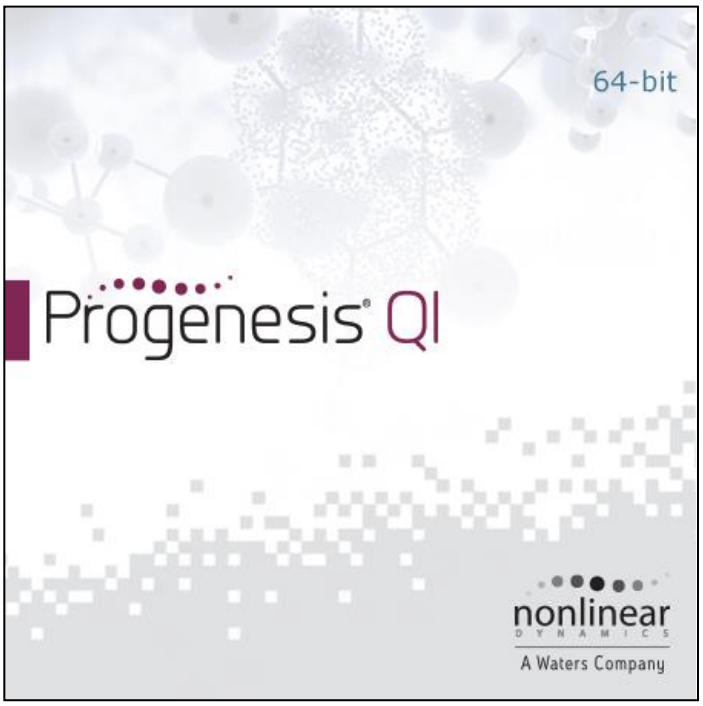


Progenesis QI
User Guide

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 features (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 57) then start at page 8.

How to use this document

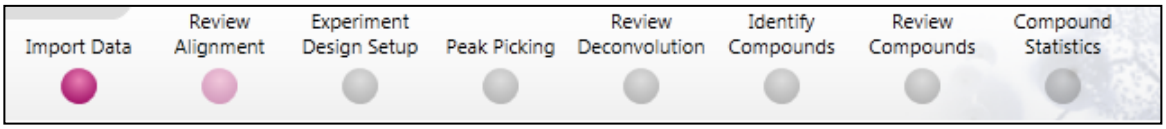
You can print this user guide to help you work hands-on with the software. The complete user guide takes about 60 minutes and is divided into two sections. This means you can perform the first half focused on run alignment and analysis then complete the second half of analysis exploring comparative differences and Compound identity at a convenient time.

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.

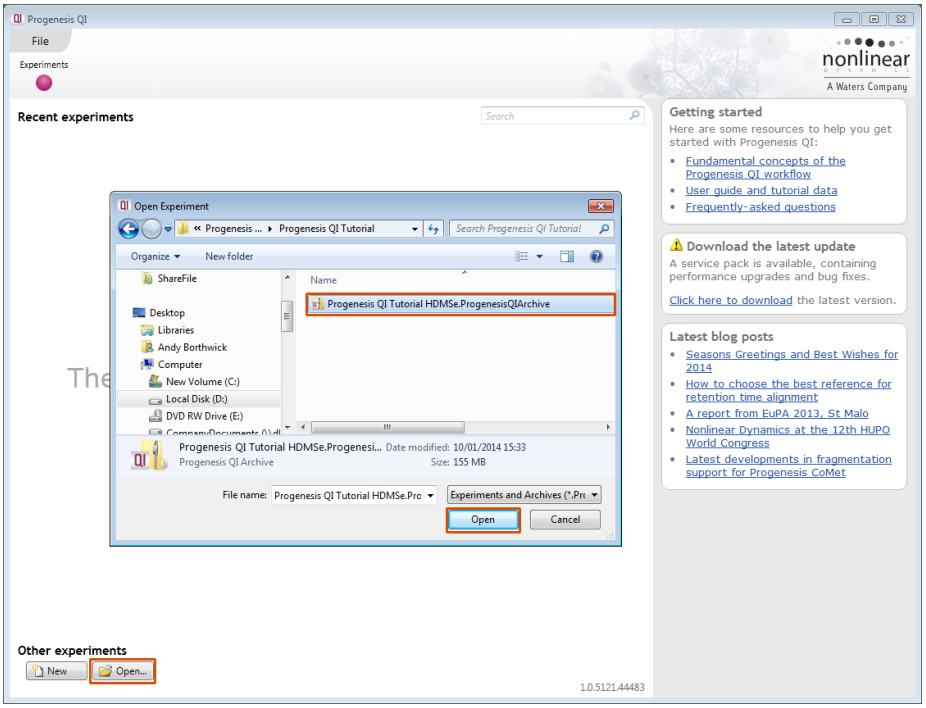


Restoring or starting a new Metabolomics Tutorial

If working with your original data files then refer to Appendix 1 page 57

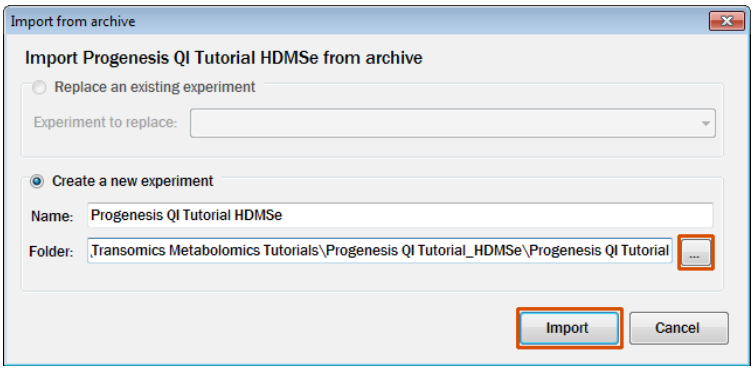
Open Progenesis QI and download the Compressed (.zip) Tutorial Archive file from the **'User guide and tutorial data**' link shown below, placing it in a **new folder** on your desktop. Before restoring the tutorial in the software **you must** first right click on the downloaded (.zip) file and extract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

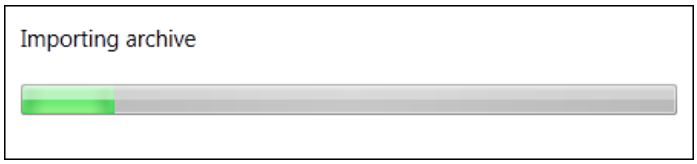


This opens the 'Import from archive' dialog.

Select the **Create a new experiment** option and select the folder in which you placed the archive, using the icon (to the right).

