Progenesis LC-MS Fractionation User Guide

Analysis workflow guidelines

for version 3.0

Contents

Introduction 3
How to use this document 3
How can I analyse my own runs using LC-MS? 3
LC-MS Data used in this user guide 3
Workflow approach to LC-MS run analysis 4
Overview of a typical Fractionation Analysis workflow 5
Analysing a single fraction 6
Stage 1: Data import and QC review of LC-MS data set for a fraction 7
Stage 2: Reference Run selection 7
Stage 3: Alignment 8
Stage 4: Filtering 10
Stage 5: Experiment Design Setup for Analysed Runs 11
Stage 6: Validation, review and editing of results 12
Stage 7: Peptide Search 13
Stage 8: Peptide Filter 14
Stage 9: Protein View 15
Completion of Fractionation Workflow Step 1 18
Fractionation Workflow Step 2 19
Stage 1 Import Fractions 19
Stage 2 Recombine Samples 21
Stage 3 Experiment Design Setup 23
Stage 4 Protein View 24
Stage 5 Saving a Multi-Fraction experiment 26
Appendix 1: Manual assistance of Alignment 27
Appendix 2: Within-subject Design 32

Introduction

This user guide takes you through the processes involved in the analysis of a fractionated label free LC-MS experiment. Where the method of fractionation can be gel based on electrophoresis: 1D, 2D, Off gel etc or chromatography using an additional LC step prior to the LC-MS.

In this example, to demonstrate the processes at each stage, an experiment using samples that have been fractionated using ion exchange chromatography is described. As each fraction constitutes the full application of the label free workflow as described in the main LC-MS userguide a shortened version of these analysis steps are described here.

It starts with LC-MS data file loading then Alignment, followed by Analysis that creates a list of interesting features (peptides) which can be explored within Progenesis Stats using multivariate statistical methods, then onto Protein identity.

How to use this document

This document is designed to be used as a guide to the processes involved in the analysis of fractionated samples. Currently a full data set is not provided as this would result in a considerable download. The initial section of the document is concerned with an abbreviated description of the main experimental workflow as applied to a single fraction (a more comprehensive description of the main analysis workflow is available in the main Userguide). The second section describes the process of recombining these individual fraction experiments into a 'Multi-fraction' experiment.

How can I analyse my own runs using LC-MS?

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 LC-MS. Instructions on how to do this are included in a section at the end of the user guide document. Alternatively if you would like to arrange a demonstration in your own laboratory contact support@nonlinear.com and we will help you.

LC-MS Data used in this user guide

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

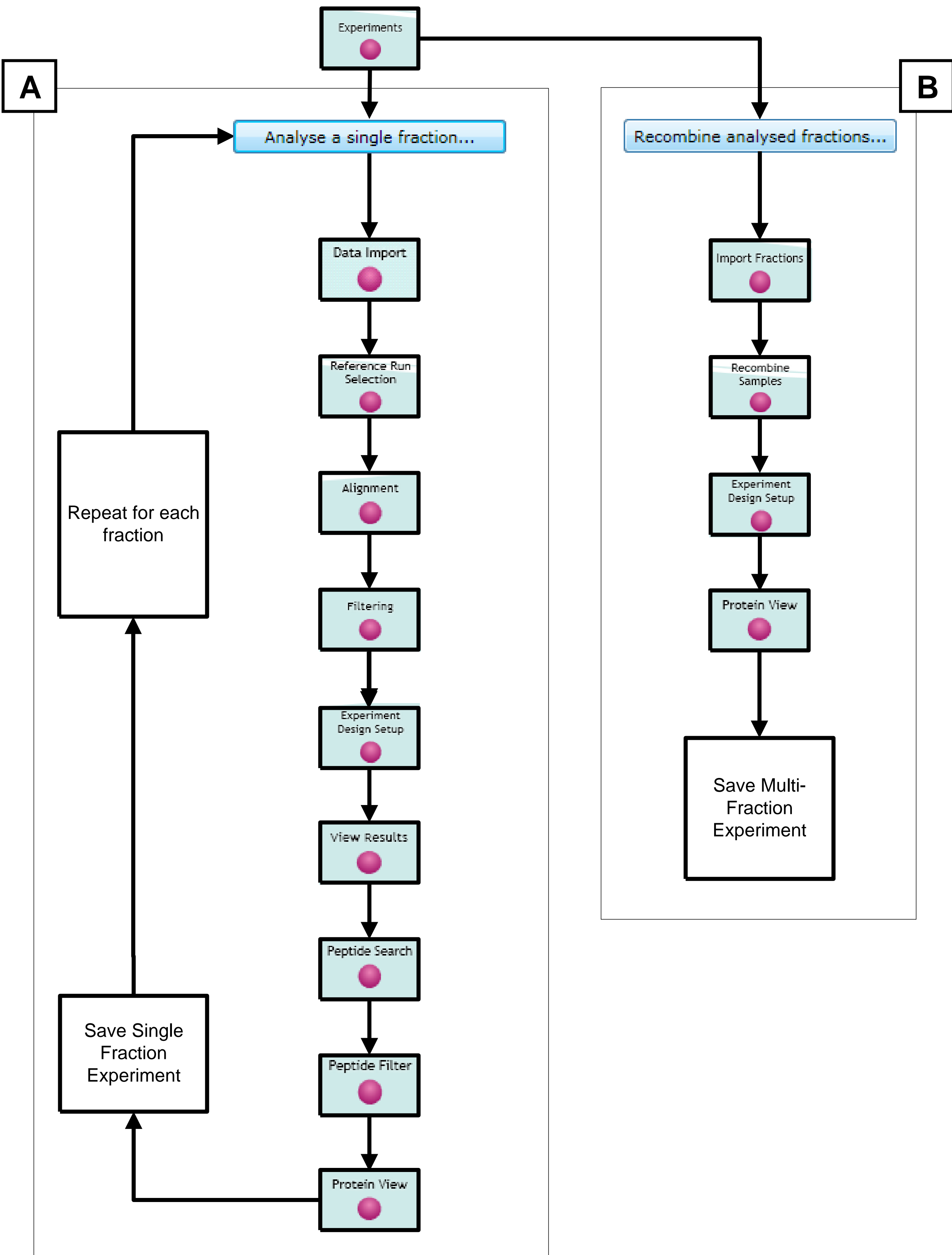
Workflow approach to LC-MS run analysis

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

Stage	Description	Page
Data Import	**LC-MS Data Import:** Selection and review of data files for single fraction analysis.	7
Reference Run Selection	**Reference Run Selection**: Select run to align to.	7
Alignment	**Alignment**: automatic and manual run alignment	8
Filtering	**Filtering**: defining filters for features based on Retention Time, m/z , Charge and Number of Isotopes.	10
Experiment Design Setup	**Experiment Design Setup**: defining one or more group set ups for analysed aligned runs	11
View Results	**View Results**: review and validate results, edit feature detection, tag groups of features and select features for further analysis	12
Peptide Search	**Peptide Search**: managing export of MS/MS spectra to, and import of peptide ids from Peptide Search engines	13
Peptide Filter	**Peptide Filter**: manage peptide ids and filters	14
Protein View	**Protein View**: validation and resolution of peptide id conflicts for data entered from Database Search engines	15
Import Fractions	**Import fractions:** import multiple analysed fractionated experiments	19
Recombine Samples	**Recombine samples:** regenerate samples from fractions	21
Experiment Design Setup	**Experiment design Setup:** define original experimental design	23
Protein View	**Protein View:** review protein and peptide identity	24

Overview of a typical Fractionation Analysis workflow

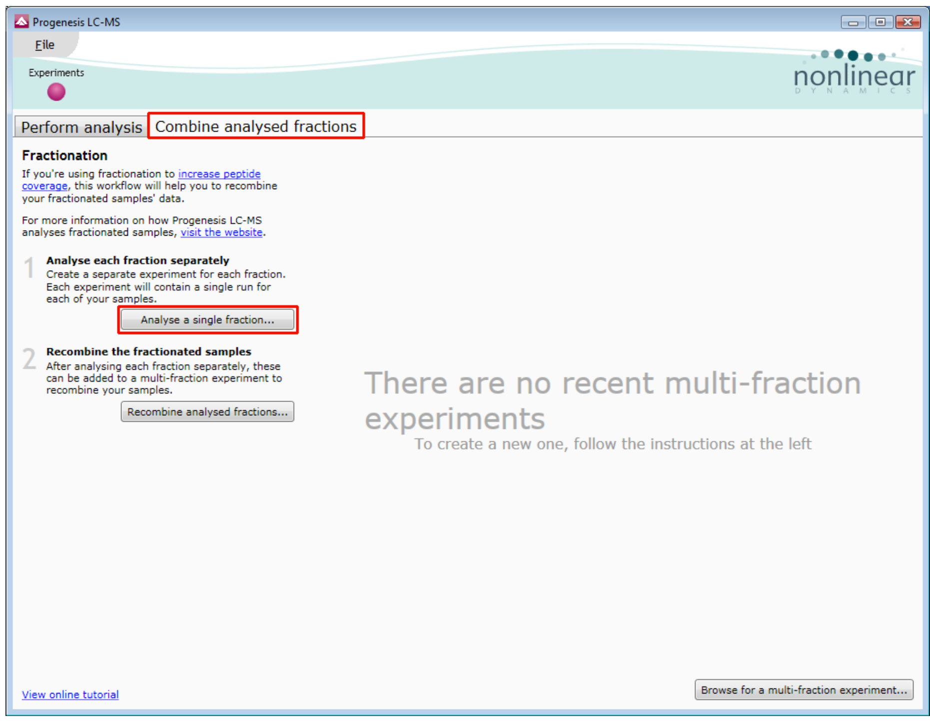
The workflow outlines the steps performed in the analysis of a typical fractionated experiment. In this example there were 6 fractions hence the main analysis workflow was performed 6 times. To analyse the data, select the **Combine analysed fractions** tab then (A) Analyse a single fraction then (B) Recombine the analysed fractions into a multi-fraction experiment.



Analysing a single fraction

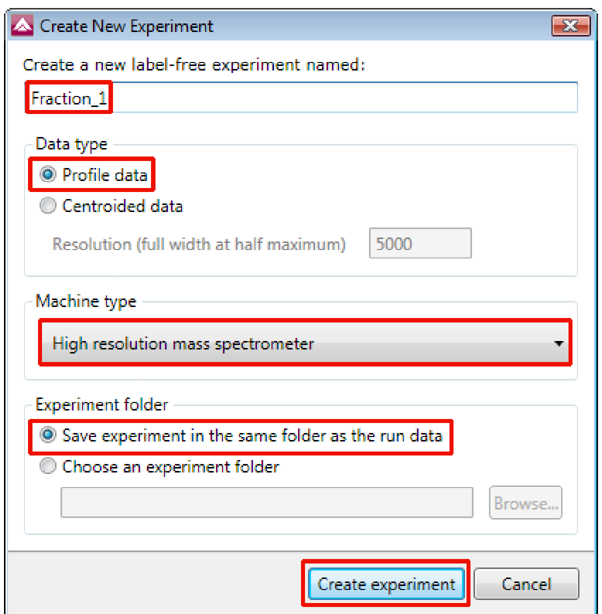
Open Progenesis LC-MS and click on the **Combine analysed fractions** tab to start the processing of Fractionated Samples.

To start the analysis of the LC-MS runs for a fraction, click on **Analyse a single fraction...**



This opens the 'Create New Experiment' dialog.

Name the fraction to be analysed (**Fraction_1**) then adjust the Data and Machine types accordingly and set the Experiment folder as required .

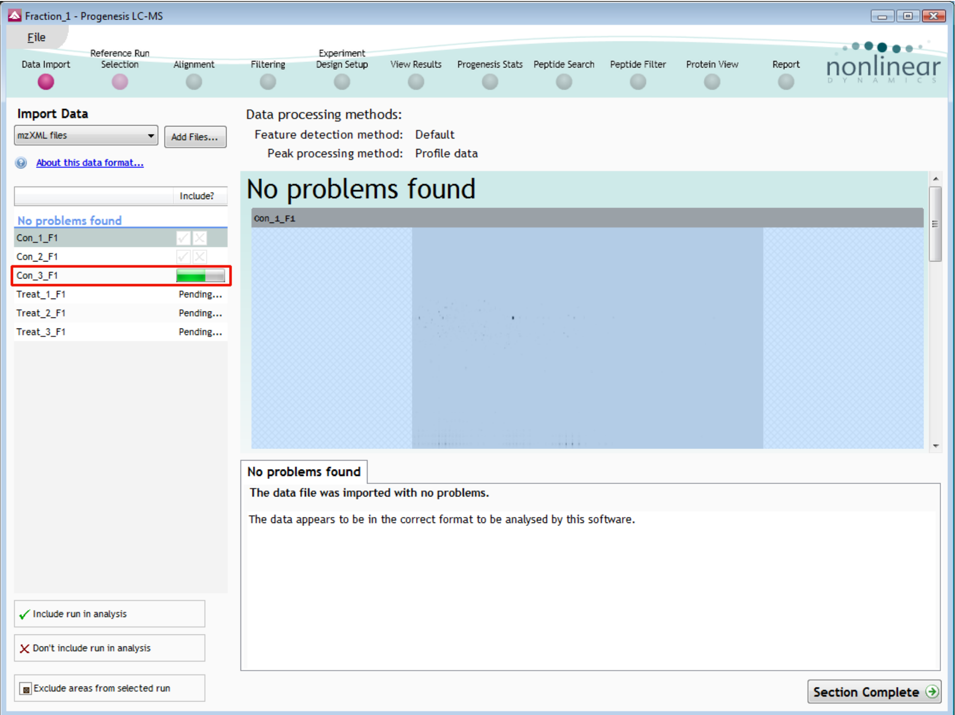


Note: current settings shown are the defaults

Click Create experiment

Stage 1: Data import and QC review of LC-MS data set for a fraction

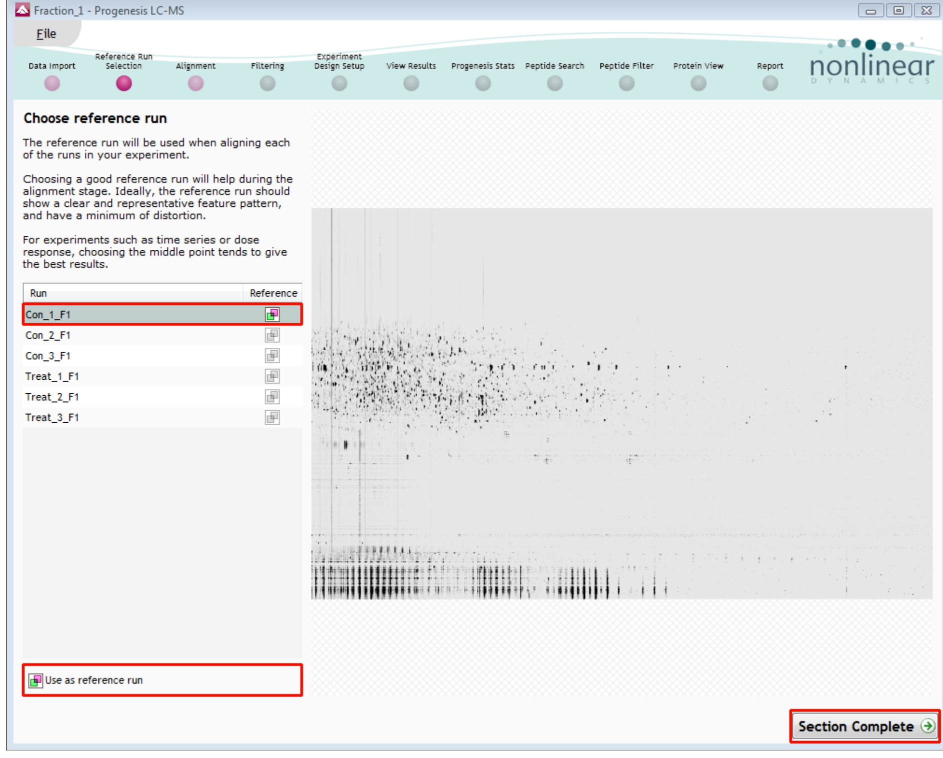
All LC-MS runs for the first Fraction are loaded into a new LC-MS experiment.



Once all the files have been imported move to the next stage in the workflow by clicking **Section Complete.**

Stage 2: Reference Run selection

This stage in the analysis workflow allows you to review and select the most appropriate Reference LC-MS run to align all the other runs to.



To select a Reference run either click on the run in the list and then click **Use as reference run** or double click on the run in the list.

