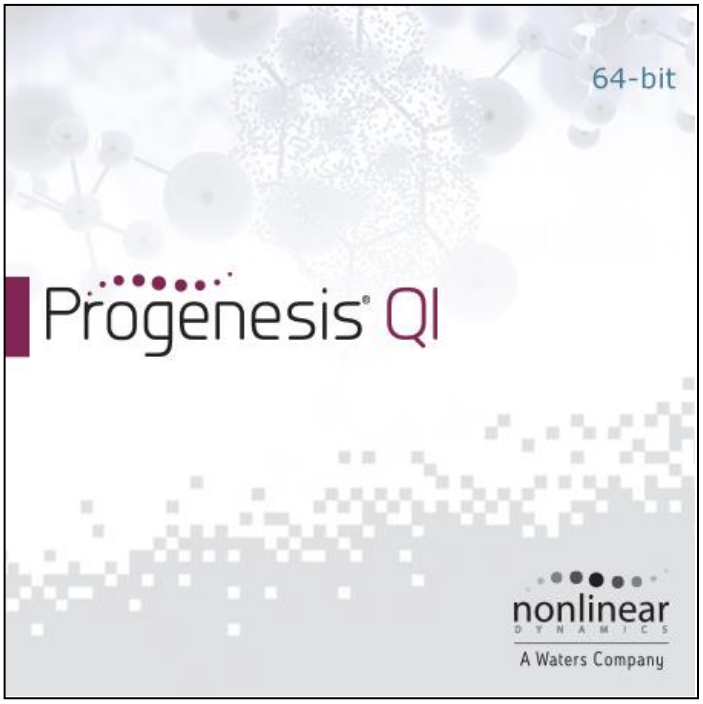


Progenesis QI
User Guide

Analysis workflow guidelines

Waters

THE SCIENCE OF WHAT'S POSSIBLE.™

Contents

Introduction

This user guide takes you through a complete analysis of 22 LC-MS runs of metabolite containing samples for 3 biological groups (6 replicate runs per group) and a set of 4 QC runs using the unique Progenesis QI workflow. It starts with LC-MS data file loading then Alignment, followed by Analysis that creates a list of interesting compounds (molecules) which are identified with compound search and then explored within Compound Stats using multivariate statistical methods.

To allow ease of use the tutorial is designed to start with the restoration of an Archived experiment where the data files have already been loaded. However, the document covers all the stages in the Progenesis QI workflow, therefore if you are using your own data files please refer to Appendix 1 (page 85) then start at page 8.

How to use this document

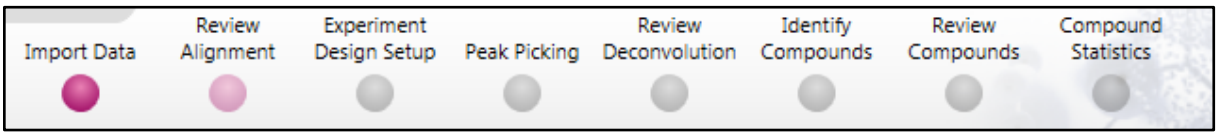
You can print this user guide to help you work hands-on with the software. The complete user guide takes about 90 minutes and is divided into two sections. This means you can perform the first half focused on run alignment and analysis then complete the second half of analysis exploring comparative differences and compound identity at a convenient time. The table of contents allows you to focus on particular areas of the workflow.

Data used in this user guide

We would like to thank Dr Giorgis Isaac at Waters Corporation, Milford USA for providing the example data which has been adapted for this user guide as well as invaluable discussion on the handling of the data.

Workflow approach to data analysis

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

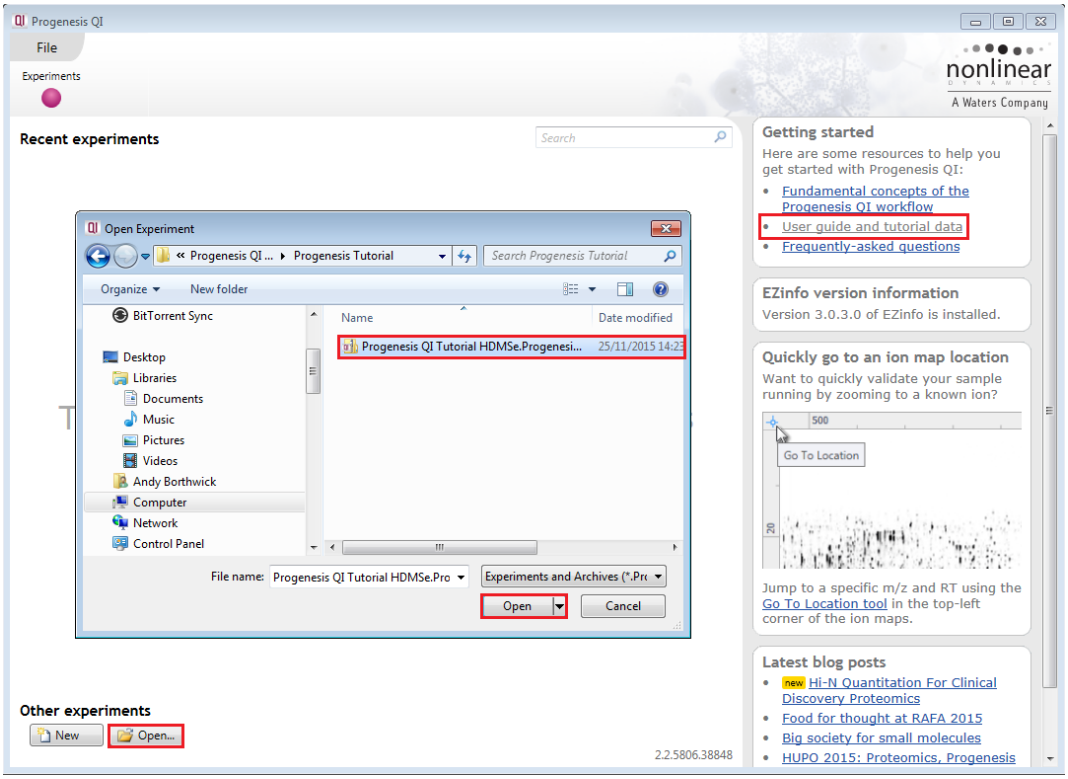


Restoring or starting a new Metabolomics Tutorial

If working with your own data files then refer to Appendix 1 page 85.

To use the Tutorial data:

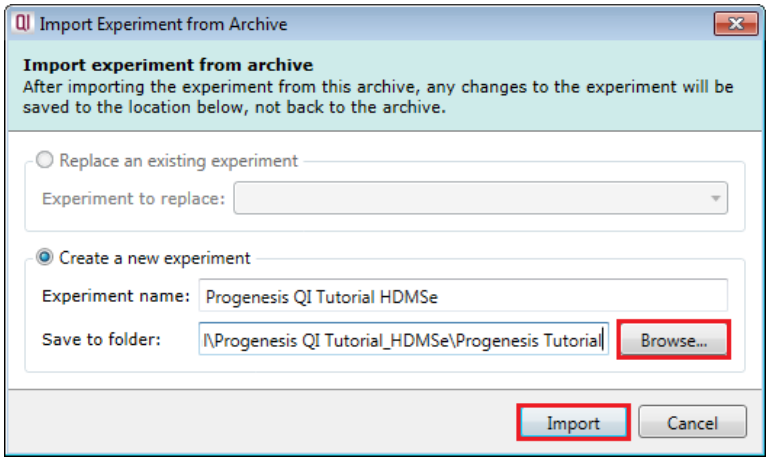
Open Progenesis QI and download the DataSet for the LC-MS user guide (.zip) from the **'User guide and tutorial data**' link shown below, placing it in a **new folder** on your desktop. Before restoring the tutorial in the software **you must** first right click on the downloaded (.zip) file and extract it to a folder, i.e. 'C:\Users\john_smith\Documents\Progenesis QI Tutorial



Now you can restore the uncompressed Progenesis QI tutorial archive file. To do this, first locate the Tutorial Archive file using the **Open** button.

This opens the 'Import from archive' dialog.

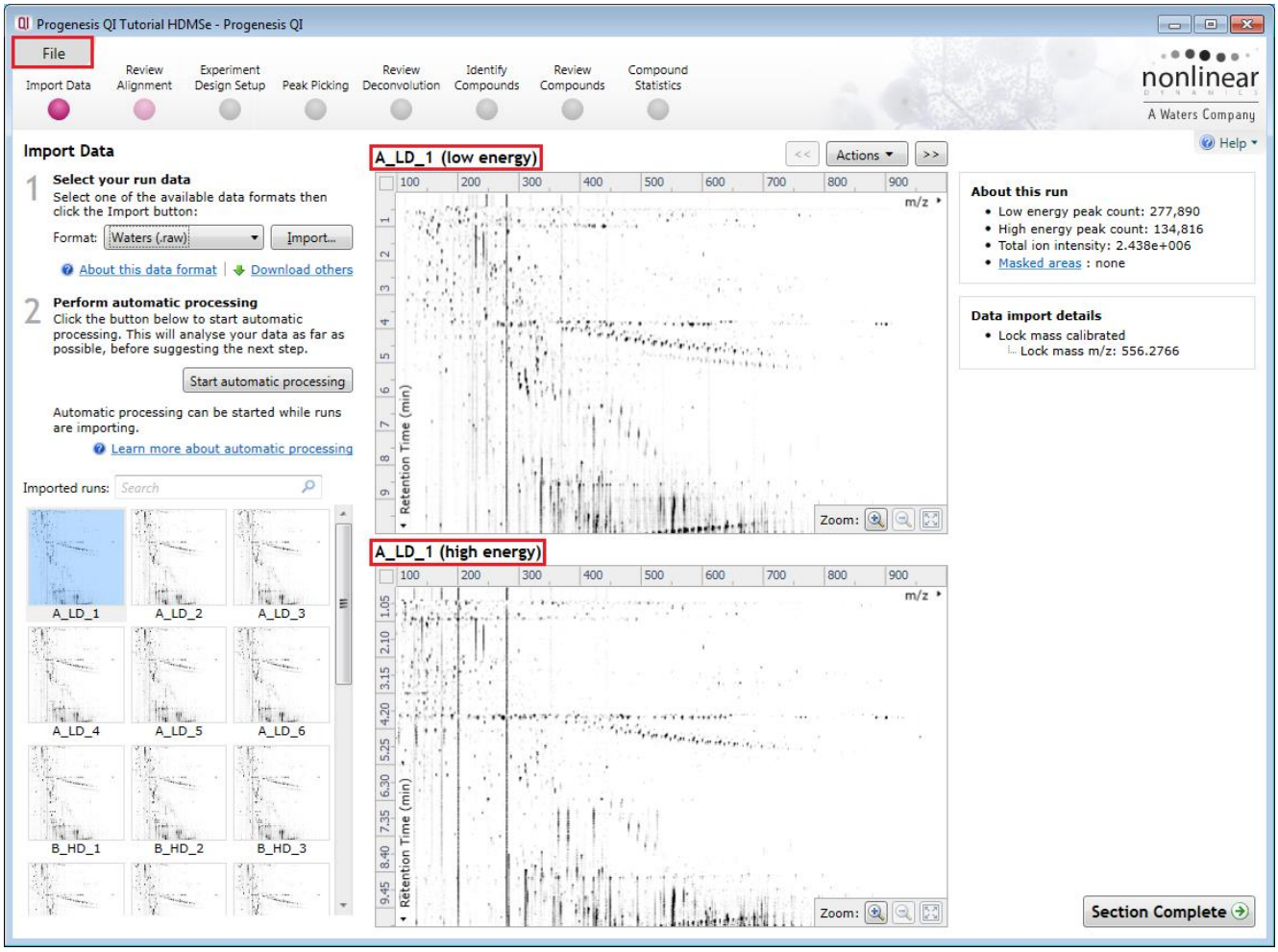
Select the **Create a new experiment** option and select the folder in which you placed the archive, using **Browse**.



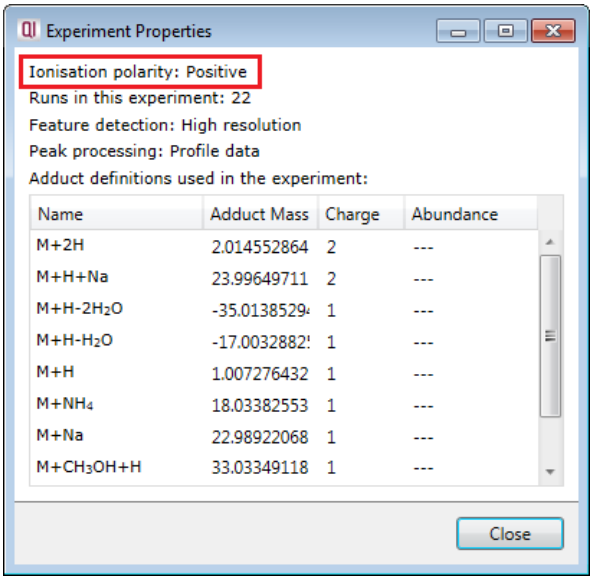
Then press **Import.**

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

The archive will restore with the software opening at the first stage in the workflow, Import Data. The panel on the left will display a 'thumbnail' for each run in the archive, indicating that they have been imported successfully.



The data file format for this tutorial was **.Raw**. It was acquired from a **SYNAPT-G2S** with the ionisation polarity set to **positive**.



You can look at the current **Experiment Properties** using the file menu. This displays details of the adduct definitions that were used during data import.

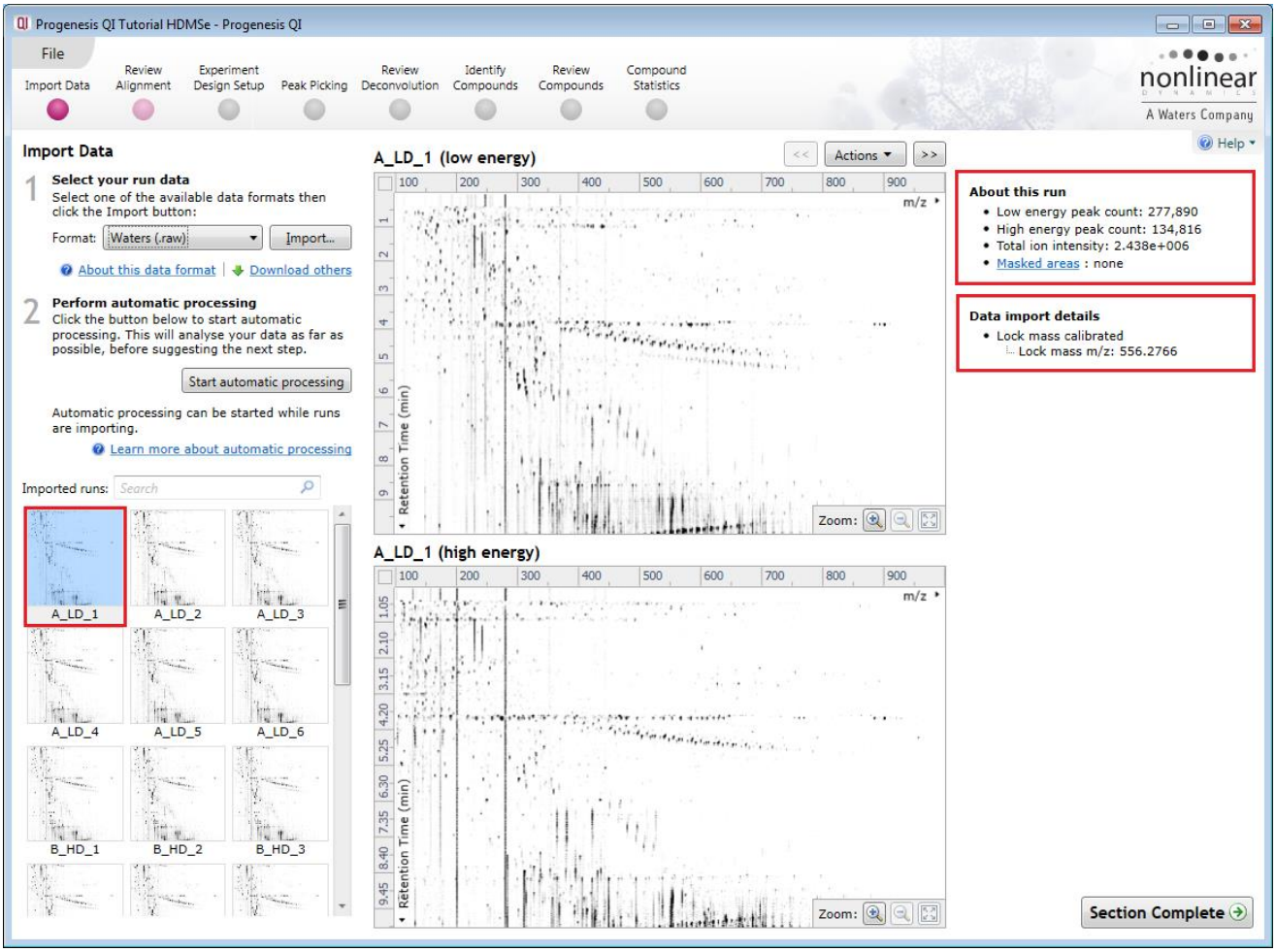
Note: total adduct abundances will be reported once Peak Picking has been performed.

As this example is using HDMSe data it contains both low and high energy exact mass data.

The HD (High Definition) refers to the presence of Ion Mobility data (Drift Time)

Stage 1: Data import and QC review of data set

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



Each run appears as a 2D ion intensity map in the list. The current Run is displayed in the main window showing both low and high energy data. At this stage you will be warned if there are any data import errors for any of the files. The files will be highlighted in red and the error will appear to the right of the screen.

Note: if you have imported one or more runs that are either: not required for the experiment or are displaying data import errors (such as incorrect polarity) these runs can be removed by right clicking on the run in the list and selecting **Remove run**.

Details about the current run are displayed, top right, showing the Low and High energy peak counts and Lock mass calibration status.

Tip: the **'Mask areas for peak picking'** facility (under **Actions**) allows you to exclude areas (usually early and/or late in the LC dimension (Retention Time) that appear excessively noisy due to capture of data during column regeneration. This is **NOT** used for this data set.

Now click **Start automatic processing**.

Stage 2A: Automatic Processing of your data

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

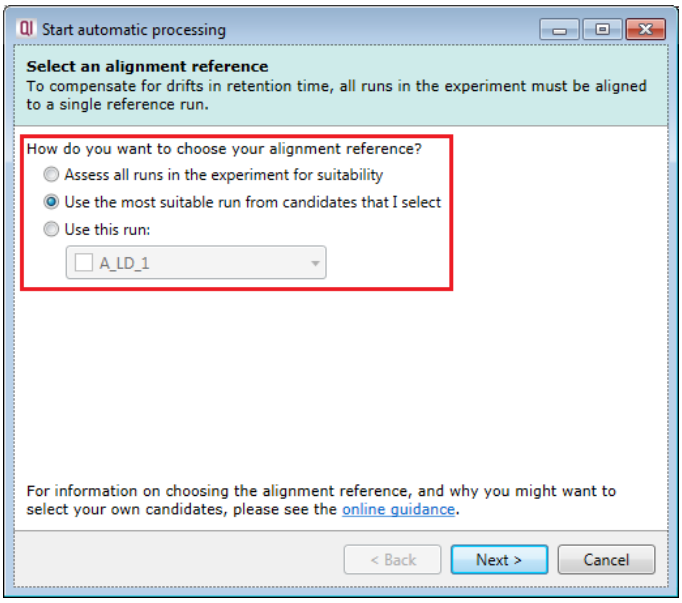
Note: for this tutorial the data has been imported already.

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

- Select an alignment reference
- Automatic alignment of all runs to a reference run
- Define an Experiment design
- Automatic peak picking for compound ion detection

In this tutorial example you have 22 HDMSe LC-MS runs.

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



Progenesis QI provides three methods for choosing the alignment reference run, as described below:

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

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

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

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

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

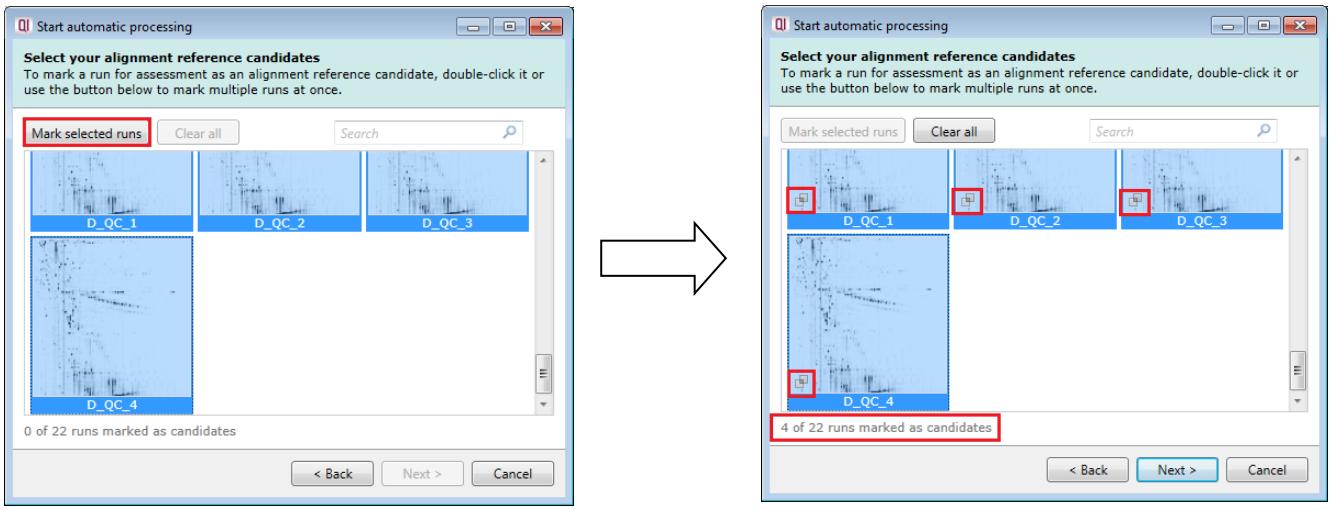
Manual selection gives you full control, but there are a couple of risks to note:

- If you choose a pending run which subsequently fails to load, alignment will not be performed.

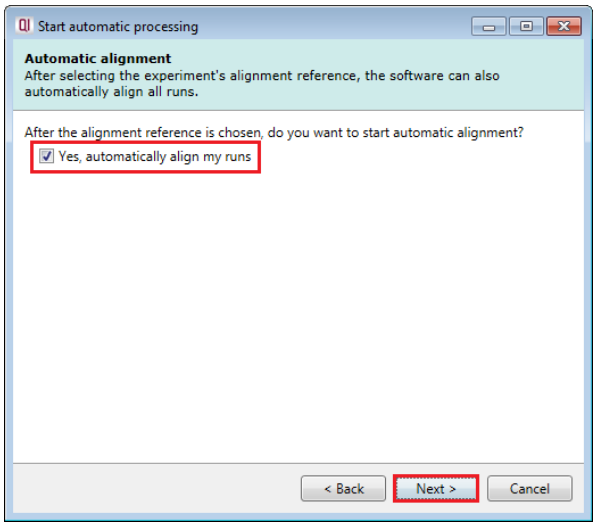
- If you choose a run before it fully loads, and it turns out to have chromatography issues, alignment will be negatively affected (for this reason we recommend that you let your reference run fully load and assess its chromatography before loading further runs).

For this tutorial we will select the **second** option and click **Next**.

Now select all four of the QC runs as the Alignment Reference Candidates and click **Mark selected runs**. These will be marked with a Candidate icon. Now click **Next**.



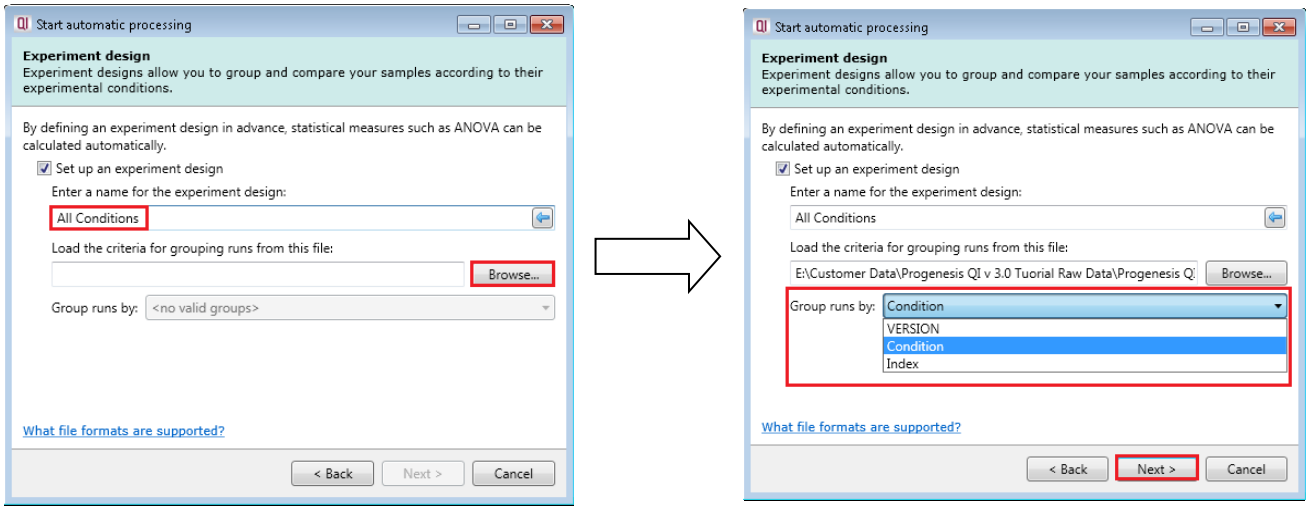
You will now be asked if you want to align your runs automatically. Check box is ticked and select **Next**.



The next page of the processing wizard will ask you to define an **Experiment Design.** To enter a design at this stage details of Experiment conditions must be available either as a MassLynx .SPL file or as a .CSV file.

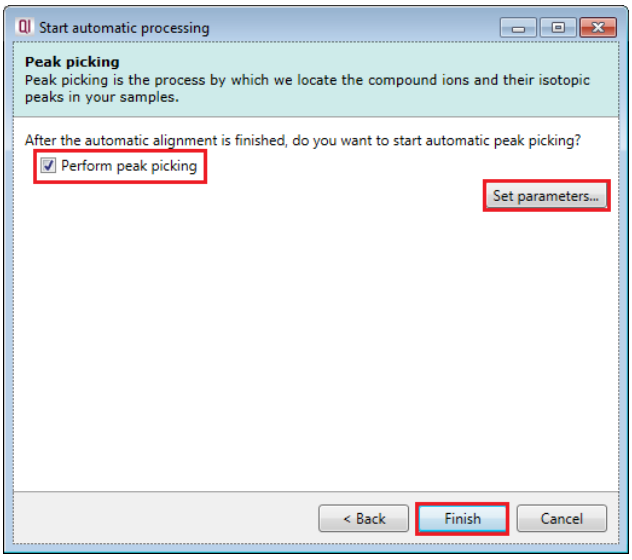
Note: if these are not available leave the Set up option unticked and click next. You can enter a design(s) following automatic processing.

For this tutorial the .spl file **Progenesis QI_HDMSe Tutorial.spl** is available in the folder you restored the Tutorial Data Set.zip in.



Select the appropriate criteria by which to group by; in this example select ‘Condition’ and then click **Next** to move forward to the next page of the processing wizard.

This asks you if you want to **Perform peak picking** and allow you to **Set parameters**.

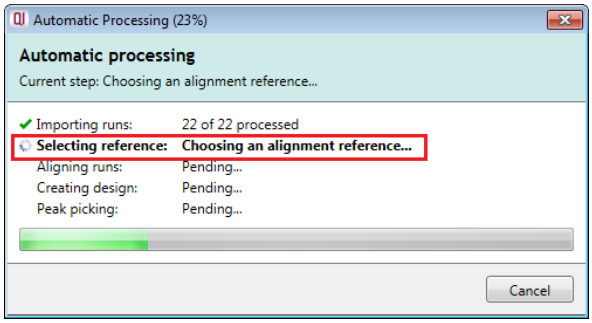


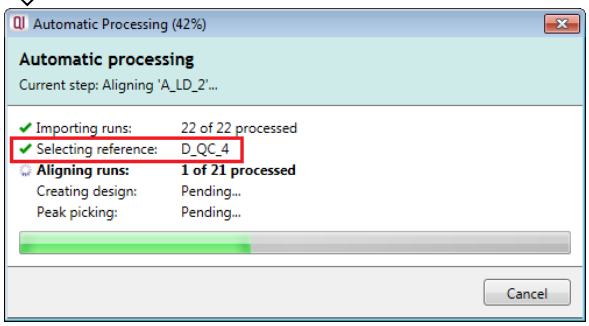
For the purposes of this User guide we will use the default settings for peak picking.

Note: for more details on setting Peak Picking parameters refer to the section on Peak Picking (page 25).

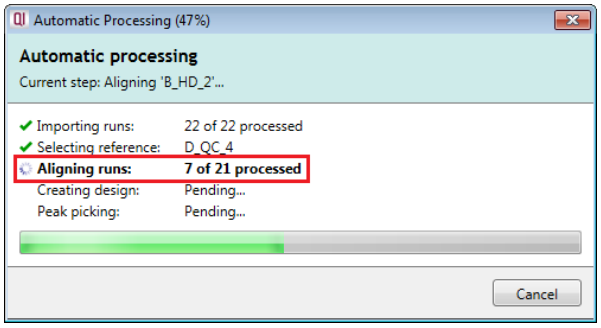
Now click **Finish** to start the automatic processing.

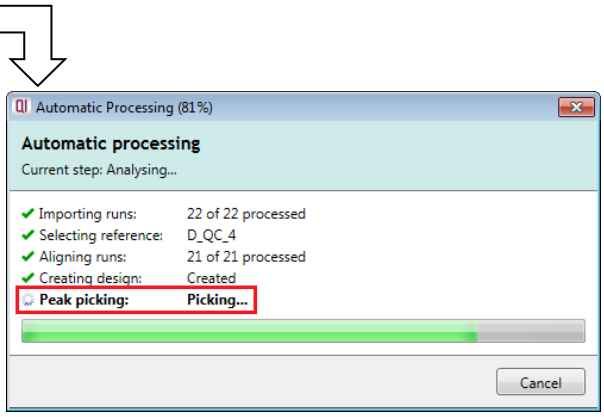
The Alignment process starts with the automatic selection of **D_QC_4** as the reference



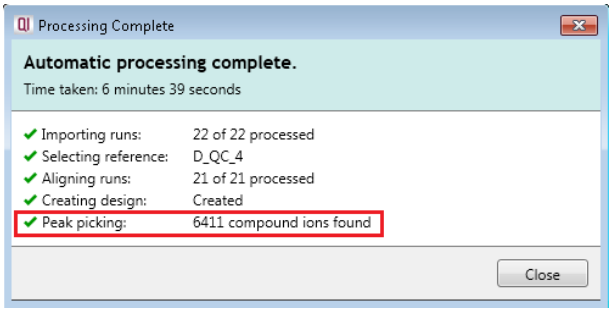


Once Alignment completes the Design is applied and Peak Picking commences.





Finally the Automatic Processing completes reporting the number of compound Ions found.

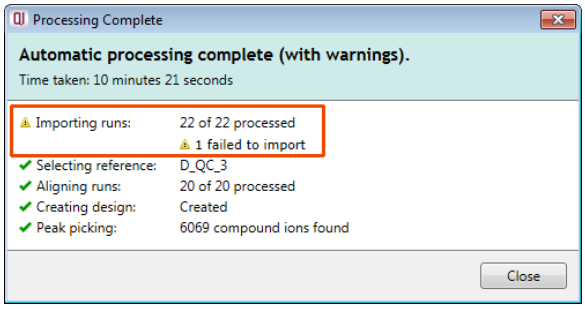


As the whole process proceeds you get information on what stage has been performed and also the % of the process that has been completed.

Note: if Progenesis QI, during import of data and/or performing the various stages of Automatic Processing, detects problems with the data then the Processing dialog will report on the arising issue(s).

For example: failing to import one or more files.

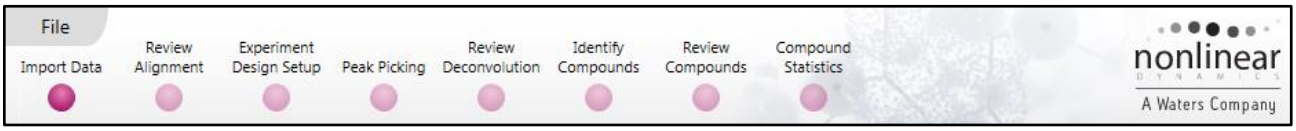
Details on the various warnings and managing failure messages are available in Appendix 2 (page 93)



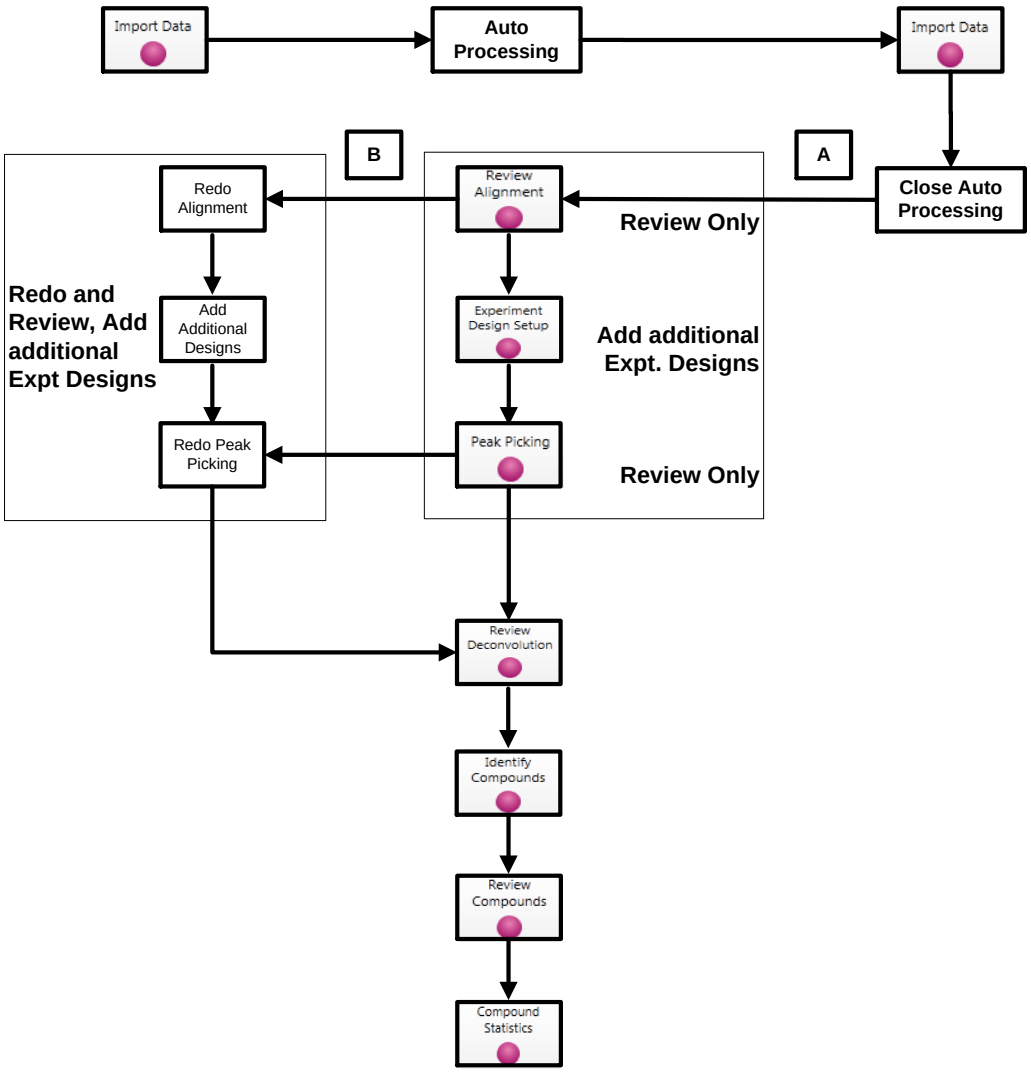
When the Automatic Processing completes click **Close**, the software remains at the Import Data stage of the workflow where you can review the chromatography of the imported runs and remove/replace any runs that have failed to import.

Stage 2B: After Automatic Processing

When Processing completes, access to the later steps in the analysis workflow will depend on the steps performed in the automatic processing. Here we outline the various paths and options you can take to proceed with the analysis following Automatic Processing.



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



In this example where the data has been automatically processed once you **Close** the completed automatic processing dialog you can either:

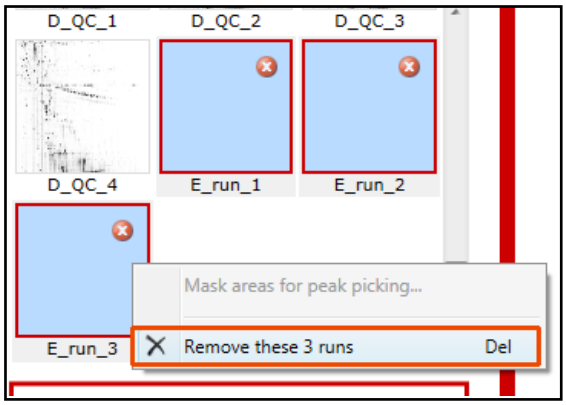
- **Option A**: Review the Chromatography at the Import Data stage (page 13) and then continue with Reviewing Alignment quality (page 15).
- **Option B**: During the review process you decide to redo the alignment (page 15) and/or Peak Picking (page 25).

Note: the tutorial data described here does not require redoing either alignment or peak picking so you can step straight to Review Deconvolution. However, this guide is designed to support the analysis of your own data sets as well, so it covers all stages of the workflow.

Tip: it is good practice to close your experiment following automatic analysis and create an archive (see FAQs for details on creating an archive), and then reopen it and continue with analysis.

Review Chromatography

Each data file appears as a 2D representation of the run. In the case of MSe and HDMSe data both high and low energy maps are displayed.

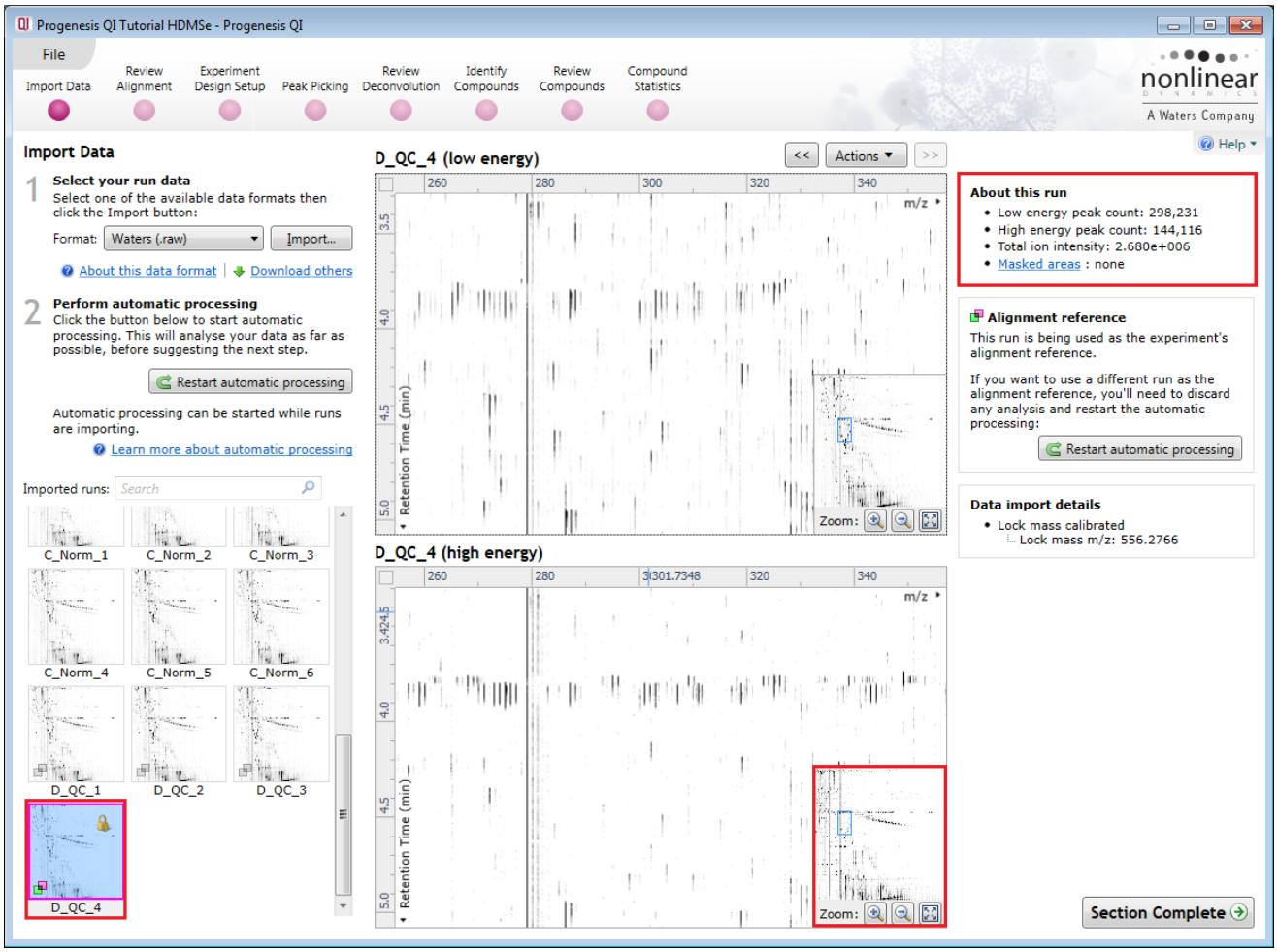


At this stage you will be warned if any of the data files have failed to import. The reasons for import failure(s) include: run is centroided when importer is expecting a profile (continuum) run, the run has a different polarity to the other runs etc.

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

Note: for this data set do not remove any files

You can examine the quality of the imported runs using the 2D representation of the runs making use of the Zoom tools (bottom right of the ion maps) and Panning tools.



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

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

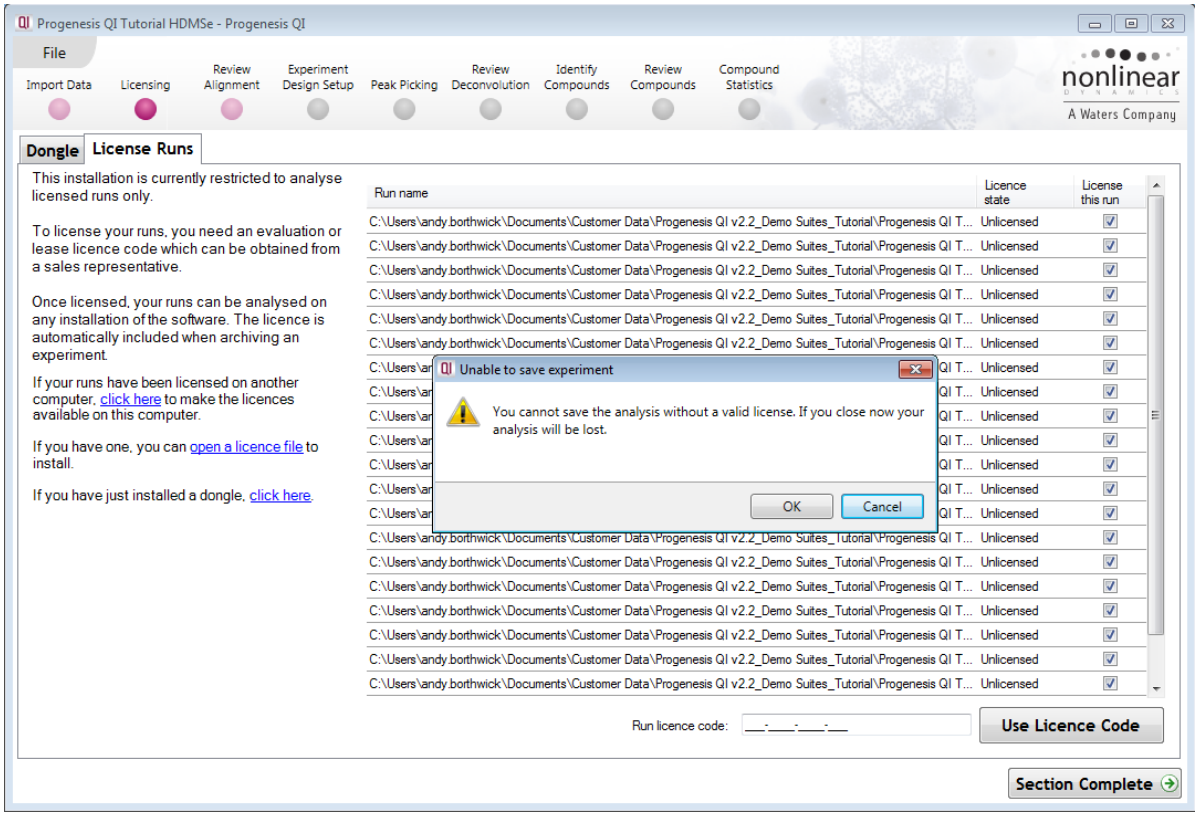
Stage 3: Licensing

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

For details on how to use Licensing go to Appendix 3 (page 96)

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

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



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

For more details on managing the licencing of your data refer to Appendix 3 (page 96)

Stage 4: Review Alignment

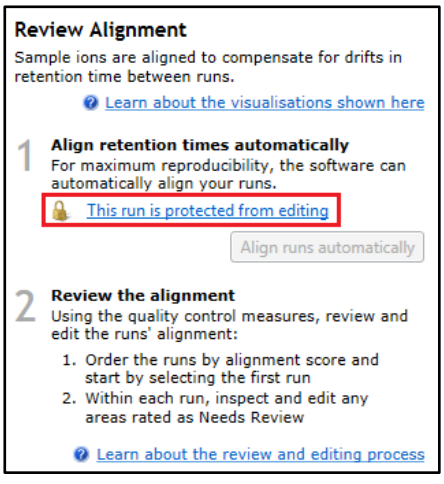
At this stage, **Review Alignment** opens displaying the alignment of the runs to the Reference run (D_QC_4).

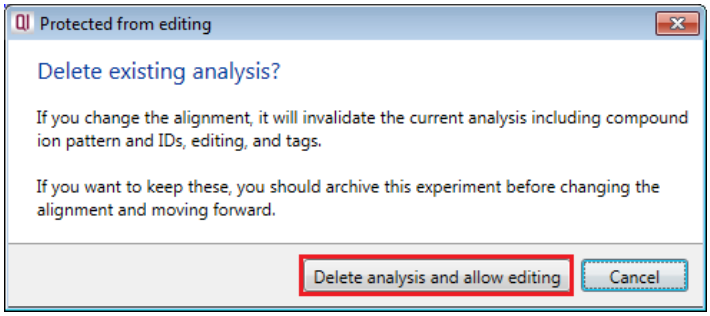
Having performed the analysis automatically there will be icons next to each run in the vectors table indicating that the run is protected from editing.

Note: for this tutorial data set, refinement of the alignment is **NOT** required, so you can move to page 19 without altering the alignment.

However, when reviewing the quality of alignment after 'Automatic Analysis' of a data set you may decide that the alignment requires editing. Reviewing and editing are described in the following sections.

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



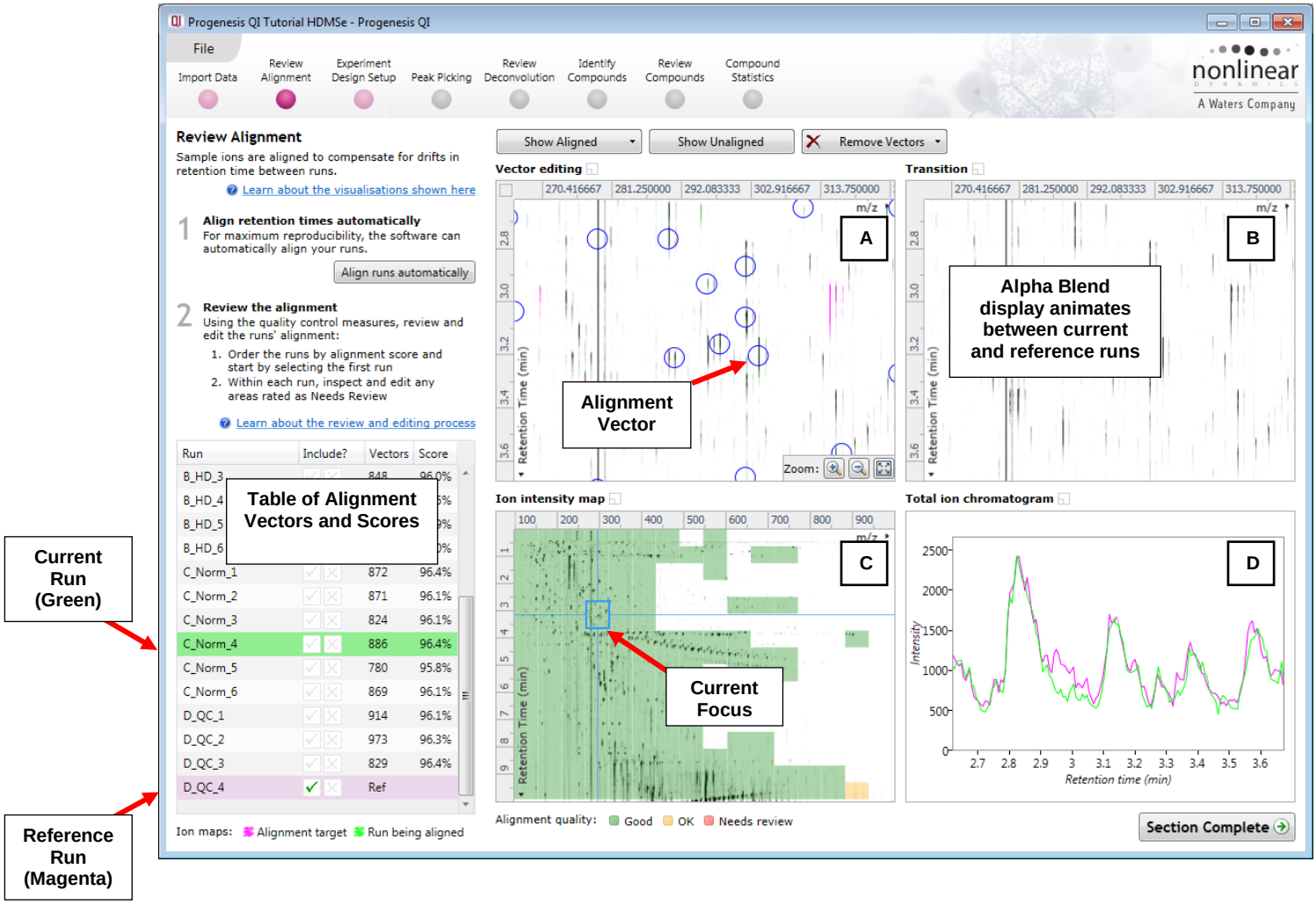


Note: as you click on the link, you will be warned that you are discarding the current analysis (all steps beyond alignment).

Details on editing alignment are described in Appendix 4 (page 97).

Layout of Alignment

To familiarize you with Progenesis QI Alignment, this section describes how to use the various graphical features used in the alignment of the LC-MS runs.



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

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

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

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

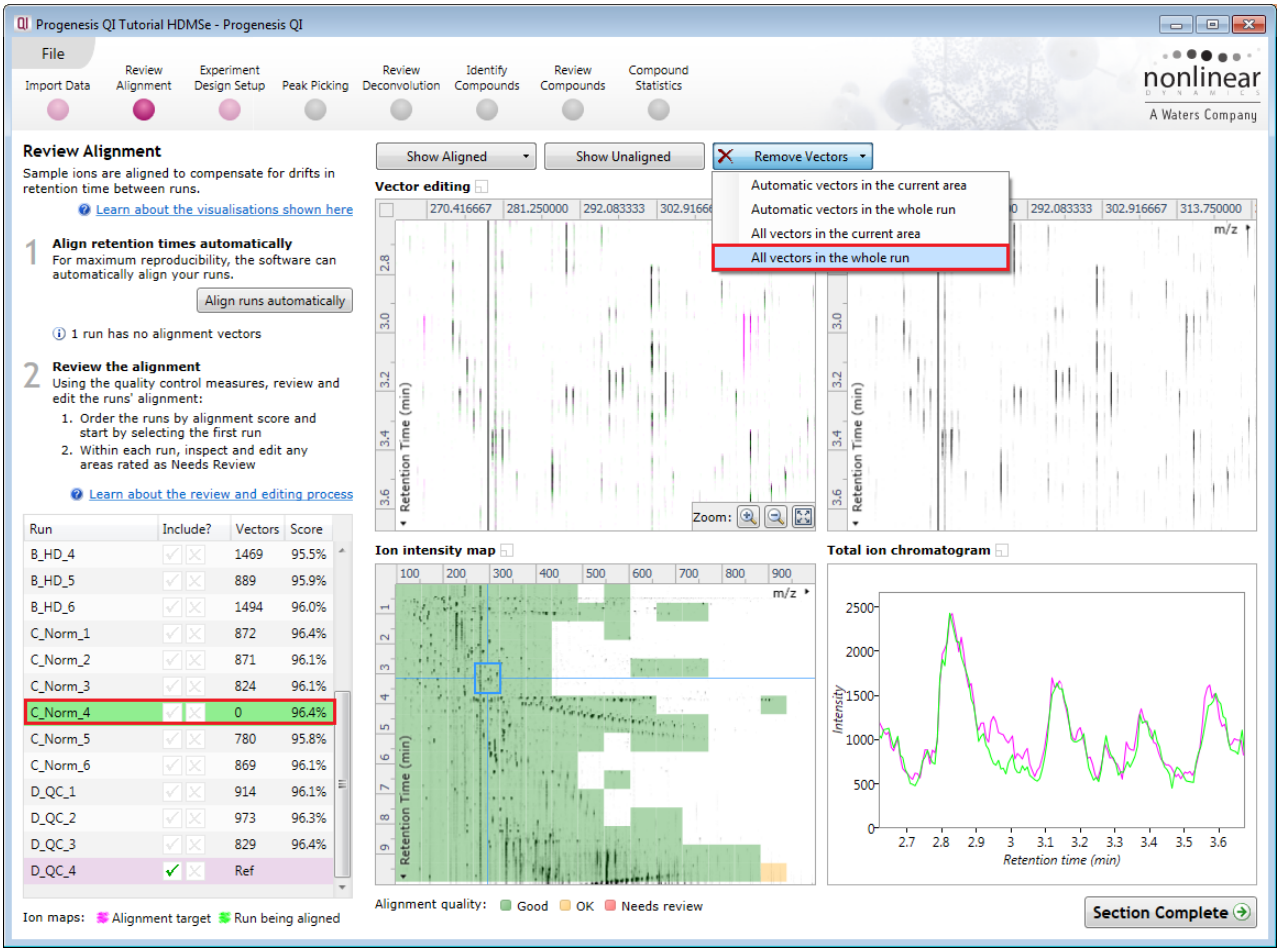
To best understand the process of alignment and the meaning of the scoring we will now undo the automatic alignment for C_Norm_4, then simulate a poorly aligned run by adding an incorrect manual vector.

Run	Include?	Vectors	Score
B_HD_4		1469	95.5%
B_HD_5		889	95.9%
B_HD_6		1494	96.0%
C_Norm_1		872	96.4%
C_Norm_2		871	96.1%
C_Norm_3		824	96.1%
C_Norm_4		886	96.4%
C_Norm_5		780	95.8%
C_Norm_6		869	96.1%
D_QC_1		914	96.1%
D_QC_2		973	96.3%
D_QC_3		829	96.4%
D_QC_4	✓	Ref	

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

- If you have not already unlocked the Analysis (note padlock icons in the runs table see right) then refer to page 15.
- In the Run table click on C_Norm_4 to make it current. You will now be looking at the alignment of C_Norm_4 to D_QC_4 in the Unaligned view.

First click on **Remove Vectors** and then select remove All vectors in the whole run.

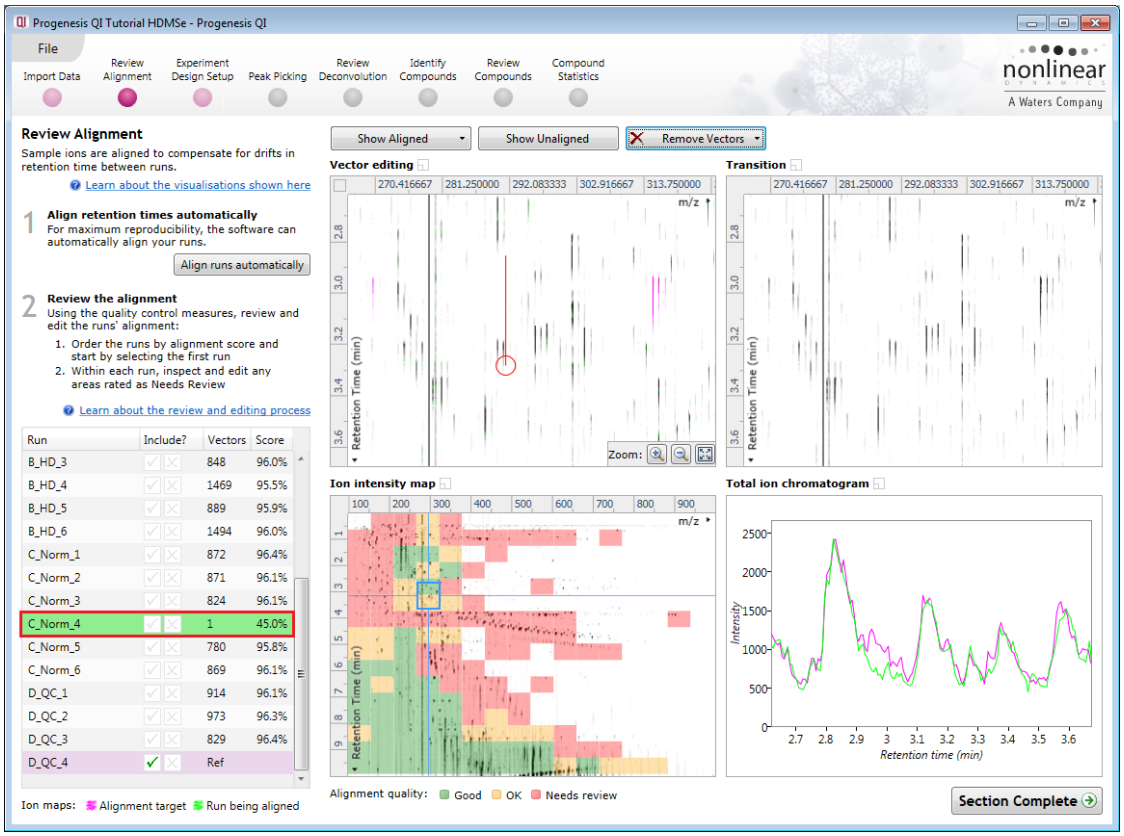


Note: the Retention Time alignment of this data is good between the runs as indicated by a high percentage score (in the absence of any vectors) and the Alignment quality showing as 'all' green.

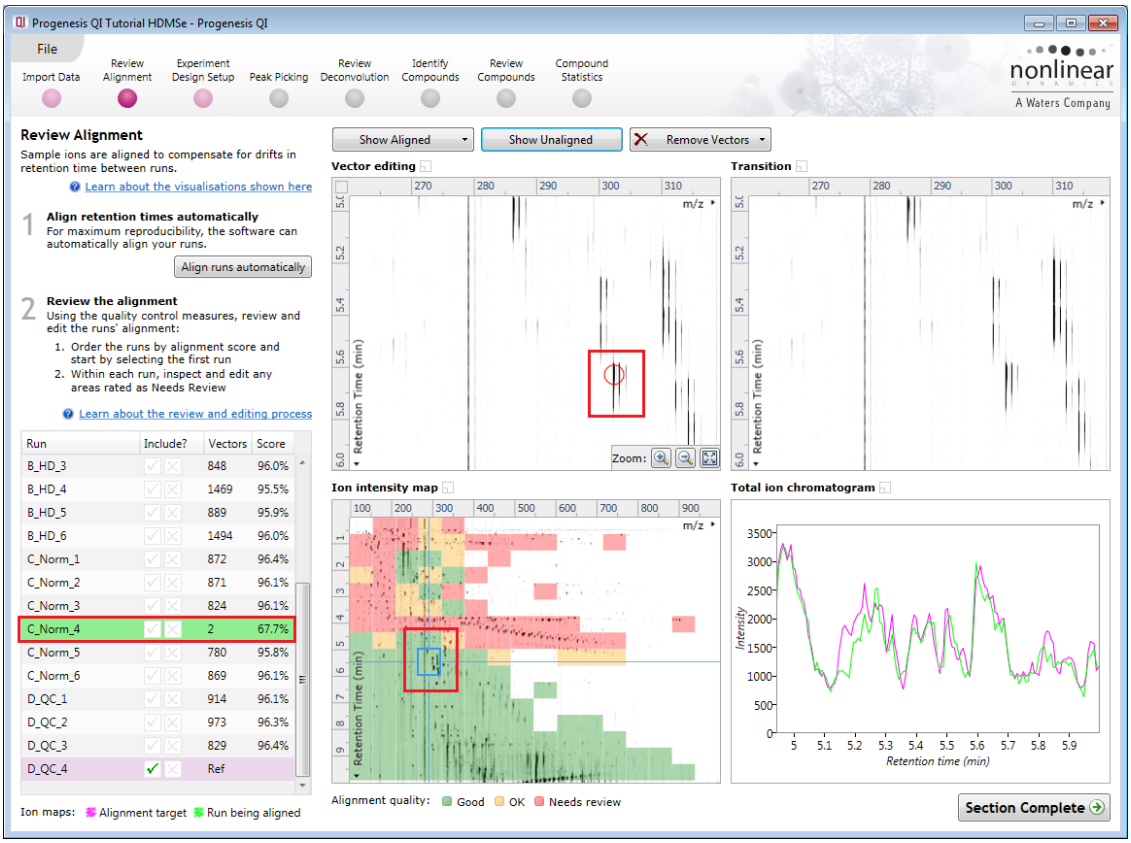
To simulate poor alignment, place a single manual vector on the Vector editing view (Window A). To do this, in the Vector Editing panel, click and drag out a single vector then release the mouse button. By doing this a single manual vector will appear with a length corresponding to the 'drag'.

