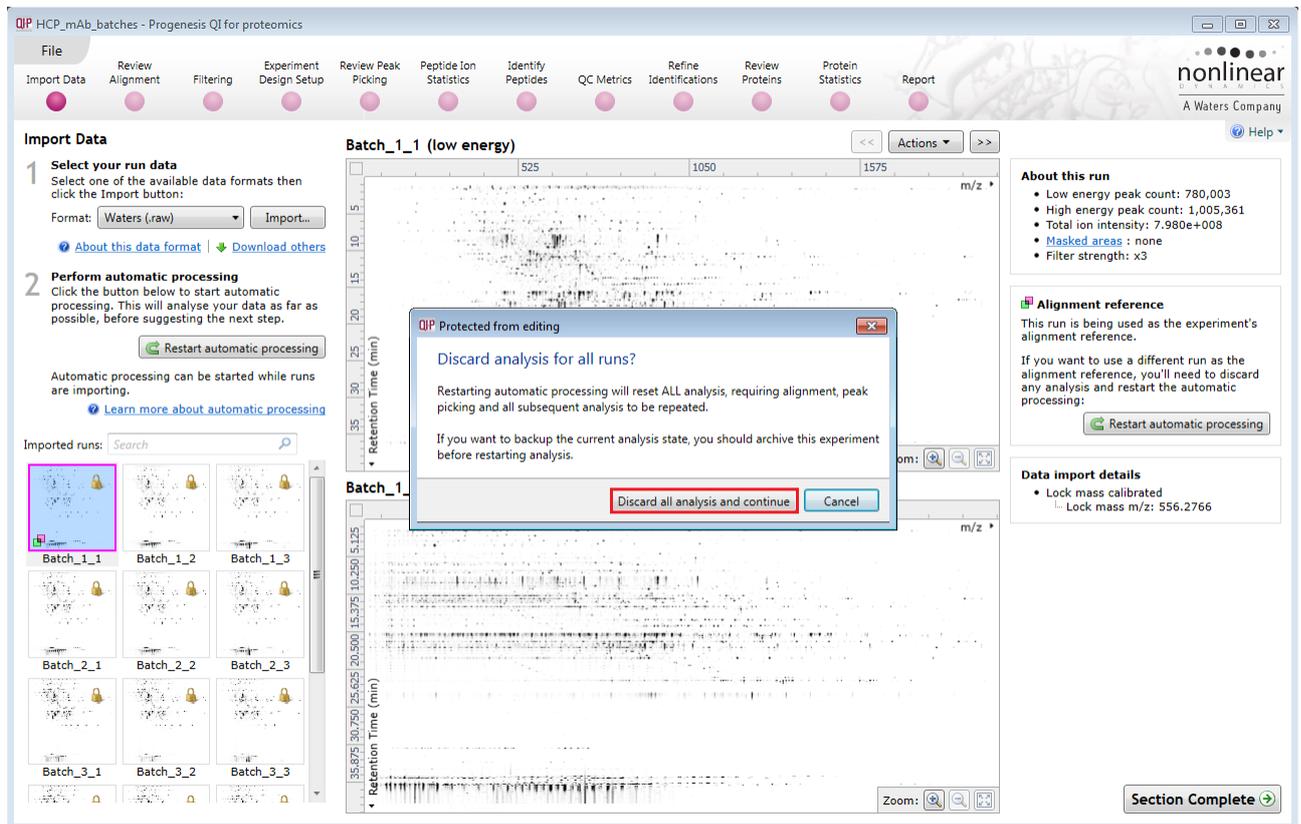
When you open the restored **HCP\_mAb\_Batches** experiment it will be already at the Identify Peptides stage but no identification has been performed.



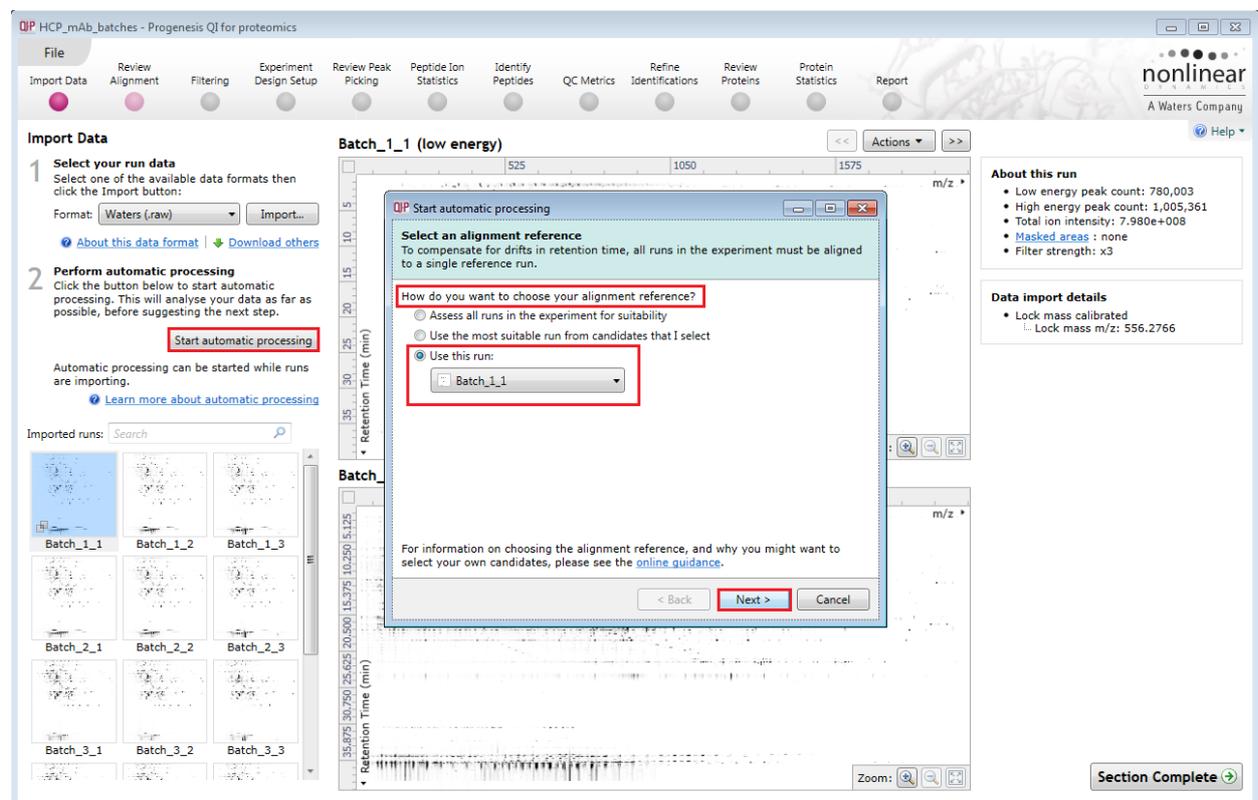
**Note:** the data has been autoanalysed and an Experiment Design applied as part of this automatic analysis.

To perform Identification using the Spectral Libraries that you made in the previous section go to Stage 11 (page 32)

The following pages describe the automatic analysis that has been performed on this data which you can repeat by first discarding the automatic analysis by clicking **Restart automatic processing**.



On clicking **Discard all analysis and continue** the first Automatic processing dialog will open asking how you want to choose the alignment reference to be determined

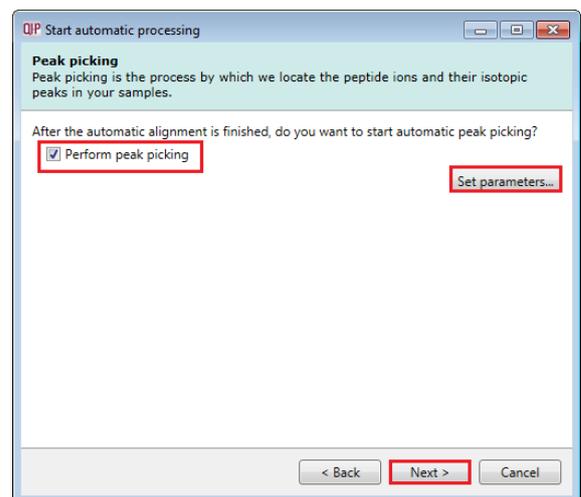
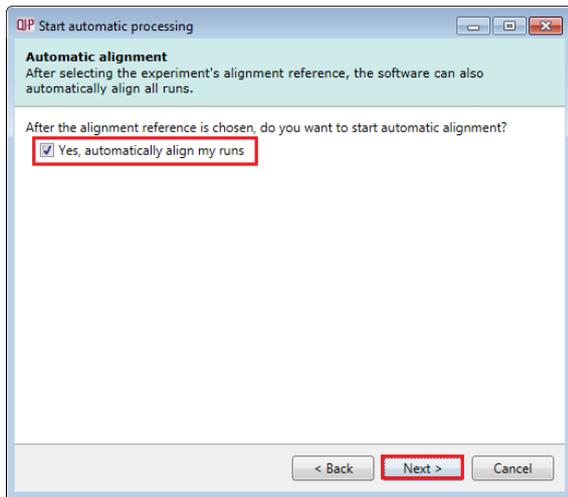


The list below describes what options were applied at each stage to automatically analyse the Batches data set.

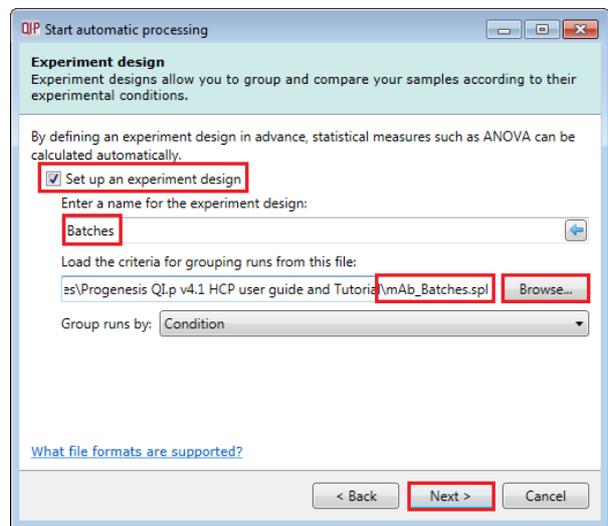
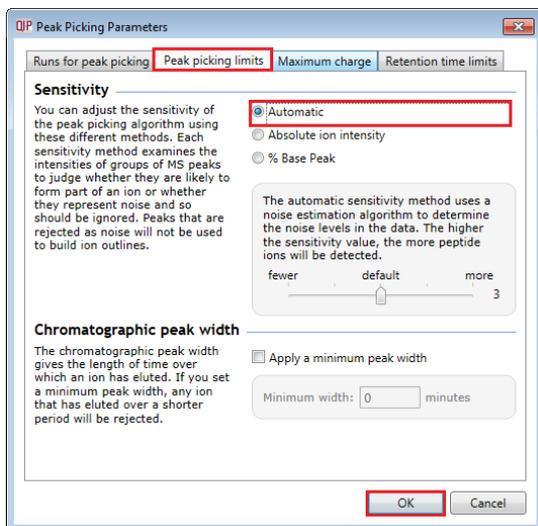
Setting up the steps for the auto processing:

The Alignment Reference was set as 'Batch 1\_1' (as shown above).

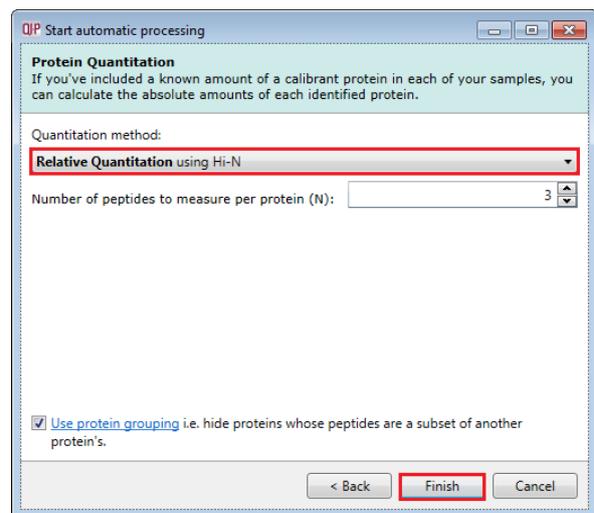
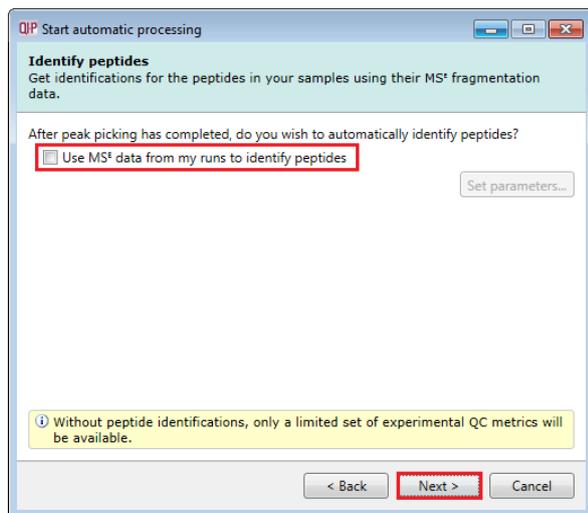
Automatic alignment was performed and the default settings for peak picking.