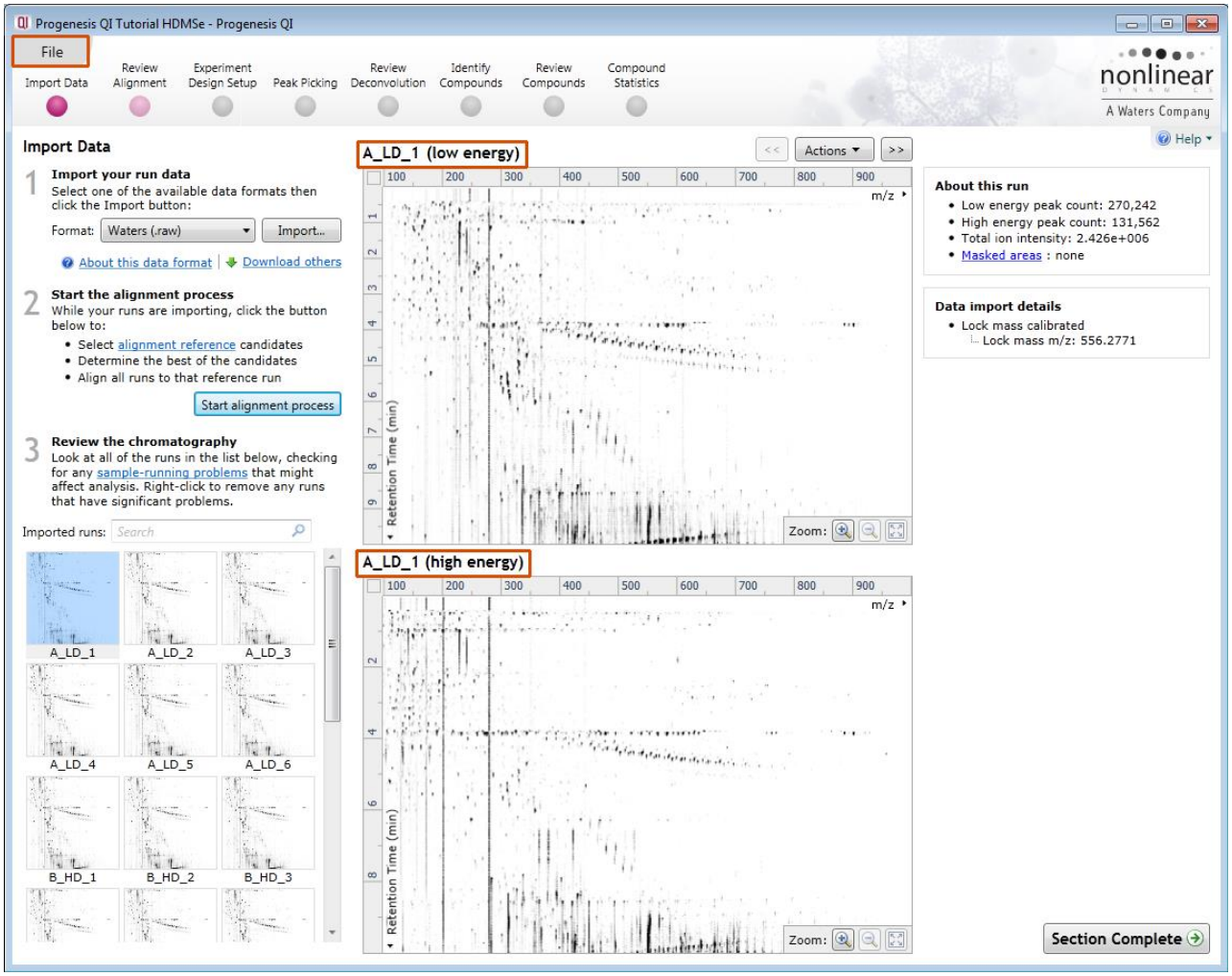


Then press **Import.**

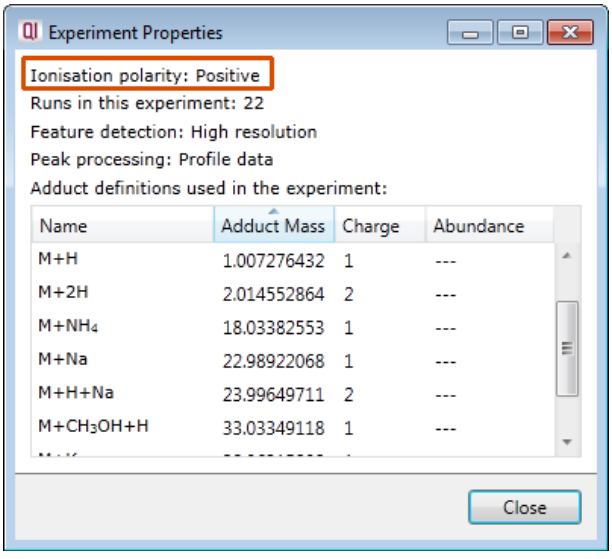


Note: use the **Replace an existing experiment** option if you want to over-write an existing version of the tutorial.

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

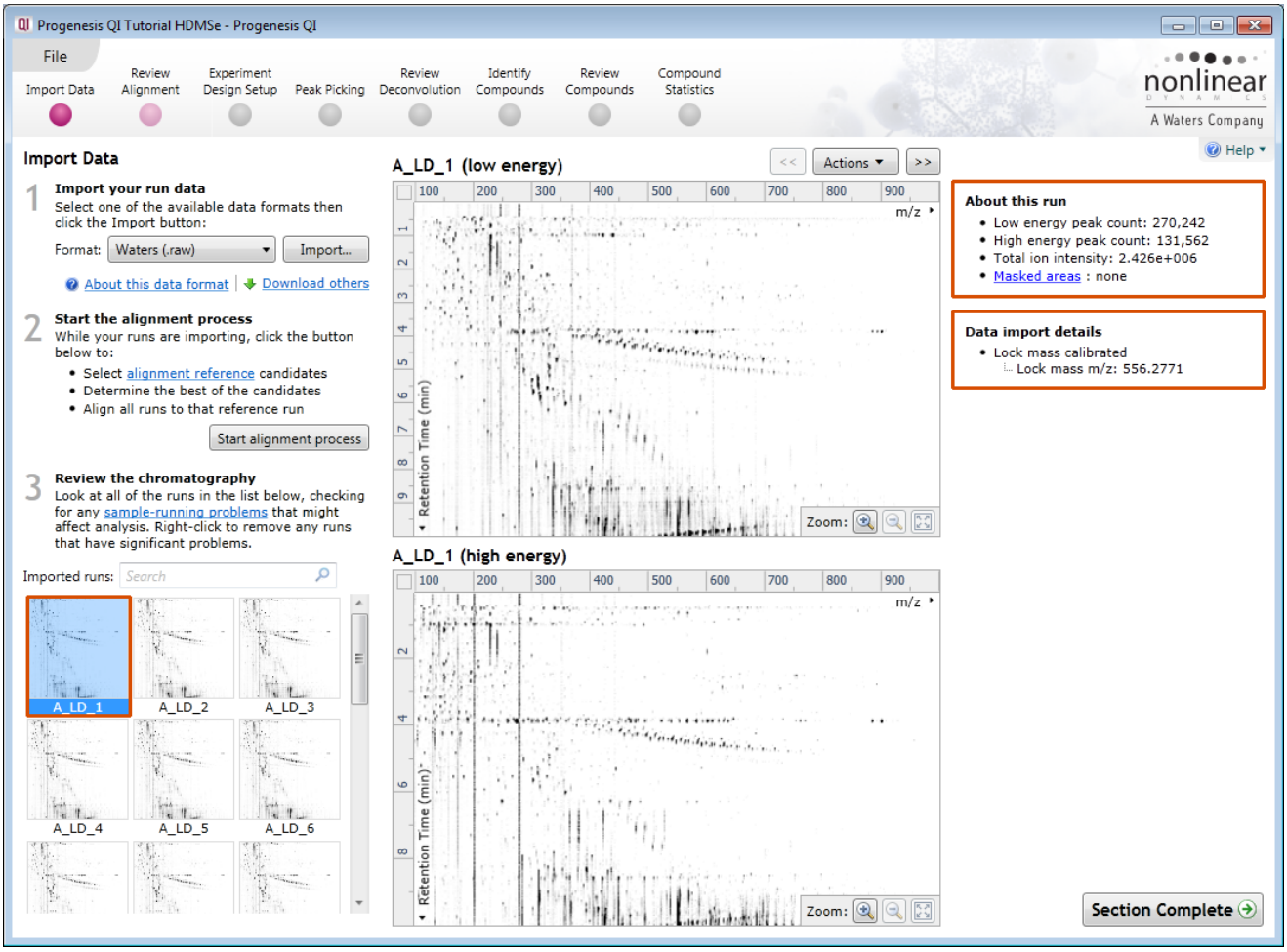
As this example is using HDMSe data it contains both low and high energy exact mass data.

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

Using the workflow icons, at the top of the screen, click on **Import Data**.

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 deleted 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.

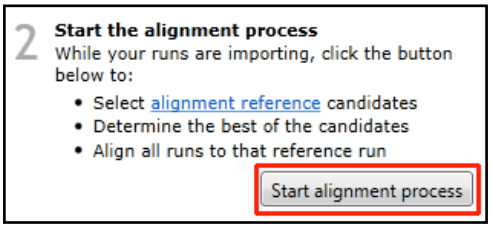
Tip: the **'Mask areas for peak picking'** facility (under **Actions**) allows you to exclude areas (usually early and/or late in the LC dimension (Retention Time) that appear excessively noisy due to capture of data during column regeneration. This is **NOT** used for this data set.

Examine the quality of the run(s) using the Zoom facility on the main window.