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

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



Now place a second vector at a different location **BUT** this time hold down the **Alt** key time click and release **without** deliberately misaligning the vector.

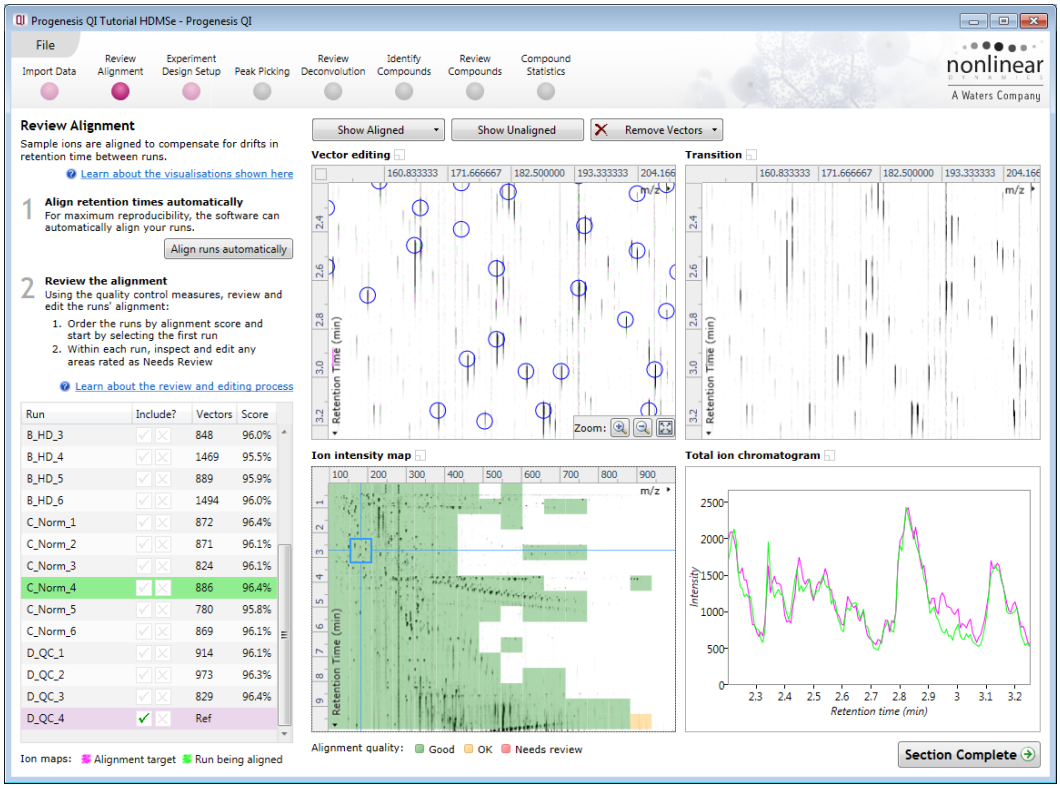


Notice how the correctly placed vector markedly improves the Score this is also reflected in the increased proportion of the Ion Intensity Map showing green indicating a good quality alignment.

Reviewing quality of alignment vectors

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

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

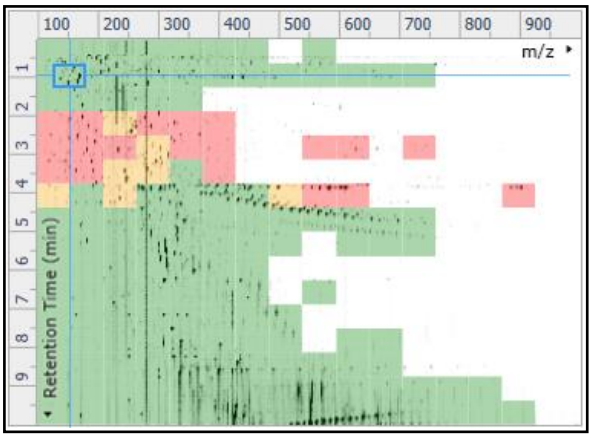


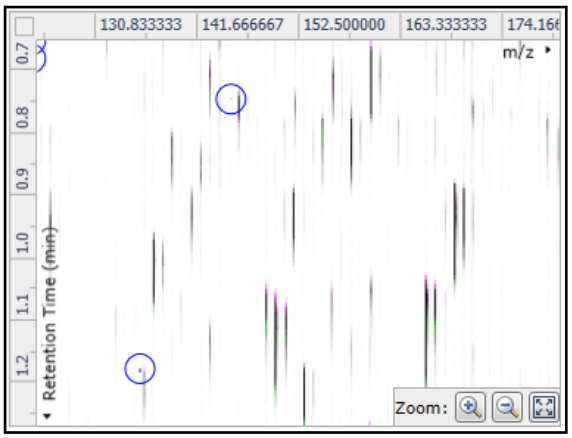
Using the **Simulated** miss-aligned example from the previous section to explain the review process for alignment, the alignment looks as below with a region of poor alignment (highlighted in red).

Reviewing Quality of Alignment

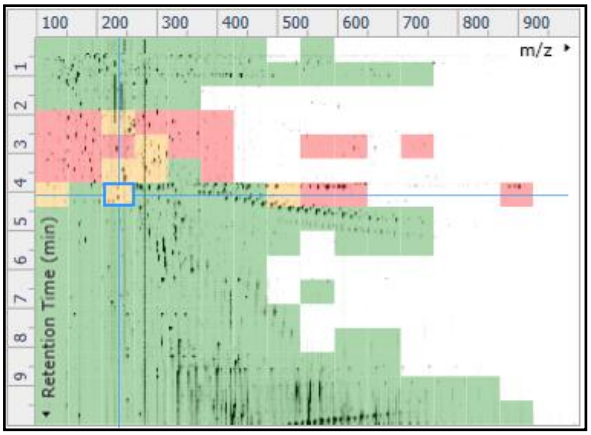
At this point the quality metric, overlaid on the Ion Intensity Map as coloured squares, acts as a guide drawing your attention to areas of the alignment. These range from Good (Green) through OK (Yellow) to Needs review (Red). Drag out a 'Focus' area that corresponds to one of the coloured squares. Three example squares are examined here.

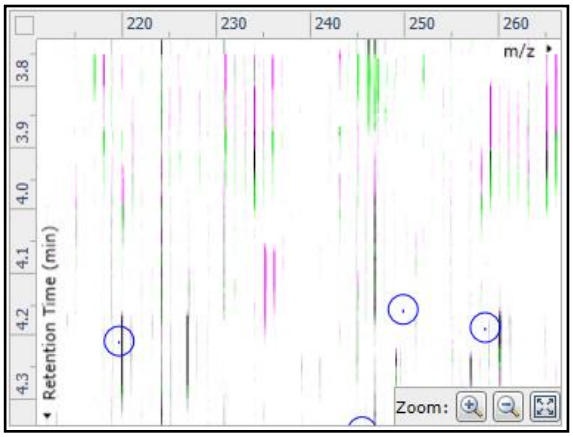
For a 'green' square the majority of the data appears overlapped (black) indicating good alignment. When viewed in the Transition view the data appears to pulse.



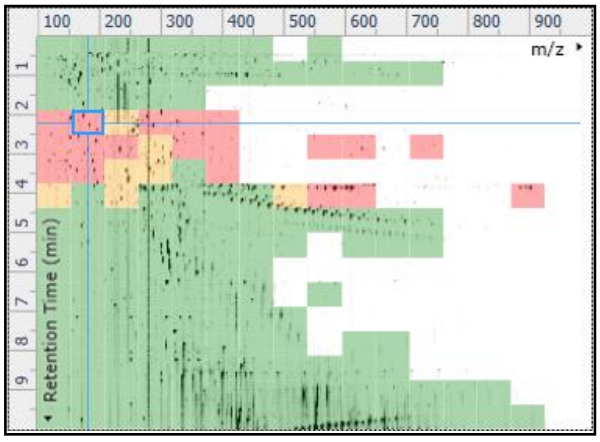


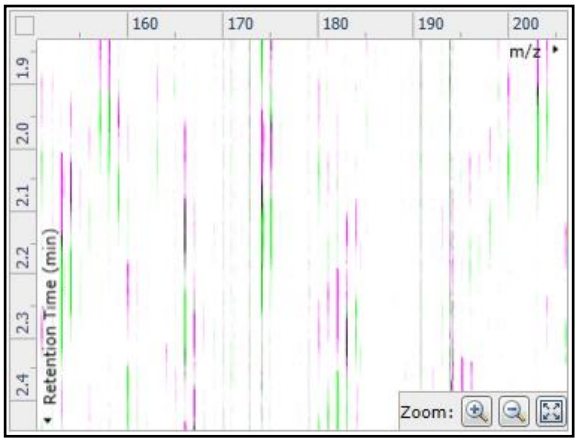
For a 'yellow' square some of the data appears overlapped (black) indicating OK alignment. When viewed in the Transition view some of the data appears to pulse.





For a 'red' square, little if any of the data appears overlapped (black) indicating questionable alignment.

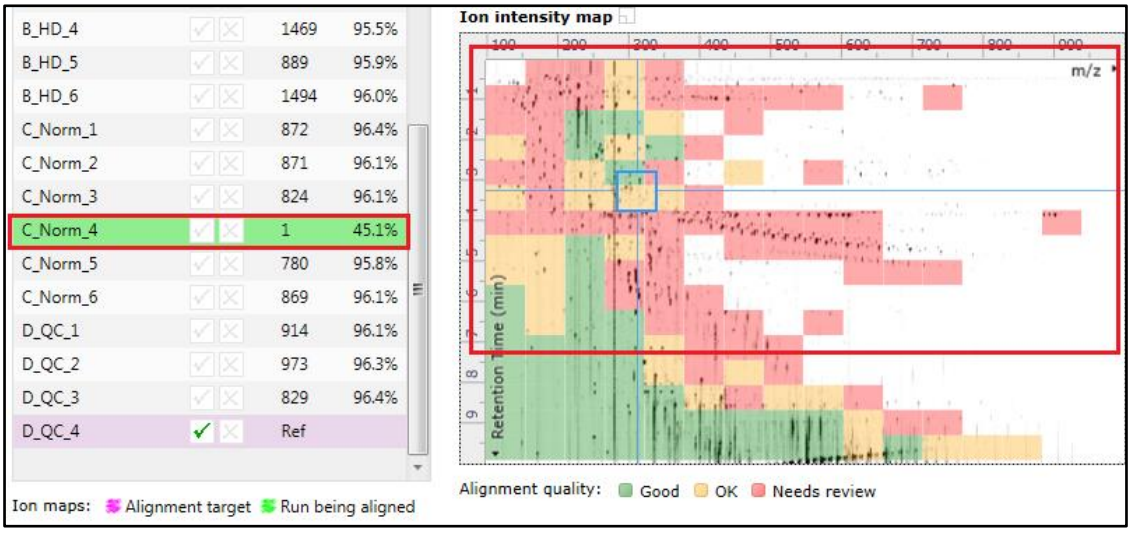




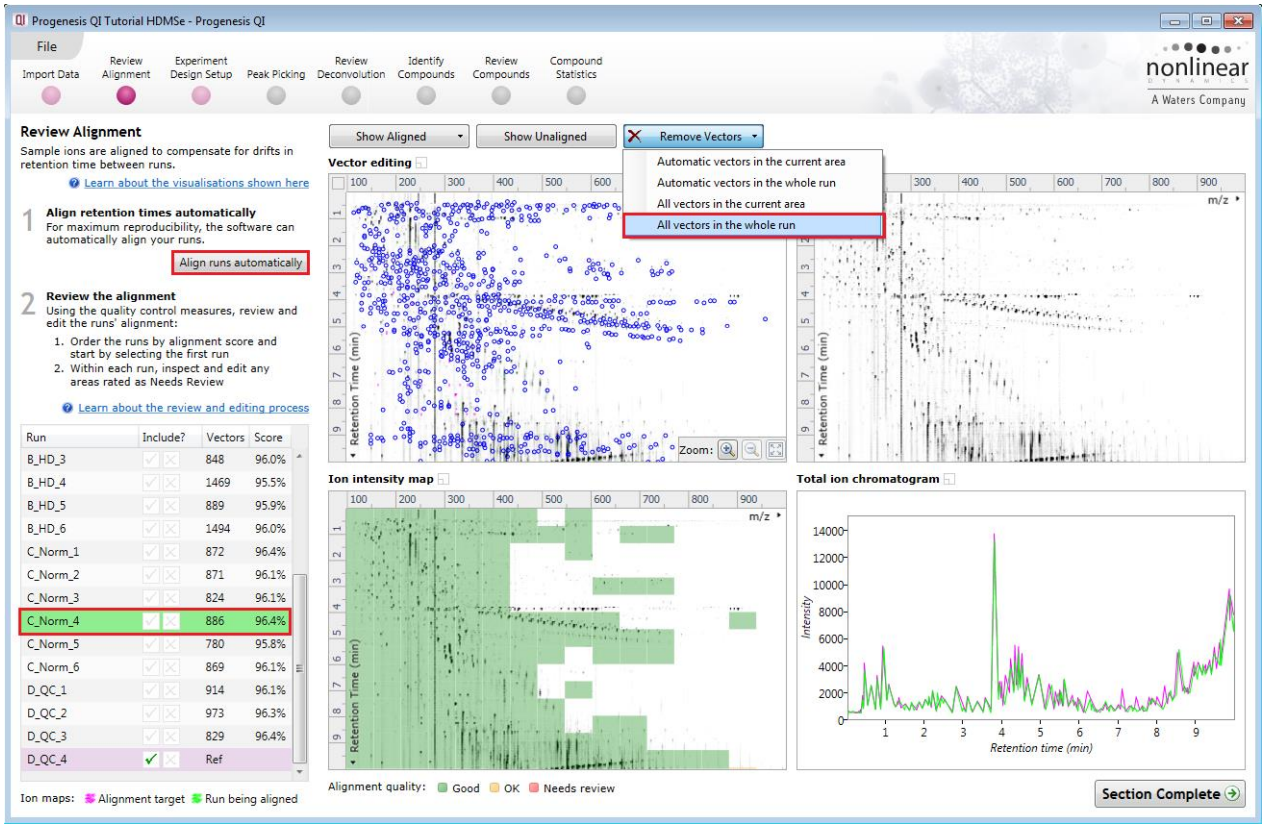
When viewed in the Transition view little data appears to pulse.

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

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



Note: a marked red area combined with a low score clearly indicates a 'miss alignment' and may require some manual intervention (see Appendix 4, page 97).



The alignment quality of this data set does not require any manual intervention so before going to the next section make sure you have Removed all manual vectors and re-performed the Automatic alignment. To do this for C_Norm_4 first select Remove 'All vectors in the whole run' and then click Align runs automatically.

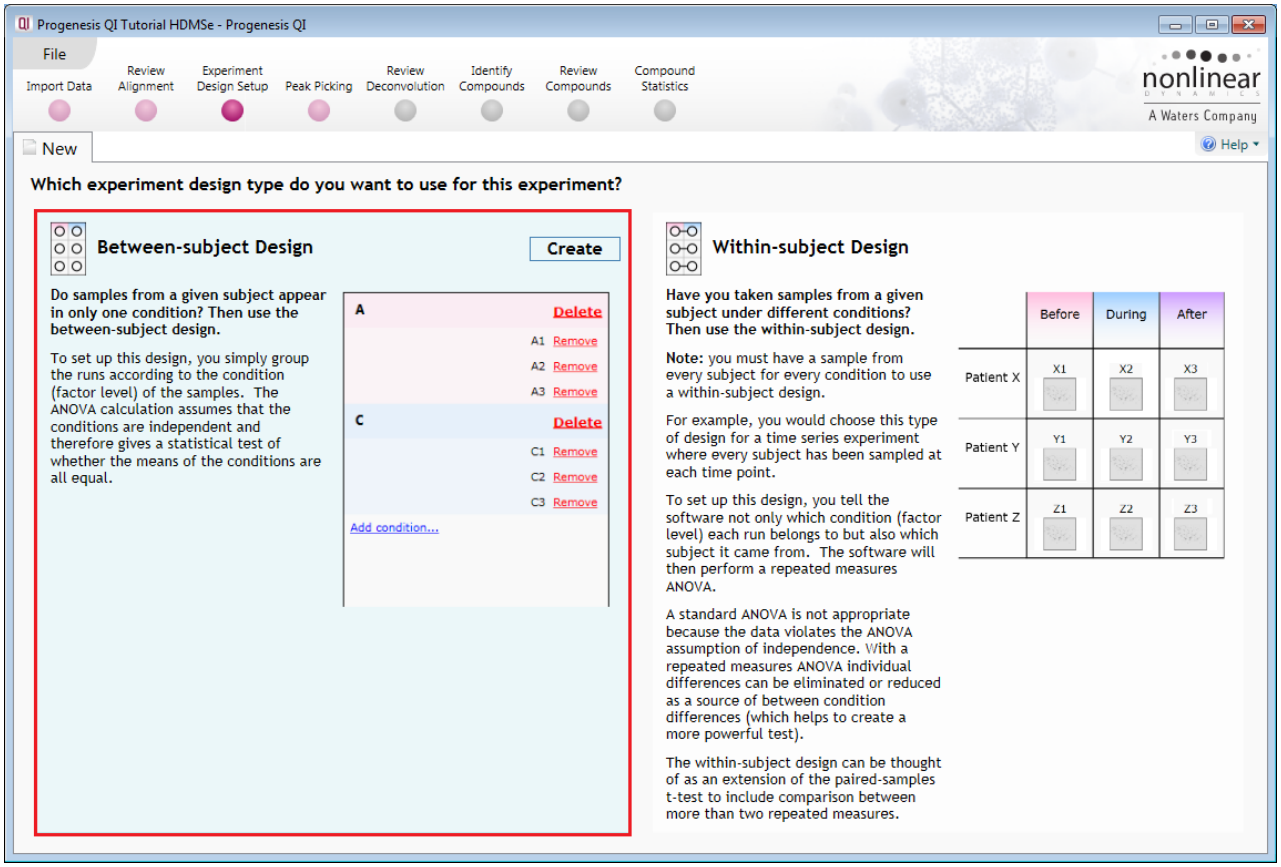
Having aligned the runs automatically click **Section Complete** to move to Experiment Design Setup.

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.

Note: If you have created a design as part of the automatic processing then it will appear as a Design tab. However if you have reperformed the analysis 'manually' then the only tab displayed will be **New**.

There are two basic types of experimental designs:



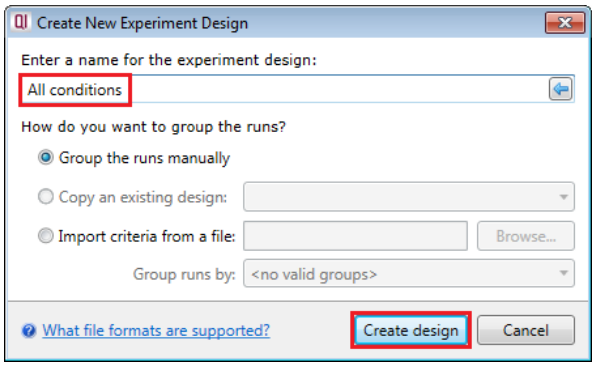
Between-subject design: here samples from any given subject appear in only one condition. (i.e. control versus various drug treatments).The ANOVA calculation assumes that the conditions are independent and applies the statistical test that assumes the means of the conditions are equal.

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

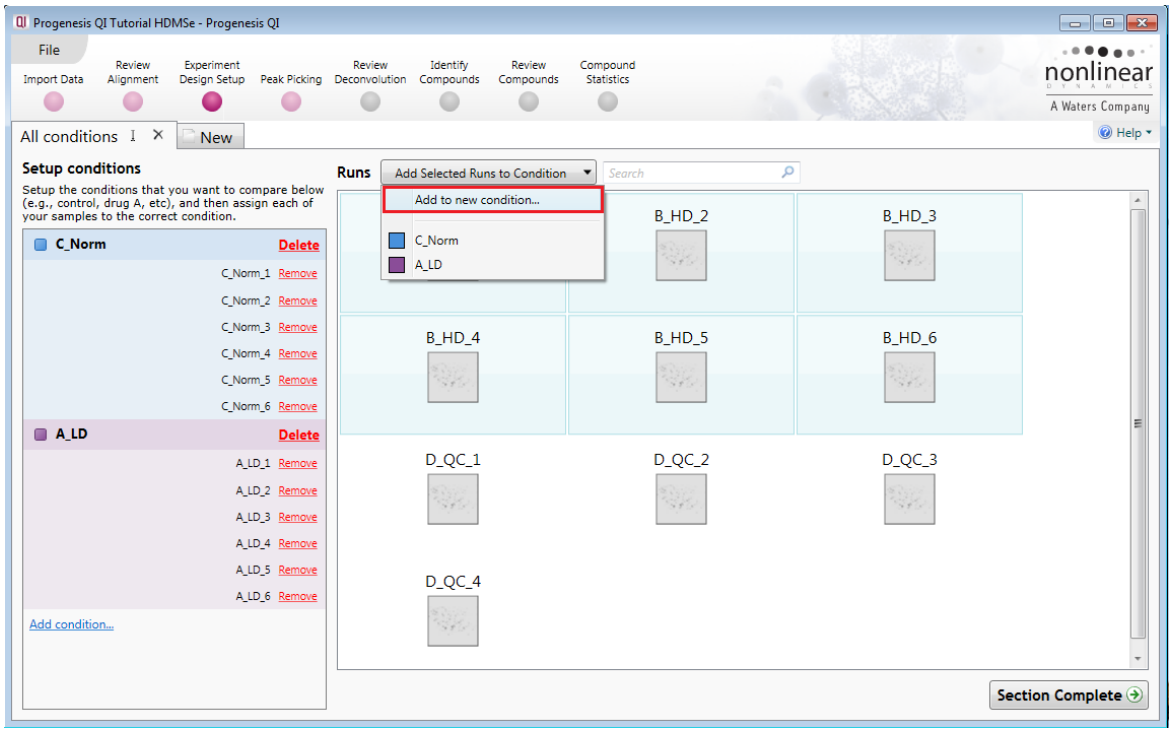
Additional information on how to apply the Within-subject Design is in Appendix 5 page 102

This experiment contains 4 conditions: A, B, C and D and uses the **Between-subject design** to group the analysed runs to reflect the Biological conditions in the original study.

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



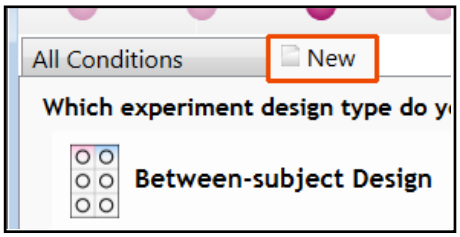
Give the new experimental design a name and then click **Create design**.



To create a new condition

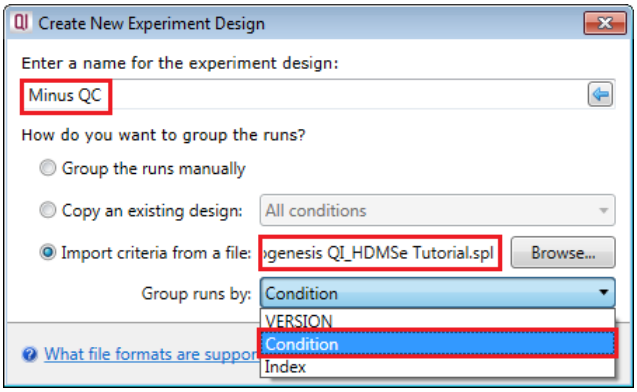
1. Select the runs for the condition by clicking on the required icon in the **Runs** panel, as shown.
2. Press the 'black triangle' next to the **Add Selected Runs to Condition** button on the main toolbar.
3. Select **Add to new condition**… from the drop down menu.
4. A new condition will appear in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

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



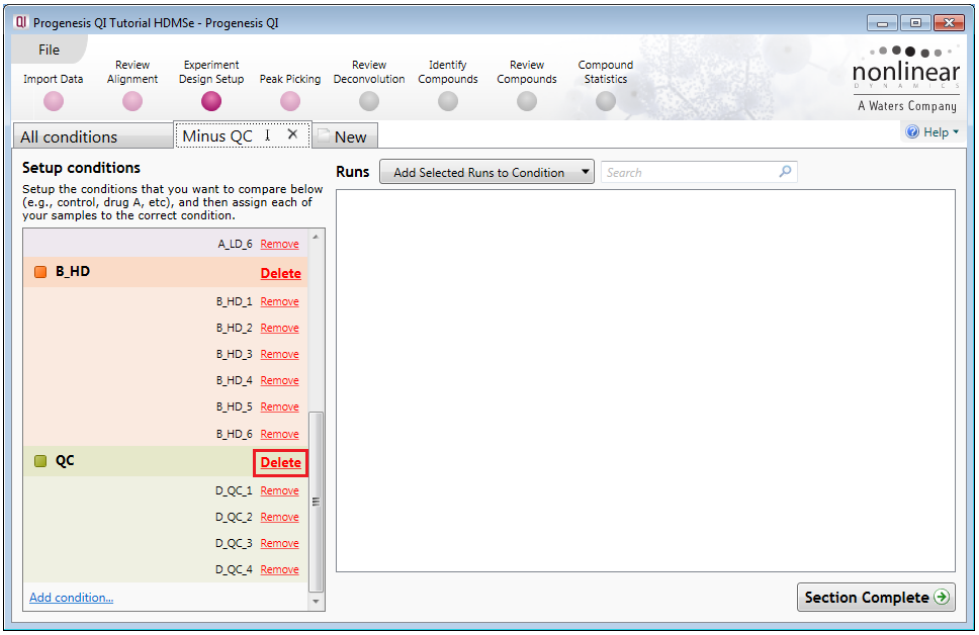
For this example there is a **Progenesis QI_HDMSe Tutorial.SPL** file available in the Experiment Archive you restored at the beginning of this tutorial exercise.

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



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

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



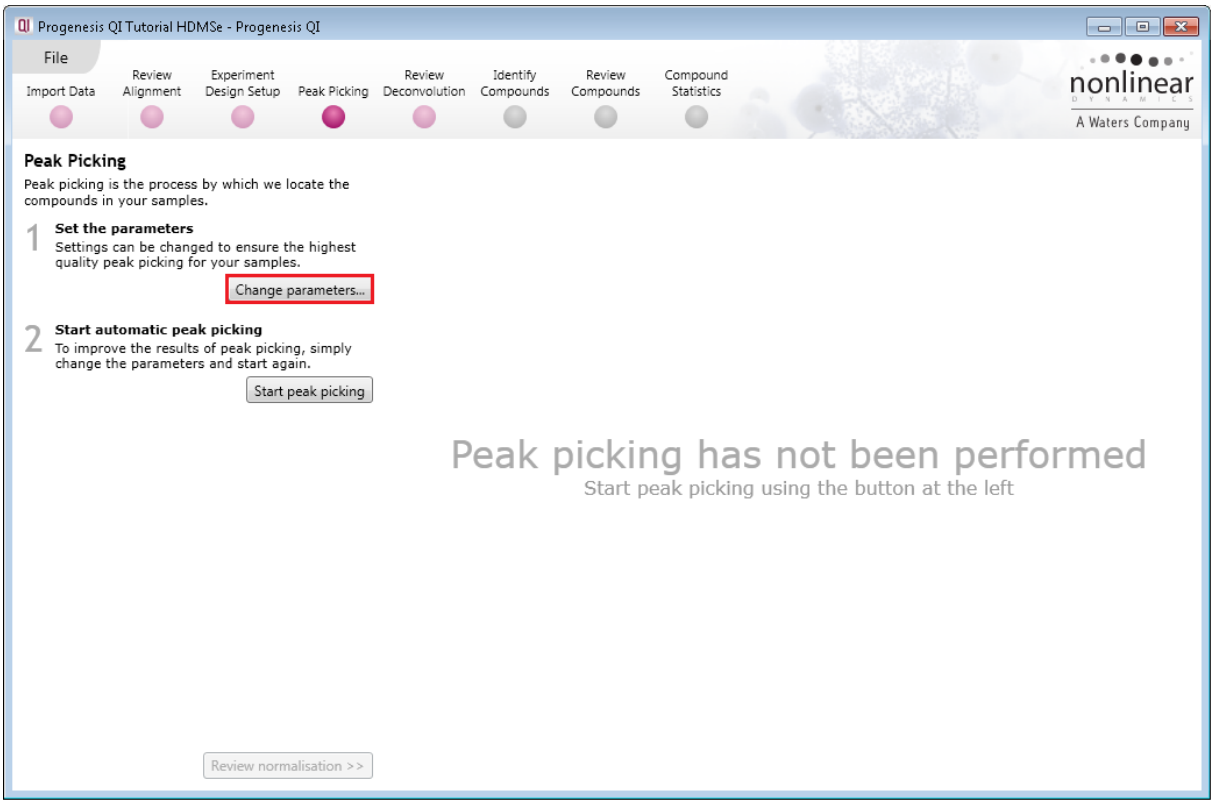
Note: On deleting a condition the runs will reappear in the Runs window.

Note: both designs are available as separate tabs.

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

Stage 6A: Peak Picking

Having set up one or more Experiment Designs for your data the Peak Picking stage will open as shown below. Here you can define thresholds for the detection of peaks across all of your aligned runs.

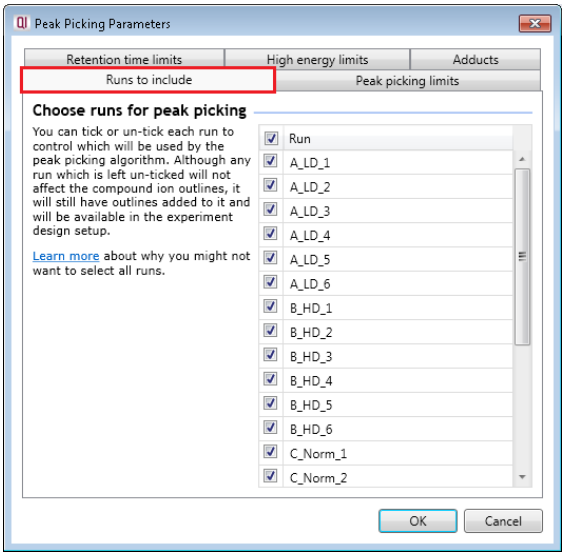


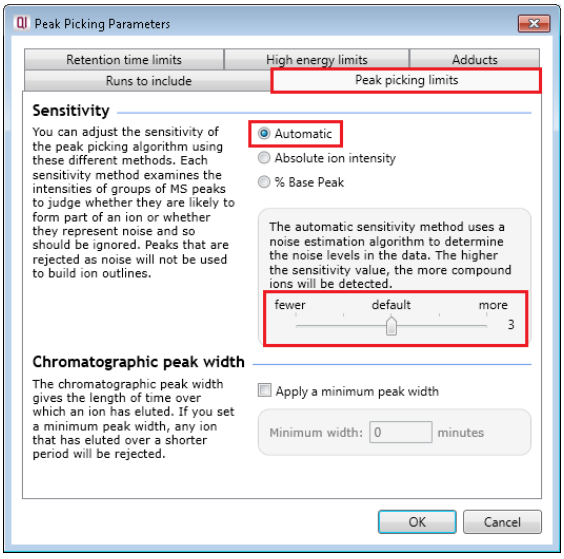
Note: if you have processed your data automatically and have NOT discarded your analysis as described on page 15 then the Peak picking will appear as shown on page 27.

If you have been reviewing the Alignment as described in the previous section then the Peak picking will look as shown above, indicating that you will need to **Start peak picking**

Peak Picking Parameters

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





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

Note: feature outlines will be added to 'un-ticked' runs; however, these runs will not contribute to the peak picking pattern.

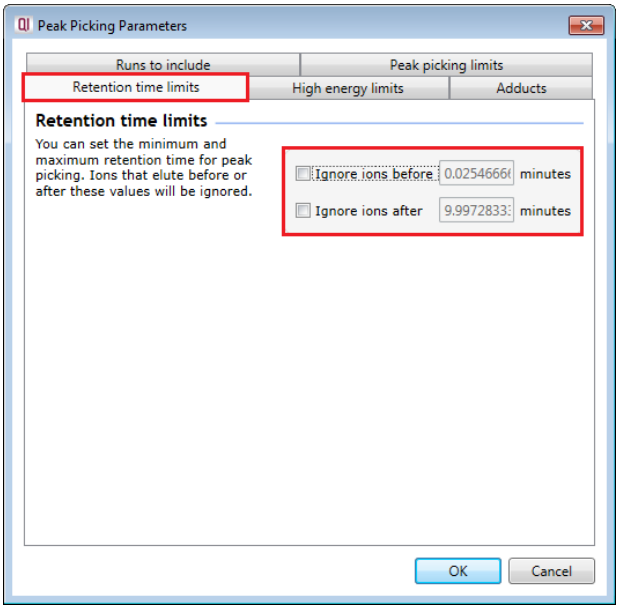
Tip: depending on run quality, a suggested minimum number of ticked runs should include at least one replicate of each experimental condition.

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

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

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

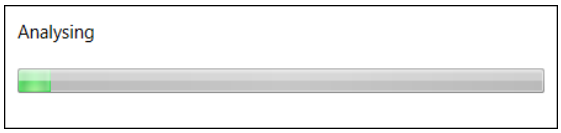
Finally you can set **Retention time limits** for the detection. The default limits are displayed if no limits are defined.



Click **OK** to close the Peak Picking Parameters dialog.

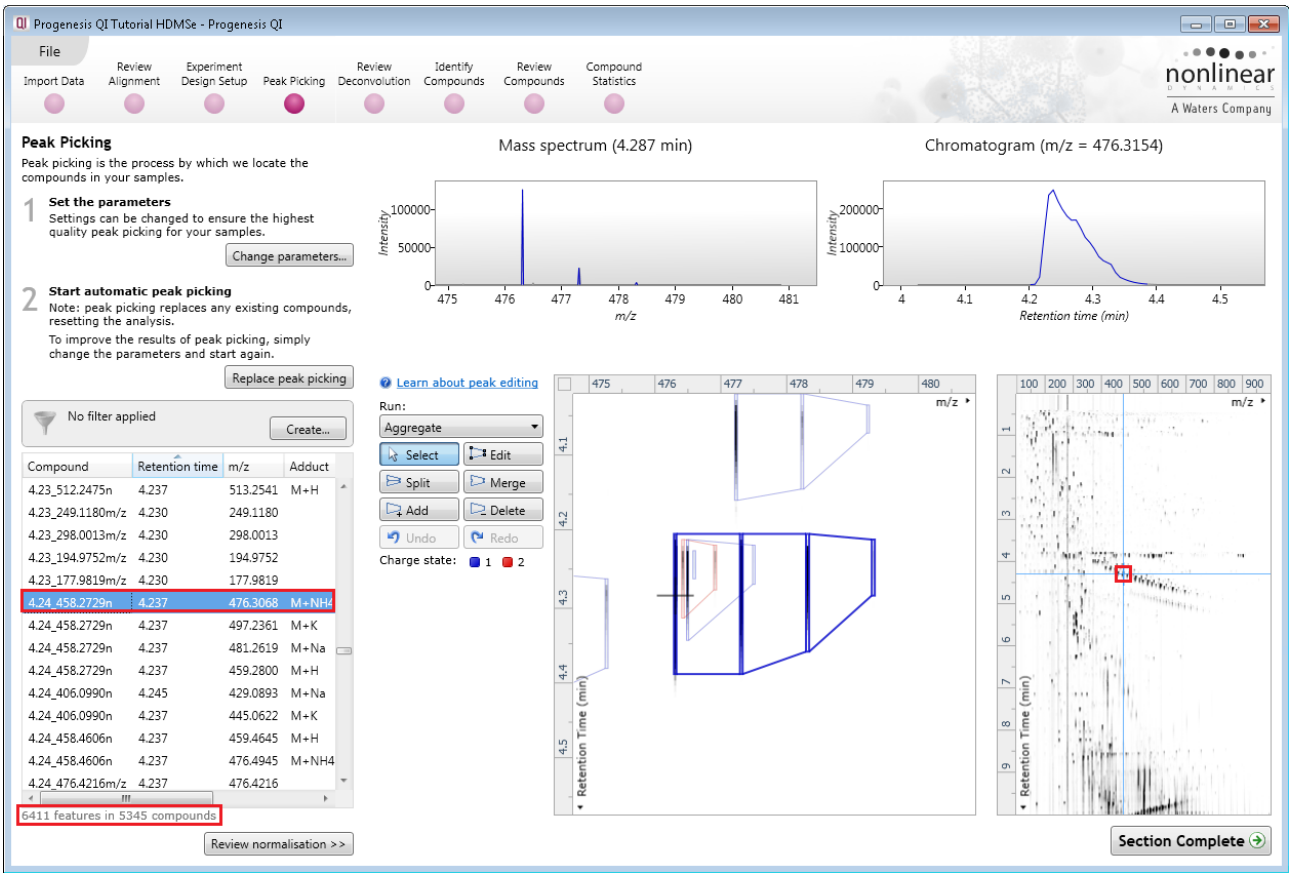
Press **Start peak picking** to start the detection process.

During peak picking, a progress bar will appear telling you that it is Analysing.

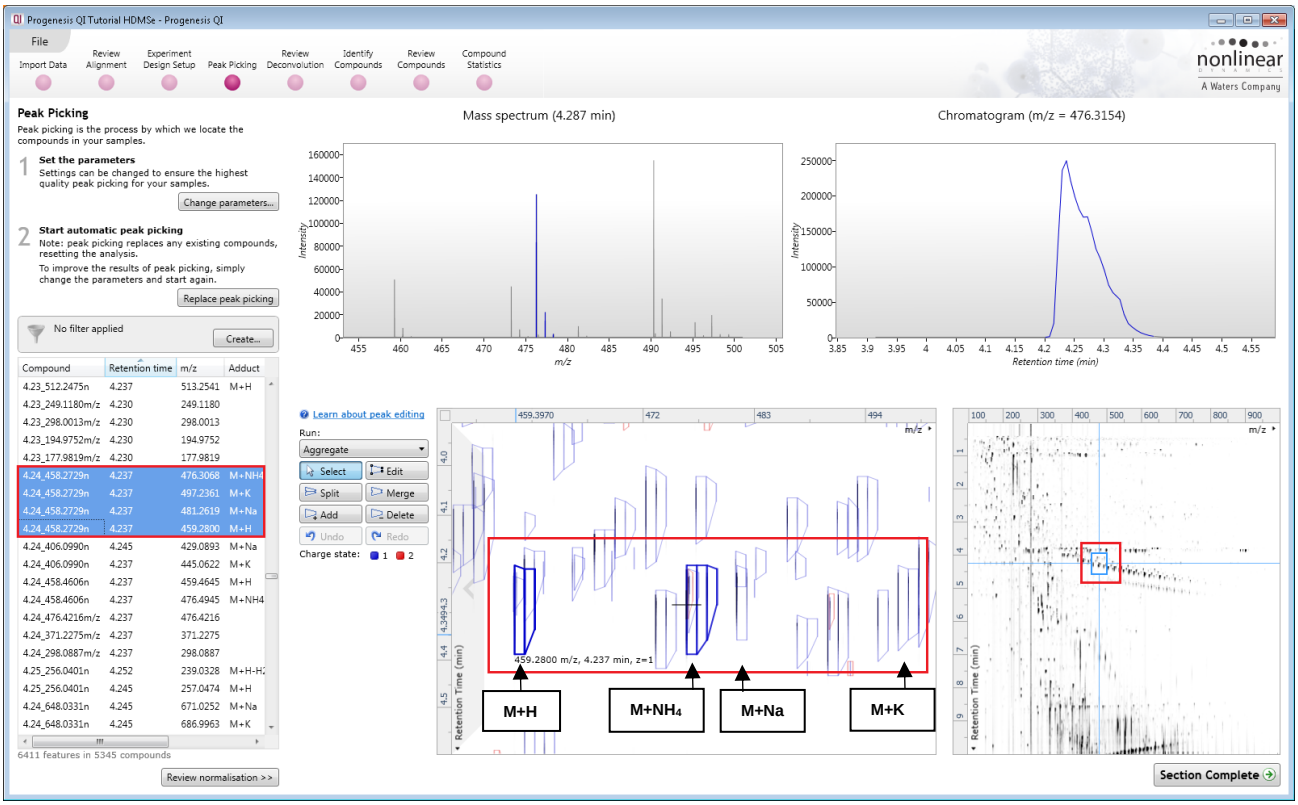


Following automatic peak picking the detected peaks are displayed in a colour according to their charge state. (**note**: number of charge states represented reflects Adduct Definitions used in the experiment)

The actual number of peaks (Compound Ions) detected is recorded on the bottom left of the screen.



To inspect the detection more closely drag out an area of interest (m/z by RT) on the full ion map**.**



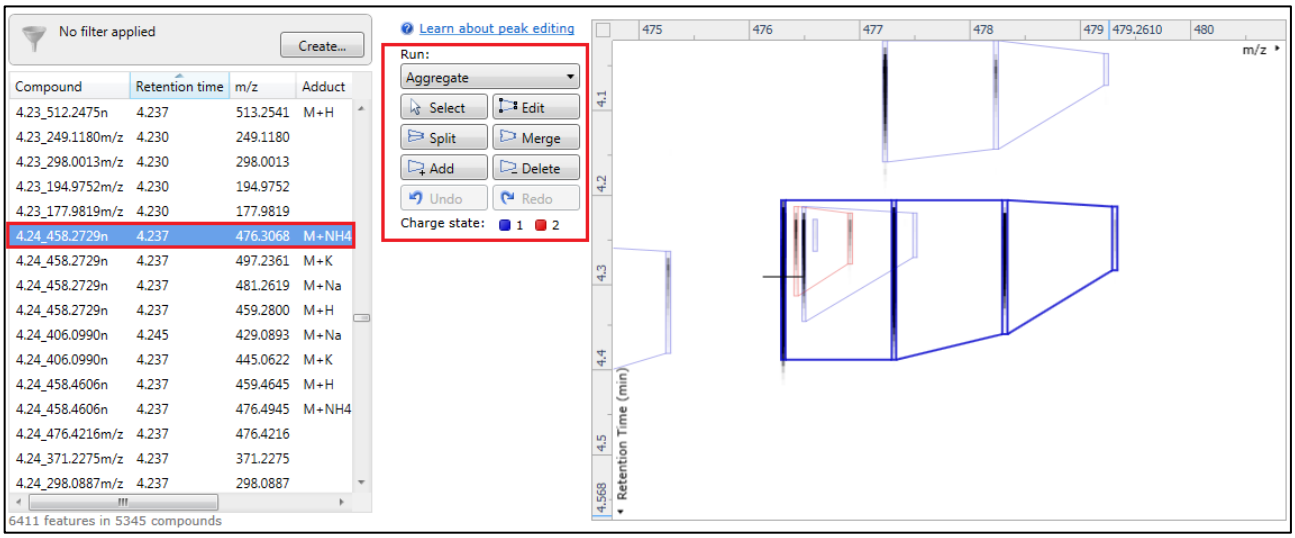
In this example the range viewed in the zoomed widow shows the 4 compound ion adducted forms (M+NH_4, M+K, M+Na and M+H) that contribute to the deconvolved neutral mass for the Compound 4.24_458.2729n in the table. Hover the cursor over each ion to show m/z and Retention Time.

Editing Peak Picking

The editing tools that are present at Peak Picking will allow you to Edit, Split, Merge, Add and Delete (**Note**: Undo and Redo are also available).

These tools can be applied to a selected Compound Ion or in the case of the Add tool it can be applied to an undetected Isotopic pattern.

For the purposes of understanding how the tools work we will focus on one isotopic pattern, the M+NH_4 for the Compound 4.24_458.2729n. Order on **Compound** in the table and scroll to m/z and RT..

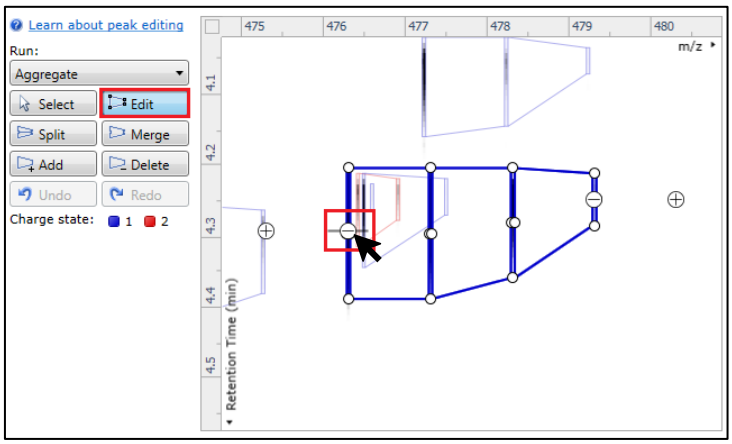


Select the compound ion to edit either by clicking on it in the table or inside the isotopic pattern on the ion map, in each case the selected ion will become centered in the view.

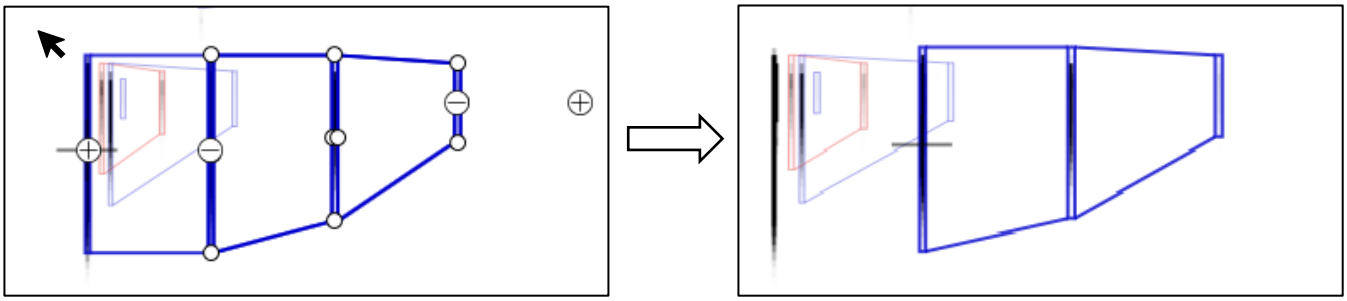
Edit tool

To remove a peak from an existing compound ion select the **Edit** tool and click on the compound ion to reveal the 'edit handles'. You can zoom in more by dragging an area around the peptide ion of interest.

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

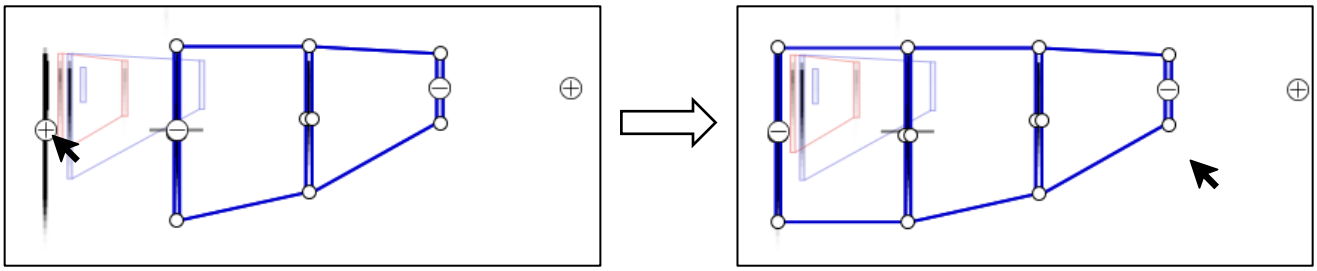


2. Click outside the boundary of the compound ion to update the view.



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

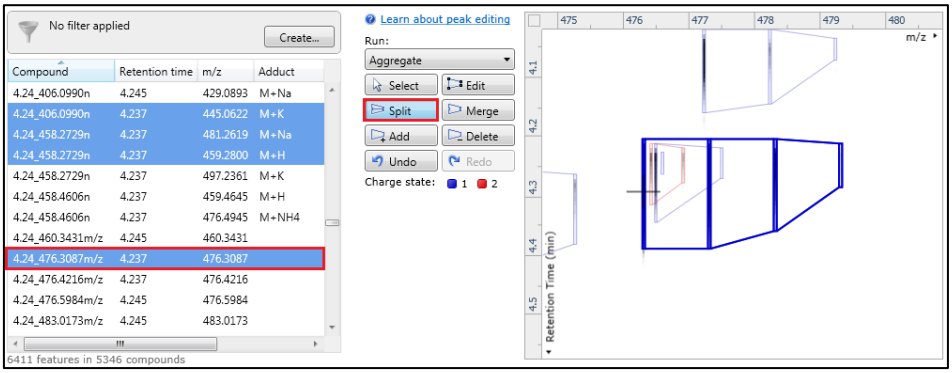
1. Click on the 'plus' handle on the peak to the left (under the plus handle) to add it



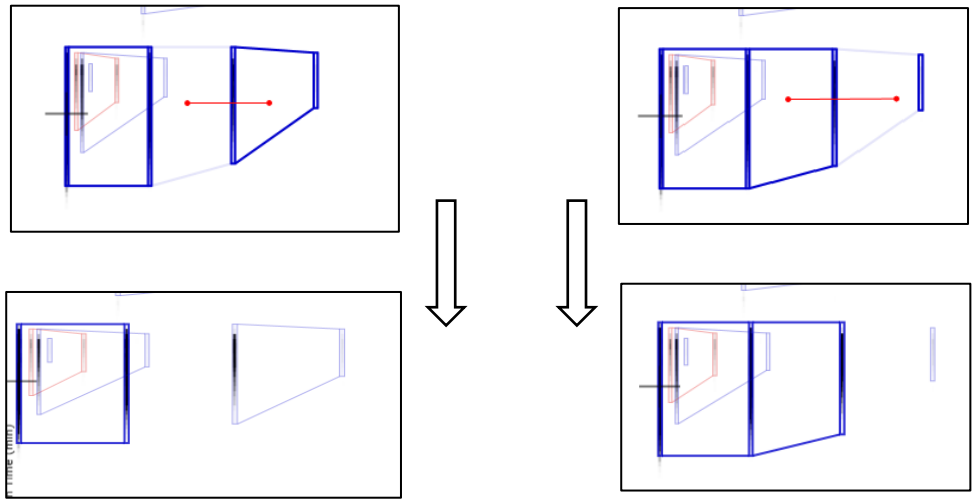
2. Then click outside the peptide ion to update the view.
3. **Note**: If you are not satisfied with the editing use the **Undo** button and retry.
4. **Note**: The **Undo** and **Redo** buttons will no longer be active if you leave and return to Peak Picking.

Split and Merge Tools

Using the edited compound ion, M+NH_4 adduct, for the Compound 4.24_458.2729n which, due to editing is now displayed as a single compound ion at 4.24_476.3087m/z.

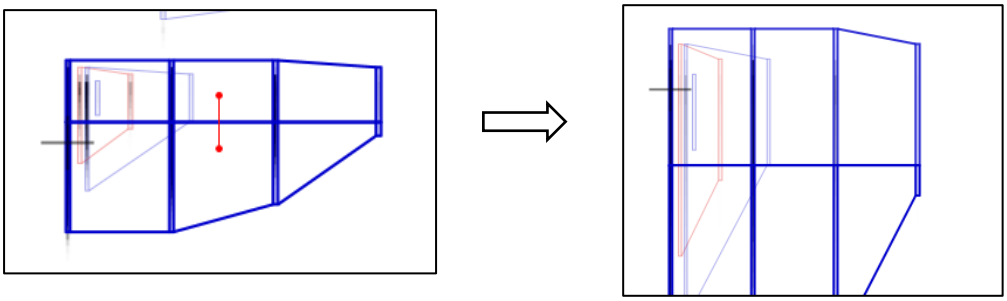


1. Click on the **Split** tool
2. To split the vertically into 2 Charge state 1 compound ions, **click and release** in the centre of the ion pattern. A red dot appears, drag the cursor to right, a line extends across the isotope envelopes. Extending this line will alter how the pattern splits.



3. To complete the split click again, the pattern will split depending on the length of the line
4. **Note**: Using the Undo tool will allow you to retry the splitting of the pattern

5. **Note**: The **Undo** and **Redo** buttons will no longer be active if you leave and return to Peak Picking
6. To split the horizontally into 2 Charge state 1 compound ions, **click and release** in the centre of the ion pattern. A red dot appears, drag the mouse up or down, a blue line extends across the isotope envelopes showing where the split will occur. Extending the red line (up and down) will alter where the pattern splits.

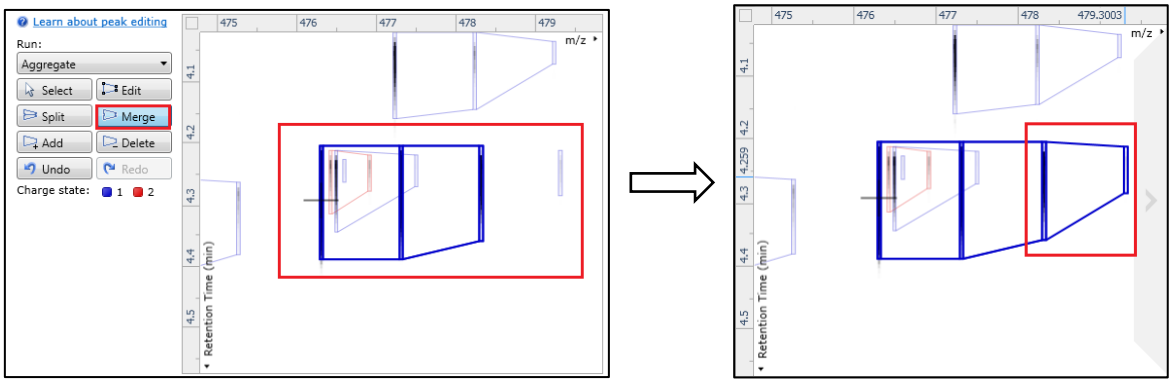


7. To complete the split click again, the pattern will split depending on the length of the red line.

Merge Tool

The merge tool can be used to combine 2 Isotopic patterns that were originally detected as separate compound ions as an example of this functionality try merging the split patterns generated above.

8. Select **Merge** tool and click on the first isotope overlay
9. Drag the cursor to the location of the second isotope overlay, if it is the correct distance for a Charge state 1 compound ion then the second overlay will appear joined to the first

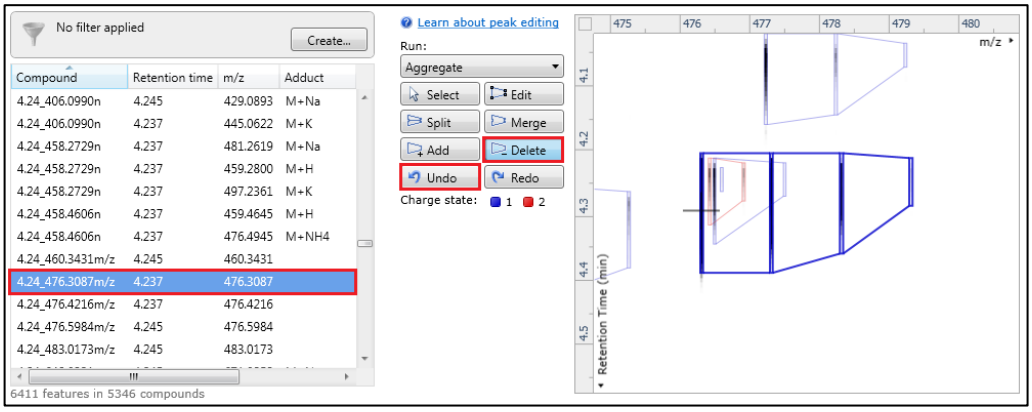


10. To complete the merge click the cursor again over the second isotope pattern
11. Before exploring the **Add** and **Delete** tools, click **Undo** to restore the original isotope pattern.

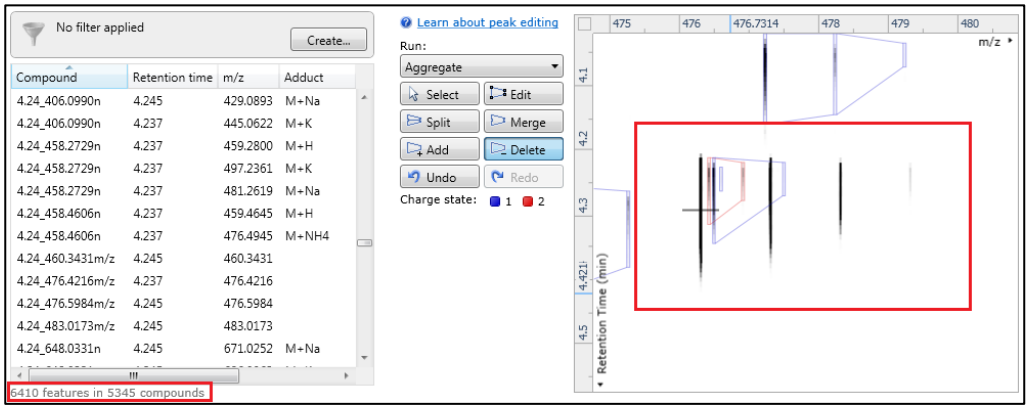
Add and Delete Tools

First use the Delete tool to remove the edited compound ion.

1. Select **Delete** and click on the compound ion in the ion map view



2. The blue, charge state 1, isotope pattern overlay will now be removed from the view.

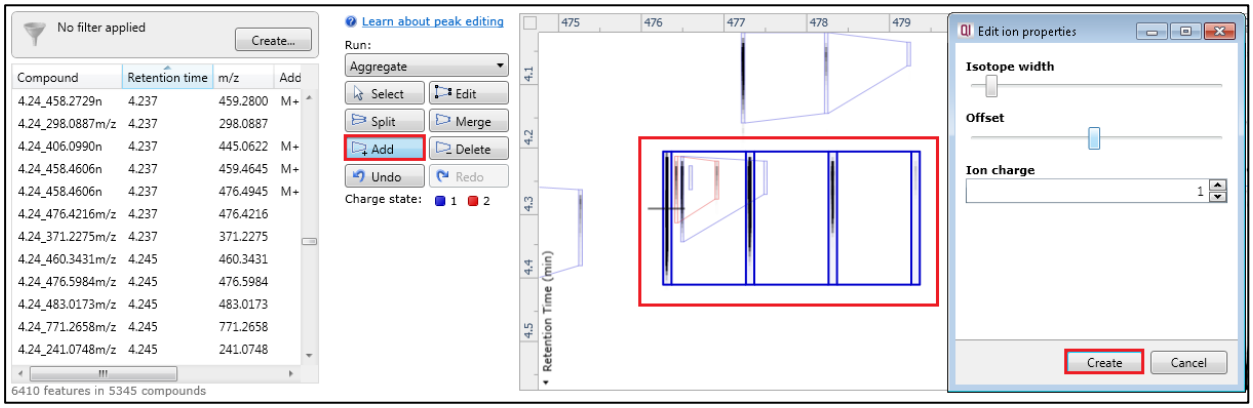


3. The isotope intensities will remain on the ion map, appearing as an ‘undetected’ isotopic pattern. The Compound entry in the table will also be deleted.

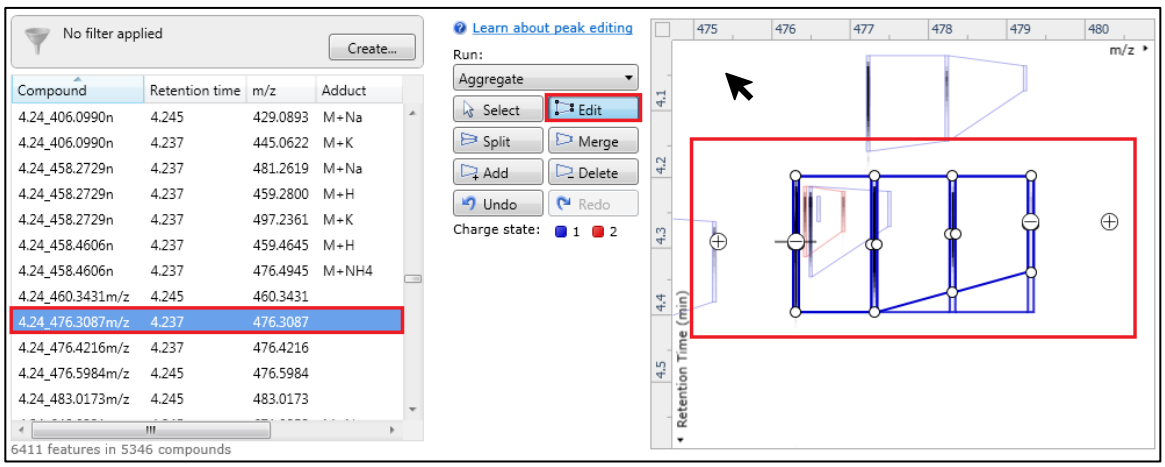
4. These can be restored using the **Undo** button, however, try using the **Add** button to create the peak picking overlay in the ion map window.

Click **Add** and then click on the top left of the first isotope and drag out an area covering the 4 isotope intensities.