The default Peak picking limits are set as Automatic (as shown below)



For Experiment design setup: use file mAb\_Batches.spl (available in same folder that the Library\_search\_data\_set.zip was extracted).

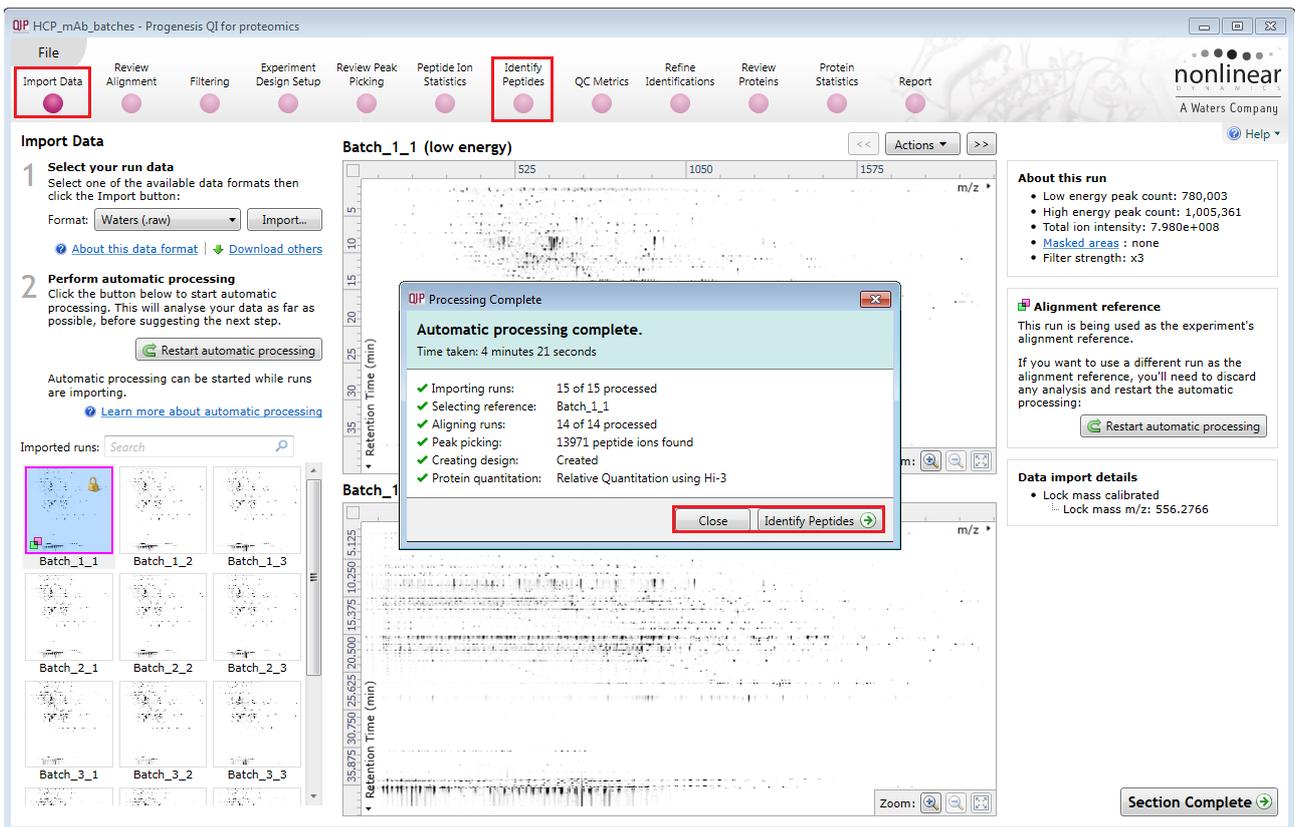


**Do not** perform identification by deselecting **Use MS<sup>E</sup> data.....** untick this option

Use default settings for Quantitation (Relative Quantitation using Hi-N) and Use protein grouping

Click **Finish** to start the analysis.

The automatic processing completes in approx. 5min (depending on your machine spec).



Having repeated the analysis you can either **Close** it, where it will remain at the Import Data stage and you can step through the Alignment and other stages or you can open it directly at the **Identify Peptides** stage. Here we will close and then click on **Review Alignment** on the workflow at the top of the screen.



**Note:** the high quality of retention time alignment of the runs as indicated by the high % Score.

Now click on Identify Peptides and go to the next section.

## Stage 11: Identifying and measuring HCP proteins using a spectral library

As an example of using the spectral libraries you have just created use it to identify and measure the Host Cell Proteins (HCP) present in the **HCP\_mAb\_Batches** of NIST Ab product. To do this use the second tutorial archive that you restored at the beginning of this tutorial.

**If you have just reanalysed this you will be currently at the Identify Peptide stage in the workflow.**

Select Spectral Library Search from the peptide identification methods

**Identify Peptides**  
Select your peptide identification method:

- Ion Accounting
- Mascot
- PLGS (\*.xml)
- SEQUEST (dta & out files)
- SEQUEST (dta & pepXml files)
- Phenyx
- Spectral Library Search
- Ion Accounting

**Features**

#	Identifications	m/z	Charge	Retention time	Tag
1	0	711.8746	4	15.56	
2	0	937.4693	2	16.68	
3	0	946.9596	2	10.75	
4	0	948.8268	3	15.57	
5	0	701.0757	4	10.02	
6	0	631.6430	3	10.75	
7	0	603.3453	1	17.43	
8	0	625.3157	3	16.63	

**Identifications for peptide ion 1**

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

709 710 711 712 713 714 715 716

Retention Time (min)

**No identification selected**  
Select an identification above to view its matched fragment ions

Section Complete

As no searches have been performed there will be no libraries to select. Click **Browse** and select the **NIST\_Product Library\_raw.msp** from the folder you restored the Tutorial data.

**Identify Peptides**  
Select your peptide identification method:

- Spectral Library Search

**Enter search parameters**  
Select the spectral library to search:  
NIST\_Product Library\_raw

Precursor tolerance: 10 ppm  
Fragment tolerance: 10 ppm  
Retention time within: 0.5 mins  
Fragments per peptide: 3 or more  
Share hits across charge states

**Identifications:**

#	Protein	Sequence	Score	Fragment matches	Peak ma
345					
346					
347					
348					
349					

709 710 711 712 713 714 715 716

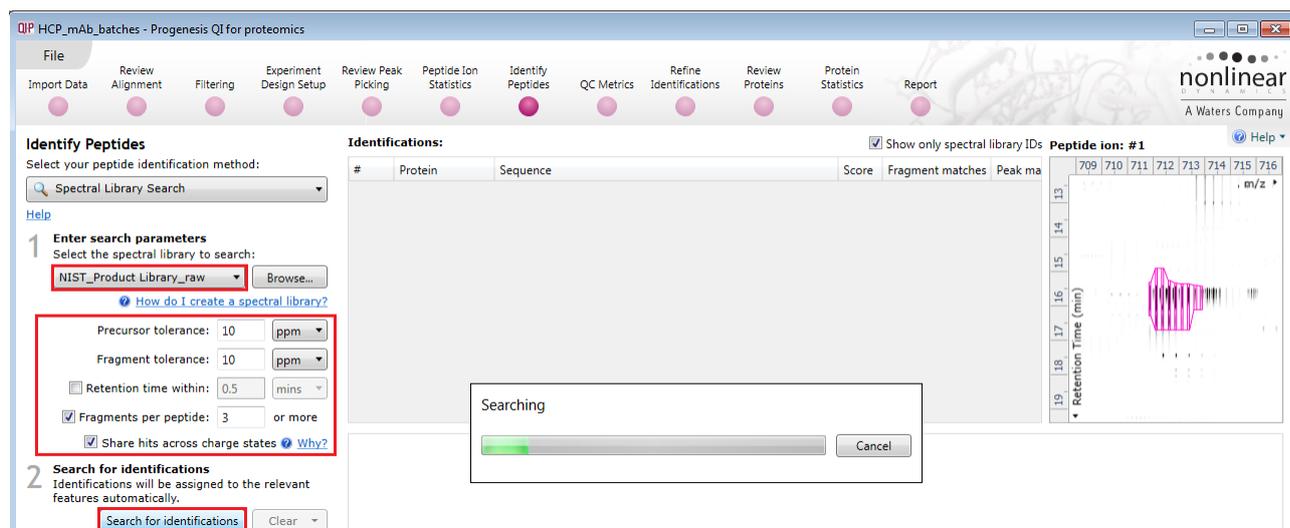
Retention Time (min)

Identified

Search for identifications using the panel at the left

Now check the tolerances as shown below

The **NIST\_Product Library\_raw** will appear as the current spectral library



The Spectral library search parameters are described below

#### Precursor tolerance

Specifies the mass error below which a precursor is considered a match to a spectral library entry.

#### Fragment tolerance

Specifies the  $m/z$  error below which a fragment is considered a match to a spectral library entry's fragments.

#### Retention time within

Specifies the retention time error, below which a precursor is considered a match to a spectral library entry. If this is not selected, or a spectral library entry has no retention time information, it has no effect.

#### Fragments per peptide

The minimum number of fragment matches required for a match to be accepted. If not selected, all matches are accepted, even those with no fragment matches. If a spectral library entry has fewer fragments than the given value, all of its fragments must match. If a spectral library entry has no fragments, matches to that entry will be accepted regardless of this value.

#### Share hits across charge states

If selected, charge state deconvolution of identifications will be performed. This means that if only some charge states of a given peptide were identified by the spectral library search, the other charge states will inherit the same identifications.

Tick the tolerances as shown above

Start the search by clicking **Search for identifications**

When the search completes order the table on Fragment matches, and select an 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.

**Identify Peptides**

Select your peptide identification method:  
Spectral Library Search

**1 Enter search parameters**  
Select the spectral library to search:  
NIST\_Product Library\_raw

Precursor tolerance: 10 ppm  
Fragment tolerance: 10 ppm  
Retention time within: 0.5 mins  
Fragments per peptide: 3 or more

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

#	m/z	Charge	Retention time	Tag
121	1422.7396	2	15.56	
122	955.6249	1	35.98	
123	488.2743	1	8.62	
124	419.7610	2	6.09	
125	563.2449	2	11.54	
126	811.4025	2	5.86	
127	448.2815	1	1.50	
128	977.6061	1	35.98	
129	613.2859	1	8.66	
130	580.2789	5	15.55	

**Identifications:**

#	Protein	Sequence	Score	Fragment matches	Peak
184	867100	STSGGTAALGCLVKDYFPEPVTVSWNSGALTSVHTFPAVLQ...	78.914	49	8014
786	867100	ESGPALVKPTQLTLTLCTFSGFSLSTAGMSVGVIRQPPGK	98.464	46	4207
121	867100	THTCPPCPAPELLGGPSVFLFPPKPK	96.981	46	2843
875	867100	DYFPEPVTVSWNSGALTSVHTFPAVLQSSGLYSLVTVPS...	90.978	46	6712
12	867100	WQQGNVFCVSMHEALHNHYTQK	92.592	44	2800
21	867100	GFYPSDIAVEWESNGQPENNYK	96.114	43	2543
42	867100	SGTASVVCLLINNFYR	97.951	41	1796
1296	867100	VYACEVTHQGLSSPVTK	89.115	40	1874
52	867100	VDNALQSGNSQESVTEQDSK	58.512	38	2134
1762	867100	WQQGNVFCVSMHEALHNHYTQK	63.554	36	2816

**Fragment matches for: THTCPPCPAPELLGGPSVFLFPPKPK**

Measured vs Library b and y ions plot.

Now click on **Review proteins** on the workflow to examine the quantified proteins. As the Product library only contained the spectra for one protein (867100) then only one protein is displayed with 31 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.

Accession	Peptides	Unique peptides	Confidence score	Anova (p)	q Value	Tag	Max fold change	Highest Mean	Lowest Mean	Description
867100	31	31	3.03E+03	0.313	0.313		1.22	Batch 5	Batch 2	Light Chain

**Selected protein: Light Chain**

View peptide measurements

ArcSinh Normalised Abundance vs Batch (Batch 1 to Batch 5)

Quantifiable proteins displayed: 1

The protein quantitation (options) are currently set to Relative Quantitation based on the 3 most abundant peptides (Hi-N).

**Note:** having searched using the NIST\_ product Library\_raw you can set the normalisation to use only the peptide ions for this protein, as the same amount of this protein is present in all the batches, refer to page 15 for details on how to change the normalisation).