Now move to the next stage in the workflow by clicking **Section Complete.**

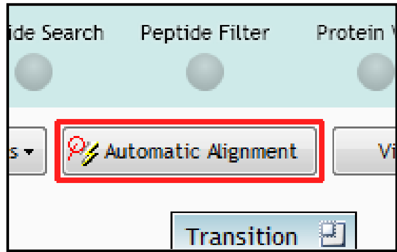
Stage 3: Alignment

At this stage Progenesis LC-MS Alignment opens displaying your data.

Generation of alignment vectors

The alignment of LC-MS runs is required in the LC (retention time) direction, this is key to correcting for the variable elution of peptides during the chromatographic separation.

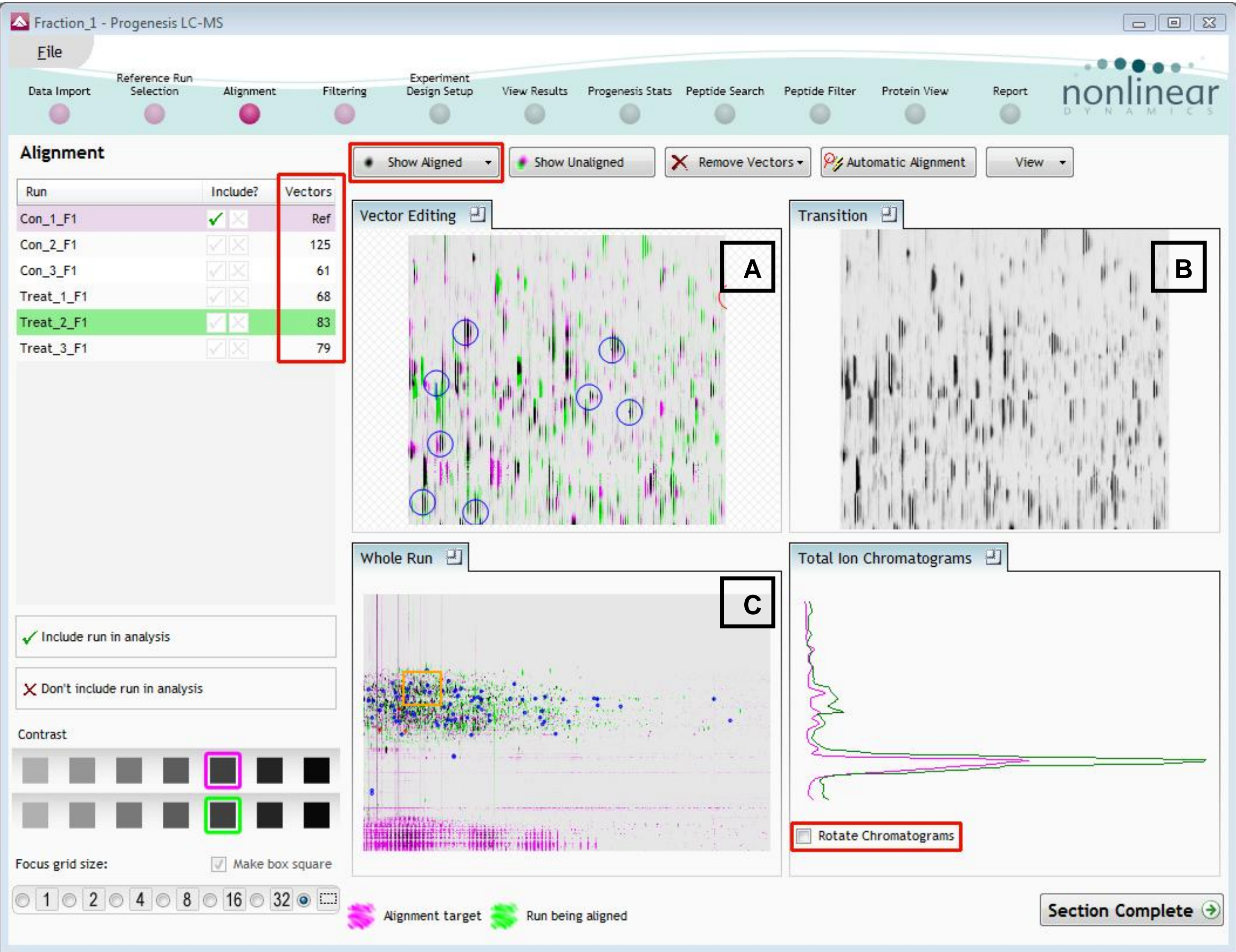
The alignment vectors are generated automatically for all the LC-MS runs by using the 'Automatic vector wizard' accessed by clicking on **Automatic Alignment** on the top tool bar, and making sure all the runs are selected before clicking OK.



Reviewing generation of alignment vectors

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

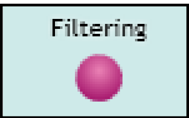
If the alignment has worked well then in Windows A and C the grid lines should show minimal distortion, Window B (Transition) will show features pulsing slightly but not moving up and down.



At this point, you should check the automatically placed (blue) vectors. This will be easier with a larger grid size. Make sure the grid size is set to 4 using the **'Focus grid size'** control at the bottom left of the window.

In each square, you can, if required edit the vectors to improve the run alignment (for more detailed information on performing the alignment of your runs refer **Appendix 1** (page 27) and also to the main LC-MS analysis Userguide.

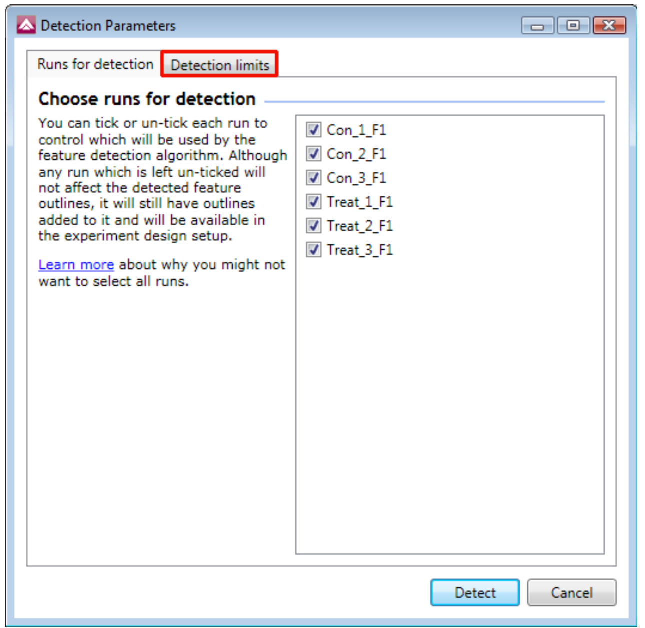
Stage 4: Filtering

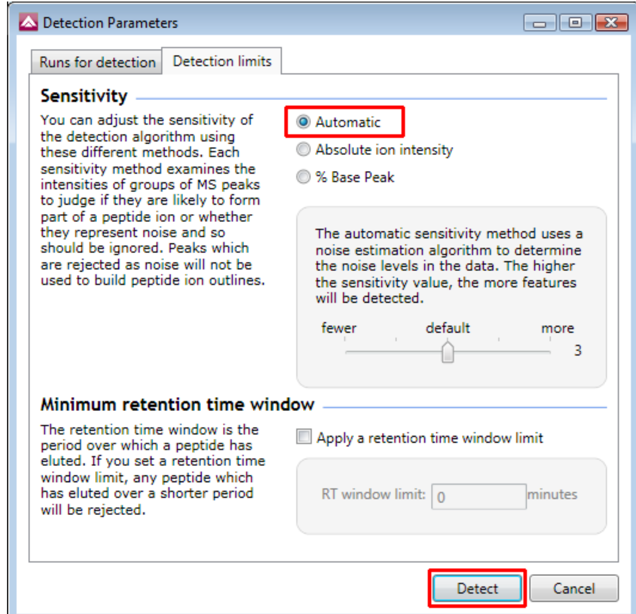


Now that you have reviewed your aligned Runs, you are ready to analyse them. Move to the **Filtering** stage, by either clicking on **Section Complete** (bottom right) or on Filtering on the workflow.

Detection Parameters

The Detection Parameters dialog, showing all the runs in the current experiment and a tick against each run, opens. This is the default setting, where the feature detection algorithm uses information from all of the runs to contribute to the pattern of feature outlines.

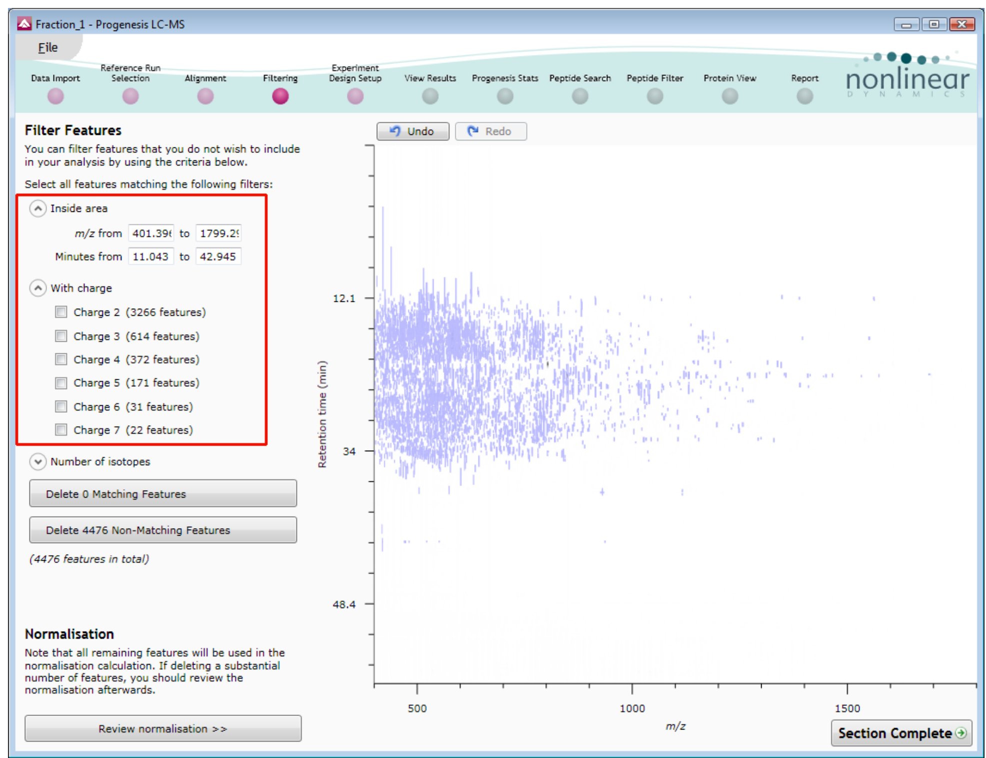




For the runs in this example the default settings for the **Automatic** method were applied.

Press **Detect** to start the detection process.

More details on the management of sensitivity are available in the How to do on **Adjusting the Sensitivity of Feature detection.**



On completion of analysis the Filtering stage will open displaying the number of features.

If required you can remove features based on position, charge state, number of isotopes or combinations of these feature properties.

The example shows the removal of features with a charge state of **less than 2 and greater than 7** and outside a defined retention time window

Having removed features the Normalisation will recalculate as you move to the next section.

Stage 5: Experiment Design Setup for Analysed Runs

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

There are two basic types of experimental designs:

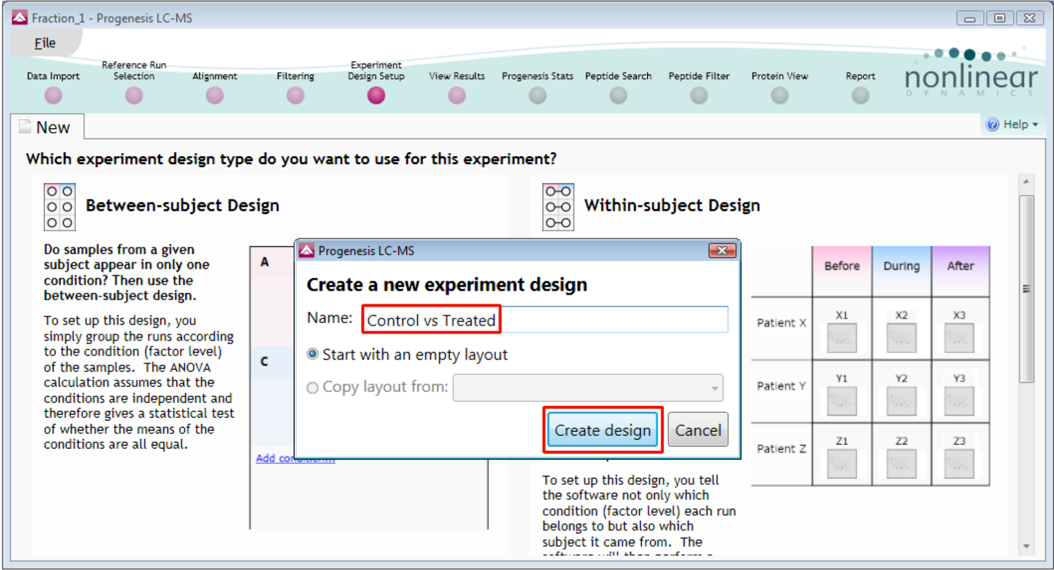
Between-subject design: where samples from any given subject appear in only one condition.

Within-subject design: where samples have been taken from a given subject under different conditions

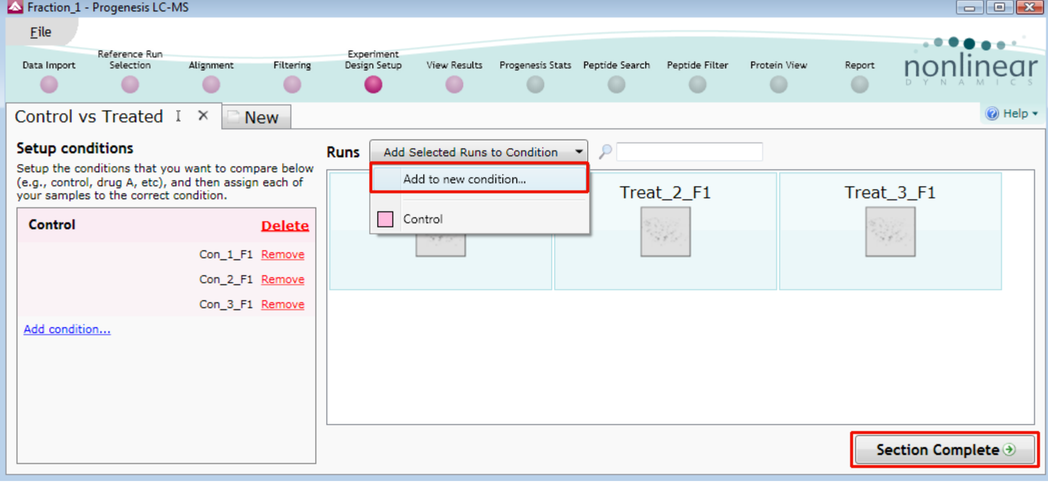
(Additional information on how to apply the Within-subject Design is in **Appendix 2** page 32)

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

Select **Between-subject** and give design an appropriate name.



Highlight the runs, to add them on to a new condition by clicking on **Add Selected Runs to Condition**



Click **Section Complete** to move to View Results.

Stage 6: Validation, review and editing of results

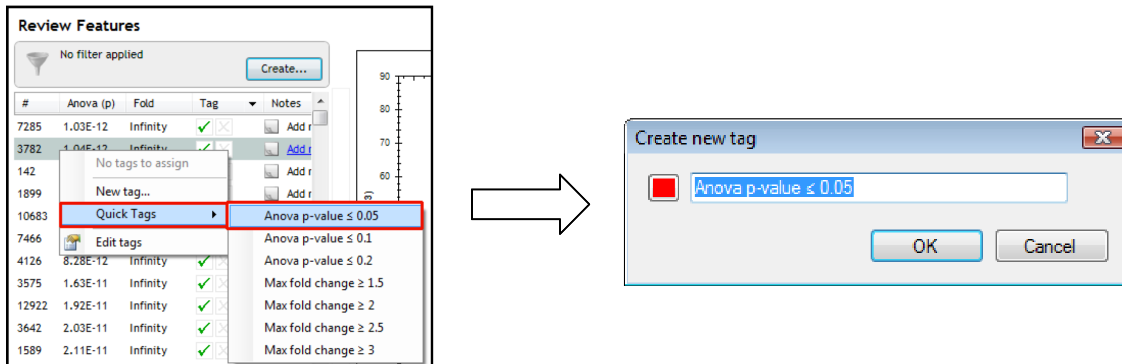
The purpose of this stage in the Workflow is to review the list of features using the visual tools provided and edit features if required.