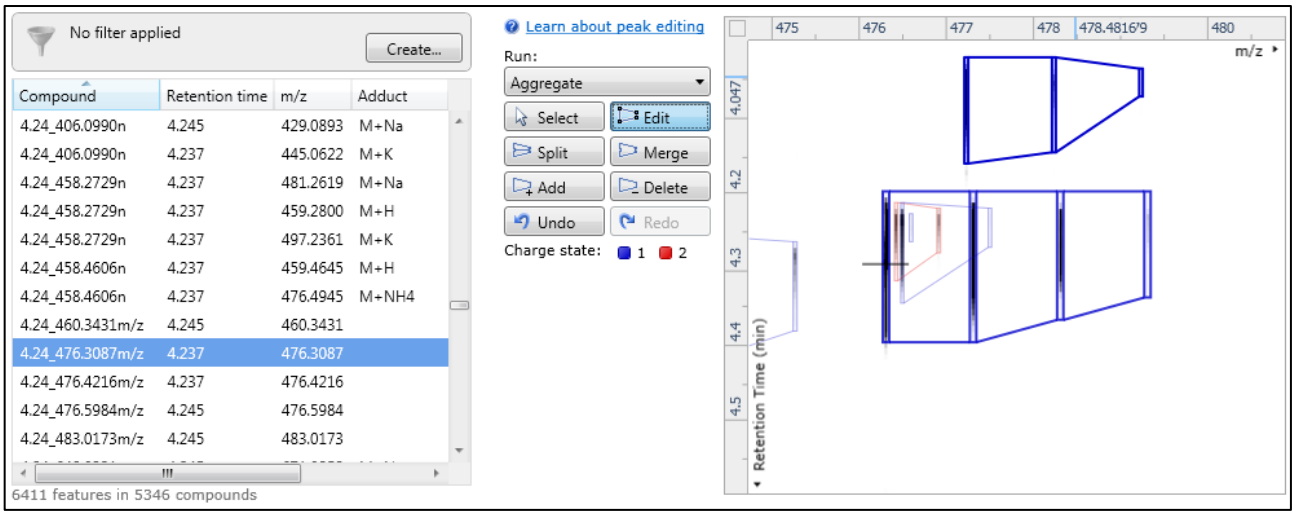
1. As you drag the overlay out it will change colour reflecting the underlying pattern of intensity



2. Use the property tools, (that appear as you release the cursor), to adjust the Isotope width and off set of the overlay on the underlying isotopic intensity pattern
3. Click **Create** to generate the new overlay and a new entry in the table
4. Click on the **Edit** tool and use it to refine the shape of the peak picking overlay by using the ‘handles provided’

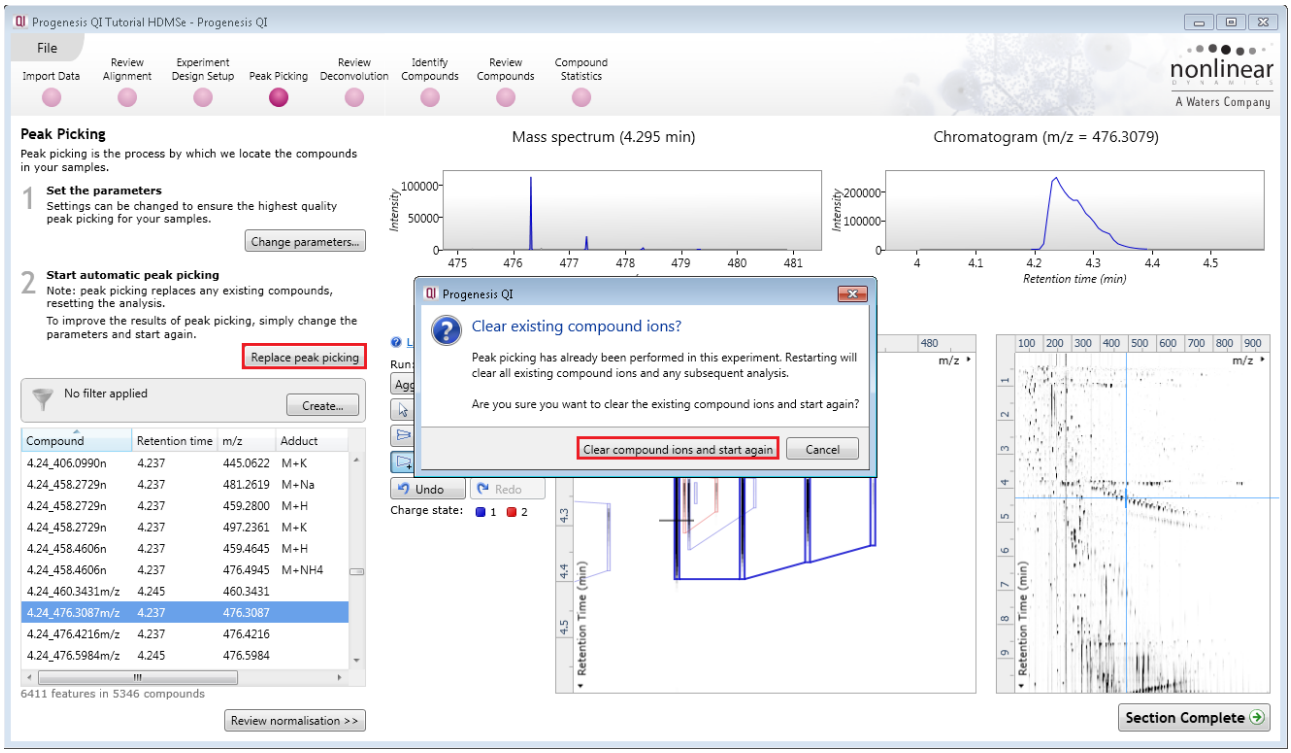


5. Finally, to set the final overlay click on an area outside the overlay on the ion map

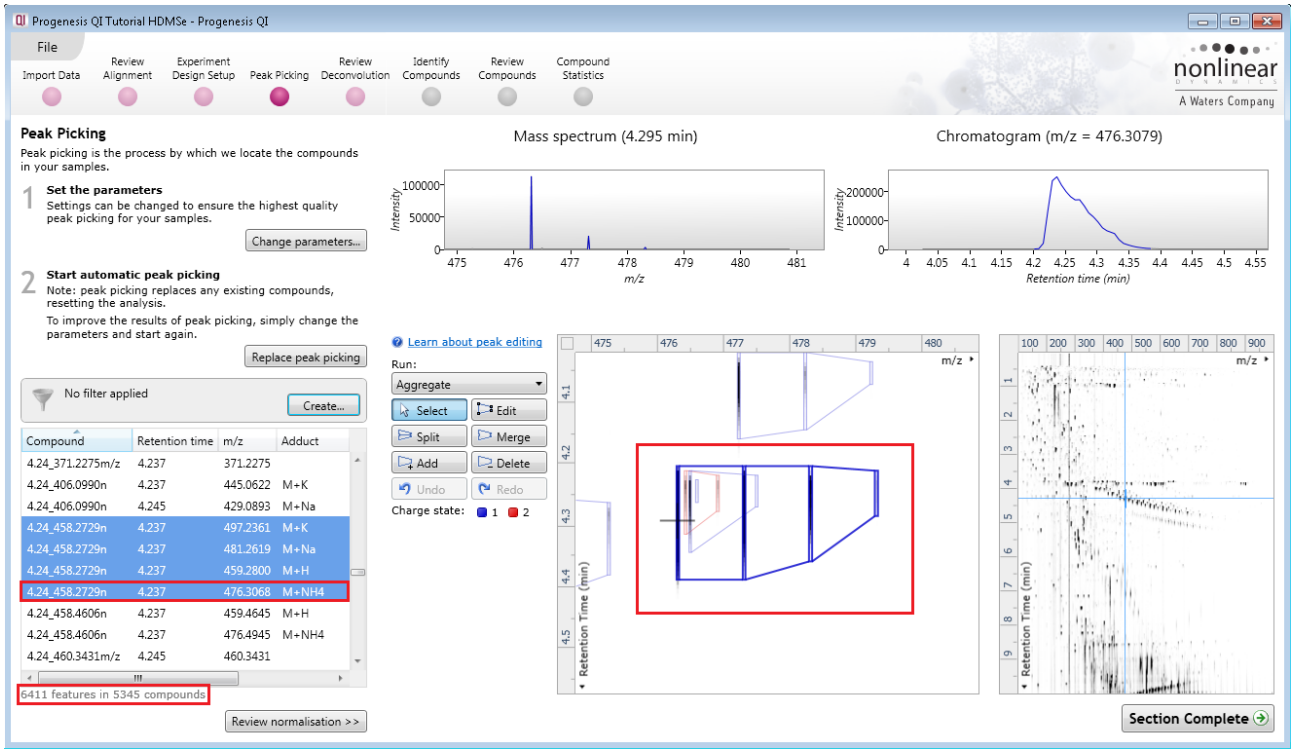


Before looking at the Normalisation if you have been using the Peak Picking Editing tools to explore their functionality you should redo the Peak Pick picking by clicking on **Replace Peak Picking**

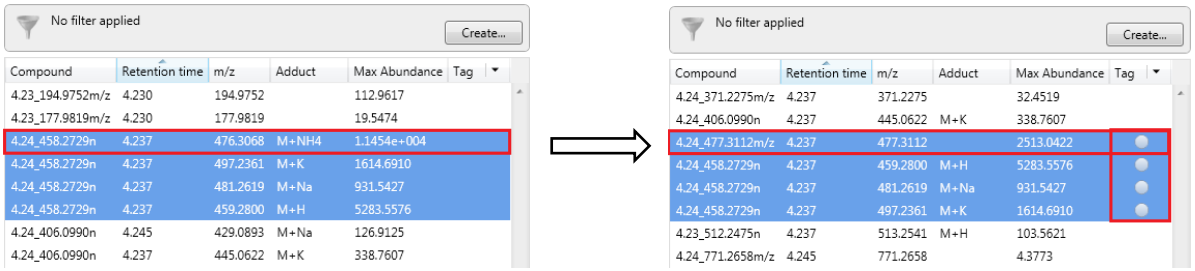
Click on **clear compound ions and start again**



When the Peak Picking has completed, the table will update the number of Compound ions (features) and Compounds reported to 6411 and 5345



Note: When you use the editing tools, the ions they are applied to automatically become tagged in the table (with a grey tag) indicating that the ion has been edited in some way.



Note: the editing of a compound adduct results in all the compound adducts for the deconvoluted compound being assigned an edit tag. In addition the edited compound ($M+NH_4$) ion loses its Adduct assignment.

At this stage all editing tags are removed when you replace the Peak Picking.

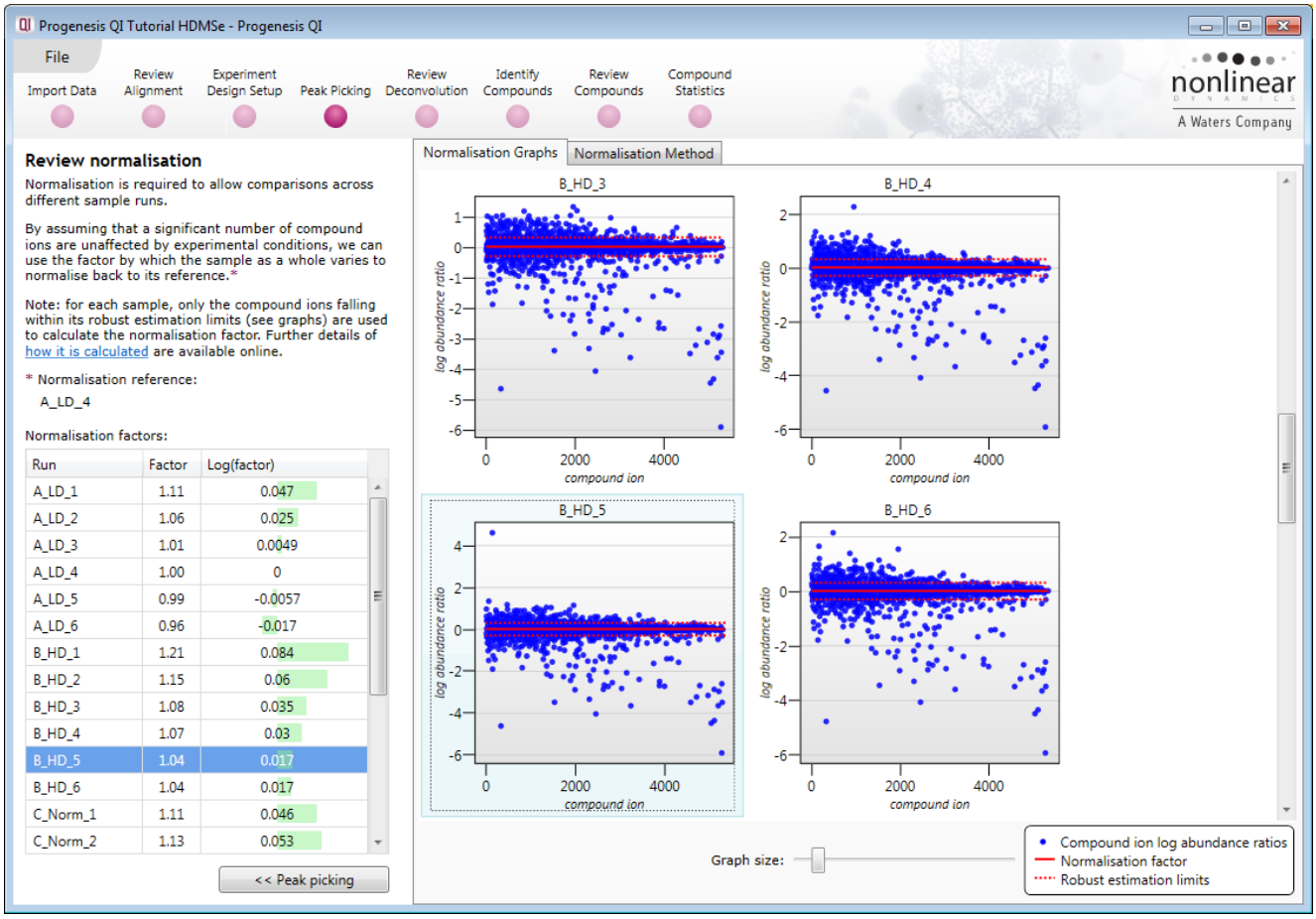
Normalisation of the data can be reviewed by clicking on **Review normalisation.**

Stage 6B: Reviewing Normalisation

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

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

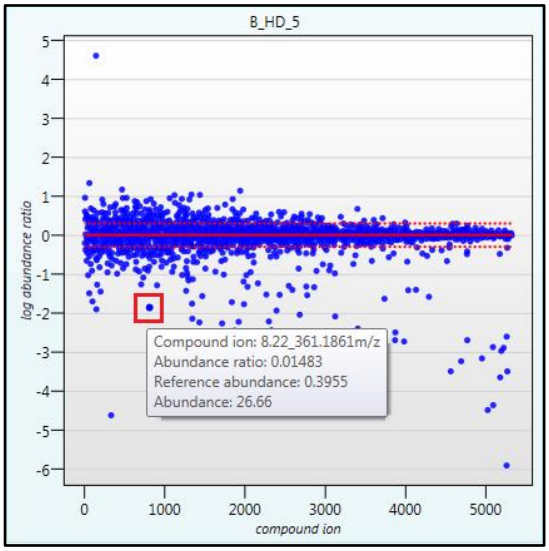
The default method is to **Normalise to all compounds**.



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

Calculation of Normalisation Factor:

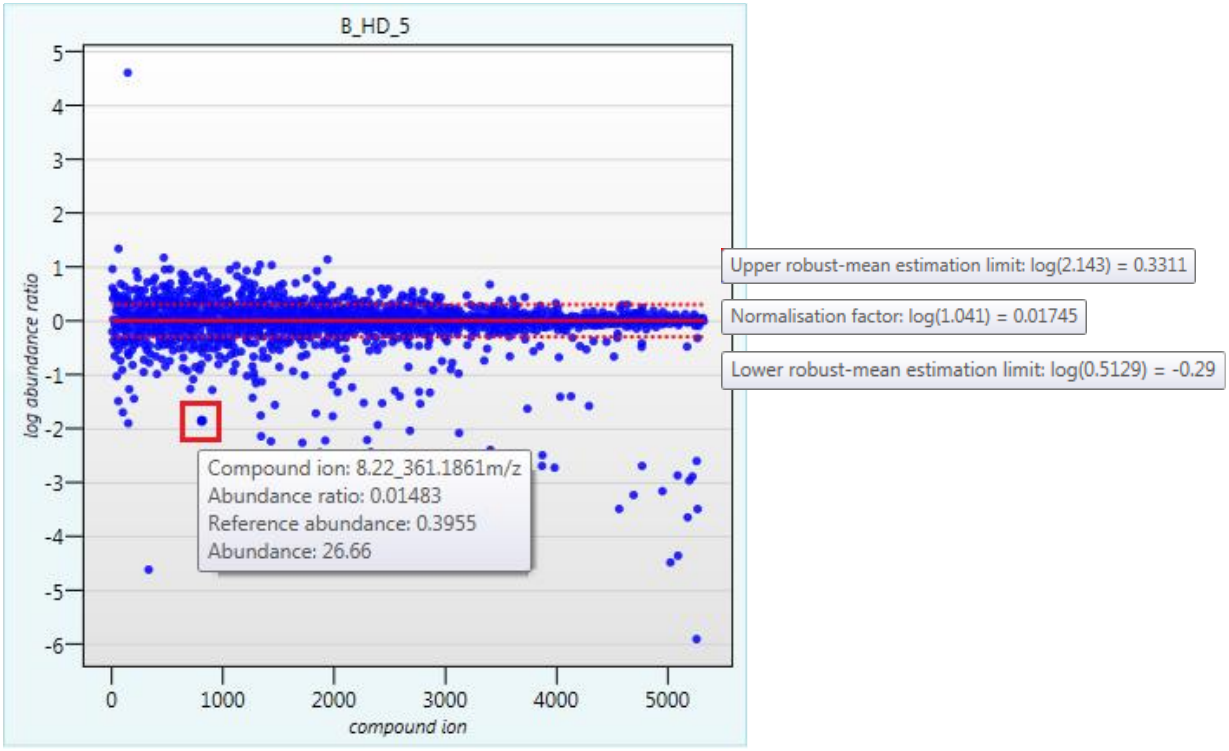
Progenesis QI will automatically select one of the runs that is 'least different' from all the other runs in the data set to be the 'Normalising reference'. The run used is shown above the table of Normalisation factors (in this example it is A_LD_4).



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

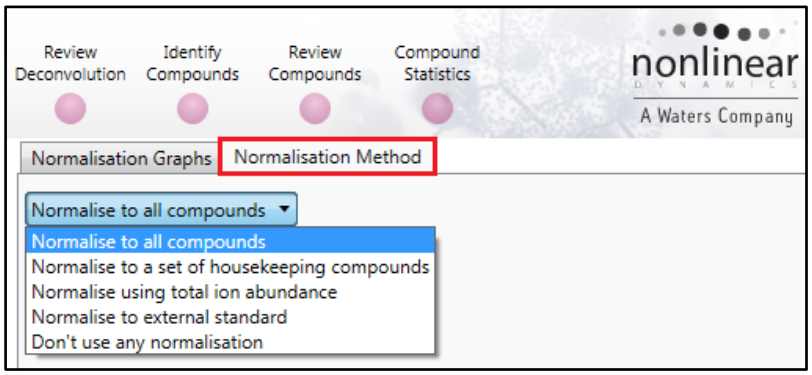
The details for individual Compound ions can be viewed as you hold the cursor over the dots on the plot. On the graph the compounds are shown ordered by ascending mean abundance.

The normalisation factor is then calculated by finding the mean of the log abundance ratios of the compounds that fall within the 'robust estimated limits' (dotted red lines). Compounds outside these limits are considered to be outliers and therefore will not affect the calculation of the normalisation factor.



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

Note: there are 5 Normalisation modes



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

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

Now return to **Peak picking** by clicking on the button on the bottom left of the screen and the press **Section Complete** to move to the Review Deconvolution stage of the workflow.

Stage 7: Review Deconvolution

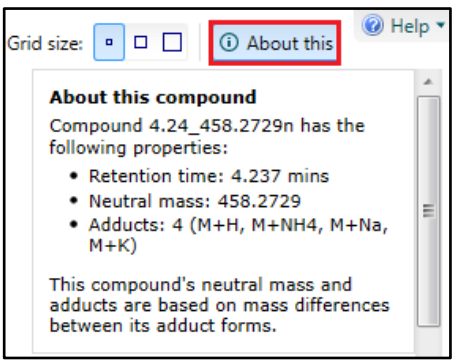
Following Peak Picking all the ions, including various adducts, for a compound are automatically recombined to provide accurate quantitation of each parent compound. Adducts of the same compound are grouped by a process called deconvolution. This stage of the workflow allows you not only to review the outcome of the deconvolution but also to add or remove adduct forms of the compound.

At this stage in the workflow you get the opportunity to review the process of Deconvolution for the Compounds. The table on the left displays the compounds ordered by the number of Adducts detected for each compound.

Note: the detected adducts are dependant on the list of expected adducts that you selected when you created the experiment

The total number of expected adducts is displayed as a montage at the top of the screen.

For an example: In **Compound 4.24_458.2729n**, 4 adducts have been detected: M+H, M+NH_4, M+Na and M+K (arrows) and displayed in the table. To find this compound easily, one may order the table by compound using the header or scroll down slightly on the Compounds table.

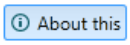


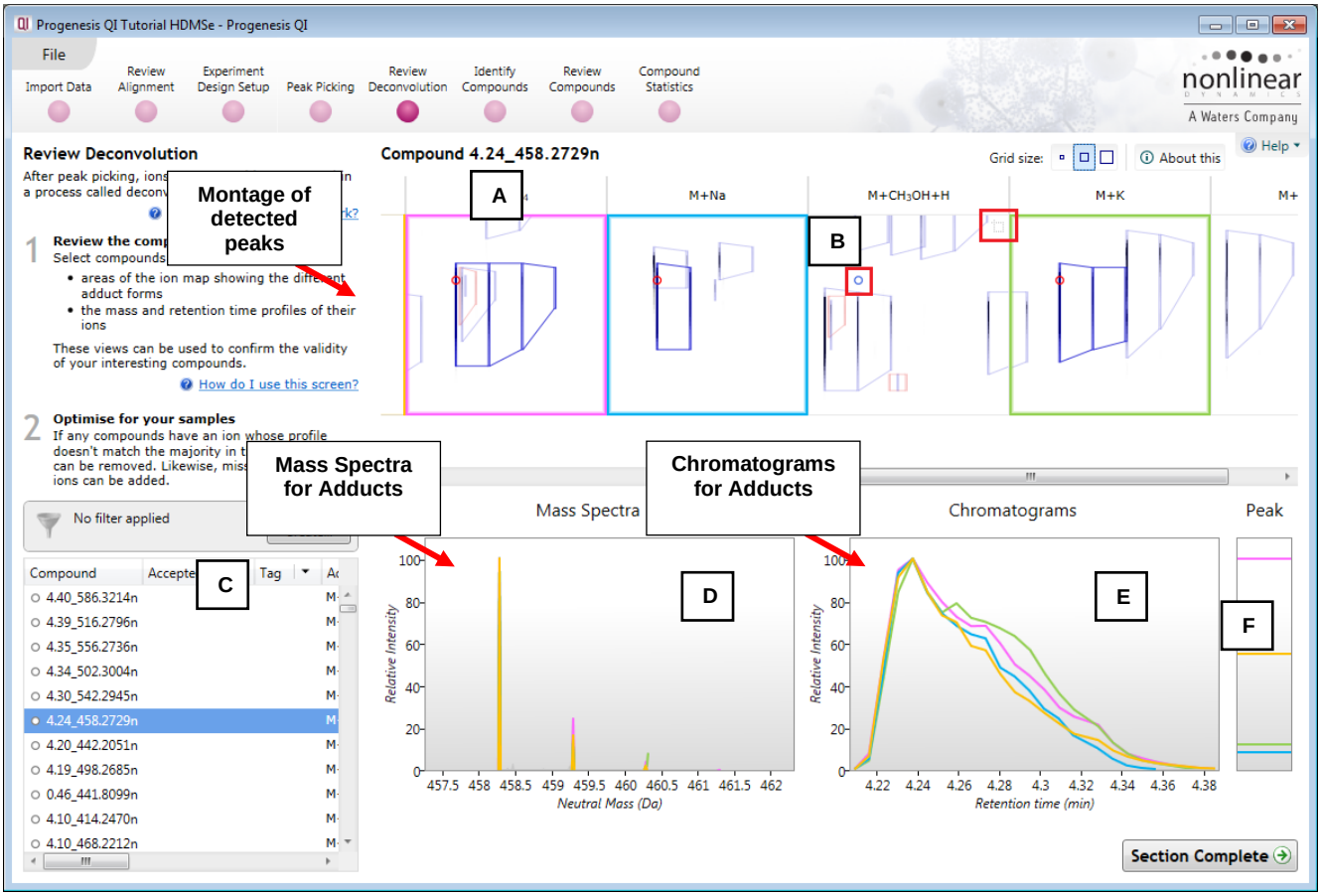
Tip: the '**About this**' panel, top right summarises the information for the current compound. If not displayed click on the **About this**.

The same details are available for the compound in the table to the bottom left.

Use the Grid size to enlarge the montage view of the adducts.

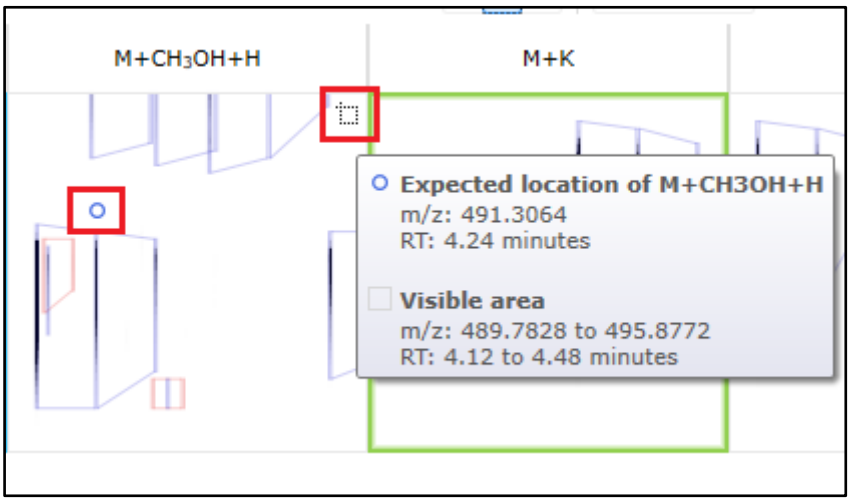
Using **Compound 4.24_458.2729n** we can look at how the information is displayed on the multiple views following deconvolution.

Note: for clarity the 'About this' pane can be hidden by clicking on 



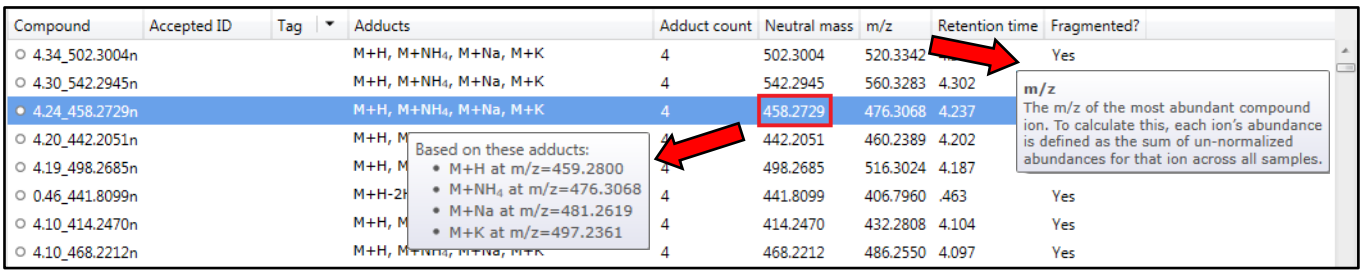
(A) Montage view showing location of detected adducts for compound

(B) Where an adduct has not been detected the expected location for the adduct is displayed on the relevant panel as a blue circle when you hover the cursor over the panel. In this case no Methanol adduct (M+CH_3OH + H) has been detected.

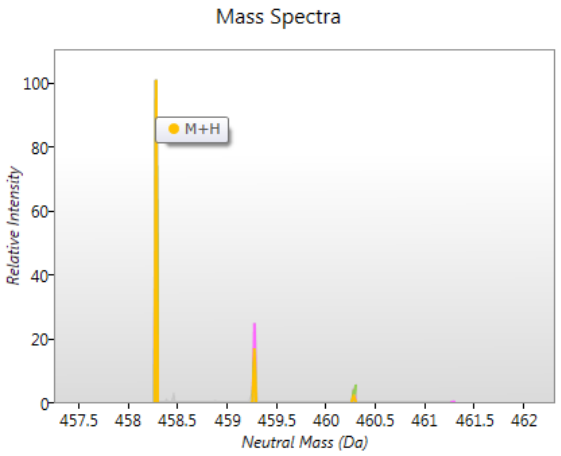


Note: hover the cursor over the icon on the top right of the panel and you will get information about the expected location and the area displayed.

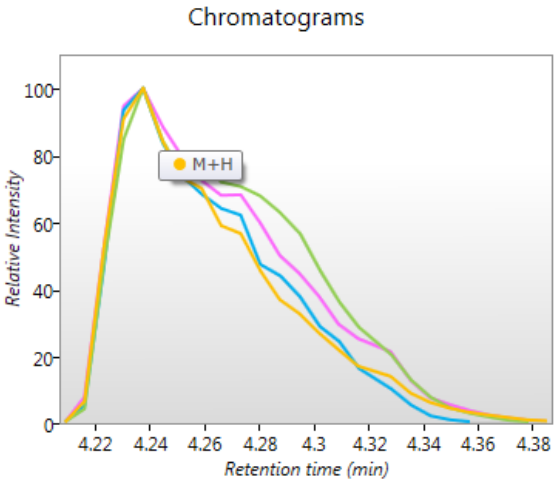
(C) The table displays the information known about the current compound with regards to the number of adducts, identity of the adducts and where deconvolution has been successful the Compound's Neutral mass is displayed. Note: when you hover the cursor over the **Neutral mass** for the current compound it displays the adducts and their m/z values that were used to generate the the value. This is shown in an expanded view of the table below. Also shown is the 'tool tip' for the displayed m/z for each compound. An example is shown below for **Compound 4.24_458.2729n**.



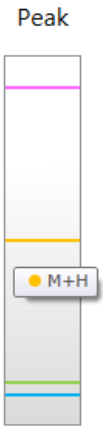
(D) This panel shows the Mass Spectra for the detected adducts of the Compound, colour coded by adduct, where the relative intensities are plotted against the **Neutral mass** scale.



(E) This panel shows the Chromatograms for the detected adducts of the Compound, colour coded by adduct, where the relative intensities are plotted against the **Retention Time** scale.



(F) This panel shows the actual Peak heights for the detected adducts of the Compound, colour coded by adduct.

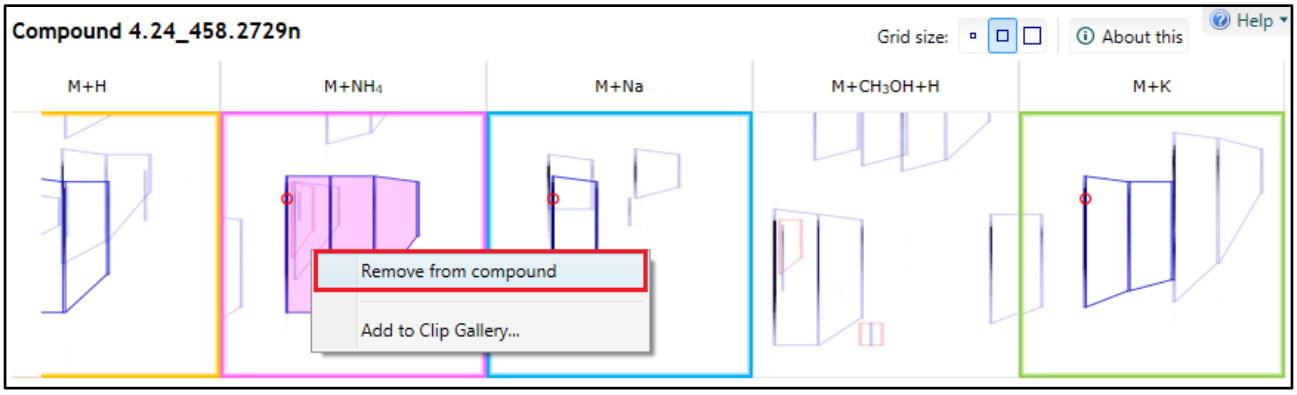


Editing Compound adducts.

The number of adducts assigned to a compound by the process of deconvolution depends on the peak detection and the adducts selected at the beginning of the workflow when the experiment is created.

The process of deconvolution can only assign a Neutral mass for a compound if two or more adducts have been detected. The accurate quantitation of a compound is dependent on summing the intensities for a compounds adducts. In addition the accuracy of the quantitation for any compound is dependant not only on the detection of the adducts but also on the correct assignment during the process of deconvolution. In a complex sample there may be a need to add or remove adducts from a compound.

To remove an adduct right click on the assigned peak in the appropriate panel of the montage. So as an example we require to remove the M+NH_4 adduct for compound **Compound 4.24_458.2729n**, right click on the peak in the adduct panel and click **Remove from compound**.



In the table the number of adducts for the compound is reduced by 1 and a tag (grey) indicating that the adducts of this compound have been edited is created.

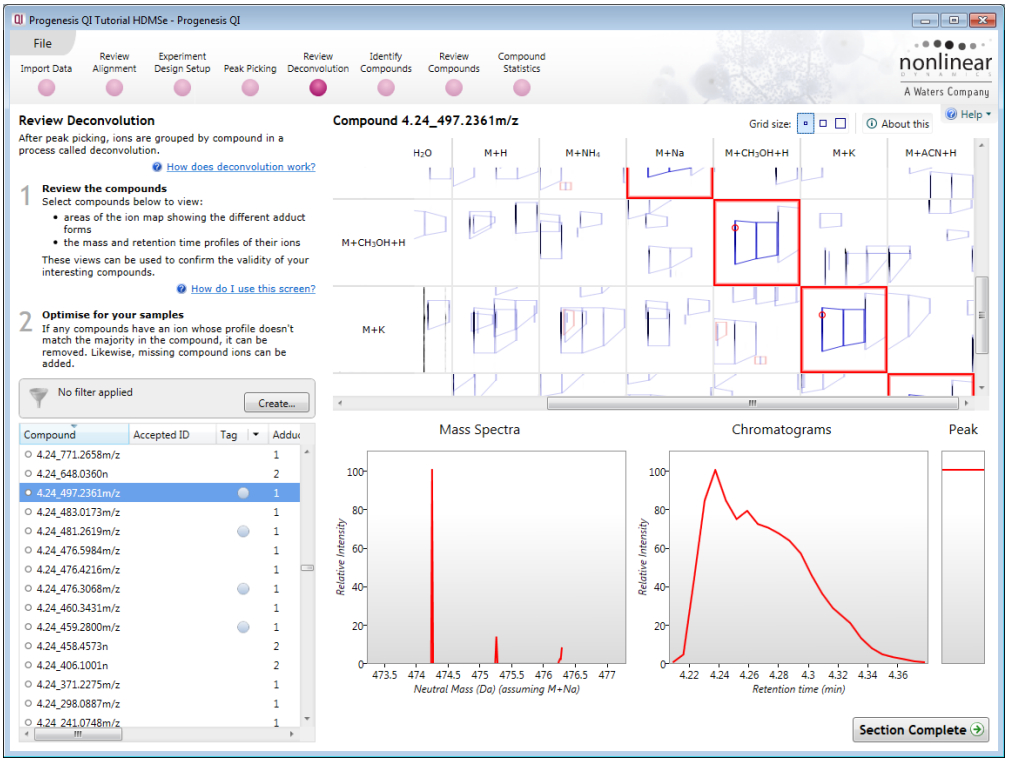
To add an adduct to a compound right click on a peak in the appropriate panel for the adduct and click **Add to compound,** again the table of compounds will update to reflect the change.

If you now remove 2 of the remaining 3 adducts the montage view will change to display the full matrix of all possible peak locations for expected second adducts dependant on the m/z values for these adducts. Also as there is only one adduct detected the Neutral mass for the compound can no longer be calculated.

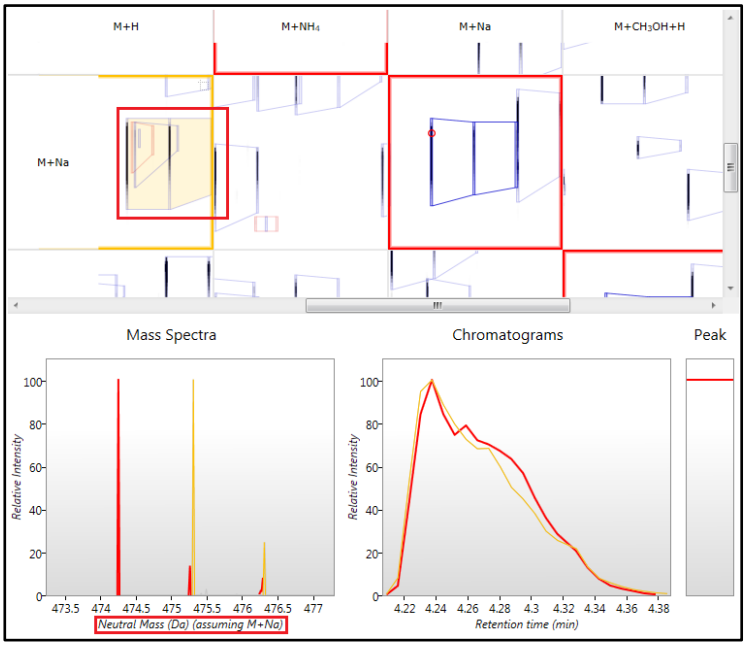
Try removing all the adducts from **Compound 4.24_458.2729n** to leave just the M+K.

The montage now turns into a matrix displaying all the possible locations for a charge state +1 Adduct in accordance with the list of expected adducts you selected at the beginning of the experiment. Also the individual Compound ions are displayed in the table i.e. **Compound 4.24_497.2361**

Note: because the peak is a charge state +1 this excludes M+2H and M+H+Na as being the compound's sole adduct on the basis of charge.

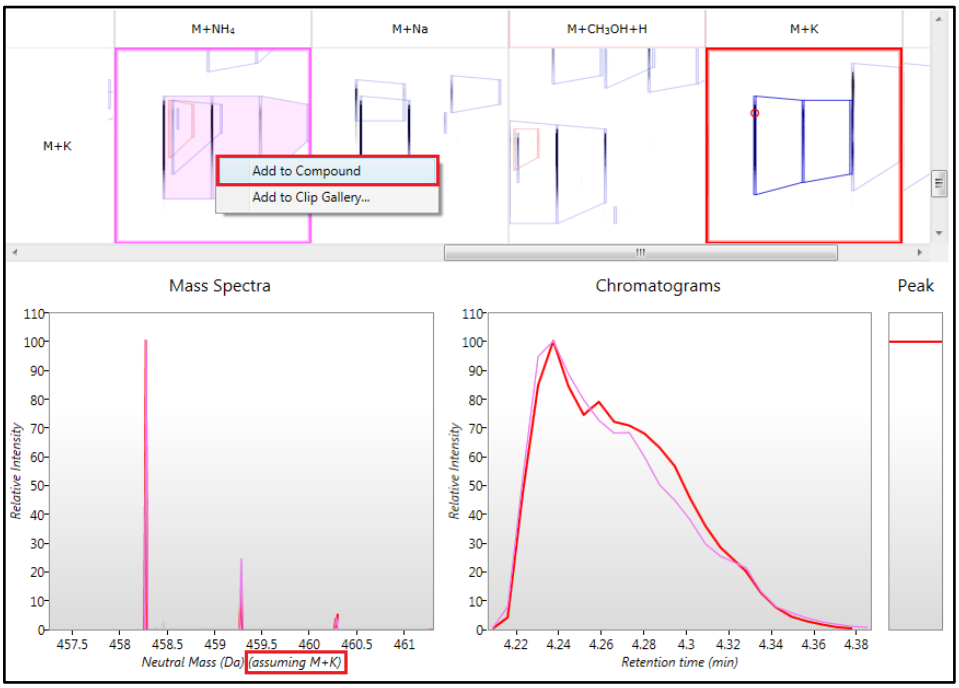


As an example to explore the possibility that **Compound 4.24_497.2361's** sole adduct is M+Na try hovering over the detected peaks for the expected adducts. In this case when you hover over M+H as the possible second adduct you can see that although the chromatograms elute at the same retention time the mass



spectra do not coincide. On this basis the second compound adduct cannot be M+H when M+Na is assumed to be the sole detected adduct.

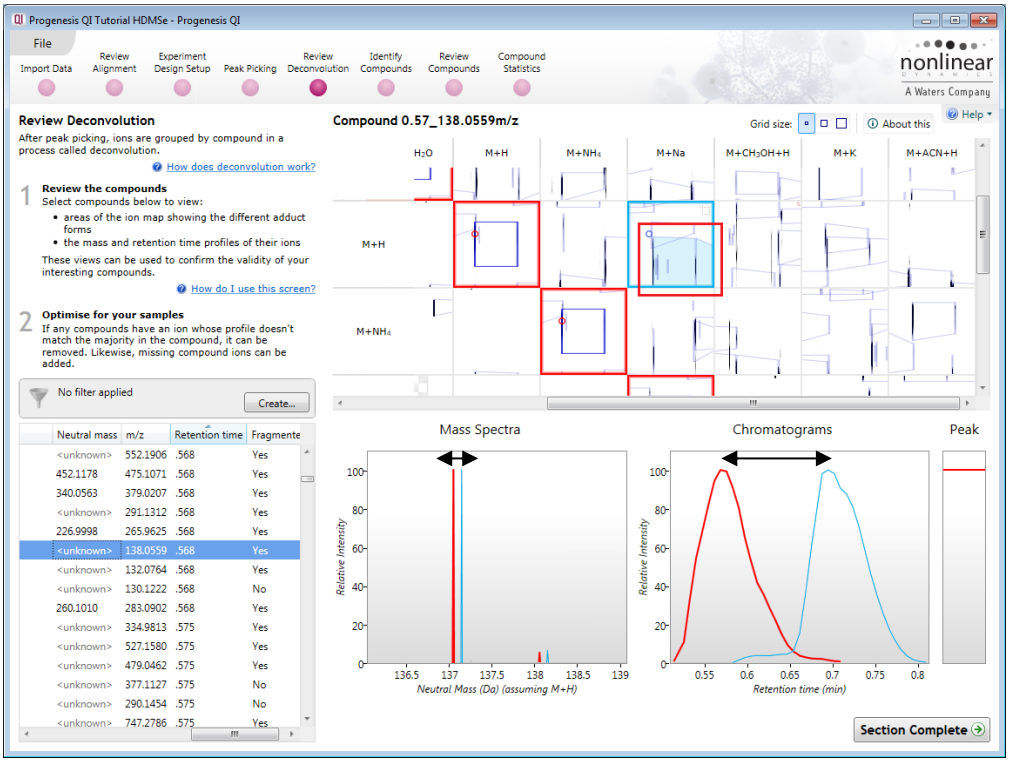
Now try M+K as the assummed sole adduct of **Compound 4.24_497.2361**. This time when you hover over the cursor over the M+NH_4 adduct both the m/z on the mass spectra and RT on the chromatogram appear to coincide on the graphs with the corresponding location for the existing single adduct M+K.



To add the M+NH_4 adduct to the compound right click then add this second adduct to the compound. As you add the second adduct the table will update to indicate that the compound has 2 adducts, a **Neutral mass** is calculated (n appears after the compound in the table) and the matrix is replaced by a single row of adducts. This can be explored for the presence of additional adducts, in this case an additional 2 adducts can also be added (M+H and M+Na).

Exploring the expected location for the second adduct of a compound

After Peak picking there will be a list of compounds with a single adduct detected. For these compounds the review deconvolution montage will be displayed as a matrix where the process of deconvolution has been unable to assign a second adduct on the basis of any of the expected adducts displaying the correct m/z and retention time at the expected locations for any of the selected adduct ions.



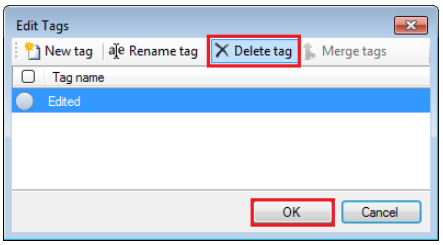
Note: As you hover the cursor over any of the expected locations for the second adduct the Mass spec and Chromatogram for the second adduct are displayed in the bottom panels. If both the m/z and RT appear to coincide (in this example the 'double headed' arrows indicate that neither coincide) with the corresponding location for the existing single adduct then Add this second adduct to the compound by right clicking on it and selecting **Add to compound**.

By adding the second adduct the montage and table will update to reflect the addition of the adduct to the compound and a Neutral mass will be assigned based on the 2 adducts. The matrix collapses to the single row displaying the locations of the two detected adducts.

Using 'Tagging' to filter data for identification

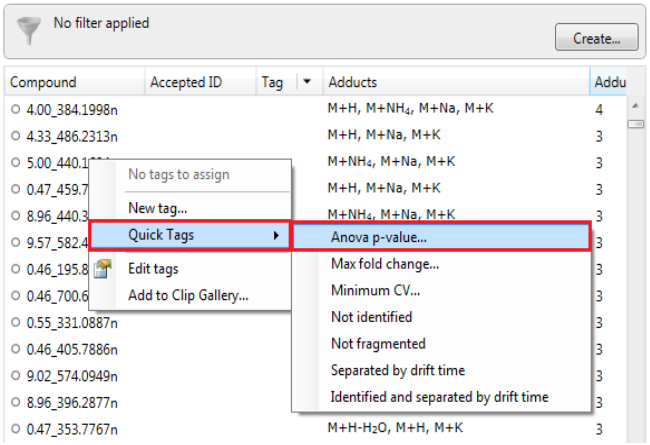
In order to identify the compounds you can focus on data that is showing significant differences between one or more groups by first creating tags for particular 'subsets' of your data. Then filter the displayed data using the tags to provide a targeted list of compounds to identify.

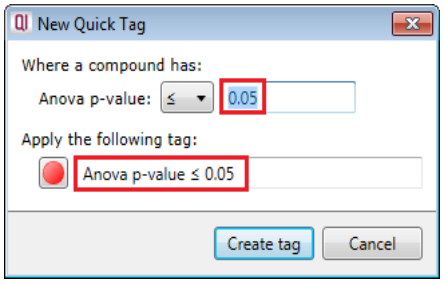
Before creating new tags you can remove the automatically generated 'Edited' you created in the previous section. Right click on the table and select 'Edit tags' click **Delete tag** and then **OK**.



As an example of the use of tags we will set out to identify the compounds that show a significant (Anova p-value < 0.05), 2 fold or greater increase in abundance for one of the conditions (A_LD, B_HD or C_Norm).

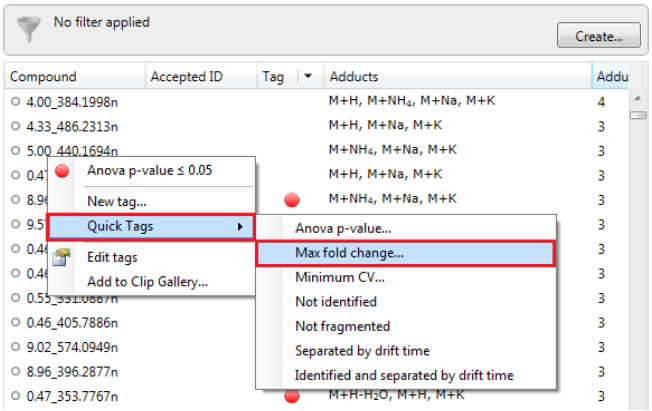
To create a tag for all compounds displaying an Anova p value <0.05, right click on the Compounds table and select Quick tags then select **Anova p-value…**

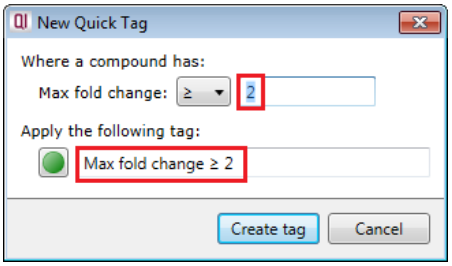




Accept the default value (≤ 0.05) and the offered name.

On pressing **Create tag** a tag appears in the table against all the compounds with an Anova p-value ≤ 0.05

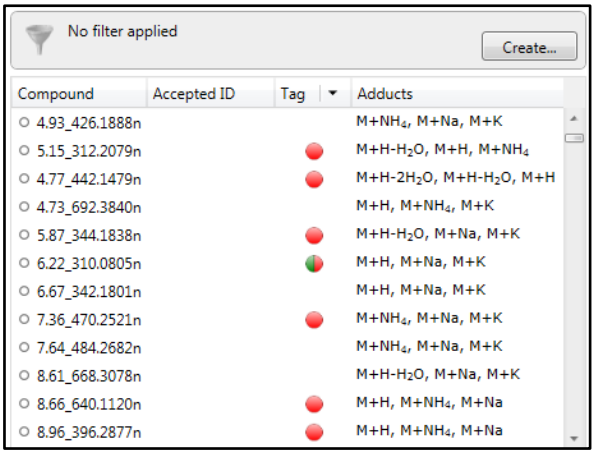




To add a second **'Quick Tag'** for those compounds with a Fold difference of 2 or greater, right click on the table to open the 'Tag' menu. Select **Quick Tags** and then **Max fold change** and accept the default value (≥ 2).

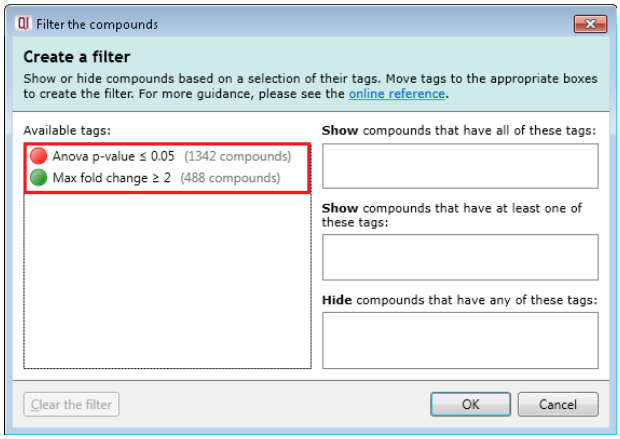
You will be offered a tag named **Max fold change ≥ 2**.

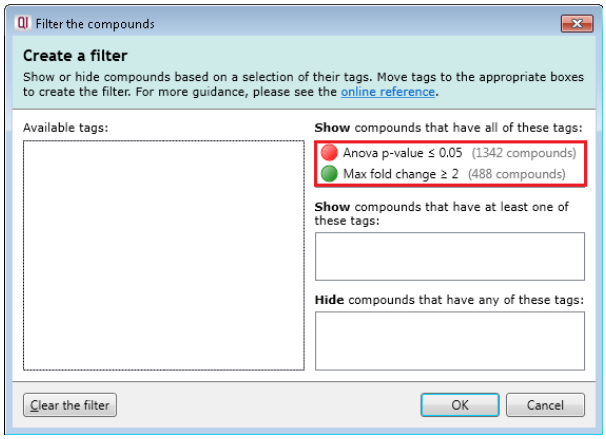
On pressing **Create tag** a second tag appears in the table against all the compounds with a Max fold change ≥ 2.



Now to get the table to display **only** those compounds which satisfy the required criteria, i.e. those compounds that display a significant 2 fold or greater increase in Mean abundance, apply a filter.

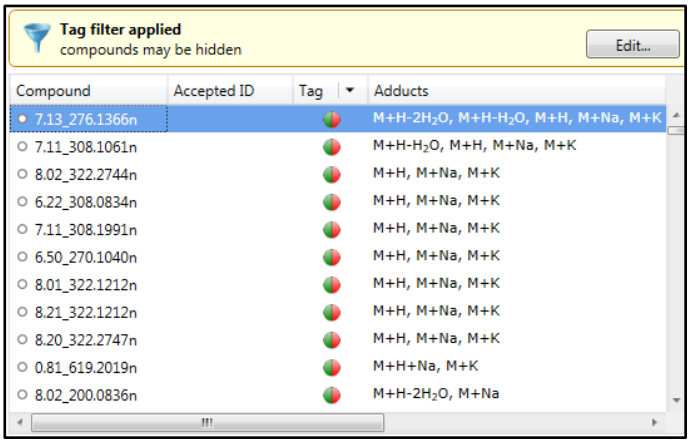
To set up a filter click **Create** to open the Filter dialog and drag the new tags on to the **Show** compounds with all of these tags.





On clicking **OK** the table now displays only those compounds with the 2 tags. This example is ordered by number of adducts (adduct count)

Note: the Tag filter panel has changed, informing you that a filter is currently applied.



With the filter still active click **Section Complete** to move to the Identify Compounds stage.

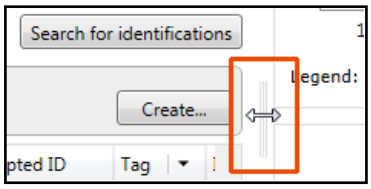
Stage 8: Identify Compounds

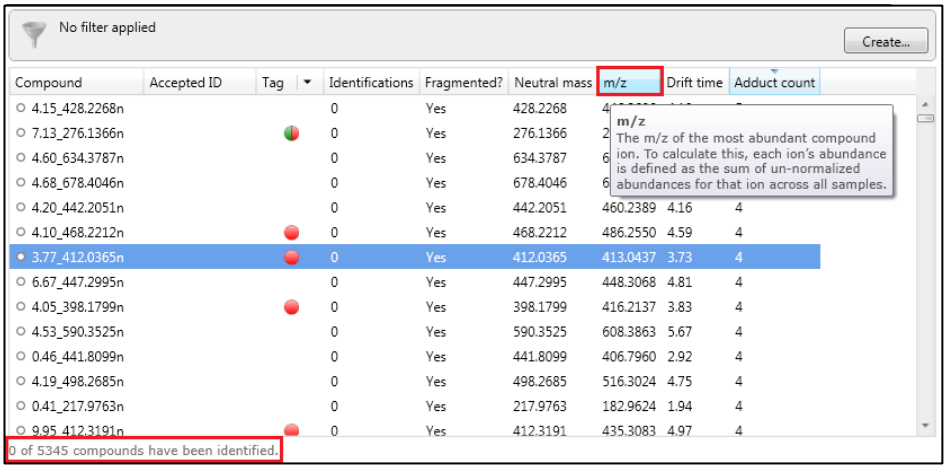
The user guide now describes how to identify compounds from **all or a subset** of compounds.

Note: that the filter you applied at the previous stage is currently active.

To display all the compounds in the table click **Edit** and then click on **Clear the filter** and then **OK.**

To Find **Compound 3.77_412.0365n,** which has a neutral mass of 412.0365, either scroll to the right in the table or expand the table by dragging the icon to the right of the table. Then order the table on descending number of adducts (Adduct count).





At this stage the view displays the pattern of fragmentation for the selected compound

Note: the m/z displayed in the table is for the most abundant compound ion (in this case the M+H adduct).

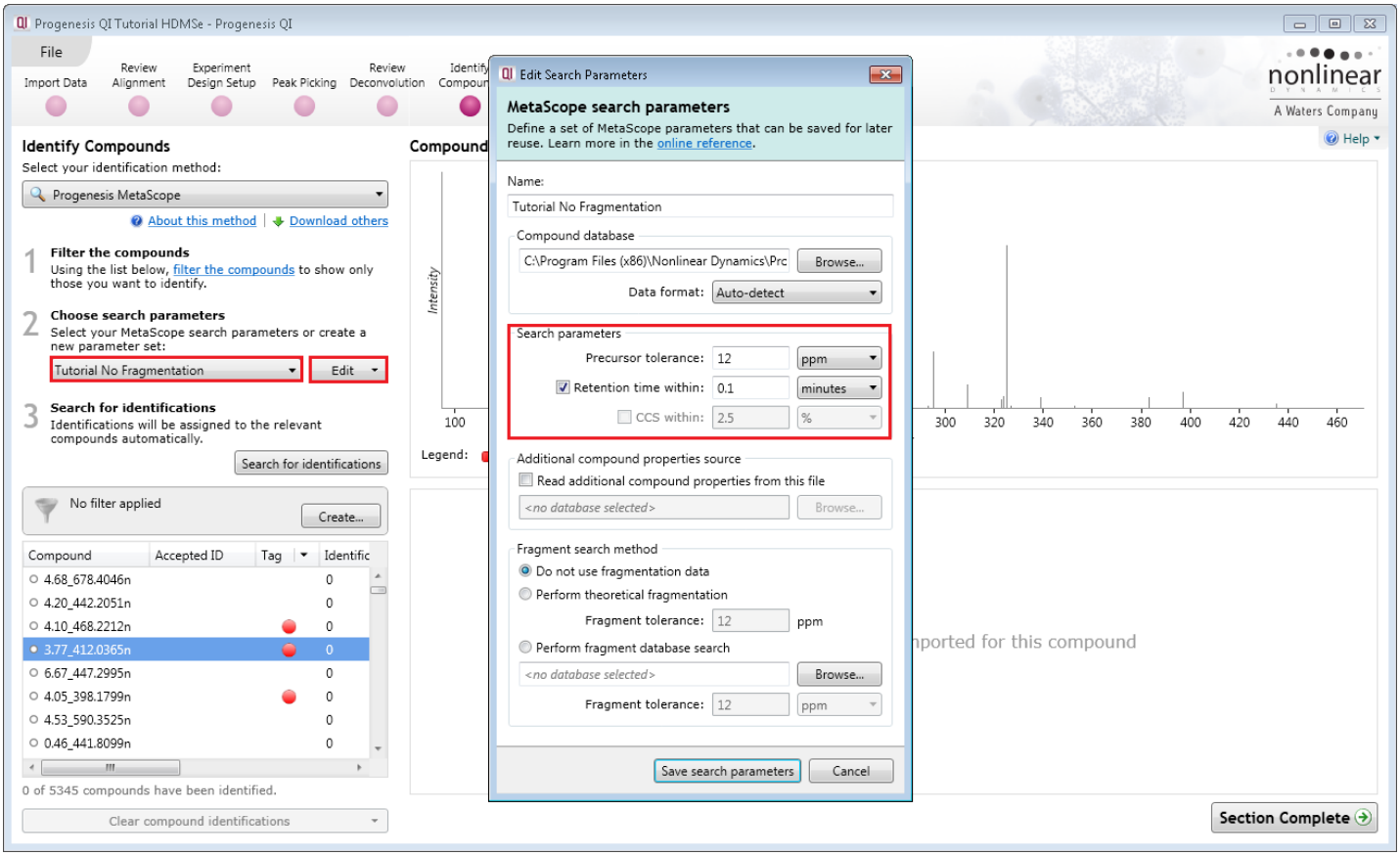
Currently there are 5345 compounds of which none have been identified.

Identification of compounds using the search method Progenesis MetaScope

This identification method is designed to support identifications from a number of different databases.

Depending on the available information in the database and the type of search being performed you can create and save re-usable **search parameter sets**.

For example: searching a database using compound values for Mass and Retention Time.



First select the identification method from the top left of the screen. In this example for the Tutorial data use Progenesis MetaScope, a flexible search engine which is designed to work with databases where you can set thresholds for Mass, Retention time, Collisional Cross Section and Theoretical Fragmentation depending on the format and content of the database.

Set the Precursor tolerance to 12ppm, in this case you can use the **Tutorial No Fragmentation** search parameters as shown above.

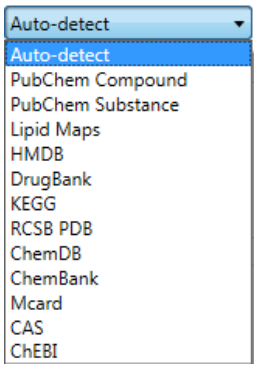
Browse to locate the database file.

Location of example Structure Data Format (SDF) files after install:

C:\Program Files (x86)\Nonlinear Dynamics\Progenesis QI\Plugins\MetaScopeSearch\SDF Databases\

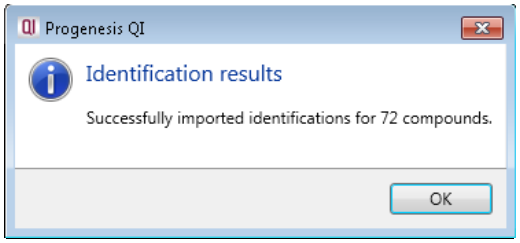
Select: Tutorial Fragmentation.sdf (if not already selected)

There are a number of Data Formats available including **Auto-detect,** which we will use here.



Save the Search Parameters and click **Search for identifications.**

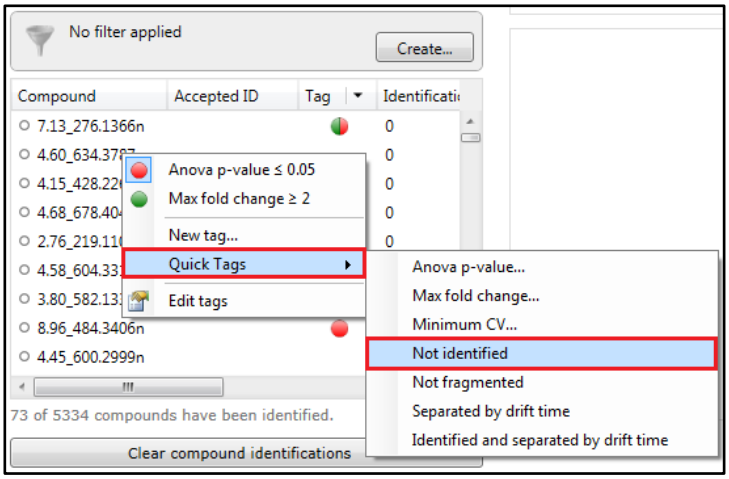
A dialog will tell you the number of compounds that you have imported identifications for.



To focus only on the identified compounds, hide all the unidentified compounds. To do this you must first tag all the unidentified compounds. Right click on the table and select the Quick Tag for **Not identified.**

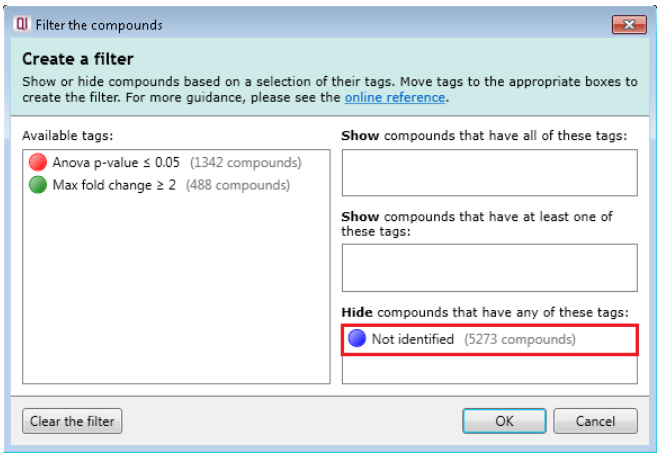
Then create a filter that hides these compounds by selecting Create and dragging the **Not identified** tag on to the **Hide** panel.