Now use the HCP\_library\_All\_raw.msp to search the batches by first browsing for the new library then Click **Search for identifications**

**Identifications:**

#	Protein	Sequence	Score	Fragment matches	Peak mass	Peptide mass	Protein mass	Mass error (Da)	Mass error (ppm)
1139	P02769	KVPQVSTPTLVEVSR	91.694	20	1638.924	1638.930	71289.430	-0.0067	-4.0761
9192	G3GVDO	LCYVALDFEQEMATAASS...	48.851	18	2549.169	2549.167	42079.017	0.0028	1.1044
9192	G3I6E6	LCYVALDFEQEMATAASS...	48.851	18	2549.169	2549.167	32332.538	0.0028	1.1044
9192	G3GX66	LCYVALDFEQEMATAASS...	48.851	18	2549.169	2549.167	27457.403	0.0028	1.1044
2850	P02769	SHCAEVEKDAIPENLPLT...	95.337	14	3510.650	3510.665	71289.430	-0.0150	-4.2818
1865	P63284	VISQNEAVDAVSNAIR	73.632	12	1654.857	1654.864	95756.274	-0.0069	-4.1769
4859	G3GX66	DLYANTVLSGGTTMYPGIA...	55.125	12	2214.059	2214.063	27457.403	-0.0036	-1.6181
4859	G3GVDO	DLYANTVLSGGTTMYPGIA...	55.125	12	2214.059	2214.063	42079.017	-0.0036	-1.6181
4869	G3H804	LAADVATVQVLNLR	78.883	11	1553.852	1553.852	28629.400	-0.0003	-0.2072
6216	P63284	VYDAEAEVLAR	97.526	11	1285.681	1285.688	95756.274	-0.0070	-5.4607
9353	P02769	LGEVGFQNALIVR	68.286	11	1478.785	1478.788	71289.430	-0.0036	-2.4052

Change the method for Protein Quantitation to **Absolute Quantitation for HCP** using Hi-N, by clicking on **Protein options** and use the peptides identified as belonging to the Chaperone Protein ClpB from E.coli, that was spiked into each sample at 200fmol, as the Calibrant (Accession P63284).

**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 shortcut to review**  
In the table, sort and filter the proteins based on their measurements, to generate a shortcut 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.

**Protein quantitation options**

Quantitation method: **Absolute Quantitation for HCP using Hi-N**

Number of peptides to measure per protein (N): 3

Calibrant accession: P63284

Amount (fmol): 200

**Selected protein: Chaperone protein ClpB**

es	Confidence score	Anova (p)	q Value	Tag	Max fold change	Highest Mean	Lowest Mean	Description
90.4	< 1.1E-16	< 1.1E-16	< 1.1E-16	Infinity	Batch 1	Batch 3		ATP synthase subunit beta OS=Cricetulus griseus GN=I79_013474 PE=3 SV=1
66.1	1.53E-07	3.05E-07		8.22E+06	Batch 1	Batch 4		Calreticulin OS=Cricetulus griseus GN=H671_3g9525 PE=3 SV=1
171	0.00106	0.00128		1.74	Batch 5	Batch 3		Chaperone protein ClpB OS=Escherichia coli (strain K12) DX=83333 GN=clpB PE=1 SV=1
301	1.09E-05	1.64E-05		2.08	Batch 1	Batch 4		Elongation factor 1-alpha 1 OS=Cricetulus griseus GN=I79_009935 PE=4 SV=1
119	2.22E-16	1.07E-15		Infinity	Batch 1	Batch 4		Heat shock cognate 71 kDa protein OS=Cricetulus griseus GN=I79_021800 PE=3 SV=1
68.2	9.66E-07	1.66E-06						Cognate 71 kDa protein OS=Cricetulus griseus GN=I79_022974 PE=3 SV=1
95.5	< 1.1E-16	< 1.1E-16						Hydrogenase A chain OS=Cricetulus griseus GN=I79_009741 PE=4 SV=1
119	< 1.1E-16	< 1.1E-16						Cricetulus griseus GN=I79_022400 PE=4 SV=1
94.3	9.11E-05	0.000						yl cis-trans isomerase OS=Cricetulus griseus GN=I79_010223 PE=3 SV=1
14.6	1.82E-13	4.85E-13						m-1 OS=Cricetulus griseus GN=I79_002954 PE=4 SV=1
16.5	3E-07	5.54E-07						Hide-isomerase A3 OS=Cricetulus griseus GN=I79_003765 PE=4 SV=1
136	2.93E-05	4.14E-05						Hide-isomerase B3 OS=Cricetulus griseus GN=I79_007616 PE=4 SV=1
81.8	1.06E-06	1.69E-06						Chaperone protein ClpB OS=Escherichia coli (strain K12) OX=80 PE=3 SV=1
834	1.45E-08	3.49E-08						G3H-P6
99.2	0.00527	0.005						Elongation factor 1-alpha 1 OS=Cricetulus griseus GN=I79_009935 PE=3 SV=1
55	< 1.1E-16	< 1.1E-16						chain OS=Cricetulus griseus GN=I79_008223 PE=3 SV=1

There are 24 quantifiable proteins identified, across the batches that you are testing

**Review Proteins**

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

**1 Set the quantification 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 this 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.

Lowest Mean	Description	Amount (fmol) - Batch 1	Amount (fmol) - Batch 2	Amount (fmol) - Batch 3	Amount (fmol) - Batch 4	Amount (fmol) - Batch 5	Amount (ng) - Batch 1	Amount (ng) - Batch 2	Amount (ng) - Batch 3	Amount (ng) - Batch 4	Amount (ng) - Batch 5
Batch 4	Protein disulfide-isomerase A3 OS=Cricetulus griseus GN=I79_003765 PE=4 SV=1	53.4	24.9	0.048	0	0	2.69	1.26	0.00242	0	0
Batch 5	Trypsin OS=Sus scrofa PE=1 SV=1	55.5	53.5	47.6	44	34.3	1.39	1.34	1.19	1.1	0.857
Batch 5	Protein disulfide-isomerase A6 OS=Cricetulus griseus GN=I79_007616 PE=4 SV=1	63.8	29.4	0.495	0.314	0.161	1.83	0.841	0.0142	0.009	0.00461
Batch 4	Heat shock cognate 71 kDa protein OS=Cricetulus griseus GN=I79_021800 PE=3 SV=1	65	29.4	0.101	0	0	1.48	0.668	0.00231	0	0
Batch 5	78 kDa glucose-regulated protein OS=Cricetulus griseus GN=I79_019946 PE=1 SV=1	92.6	43.5	2.66	0.584	0.0863	6.72	3.16	0.193	0.0424	0.00626
Batch 5	Pyruvate kinase OS=Cricetulus griseus GN=I79_004880 PE=3 SV=1	109	47.8	0.214	0.0003	0	5.67	2.49	0.0111	1.56E-05	0
Batch 5	Tubulin beta-3 chain OS=Cricetulus griseus GN=I79_000194 PE=3 SV=1	140	67.8	2.56	0.511	0.186	12.2	5.88	0.222	0.0443	0.0144
Batch 4	Perolein-1 OS=Cricetulus griseus GN=I79_002954 PE=4 SV=1	148	71.9	1.57	0	0	3.33	1.62	0.0353	0	0
Batch 5	60 kDa heat shock protein, mitochondrial OS=Cricetulus griseus GN=I79_011398 PE=3 SV=1	148	66.8	1.15	0.0187	0	8.41	3.79	0.0652	0.00106	0
Batch 5	Peptidyl-prolyl cis-trans isomerase OS=Cricetulus griseus GN=I79_0001023 PE=3 SV=1	153	82.3	12.1	6.16	4.03	2.58	1.39	0.204	0.104	0.0682
Batch 4	Calreticulin OS=Cricetulus griseus GN=H71_3g929 PE=3 SV=1	155	68.3	0.0709	1.98E-05	0.000105	7.52	3.31	0.00246	9.44E-07	5.06E-06
Batch 5	Chaperone protein Cbp OS=Escherichia coli strain K12 GN=K1333 GN=nc8 PE=1 SV=1	200	200	200	200	200	19.2	19.2	19.2	19.2	19.2
Batch 5	Heat shock cognate 71 kDa protein OS=Cricetulus griseus GN=I79_022974 PE=3 SV=1	233	111	3.82	0.0946	0	5.38	2.57	0.0908	0.0126	0
Batch 4	Actin, cytoplasmic 1 OS=Cricetulus griseus GN=I79_001666 PE=3 SV=1	279	132	4.77	0	0	11.8	5.57	0.301	0	0
Batch 5	Serum albumin BSA	473	232	14.8	1	0.178	33.7	16.5	1.05	0.0713	0.0127
Batch 5	Elongation factor 1-alpha 1 OS=Cricetulus griseus GN=I79_009935 PE=4 SV=1	706	520	496	354	298	39.2	28.9	27.6	19.7	16.5

**Selected protein: Pyruvate kinase OS=Cricetulus griseus GN=I79\_004880 PE=3 SV=1**

Quantifiable proteins displayed: 24

**Note:** having used Absolute Quantification for HCP using Hi-N, the spiked protein (Calibrant (Accession P63284) will show the same Amount (either fmol or ng) across all Batches under test.

Now order the data at the **Review Proteins** stage, starting with Batch 1. The first protein in the list shows decreasing presence as you go from Batch 1 to 2 to essentially undetected in Batch 3.

**Review Proteins**

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

**1 Set the quantification 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 this 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.

Description	Amount (fmol) - Batch 1	Amount (fmol) - Batch 2	Amount (fmol) - Batch 3	Amount (fmol) - Batch 4	Amount (fmol) - Batch 5	Amount (ng) - Batch 1	Amount (ng) - Batch 2	Amount (ng) - Batch 3	Amount (ng) - Batch 4	Amount (ng) - Batch 5
L-lactate dehydrogenase A chain OS=Cricetulus griseus GN=I79_009741 PE=4 SV=1	19.4	9.37	0	0	0	0.34	0.113	0	0	0
Annexin OS=Cricetulus griseus GN=I79_022882 PE=3 SV=1	25.6	10.5	0	0	0	0.699	0.287	0	0	0
Tubulin beta chain OS=Cricetulus griseus GN=I79_008223 PE=3 SV=1	30.7	13.6	0	0	0	1.54	0.68	0	0.712	0
Alpha-enolase OS=Cricetulus griseus GN=I79_019072 PE=4 SV=1	35	37.8	40.3	40.2	33.7	0.618	0.668	0.713	0.712	0.597
ATP synthase subunit beta OS=Cricetulus griseus GN=I79_013474 PE=3 SV=1	35.6	16.8	0	0	0	2.05	0.951	0	0	0
ATP synthase subunit alpha, mitochondrial OS=Cricetulus griseus GN=I79_007391 PE=4 SV=1	37.6	16.9	0.584	0.453	0.0795	0.264	0.119	0.0041	0.00218	0.000557
Nucleolin OS=Cricetulus griseus GN=I79_022400 PE=4 SV=1	38.5	17.7	0	0	0	2.07	0.93	0	0	0
14-3-3 protein eta OS=Cricetulus griseus GN=I79_011108 PE=3 SV=1	45.7	22.4	0.394	0	0	1.25	0.613	0.0108	0	0
Protein disulfide-isomerase A3 OS=Cricetulus griseus GN=I79_003765 PE=4 SV=1	53.4	24.9	0.048	0	0	2.69	1.26	0.00242	0	0
Trypsin OS=Sus scrofa PE=1 SV=1	55.5	53.5	47.6	44	34.3	1.39	1.34	1.19	1.1	0.857
Protein disulfide-isomerase A6 OS=Cricetulus griseus GN=I79_007616 PE=4 SV=1	63.8	29.4	0.495	0.314	0.161	1.83	0.841	0.0142	0.009	0.00461
Heat shock cognate 71 kDa protein OS=Cricetulus griseus GN=I79_021800 PE=3 SV=1	65	29.4	0.101	0	0	1.48	0.668	0.00231	0	0
78 kDa glucose-regulated protein OS=Cricetulus griseus GN=I79_019946 PE=1 SV=1	92.6	43.5	2.66	0.584	0.0863	6.72	3.16	0.193	0.0424	0.00626
Pyruvate kinase OS=Cricetulus griseus GN=I79_004880 PE=3 SV=1	109	47.8	0.214	0.0003	0	5.67	2.49	0.0111	1.56E-05	0
Tubulin beta-3 chain OS=Cricetulus griseus GN=I79_000194 PE=3 SV=1	140	67.8	2.56	0.511	0.186	12.2	5.88	0.222	0.0443	0.0144
Perolein-1 OS=Cricetulus griseus GN=I79_002954 PE=4 SV=1	148	71.9	1.57	0	0	3.33	1.62	0.0353	0	0

**Selected protein: L-lactate dehydrogenase A chain OS=Cricetulus griseus GN=I79\_009741 PE=4 SV=1**

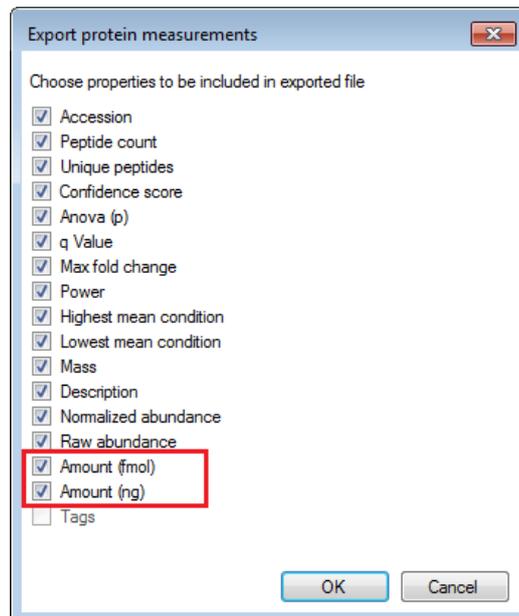
Quantifiable proteins displayed: 24

As you order on Batch 3, 5 of the proteins originally present in Batch 2 are no longer measurable in Batch 3, this rises to 13 in Batch 5.