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

Details on how to use the various views and table are described in detail in the Main analysis workflow userguide.

For the purposes of this example we require to identify all those features that demonstrate a significant Anova value (p<0.05) between the 2 conditions being studied. We will create a Tag identifying just those features.

Right click on a feature in the table and select **Quick Tags** then Anova p-value =<0.05 and accept the name

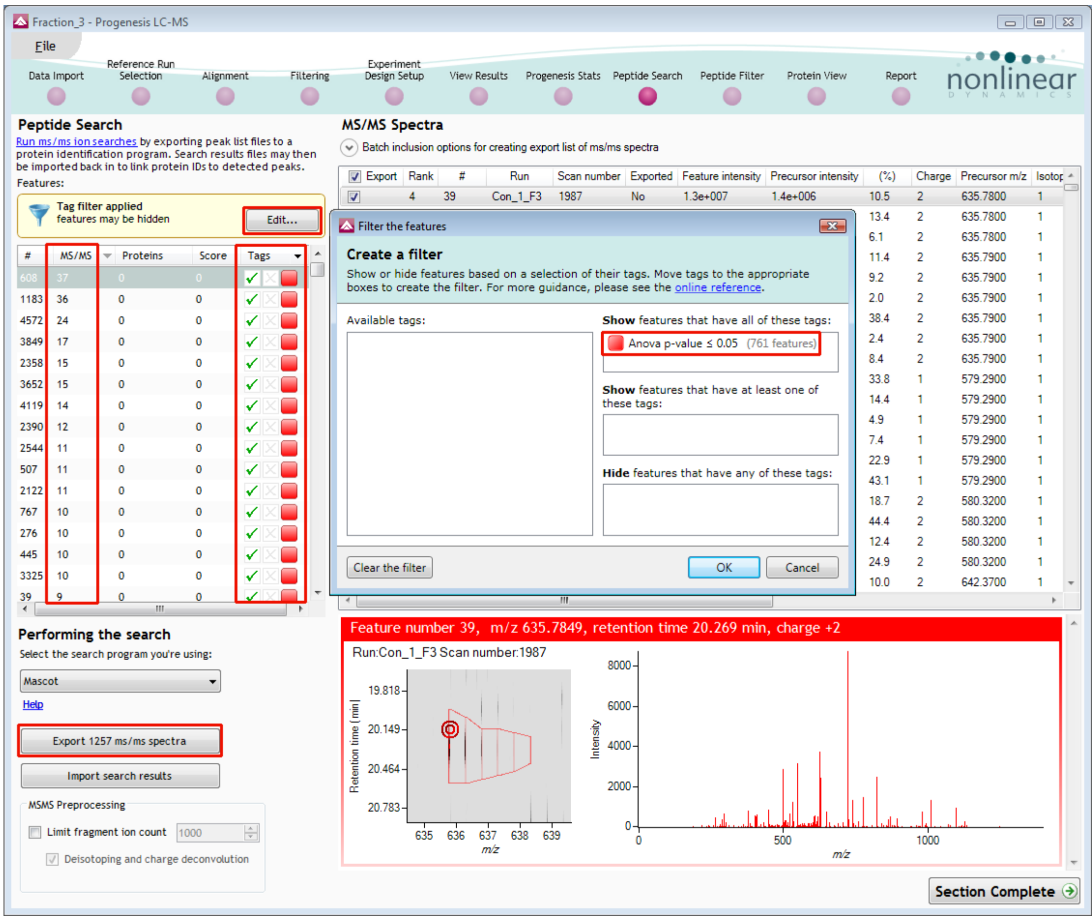


Now move to the **Peptide Search** stage in the workflow using the icon on the workflow.

Stage 7: Peptide Search

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

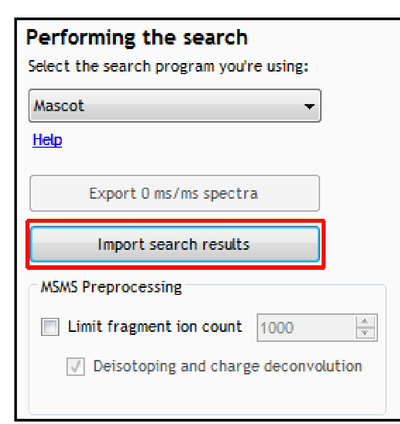
Determining protein identification is dependant on the availability of MS/MS data for the LC-MS runs. This data may be available but limited if the LC-MS was performed in a data dependant MS/MS detection mode due to under sampling. Under these conditions MS/MS data acquisition is dependant on thresholds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



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

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

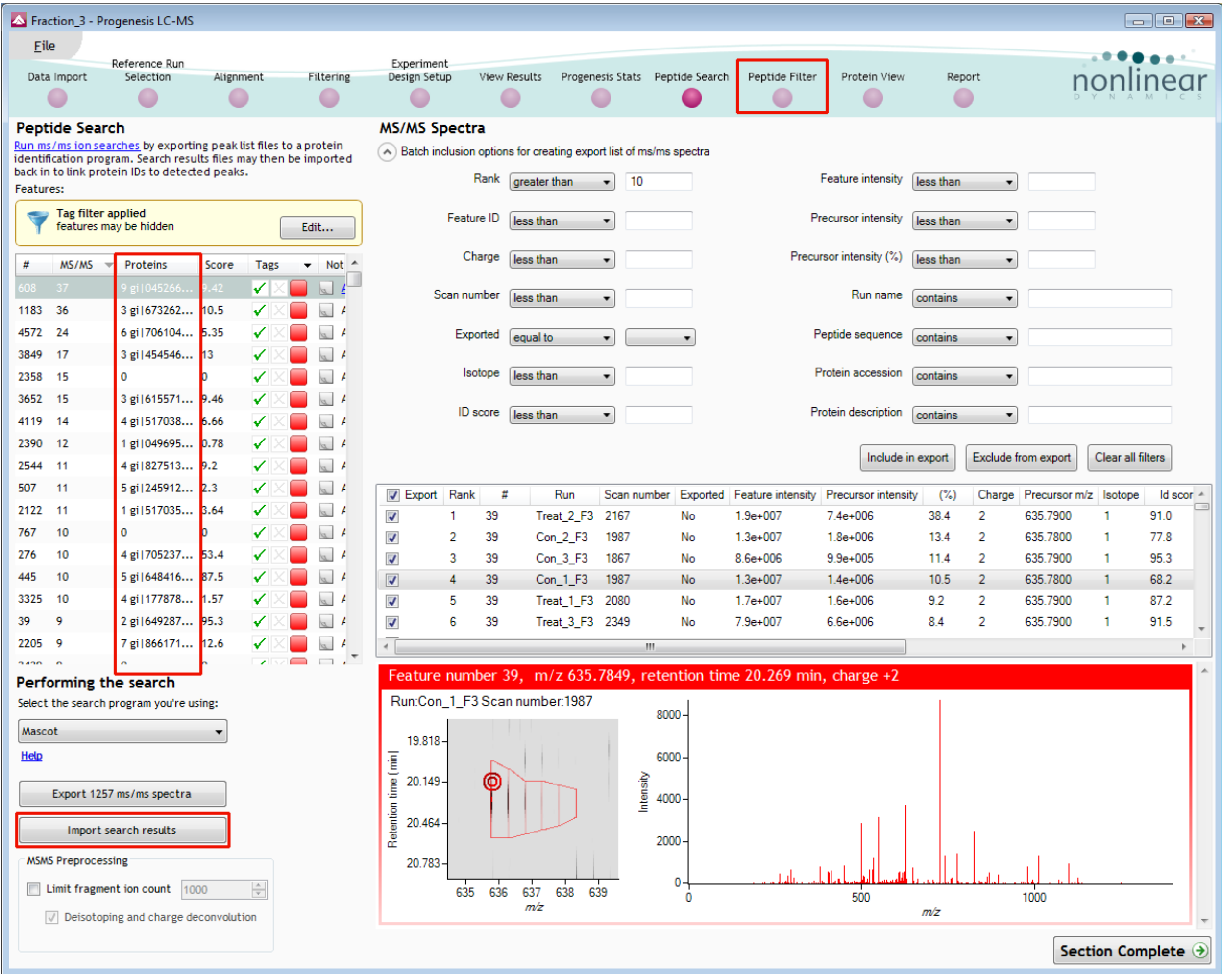
The first step is to decide which MS/MS scans you wish to export to be identified. By default this is all the available spectra for the Features displayed in the Features list. Using the tag created in the previous section you can filter the table to only those showing a significant change (Anova p=<0.05) between the conditions This number of spectra to be exported is visible on the Export button.



Performing an MS/MS Ion Search

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

On importing the Search results the Features table updates to reflect the identified proteins and the relevant score for each searched feature.



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

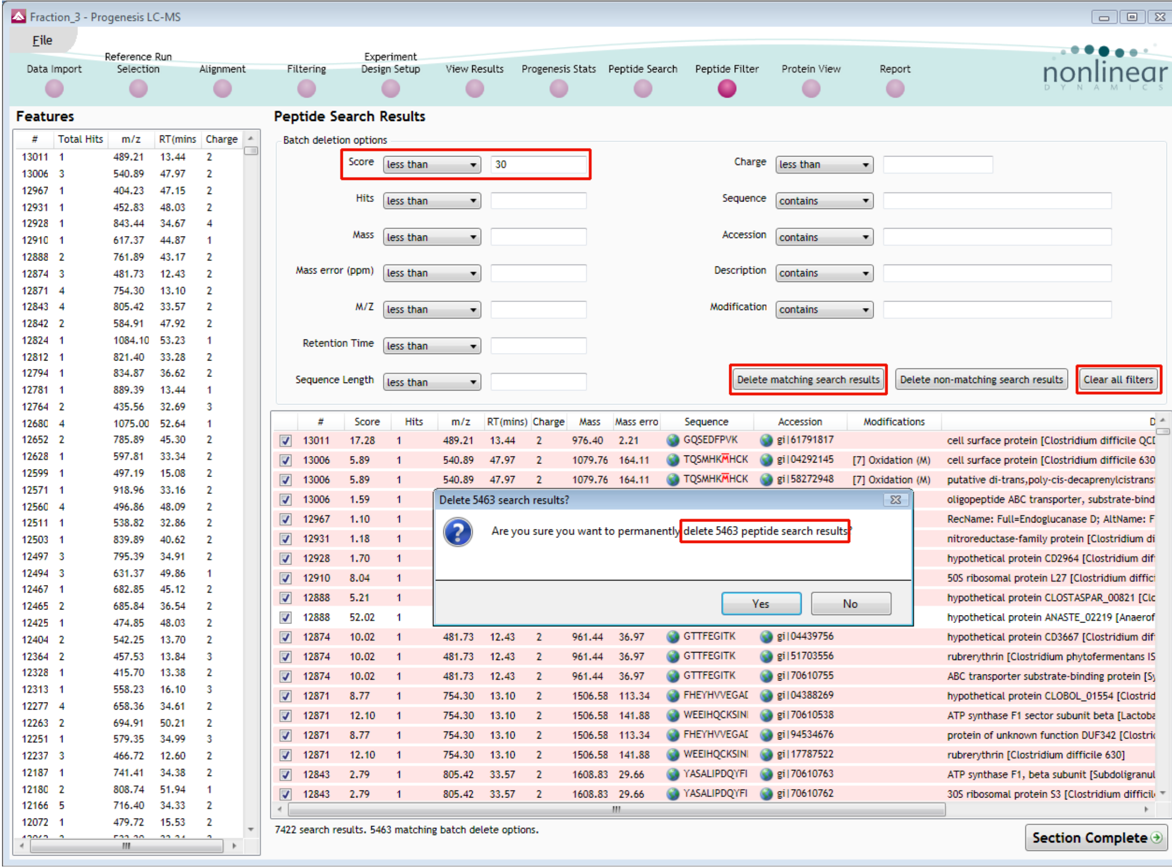
Stage 8: Peptide Filter

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

As an **example** 'Acceptance Criteria' on which to base the sequential filtering of the Peptide results, the following thresholds will be applied:

- Remove identifications with a Score less than 30
- Remove identifications where less than 2 hits were returned
- Remove all identifications where the Protein Description **Contains** 'hypothetical'
- Remove all identifications where the Protein Description **Doesn't contain** ‘Clostridium difficile’

On the Batch detection options panel, set the Score to less than 30, then **Delete matching search results**.



Note: the search results matching the filter criteria turn pink and the number of search results being deleted is displayed

Now **Clear all filters** and then apply the next filters as described in the previous page.

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

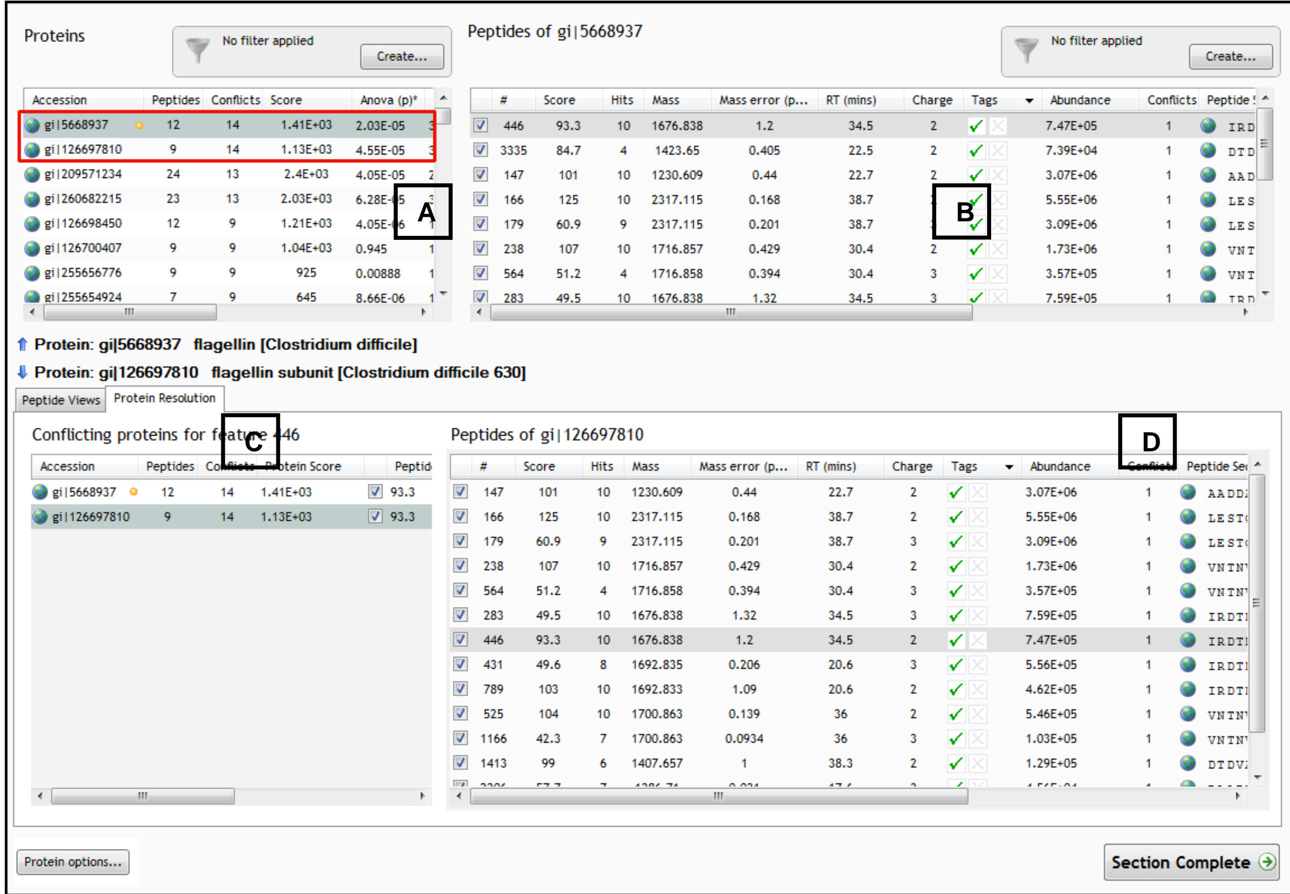
Stage 9: Protein View

The Protein View combines the quantitative LC-MS data with the qualitative MS/MS results at the protein level, highlighting proteins of interest between experimental groups. This stage allows you to examine the behaviour of the identified peptides and resolve any conflicts for the various peptide assignments at the protein level.