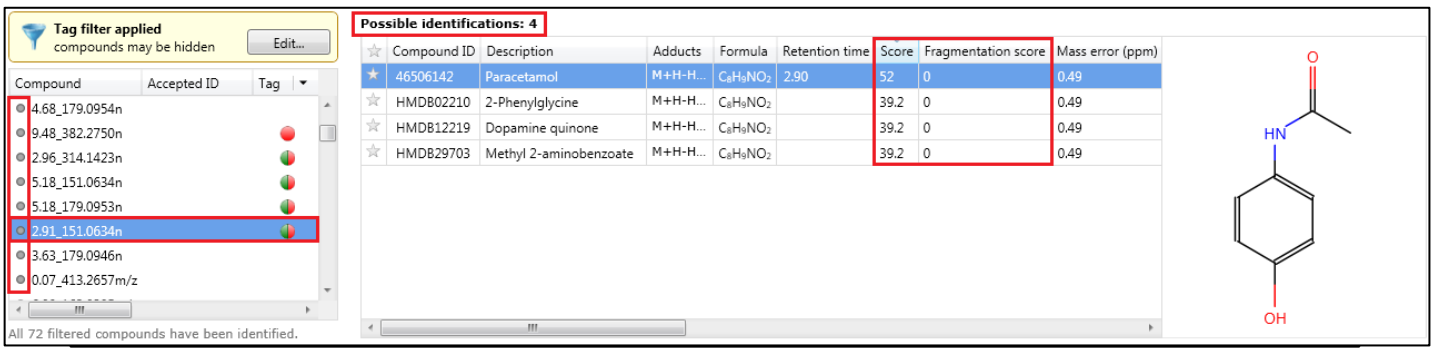
Click **OK** to apply the filter.



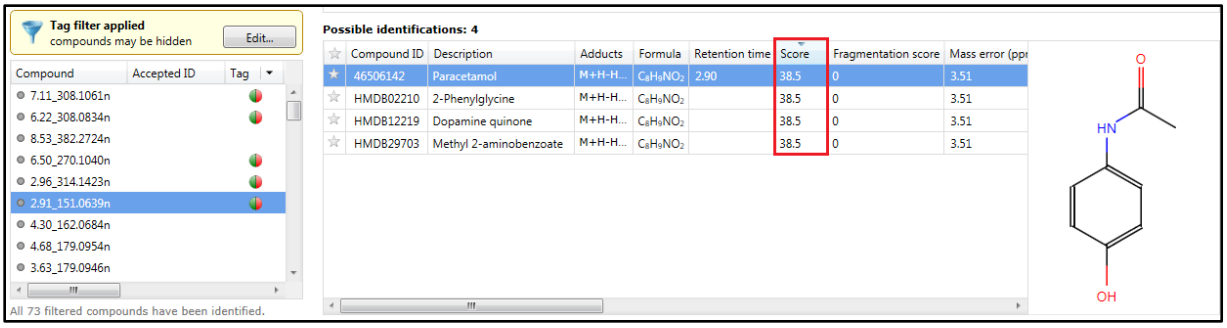
Having imported the search results you can see there are a number of possible identifications for the identified compounds.

Possible identifications for the current compound are ranked on an overall **Score** based on, **Mass error**, **Retention Time error** (if available), **Isotope Similarity** (calculated from the comparison of the measured isotope distribution for the compound vs the theoretical based on the compound formula, **Fragmentation Score** and the compound's **Collisional Cross-Section** (CCS) if available.





Find **Compound 2.91_151.0634n.** The solid grey icon to the left of the **Compound** column indicates possible identifications have been made for this compound. With the available Retention time data for Paracetamol this would indicate that Paracetamol is a stronger match.

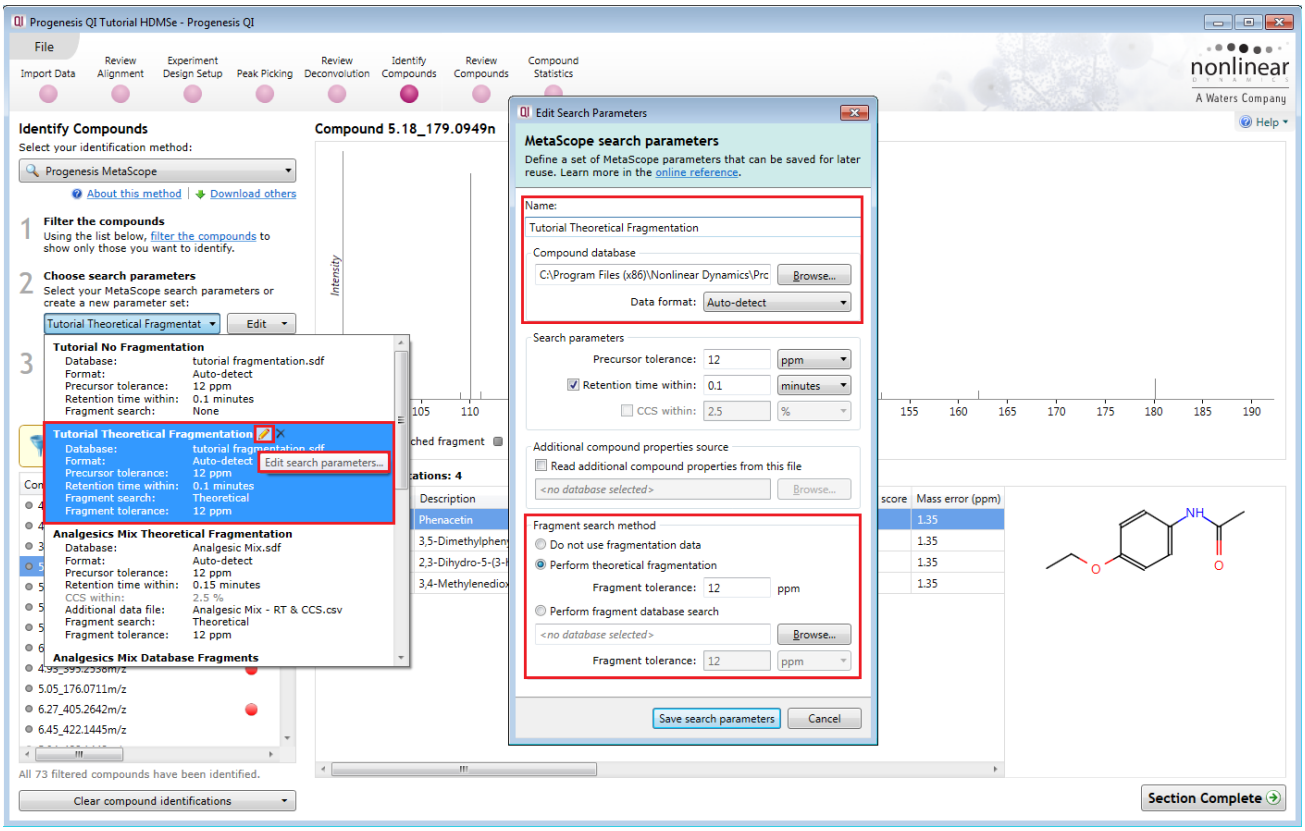


Note: when the Retention time data is **not** included as part of the search scoring (by unticking the option on the search parameters) then there is no difference in the scoring for the 4 identification candidates in the table to the right.

Incorporating Theoretical Fragmentation

To improve the confidence in the compound identification you can choose to perform theoretical fragmentation of a candidate list of compounds and then match the resulting 'in silico' fragmentation against the measured/observed fragments for a compound. The candidate molecules are selected from a compound database based on the exact mass (within a specified error range given in (ppm).

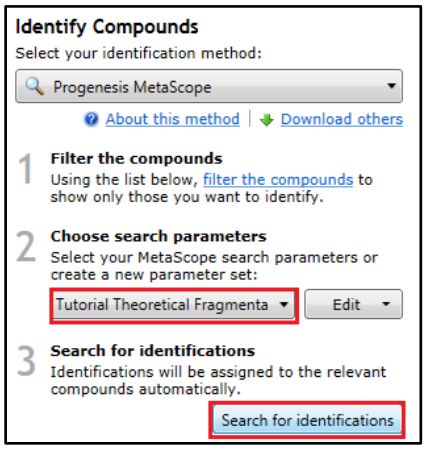
The compound database is in SDF format providing the structural information for each candidate compound. Using this list of candidates, the fragmentation algorithm generates all possible fragments for a candidate compound in order to match the fragment mass with the measured peaks.



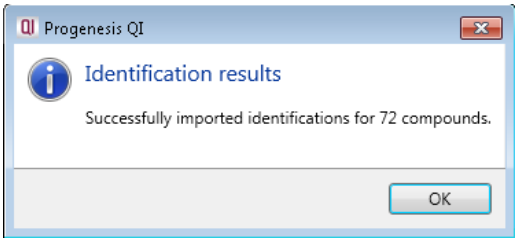
To perform theoretical fragmentation using the **tutorial fragmentation.sdf** as the source of the candidate compounds select the second Search Parameter Set (Tutorial Theoretical Fragmentation) from the drop down

Note: you can review/edit the parameter set, either by clicking on the 'pencil' icon in the drop down or using the **Edit** option to the right of the drop down.

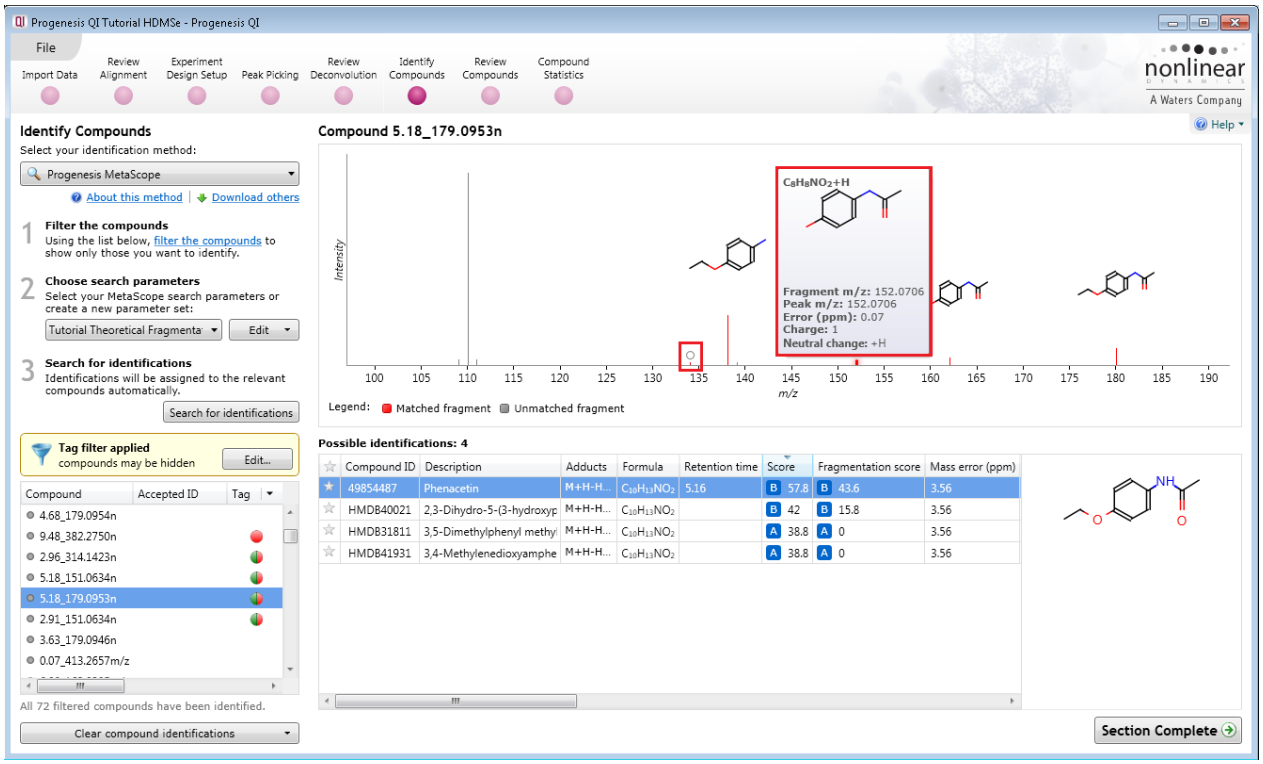
Ensure the **Perform theoretical fragmentation** from the Fragment search method is selected and then save search parameters. To perform the Theoretical Fragmentation search click on **Search for identifications**



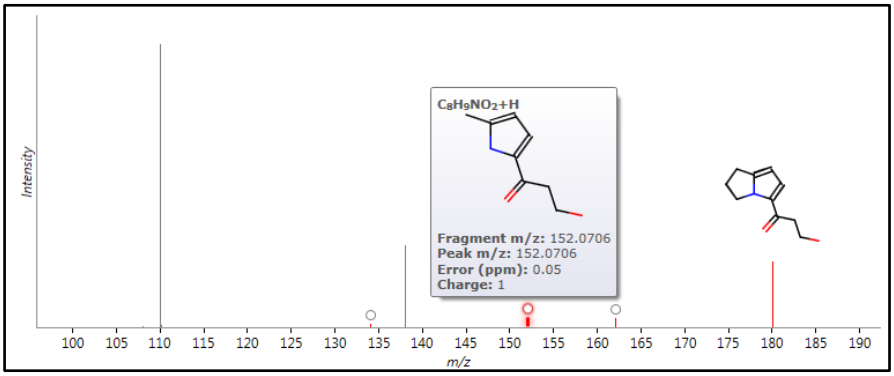
As before search 72 results will be returned.



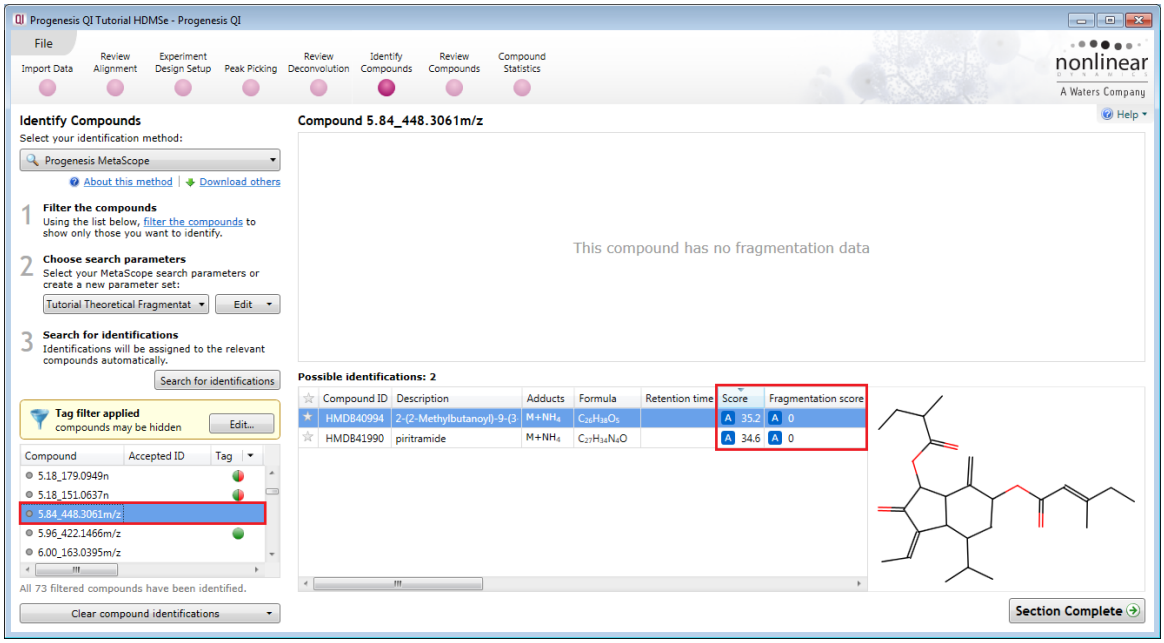
Where fragments have been measured for a compound and a match is made with the theoretical fragments from each candidate compound a solid coloured line appears on the Fragmentation graph. This is accompanied with either an image of the fragment displayed (if there is room) or an open symbol at the top of the line which enables display of the fragment when the cursor is held over it (as shown below for **Compound 5.18_179.0953n**).



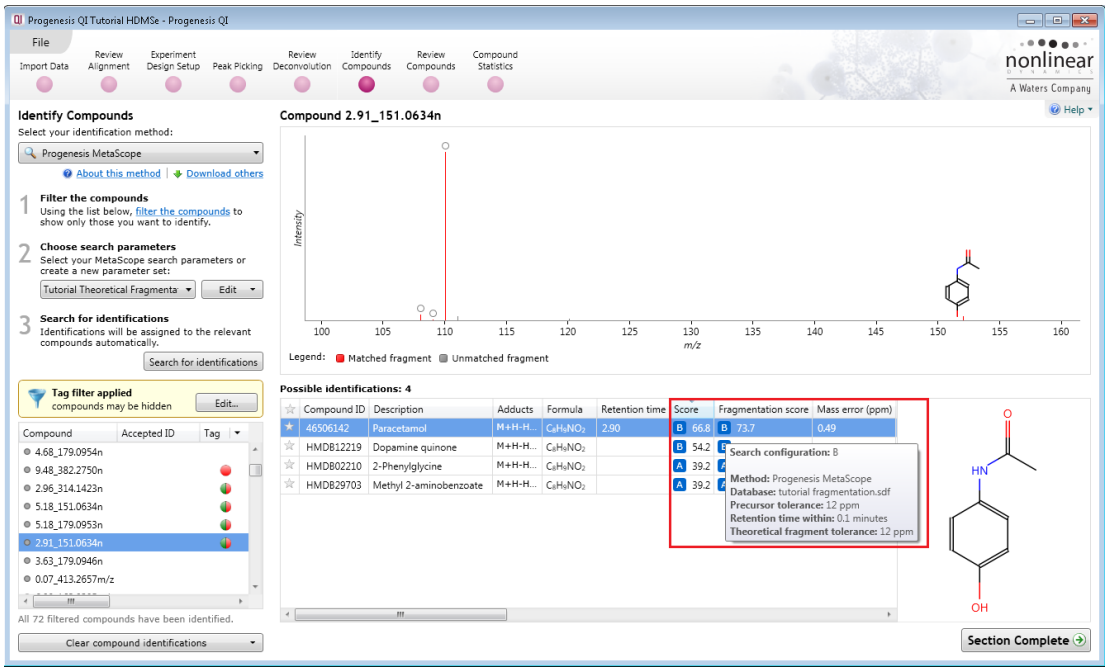
The fragmentation score for each possible identity appears in the table. When you click on one of the possible identifications with a lower fragmentation score the Fragmentation graph displays which fragments contribute to that score for example the **2nd ranked compound** Fragmentation matches (see below).



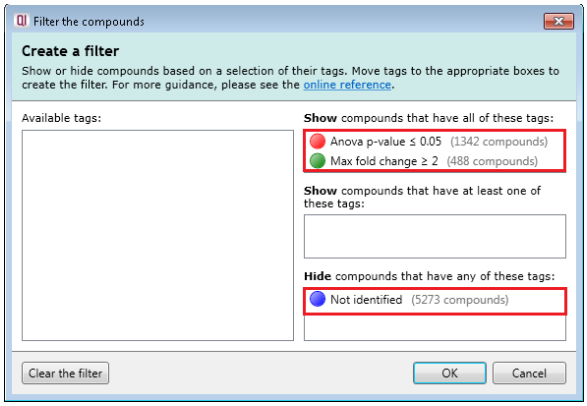
Note: a fragmentation score of 0 indicates either that **no** match was achieved between the theoretical fragments (of a candidate compound) and the measured fragments or in fact the compound has no fragmentation data. For example, in **Compound 5.84_448.3061m/z**:



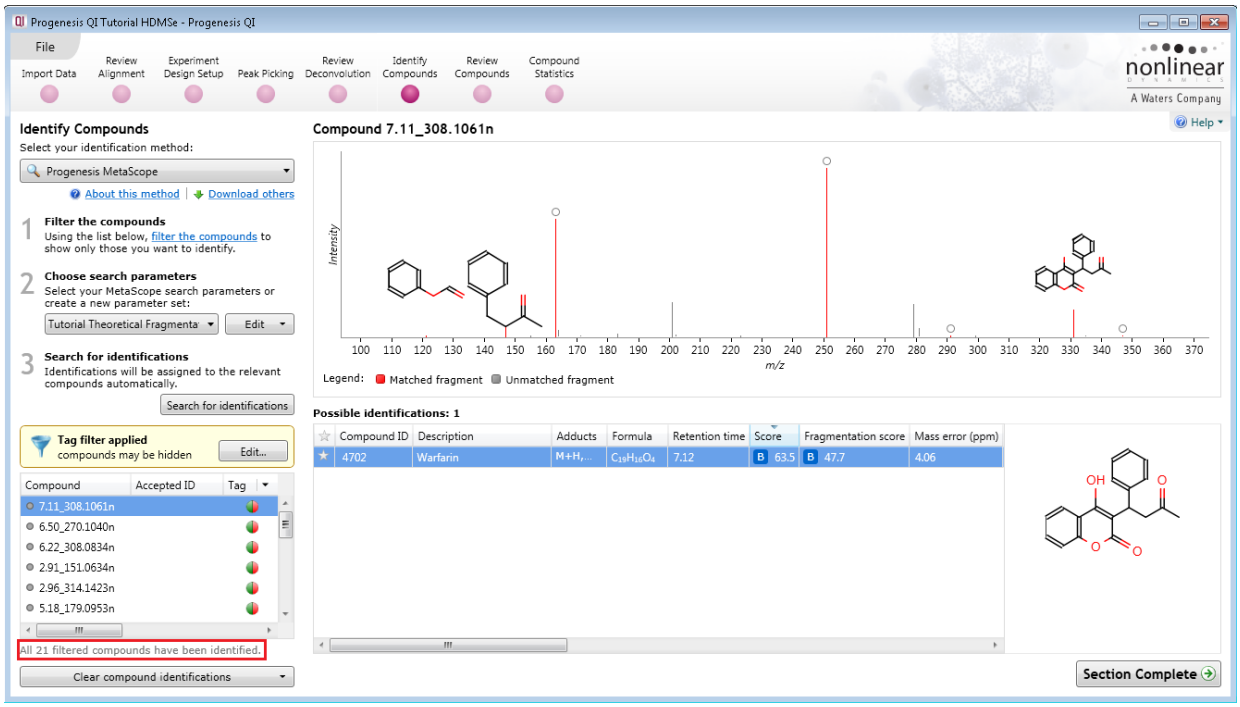
Note: where multiple searches have been performed the search yielding the highest overall score for each candidate is displayed. Also the search parameters giving rise to these scores are displayed as tooltips when the cursor is held over the Search Parameter indicator.



The table of compounds can be reduced to displaying only those compounds that have been identified, show a significant difference between the conditions, and are altered by at least two-fold, by applying a 'Tag Filter' based on the assigned 'Quick tags.

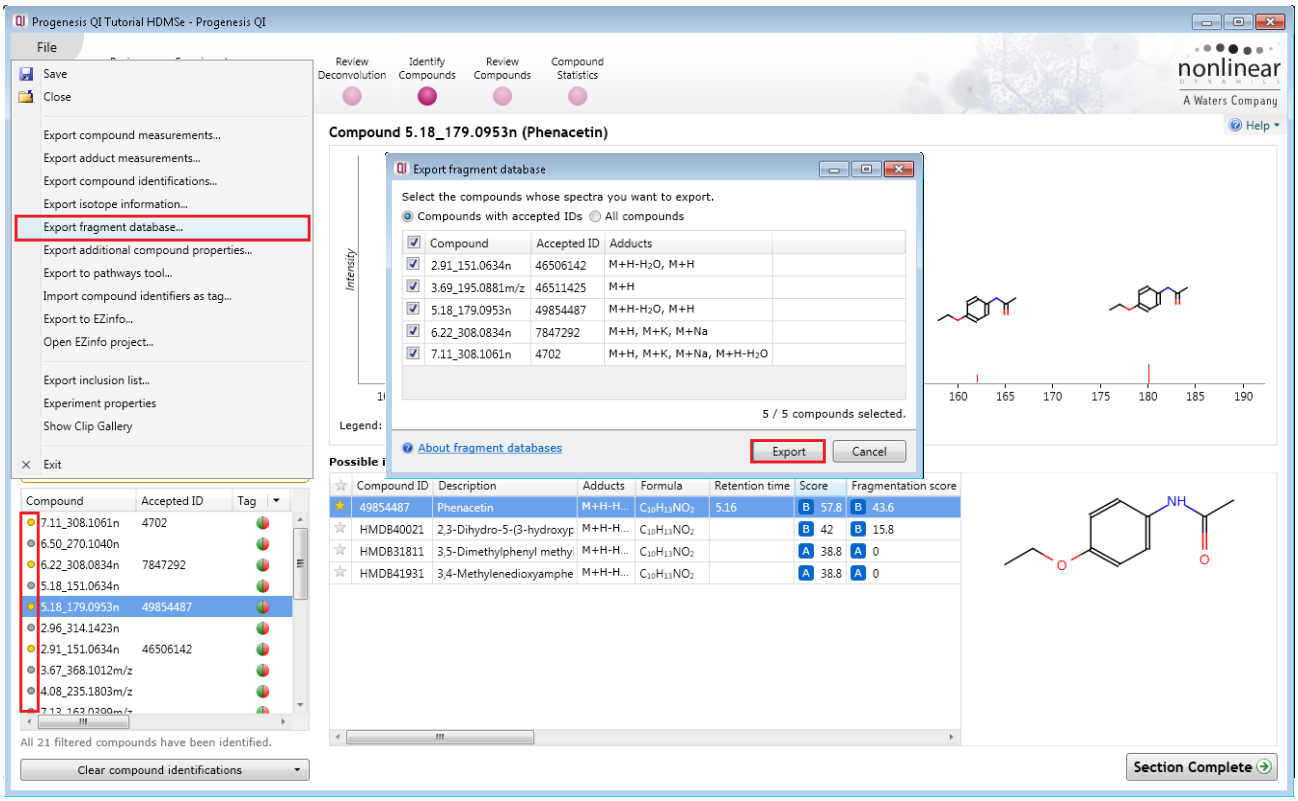


The table now displays only information for 21 identified compounds.



Creating and using databases

The identification data can be stored and used to perform fragment database searches. This allows you to build up your own (local) databases of fragment information. These can then be used to perform future identifications based on fragment matches when using the same compound database. To create and subsequently export data to such a database you must indicate compounds to export by accepting their identity.

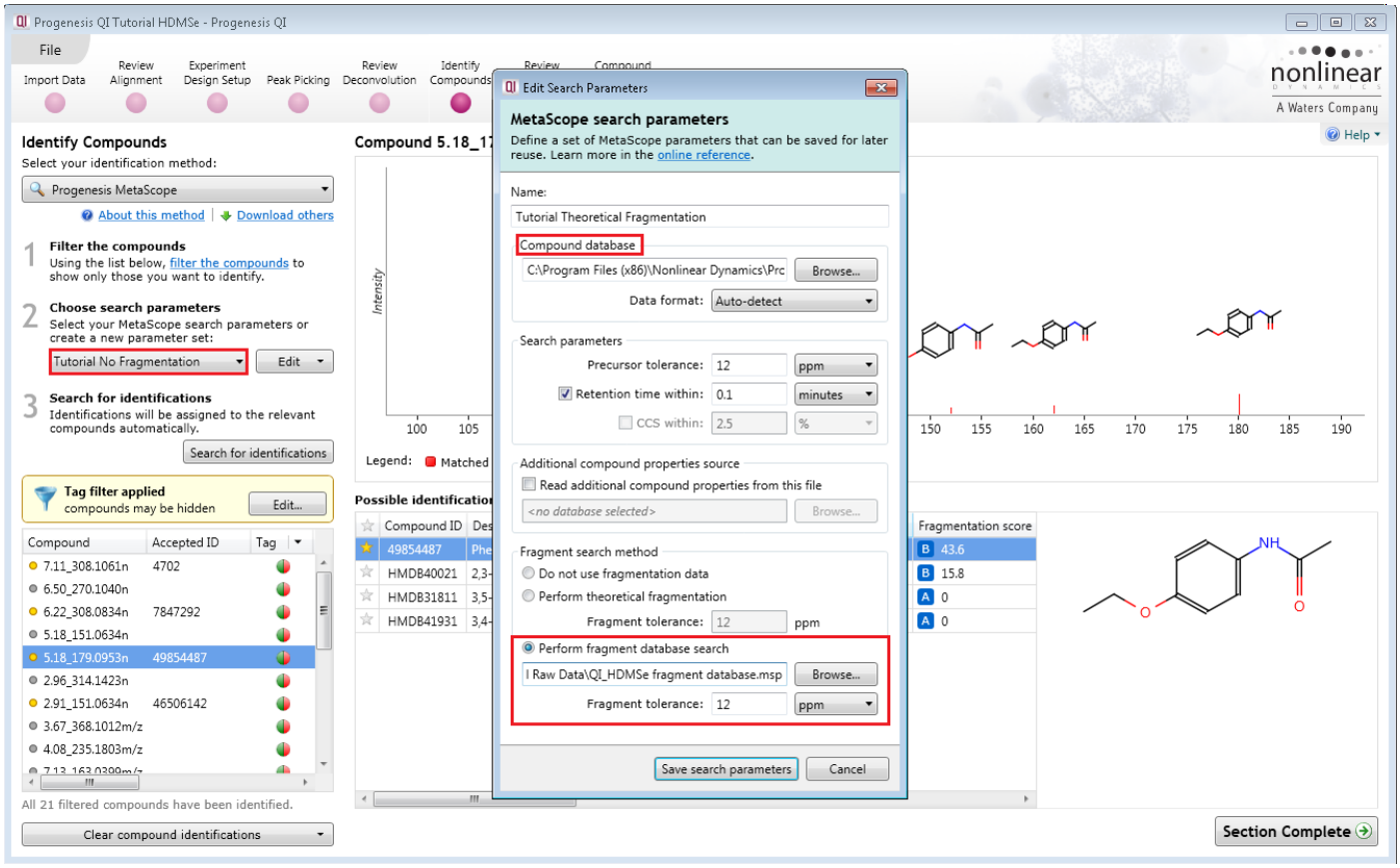


Using the Filter set created for the 21 compounds in the previous section, accept the identity of a compound by clicking on the 'star' icon to the left of the Compound ID, it will turn yellow. Do this for a number of compounds displaying an acceptable fragmentation score.

Then select **Export fragment database…** from the file menu as shown above.

When you export these accepted identifications the file is saved as a Mass Spectrometry Profile (msp) file.

i.e. **QI_HDMSe fragment database.msp** (example: in the folder you restored the Tutorial Data Set.zip).



You can add to this file as you gather additional information from other experiments.

Now as an example re-perform the Tutorial Theoretical Fragmentation but this time instead of selecting the **Perform theoretical fragmentation** option select the **Perform fragment database search** option. Then locate the .msp file you have just created.

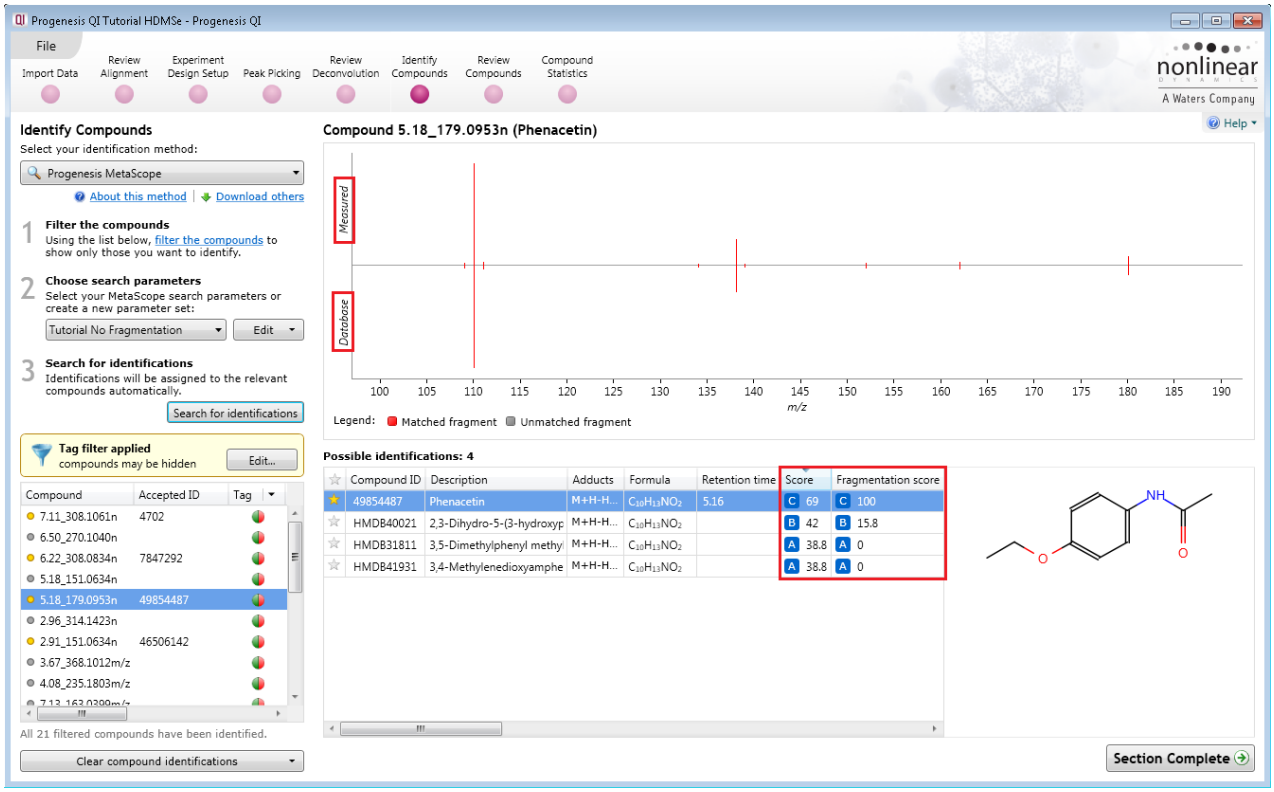
Then save the Search parameters and perform the Fragment database search by clicking on **Search for identifications**.

Note: you must use the same Compound database that was used when the original msp file was generated. This is because the initial step of Progenesis QI using a fragment database is to find all possible identifications based on neutral mass (RT and CCS used if available). Then for each possible identification (candidate) Progenesis QI searches the fragment database for a record with a matching compound ID. It then compares the stored fragment data to the current experimental fragmentation data.

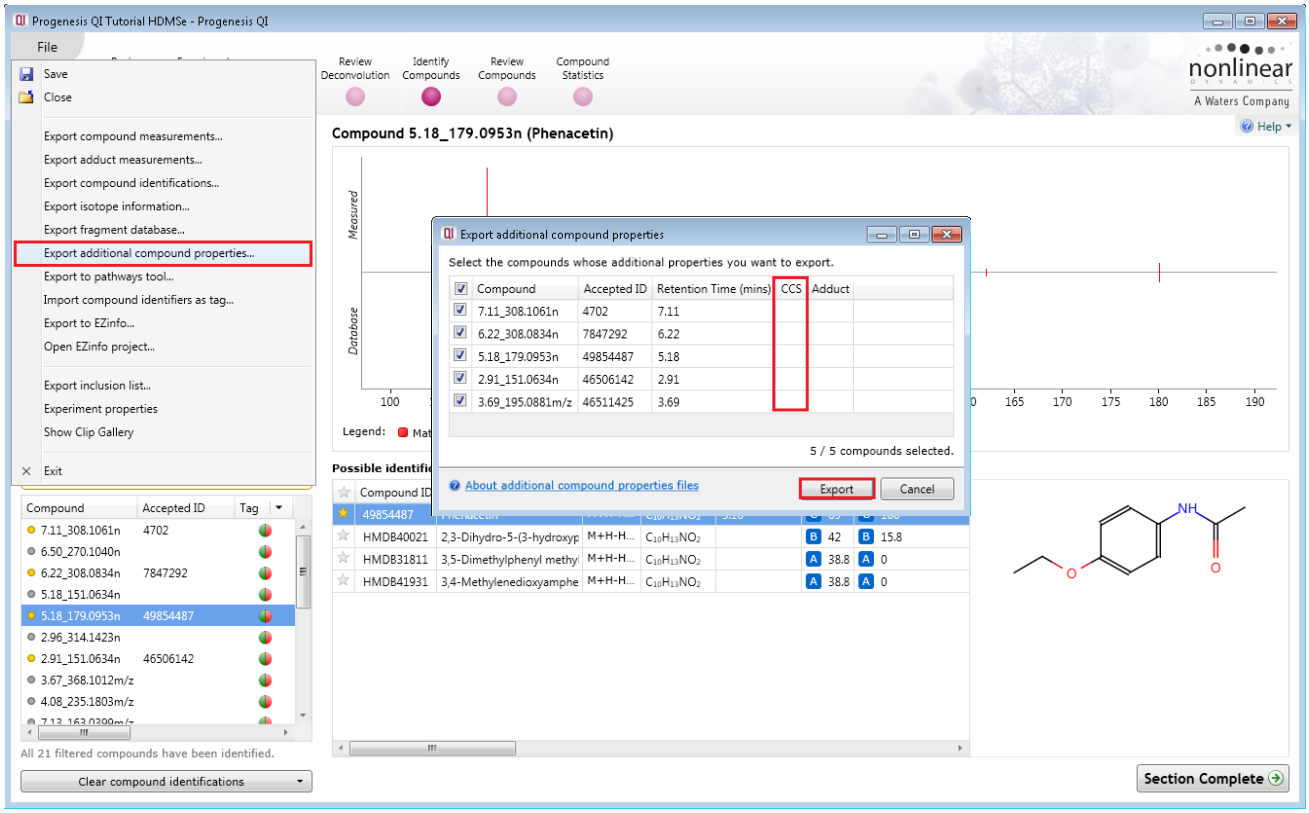
On performing the search you will get a 'mirror plot' where the upper half represents the measured/observed fragmentation and the lower half shows the currently stored fragmentation in the database. This gives rise to a **Fragmentation score** describing how good the match is between the observed fragmentation and the stored fragmentation in the database.

The fragmentation score is based on the number of fragments matched and their intensity. This excludes any contribution from the matching of the precursor.

The current search with the highest score is designated **C** in this example (**Compound 5.18_179.0953n)**.



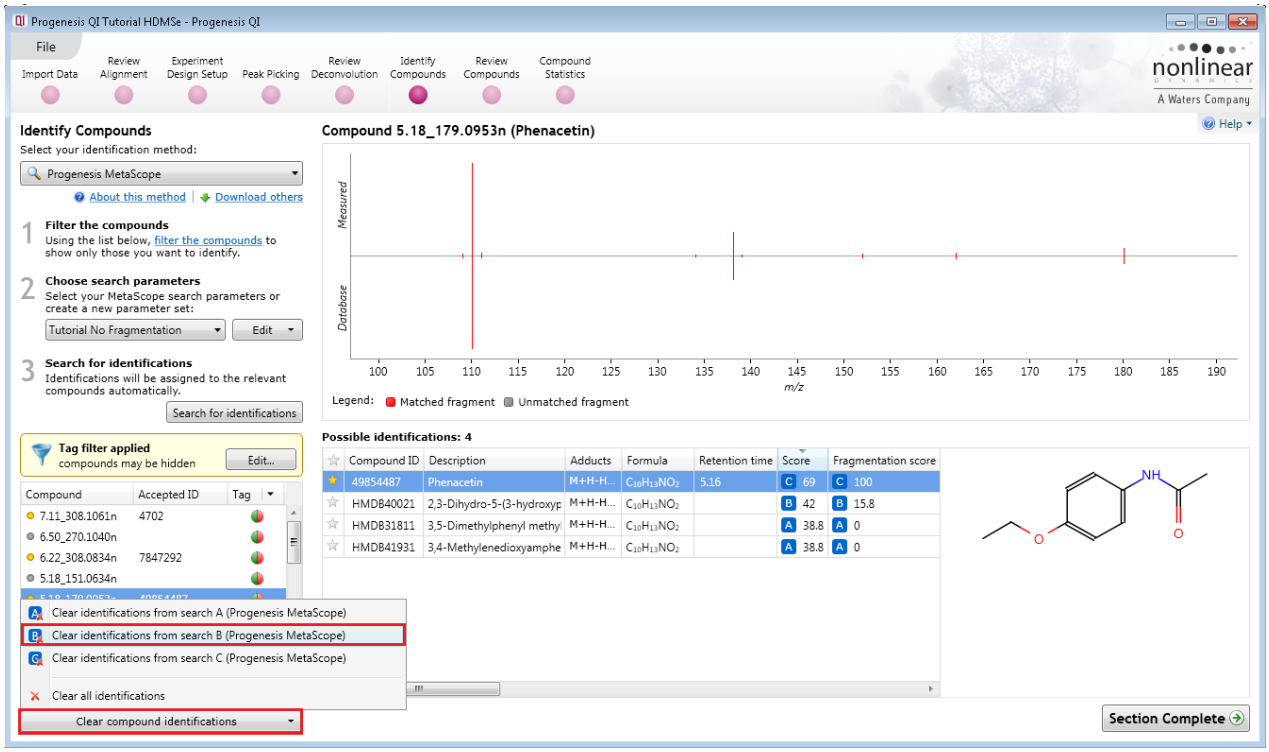
You can also manage and search against additional compound properties. The advantage of this is that you can create a database of compound data that retains information on an identified compound for m/z, retention time. Export and save this **Additional Compound properties** file.



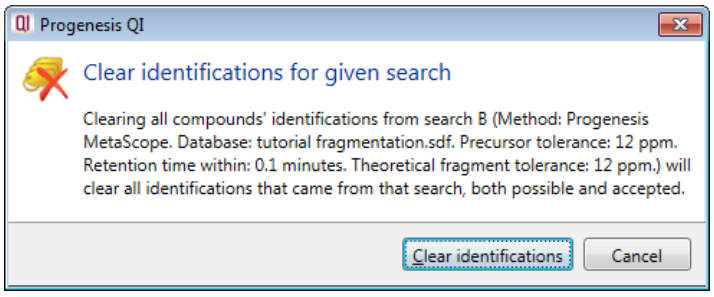
Note: if the data format contains Drift time data and it has been calibrated then the value for a compounds Collisional Cross Section (CCS) may also be stored as an additional compound property and used in the process of compound identification.

Deleting Compound Identifications

Before using the **additional compounds properties** file, you can explore how to clear compound identifications. If more than one Search Configuration has been used then you can remove identifications based on the Configuration. To do this click on **Clear compound identifications** and select the configuration.

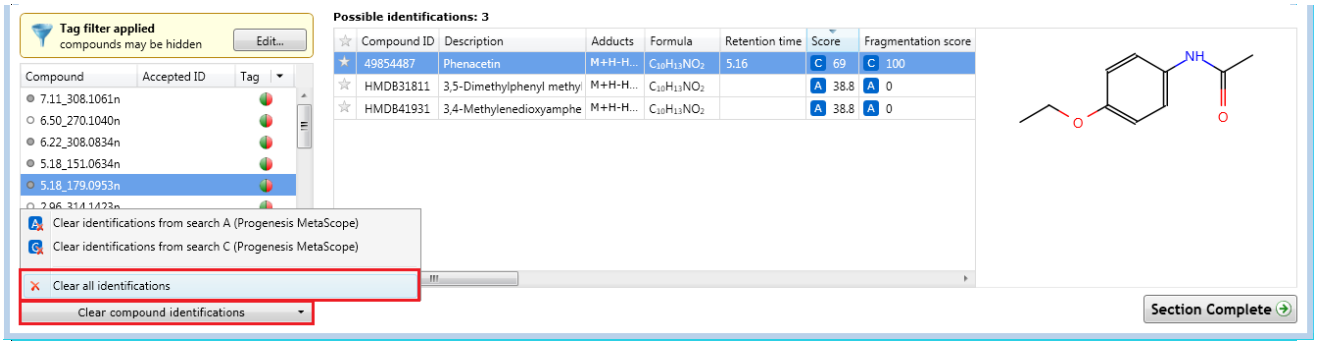


When deleting identifications by Search configuration (in this example B) you will be warned which ones will be cleared.

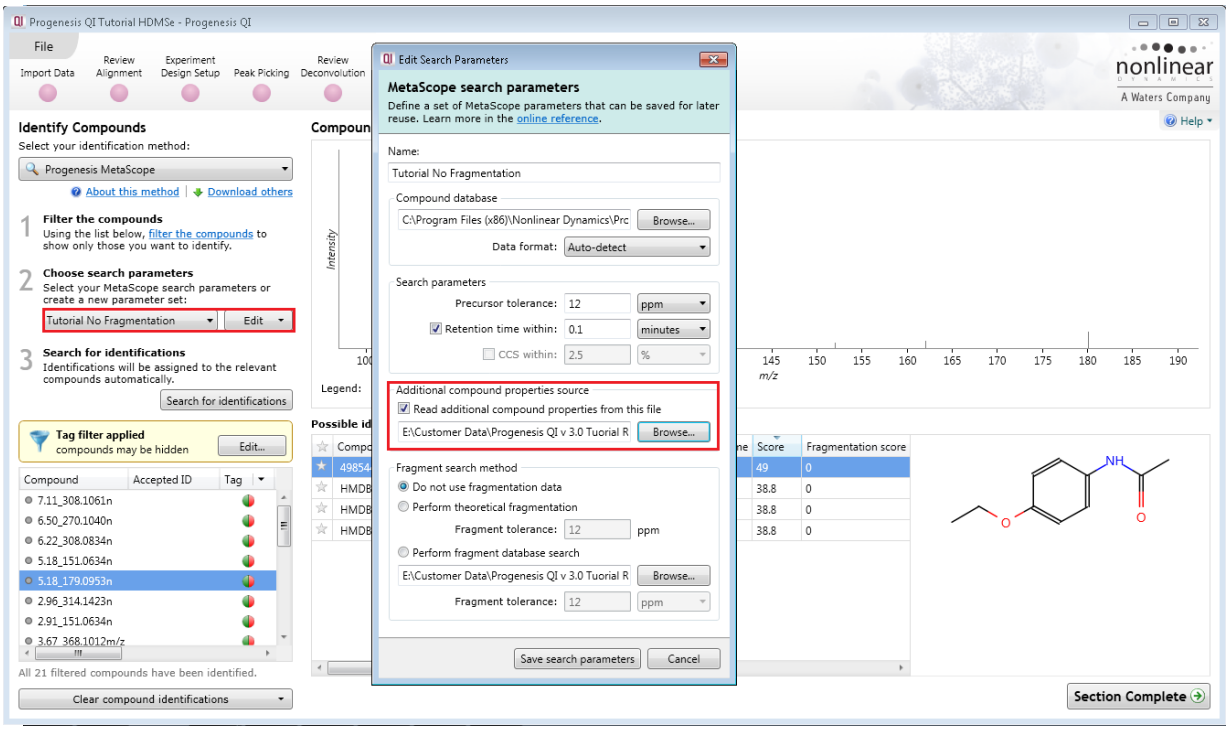


Following deleting the identifications based on Search configuration B there will be 3 possible remaining identifications for **Compound 5.18_179.0953n**

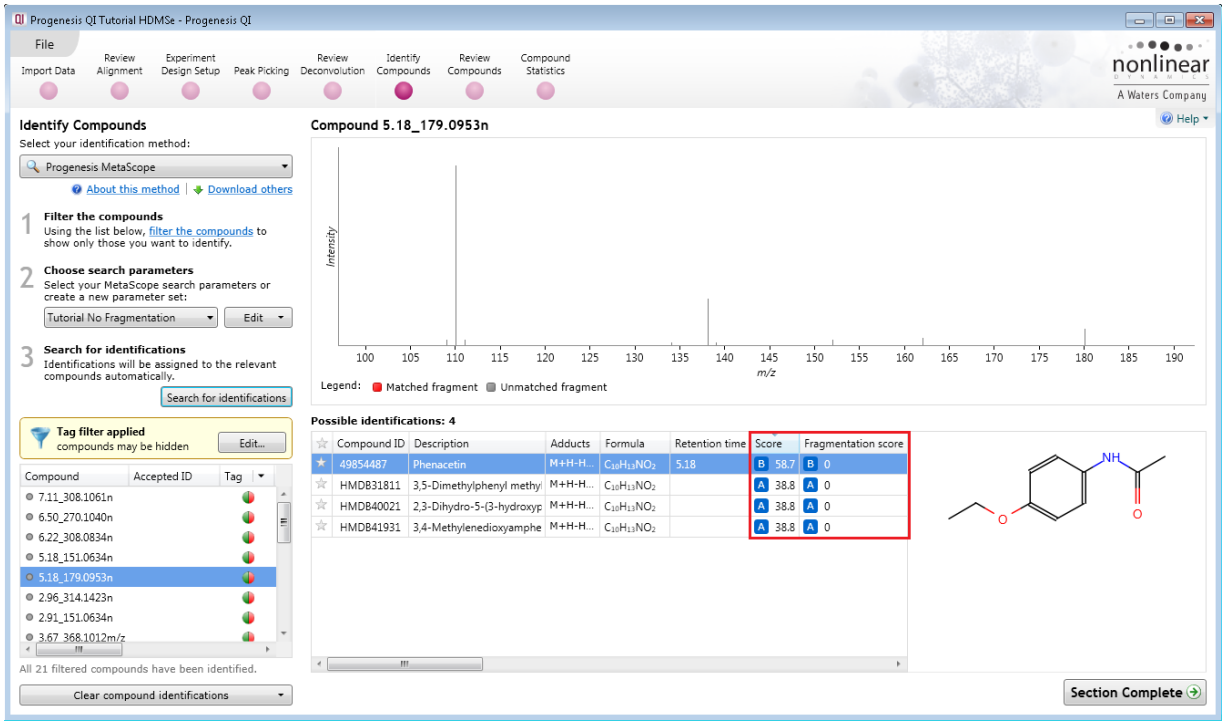
Now clear all the compound identifications and re-perform the **Tutorial No Fragmentation** search as described on page 45.



To make use of the **Additional compound properties file**, you created on page 46, either edit, copy or make a new set of search parameters using this file as the source. In the example below we have edited the existing search parameters for the **Tutorial no Fragmentation** search parameters.



When you re-perform the search a higher overall score is achieved using the additional scored compound data.



Note: the saved **Additional Compound Properties** file is the source of the 'in house' saved compound properties (Retention time and CCS) while the Compounds database file provides the neutral mass and structure details.

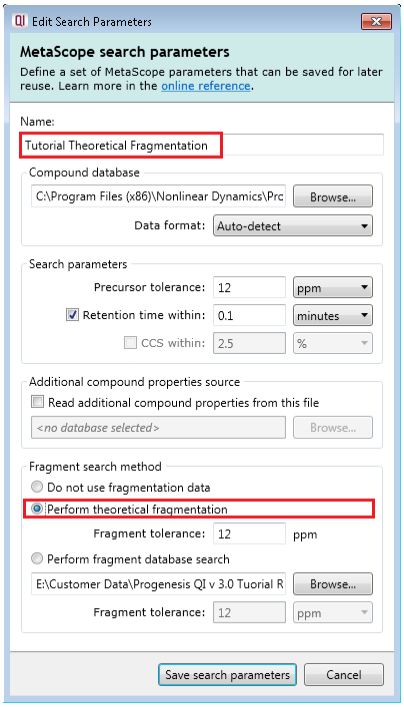
Each matching component (where available) contributes 20% to the overall score: Mass error, Retention time, Isotope similarity, Fragmentation score and Collisional Cross-Section (CCS).

i.e. if only mass error and isotope similarity are available then the maximum score will be out of 40.

Additional Identification Methods

In addition to Metascope and Metlin, Progenesis QI v3.0 has identification methods for Elemental Composition, LipidBlast, ChemSpider and NIST. With the exception of NIST, these methods are installed by default and are available at the Identify Compounds stage from the identification method drop down.

As an example of using **Elemental Composition**: first **Clear all compound identifications** that you carried out in the previous section**,** and leave the current Tag Filter applied.

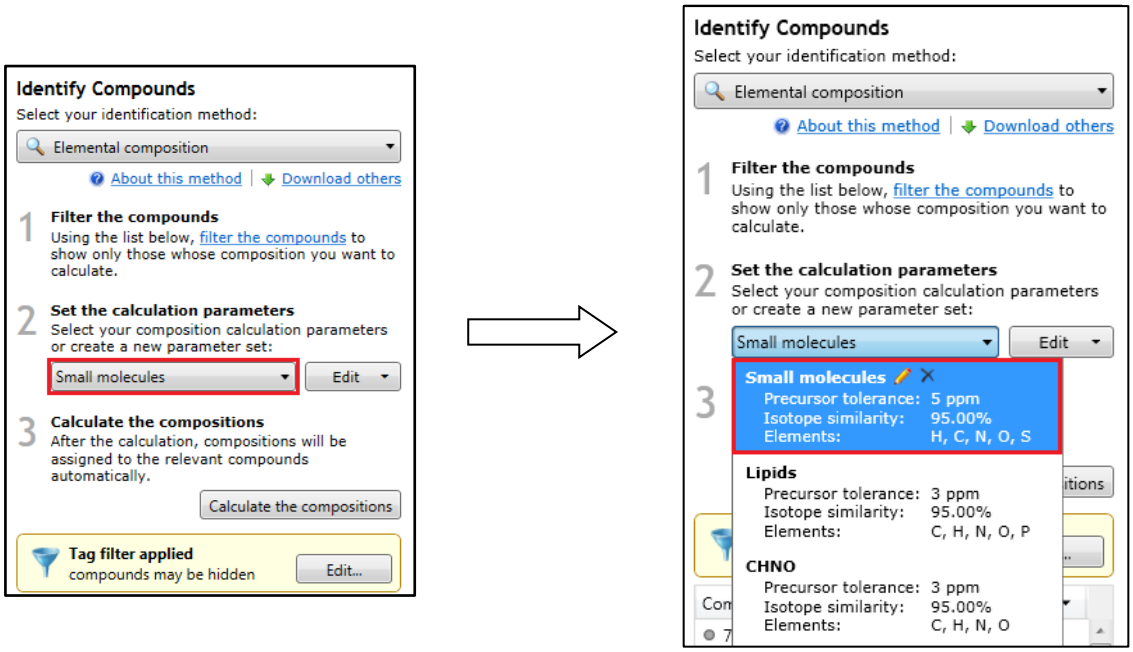


Then with MetaScope still selected as the current identification method return the **Tutorial Theoretical Fragmentation** search parameters to the default settings as shown to the right by ticking **Perform theoretical fragmentation**. Save the search parameters.

Finally re-perform the search for identifications using the Tutorial Theoretical Fragmentation parameter set.

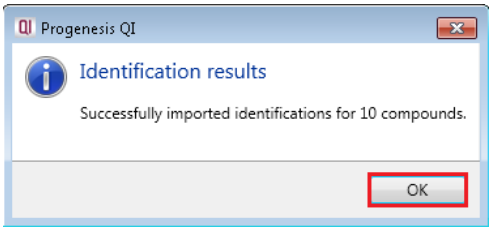
Then select **Elemental composition** as the Identification methods.

There are 3 default parameter sets: Small molecules, Lipids and CHNO each containing a list of Elements to be considered and thresholds for **Precursor tolerance (ppm/Da)** and **Isotope similarity (%).**

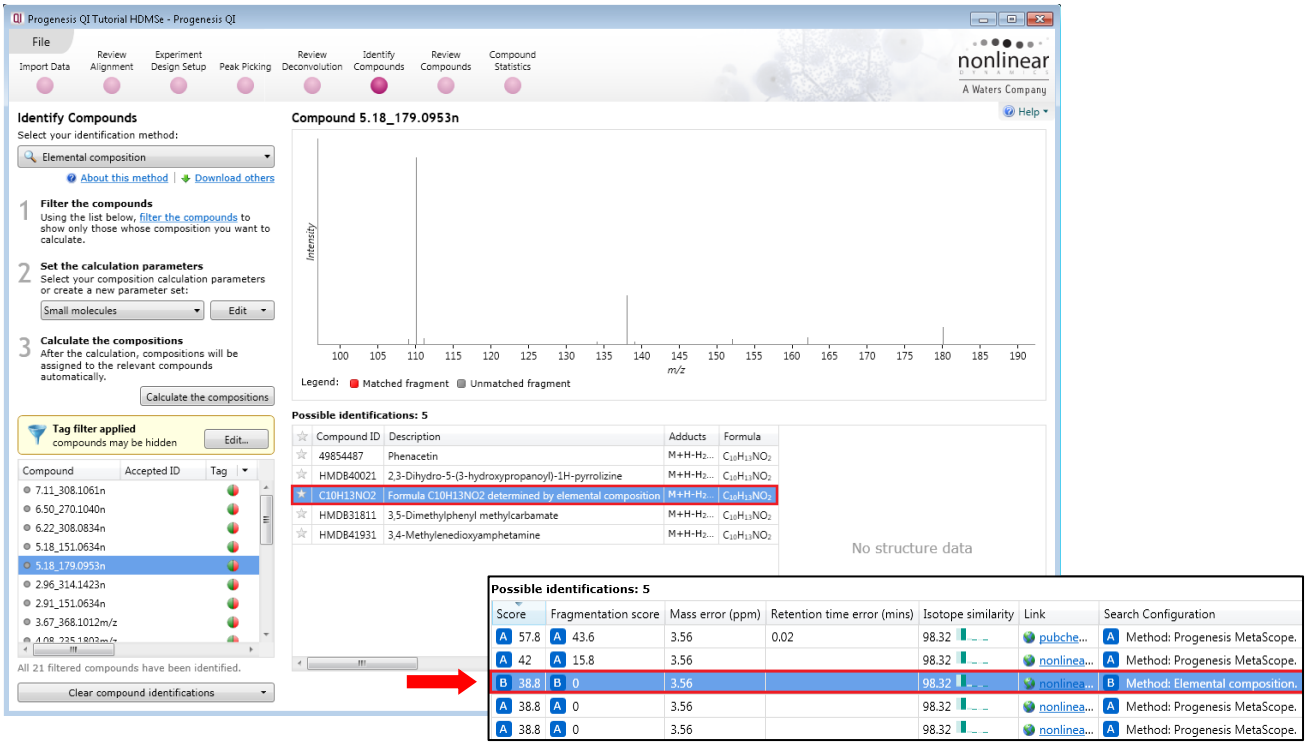


Select the **Small molecules** parameter set and click **Calculate the compositions**.

Click OK to import the 10 identifications.



The **Possible identifications** table for the current compound updates to include any compositions that meet the current calculation parameters.

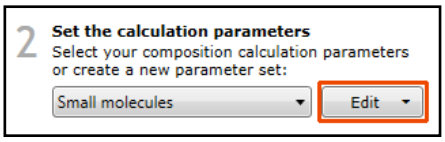


Note: For the example shown an additional compound ID appears as you accept the identifications based on the calculation of the Elemental Composition.

Scrolling along to the right the actual Mass error and Isotope similarity values are displayed. The **B** indicates the type of search applied.

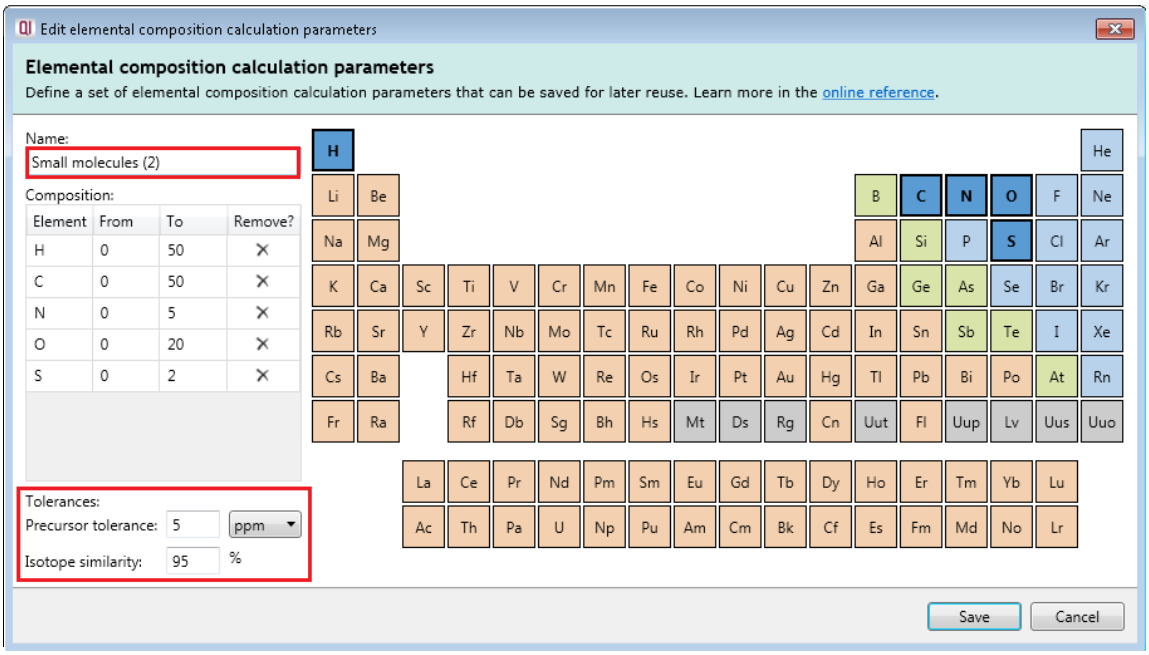
Note: in this case the only elemental composition returned (using the current thresholds) corresponds to the formula returned for the other possible identifications.

To edit/create/copy a parameter set select from the **Edit** drop down to the right.