Protein measurements are exported as .csv the content being dependant on the choices made on the dialog shown to the right.

You can export these measurements to Excel by clicking on **Export Protein Measurements**

Alternatively, right click on the table and select **Add to clip gallery...** (refer to Appendix 7 (page 59) for more details)



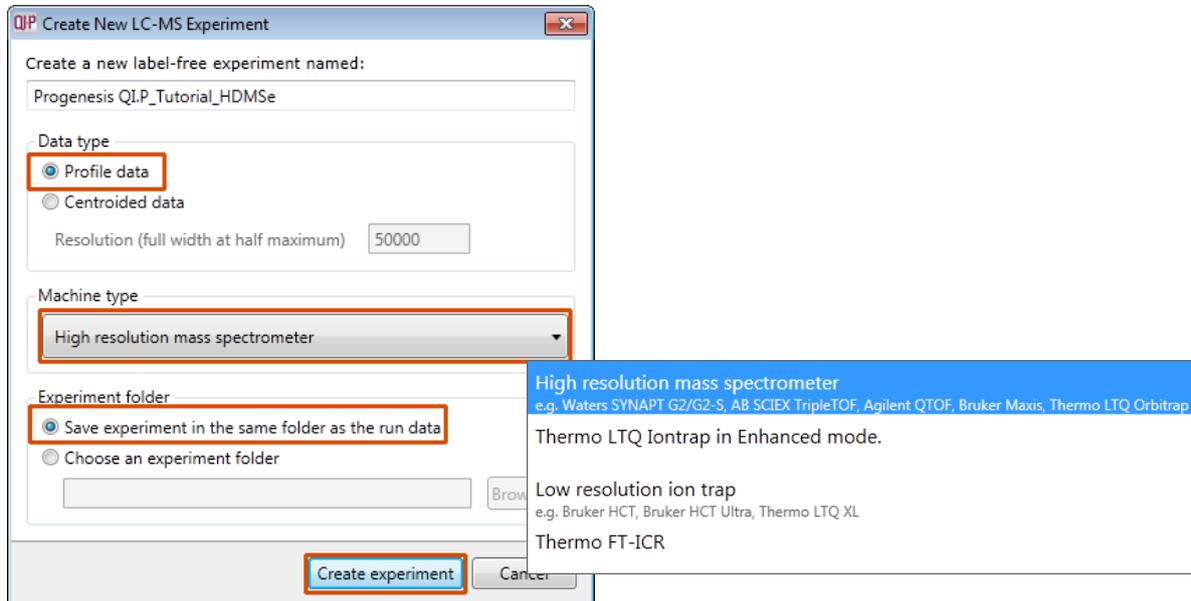
You can clear the current searches and repeat the searches with the other libraries provided or try the ones you have created.

## 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'.

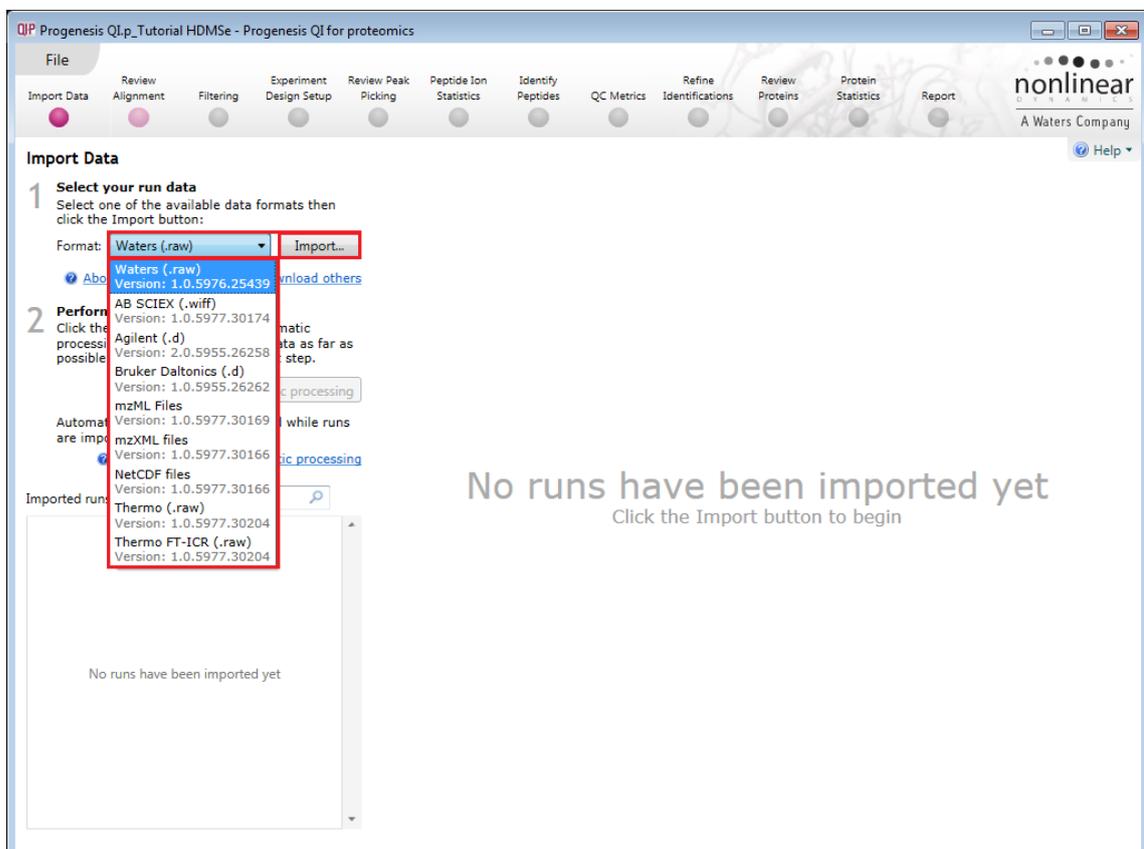
**Note:** if you have converted or captured the data as centroided then select Centroided data and enter the Resolution for the MS machine used.



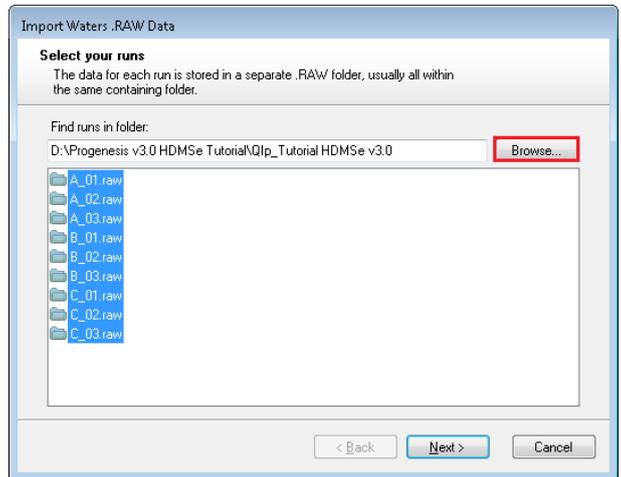
Click **Create experiment** to open the LC-MS Data Import stage of the workflow.

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

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

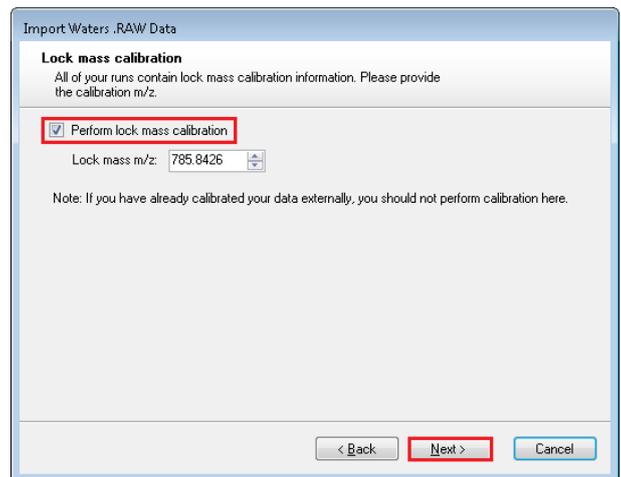


Locate and select all the .RAW folders (A\_01 to C\_03).



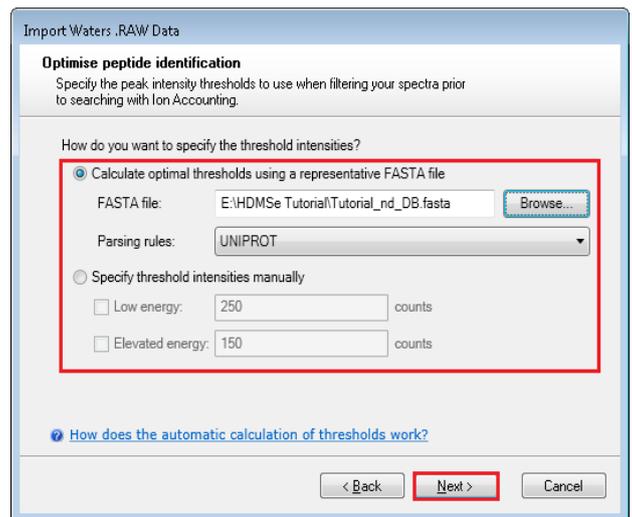
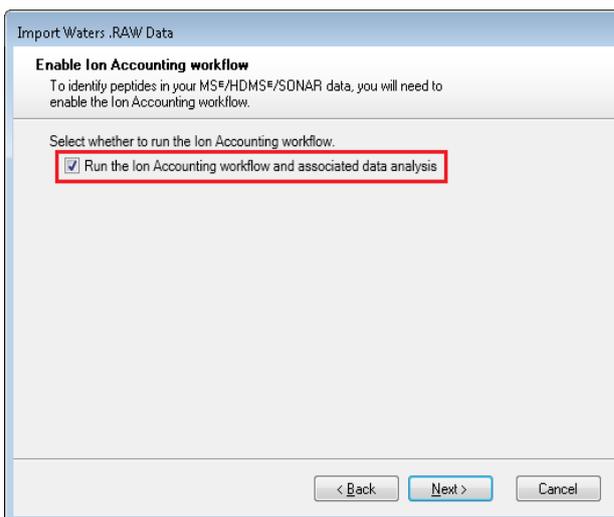
On importing, the lock mass calibration is read and presented on this dialog

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.



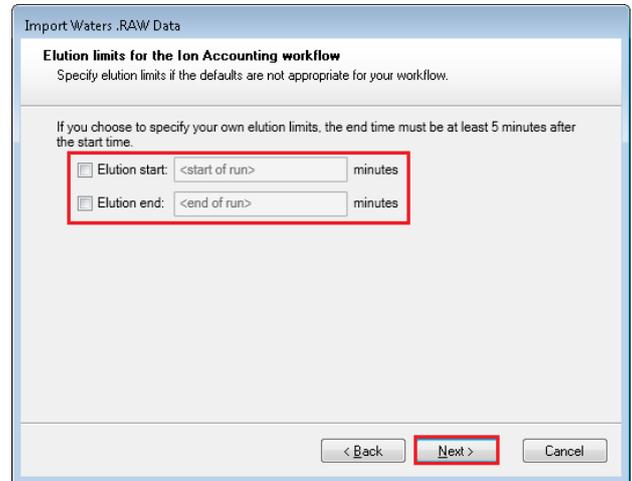
If you choose to determine the thresholds automatically then Progenesis determines appropriate thresholds by sampling each run and finding the thresholds that yield the most protein identifications in the sample area. For each run, Progenesis performs the following steps:

- 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.

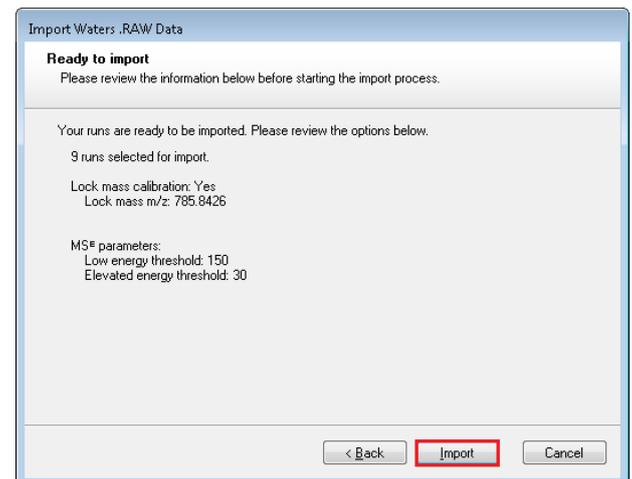
- Finally, it selects the thresholds that resulted in the largest number of protein identifications and applies those to the whole run. If more than one set of thresholds results in similar numbers of identifications (within 10% of the maximum), it will choose the highest thresholds as a way of optimising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

**Note:** for HDMS<sub>e</sub> 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.

The screenshot shows the Progenesis QI software interface. The 'Import Data' section is active, with step 1 'Select your run data' and step 2 'Perform automatic processing'. The 'Start automatic processing' button is highlighted with a red box. Two 2D chromatograms are displayed: 'A\_01 (low energy)' and 'A\_01 (high energy)'. A 'Section Complete' button is visible in the bottom right corner.