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

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

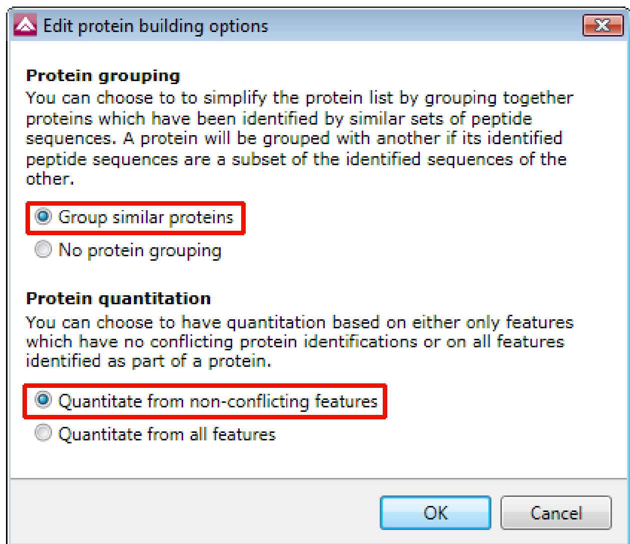
Note: the look of the tables (with regards to ordering) in the following section may vary slightly.



In the above example the conflict would be resolved in favour of the protein with 12 peptides as the flagellin subunit does not contain any unique peptides as compared to flagellin.

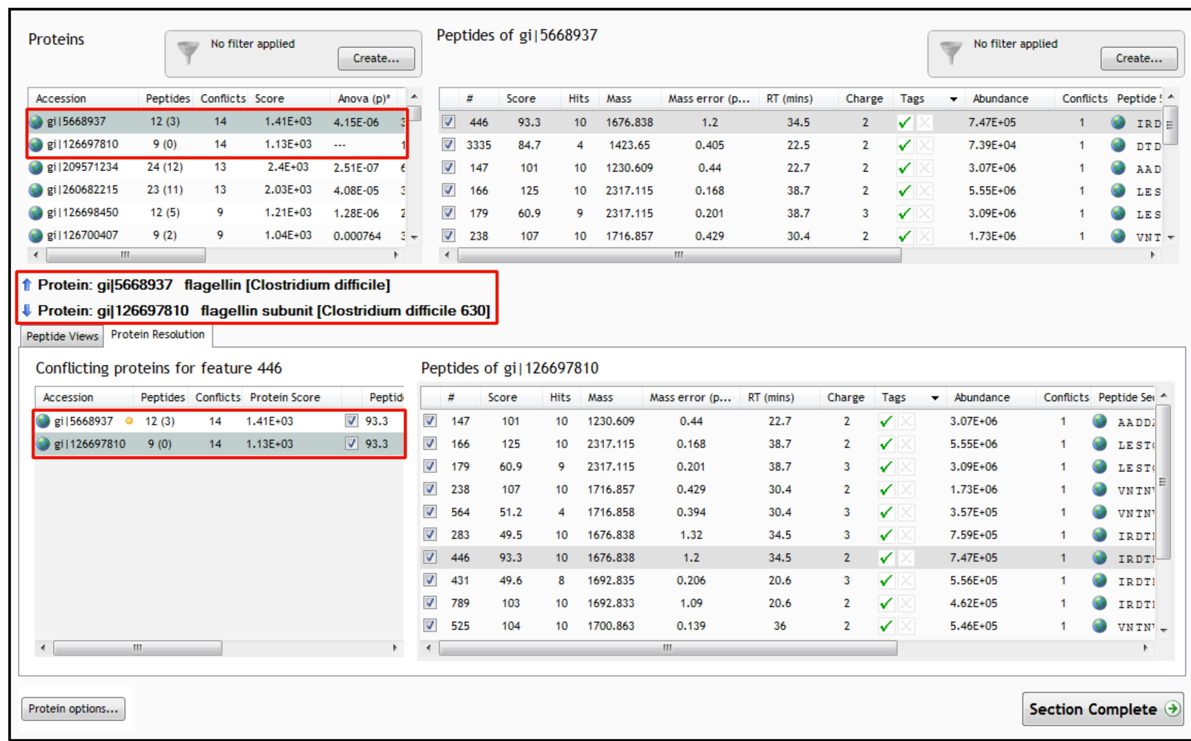
Note: the number of conflicts you have to resolve will depend on the scope and stringency of the filters you apply at the **Peptide Search** stage,

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

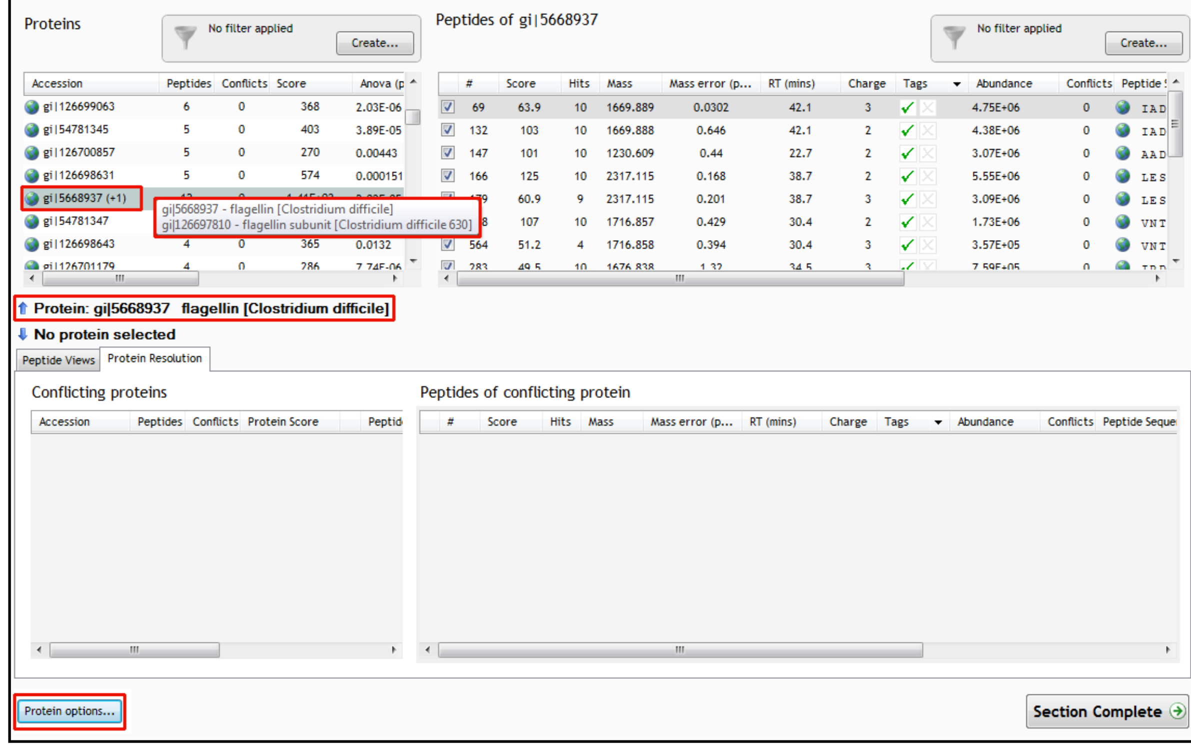


Having performed the conflict resolution with **Group similar proteins** and **Quantitate from non-conflicting features** now switch off the protein grouping.

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



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

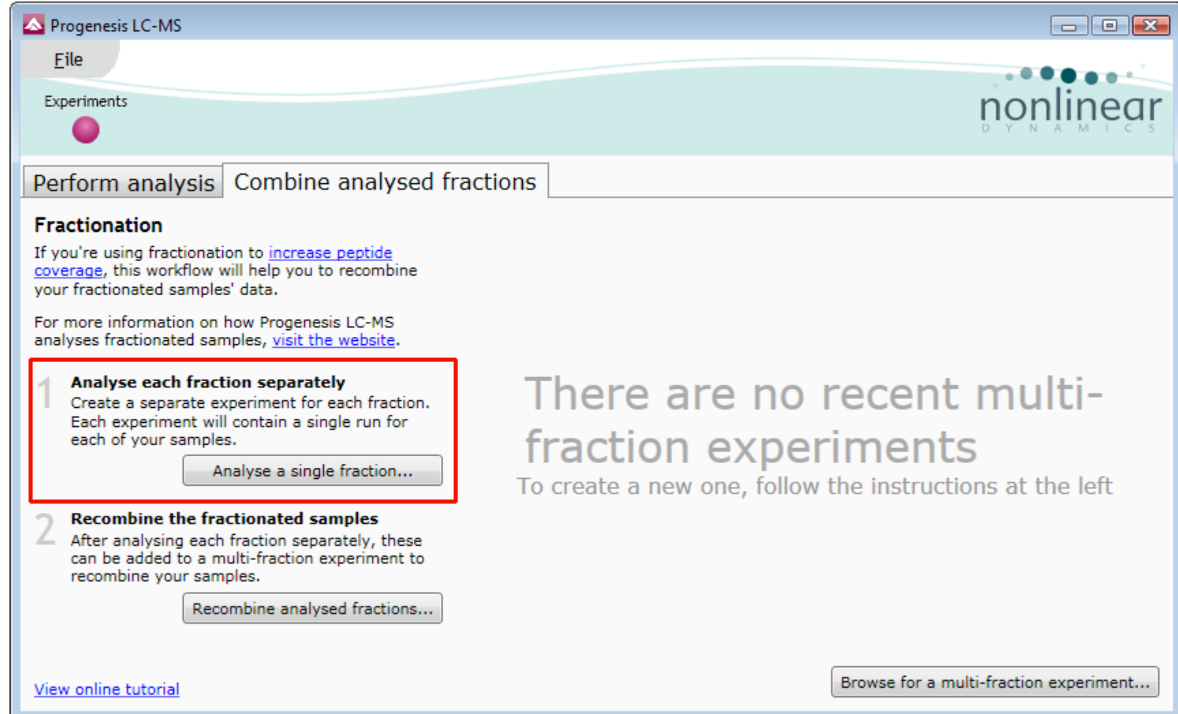


Note: the flagellin subunit has **no unique** peptides (brackets after the peptides field in the Proteins table as shown above) as they are all present in flagellin protein hence the reason for grouping. As a result all the conflicts are internal to the group.

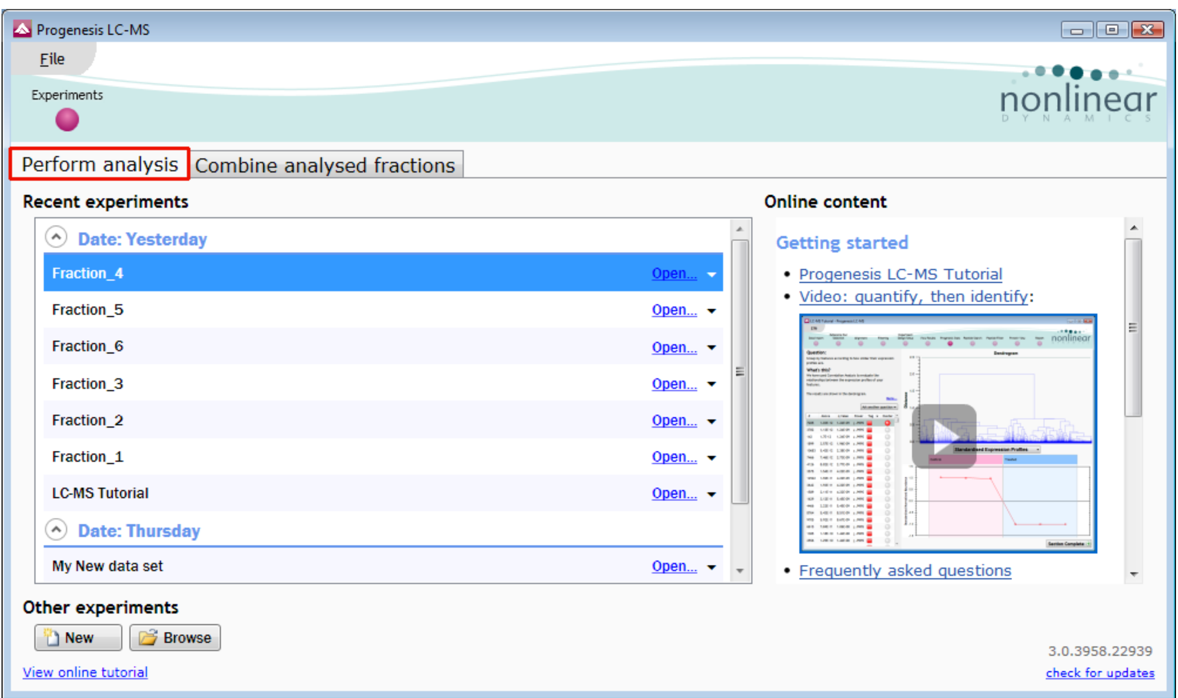
Before recombining the fractions make sure all the single fraction experiments are complete.

Completion of Fractionation Workflow Step 1

The first stage in the analysis of a Fractionated experiment is completed when you have analysed all the single-fraction experiments, in this example there are 6 fractions, therefore 6 experiments. For each experiment this includes the identification of proteins and resolution of any peptide conflicts as described in the previous sections.



The six experiments will appear in the Perform Analysis tab and can be accessed individually.

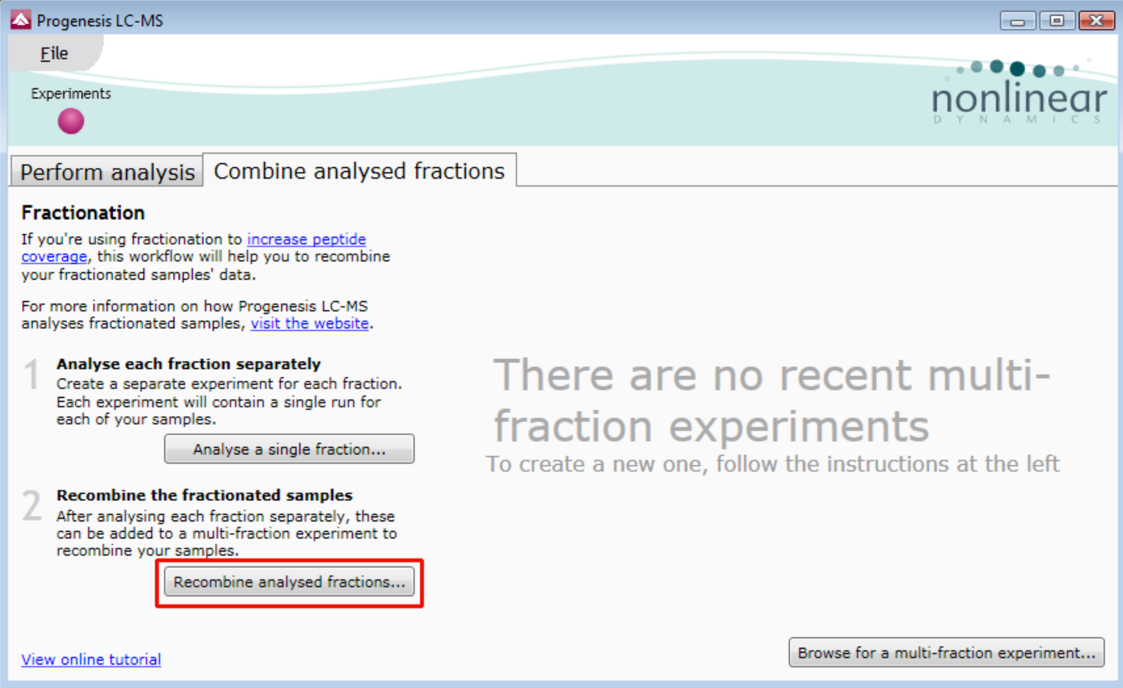


The next section describes the workflow involved in the 'Combining' of these single fraction experiments to generate a 'multi-fraction' experiment.