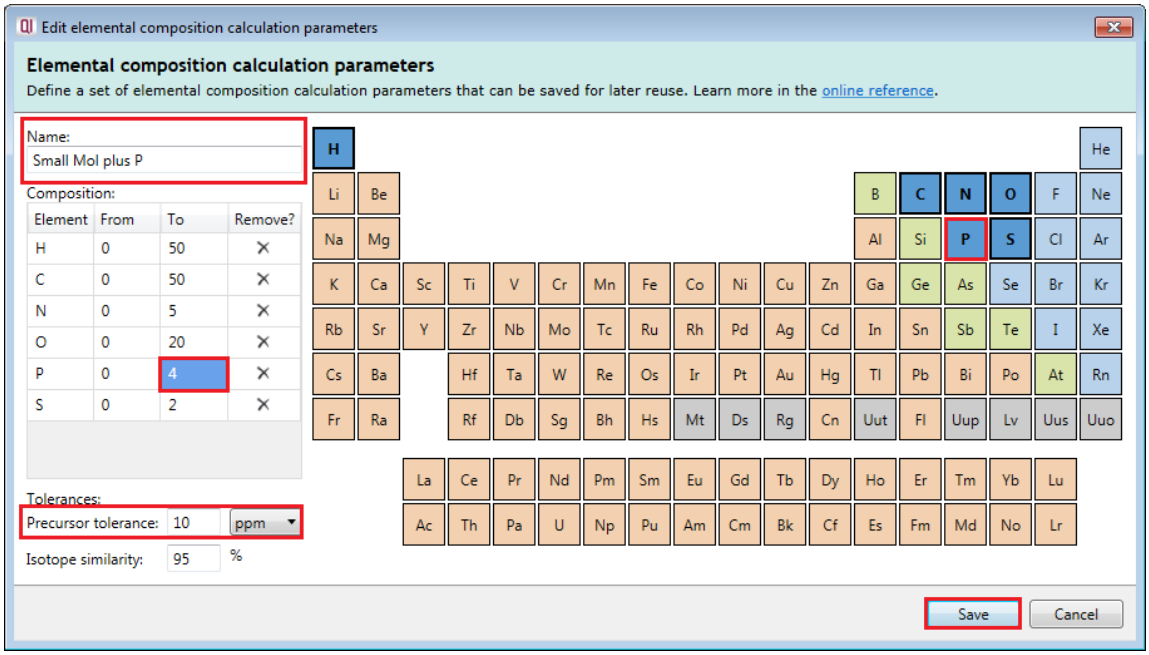
As an example, create a new parameter set by first copying the existing default set for small molecules.

Select **Create a copy…** from the Edit drop down, the edit dialog for the calculation parameters will appear.



Rename the Parameter set (Small Mol plus P), then add Phosphorous (P) to the composition by clicking on it in the Periodic table.

Then set the maximum number of Phosphates for the composition by adjusting the value in the 'To' column (i.e. 4). Increase the Precursor tolerance to 10ppm.

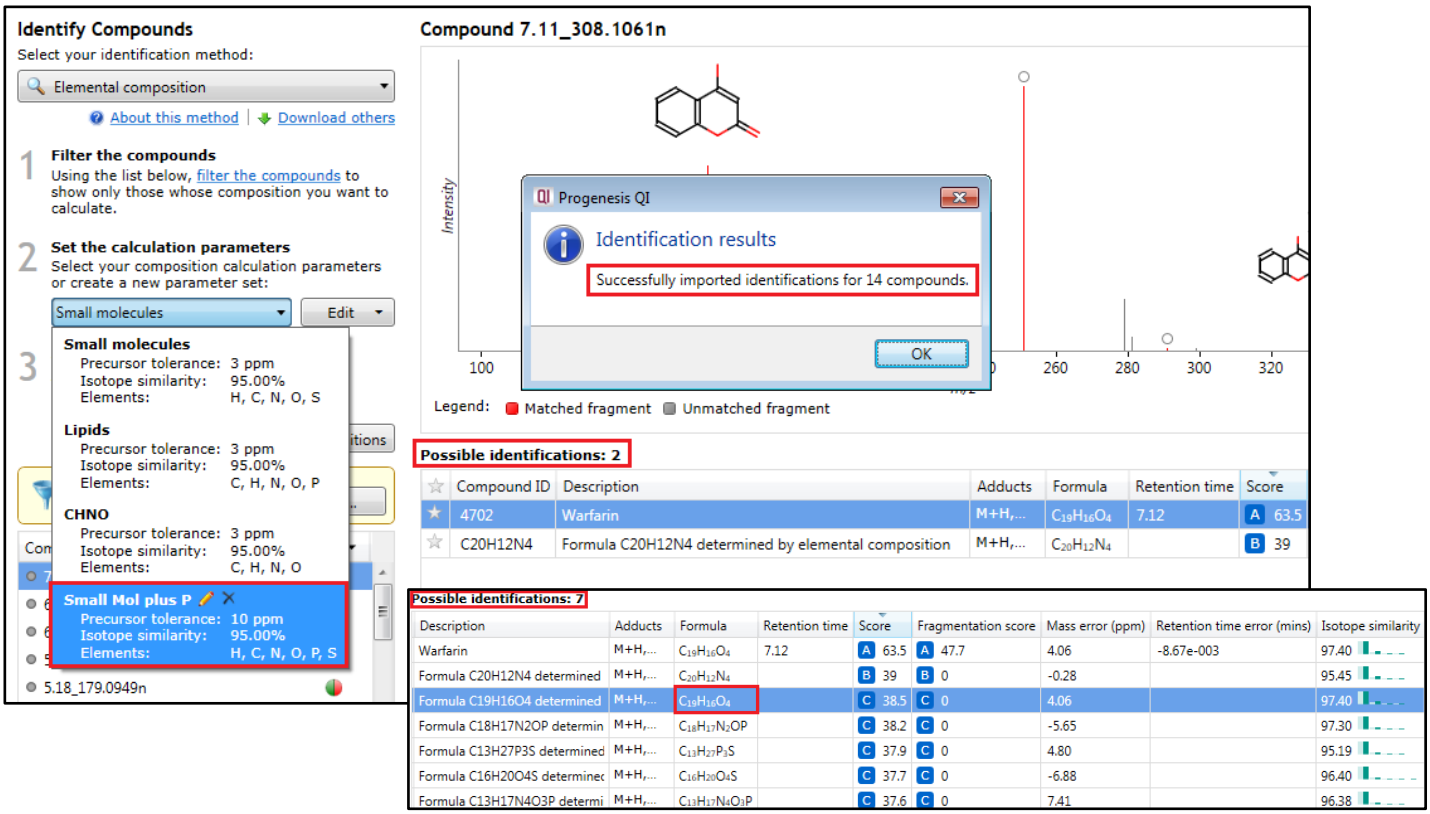


Finally **Save** the new parameter set.

Note: on saving, the new parameter set becomes current and its parameters are listed in the drop down.

Make compound 7.11_308.1061n current in the table and click **Calculate the composition**.

When the compositions are calculated using the adjusted composition and tolerances an additional 5 possible ids appear in the list.



Note: Adding phosphorous to the composition parameter set and increasing the Precursor tolerance returns more possible identifications including the formula for Warfarin which had previously been excluded due to the previously low Precursor tolerance.

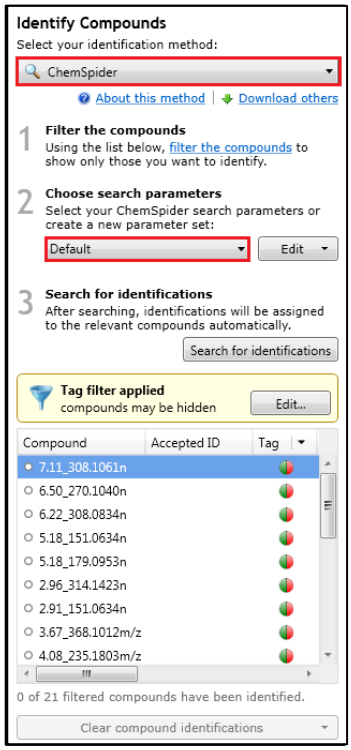
Running a ChemSpider Search

ChemSpider is a web-based chemical structure database with access to over 32 million structures from hundreds of data sources. This method makes use of those ChemSpider web services, automatically exporting data from Progenesis QI to ChemSpider for searching according to the parameters you select, importing the results, and assigning them against the correct compounds within the software.

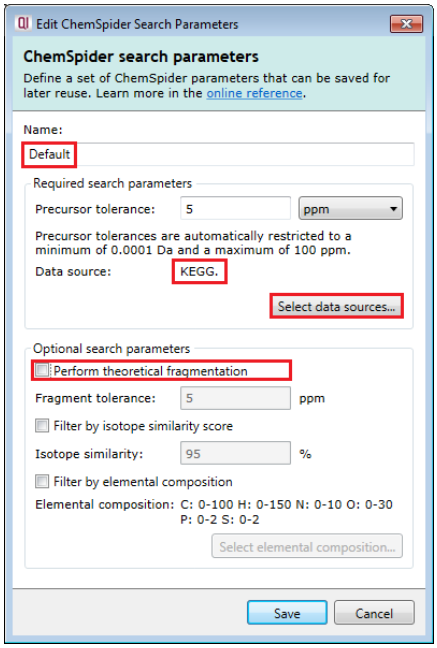
Note: use of ChemSpider requires internet access and is a service provided to users with a valid Progenesis QI support contract.

Select ChemSpider as the identification method.

If there are already compound identifications present from the previous sections, before performing a search first click on **Clear compound identifications** and click **Clear all identifications**.



ChemSpider can access a large number of data sources when you Search for identifications.



With the search parameters set to Default Click **Edit.**

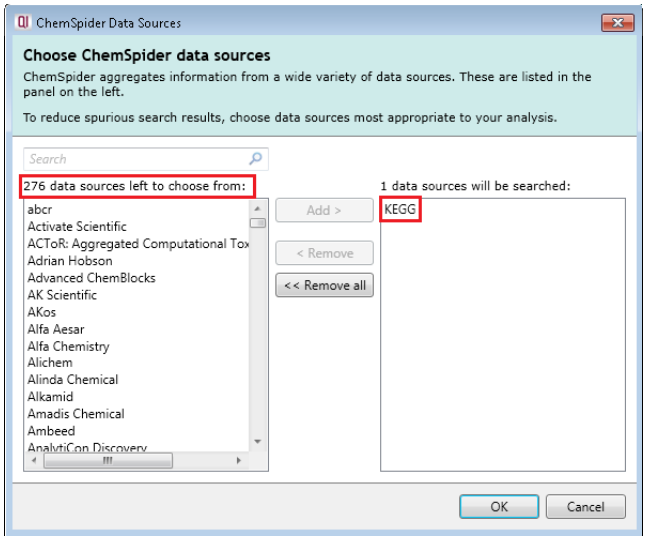
The first time you use ChemSpider, KEGG is selected as the default Data Source.

Untick the **Perform theoretical fragmentation** option

To add additional data sources or change the current data source Click **Select data sources…** .

The ChemSpider data source dialog opens. This enables you to search and select from the 276 data sources currently accessed by ChemSpider.

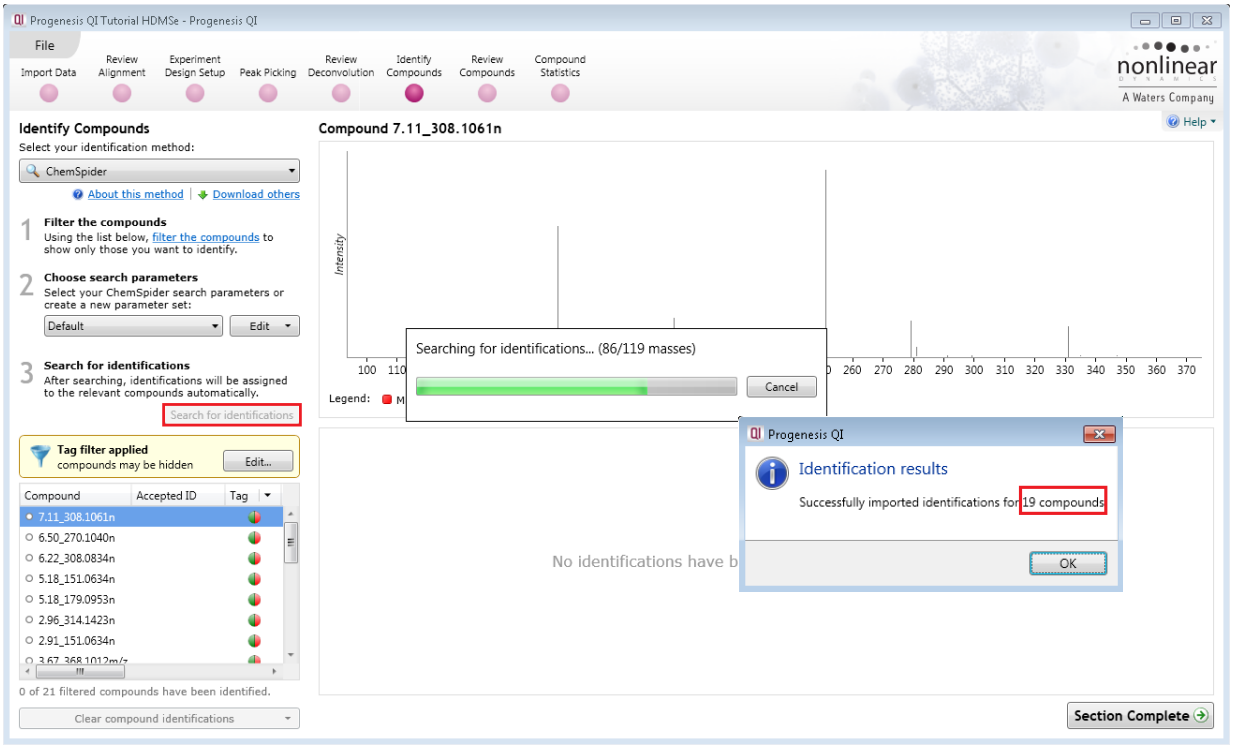
Add the most appropriate source(s) for your current search to the right hand panel.



Tip: since the ChemSpider search is based on accurate mass matching, searching a large number of data sources can take an appreciable amount of time.

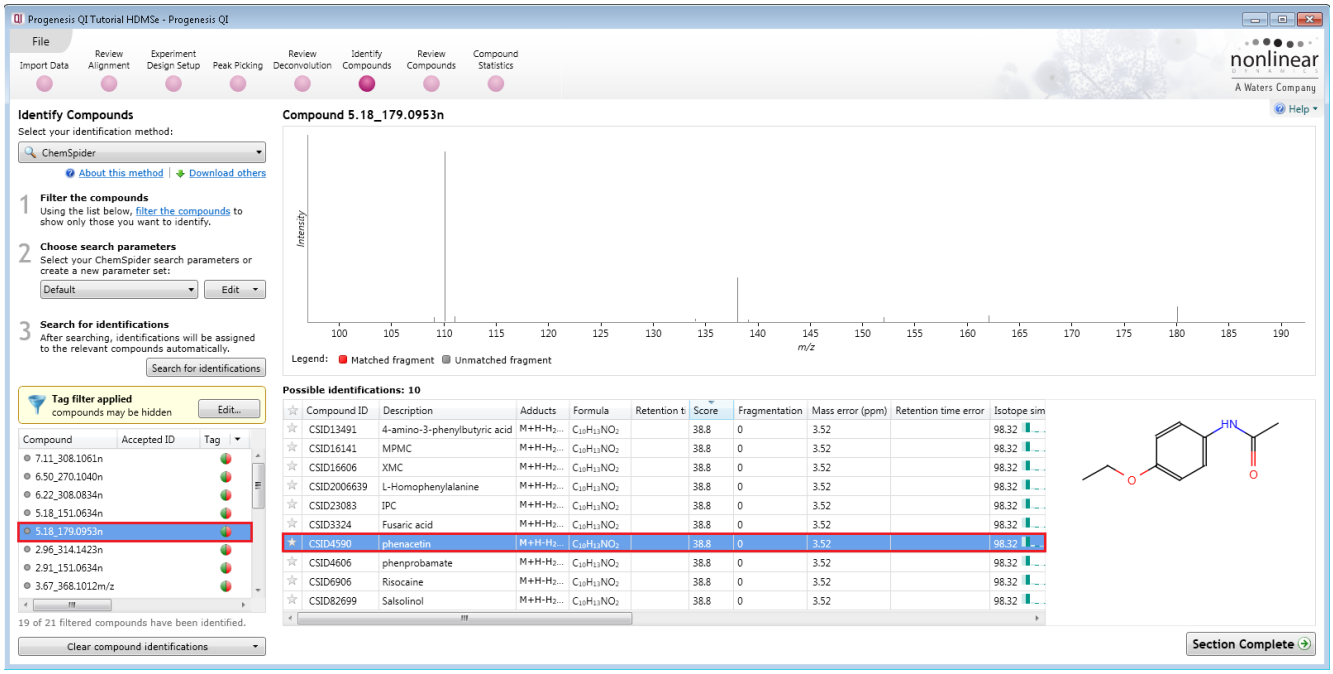
The selected Data sources will be added to the Search Parameters. **Note**: if more than five are selected then a number is reported for those used in addition to the 5 displayed.

Then click **Search for identifications.**

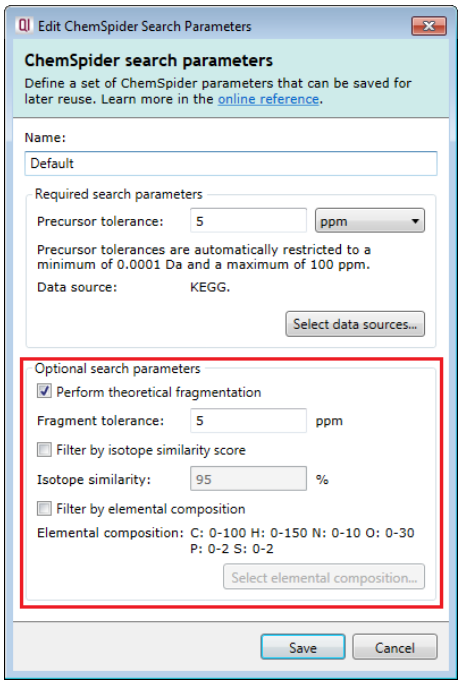


The progress bar reports the number of masses being searched. When the search completes the number of compounds with imported identifications is displayed.

Note: although Phenacetin is identified as one of the possible identities for **Compound 5.18_179.0953**, based on a mass error tolerance of 5ppm there are 9 other possibilities with the same mass error.



The ChemSpider search parameters includes 3 additional options that can be selected to modify the search:



Theoretical fragmentation can be performed on the search results returned by the ChemSpider search using a specified Fragment tolerance.

Isotope similarity filtering discards ChemSpider results where the isotope similarity score falls below a specified cut-off. **Note**: the score is calculated by comparing theoretical and measured isotope patterns as described for the MetaScope search engine.

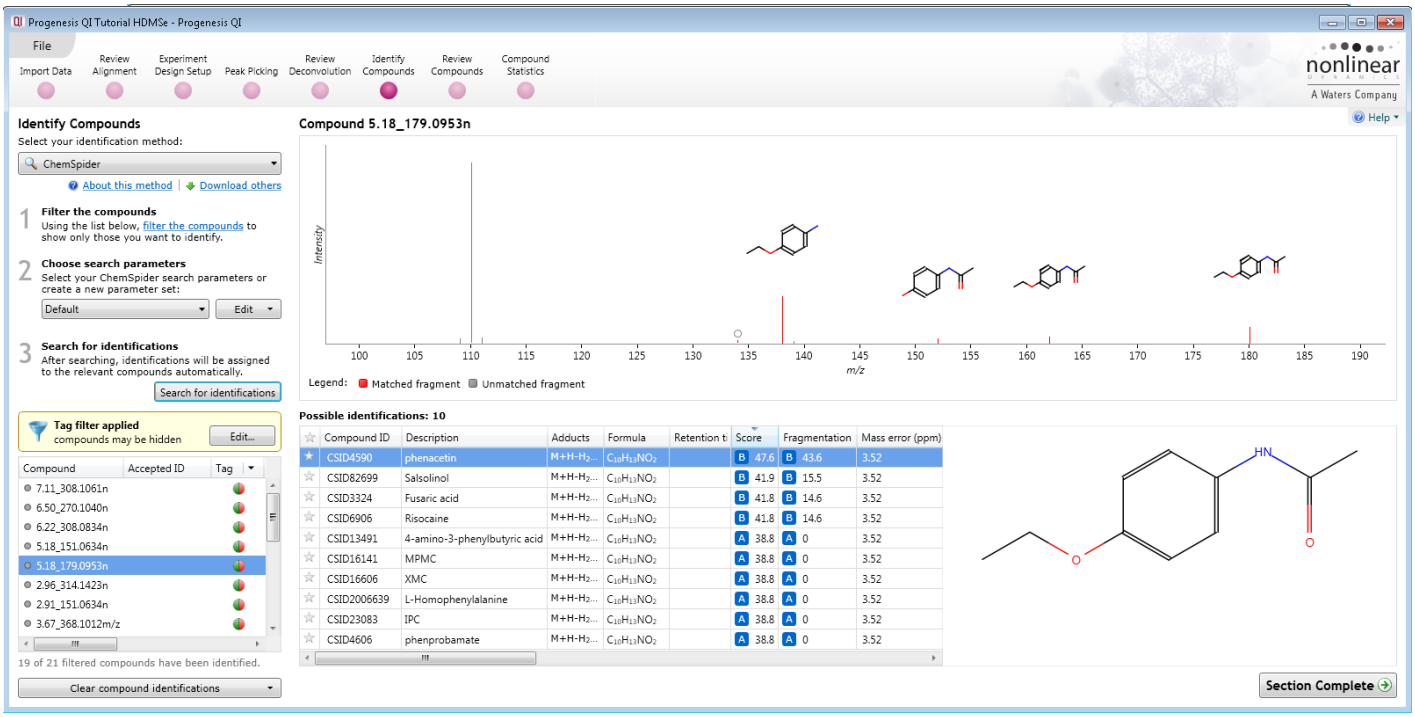
Select elemental composition which compares the formula returned by ChemSpider to a specification based on a list of elements and their allowed count ranges.

Performing Theoretical Fragmentation on ChemSpider Results

As an example, tick the **Perform theoretical fragmentation** option with the **Fragment tolerance** set to 5ppm.

Save the parameter set and re-perform the search.

When the ChemSpider search is re-performed, using the additional information provided by performing theoretical fragmentation on the results this now provides a level of order to the 10 possible identifications for **Compound 5.18_179.0953** making Phenacetin to be the more likely compound identification.

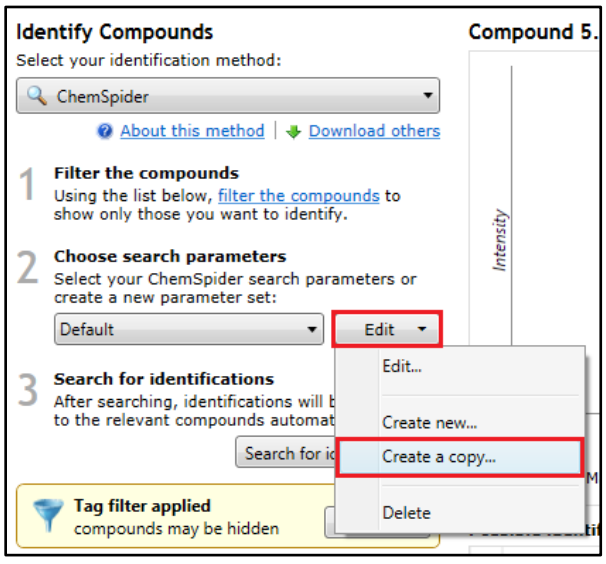


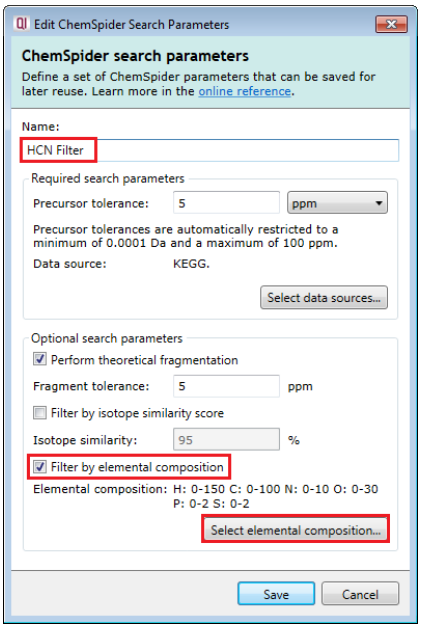
Filtering using ChemSpider Results using Elemental composition

To apply a filter based on a specified elemental composition to ChemSpider you must first set up the filter as a new or edited parameter set.

Tip: a simple way to create a new search parameter set is to **Create a copy**, rename and edit an existing set.

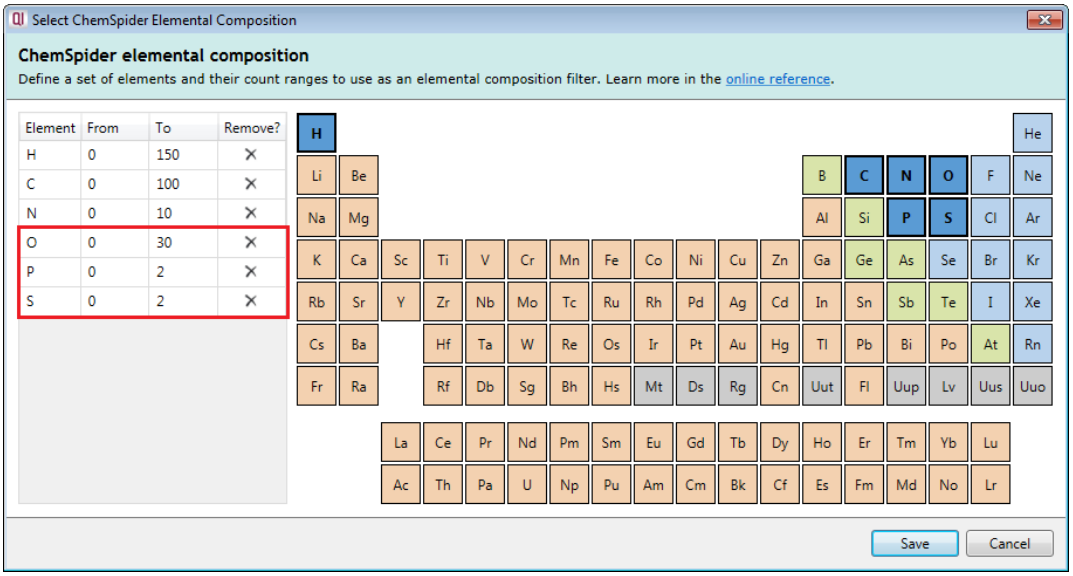
With the Default set as current, select **Create a copy** from the Edit menu. In the parameters dialog rename this set (i.e. HCN Filter) and tick **Filter by elemental composition** and untick Perform theoretical fragmentation**.**





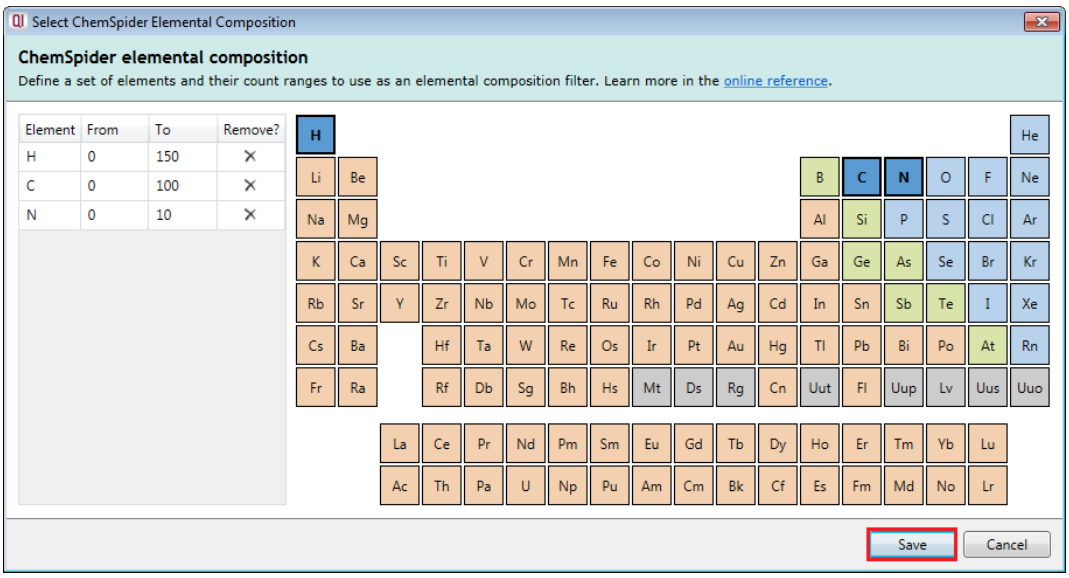
Click **Select elemental composition…**to edit the current Elemental composition.

The existing composition is displayed.



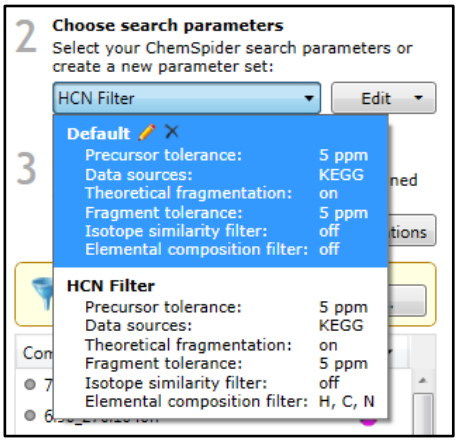
Remove elements either by clicking on the ***x*** in the table or on the element(s) in the periodic table. Then adjust the ranges (From and To) of atoms for each element in the formula using the table.

Finally **Save** the adjusted composition.

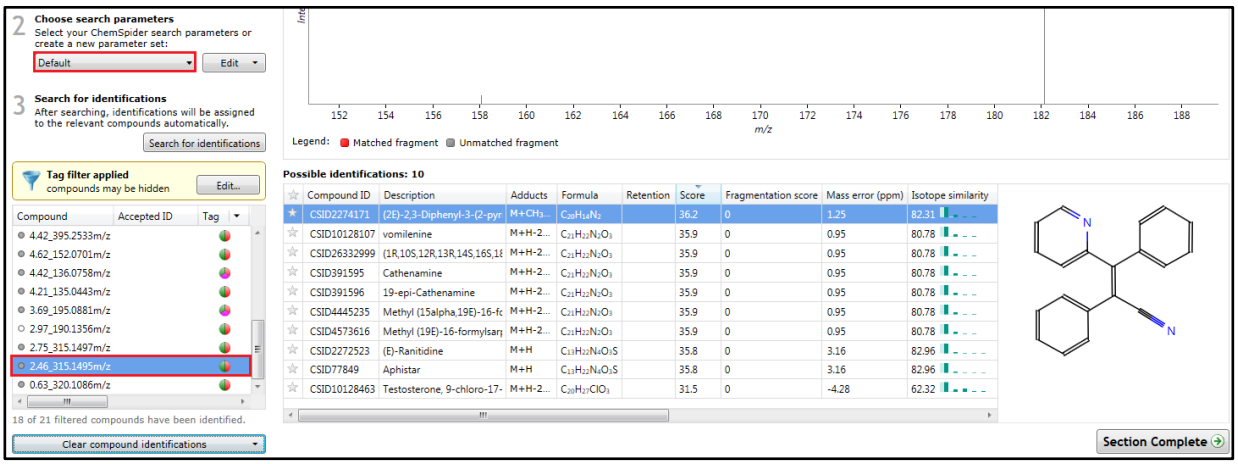


The amended composition appears in the HCN Filter parameter set. Save this to make it current.

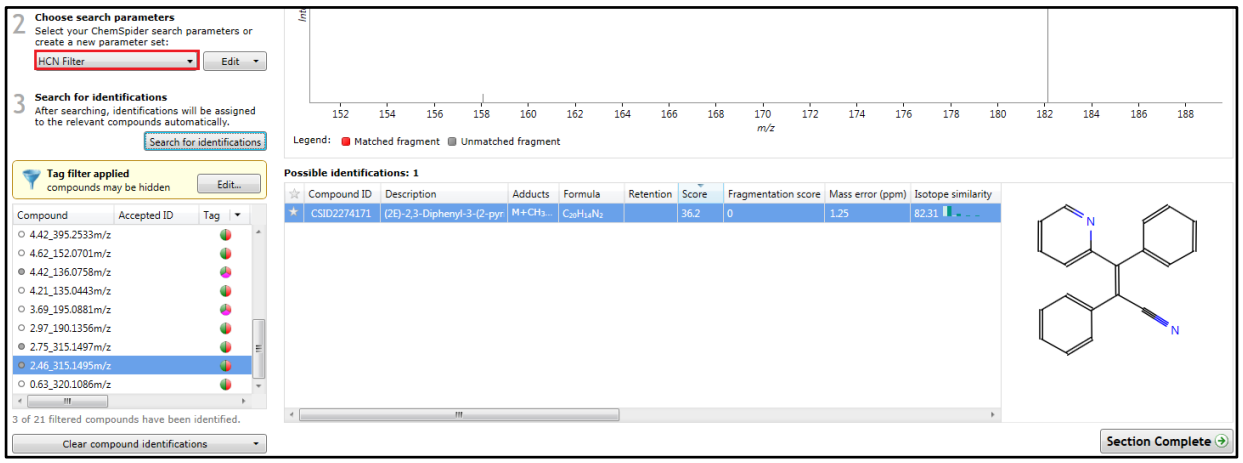
The new set now displays the Elemental Composition Filter



As an example, using the Default search parameters ChemSpider returns 10 possible identities for **Compound 2.46_315.1495**



Clear the compound identifications and change the search parameters to HCN Filter and click **Search for identifications**.



Application of the filter to the ChemSpider results reduces the possible identities to one based on restricting the composition to HCN.

Running a LipidBlast Search

As the samples for this tutorial were prepared from Urine, using LipidBlast as your primary method of compound identification is **not** appropriate.

For details on how to setup and use LipidBlast refer to Appendix 6 (page 104)

Running a NIST MS/MS Search

Select NIST MS/MS library from the Identifications Methods.

Note: to use this compound search method you will need to purchase the NIST MS/MS library. Hence using this section is dependent on you having purchased and downloaded the NIST MS/MS library plugin.

The NIST library search plugin bundles the NIST MS/MS libraries and performs a combination of neutral mass and MS/MS based searches. It uses the same search and scoring algorithms as Progenesis MetaScope.

Firstly, delete all your compound identifications using "**Clear compound identifications**".

Set the required search parameter for the both the Precursor and the Fragment mass error tolerances for matching each compound and its observed fragments, the default tolerance is 5ppm.

Precursor tolerance

This value, in ppm or Da, is the maximum difference between the mass for the compound in the NIST library and the observed mass.

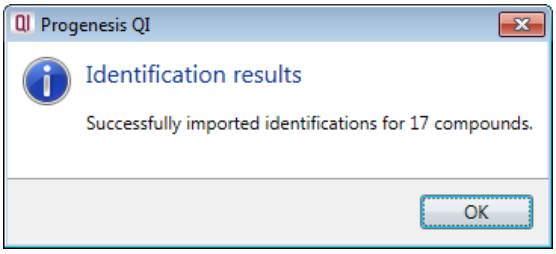
Fragment tolerance

This threshold (in ppm or Da) is the maximum difference between the mass of a fragment in the NIST MS/MS library and the observed mass of the fragment.

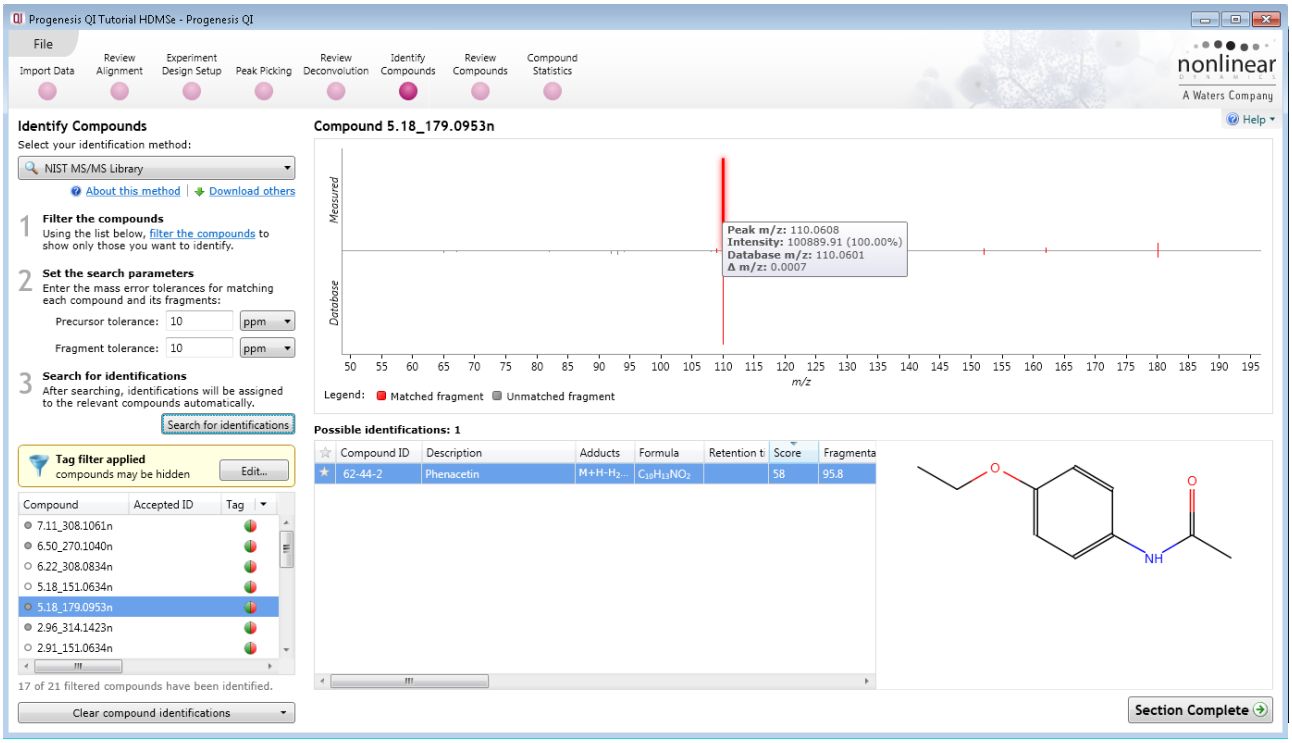
For this example, set both tolerances to 10ppm.

Then click **Search for identifications.**

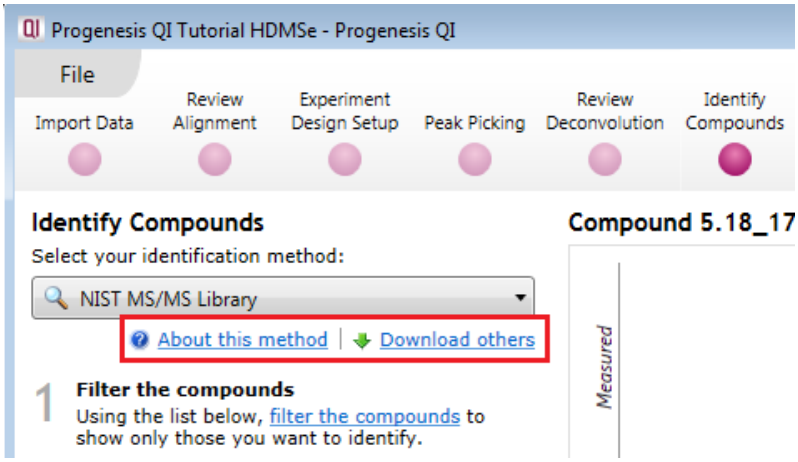
The progress bar reports the number of masses being searched. When the search completes the number of compounds with imported identifications is displayed.



You can visualise the matching of measured to database-predicted fragments for each potential ID using the 'mirror plot' on the same screen, after clicking on a search result. Matched fragments are shown in red, unmatched in grey; and hovering the cursor over a fragment match will display relevant information in a tooltip.



For all the identification methods described in this section additional information and guidance can be obtained **online** in the form of FAQs accessed by the links imbedded in the software.

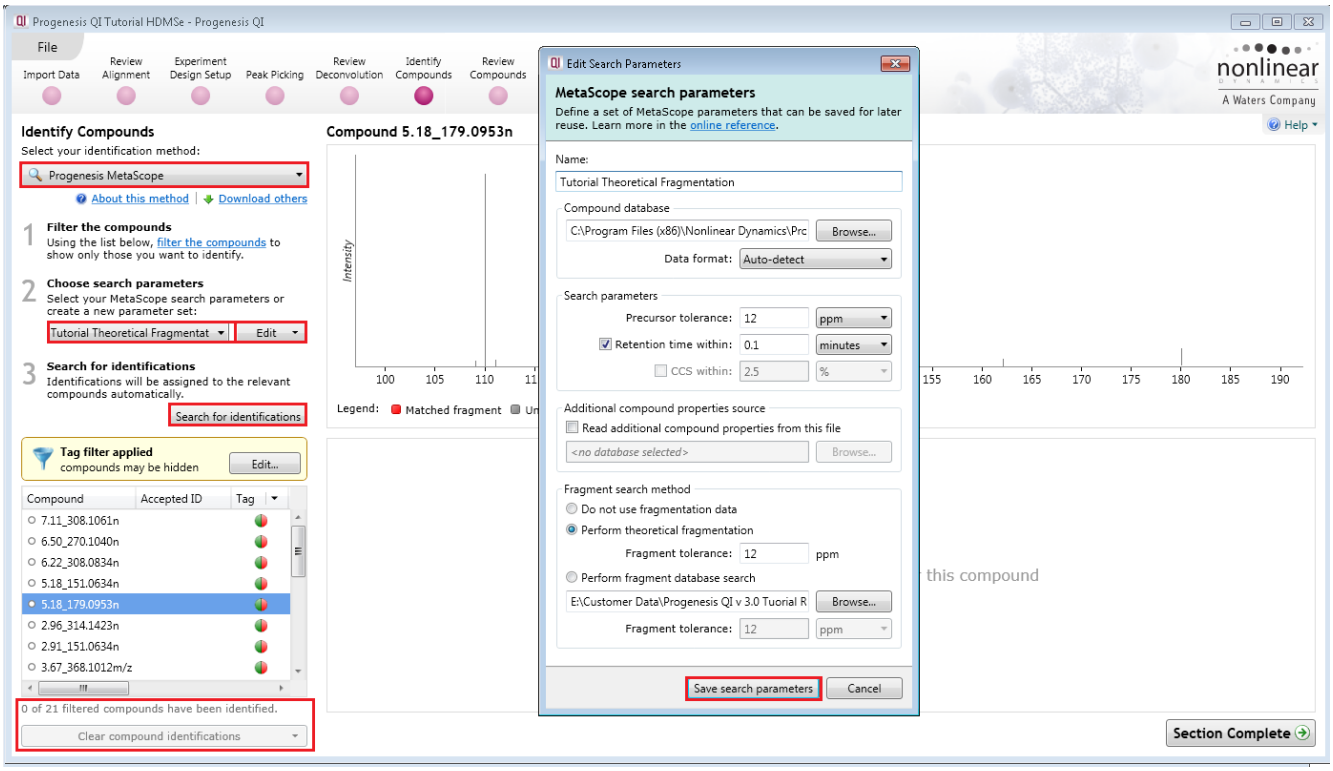


For each identification method the **About this method** link will take you to relevant page(s) for the current method.

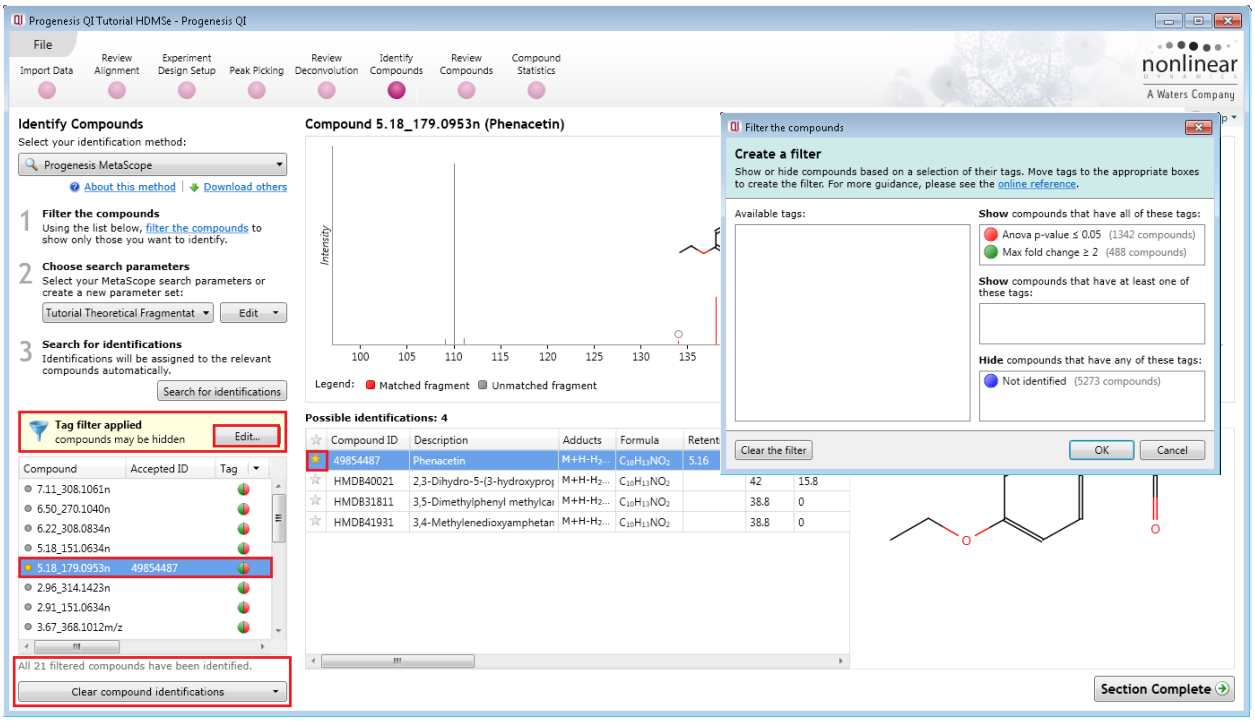
Stage 9: Review Compounds

The Review Compounds stage allows you to examine the behaviour of all or subsets of compounds based on tag filters.

Note: having explored the various methods for identifying compounds, before you use **Review Compounds** return to the Identify Compounds stage and click **Clear compound identifications (if not already clear),** then select Progenesis MetaScope as the identification method, select Tutorial Theoretical Fragmentation as the current search parameters and check they are set as shown below using the **Edit** drop down.

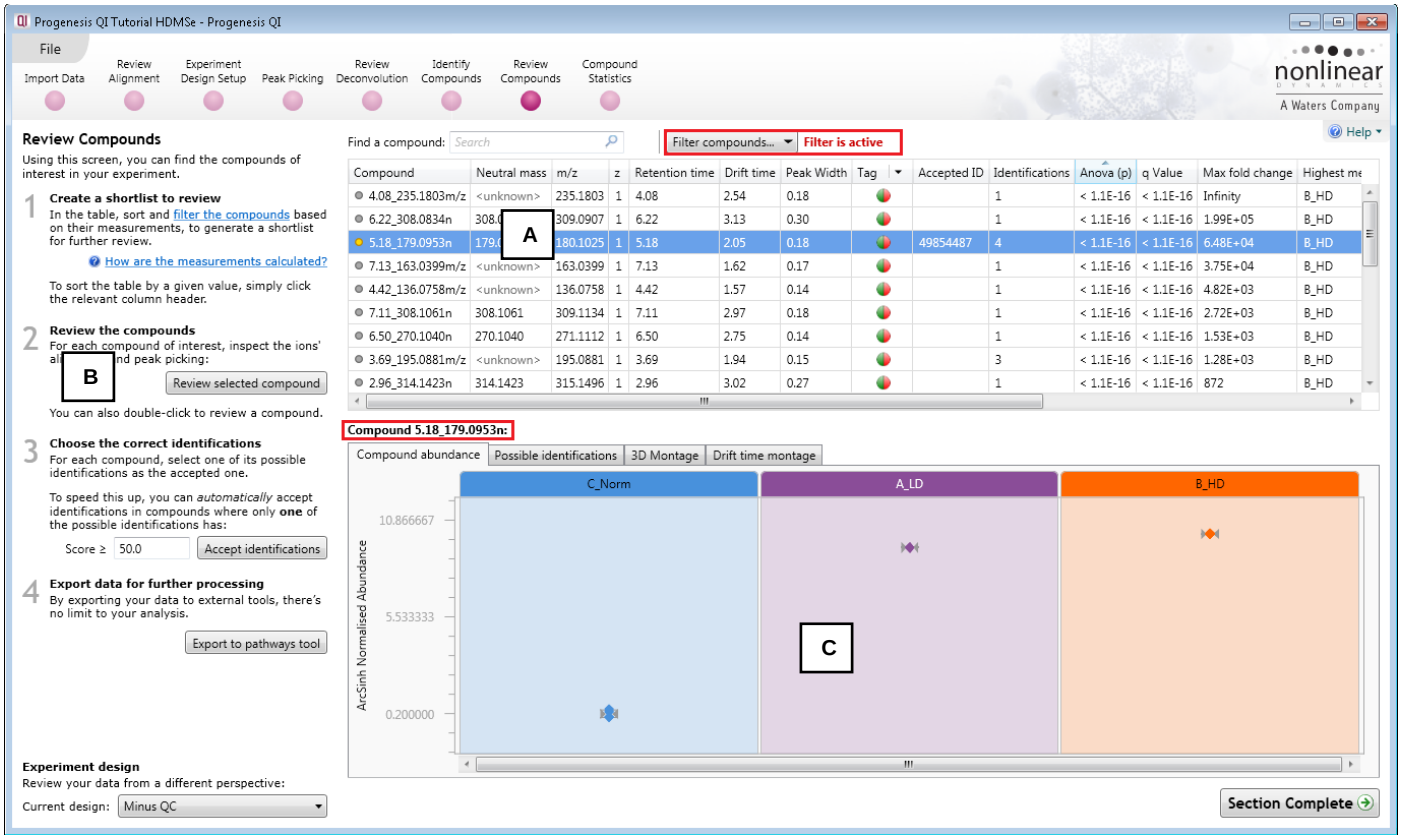


Click **Search for identifications.** Click **Edit** on the applied tag filter the tags created earlier should still be applied, if not set them up as described in tagging page 42. The display should look as below.



Finally to accept Phenacetin as **Compound 5.18_179.0953n's** identity, click on the 'star' to the left of the Compound ID.

Now move to **Review Compounds** by clicking on the Workflow.

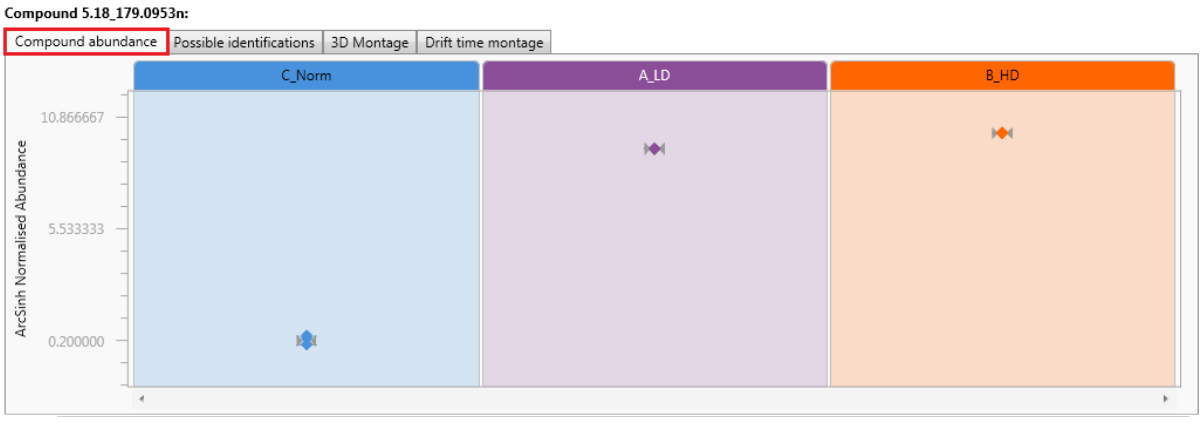


Window A displays the main table of Compounds with identifications as well as those that remain **unknown** after the identifications have been imported.

Window B Provides tools to set thresholds for the acceptance of identifications based on a score.

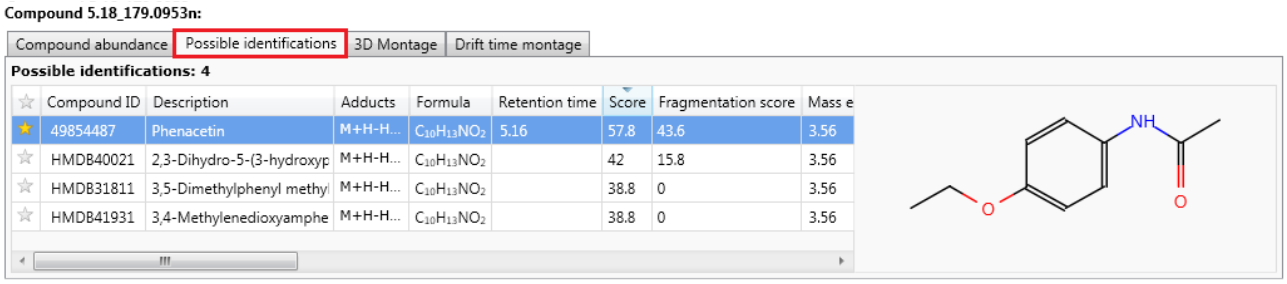
Window C displays either: a **Compound abundance** plot, list of **Possible identifications, 3D Montage**, **Drift Time Montage (for data collected with drift time)**, for the current compound highlighted in Window A.

Using **Compound 5.18_179.0953n** as an example the 4 views for Window C show:

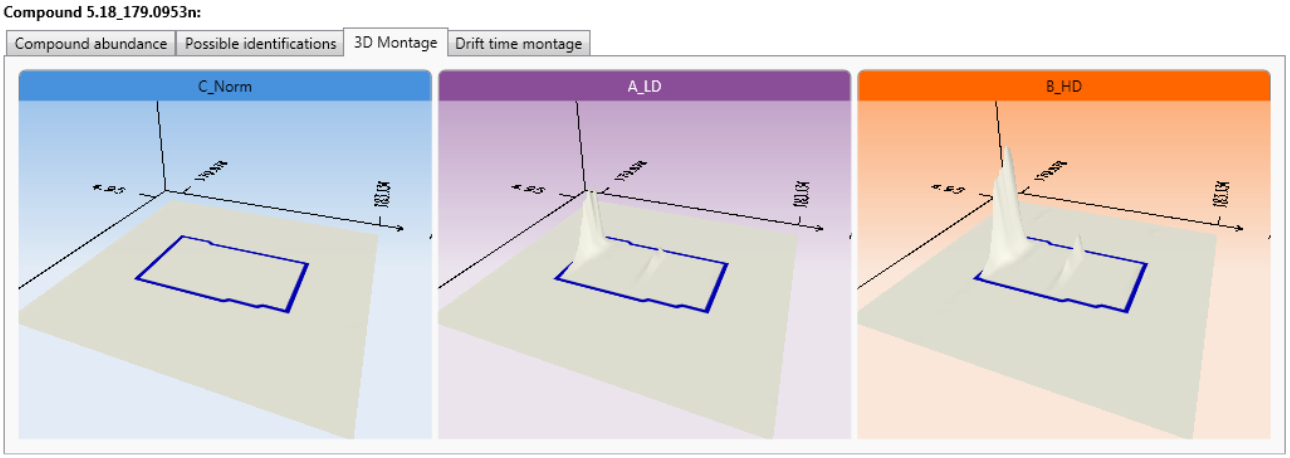


The **Compound Abundance** displaying normalised values for each run, a mean value and 3 standard deviations.

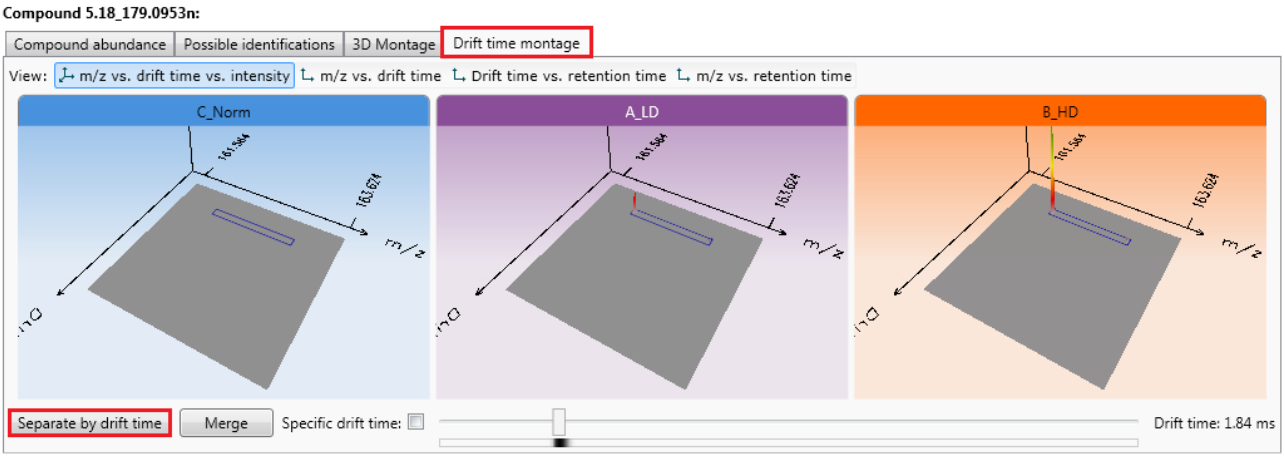
Possible identifications for current compound with Overall Score based on, Mass error, Retention Time (if available), Isotope Similarity based on the comparison of the measured isotope distribution for the compound vs the expected based on the compound formula, Fragmentation Score and CCS if available.



The **3D montage** is based on the first run in each group.



The **Drift time montage** tab is **only** shown for HDMSe data.

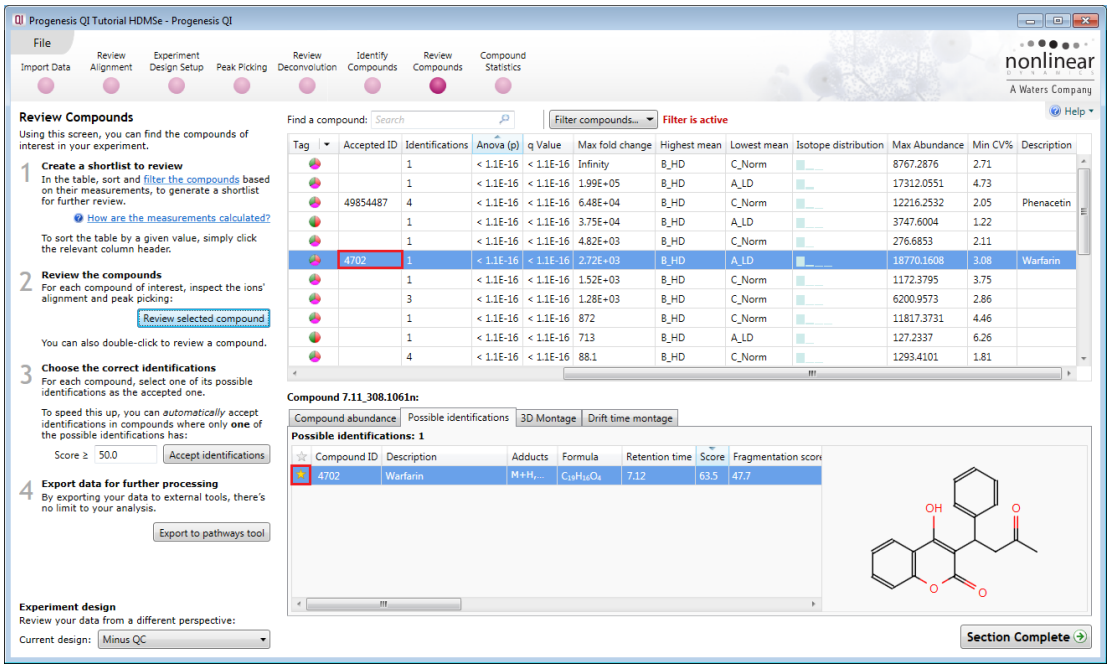


Setting a Compound's identity

If you have not already set or accepted a compounds identity at the **Identify Compounds** stage, or wish to amend it as a result of reviewing the data displayed here then you have the same functionality available to you in the **Possible Identifications** tab

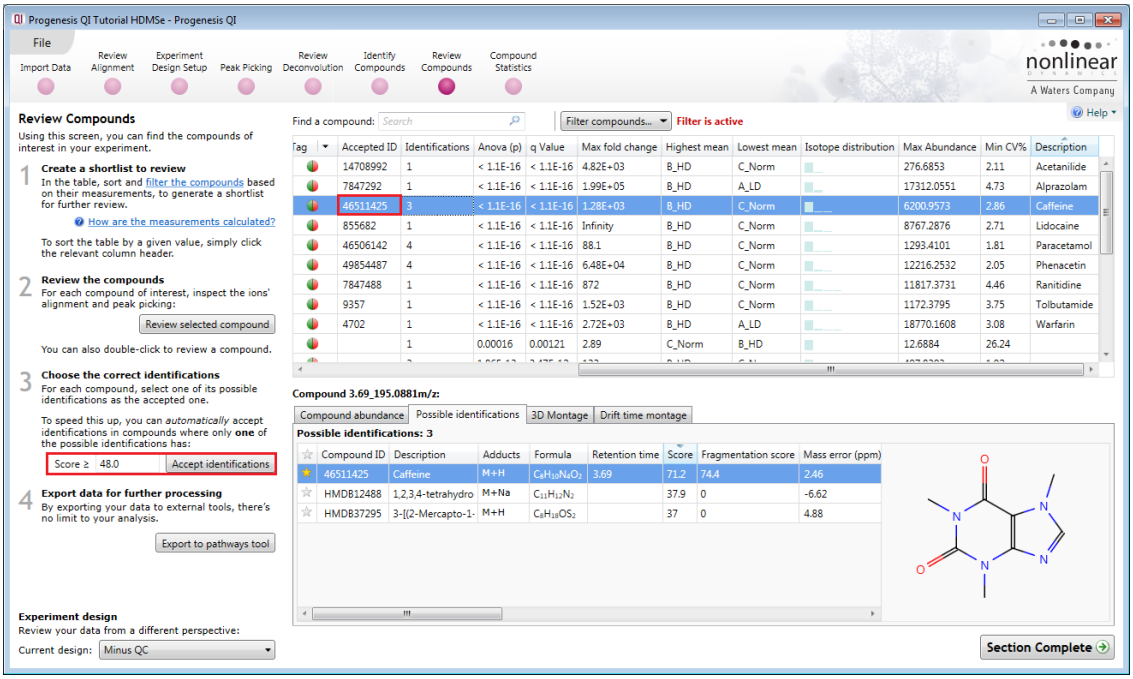
For example, to accept **Warfarin** as the identity for **Compound 7.11_308.1061n** click on the 'star'.