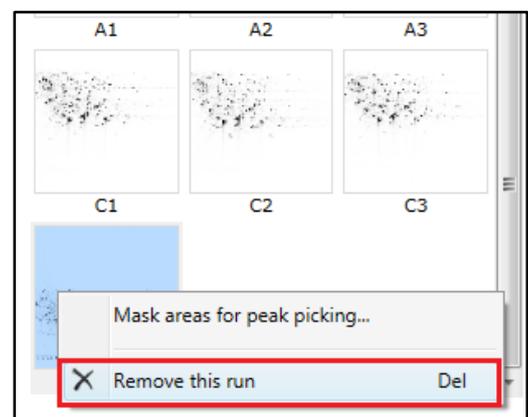
For details of setting up the steps in the automatic processing wizard return to Stage 2A page 9.

## 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

**Import Data**

**1 Select your run data**  
Select one of the available data formats then click the Import button:  
Format: Waters (.raw) Import...

**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.  
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)**  
m/z  
Retention Time (min)

**A\_01 (high energy)**  
m/z  
Retention Time (min)

**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.

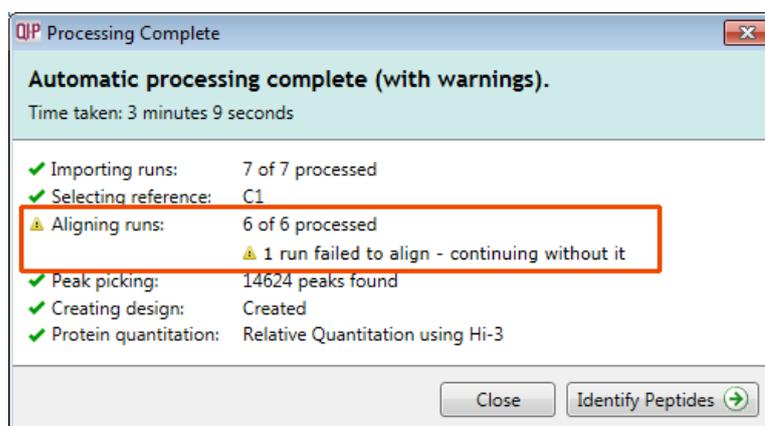
**Note:** you will be offered the automatic alignment if you have not performed it automatically already.

Now move to the next stage in the workflow (page 9 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.



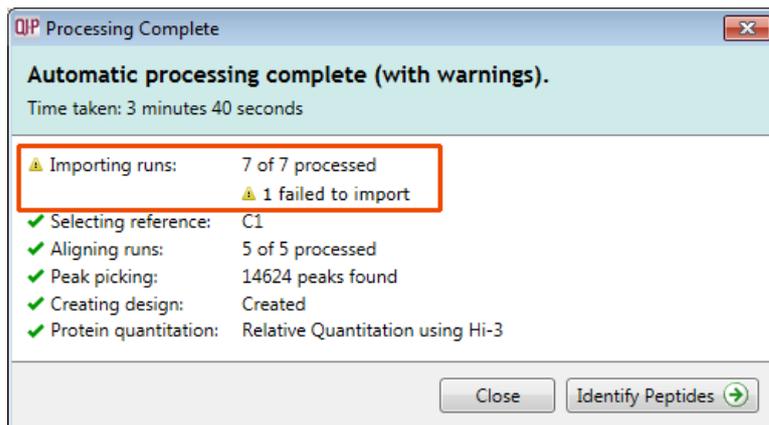
**Note:** in this example the run that failed to align will not contribute to the peak picking and will be excluded at the alignment stage (a cross appears in the include column).

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.

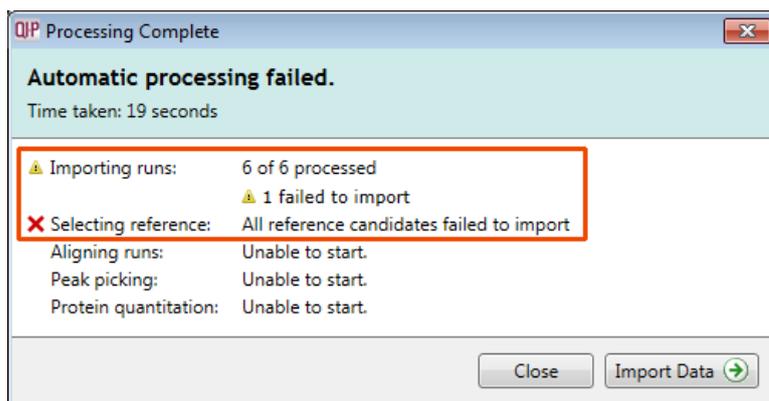
If some runs in a data set fail to import (but not all), the automatic processing will continue informing you that one or more runs have failed to import.



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

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.**

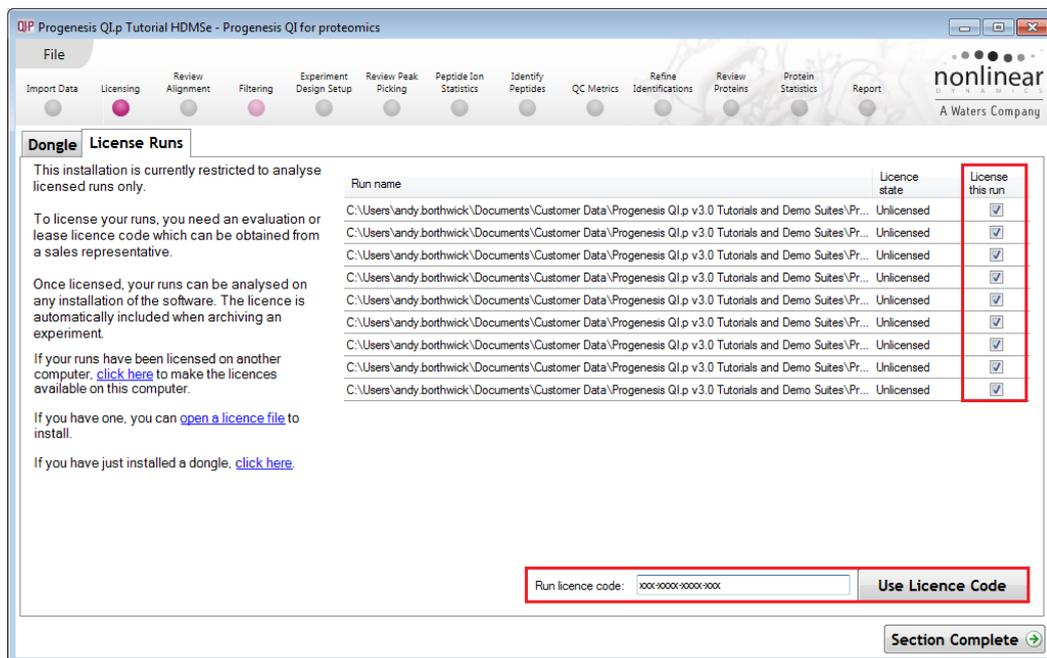
To use this page to License your Runs you must first either obtain an **'Evaluation' Licence Code from a Sales Person or purchase a licence code directly**.

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

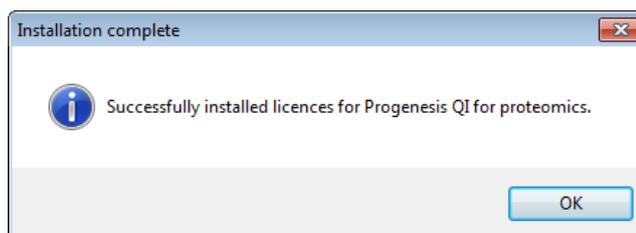
The runs in your experiment will be listed as shown below.

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>B_02</b>	<input checked="" type="checkbox"/>	0	<b>40.29</b>
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**

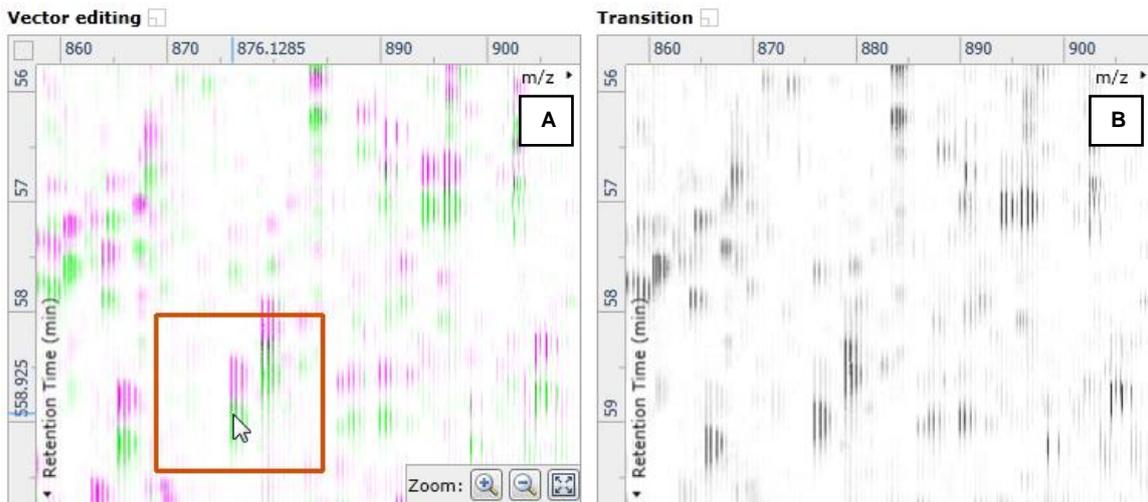
- Click on Run B\_02 in the **Runs** panel, this will be highlighted in green and the reference run (A\_01) will be highlighted in magenta.
- You will need to place approximately 5 - 10 **alignment vectors** evenly distributed from top to bottom of the whole run (RT range).
- First drag out an area on 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.

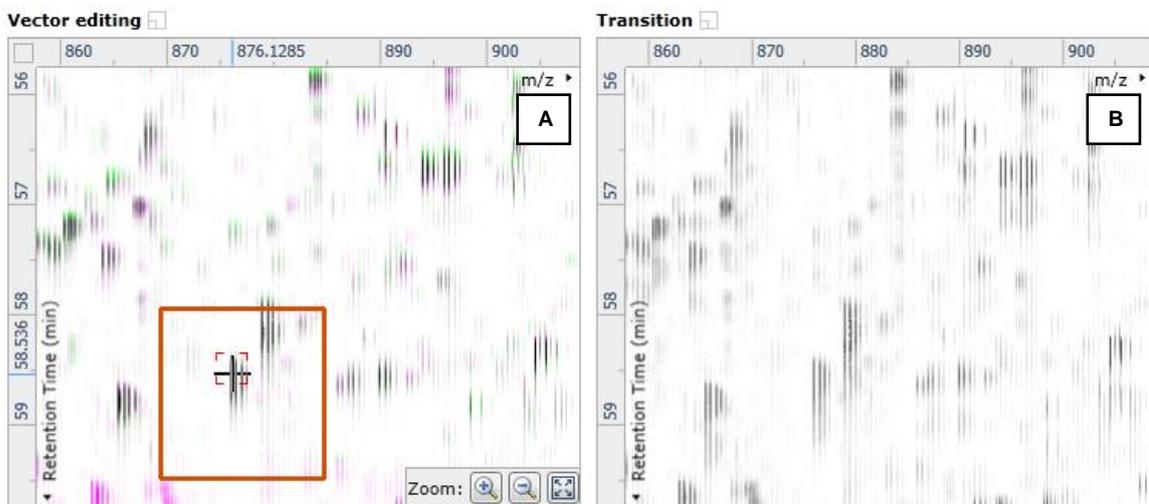
**Note:** the Ion Intensity Map gives you a colour metric, visually scoring the current alignment and an overall score is placed next to the Vectors column in the table. With each additional vector added this score will update to reflect the 'changing' overall quality of the alignment. The colour coding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