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

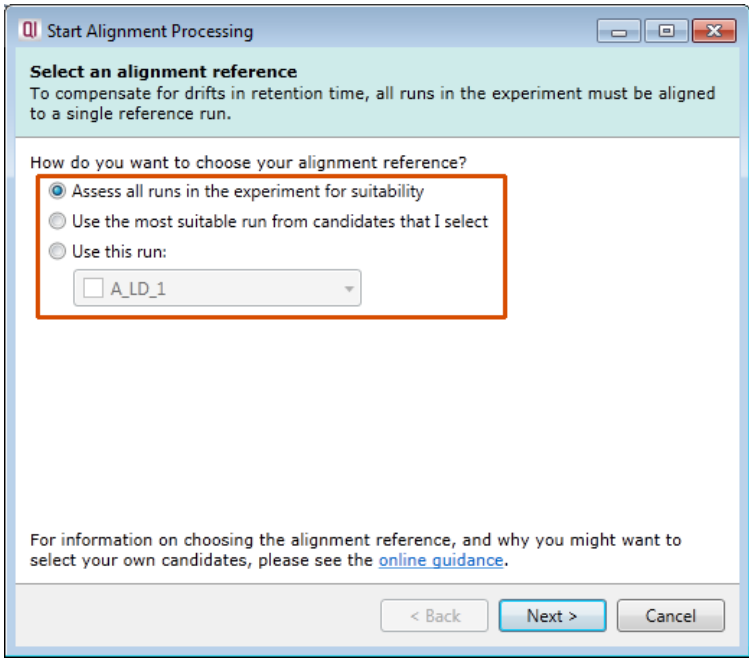
Stage 2: Automatic Alignment of your data

The process of alignment of your data can be started before the import of your data has been completed by clicking on **Start alignment process.**



During this process the software will Align all your runs to a Reference run which can either be selected automatically by the software or manually selected by you.

In this tutorial example you have 22 runs which have already been imported so to start the process click on **Start alignment process**



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.
   The run with the greatest similarity to all other runs is chosen 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 can take a long time.

2. **Use the most suitable run from candidates that I select**
   This method asks you to choose a selection of reference candidates, and the automatic algorithm chooses the best reference from these runs.

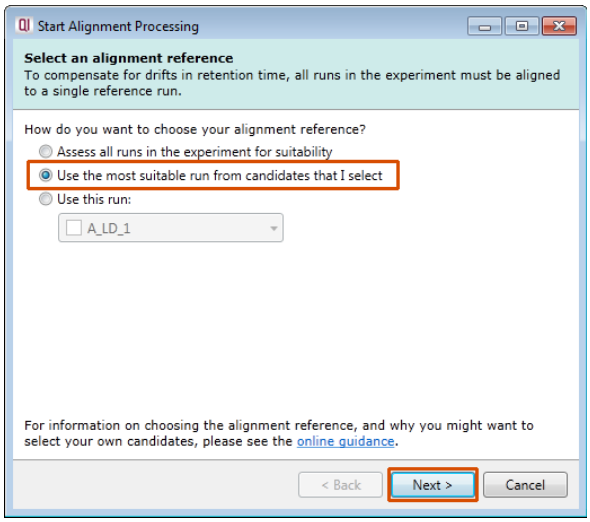
When you have some prior knowledge of your runs suitability as references:
   runs from pooled samples (QC samples)
   runs for one of your experimental conditions will contain the largest set of common compound ions.

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

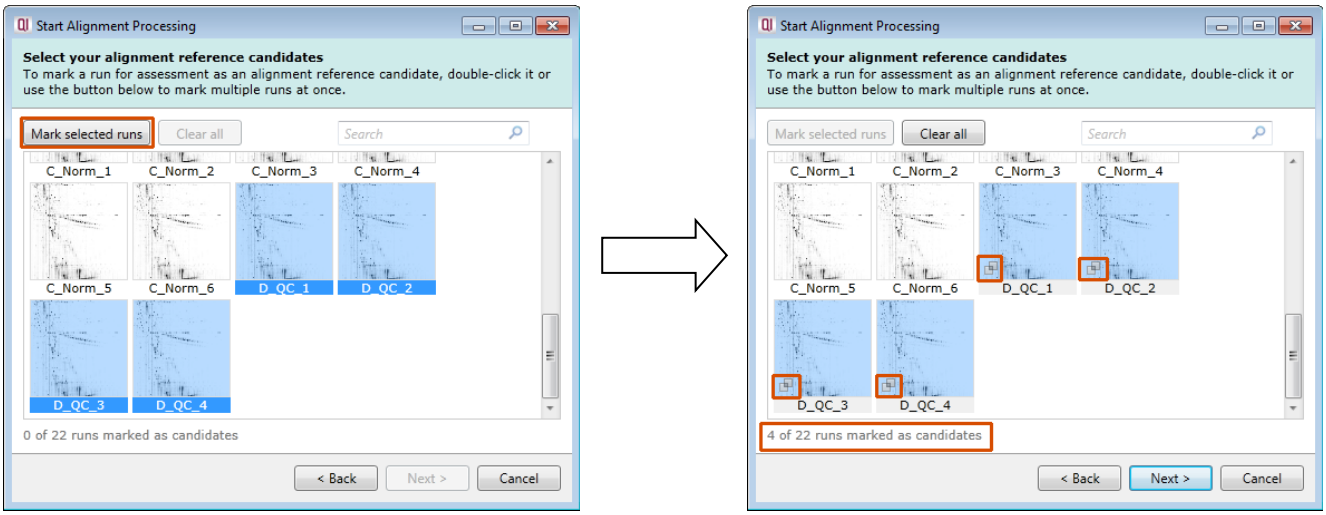
Manual selection gives you full control, but there are a couple of risks to note:

- 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).

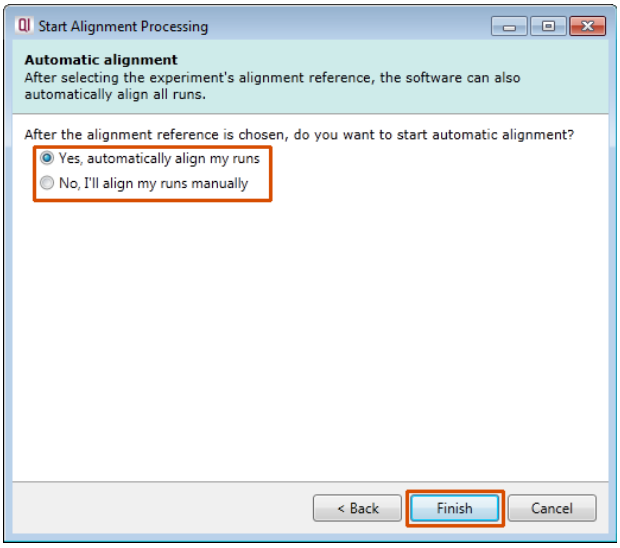
- For this tutorial we will select the **second** option (See Appendix 1, page 57 for more details on using the other options). Select option 2 and click **Next**.



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**.