Fractionation Workflow Step 2

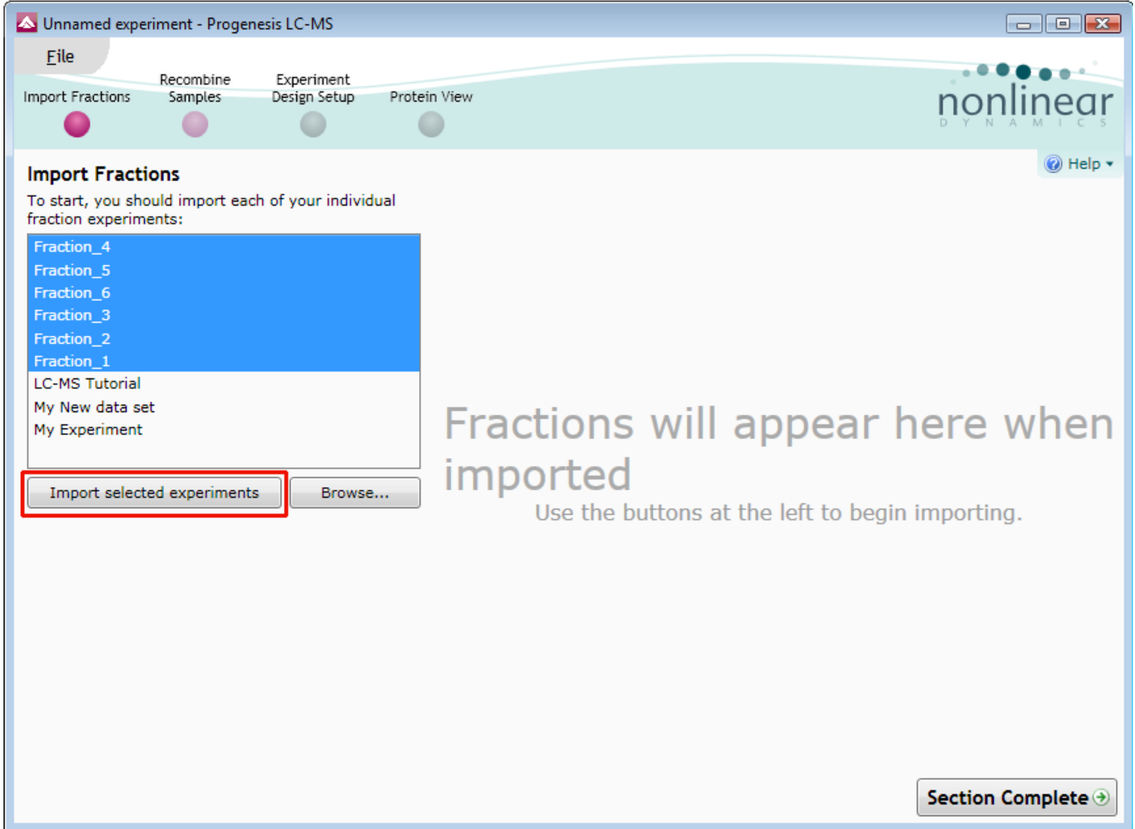
The combining of the single fraction experiments is performed in the second stage of the Fractionation workflow.

Select **Recombine analysed fractions...**



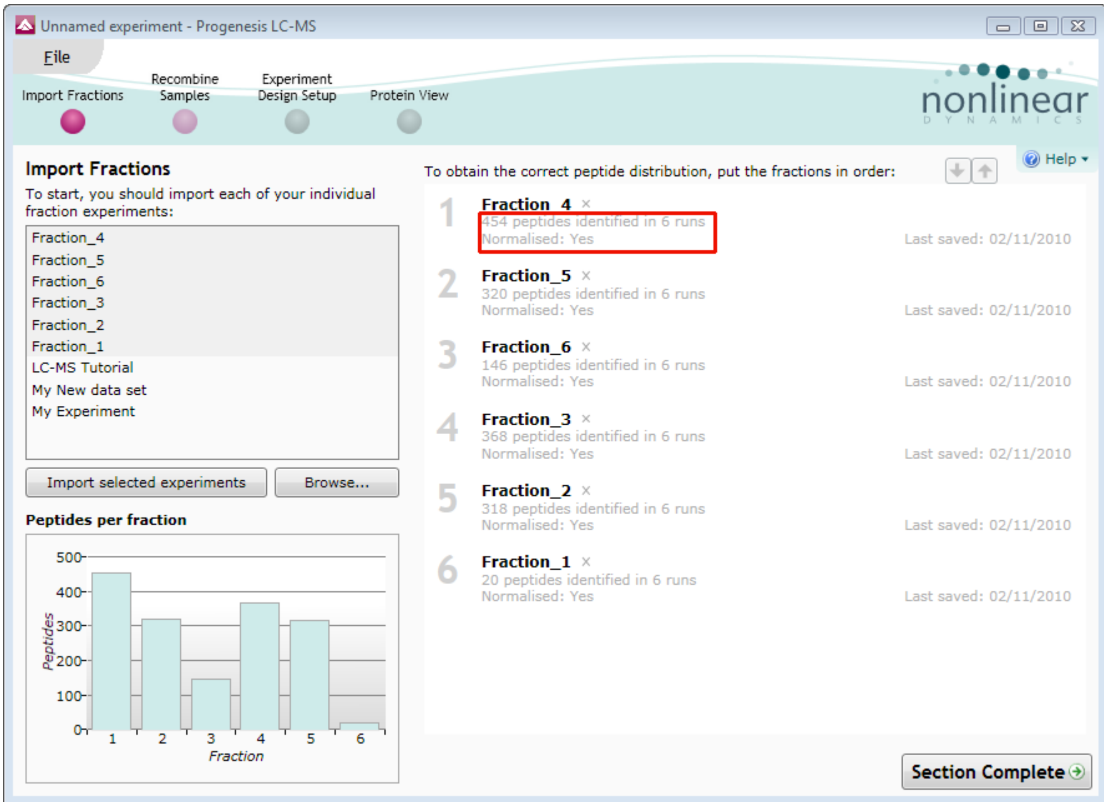
Stage 1 Import Fractions

The Import Fractions stage of the work opens, select the experiments that correspond to the (6) fractions.

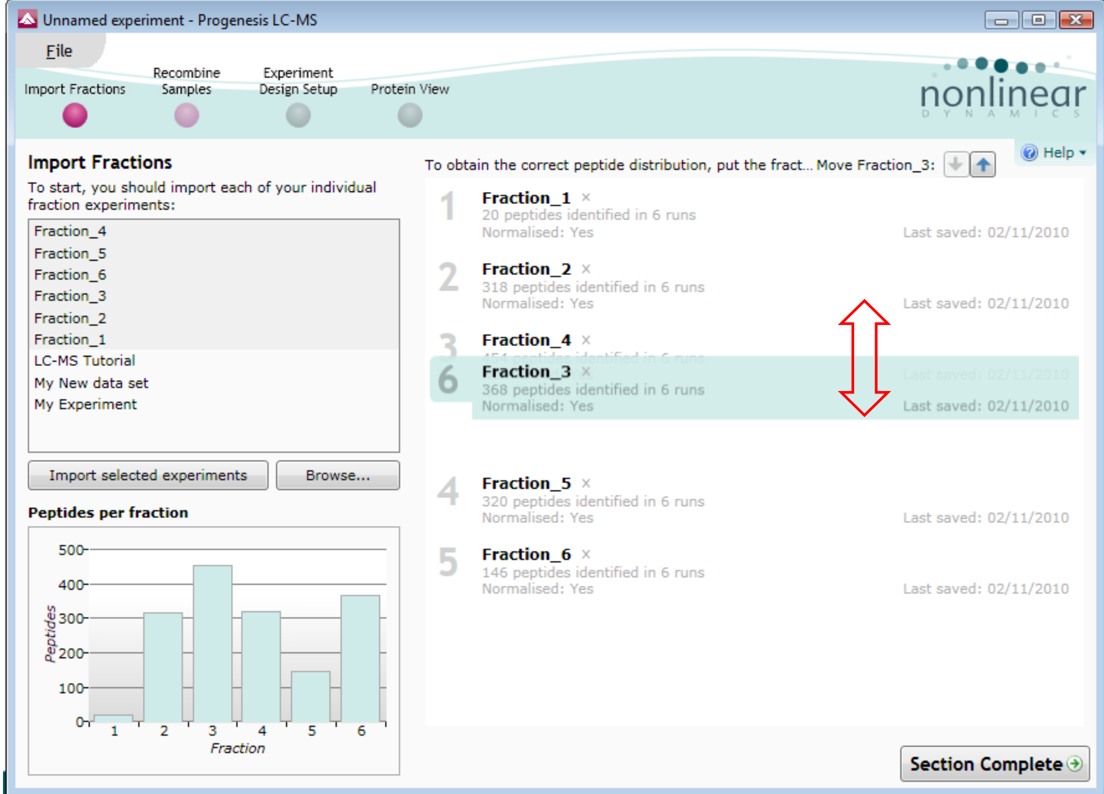


Click **Import selected experiments**

Details for each single fraction experiments appears in the panel showing numbers of identified peptides.



You can adjust the order of the single fraction experiments to reflect the order of the fractions by dragging the single fraction experiments to the correct position.



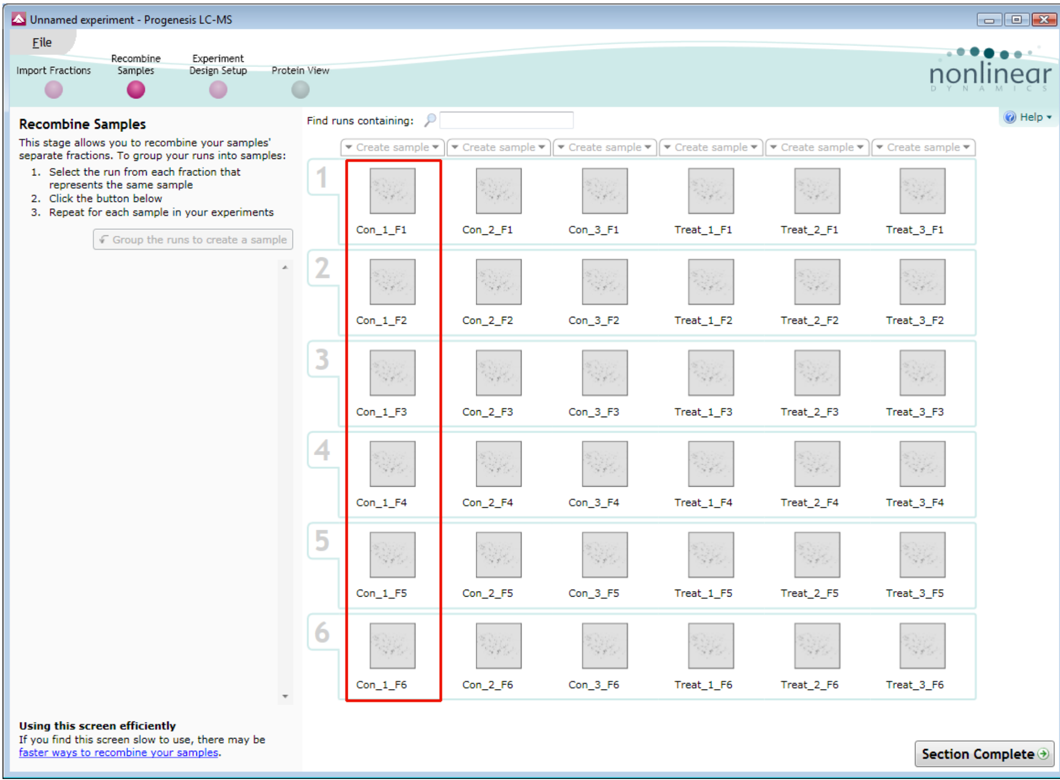
Note: the graph of 'peptides per fraction' updates to reflect the new order.

To move to the **Recombine Samples** stage click **Section Complete**.

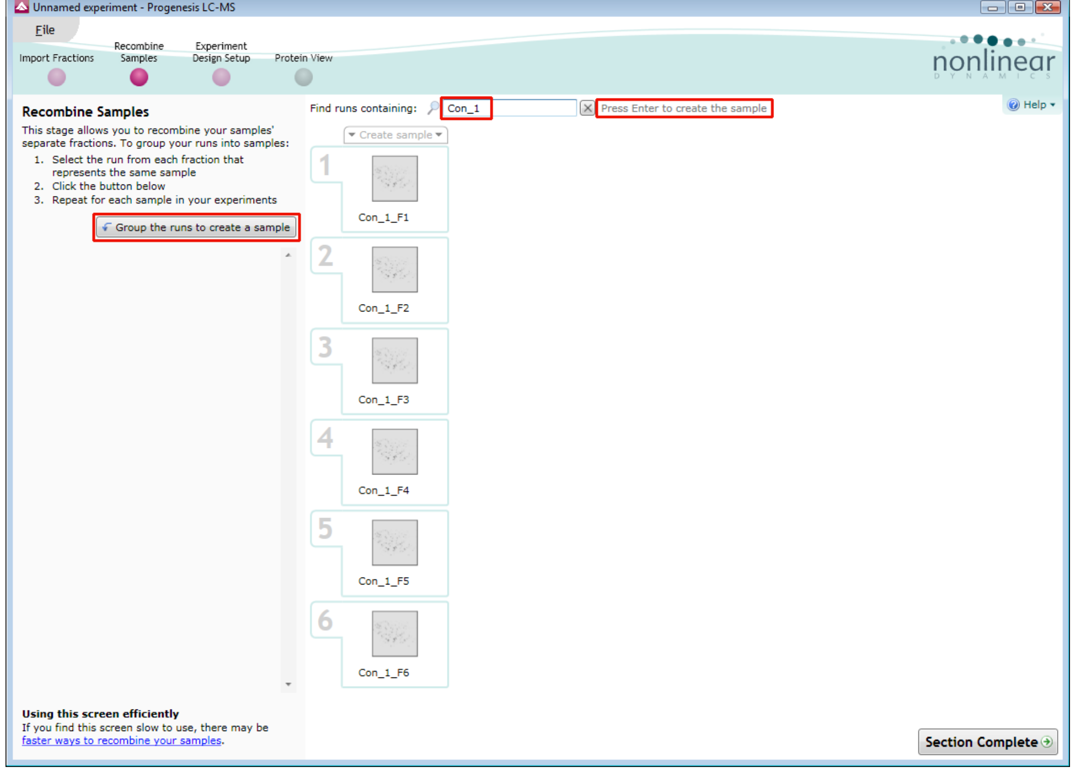
Stage 2 Recombine Samples

At this stage you will recombine the samples by selecting the runs that correspond to each sample from the single fraction experiments.

Note: how efficiently you use this page will depend on how methodically you have named the various sample runs. For this example Samples are Con_1, 2 and 3 and Treat_1, 2 and 3

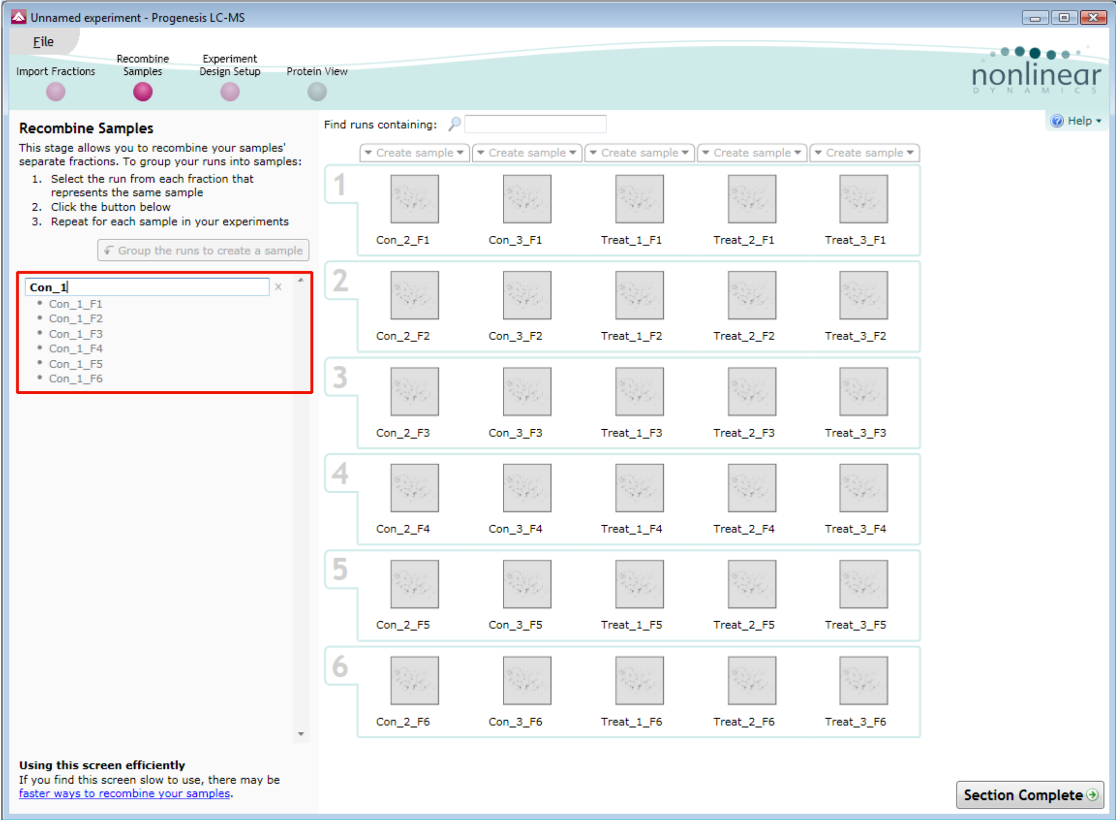


Typing **Con_1** in the **Find runs containing** search box will locate the runs corresponding to sample Con_1.