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

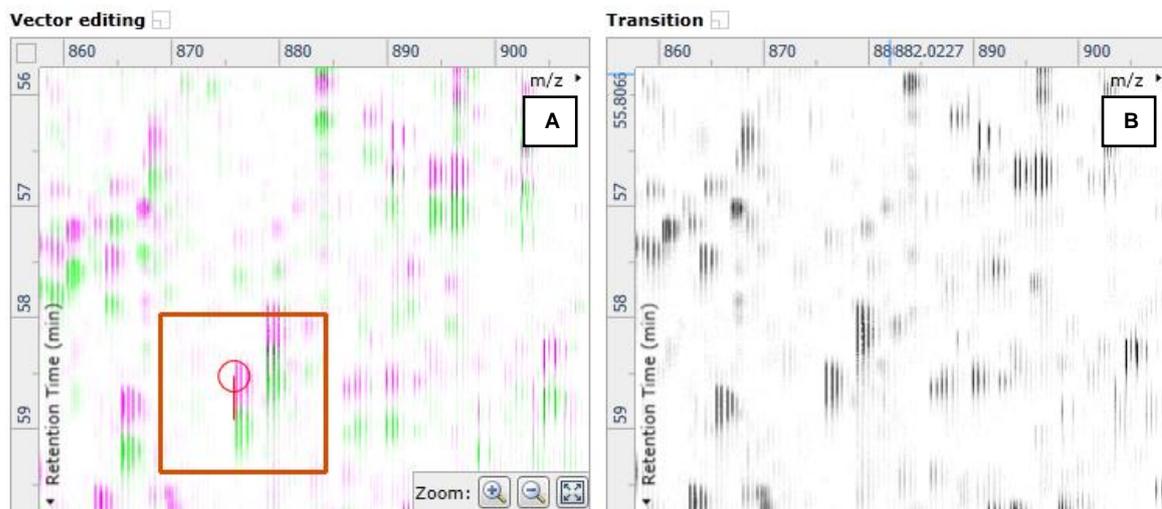


- As you are holding down the left mouse button (depending on the severity of the misalignment), the alignment vector will automatically find the correct lock. If not, drag the green peptide ion over the corresponding magenta peptide ion of the reference run. The red box will appear as shown below indicating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



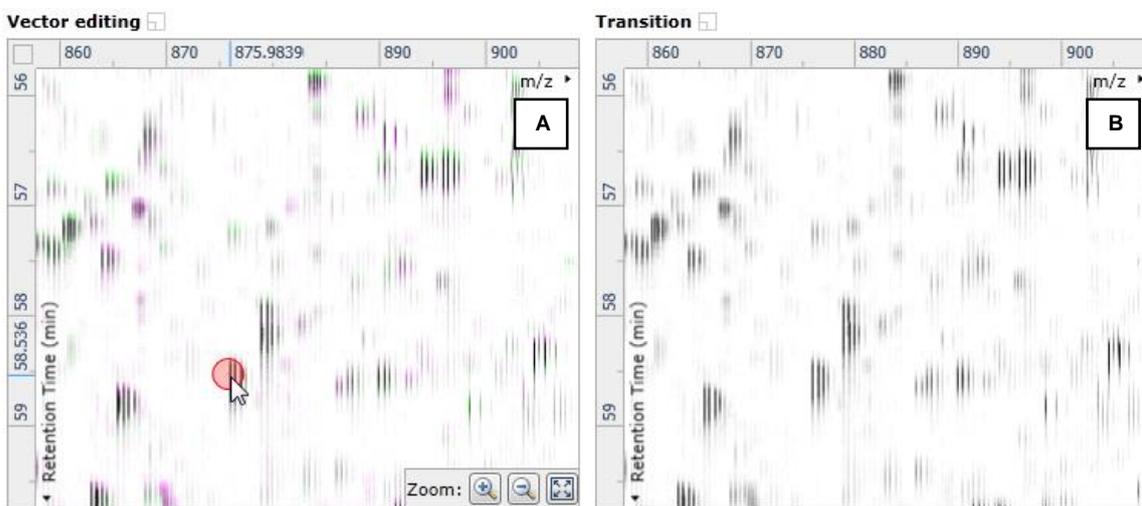
**Tip:** while holding down the mouse button hold down the **Alt** key. This will allow smooth movement of the cursor as the **Alt** key allows you to override the 'automatic alignment' performed as you depress the mouse button.

- 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.

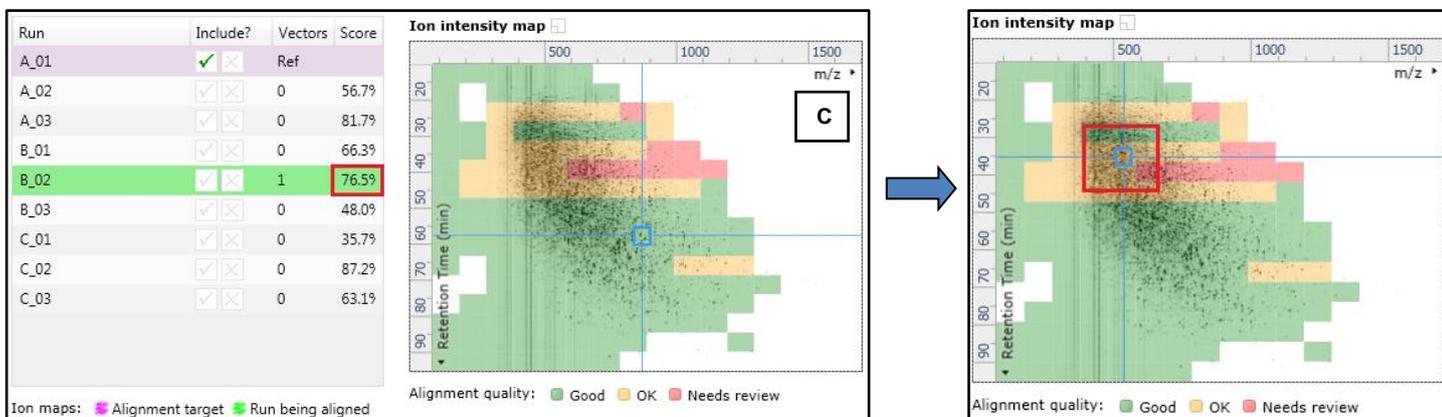


**Note:** an incorrectly placed vector is removed by right clicking on it in the **Vector Editing** window and selecting delete vector.

- 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.

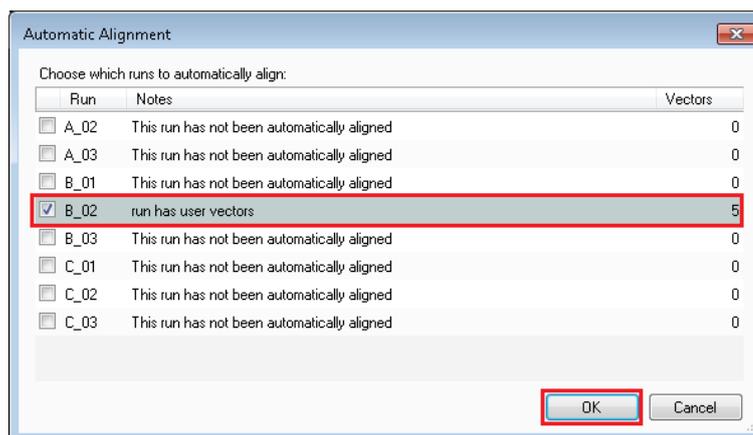


10. The shift in the Retention Time (RT) is as a result of incorrect running of the chromatography. In many of these cases if the Automatic Alignment fails to generate a good alignment then removing all the alignment vectors for this run and placing a single manual vector to act as a 'seed' for the Automatic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

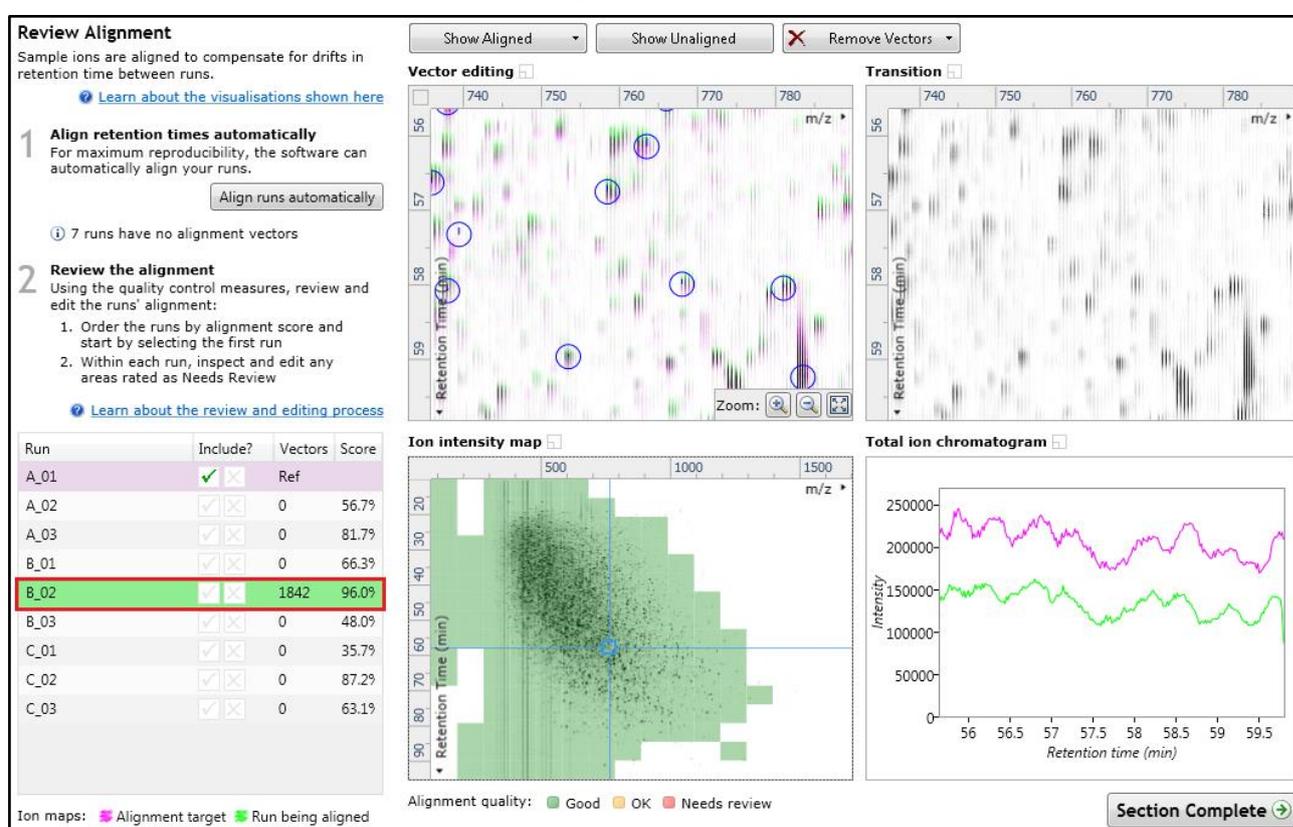
11. In the case of the example shown above placing a small number of vectors from the top to the bottom of the run is sufficient to markedly improve the alignment.



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).



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.

## 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.

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

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: \_\_\_\_\_

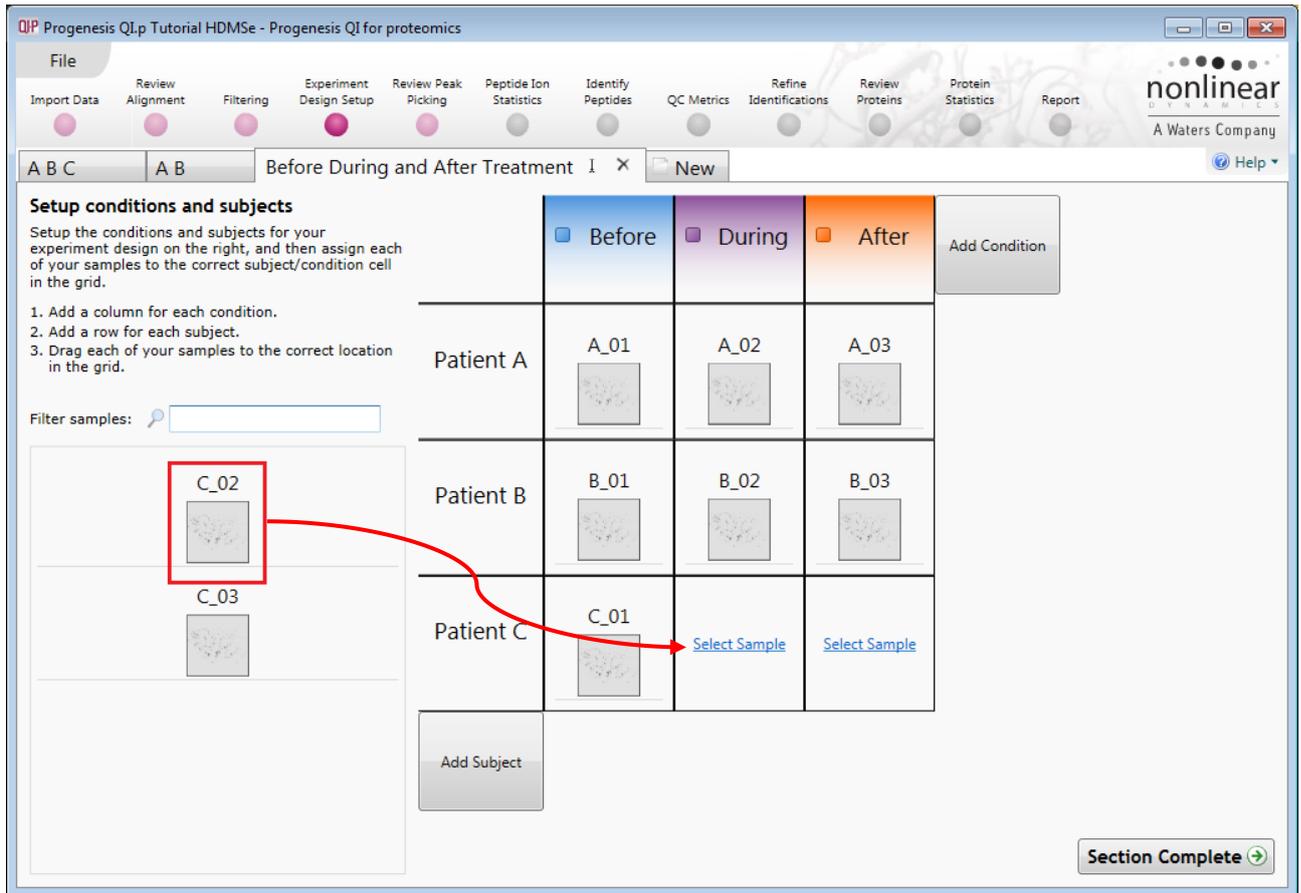
Create design Cancel

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.