This updates the compound's 'Accepted ID' and Description.



Rather than review all the possible compound identifications you can set a threshold for the score. This allows the automatic acceptance of compound identity where **only** one of the possible identifications is greater than the defined score.

As an example, set the threshold as **48** and then order on **Accepted ID and Description**.



Note: accepting a score threshold causes the Accepted ID and Description fields to be populated when there is a single ID with a score ≥ 48 for the compound.

In this example Paracetamol (2.91_151.0634n) was accepted manually. Also the identification of Aspirin (7.12_163.0399) was rejected by clicking on the star icon for possible identifications.

Compound Validation View

Further information about a compound and it's adduct forms, which enables you to confirm the validity of the measurements, is available in the Compound Validation View.

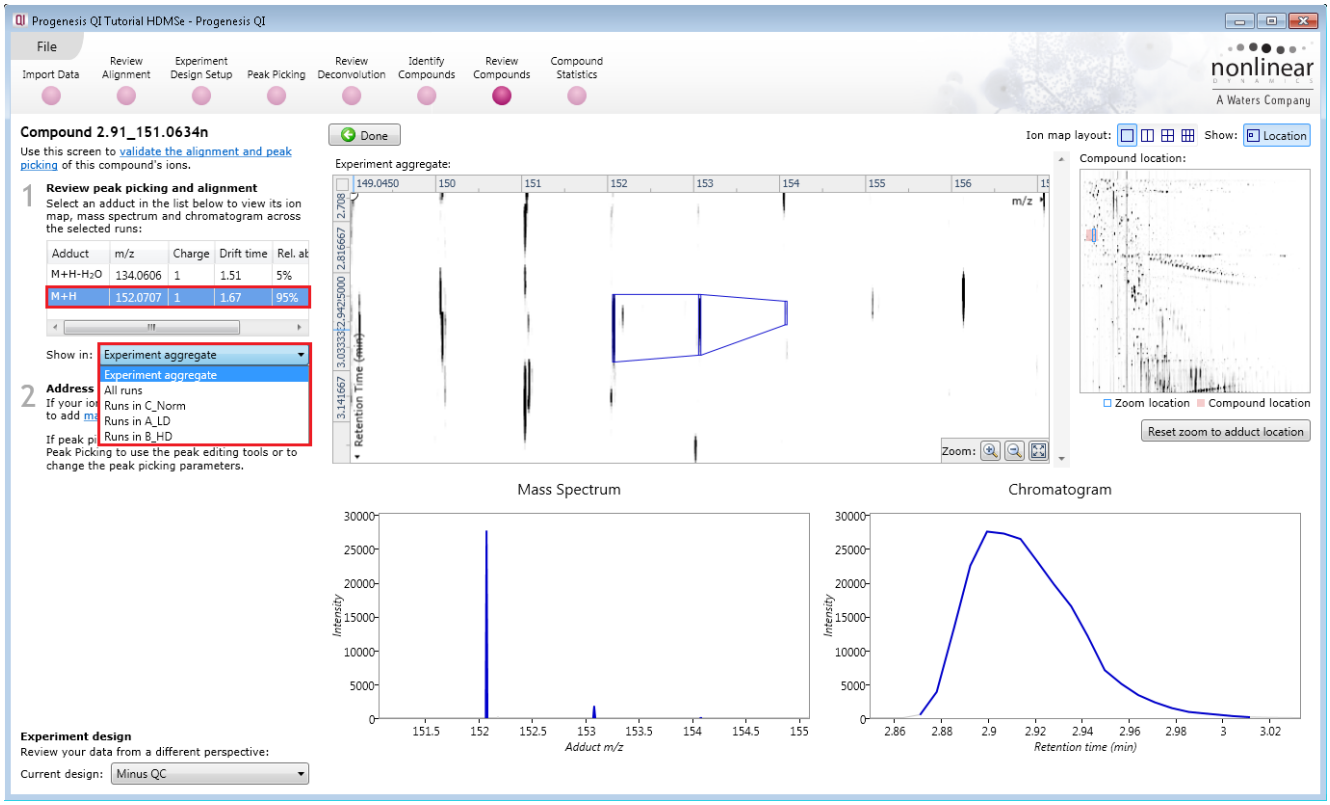
In this view:

- A montage of ion maps helps to validate the ions' peak picking and alignment
- Graphs of mass spectra and chromatograms provide further confirmation of correct peak picking

To open the view either double click on the Compound in the Review Compounds screen or click on Review selected compound

Try double clicking on the compound **2.91_151.0634n** (note: you can accept this compound's identity, as Paracetamol by clicking on the 'star' in the table)

As the **Compound Validation** view opens you can select which adduct of the compound to review first, in this example (M+H).



Use the **Show Location** option at the top right of the screen to view the ion's location within the full ion map. **Note**: you can 'toggle' this view on and off and reset the zoom for the adduct location.

Below the list of adducts, a drop-down list lets you select whether you want to see the selected adduct:

- on the experiment aggregate
- across all runs, or
- only on runs within a single experiment condition

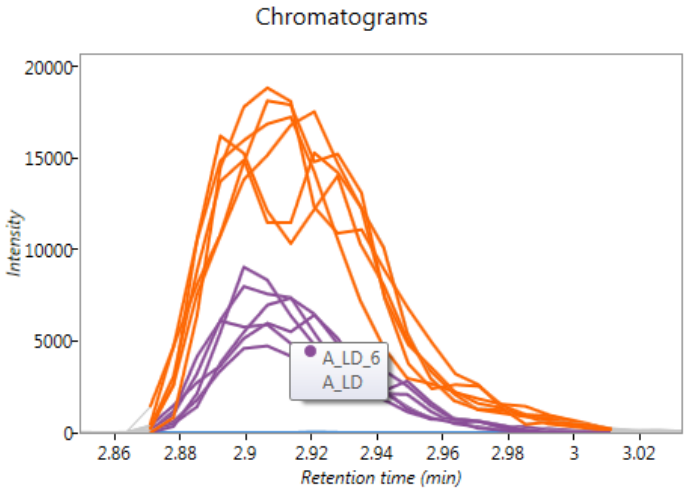
The **Experiment aggregate** option is most useful for validating the ions' peak picking, as the aggregate is generated immediately prior to peak picking and is used as the input to that process.

To validate the alignment you will find it best to select to select the **All runs** option and then adjust the ion map layout to display multiple runs

Use the Ion map layout and zoom tools to set up the views as shown above. The multi-panels allow you to confirm the alignment across all the runs.

Note: that when the **All runs** are selected the mass spectra and chromatograms are shown for all the runs in the experiment.

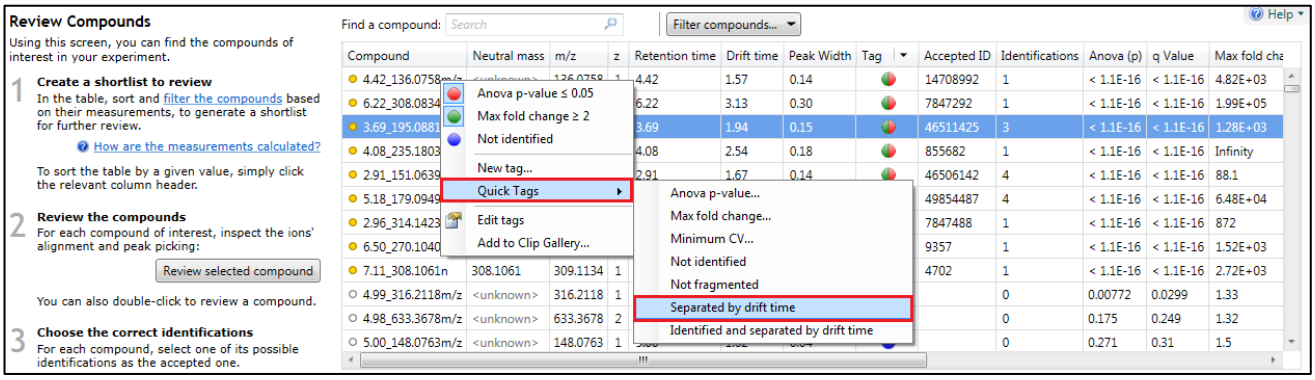
In the chromatograms view each line corresponds to a single run, coloured according to the experiment condition to which it belongs. If you hover the cursor over a line, the name of the run which generated it will appear. When showing all runs, these graphs give further validation of how the adduct form's abundance is changing between experimental conditions.



Compounds separated by Drift Time

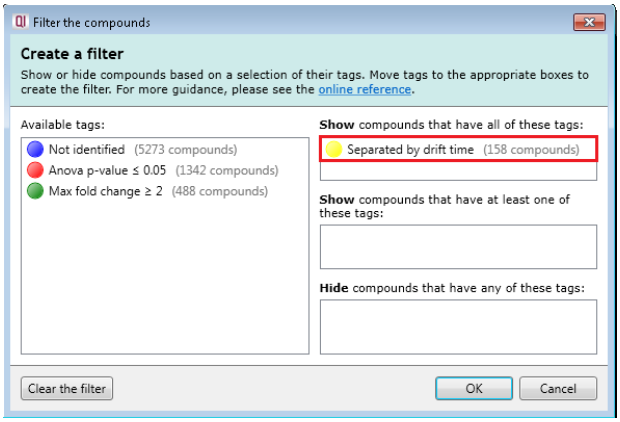
To view compounds separated by drift time, first clear the existing Filter by clicking on **Filter compounds** and then **Clear the filter.**

Then right click on the Review Compounds Table and create a Quick Tag for those compounds Separated by drift time.

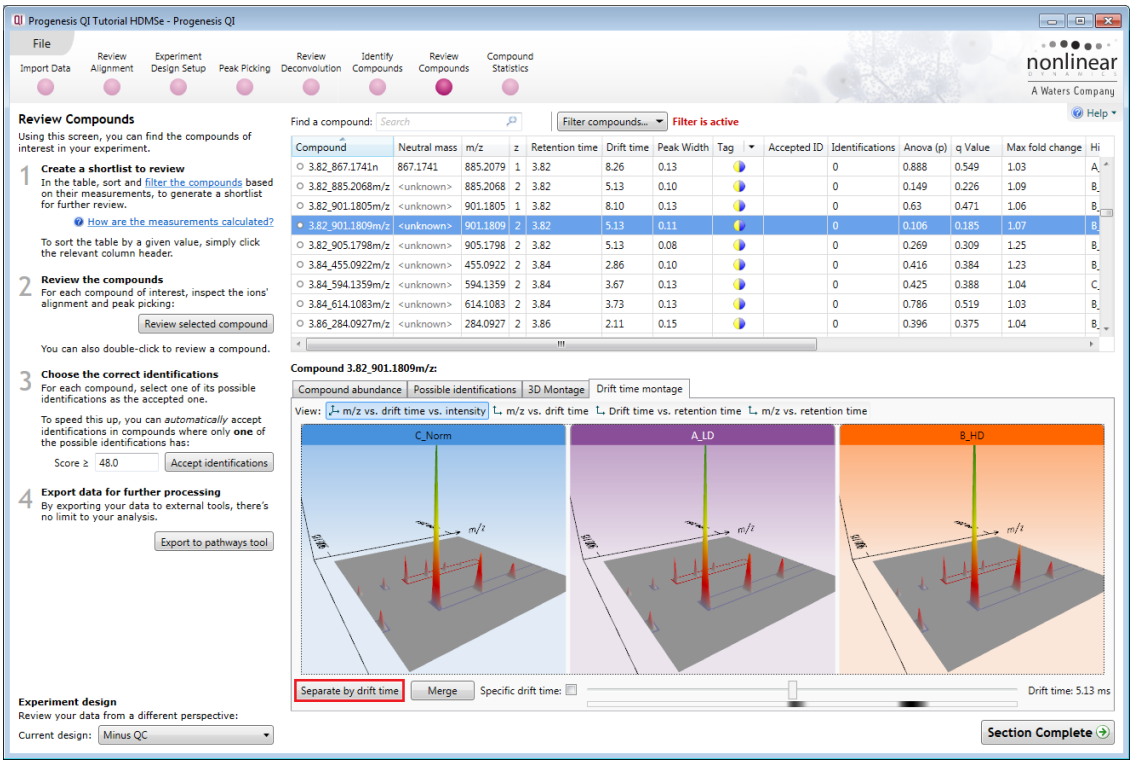


Once you have created the tag then filter the table to show only those compounds Separated by Drift Time.

To do this click on **Filter compounds**… and drag the new Tag on to the **Show** panel and click OK.

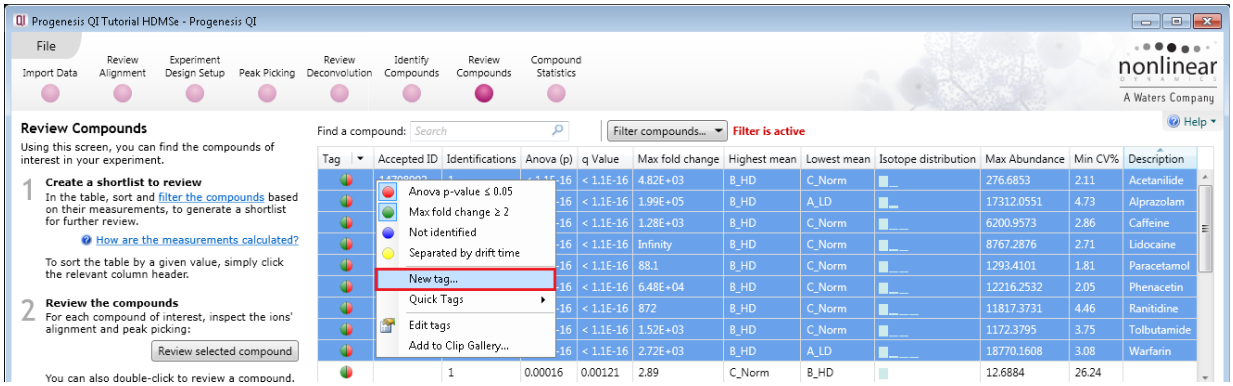


The table now shows the only the compounds that are separated by drift time. Click on **Separate by drift time** to view the detected compound ions with the same m/z and RT.

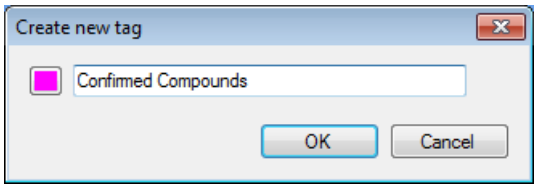


Exporting compound data

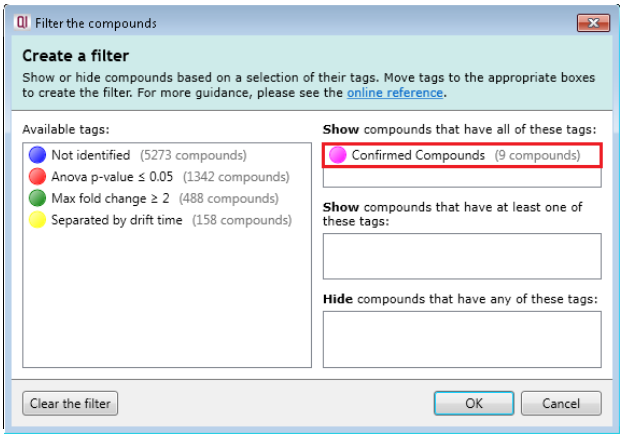
Compound data can be exported in a csv file format. You can either export the **compound measurements** or the **compound identifications.** As an example of data export, first clear any existing filters and then order the table on **Description** and highlight all the compounds which have a description.



Then right click on the highlighted compounds and create a **New tag…** called **'Confirmed Compounds'**

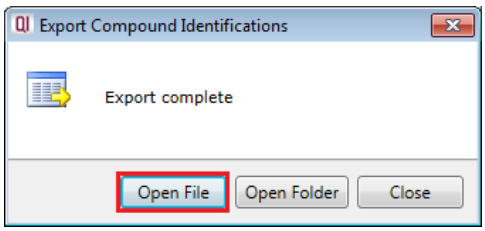


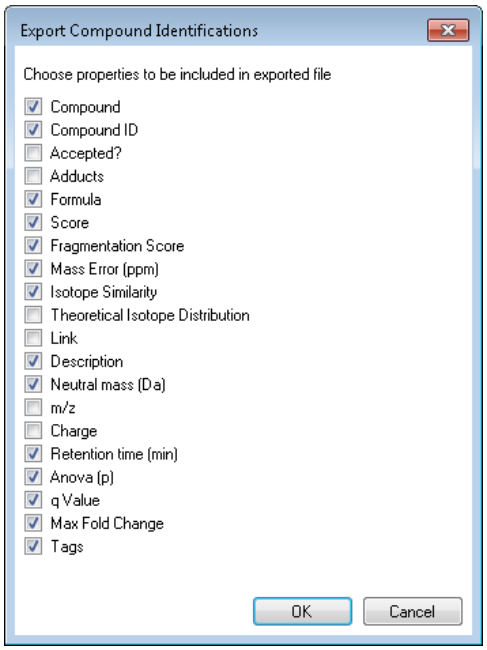
Click on Filter Compounds then drag the new tag onto the Show panel of the **Filter compounds** dialog and click **OK**.



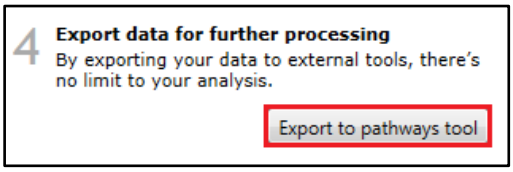
Click on the File menu and select **Export compound identifications**, adjust the properties to be included in the export and click OK.

When you have saved the file, a dialog opens allowing you to open the file if required:





Exporting Compound Data to Pathways Tool(s)



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

Currently Progenesis QI supports the export to:

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

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

Support for the export of compound identity and expression data to either pathways tool is provided as standard.

Note: the use of IPA requires a local licensed copy, for licensing details please visit the IPA website.

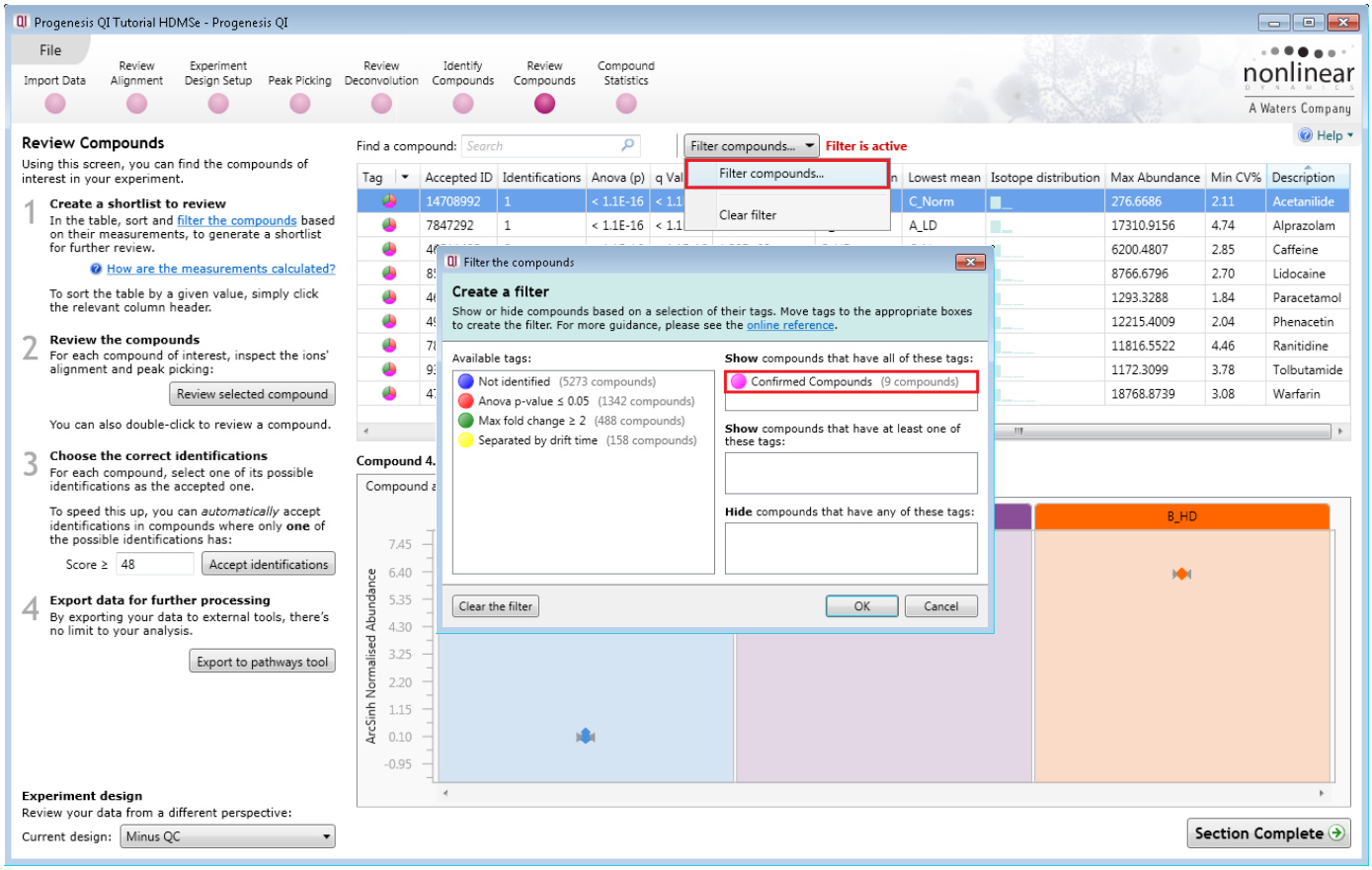
The following section describes an example of the processes involved in exporting to and using **IMPaLA**. For additional details on the export of data to IPA refer to Appendix 8 (page 106)

Using IMPaLA

Typically, only accepted identifications are exported to pathway analysis tools. This is the default and recommended option.

First select the compound data to export to the pathways tool using tag filtering to 'focus' the set to export. For this example you can use the **'Confirmed Compounds'** set described in the previous section.

To ensure that the 9 Confirmed Compounds are the only ones displayed in the table set up the filter as shown below:

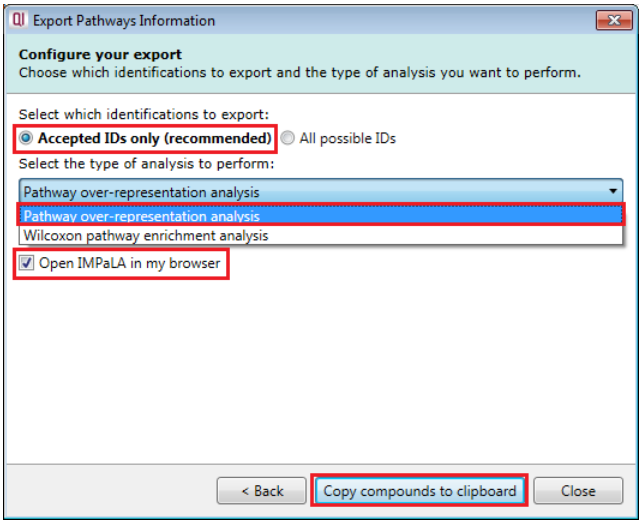


Then click **Export to pathways tool**, then select IMPaLA from the Pathways tools and click **Next.**

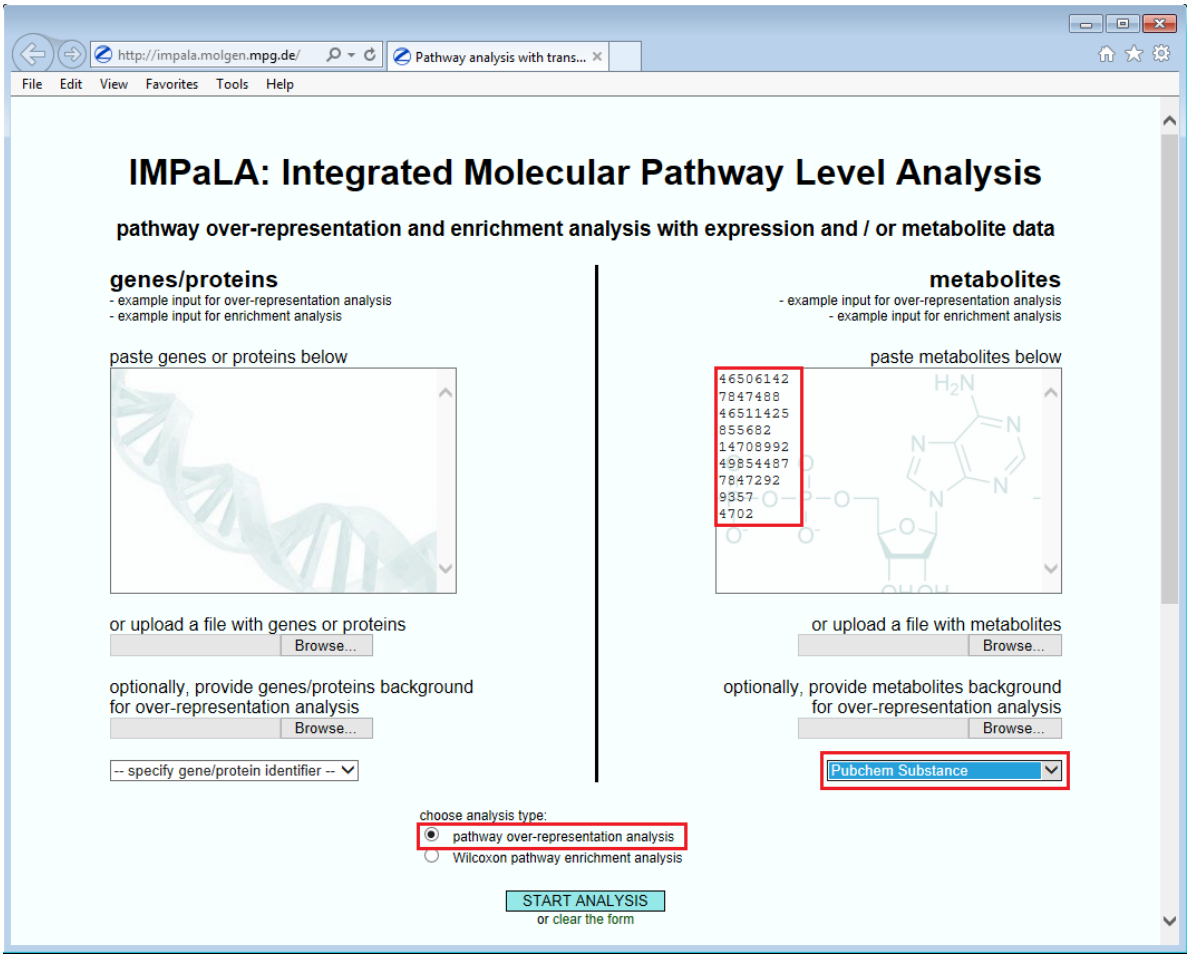
Ensure that Accepted IDs only is selected.

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

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



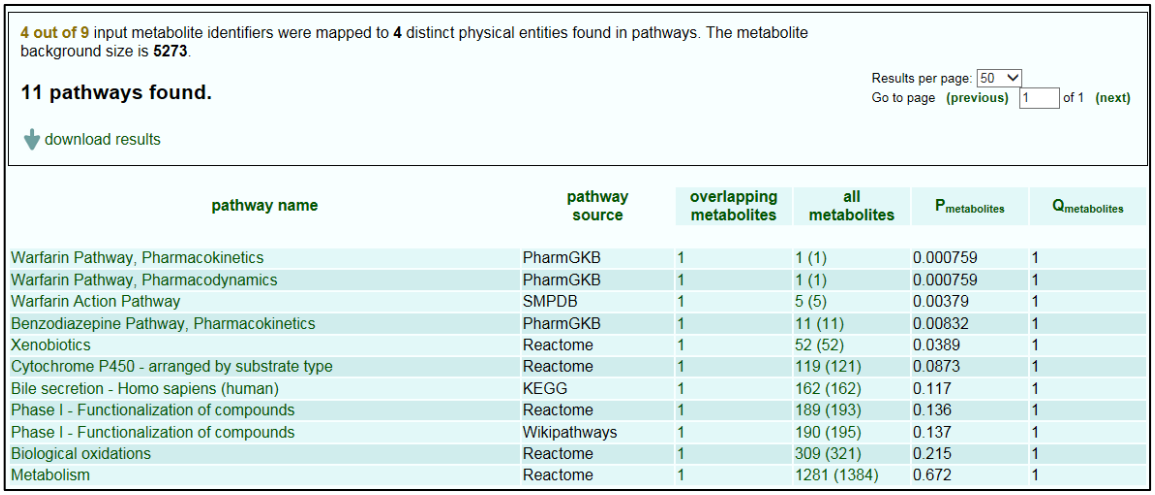
When IMPaLA opens, paste the metabolites into the right hand panel.



Select specific metabolite identifier

Click **Start Analysis**

Analysis results list is returned:



4 out of 9 input metabolite identifiers were mapped to **4** distinct physical entities found in pathways. The metabolite background size is **5273**.

11 pathways found.

Results per page: 50
Go to page (previous) 1 of 1 (next)

download results

pathway name	pathway source	overlapping metabolites	all metabolites	$P_{metabolites}$	$Q_{metabolites}$
Warfarin Pathway, Pharmacokinetics	PharmGKB	1	1 (1)	0.000759	1
Warfarin Pathway, Pharmacodynamics	PharmGKB	1	1 (1)	0.000759	1
Warfarin Action Pathway	SMPDB	1	5 (5)	0.00379	1
Benzodiazepine Pathway, Pharmacokinetics	PharmGKB	1	11 (11)	0.00832	1
Xenobiotics	Reactome	1	52 (52)	0.0389	1
Cytochrome P450 - arranged by substrate type	Reactome	1	119 (121)	0.0873	1
Bile secretion - Homo sapiens (human)	KEGG	1	162 (162)	0.117	1
Phase I - Functionalization of compounds	Reactome	1	189 (193)	0.136	1
Phase I - Functionalization of compounds	Wikipathways	1	190 (195)	0.137	1
Biological oxidations	Reactome	1	309 (321)	0.215	1
Metabolism	Reactome	1	1281 (1384)	0.672	1

Click to explore returned analysis.

Note: the success of a Pathways analysis is dependent on the organism under study being available to search using **IMPaLA.** This data set is not truly a suitable example but returns results that at least demonstrate the process.

Now move to the **Compound Statistics** section by clicking section complete.

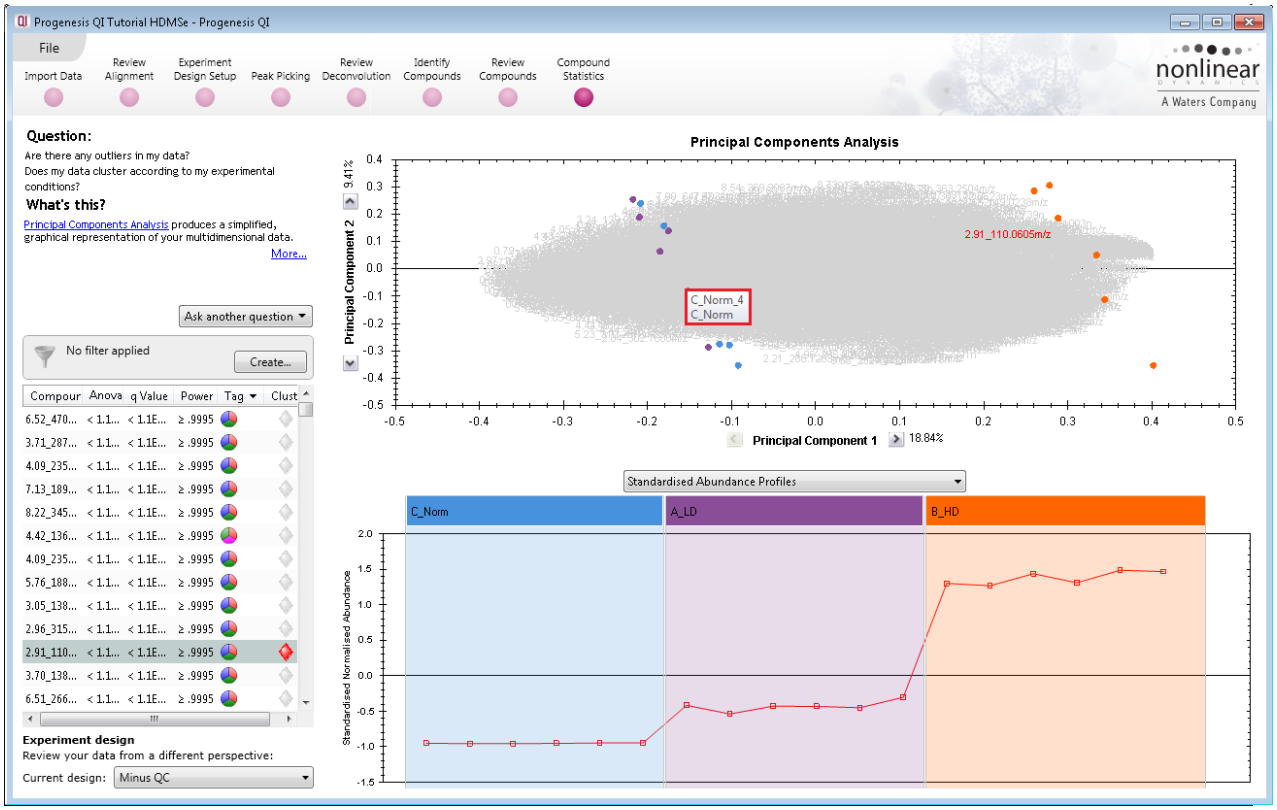
Stage 10: Compound Statistics

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

Compound Statistics opens calculating the Principal Components Analysis (PCA) for the active 'tag' filter if one exists.

Note: make sure you have cleared any Tag Filters before performing Principal Components Analysis (PCA)

As an example, we will start by examining the PCA for all 5345 compounds.



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

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

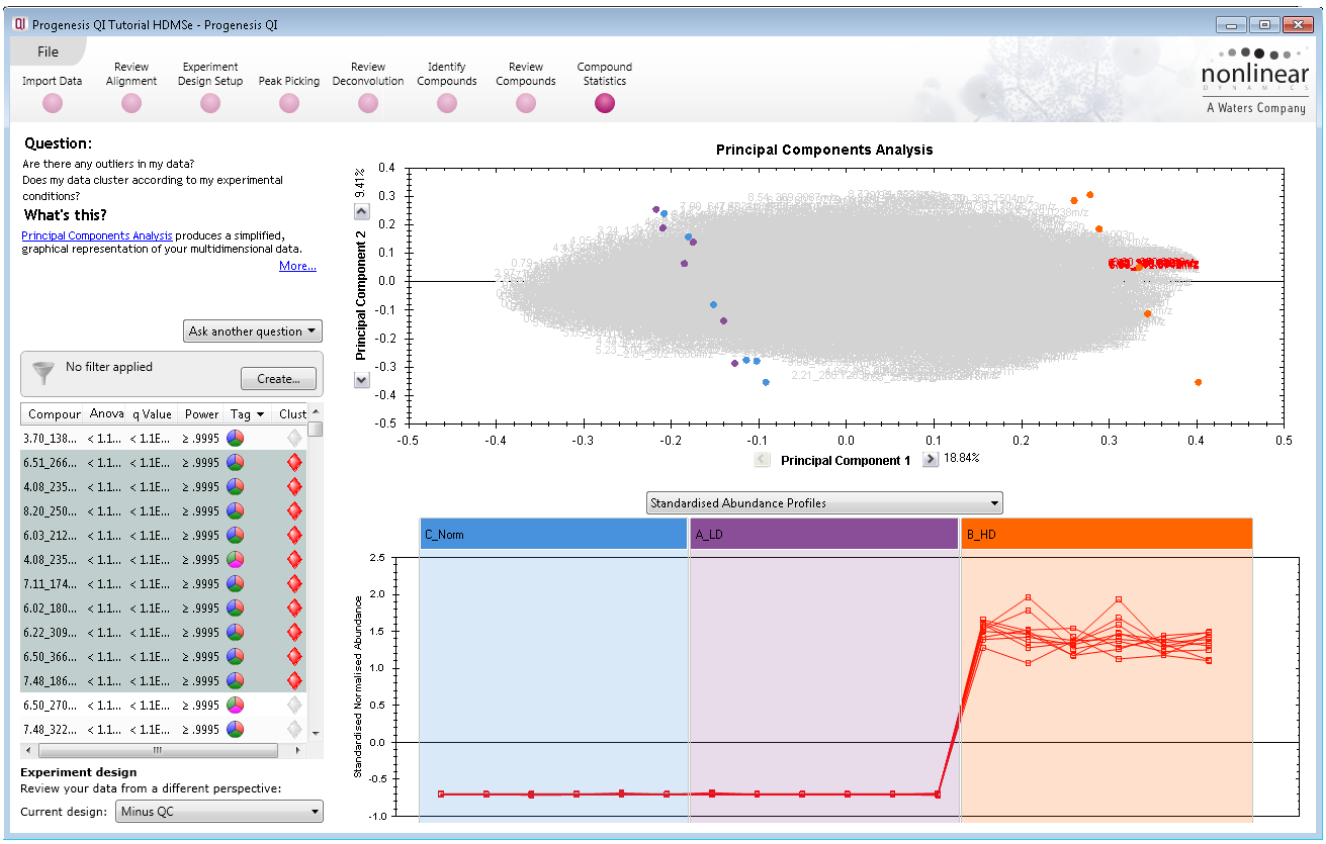
Principal Component Analysis (PCA)

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

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

It answers this question by: *'Using Principal Components Analysis (PCA) to produce a simplified graphical representation of your multidimensional data'.*

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



Note: the table in the Statistics view contains additional columns:

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

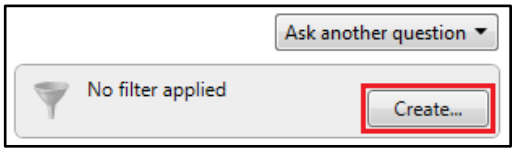
Power: can be defined as the probability of finding a real difference if it exists. 80% or 0.8 is considered an acceptable value for power. The Power Analysis is performed independently for each compound, using the abundance variance, sample size and difference between the means.

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

Note: Power analysis is discussed in Appendix 8 (page 109)

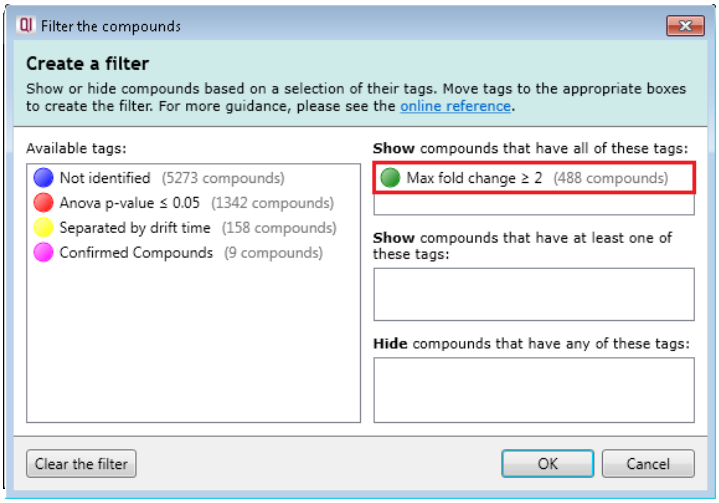
Correlation Analysis

Use the tags created in Review Compounds to filter the compounds displayed in the table. We are going to explore the Correlation Analysis for all the Compounds (with possible identifications) that display a significant 2-fold or greater difference in abundance.



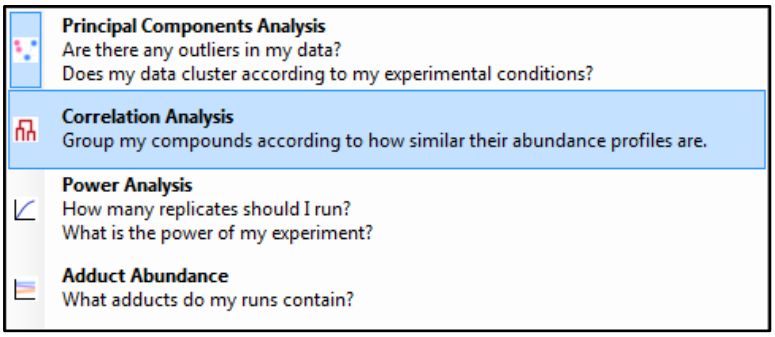
To filter the data click **Create**

Set up the filter as shown below



On pressing OK the PCA will recalculate.

To set up the **Correlation Analysis** using this filtered data set click on **Ask another question** (above the table). A selection of 4 tools will appear in the form of questions



Select the second option to explore 'feature correlation based on similarity of abundance profiles'

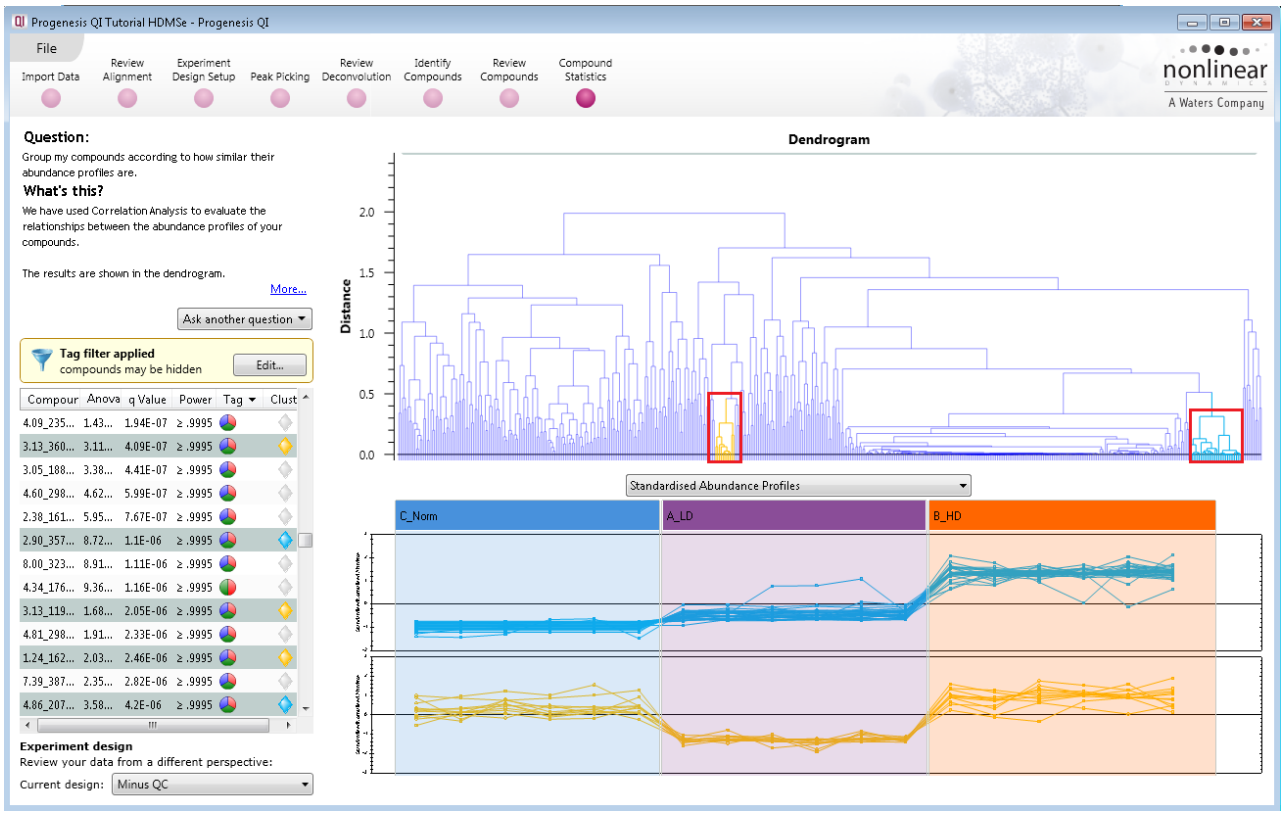
This time the statistically based question(s) being asked is:

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

The question is answered by:

'Using Correlation analysis to evaluate the relationships between the (selected) compounds' expression profiles'.

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

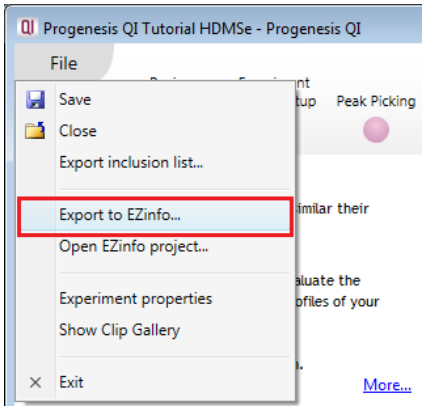


Correlation Analysis enables the grouping of compounds together according to how similar their abundance profiles are.

Clicking on a branch on the Dendrogram selects the compounds on the table. You can then tag this group, of 'potentially' related compounds, by right clicking and creating a 'New tag' for them. Then use the tag to focus on them at the **Review Compounds** stage.

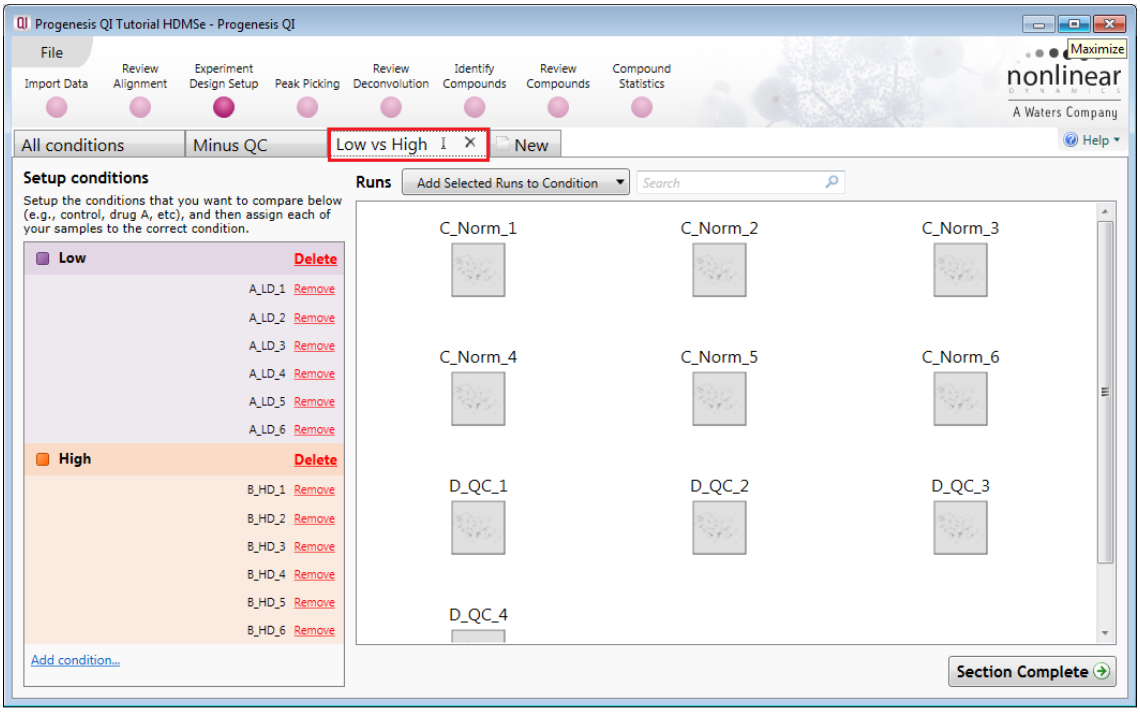
Exporting to EZinfo

To export data to EZinfo select the option from the file menu.

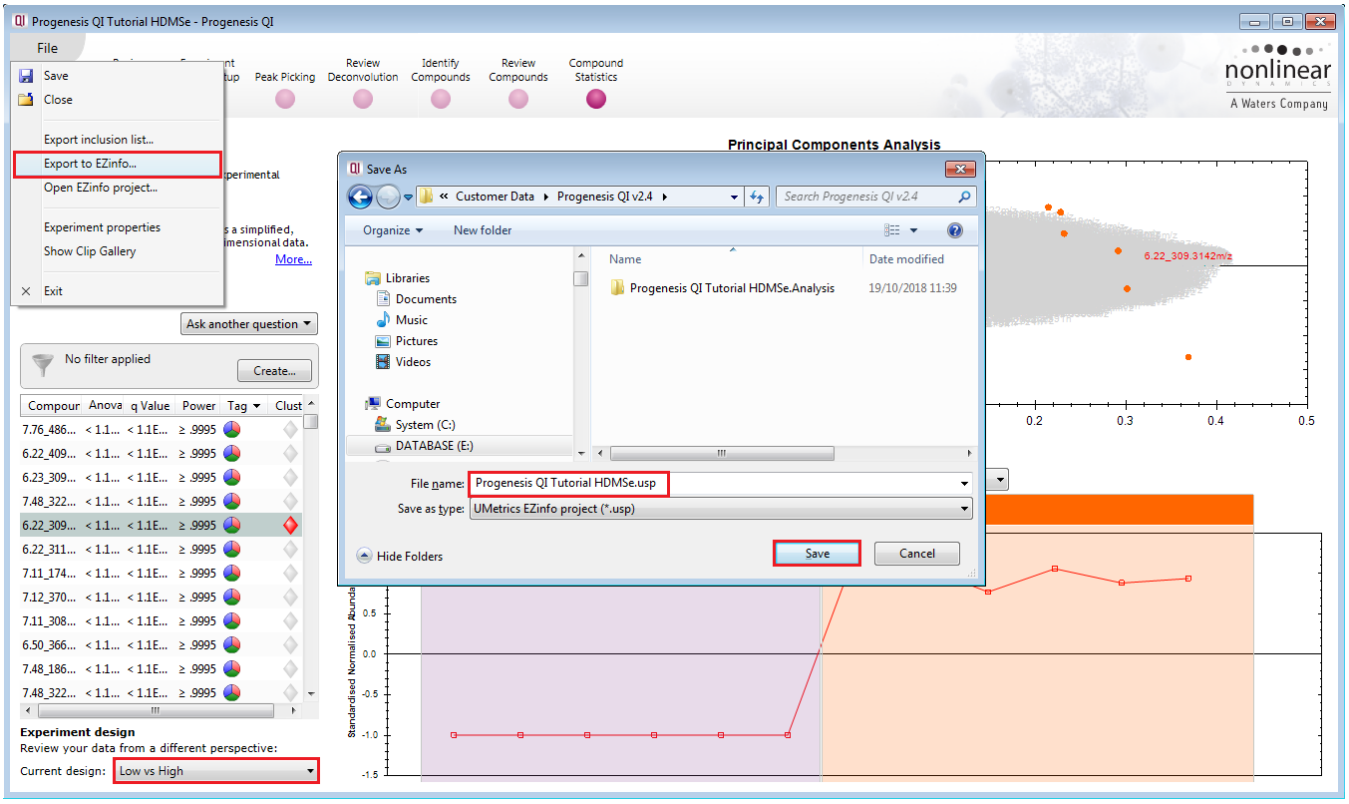


Note: EZinfo only appears in the Progenesis QI File menu if it has been installed. Therefore, this section is only relevant to users who have access to EZinfo.

For this example, before exporting to EZinfo, please ensure that you have created an Experimental Design Setup to compare the A_LD and B_HD conditions. To do this, return to Experiment Design Setup set up and create an additional design for 'Low vs High' as shown below.



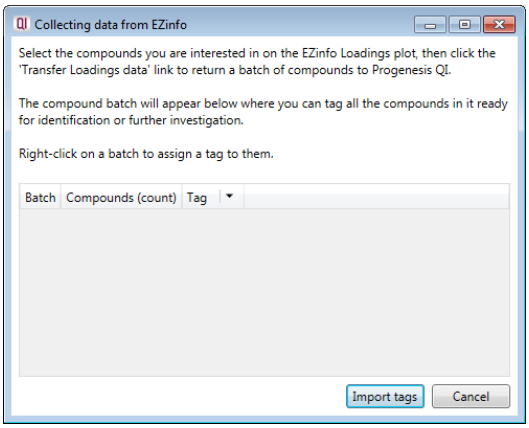
Return to **Compound Statistics,** make sure that there is no Filter applied, then select **Export to EZinfo.** You will be asked to save the project before EZinfo opens.



Save the EZinfo project in the same folder as your Progenesis QI experiment.

A dialog for **Collecting data from EZinfo** opens in Progenesis QI and EZinfo opens displaying the exported data.

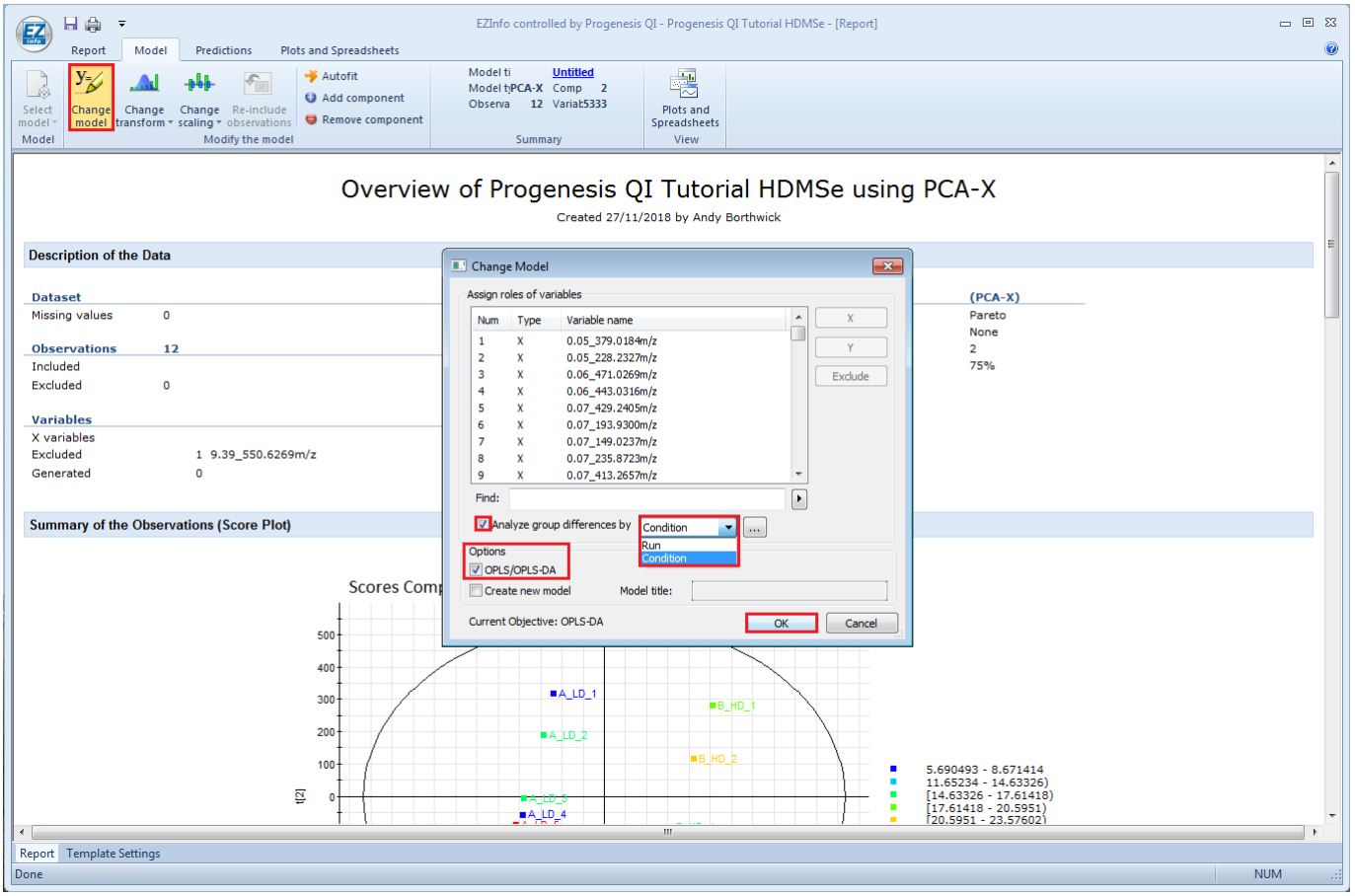
Note: the dialog may be obscured by the EZinfo application as it opens.



Ensure that Pareto Scaling is currently selected, if it is not current, then use **Change Scaling** in the 'Model tab' to set it to Pareto.

As we have specifically exported data to EZinfo for the comparison of 2 conditions we can generate an S-plot from a discriminate analysis (OPLS).

To do this in EZinfo with the exported data, first change the model of the data by selecting **Change Model**

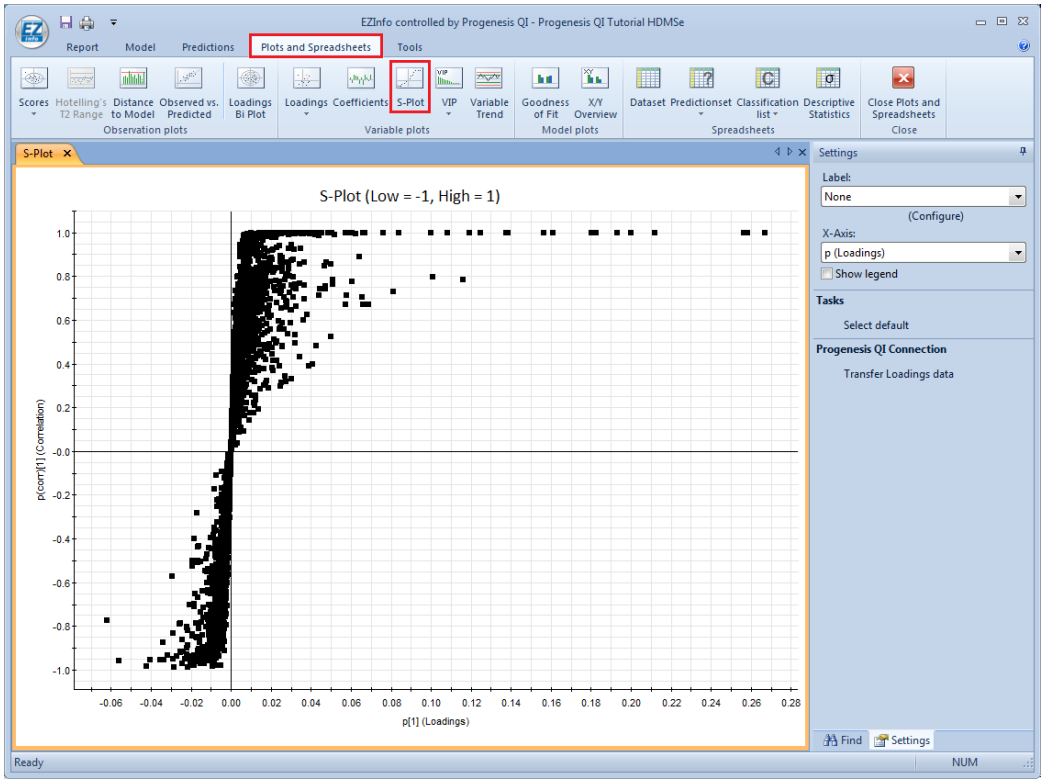


Select analyse group differences by **Condition**.

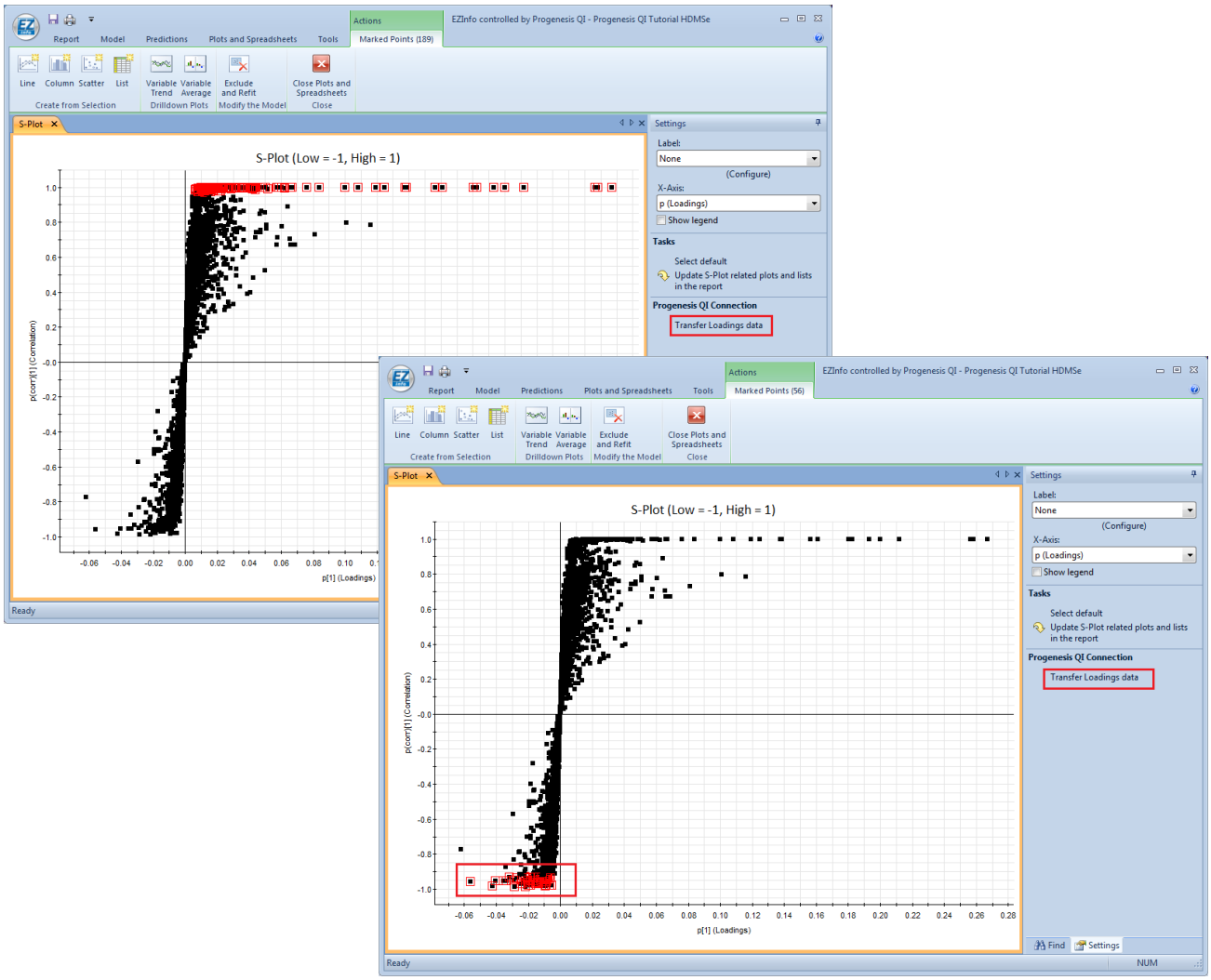
Set the options to OPLS/OPS-DA and click OK

The PCA will update the new model.