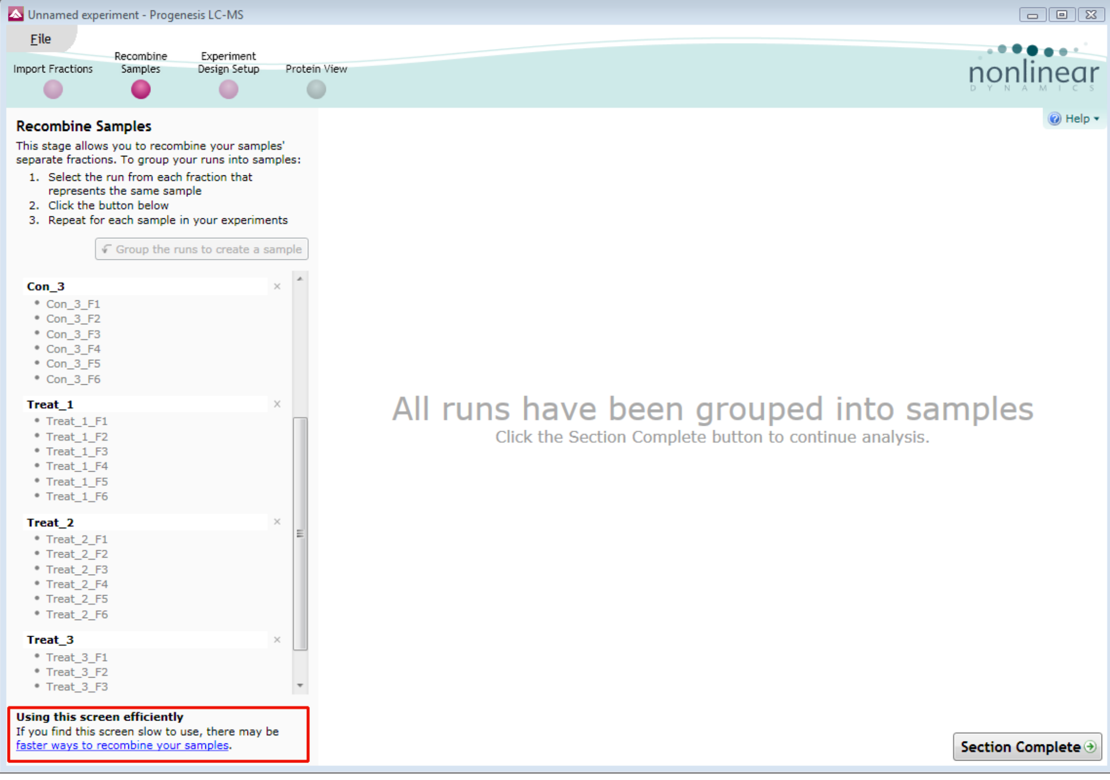


Click '**Enter**' or **Group the runs to create a sample** in the left hand panel

Note: using Enter will set the sample name as **Con_1**, overtype to rename as required.



Repeat for the remaining samples.



Note: as mentioned before other ways of Recombining the samples can be applied, depending on the naming conventions used ; use the link, bottom left, to see the alternatives.

Having completed the recombination of the samples click **Section Complete**.

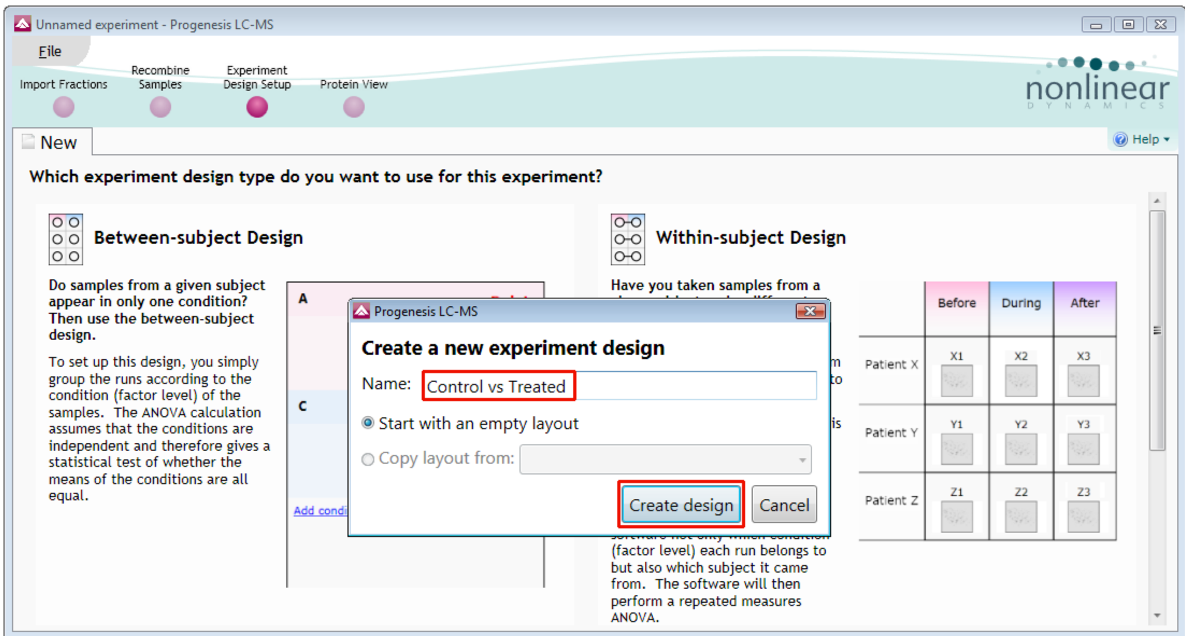
Note: at this point the data is re-normalised to account for the fractionation of the samples.

Stage 3 Experiment Design Setup

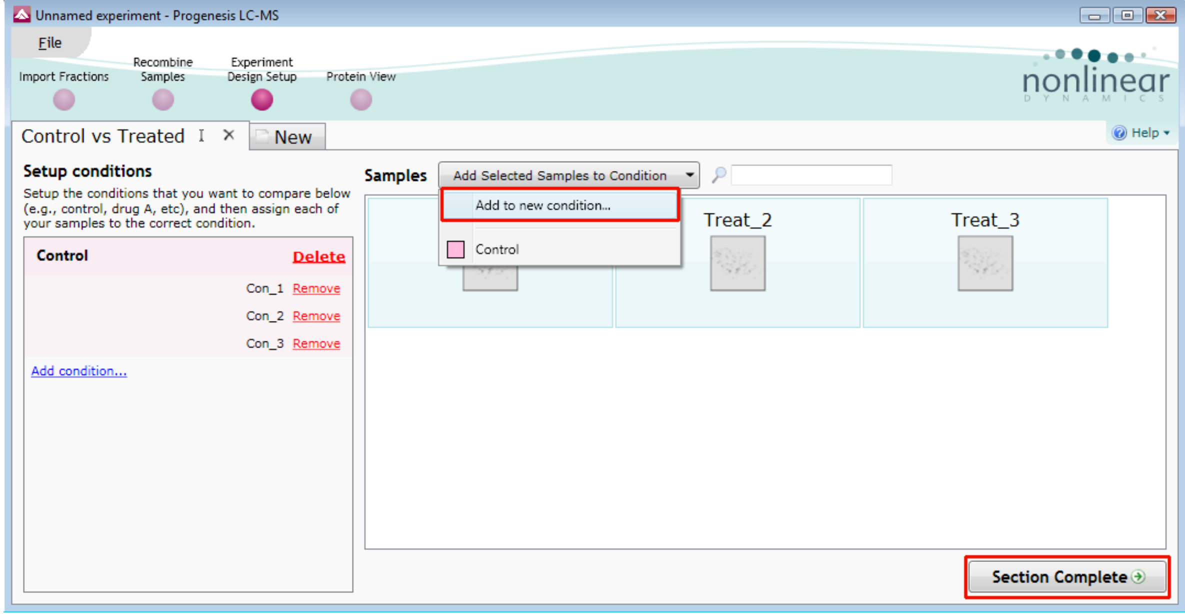
Having recombined your samples you can now define the experimental designs most appropriate for your original experiment.

This experiment contains 2 conditions: Control and Treated and uses the **Between-subject design** to group the analysed samples to reflect the Biological conditions in the original study.

Select **Between-subject** and give design an appropriate name.



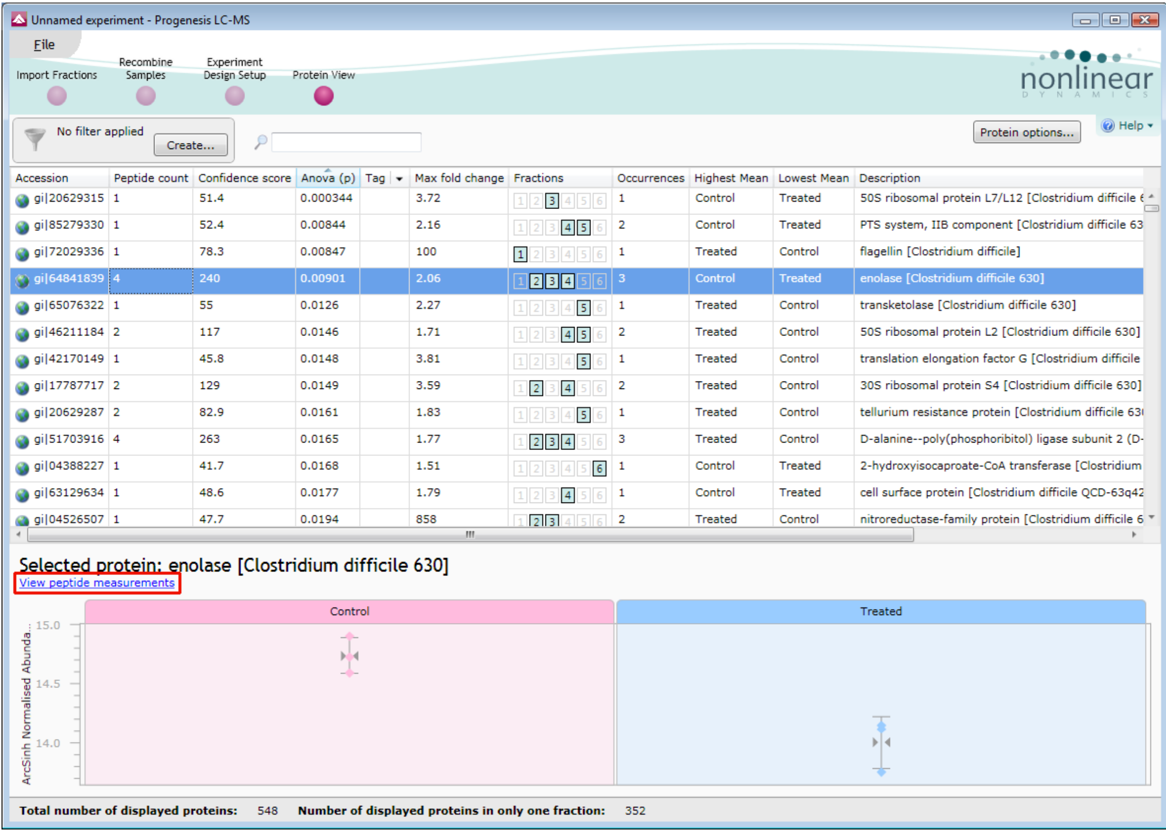
Highlight the samples, to add them on to a new condition click on **Add Selected Samples to Condition**



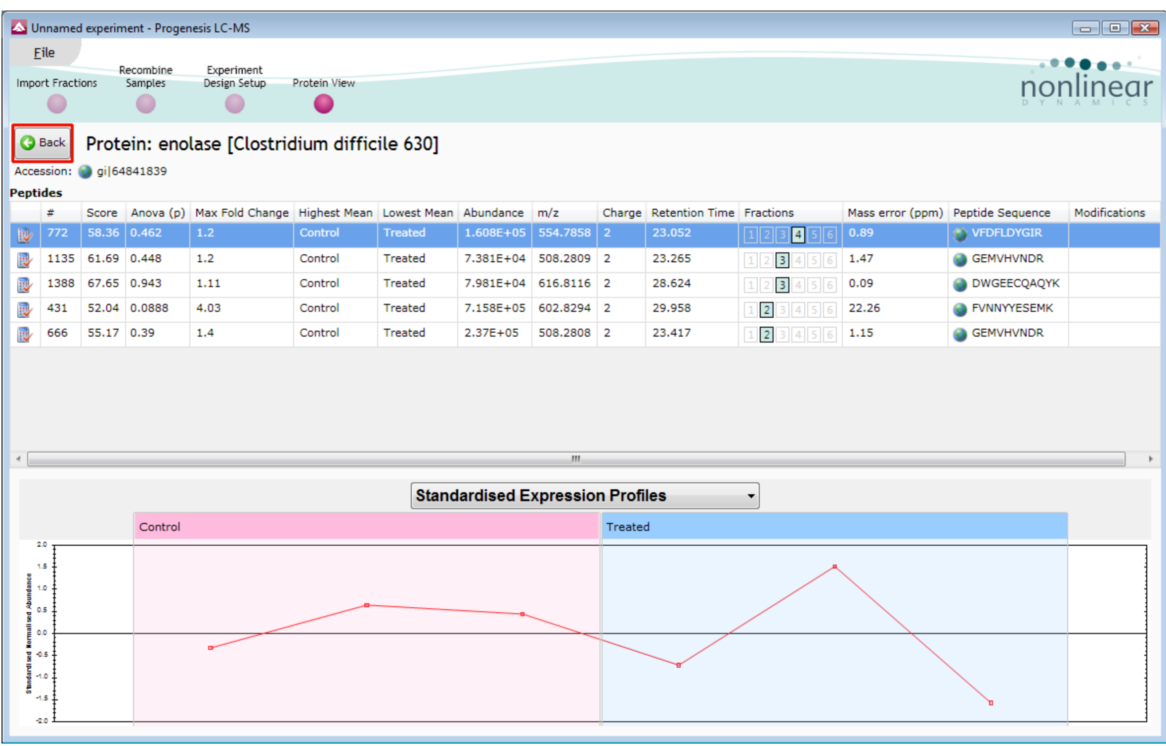
Click **Section Complete** to move to Protein View

Stage 4 Protein View

The recombined data can now be viewed at the level of the Proteins.