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.

**Currently no Conflict resolution was performed on the HCP\_Spectral Library\_Created data set.**

**Example of performing conflict resolution on G318R9** (left hand table), which has a total number of 19 peptides, 17 of which are uniquely assigned to this protein, the remaining 2 have a total of 10 conflicting protein assignments.

The screenshot displays the Progenesis QI software interface. At the top, the 'Refine Identifications' stage is active. The main window is divided into two tables. The left table, titled 'Proteins', lists identified proteins with columns for Accession, Peptides, Unique, Conflict, Score, Tag, and Abundance. The right table, titled 'Peptide ions of G318R9', lists individual peptide ions with columns for #, I, S, Score, Hits, Correlation, Mass, Mass error, RT, Charge, Tag, Abundance, Conflicts, and Peptide Sequence. The 'Conflicts' column in the right table is highlighted in red. Below the tables, there are two plots: 'Standardised Expression Profiles' and 'Peptide ions of selected protein'. The 'Protein resolution' tab is selected, and the 'Refine Identifications' button is visible at the bottom left.

On the peptides table order on conflicts and select the first peptide (which has 3 conflicts) and click on the Protein resolution tab to show the conflicting proteins

The Conflicting proteins are ordered on the basis of Protein Score in the bottom left table

The screenshot shows the 'Conflicting proteins for peptide ion 1240' table. The table has columns for Accession, Peptides, Unique, Conflict, Protein Score, Peptide Score, and Tag. The proteins listed are G318R9, G3H4Z8, G31D7, and G3HZE5. The G3H4Z8 row is highlighted in red. The 'Protein resolution' tab is selected, and the 'Refine Identifications' button is visible at the bottom left.

Untick the peptide on the bottom right table to favour the current protein based on the highest protein score  
Then move down the Conflicting proteins table to the next protein.

Here also untick the peptide in the bottom right table.

The screenshot shows the Progenesis QI interface. At the top left, a table lists proteins with columns for Accession, Peptides, Unique, Conflict, Score, Tag, and Abundance. The protein G31D7 is highlighted. Below this, a table lists peptide ions for G31D7 with columns for #, I, S, Score, Hits, Correlation, Mass, Mass error (p...), RT (mins), Charge, Tag, Abundance, Conflict, and Peptide Sequence. The peptide ion 1240 is highlighted, and its Conflict count is 2, enclosed in a red box. Below the peptide ion table, there are two sub-tables: 'Conflicting proteins for peptide ion 1240' and 'Peptide ions of G31D7'. The 'Conflicting proteins' table shows G31D7 with a conflict count of 2. The 'Peptide ions of G31D7' table shows the peptide ion 1240 with a conflict count of 2, also enclosed in a red box.

**Note:** as you untick each Peptide ion for the conflicting protein the number of conflicts is reduced in the top 2 tables

Repeat this process for all the conflicting proteins for this peptide ion

This screenshot shows the same Progenesis QI interface as the previous one, but with the peptide ion 1240 now having a conflict count of 0, enclosed in a red box. The 'Conflicting proteins for peptide ion 1240' table now shows G31D7 with a conflict count of 0. The 'Peptide ions of G31D7' table shows the peptide ion 1240 with a conflict count of 0, also enclosed in a red box.

Now move to the second Peptide ion (with 3 conflicts) on the top right table

Then move through the Conflicting proteins on the bottom left table unticking the peptide ions on the right hand tables as appropriate

This screenshot shows the Progenesis QI interface with the peptide ion 30114 highlighted, which has a conflict count of 3, enclosed in a red box. The 'Conflicting proteins for peptide ion 30114' table shows G3H4Z8 with a conflict count of 3. The 'Peptide ions of G3H4Z8' table shows the peptide ion 30114 with a conflict count of 3, also enclosed in a red box.

The tables will update to reflect the resolved conflicts.

Accession	Peptides	Unique	Conflict	Score	Tag	Abundance
G3H5Q0 (+2)	12	1	12	123		9.46E+04
G3IDL7	8	6	6	106		9.39E+04
G3BR9	19	18	4	207		6.73E+04
G3IDL8	4	0	9	37.6		---
G3RPL1	3	0	9	35.6		---
G3BH0	3	0	8	27.1		---

#	I	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflicts	Peptide Sequence
1240	1	8.95	2	0.999	1227.617	-3.36	5.21	2		7.4E+04	0	VEIANDQGNR
30114	0	---	---	0.997	1227.61	-8.49	5.21	3		1.07E+03	0	VEIANDQGNR
2682	0	9.37	2	1.000	1658.876	-7.15	13.8	3		4.13E+04	2	IINEPTAAAIAYGLDK
1210	2	9.37	2	0.999	1658.875	-7.55	13.8	2		1.58E+05	2	IINEPTAAAIAYGLDK
1703	1	8.79	2	1.000	1216.616	-6.46	12.3	2		5.69E+04	0	DAGTIAGLNVMR
1071	2	0.00	2	0.000	1535.786	-7.78	16.7	2		8.02E+04	0	TPDPPTTSMVLTW

Protein: G3BR9 78 kDa glucose-regulated protein OS=Cricetulus griseus GN-I79\_019946 PE=1 SV=1  
 Protein: G3HZE5 Heat shock 70 kDa protein 1A OS=Cricetulus griseus GN-I79\_016446 PE=3 SV=1

Accession	Peptides	Unique	Conflict	Protein Score	Peptide Score	Tag
G3BR9	19	18	4	207		
G3H4Z8 (+4)	9	2	15	108		
G3IDL7	8	6	6	106		
G3HZE5	3	0	8	31.5		

#	I	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict	Peptide Sequence
2932	2	9.06	2	1.000	1486.684	-6.61	9.81	2		4.25E+04	2	ITFSYVAFDTER
12860	0	8.03	2	0.998	803.432	-8.28	3.13	2		2.13E+03	2	ITINDK
7832	0	7.21	2	0.999	2773.307	-4.64	15.5	3		2.84E+04	2	QTQFTTYSNDQPGVLIQVVEGER
12769	0	7.21	2	0.999	2773.323	1.26	15.5	2		1.92E+04	2	QTQFTTYSNDQPGVLIQVVEGER
30114	0	---	---	---	1227.61	-8.49	5.21	3		1.07E+03	0	VEIANDQGNR
1240	1	8.95	2	---	1227.617	-3.36	5.21	2		7.4E+04	0	VEIANDQGNR

Repeat this process until there are no remaining conflicts on the top 2 tables for the current protein G318R9.

Progenesis QI for Proteomics

Proteins: No filter applied

Accession	Peptides	Unique	Conflict	Score	Tag	Abundance
G3H5Q0 (+2)	12	1	12	123		9.46E+04
G3IDL7	7	6	2	87.7		9.39E+04
G3BR9	19	19	0	207		6.73E+04
G3IDL8	4	0	9	37.6		---
G3RPL1	3	0	9	35.6		---
G3BH0	3	0	8	27.1		---

Protein: G318R9 78 kDa glucose-regulated protein OS=Cricetulus griseus GN-I79\_019946 PE=1 SV=1  
 Protein: G3IDL7 Heat shock cognate 71 kDa protein OS=Cricetulus griseus GN-I79\_021800 PE=3 SV=1

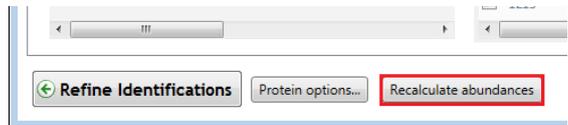
Accession	Peptides	Unique	Conflict	Protein Score	Peptide Score	Tag
G3BR9	19	19	0	207		9.37
G3H4Z8 (+4)	8	2	11	89.2		9.37
G3IDL7	7	6	2	87.7		9.37

#	I	Score	Hits	Correlation	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict	Peptide Sequence
1240	1	8.95	2	---	1227.617	-3.36	5.21	2		7.4E+04	0	VEIANDQGNR
2682	0	9.37	2	---	1658.876	-7.15	13.8	3		4.13E+04	0	IINEPTAAAIAYGLDK
1210	2	9.37	2	0.999	1658.875	-7.55	13.8	2		1.58E+05	0	IINEPTAAAIAYGLDK
1703	1	8.79	2	1.000	1216.616	-6.46	12.3	2		5.69E+04	0	DAGTIAGLNVMR
1071	2	0.00	2	0.000	1535.786	-7.78	16.7	2		8.02E+04	0	TPDPPTTSMVLTW

Refine Identifications Protein options... Recalculate abundances Section Complete

To Resolve the conflicts for the whole data set work through this process with all of the proteins in the top left hand table, that display conflicts, until there are no remaining conflicts.

**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 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 2) then this option will be applied.

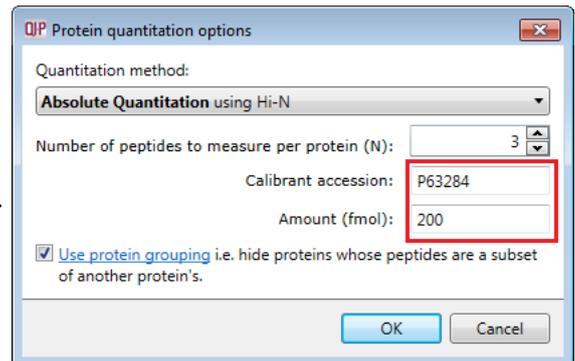
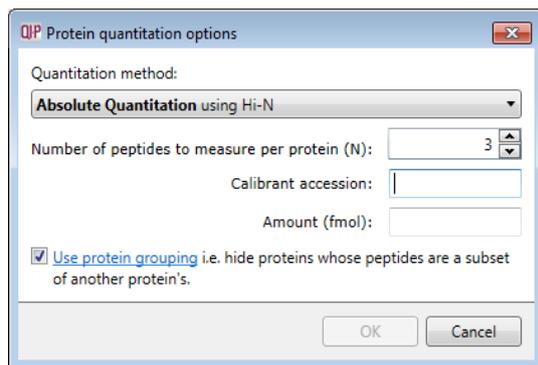
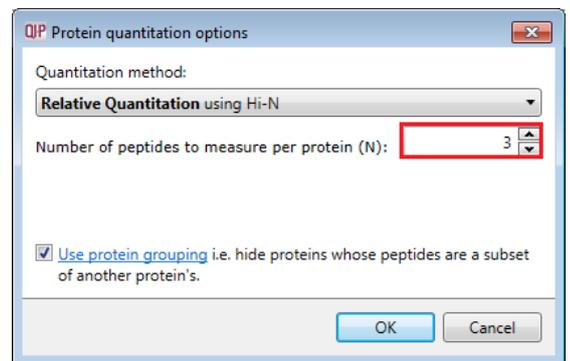
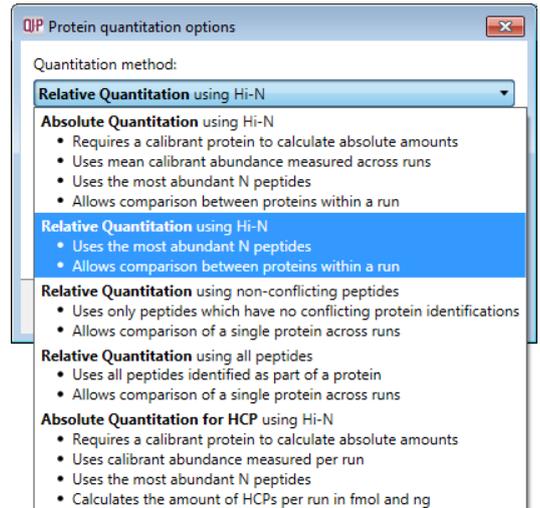
Hi-N in Progenesis QI for proteomics is an implementation of Hi-3 as described by Silva *et al.* [References]. After peptide and protein identification, the abundance of each peptide is calculated from all its constituent peptide ions.

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.