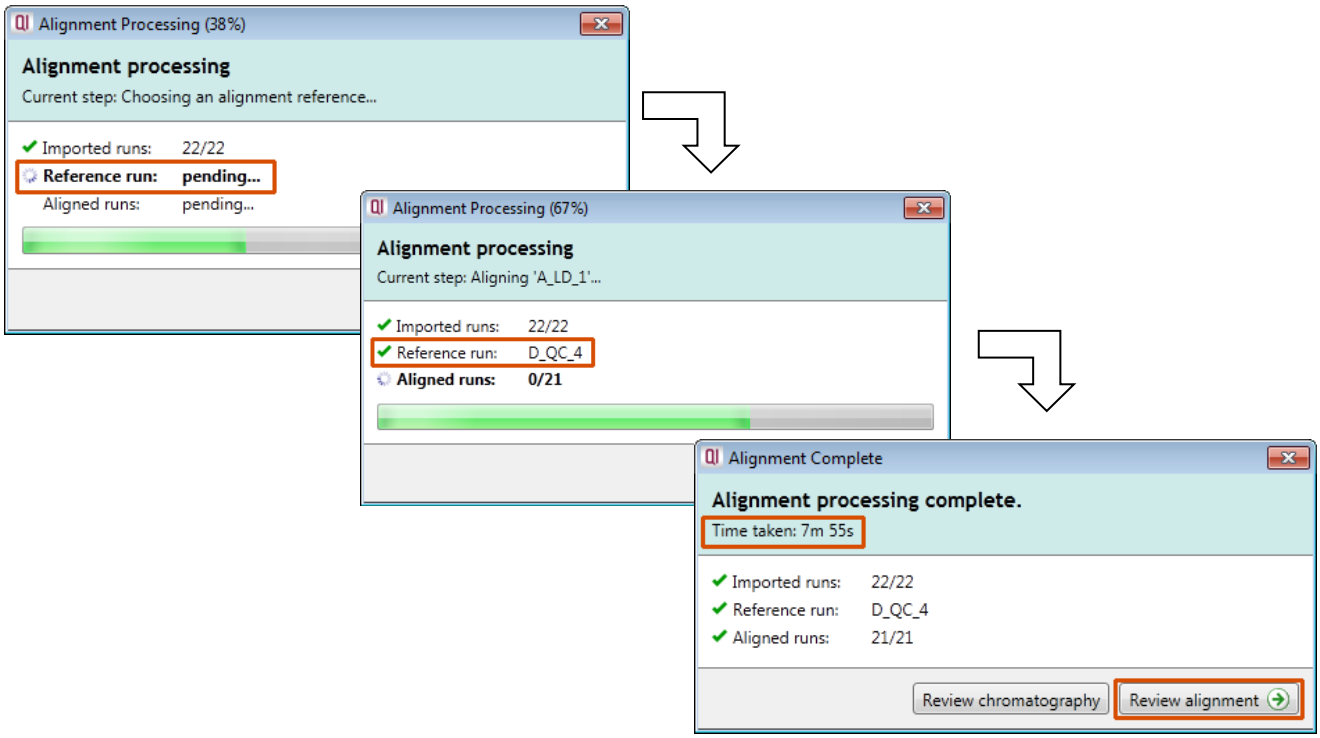


You will now be asked if you want to align your runs automatically or manually.



Select automatically and click finish.

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



Once the Reference run has been chosen the automatic alignment is then performed. As the whole process proceeds you get information on what stage has been performed and also the % of the process that has been completed.

When the Alignment completes you can either review the chromatography or go to the Review Alignment stage using the options on the Alignment Dialog.

Click **Review Alignment**.

Stage 3: Licensing

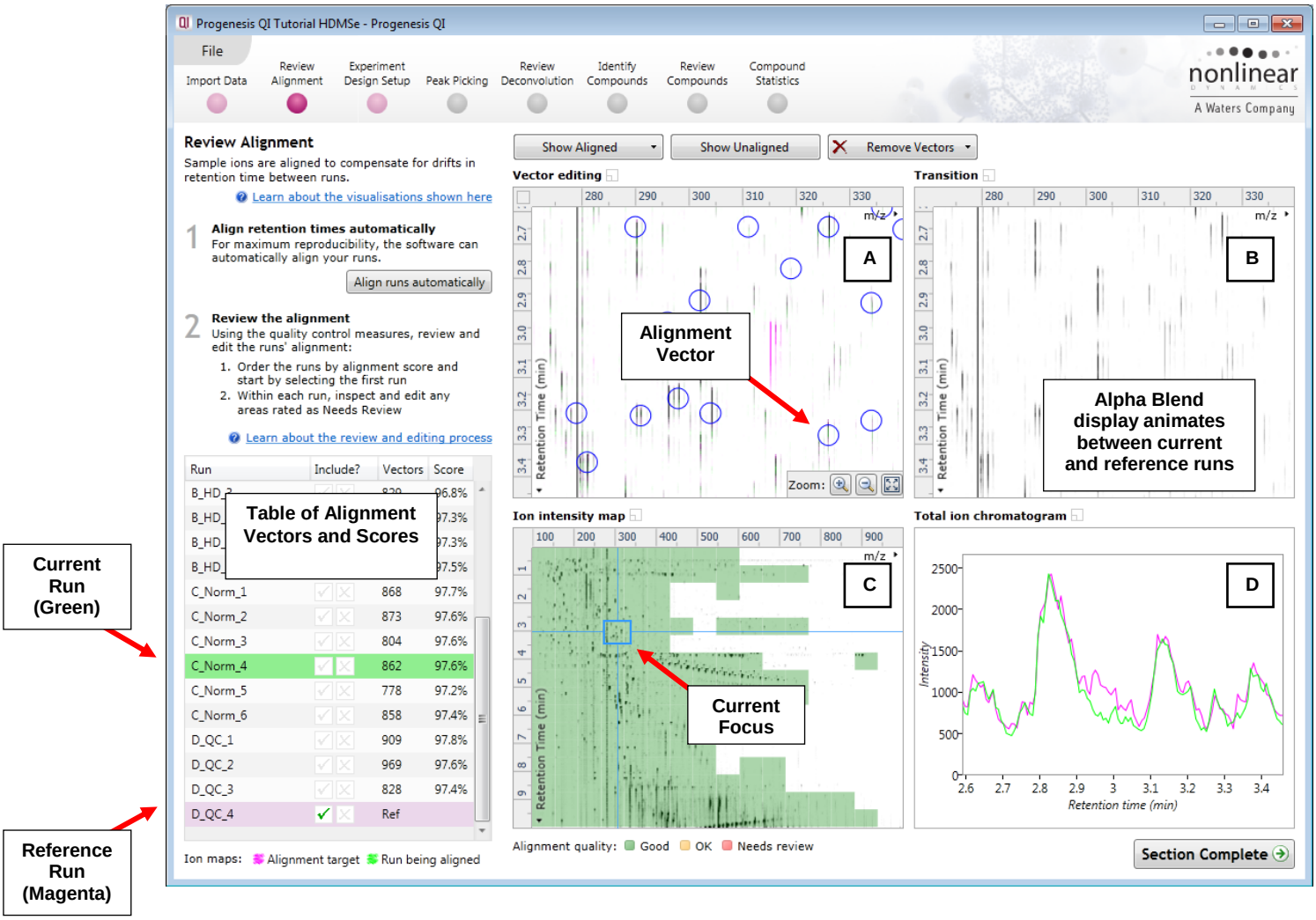
This stage in the analysis workflow will **only** appear if you are using 'Unlicensed' data files to evaluate the software and have no dongle attached.

For details on how to use Licensing go to Appendix 2 (page 66)

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

Stage 4: Review Alignment

At this stage **Review Alignment** opens displaying the alignment of the runs to the Reference run (D_QC_4). Change the current run to C_Norm_4 by clicking on it in the table.



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.

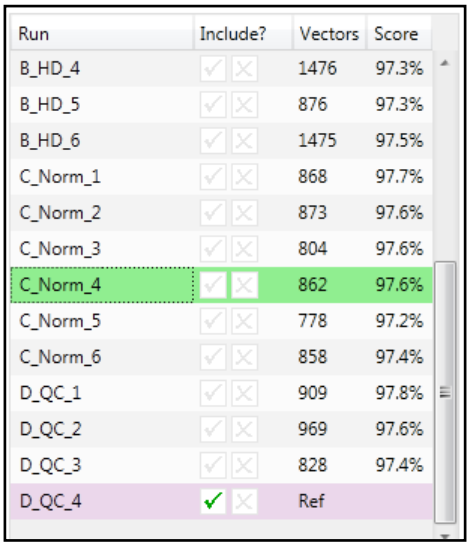
Transition (Window B): uses an **alpha blend** to animate between the current and reference runs. Before the runs are aligned, the features appear to move up and down. Once correctly aligned, they will appear to pulse. During the process of adding vectors, this view can be used to zoom in and also navigate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