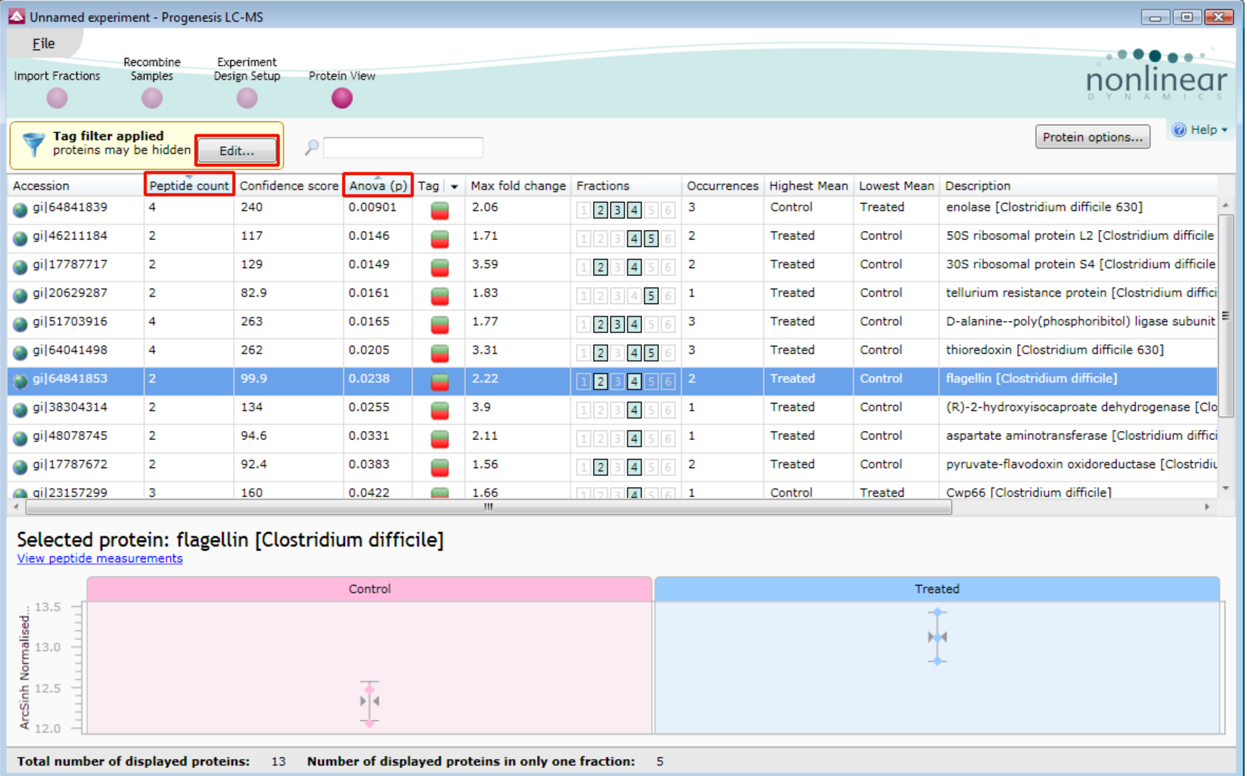
And at the peptide level when you click **view peptide measurements**



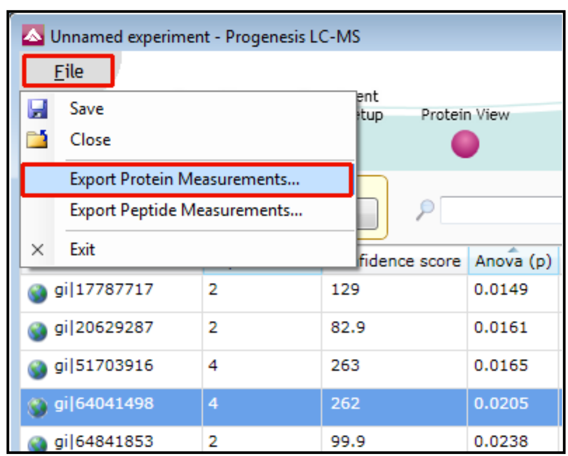
Click back to return to the **Protein View**

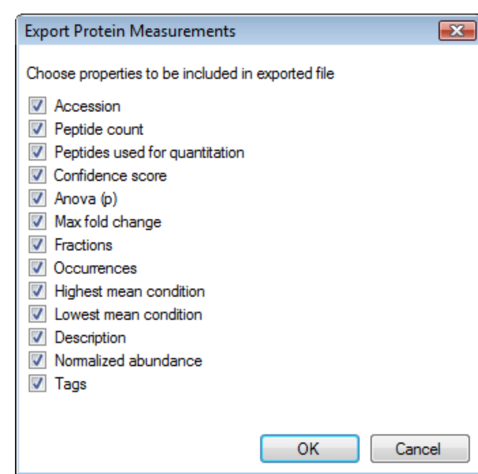
Using the combination of ordering of the proteins properties and Protein Tags you can generate a list of proteins based on similar properties and thresholds.

i.e. the list can be reduced to show only those with a p value less than 0.05 and containing 2 or more peptides



Using the 'Tag' filters the list is reduced to the relevant proteins. Details of these proteins can be exported by selecting **Export Protein Measurements** from the **File** menu.



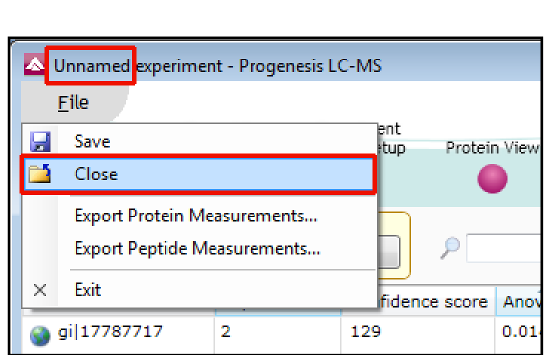


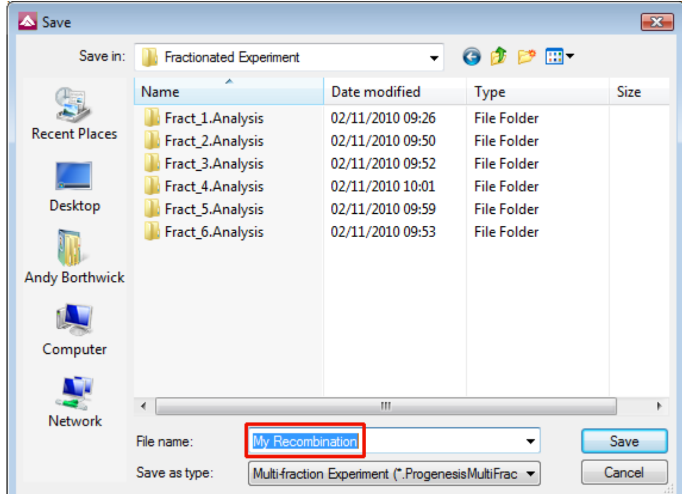
A dialog allows you to control the details of the output file.

Finally your recombined data can be saved as a **Multi-fraction experiment**.

Stage 5 Saving a Multi-Fraction experiment

When you opened the Fractionation workflow and started working with the recombination of your 'Single Fraction Experiments' the workflow recognises the current experiment as '**Unnamed**' this status will change as you close and/or save the experiment.





On saving the new mult-fraction experiment appears on the **Combine analysed fractions** page

Details of the new experiments location (folder), number of fractions, samples and total identified proteins.

Note: you can reopen the experiment by either double clicking on it or using **open.**

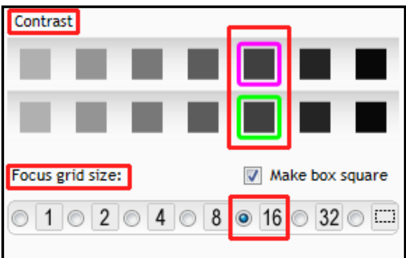
This completes a guided tour through using Progenesis LC-MS to analyse fractionated data.

Appendix 1: Manual assistance of Alignment

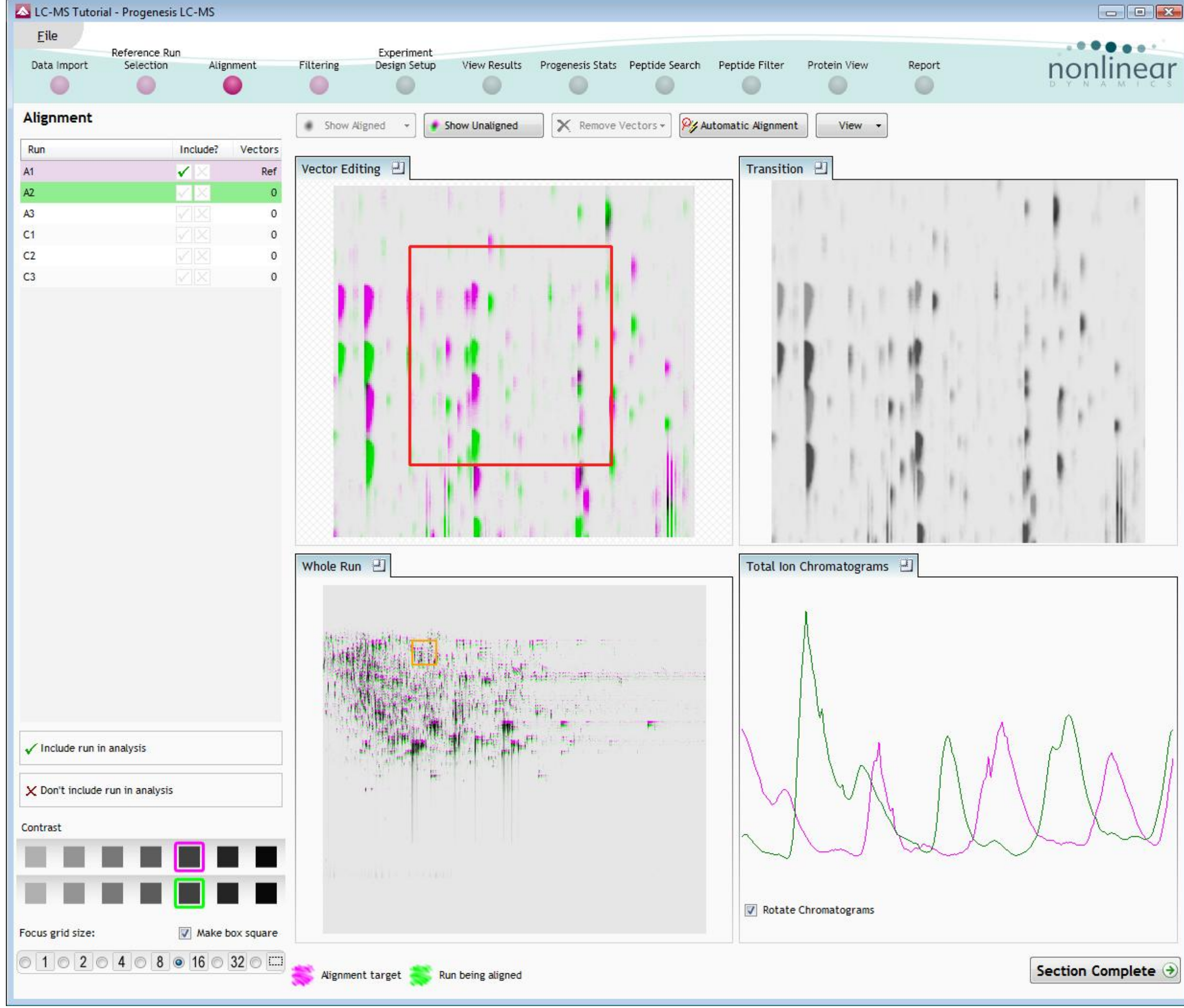
Approach to alignment

To place manual alignment vectors on a run (A2 in this example):

1. Click on Run A2 in the **Runs** panel, this will be highlighted in green and the reference run (A1) will be highlighted in magenta.
2. You will need approximately 5 - 10 **alignment vectors** evenly distributed from top to bottom of the whole run.
3. First ensure that the size of the focus area is set to **8 or 16** in the Focus grid size on the bottom left of the screen.



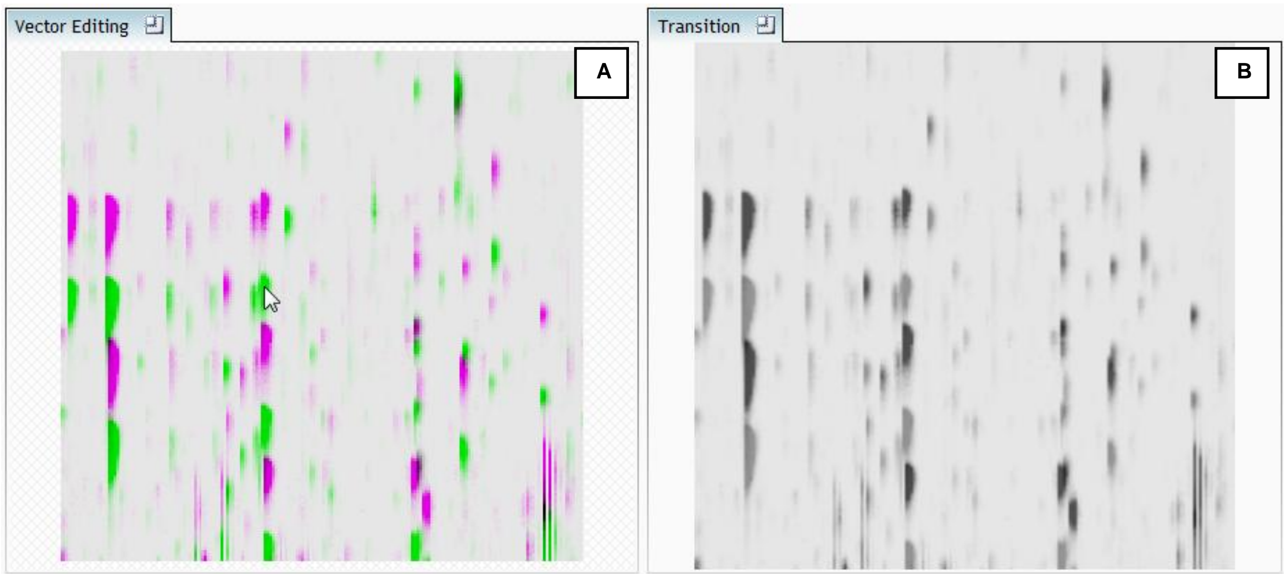
Click on an area (see below) in the **Whole Run** window (C) to refocus all the windows. Adjust Contrast as required



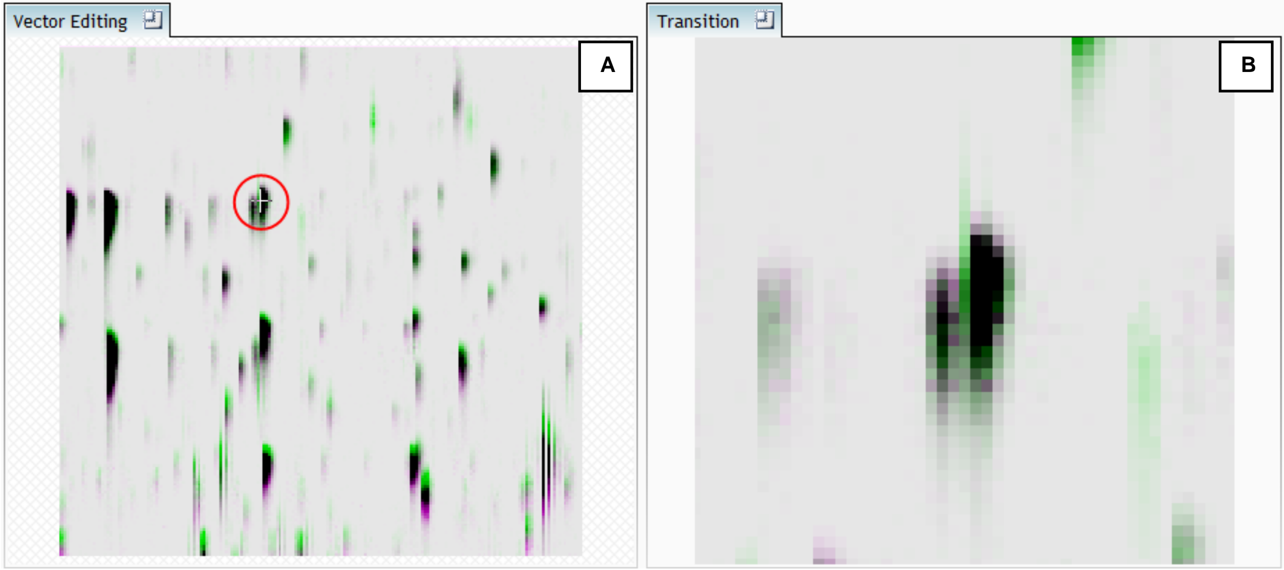
Note: the features moving back and forwards between the 2 runs in the **Transition** view indicating the misalignment of the two LC-MS runs

Note: The **Total Ion Chromatogram** view also reflects the misalignment of the 2 runs for the current Retention Time range (vertical dimension of the current Focus grid in the **Whole Run** view.