Now Select **Plots and Spreadsheets** on the main tool bar and click on the **S-Plot**

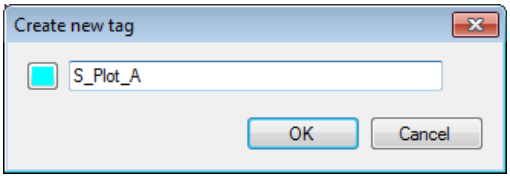


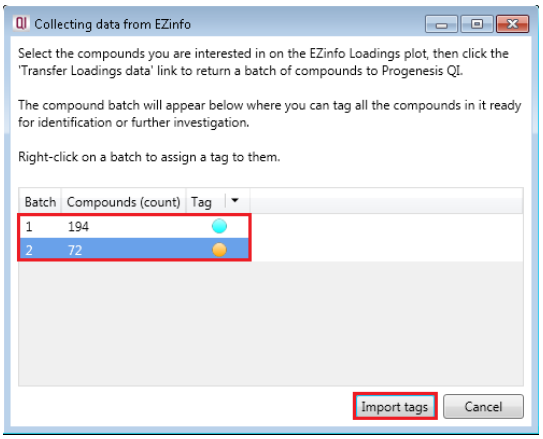
Select a set of points and click on **Transfer Loadings data**. Repeat this process for other sets of data.



As you transfer each group of selected data from the loadings plot a new batch appears in the **Collecting data from EZinfo** dialog.

Select and right click on each batch in turn and create a **New tag… ,** naming it appropriately.

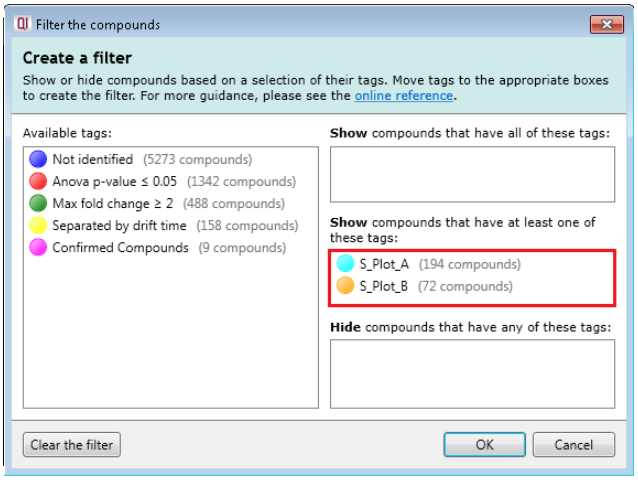




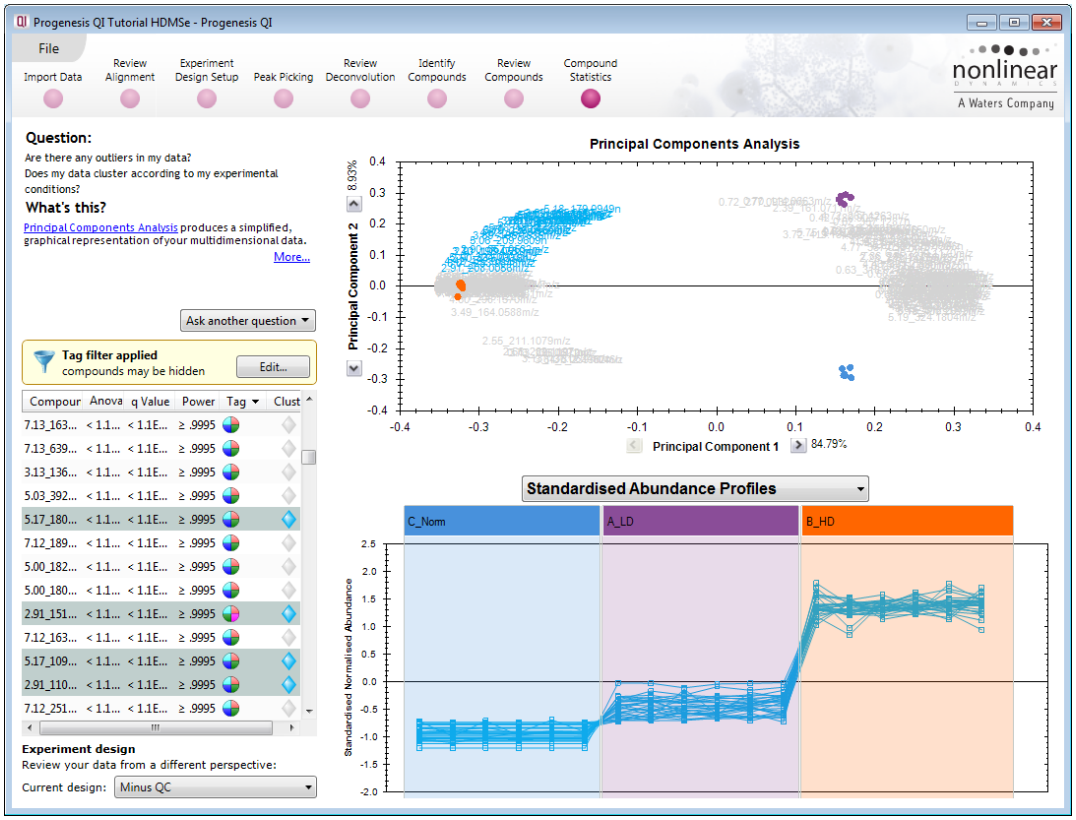
Click **Import tags**

Close the dialog and EZinfo. The tags will now be available in to use in Progenesis QI.

Set a new Tag Filter in **Compound Statistics** to **Show** compounds that have at least one of these tags.



Click OK to create the new PCA based on these tags.

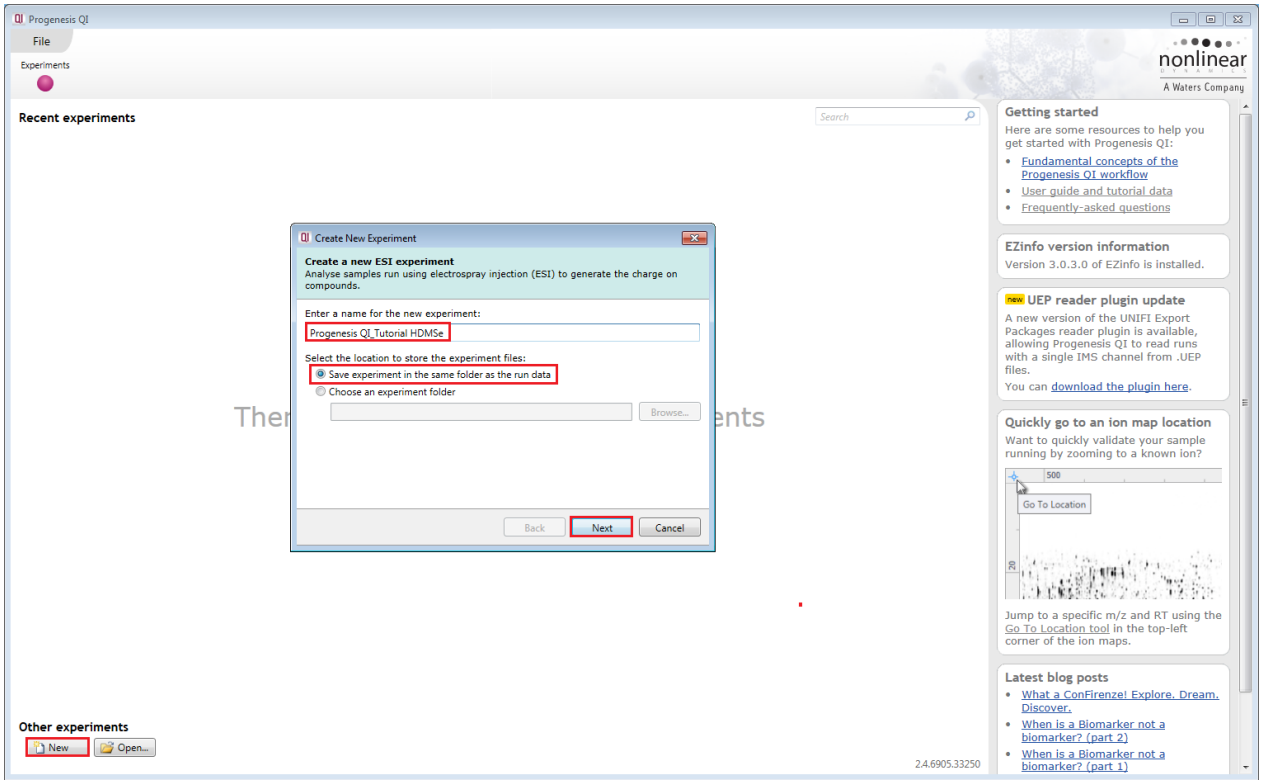


Note: as you close Progenesis QI, EZinfo will also close.

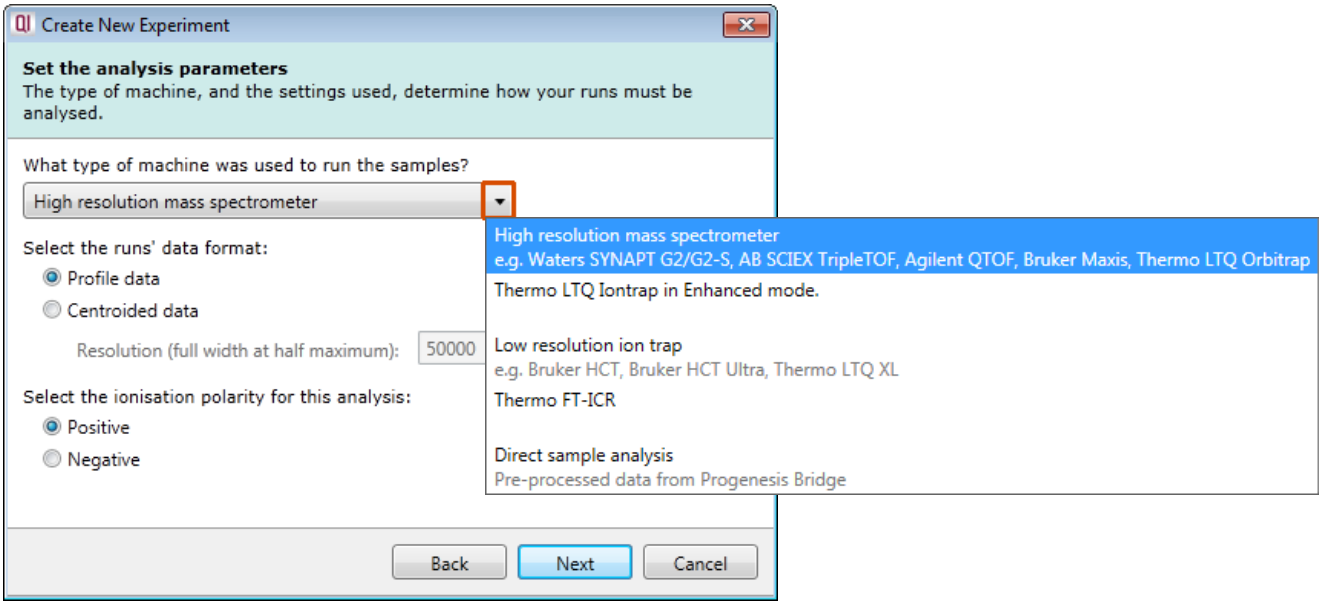
Appendix 1a: Stage 1 Data Import and QC review of data set

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

To create a new experiment with your files select **New** give your experiment a name and then select a location to store the experiment files. Click **Next**.

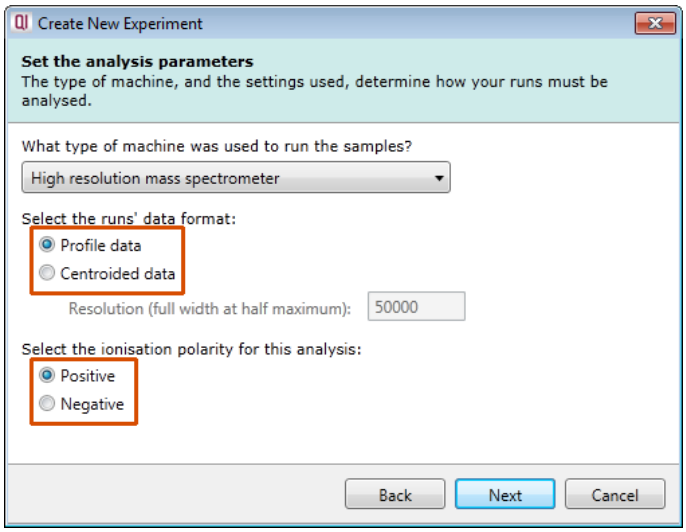


Now select the machine type that was used to run the samples.

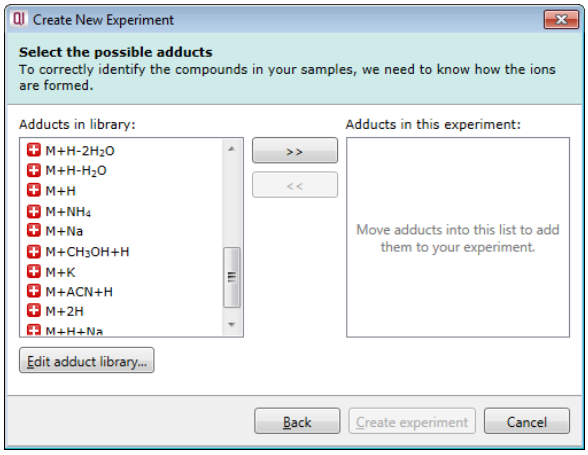


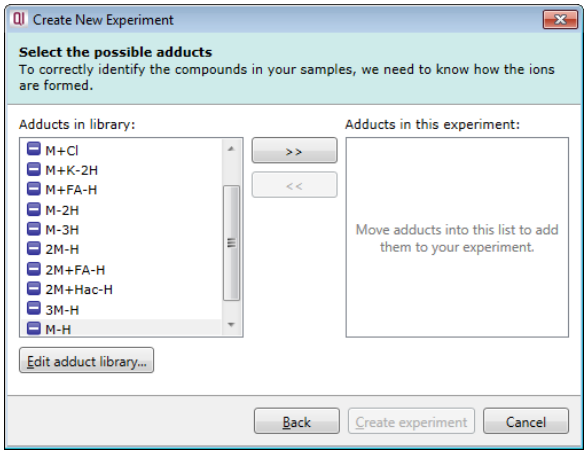
Note: if you have converted or captured the data as centroided then select Centroided data and enter the Resolution (full width at half maximum) for the MS machine used.

In this example the data was captured in **Profile** mode.

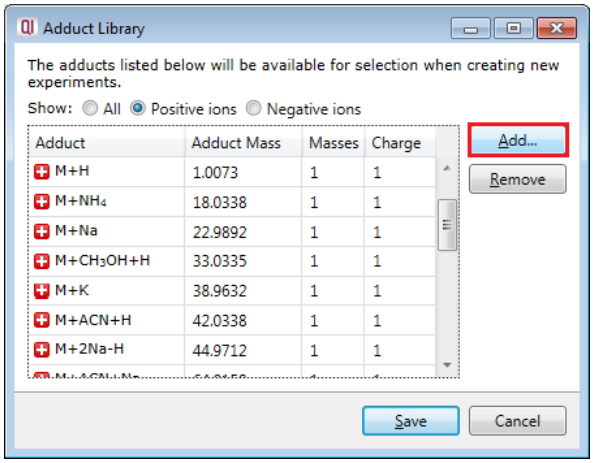


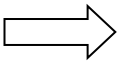
Finally select the ionisation polarity, either **positive** or **negative**, this will determine which list of possible adducts is available on the next page.

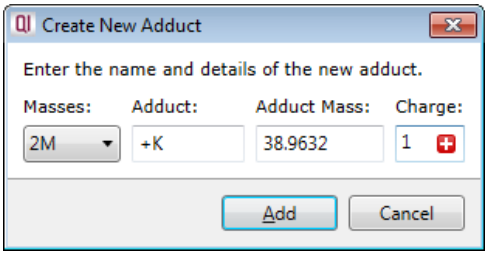




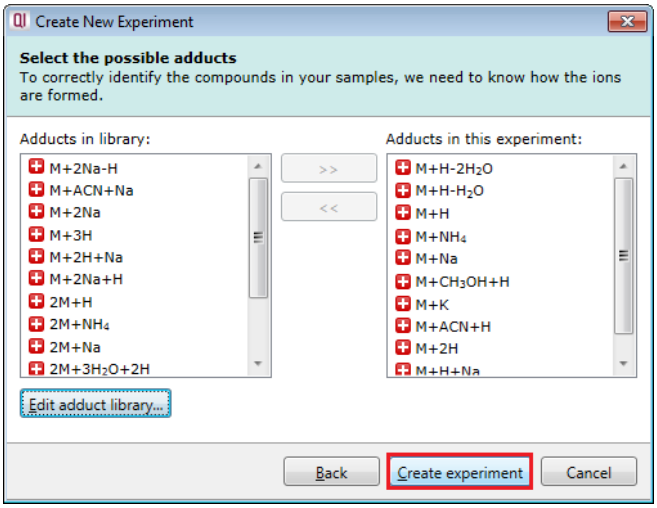
To add an adduct to the library select **Edit adduct library…** and then click **Add**. Now provide the Mass (monomer, dimer or trimer), Adduct, Adduct Mass and Charge of the new adduct and then add it to the library.







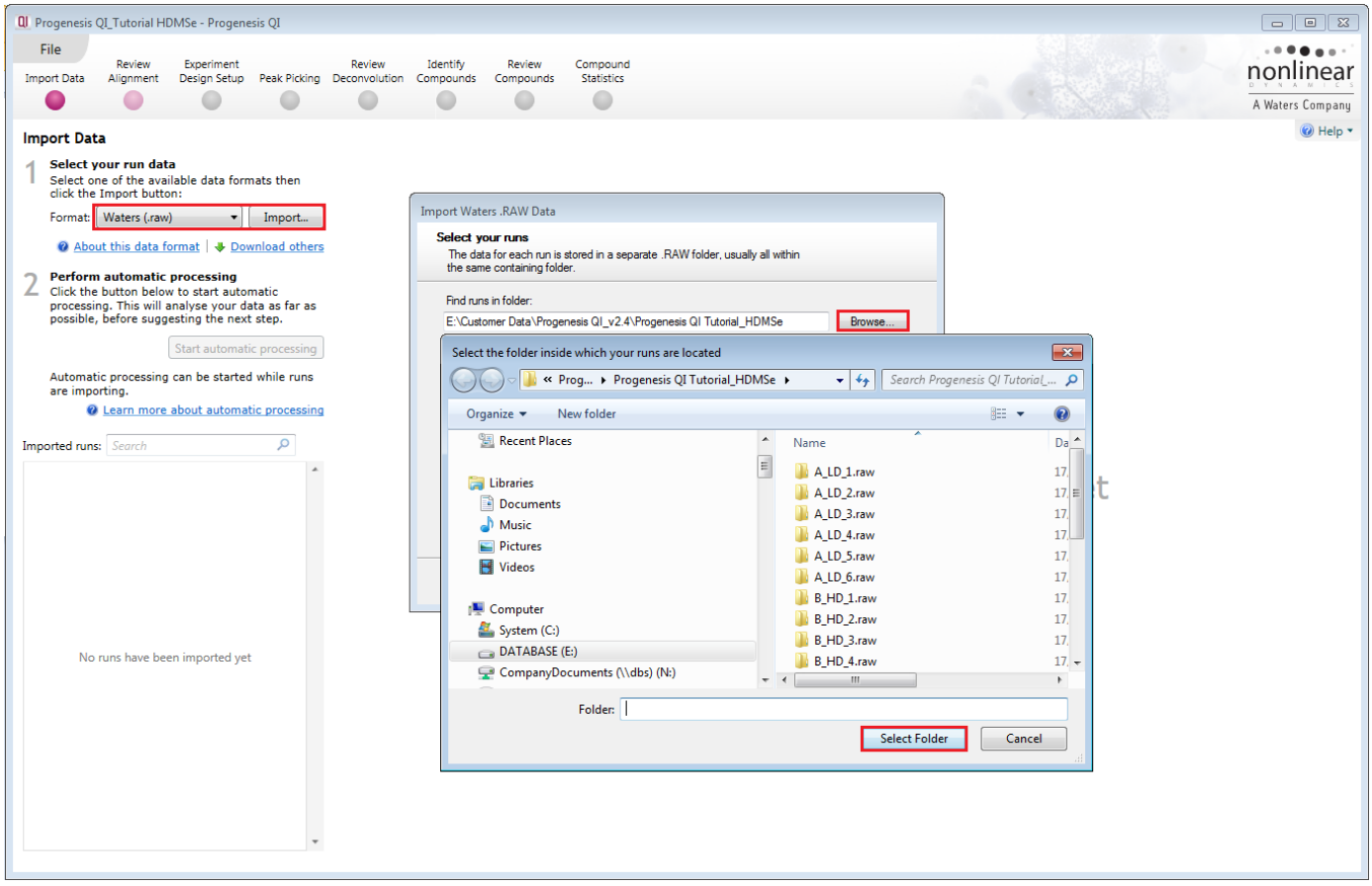
Now select the expected adducts based on your knowledge of your experimental conditions. In this example the polarity was positive and a basic set of Adducts was selected as shown on the right panel below.



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

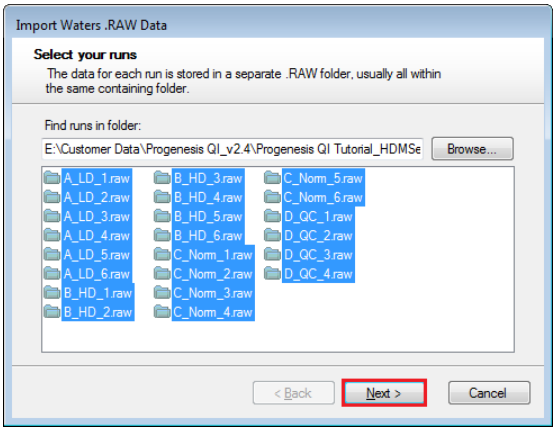
Select the 'Import Data file format', in this example they are **Waters (.raw)** files

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



Use Browse to locate and select the folder containing all the Data files, in this example (A_LD_1 to D_QC_4).

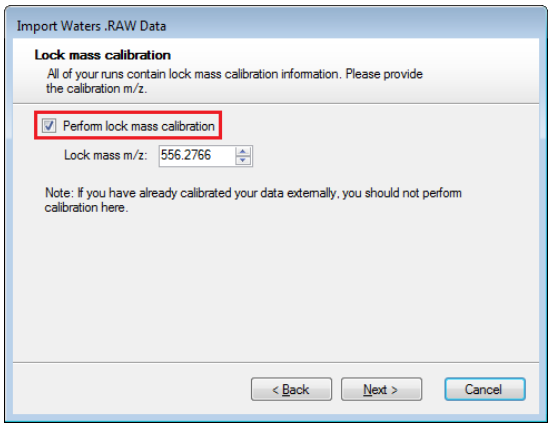
Select the required **.raw** folders for your experiment and click **Next.**



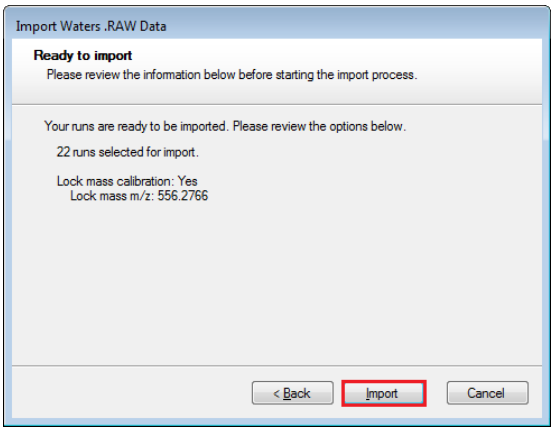
The folders will be examined and the Dialog will ask you for Lock Mass Calibration details.

If this has already been performed externally then untick this option.

Note: for the Tutorial data Lock mass calibration is required.

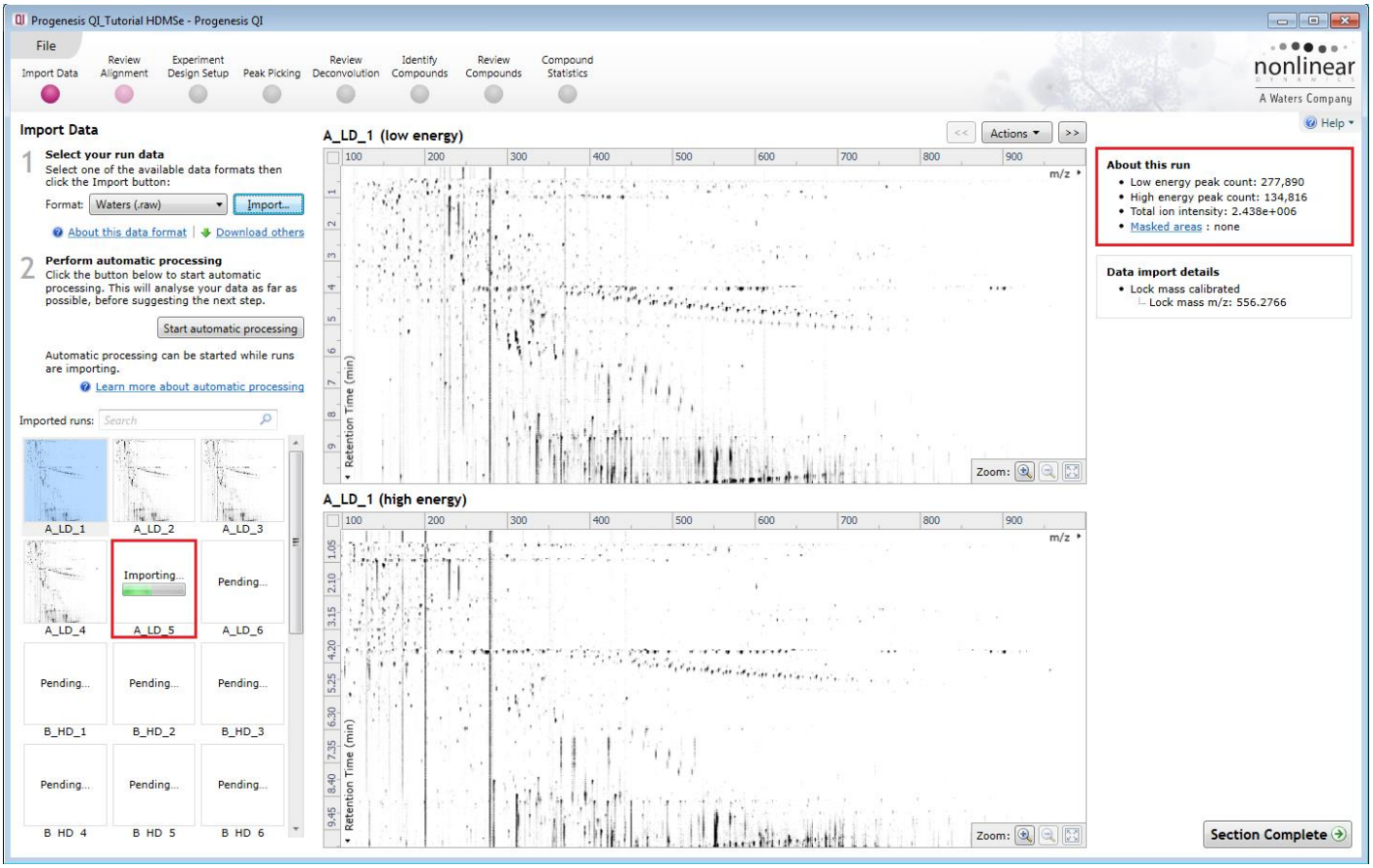


Finally a summary page appears. Click **Import**



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

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



Each data file appears as a 2D representation of the run. At this stage, assuming you have selected ‘profile data’ (page 79) you will be warned if any of the data files have been ‘centroided’ during the data acquisition and conversion process.

Note: as each data file is loaded the progress is reported in the **Import Data** list. The dialog below the run reports on the QC of the imported Data files. In this case ‘No problems found’ with this data file.

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

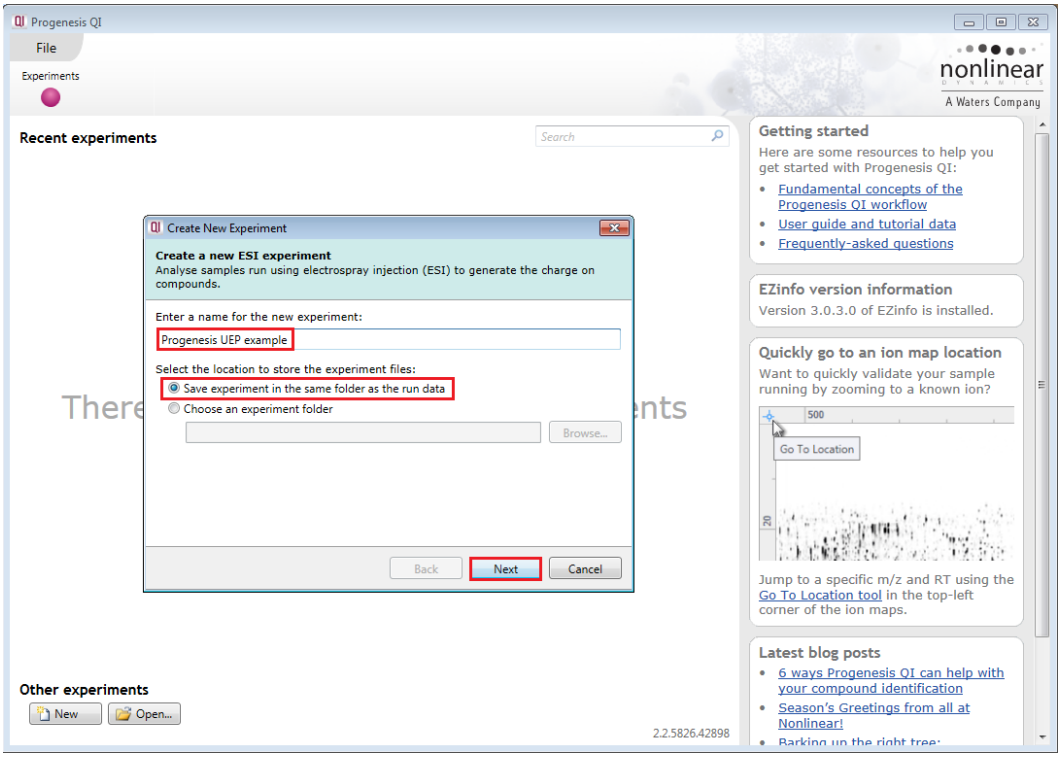
Note: as the loading process starts you can also **set up and start the Automatic Processing** before the loading has completed. This is a 4 stage process that involves the selection of an Alignment Reference (either automatically or manually), then the Automatic Alignment of all your runs to this Reference run followed by setting up of an Experiment Design and Peak Picking.

Click **Start automatic processing** to start the automatic alignment of your runs.

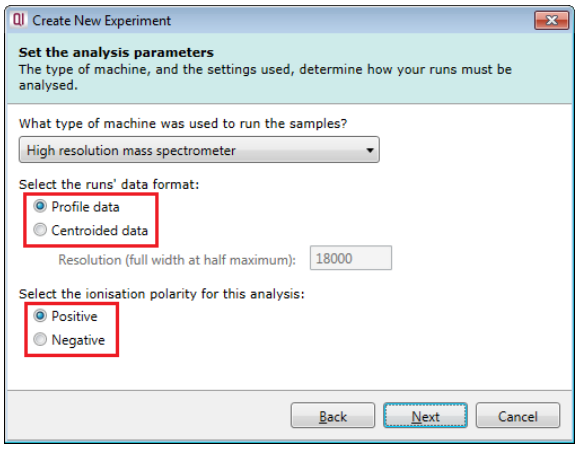
Appendix 1b: Stage 1 Data Import from UNIFI (Waters Only)

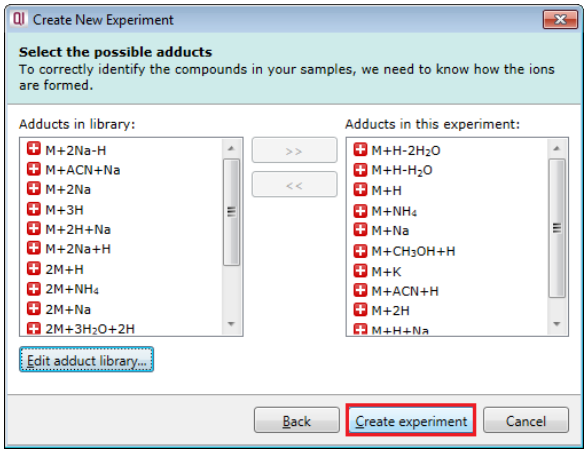
LCMS data that has been acquired fron Waters Vion and Xevo machines where the acquisition is managed through the UNIFI Scientific Information System, is exported **Uncompressed** from UNIFI as a Unifi Export Package (UEP) format file.

To create a new experiment in Progenesis QI, using a .UEP file, select **New** give your experiment a name (Progenesis UEP example) and then select a location to store the experiment files. Click **Next**.



Now select the machine type, polarity that was used to run the samples. Then select the expected adducts based on you knowledge of your experimental conditions.

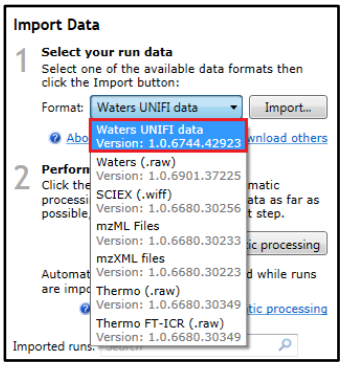




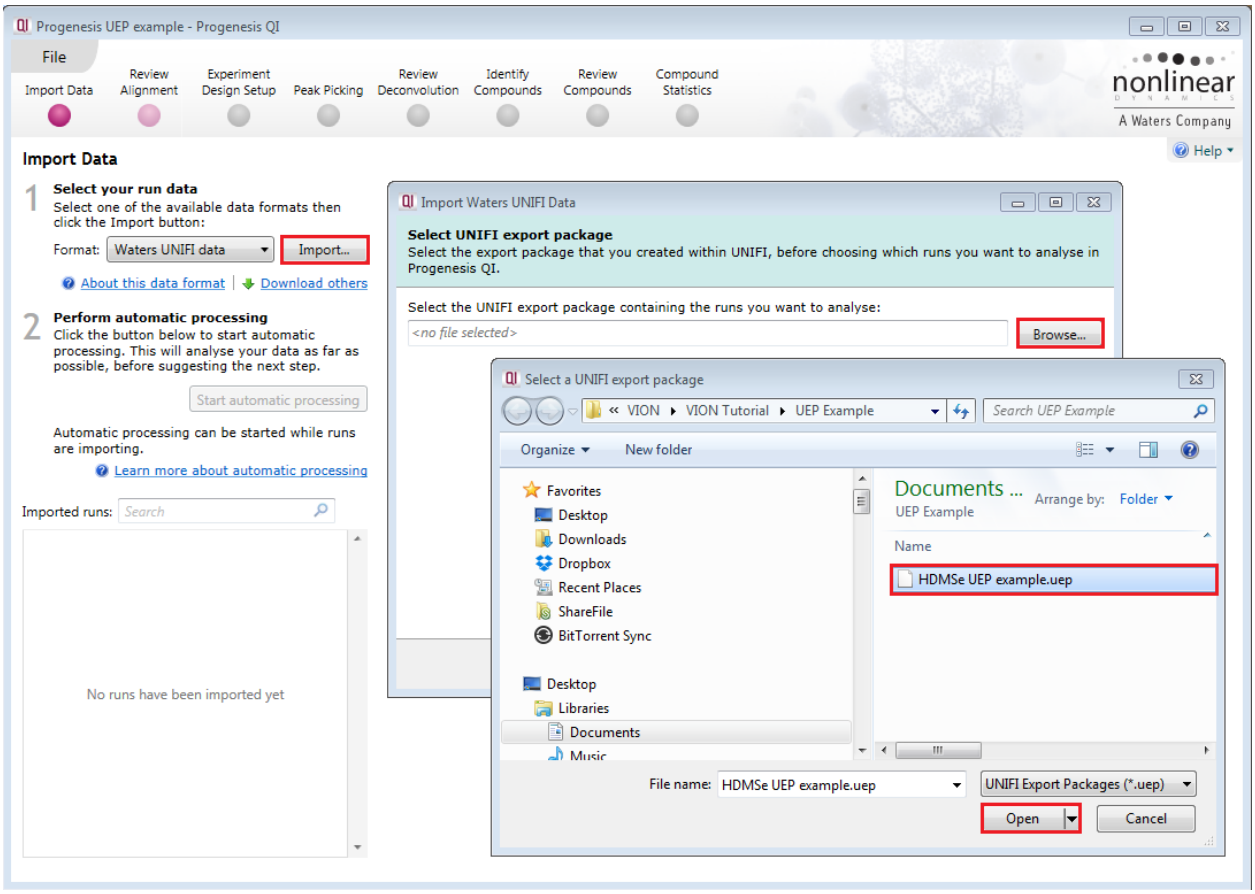
Click **Create experiment** to generate the experiment.

When Import data opens select **Waters UNIFI data** as the import data format.

Then click **Import.**



Browse to the location of the UEP and open the file

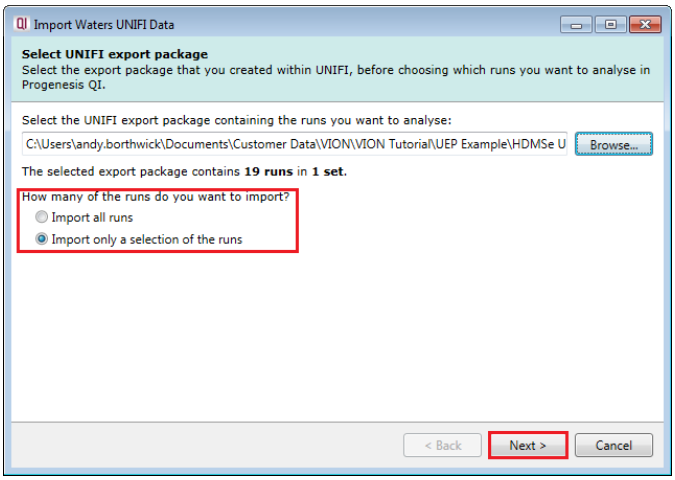


Depending on the data in UNIFI the exported package may contain 1 or more sample sets. In this example there is only one, containing 19 runs.

You have the option of importing all the samples or a selection.

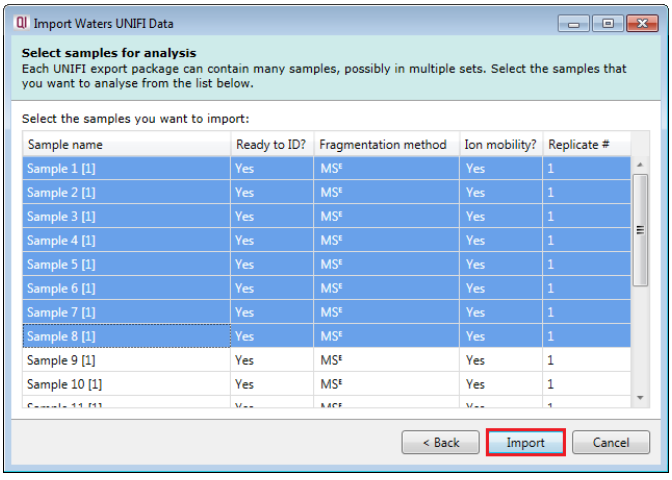
In this example we will import a selection of the runs.

Click **Next**.

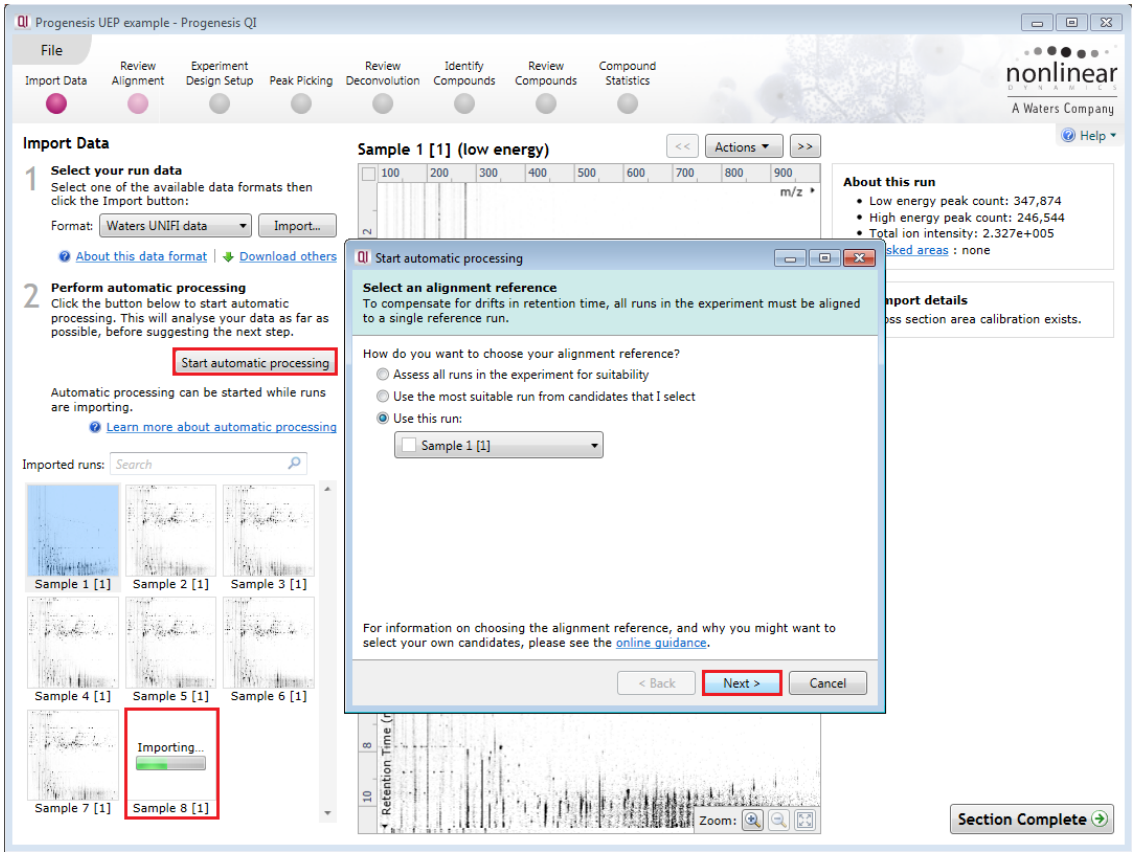


Select the runs to import.

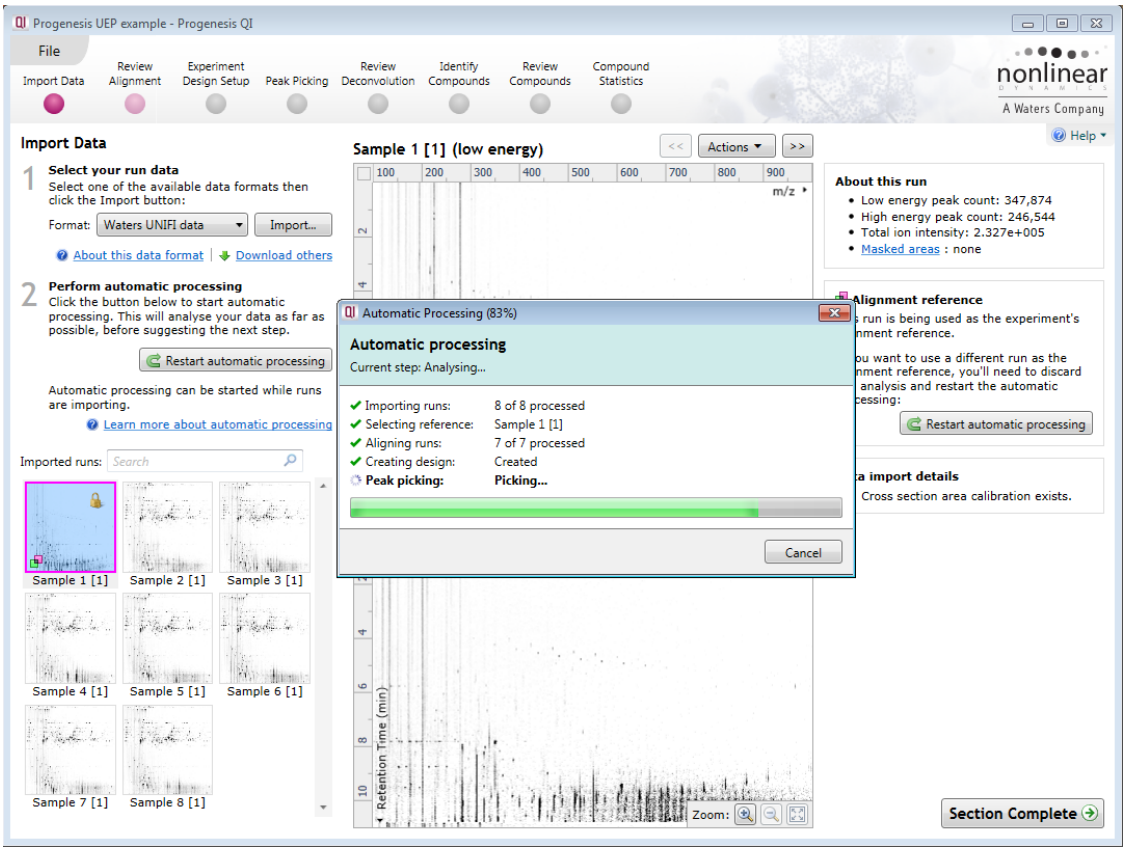
Click **Import**.



As importing from the UEP starts you can set up the **automatic processing** to run on the data as described in Stage 2A of this guide (page 8).



Once the importing is complete, the Automatic processing of the data proceeds.

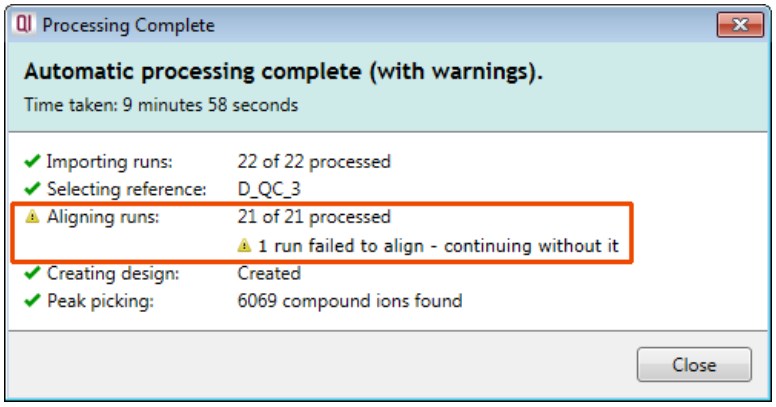


Once the analysis completes refer to page 12 in the guide.

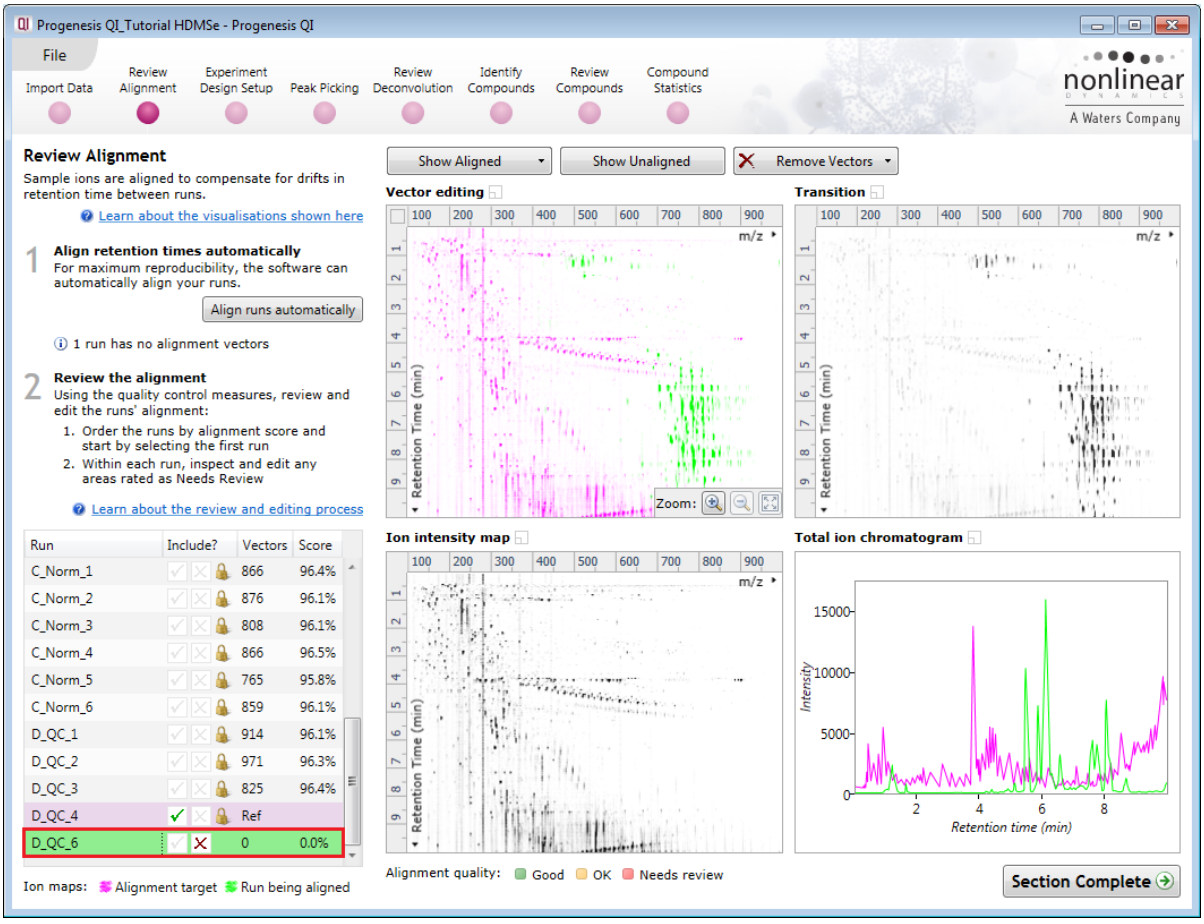
Appendix 2: Stage 2 Processing warnings and failures

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 warn you that runs failed to align. On completion of the Automatic processing go to the **Review Alignment** stage to investigate the problem.



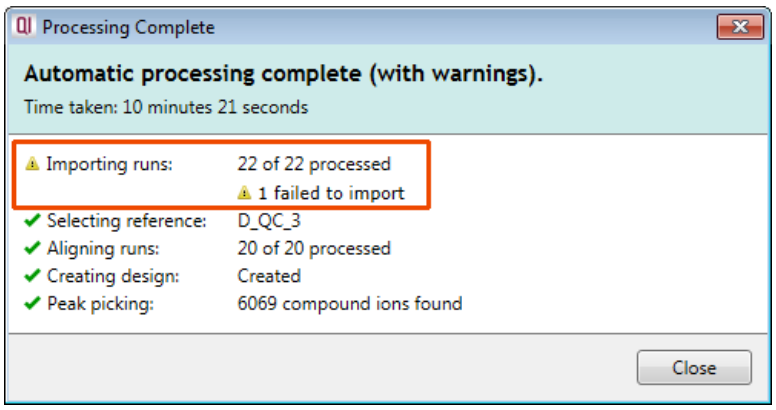
Note: in this example the run that failed to align will not contribute to the peak picking and will be excluded at the alignment stage (a cross appears in the include column).



You can either remove the run from the experiment at **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.

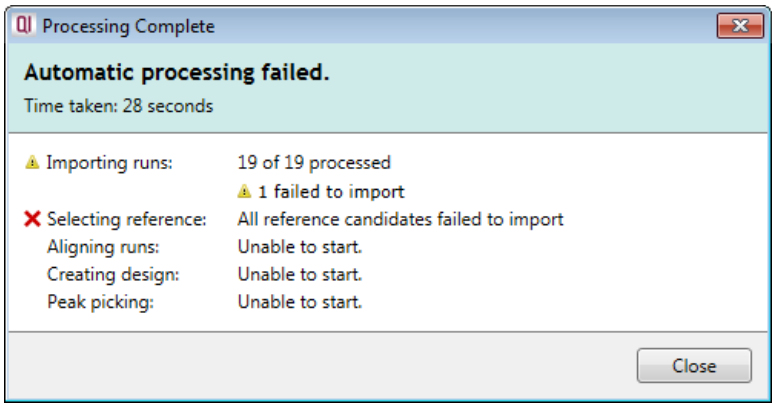
If some runs in a data set fail to import (but not all), the automatic processing will continue informing you that one or more runs have failed to import.



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

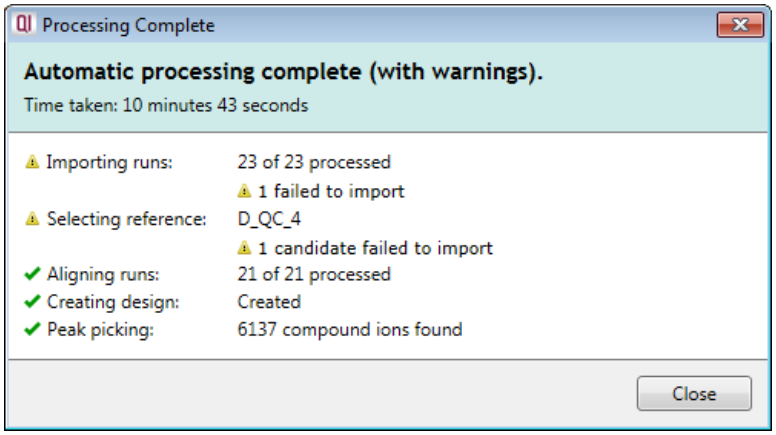
Note: adding additional runs will then be aligned and peak picking should be re-done to include data from the added runs in the generation of the aggregate.

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).

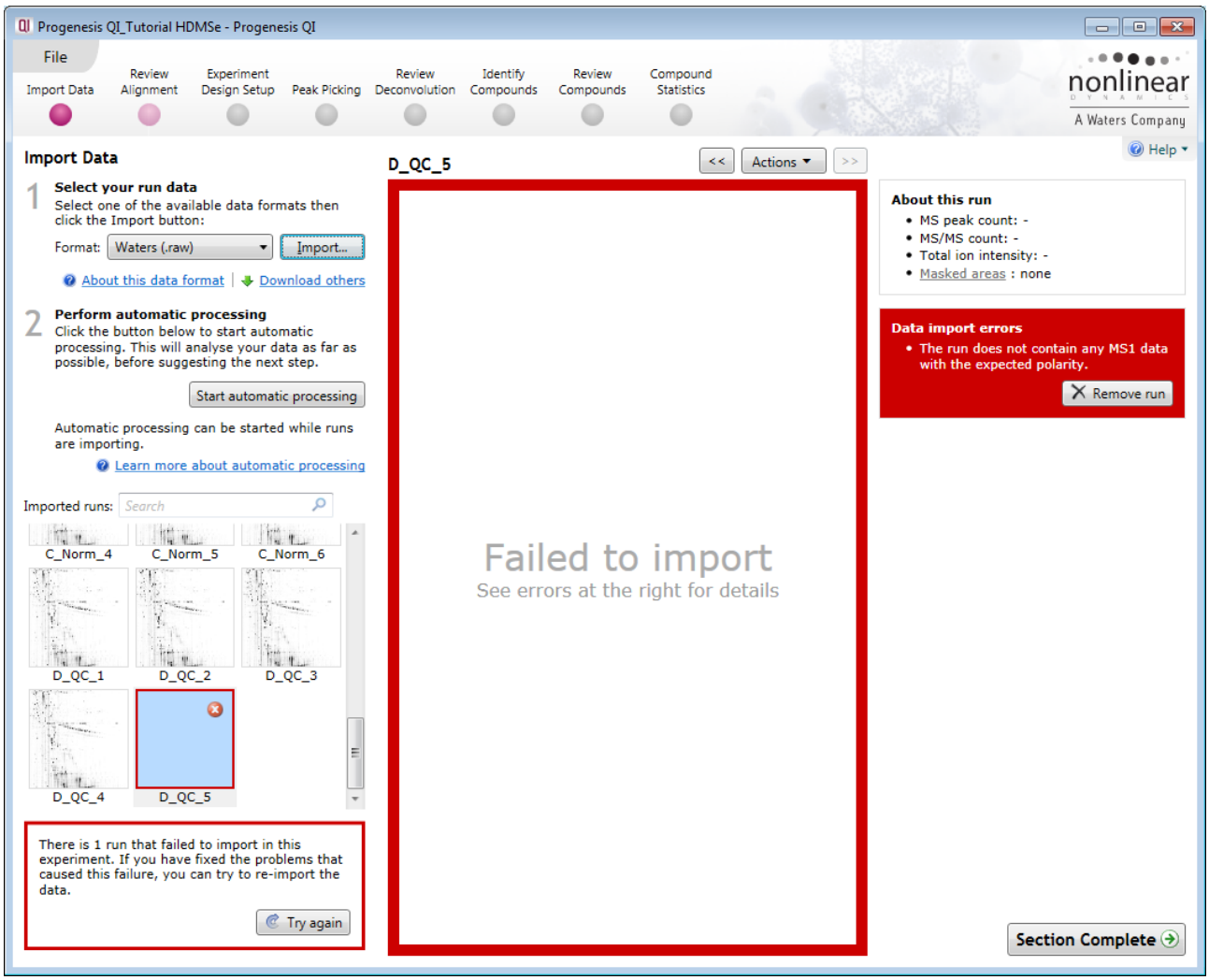


In this case, the processing dialog has halted indicating that all of the alignment reference candidates failed to import. To proceed restart the process with other runs.

Finally if you have chosen to select the Alignment Reference from a number of QC runs and one of them fails to import then processing will continue as there are alternative runs that did load that can be automatically selected as the reference.



An example of a failed import is shown below



In this example the run has the wrong polarity compared with the other runs. You can right click on the run and remove it or just click on Remove run to the right of the main display.

Click **Start automatic processing** to start the automatic alignment of your runs.

Note: At this stage you have the option to Review the Chromatography or go straight to the review of the Automatic Alignment of your data.

Appendix 3: Licensing runs (Stage 3)

When setting up a **New experiment** if you are evaluating Progenesis QI 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.

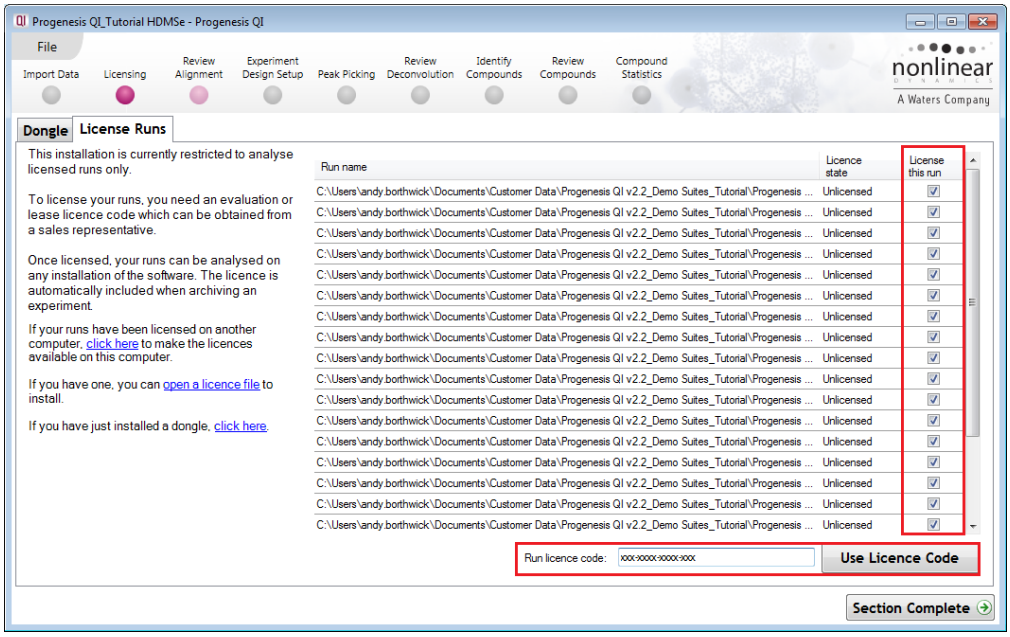
To use this page to License your Runs you must first either obtain an **'Evaluation' Licence Code from a Sales Person or purchase a licence code directly**.

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

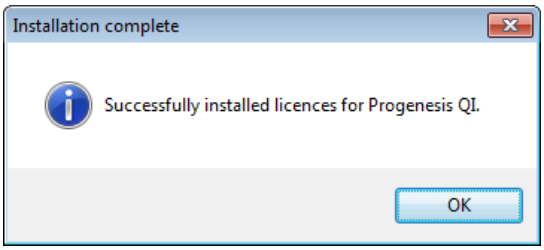
The runs in your experiment will be listed as shown below.