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

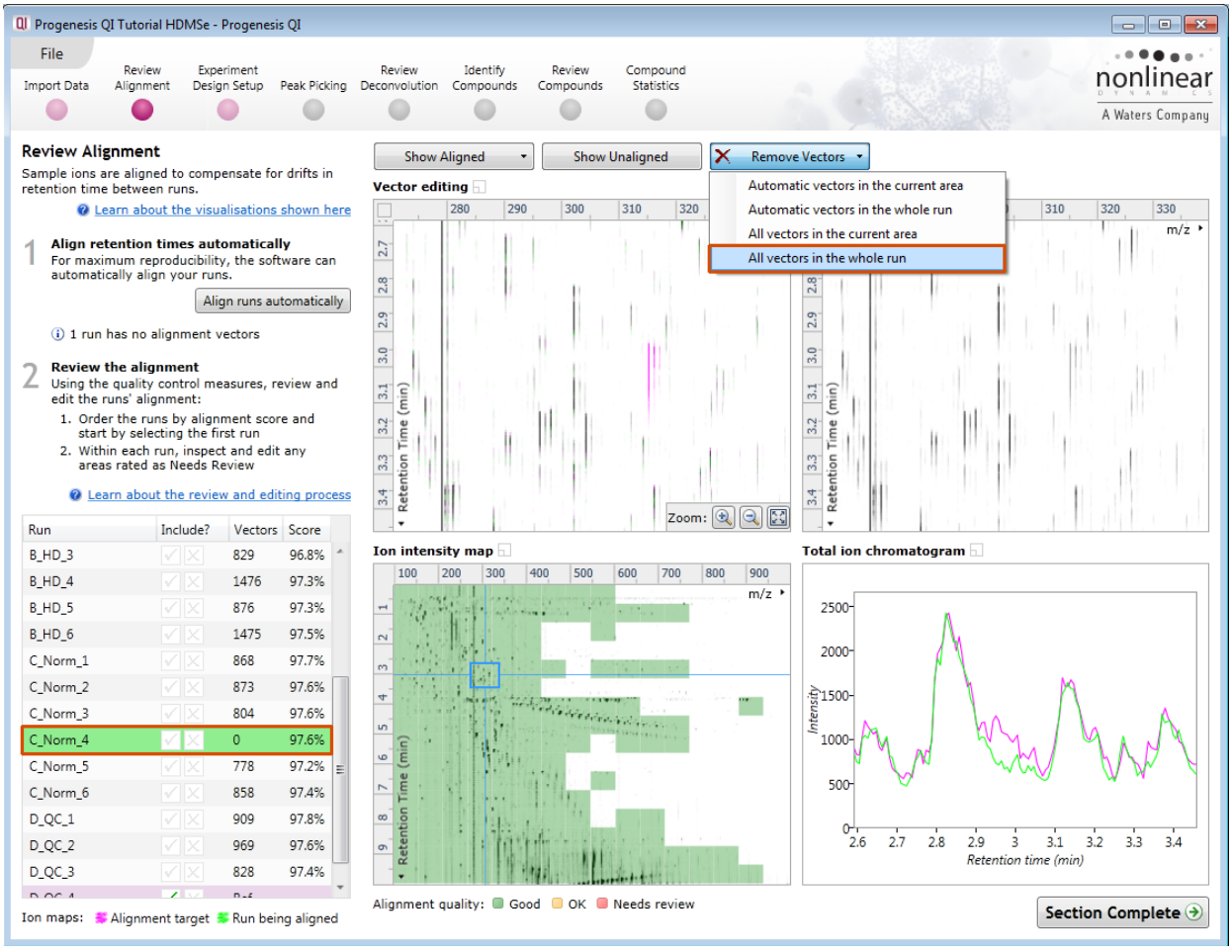


Run	Include?	Vectors	Score
B_HD_4		1476	97.3%
B_HD_5		876	97.3%
B_HD_6		1475	97.5%
C_Norm_1		868	97.7%
C_Norm_2		873	97.6%
C_Norm_3		804	97.6%
C_Norm_4		862	97.6%
C_Norm_5		778	97.2%
C_Norm_6		858	97.4%
D_QC_1		909	97.8%
D_QC_2		969	97.6%
D_QC_3		828	97.4%
D_QC_4	✓	Ref	

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

- 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.

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



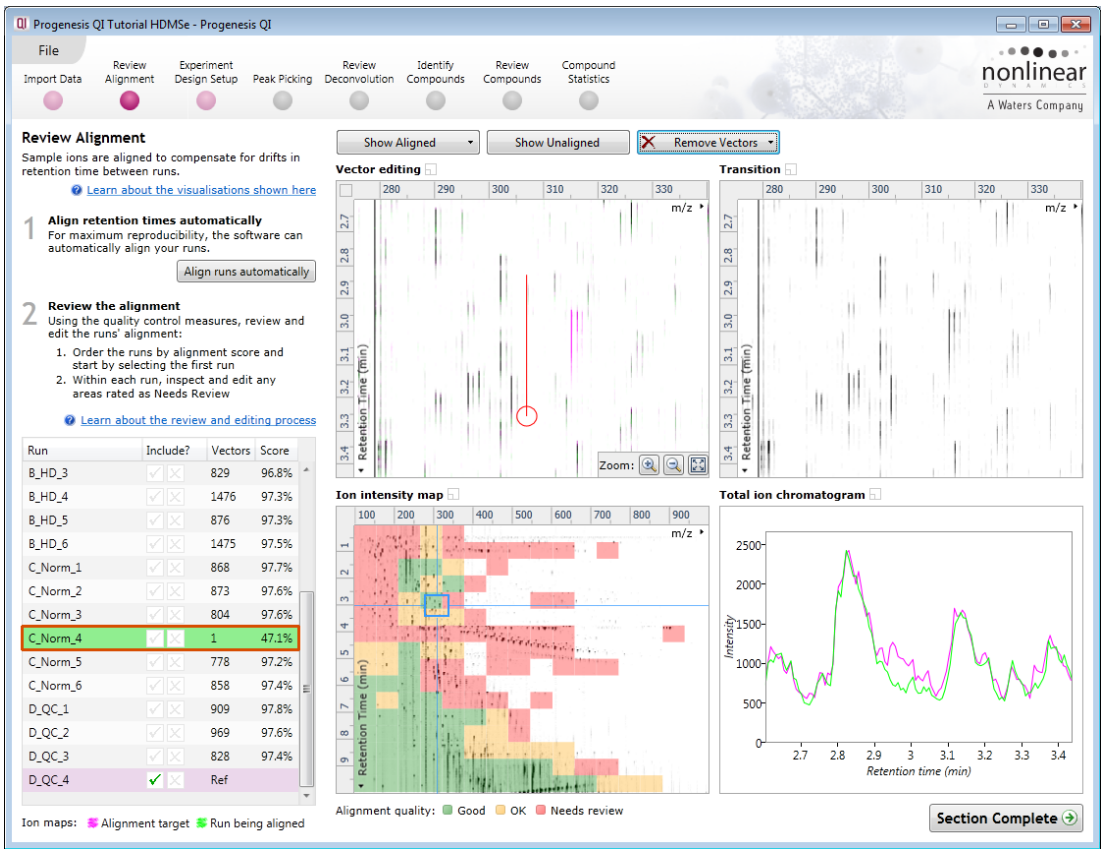
Run	Include?	Vectors	Score
B_HD_3		829	96.8%
B_HD_4		1476	97.3%
B_HD_5		876	97.3%
B_HD_6		1475	97.5%
C_Norm_1		868	97.7%
C_Norm_2		873	97.6%
C_Norm_3		804	97.6%
C_Norm_4		0	97.6%
C_Norm_5		778	97.2%
C_Norm_6		858	97.4%
D_QC_1		909	97.8%
D_QC_2		969	97.6%
D_QC_3		828	97.4%

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.

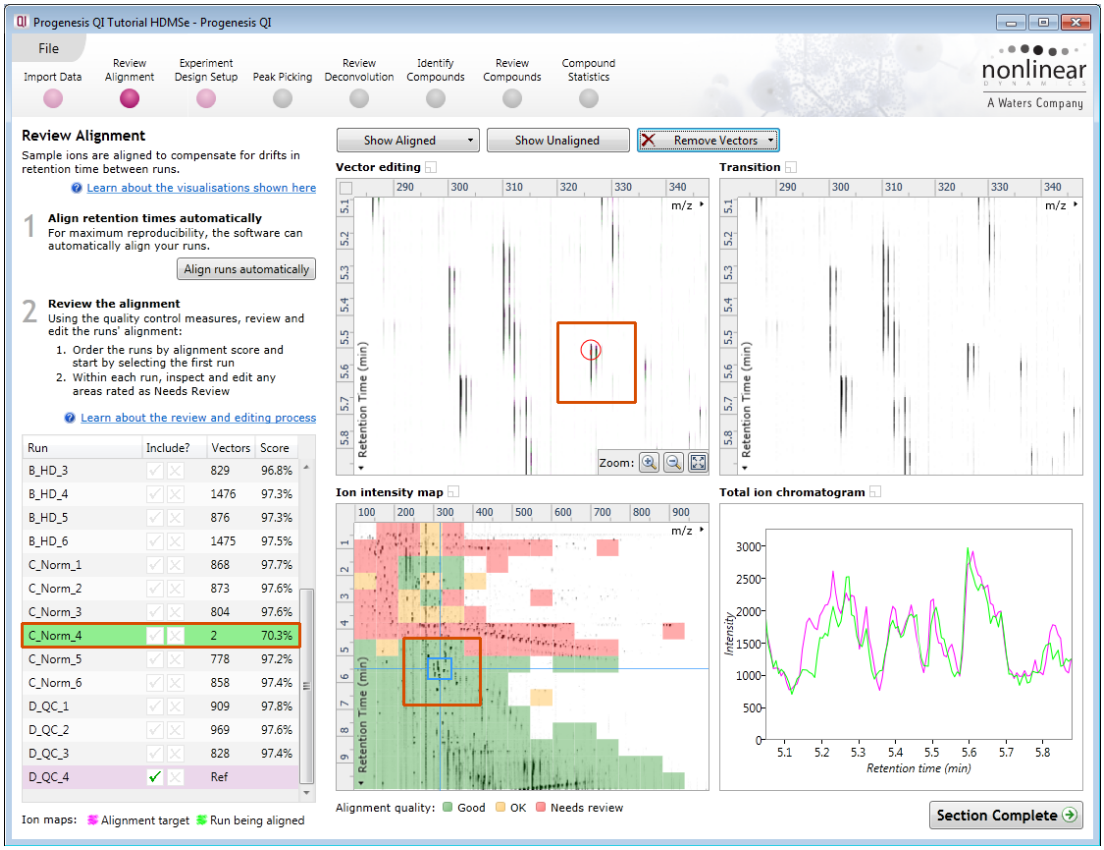
To simulate poor alignment, place a single manual vector on the Vector editing view (Window A). To do this click and drag out a single vector then release the mouse button. By doing this a single manual vector will appear with a length corresponding to the 'drag'.