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

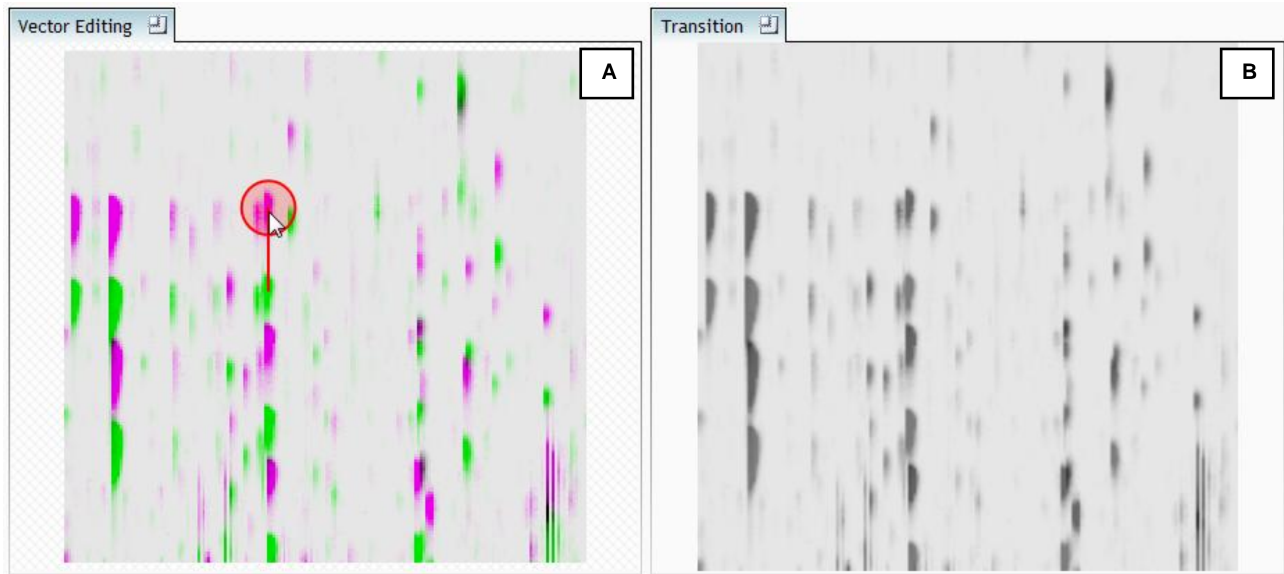


5. As you are holding down the left mouse button drag the green feature over the corresponding magenta feature of the reference run. The red circle will appear as shown below indicating that a positional lock has been found for the overlapping features.



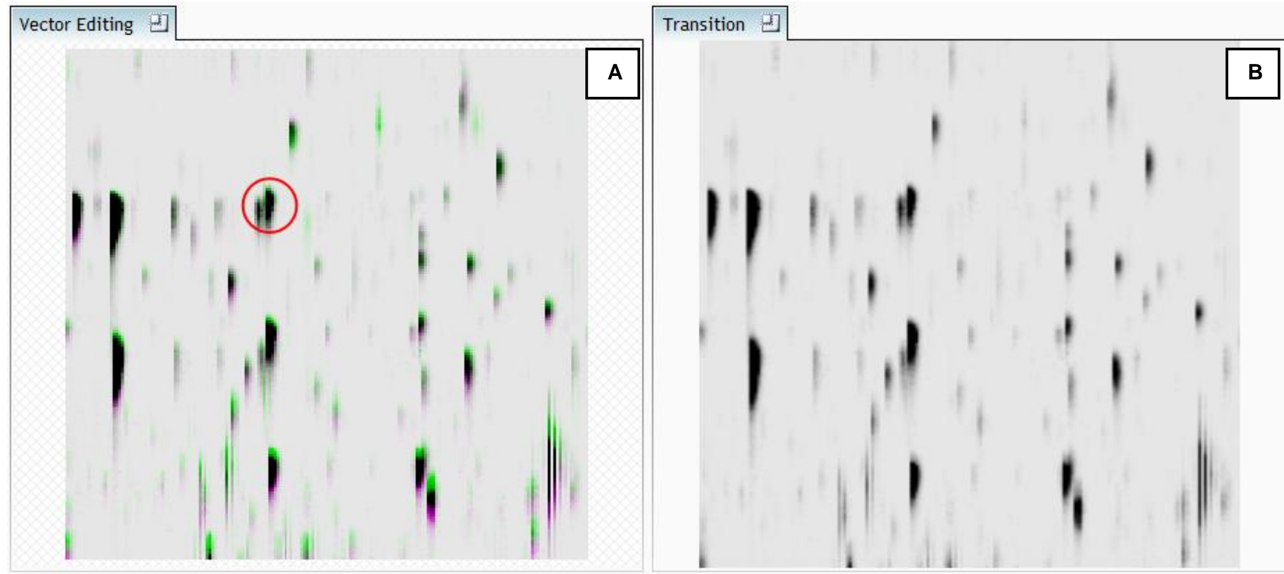
Note: as you hold down the mouse button, window B zooms in to help with the alignment.

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

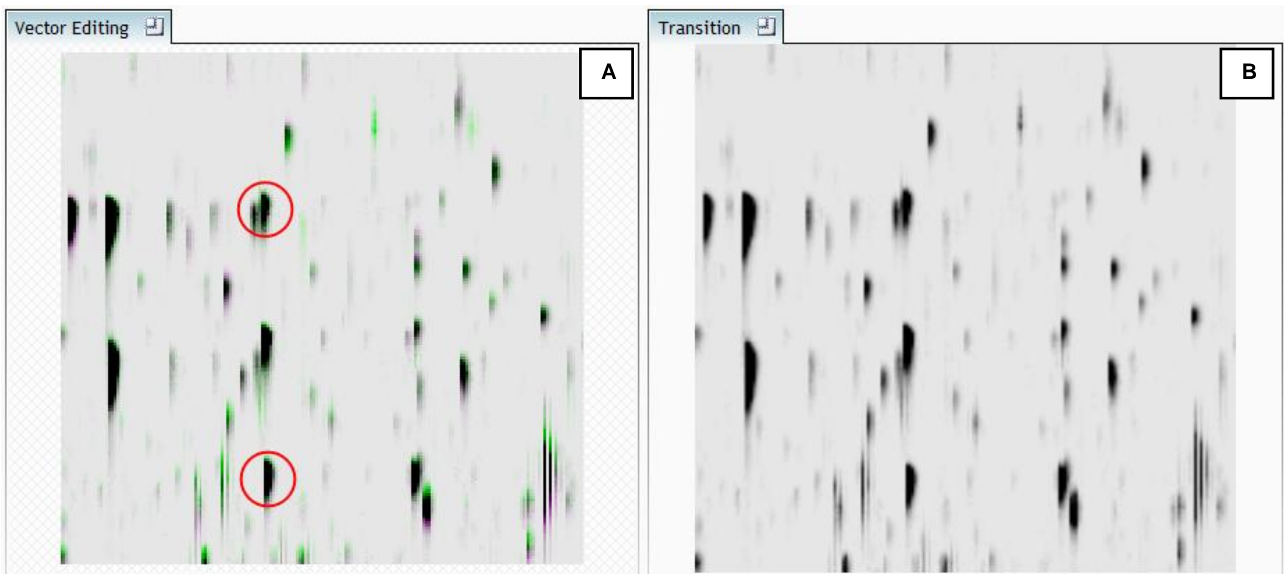


Note: an incorrectly placed vector is removed by right clicking on it in the **Vector Editing** window

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



8. Additing an additional vector will improve the alignment further. **Note** this time as you click to add the vector it 'jumps' automatically to the correct position using the information from the existing alignment vector.



Repeat this process moving the focus from top to bottom on the **Whole Run** view

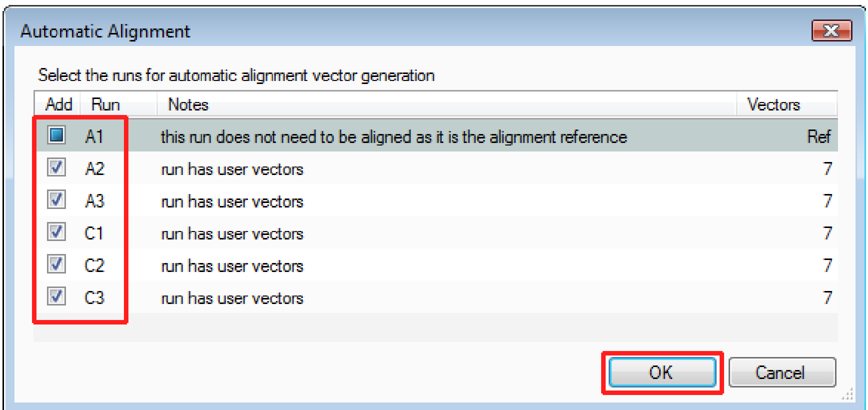
Note: the number of vectors you add is recorded in the **Runs** table

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

The number of manual vectors that you add at this stage is dependant on the misalignment between the current run and the Reference run. In many cases only using the Automatic vector wizard will achieve the alignment.

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

11. Then select Automatic Alignment to bring up the Automatic Alignment dialog and click **OK**. The automatic alignment process will begin, using the manual vectors you have added to aid in automatic vector placement.



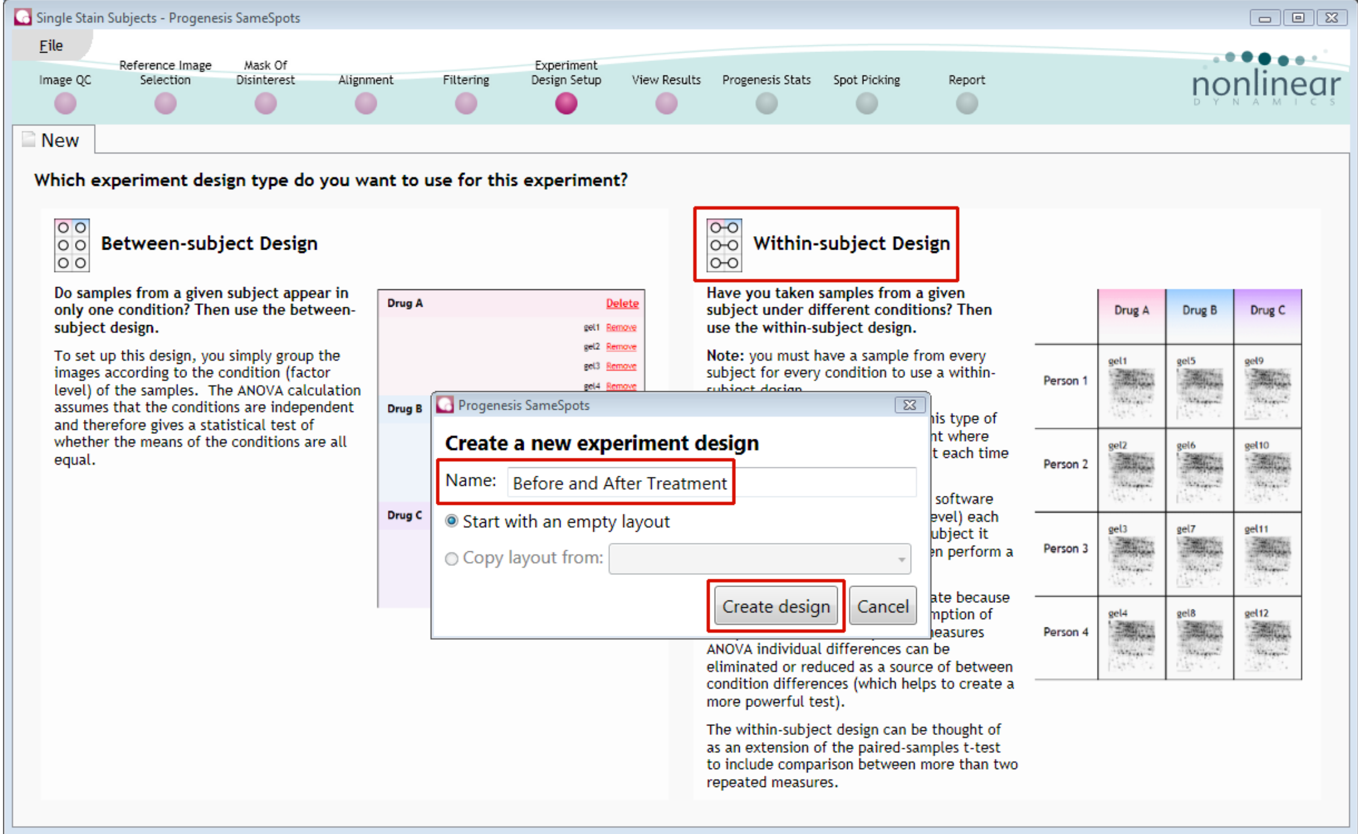
Note: the tick boxes next to the 'Run name control' which control whether vectors will be generated for each run.

To review the vectors, automatic and manual return to page 9

Appendix 2: Within-subject Design

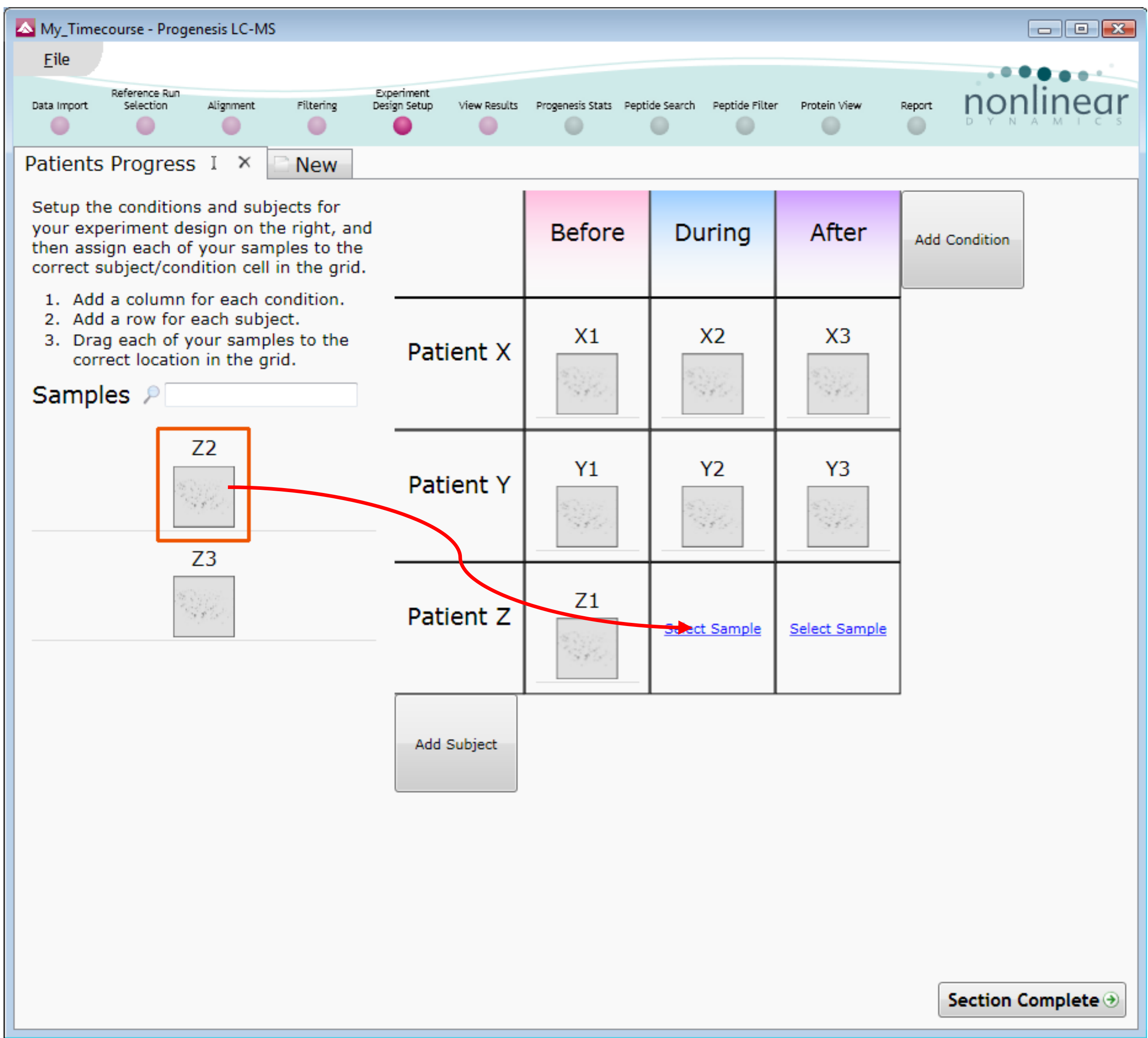
To create a **Within-subject Design** for your data set select this option on the **Experiment Design Setup** page and enter the name of the design.

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



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

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



You can create additional Experimental Designs using the New tab

All of these Experimental Designs are available at all the following stages in the LC-MS workflow