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 (using tutorial data)

To place manual alignment vectors on a run (A_LD_1 in this example):

Note: this data is already well aligned so a single vector has already been placed to simulate misalignment so as to allow the demonstration of the use of manual vectors.

1. Click on Run A_LD_1 in the **Runs** panel, this will be highlighted in green and the reference run (D_QC_4) will be highlighted in magenta.
2. You will need approximately 5 alignment vectors evenly distributed from top to bottom of the whole run.
3. First drag out an area on the **Ion Intensity Map** (C), this will reset the other 3 windows to display the same 'zoomed' area.

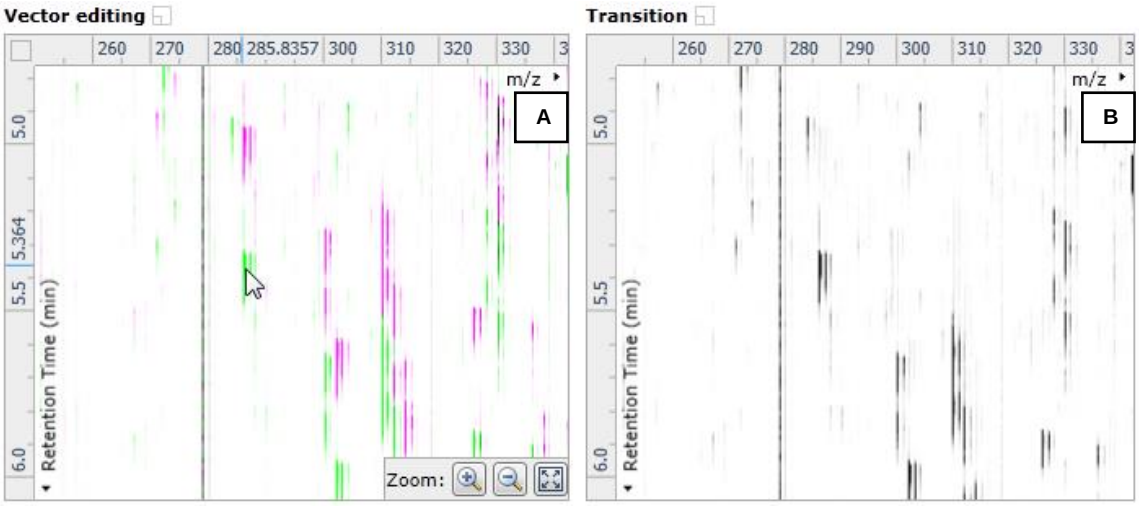
Note: when you click anywhere in the Ion Intensity Map this will **reposition the focus** and update the other views accordingly.

Note: the features moving back and forwards between the 2 runs in the **Transition** window (B) indicating the misalignment of the two runs

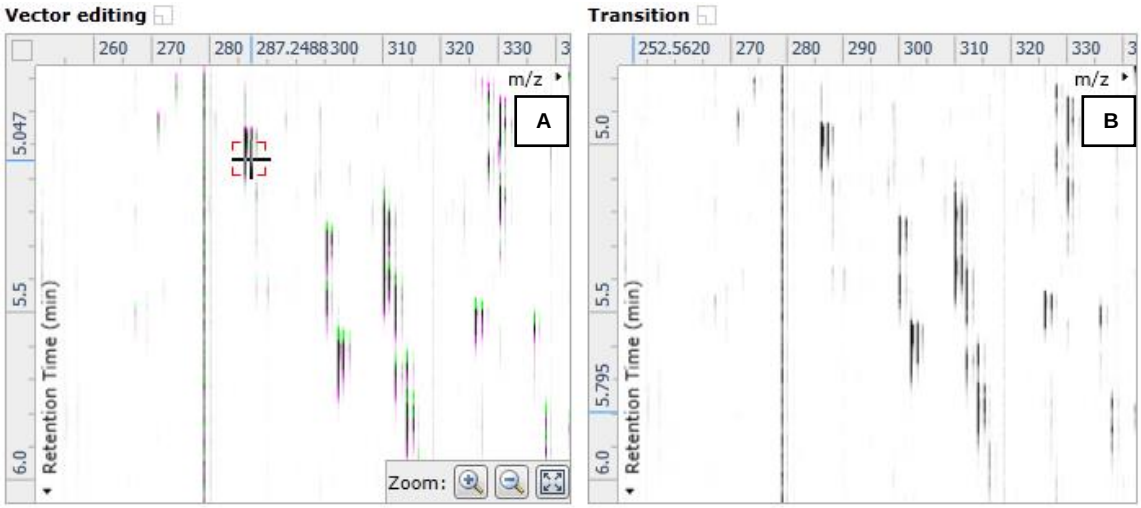
Note: the Ion Intensity Map gives you a colour metric, visually scoring of the current alignment and an overall score is placed next to the Vectors column in the table. With each additional vector this score will update to reflect the overall quality of the alignment. The colour coding on the Ion intensity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

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

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

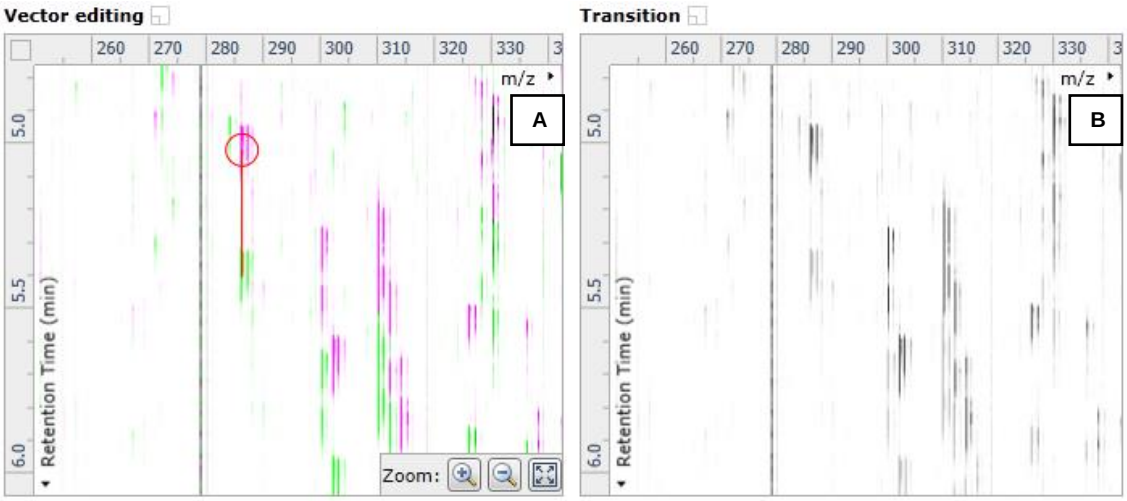


5. As you are holding down the left mouse button (depending on the severity of the misalignment), the alignment vector will automatically find the correct lock. If not, drag the green feature over the corresponding magenta feature of the reference run. The red box will appear as shown below indicating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



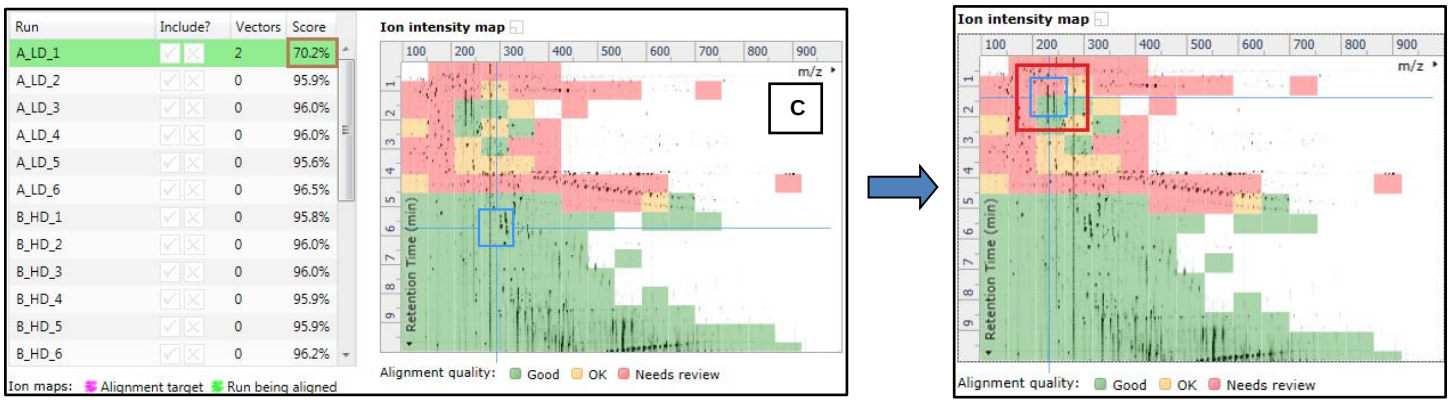
Tip: while holding down the mouse button hold down the **Alt** button. This will allow smooth movement of the cursor as the **Alt** key allows you to override the 'automatic alignment' performed as you depress the mouse button.

6. On releasing the left mouse button the view will 'bounce' back and a red vector, starting in the green 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 and selecting delete vector

7. 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.



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

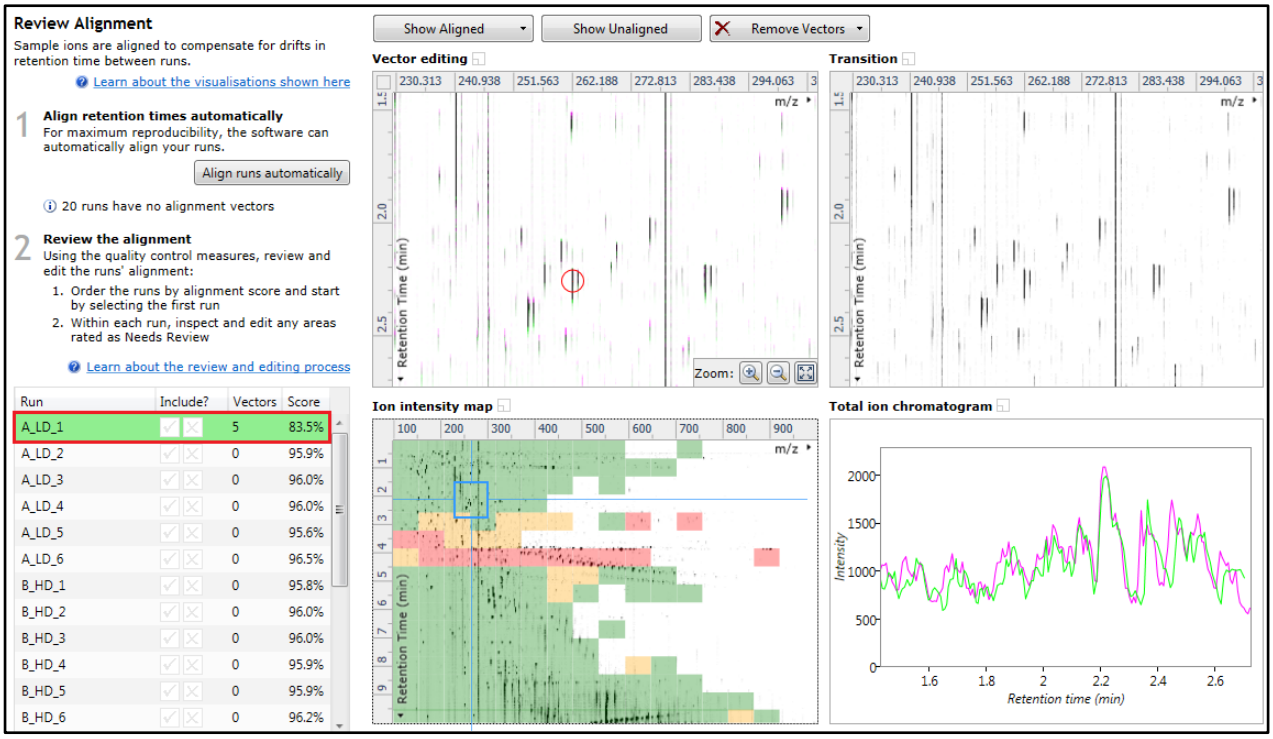
9. Repeat this process moving the focus from top to bottom on the **Ion Intensity Map** view the number you add is dependent on the length of gradient used and the severity of the misalignment. In many cases a single vector per alignment is all that is required to improve the performance of the Automatic Alignment

Note: the manual vectors are red to distinguish them from the automatic vectors which blue

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

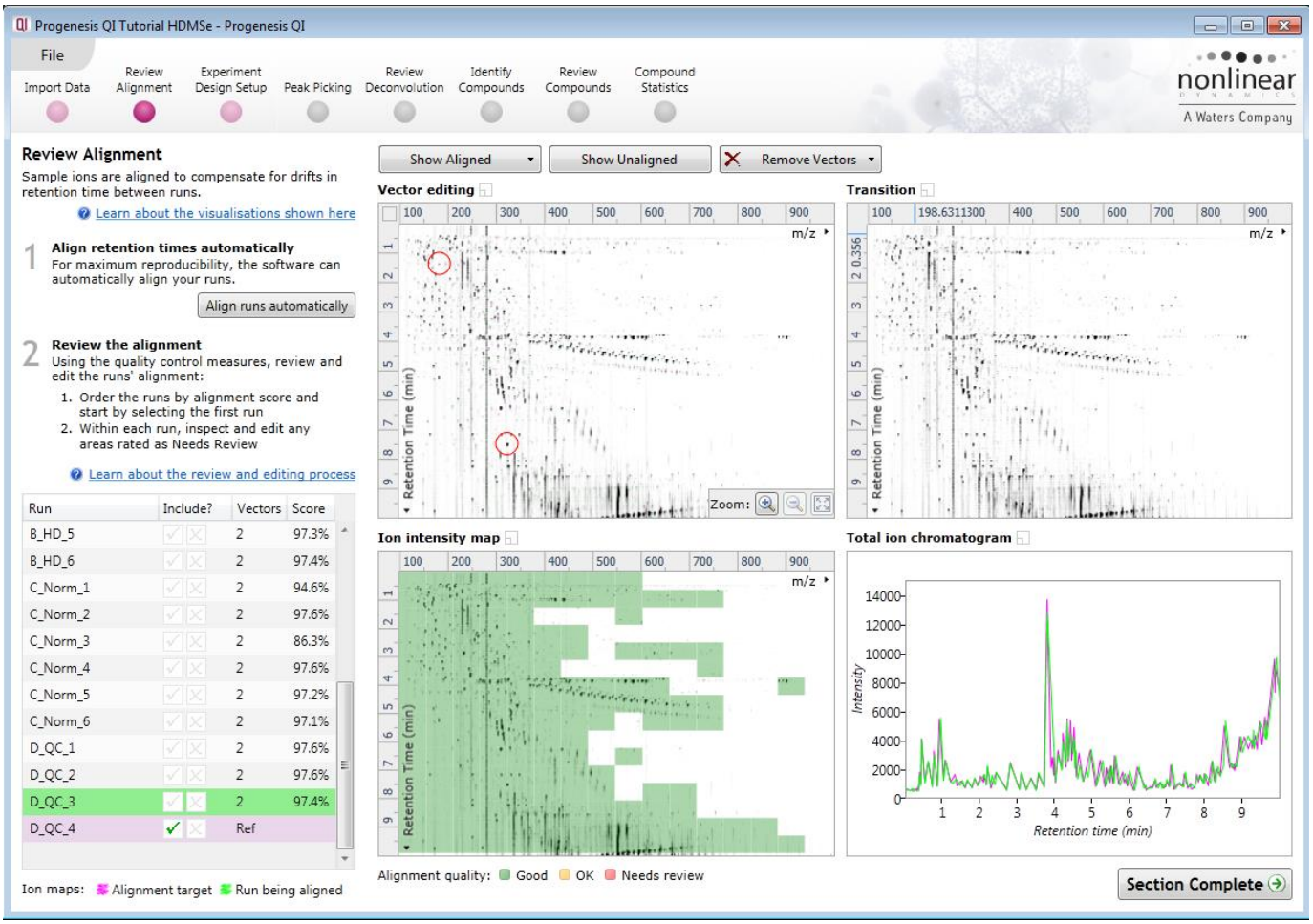
The number of manual vectors that you add at this stage is dependent on the misalignment between the current run and the Reference run. The effect of adding the manual vectors can be seen when you press show aligned

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

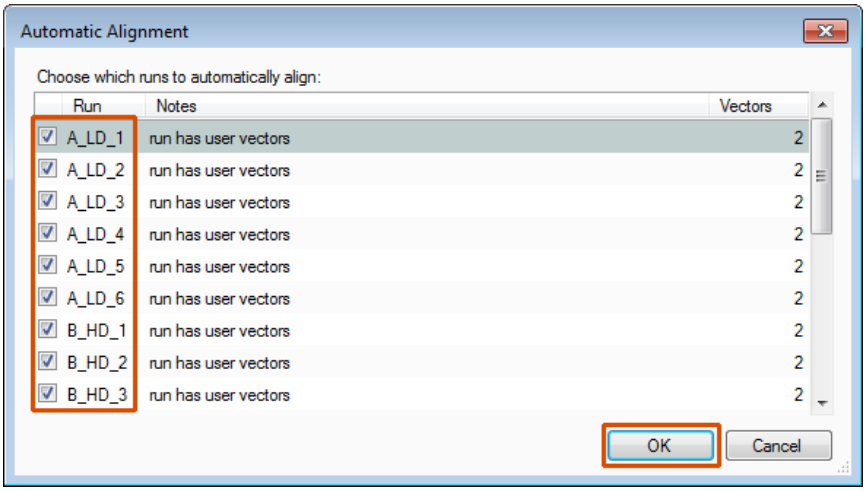


In many cases only using the Automatic vector wizard will achieve the alignment.

Also the 'ease' of addition of vectors is dependent on the actual differences between the 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.



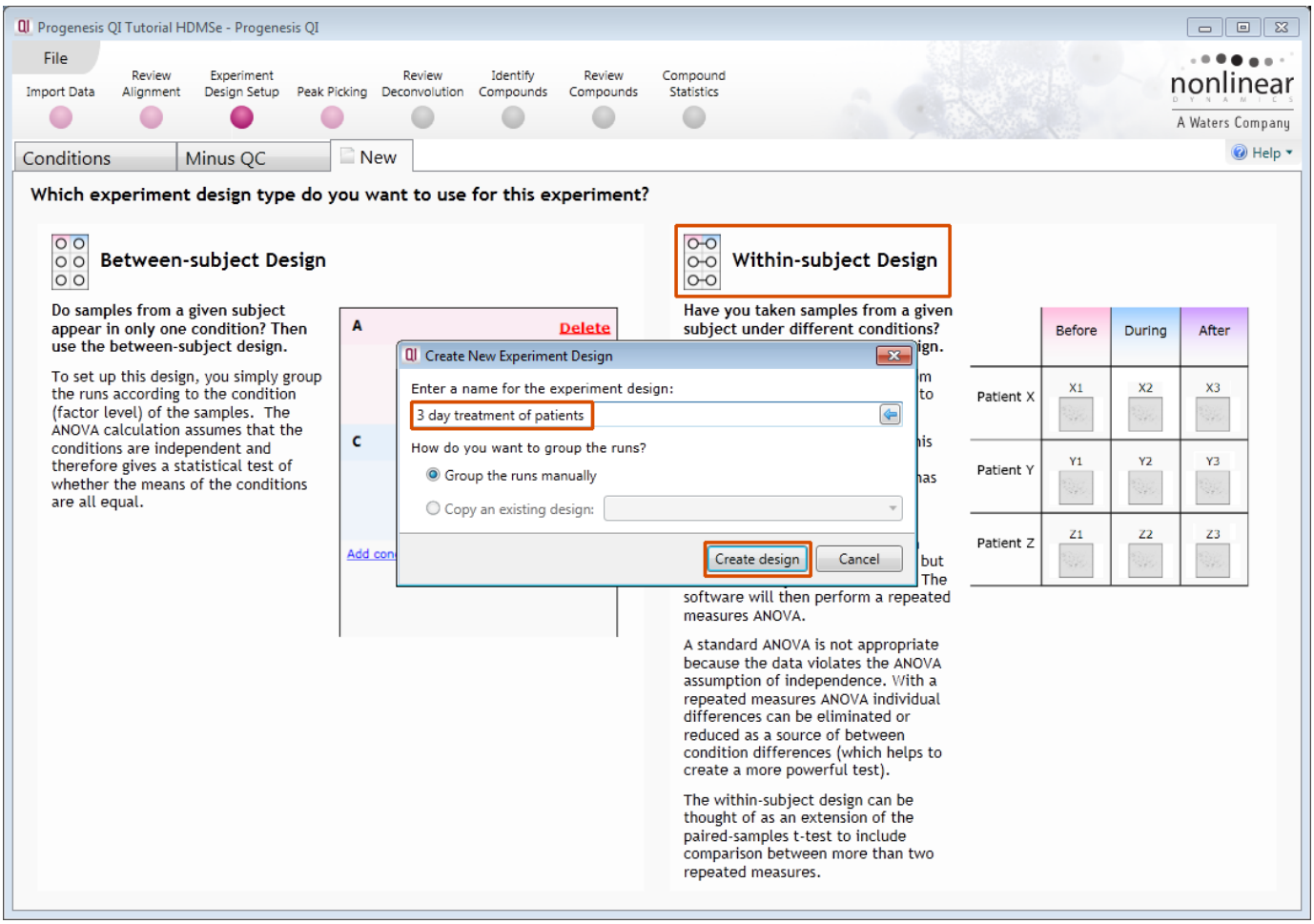
Note: the tick box next to the **Run** name controls whether automatic vectors will be generated for each run.

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.

Appendix 5: Within-subject Design

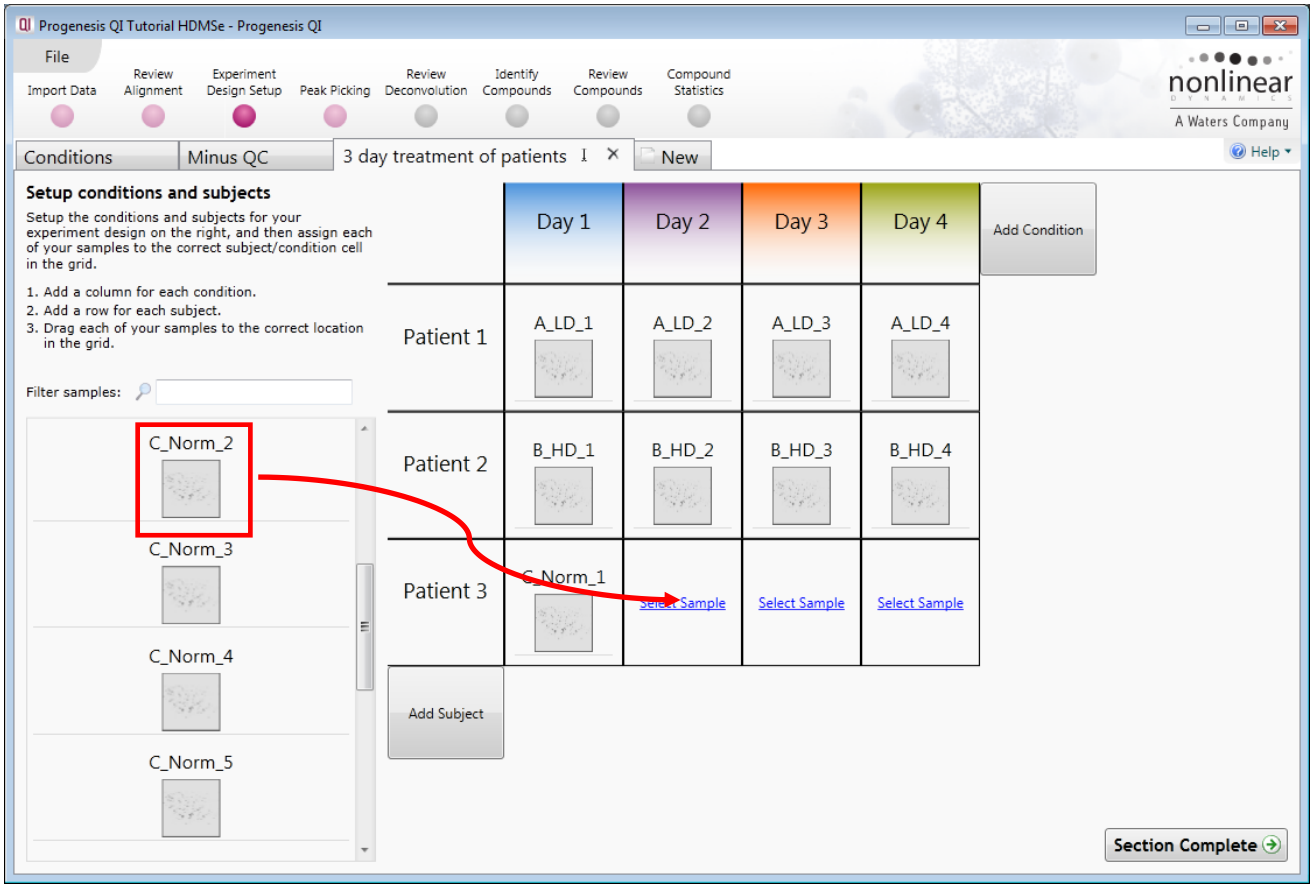
To create a **Within-subject Design** for your data set select this option on the **Experiment Design Setup** page and enter the name of the design.

In this example there are 3 Subjects (i.e. patients A, B and C) who have been individually sampled: Before Treatment: Day1 and then at 3 times following treatment: Day2, Day3 and Day4.



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.

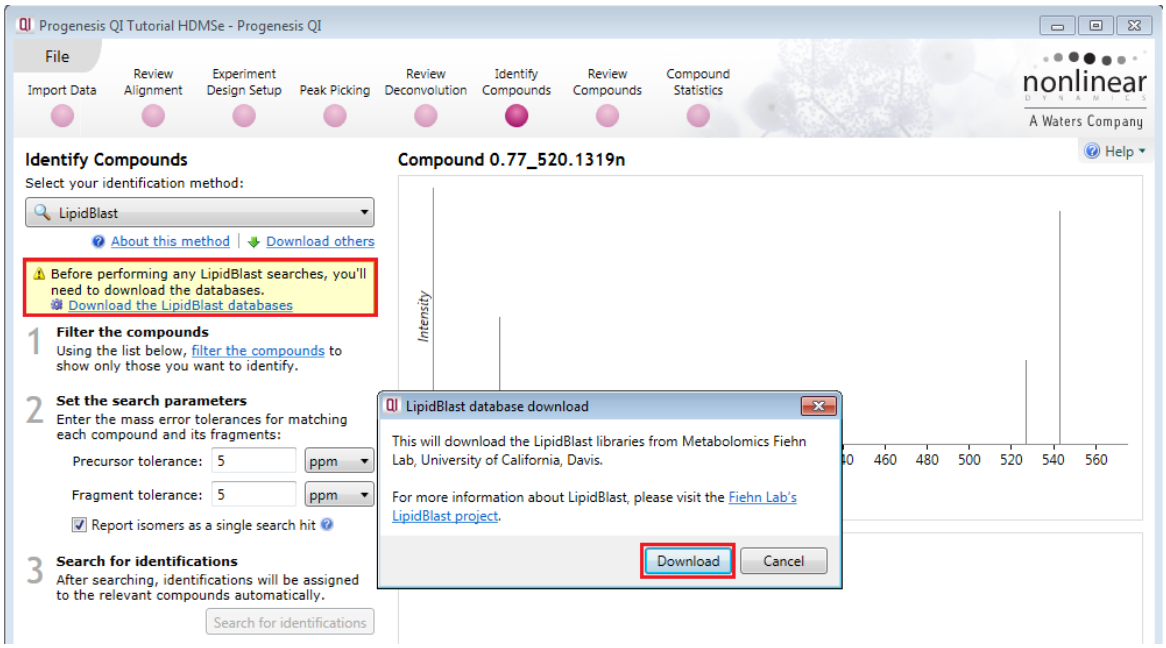


You can create additional Experimental Designs using the New tab

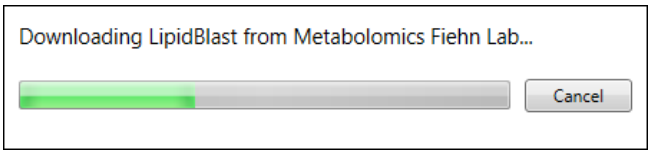
All of these Experimental Designs are available at all the following stages in the Progenesis QI workflow.

Appendix 6: Using LipidBlast for Compound Identification

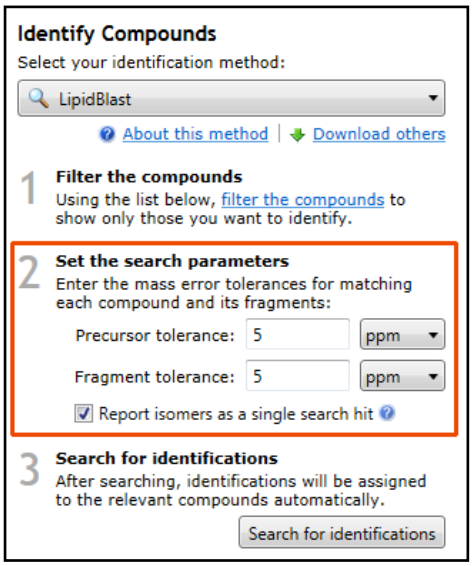
When using LipidBlast for the **first** time you will require internet access to download the LipidBlast databases. Click on the download link.



Dialog will open acknowledging the source of the LipidBlast libraries, click Download.



When performing a LipidBlast search the following parameters can be set:



Precursor tolerance

Set the allowable threshold (in ppm or Da) for the difference between the database neutral mass and your measured neutral mass (default is 5ppm).

Fragment tolerance

Set threshold (in ppm or Da) that determines whether a database ms/ms peak is considered to match a measured peak. The difference in m/z must fall below the threshold to be considered a match.

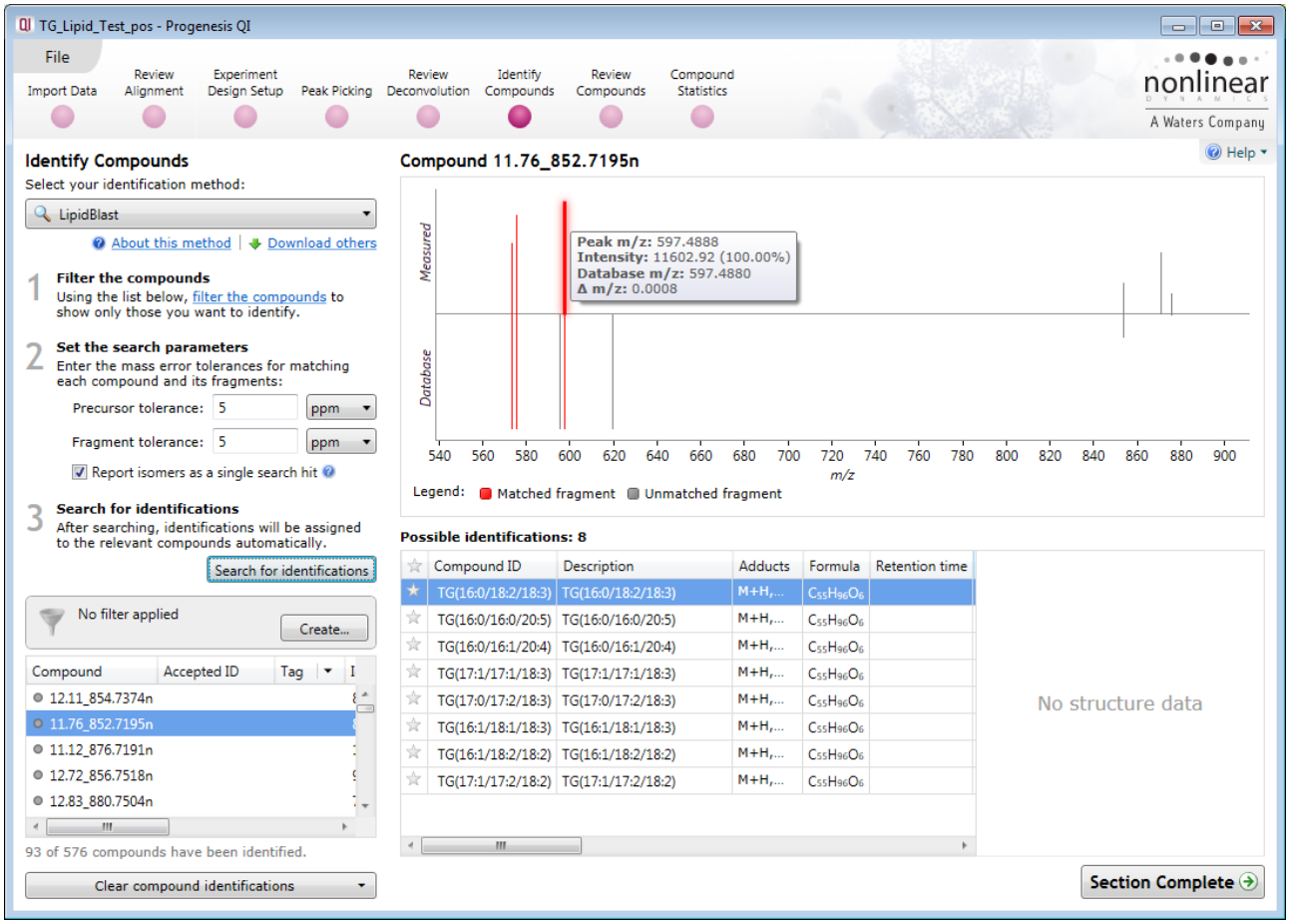
Report isomers as a single search hit

The search will merge hits that differ only in double bond positioning and/or geometry thus reporting isomers as a single search hit.

For example, if a search returns two identifications, **GPCho(16:0/20:3(8Z,11Z,14Z))** and **GPCho(16:0/20:3(5Z,8Z,11Z))**, these will be reported as a single hit **GPCho(16:0/20:3)**. When not selected, all lipid forms will be returned as separate search hits.

Having set the parameters click **Search for identifications**

On performing the search, you will get a 'mirror plot' where the upper half represents the measured/observed fragmentation and the lower half shows the currently matched fragmentation spectrum in the database.



Note: a matched fragment is indicated in red, spanning the x axis. If you hold the cursor over a matched fragment details of the measured and database m/z are reported as well as the mass difference.

Where fragmentation data is available then the order will be based on the fragmentation score. In the absence of fragmentation data the score will be based on mass error and Isotope similarity.

Note: LipidBlast does not contain structural data. Clicking on the URL in the in the **Link** field, of the possible identifications table, takes you to Lipid Maps where a structural diagram is available.

Appendix 7: Exporting compound identities to IPA

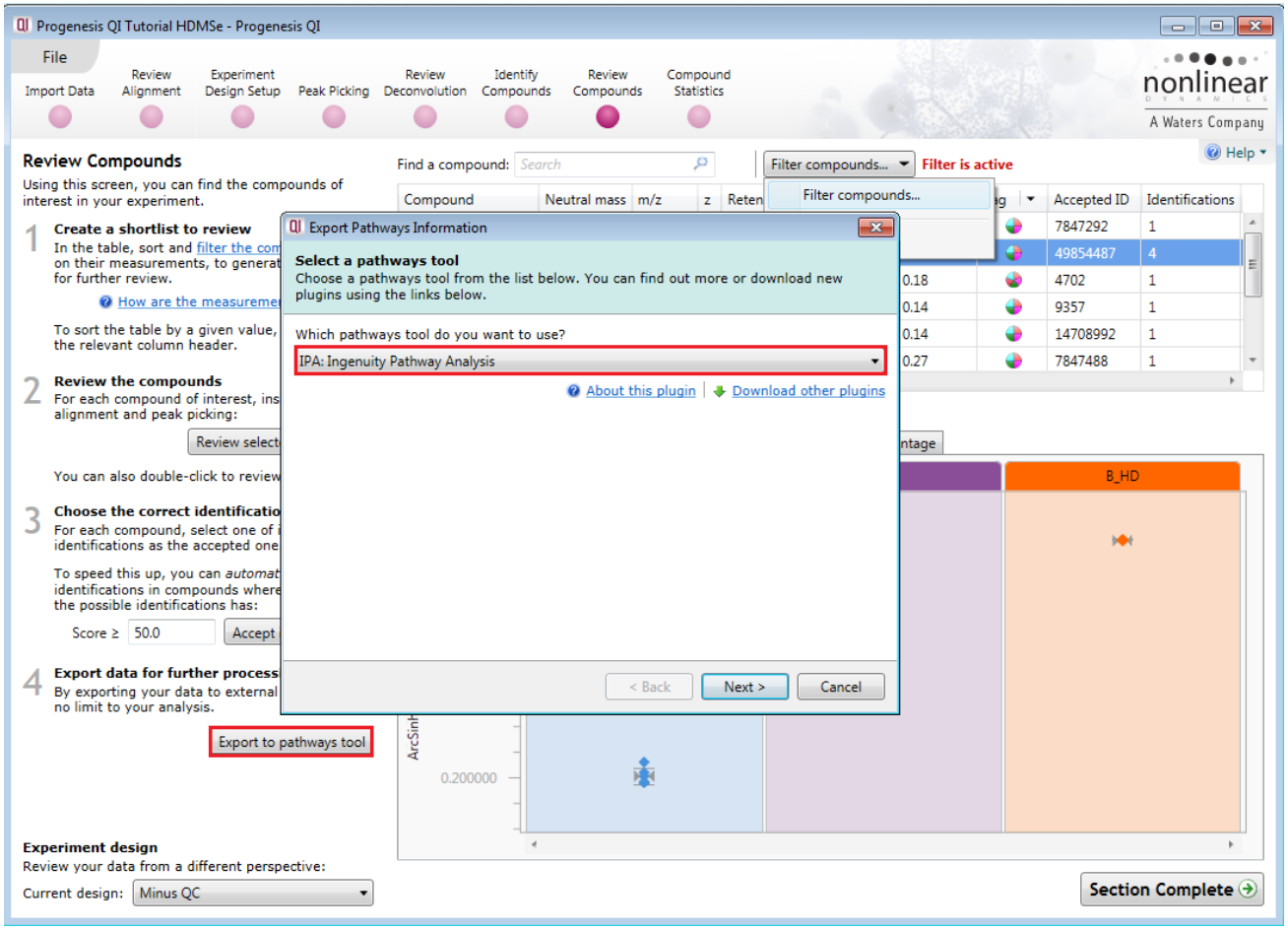
IPA (Ingenuity Pathway Analysis): is a commercial web-based software application for the analysis, integration and interpretation of data derived from any experiments that generate gene and chemical lists with expression data, such as metabolomics and proteomics.

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

Note: the use of IPA requires a licensed copy, for licensing details please visit the IPA website.

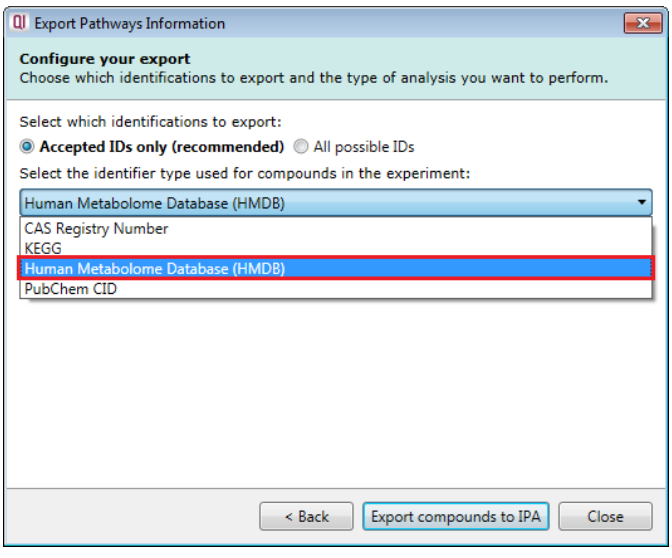
First select the compound data to export to the pathways tool using tag filtering to 'focus' the set to export.

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



Export of the **Accepted IDs only** is recommended then select the appropriate identifier type used for compounds in the experiment.

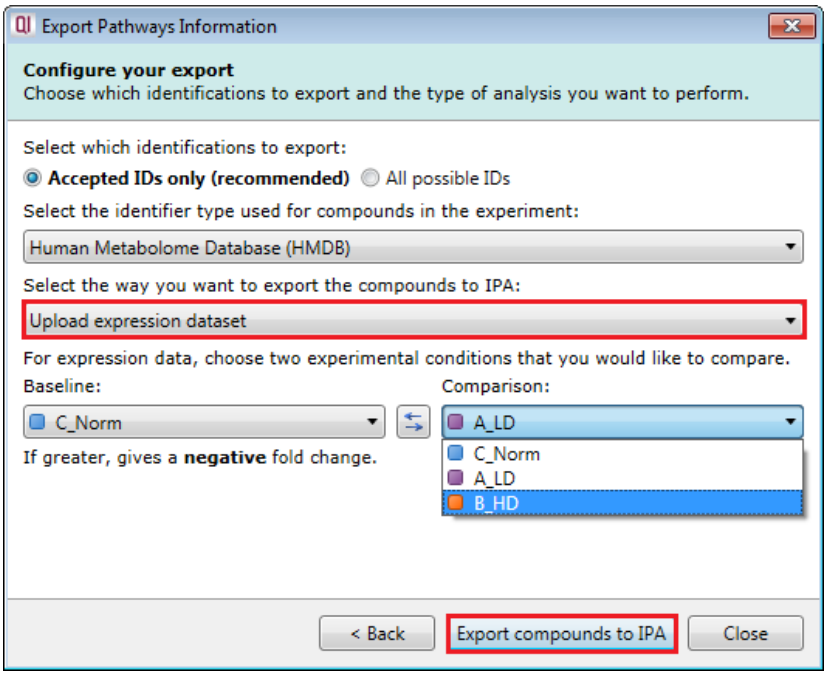
Note: for compounds with different identifier types, filter the compound list to contain only compounds with single identifier type at a time, then perform the export operation and merge the resulting lists in IPA.



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

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

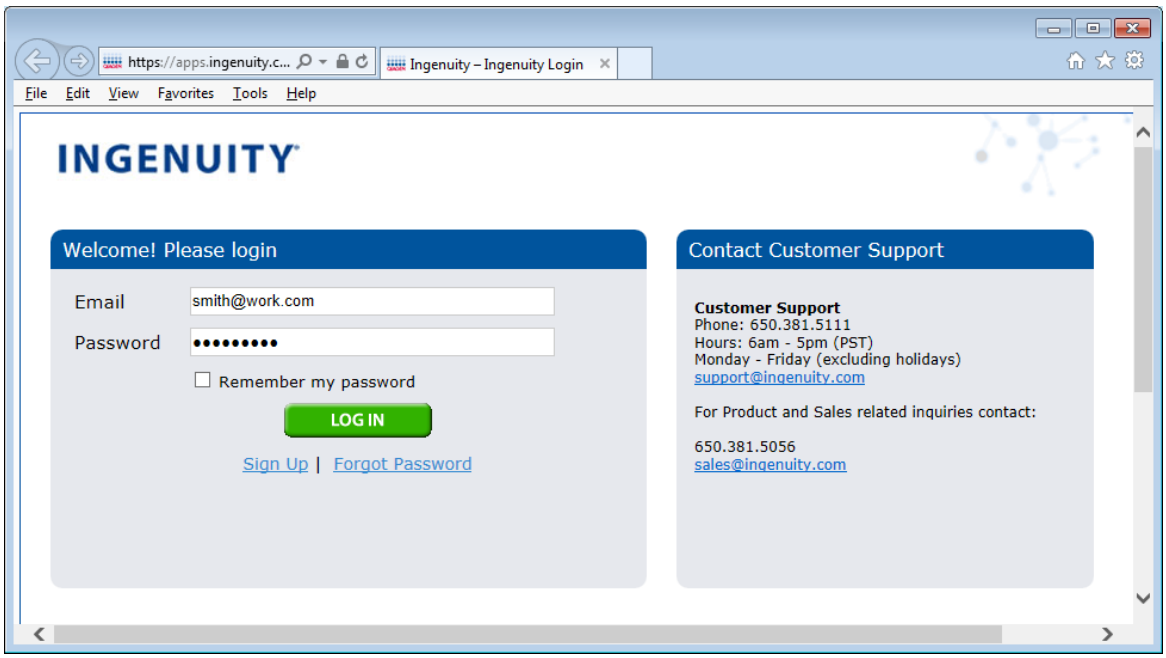
If you require to perform enrichment analysis or other types of analyses that use compound expression data, select the option to **Upload expression dataset** to IPA.



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

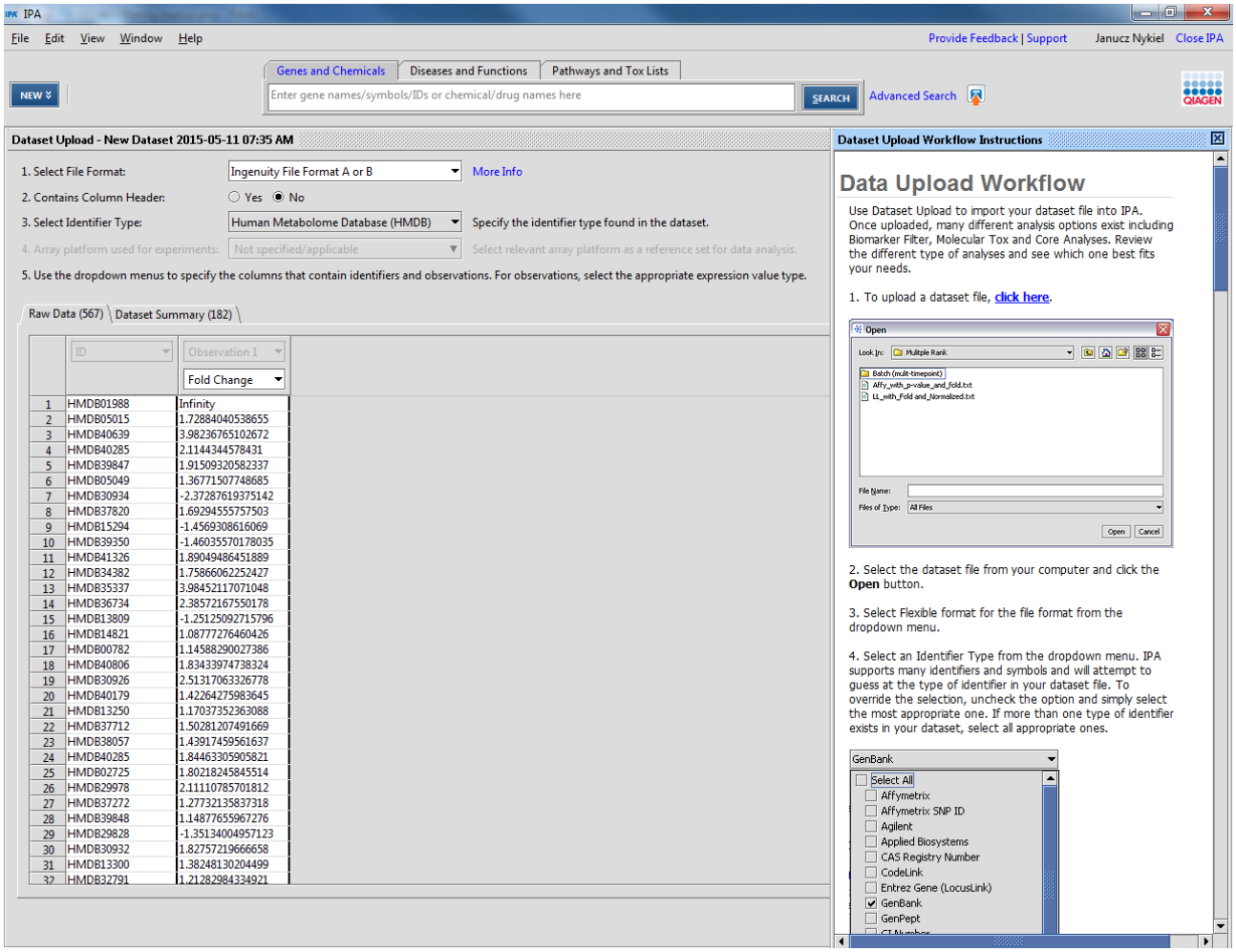
The operation exports data from a single observation at a time - select the two experimental conditions (taken from the current **experiment design as set in Progenesis QI** you wish to compare as that observation).

Click **Export compounds to IPA** .



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

IPA will open displaying the imported data from Progenesis.

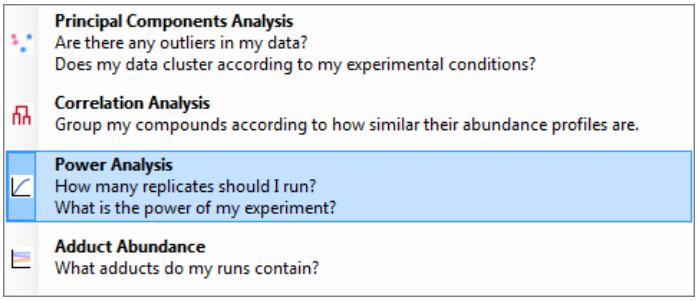


You can now explore your compound expression data using the tools available in IPA. Instructions on how to manage and explore your imported data set are provided by IPA.

Appendix 8: Power Analysis

Power analysis is a statistical technique, which is used to gauge how many replicates are needed to reliably observe the abundance differences in your data. It is available through the Compound Stats section of the workflow.

To perform a power analysis of the data click on **Ask another question** at the top of the table in the Compound Stats screen. A selection of 4 tools will appear in the form of questions.



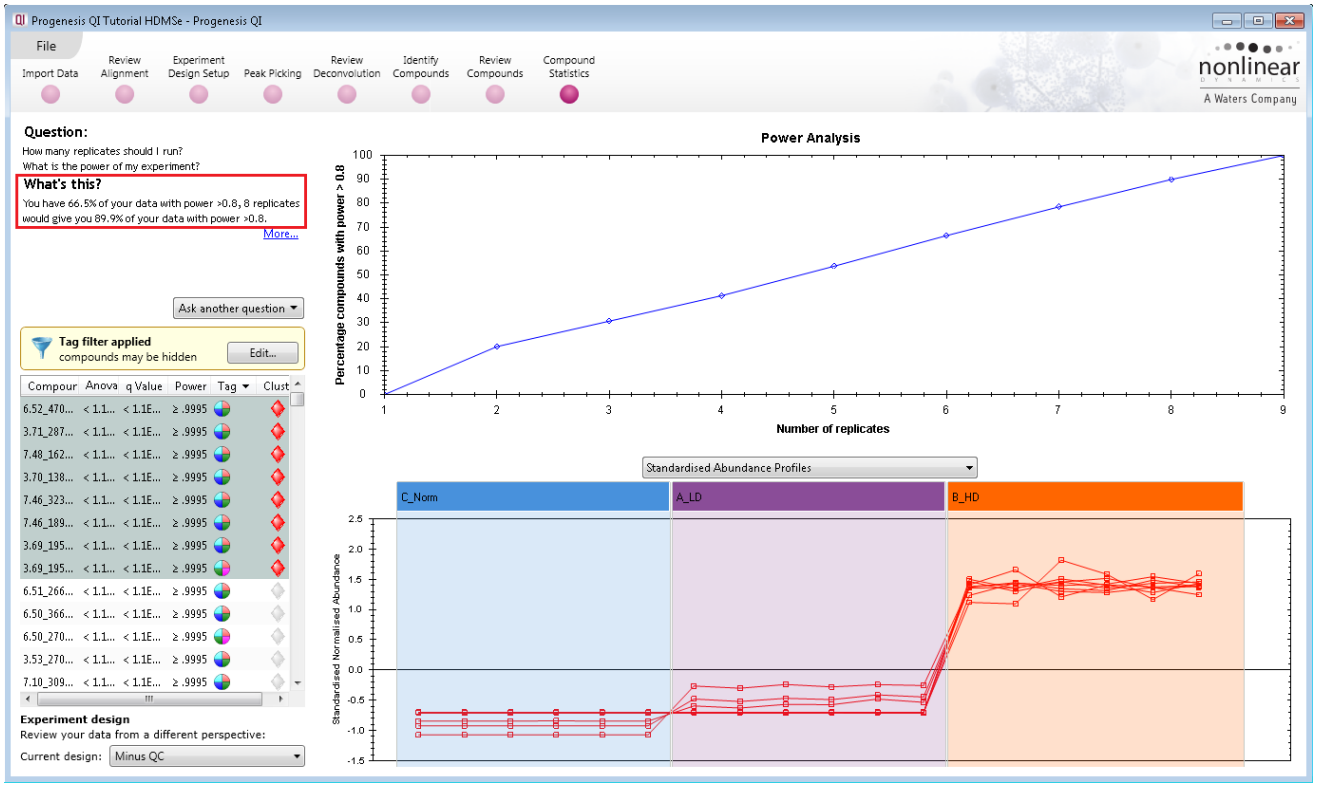
Select the option

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

It answers this question by informing you:

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

Using only the compounds tagged **Anova p<0.05 (i.e. 1342 compounds),** as an example, view the power analysis.



This is displayed graphically showing that 66.5% of the 1342 compounds have a power of 80% or that 8 replicates would give you 89.9% of your data with power > 0.8.

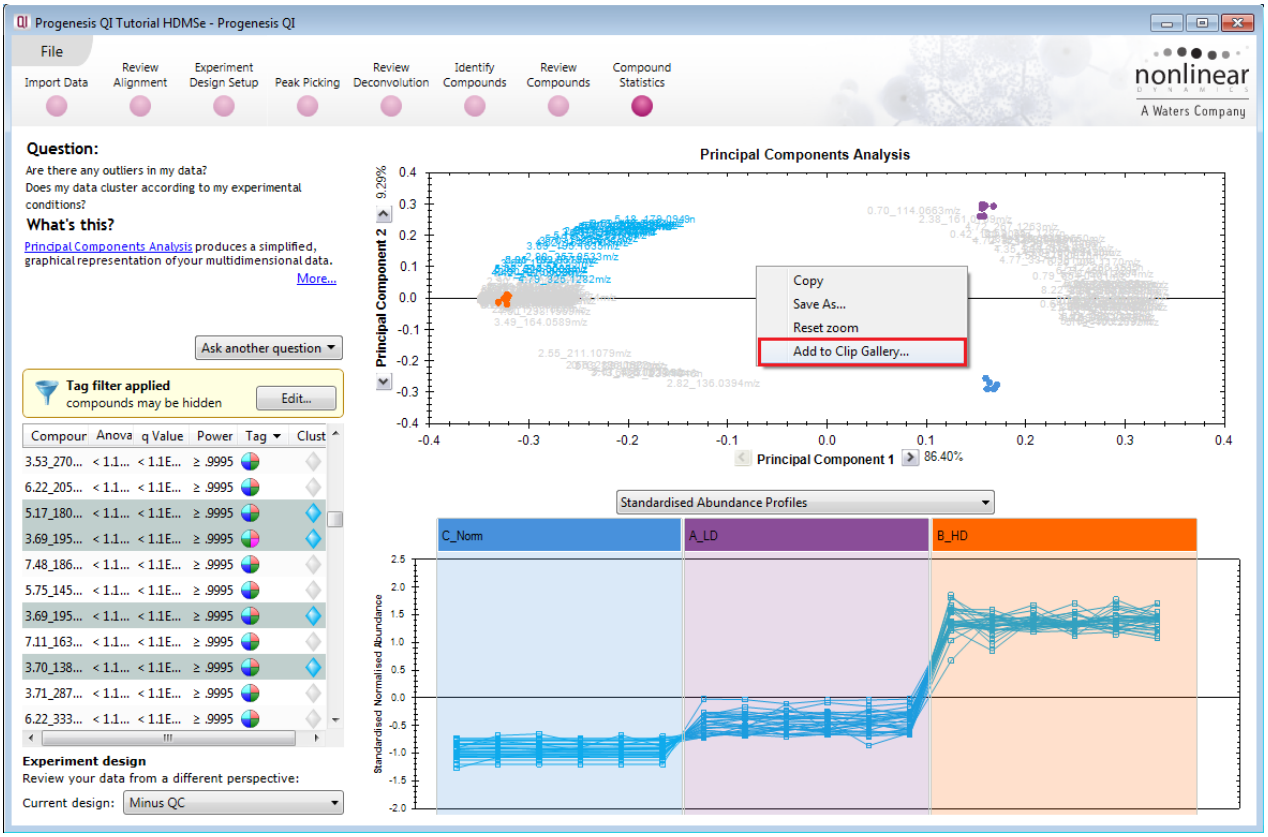
Appendix 9: Using Clip Gallery to Save and Export Pictures and Data

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

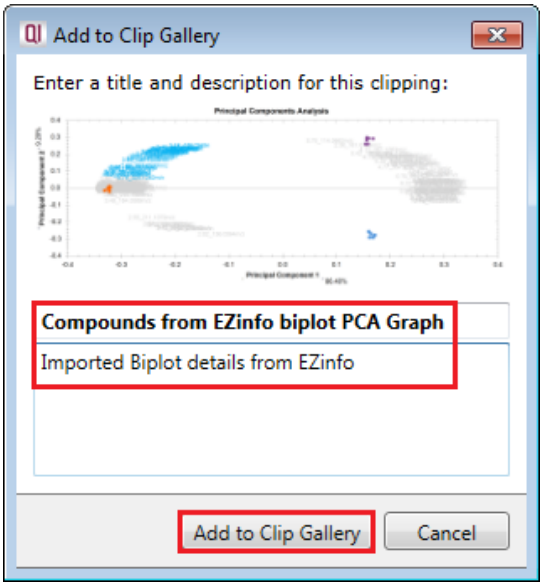
The saved images are retained as part of the experiment and are stored accordingly. This facility allows you to capture (high resolution) images that can be used in the development of specific reports and/or used as part of the process of publishing your experimental findings.

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

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

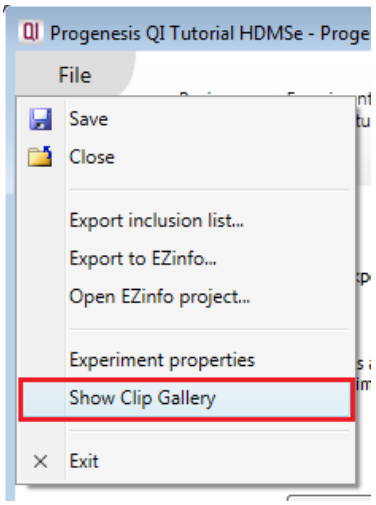


This will open a dialog displaying what is to be saved and allows you to alter the title and provides a description of the image for later reference.

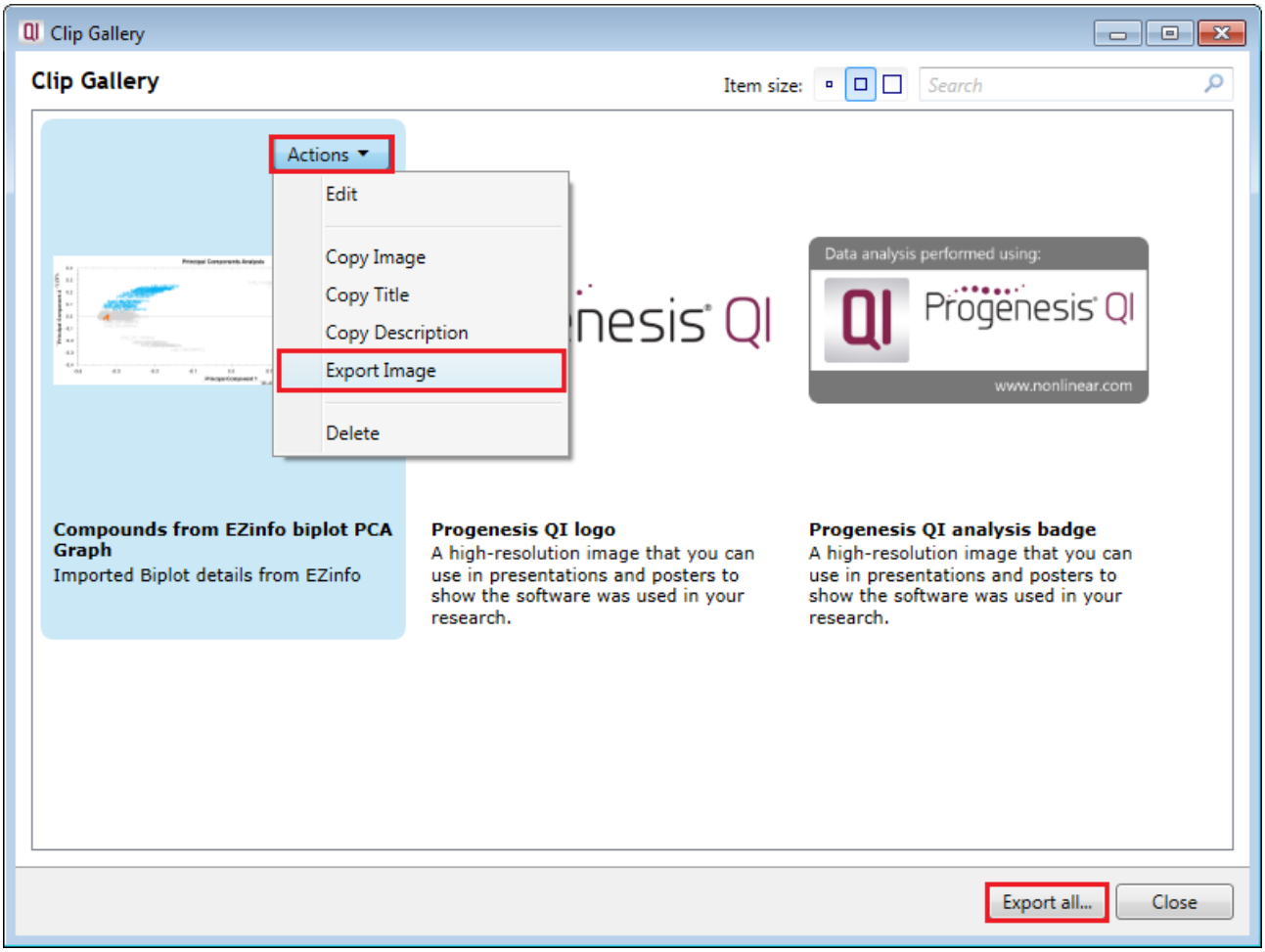


Enter details as required and click **Add to clip gallery**

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



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



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

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