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 click and release **without** deliberately misaligning the vector.

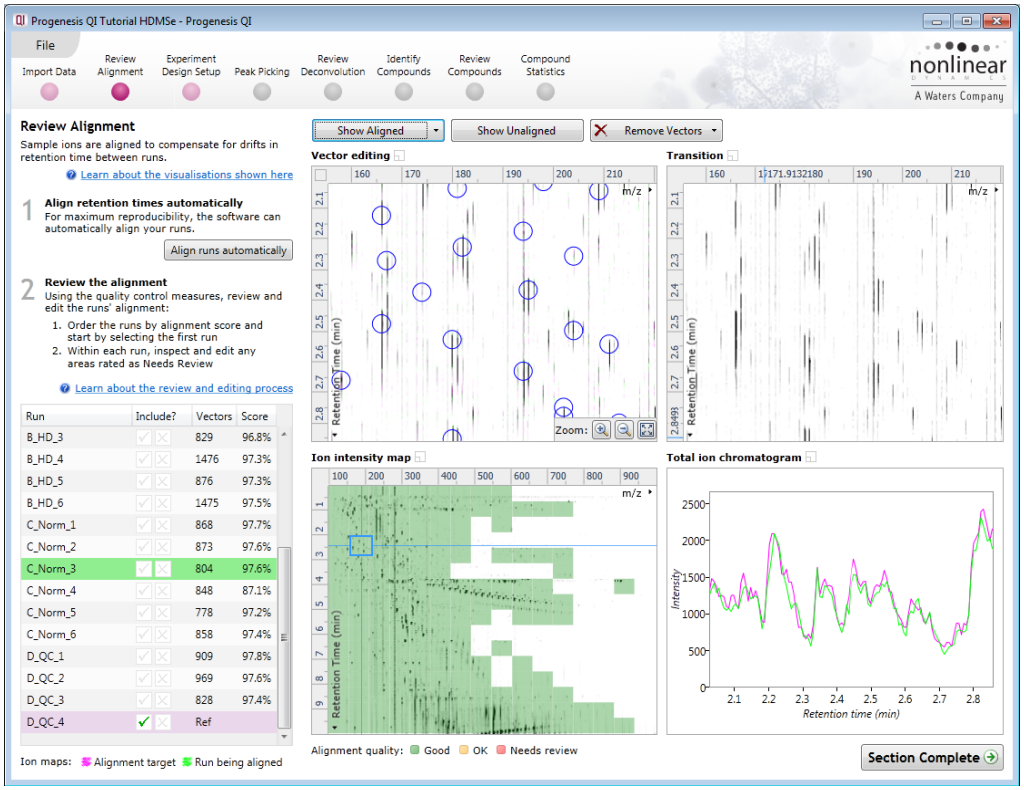


Notice how the correctly placed vector markedly improves the Score this is also reflected in the increased proportion of the Ion Intensity showing green indicating a good quality alignment.

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.

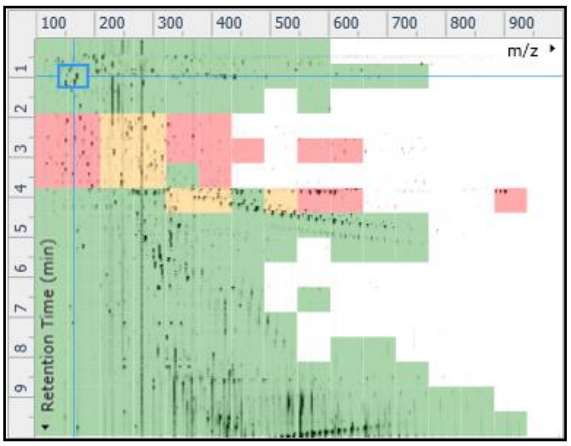


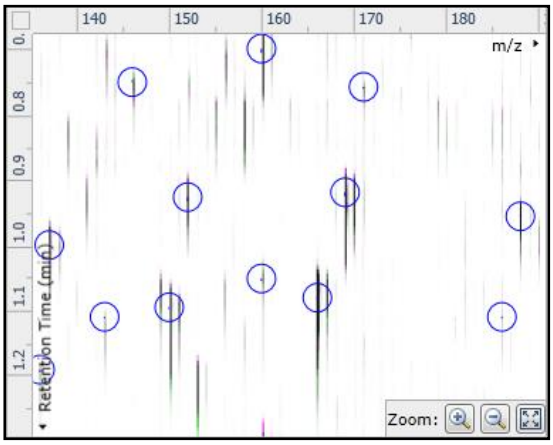
Using the **Simulated** miss-aligned example from the previous section to explain the review process for alignment, the alignment looks as below with a region of poor alignment (highlighted in red).

Reviewing Quality of Alignment

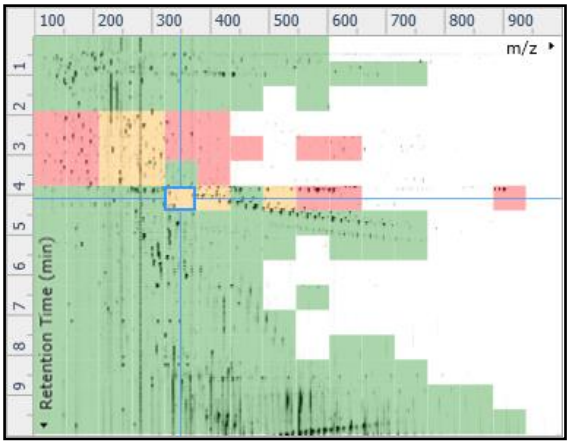
At this point the quality metric, overlaid on the Ion Intensity Map as coloured squares, acts as a guide drawing your attention to areas of the alignment. These range from Good (Green) through OK (Yellow) to Needs review (Red). Drag out a 'Focus' area that corresponds to one of the coloured squares.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

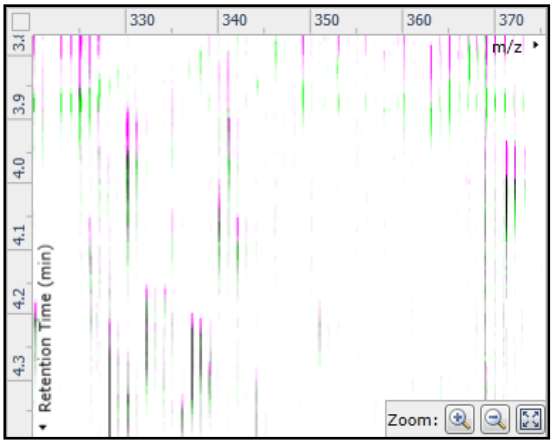
For a 'green' square the majority of the data appears overlapped (black) indicating good alignment. When viewed in the Transition view the data appears to pulse.