Description	Amount (fmol) - Batch 1	Amount (fmol) - Batch 2	Amount (fmol) - Batch 3	Amount (fmol) - Batch 4	Amount (fmol) - Batch 5
L-lactate dehydrogenase A chain OS=Cricetulus griseus GN=I79_009741 PE=4 SV=1	19.6	8.62	0	0	0
Nucleolin OS=Cricetulus griseus GN=I79_022400 PE=4 SV=1	40.1	16.8	0	0	0
Peptidyl-prolyl cis-trans isomerase OS=Cricetulus griseus GN=I79_001023 PE=3 SV=1	155	77.1	8.95	5.71	4.83
Peroxisome oxidin-1 OS=Cricetulus griseus GN=I79_002954 PE=4 SV=1	150	67.7	1.2	0	0

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.

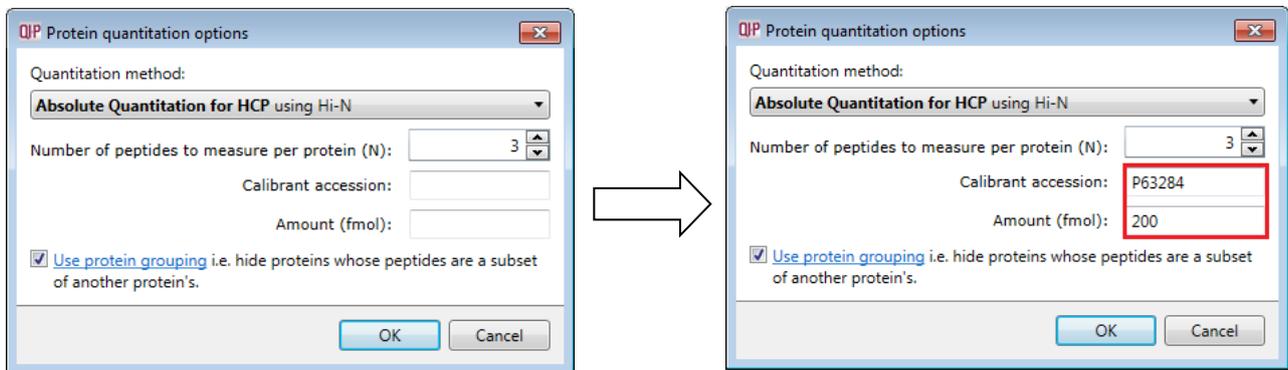
To generate values for **Absolute Quantitation for HCP** (Host Cell Proteins) select the 5<sup>th</sup> option on the drop down at the beginning of this section.

**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) - Batch 1	Amount (fmol) - Batch 2	Amount (fmol) - Batch 3	Amount (fmol) - Batch 4	Amount (fmol) - Batch 5	Amount (ng) - Batch 1	Amount (ng) - Batch 2	Amount (ng) - Batch 3	Amount (ng) - Batch 4	Amount (ng) - Batch 5
L-lactate dehydrogenase A chain OS=Cricetulus griseus GN=i79_009741 PE=4 SV=1	194	9.17	0	0	0	0.24	0.113	0	0	0
Nucleolin OS=Cricetulus griseus GN=i79_022400 PE=4 SV=1	39.5	17.7	0	0	0	2.07	0.93	0	0	0
Peptidyl-prolyl cis-trans isomerase OS=Cricetulus griseus GN=i79_001023 PE=3 SV=1	153	82.3	12.1	6.16	4.03	2.58	1.39	0.204	0.104	0.0682
Peroussidou-1 OS=Cricetulus griseus GN=i79_002954 PE=4 SV=1	148	71.9	1.57	0	0	3.33	1.62	0.0353	0	0

As 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.

**How the values for the calibrant behave across runs comparing Absolute Quant and Absolute Quant for HCP using the same calibrant**

Description	Amount (fmol) - Batch 1	Amount (fmol) - Batch 2	Amount (fmol) - Batch 3	Amount (fmol) - Batch 4	Amount (fmol) - Batch 5
Calreticulin OS=Cricetulus griseus GN=H671_3g9525 PE=3 SV=1	158	65.1	0.0418	1.92E-05	0.000136
Chaperone protein ClpB OS=Escherichia coli (strain K12) OX=83333 GN=clpB PE=1 SV=1	203	188	151	196	202
Elongation factor 1-alpha 1 OS=Cricetulus griseus GN=i79_009935 PE=4 SV=1	715	489	371	344	390

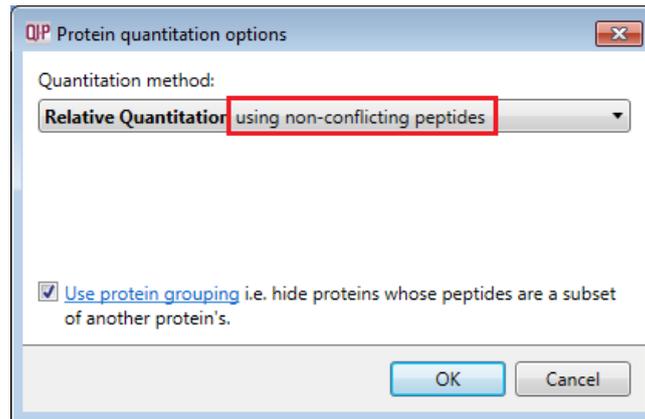
Description	Amount (fmol) - Batch 1	Amount (fmol) - Batch 2	Amount (fmol) - Batch 3	Amount (fmol) - Batch 4	Amount (fmol) - Batch 5	Amount (ng) - Batch 1	Amount (ng) - Batch 2	Amount (ng) - Batch 3	Amount (ng) - Batch 4	Amount (ng) - Batch 5
Calreticulin OS=Cricetulus griseus GN=H671_3g9525 PE=3 SV=1	155	68.3	0.0509	1.95E-05	0.000105	7.52	3.31	0.00246	9.44E-07	5.06E-06
Chaperone protein ClpB OS=Escherichia coli (strain K12) OX=83333 GN=clpB PE=1 SV=1	200	200	200	200	200	19.2	19.2	19.2	19.2	19.2
Elongation factor 1-alpha 1 OS=Cricetulus griseus GN=i79_009935 PE=4 SV=1	706	520	496	354	298	39.2	28.9	27.6	19.7	16.5

**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.

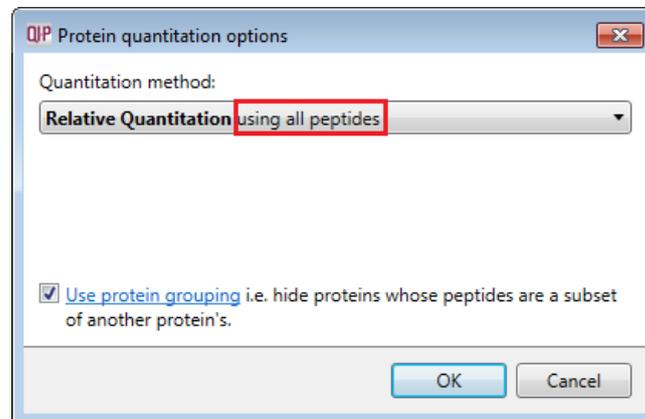
Naturally, if the conflicted peptides are not among the N most abundant in either protein initially, then this will not cause any difference in the result.

**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.

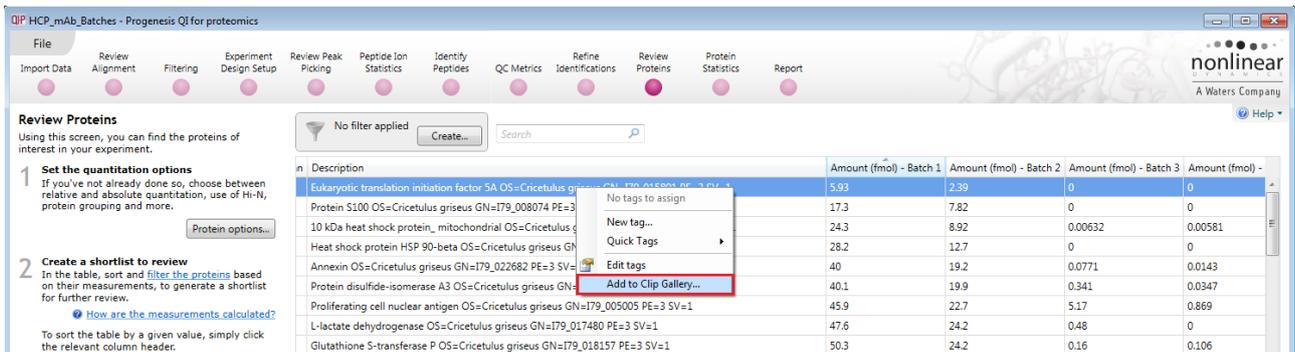
Now move to the Review Proteins section by clicking on **Review Proteins** icon on the workflow at the top of the screen.

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

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

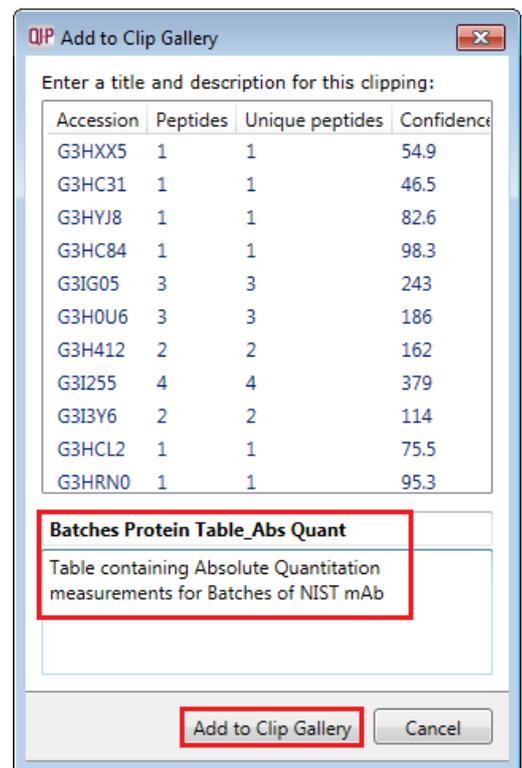
The saved images of the Views and the tables are retained as part of the experiment and are stored accordingly. This facility allows you to capture (high resolution) images that can be used in the development of specific reports and/or used as part of the process of publishing your experimental findings.

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...**



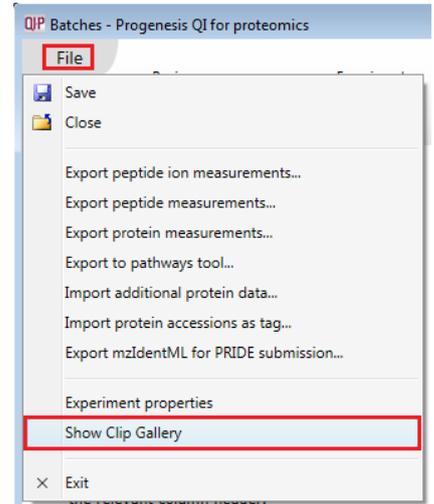
This will open a dialog displaying what is to be saved and allows you to alter the title and provide a description of the item for later reference.

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

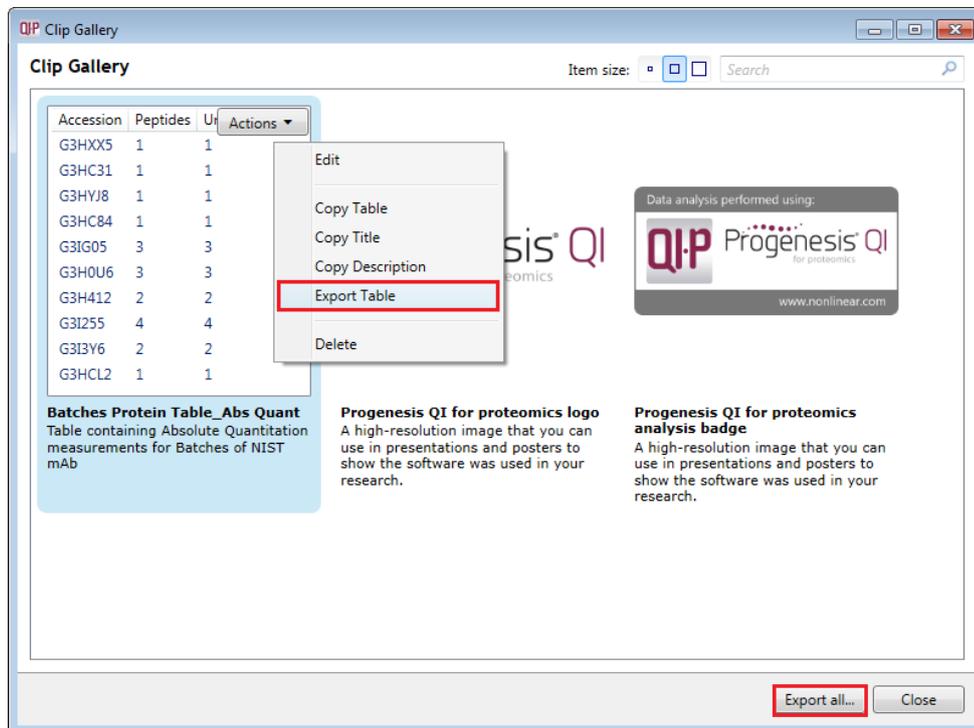


To view, edit and/or export from the clip galley the gallery can be accessed from the **File** menu.

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.

## 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<sup>F</sup>