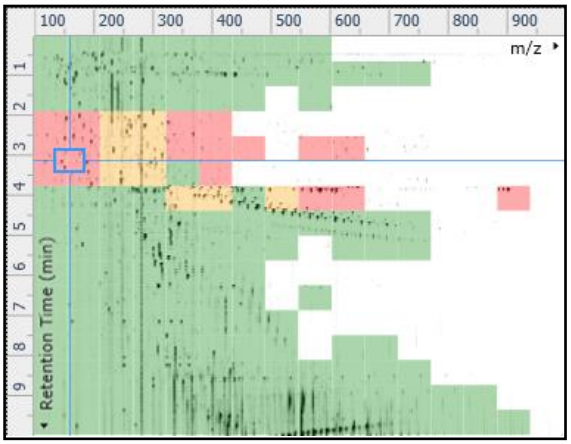


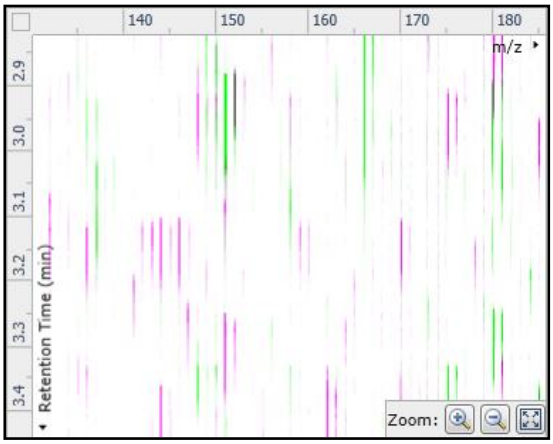
For a 'yellow' square some of the data appears overlapped (black) indicating OK alignment. When viewed in the Transition view some of the data appears to pulse.





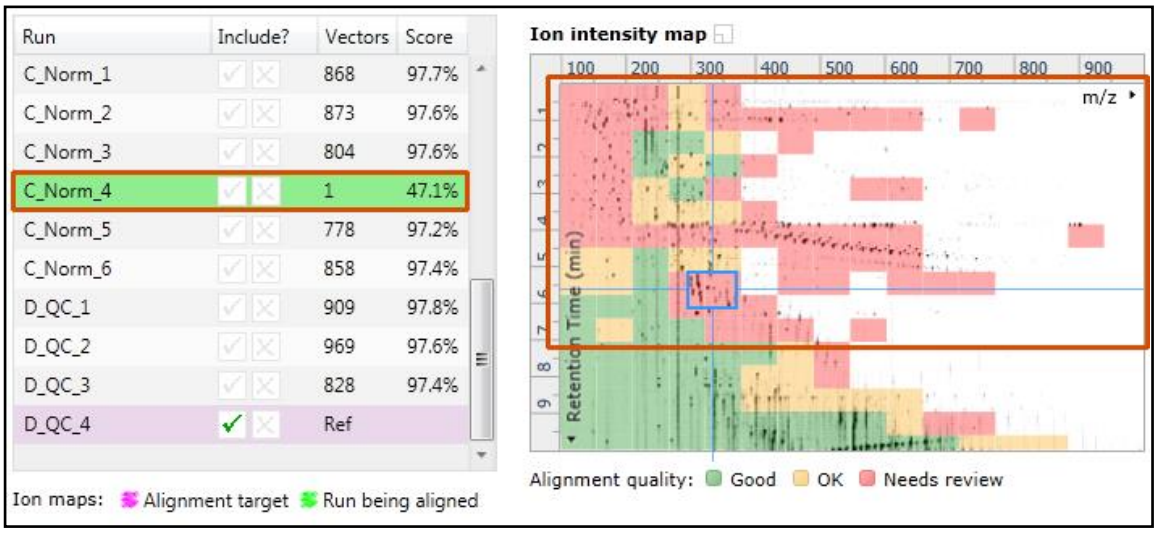
For a 'red' square little if any of the data appears overlapped (black) indicating questionable alignment. When viewed in the Transition view little data appears to pulse.





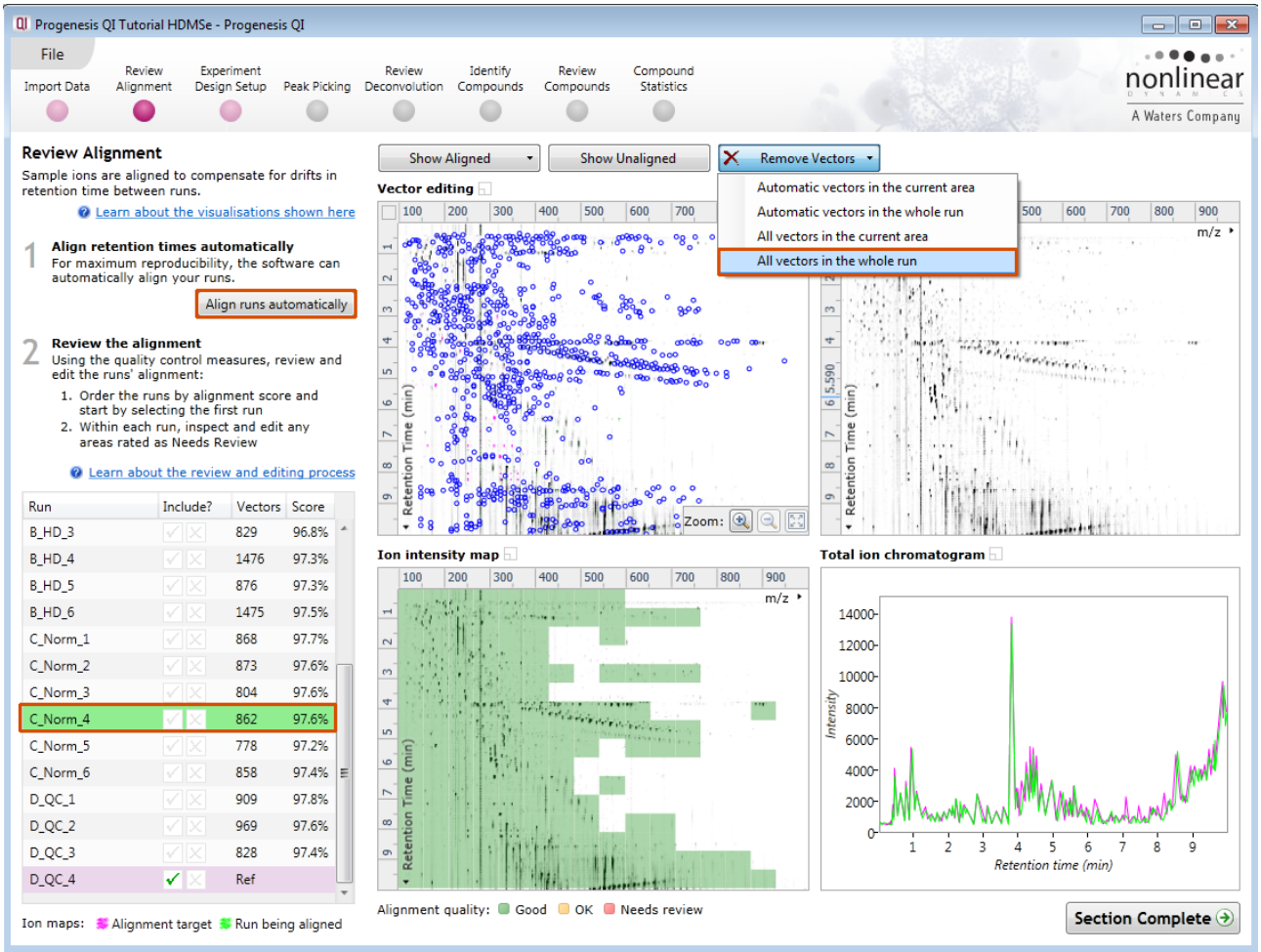
Note: the coloured metric **should be used as a guide**. In cases where there are a few 'isolated' red squares this can also be indicative of 'real' differences between the two runs being aligned and should be considered when examining the overall score and surrounding squares in the current alignment.

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 3, page 67).

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.

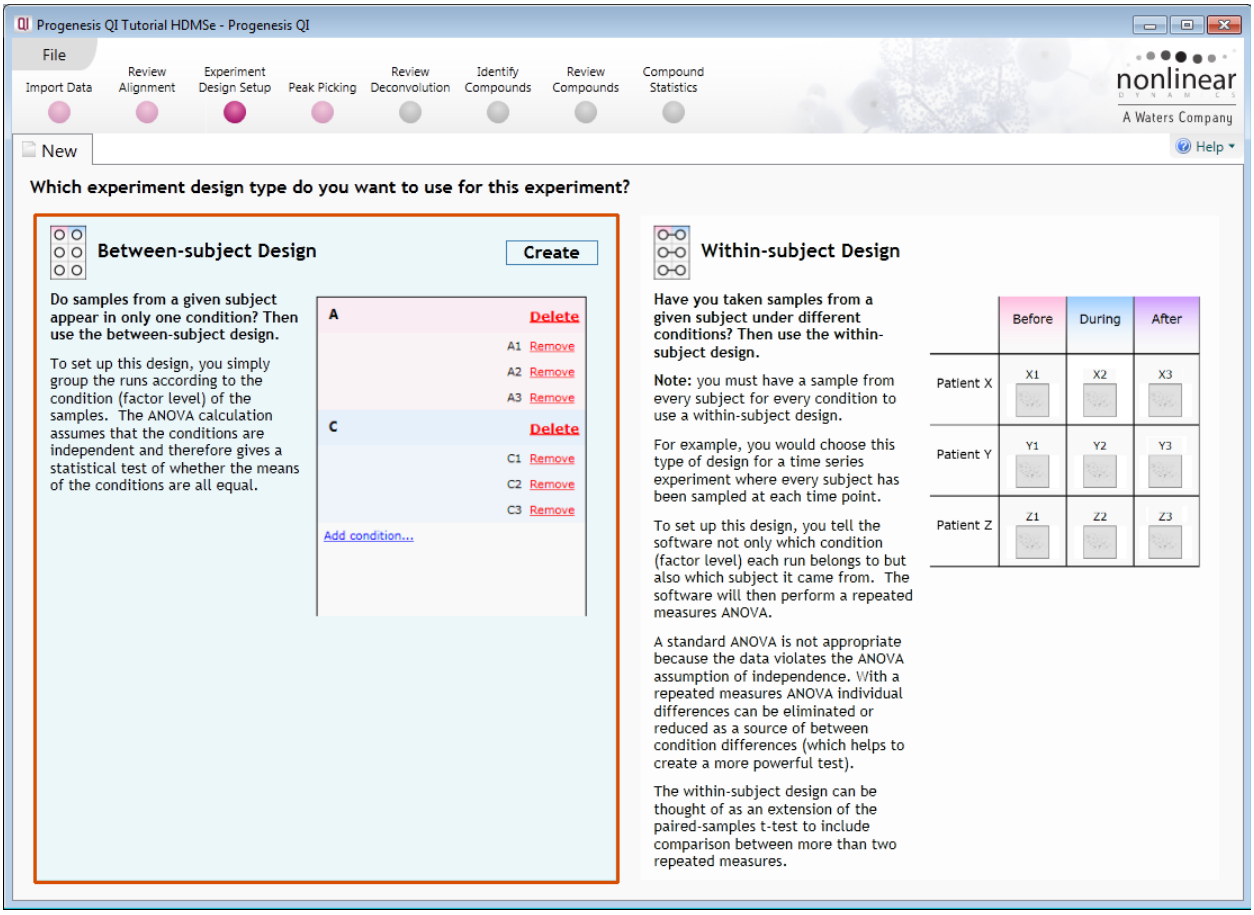


Having aligned the runs automatically click **Section Complete** to move to Experiment Design Setup.

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.

There are two basic types of experimental designs:



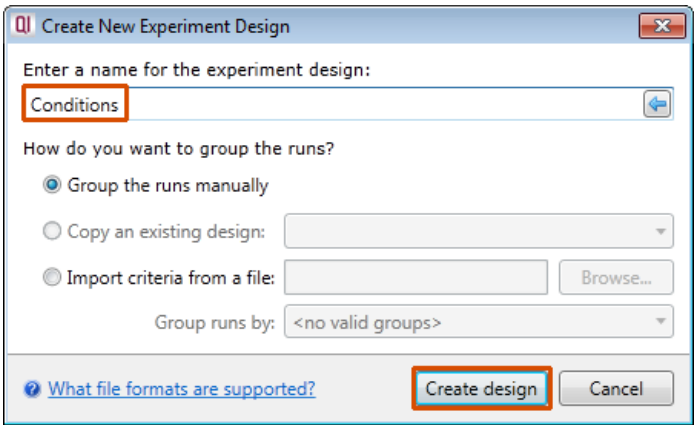
Between-subject design: here samples from any given subject appear in only one condition. (i.e. control versus various drug treatments).The ANOVA calculation assumes that the conditions are independent and applies the statistical test that assumes the means of the conditions are equal.

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.

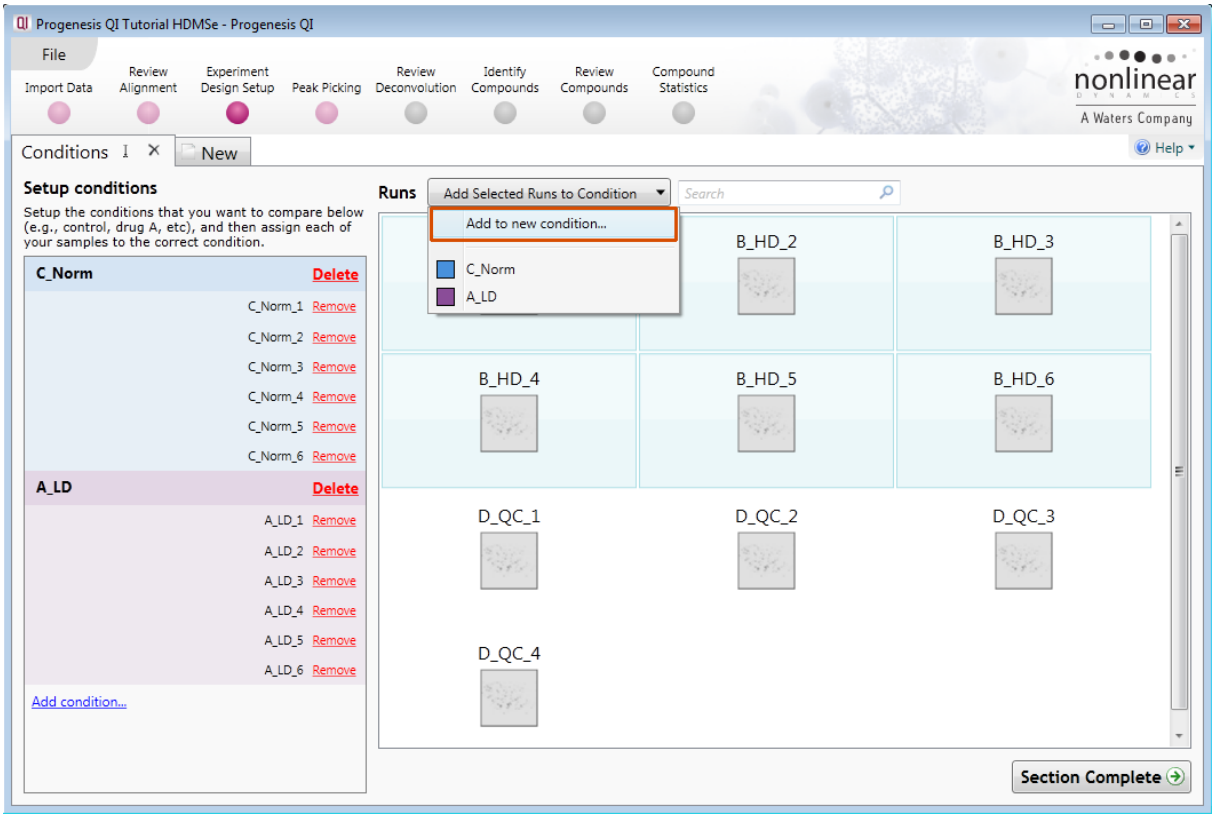
Additional information on how to apply the Within-subject Design is in Appendix 4 page 72

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**.



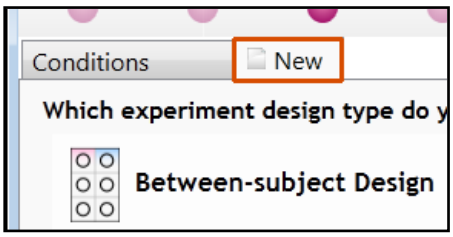
To create a new condition

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.

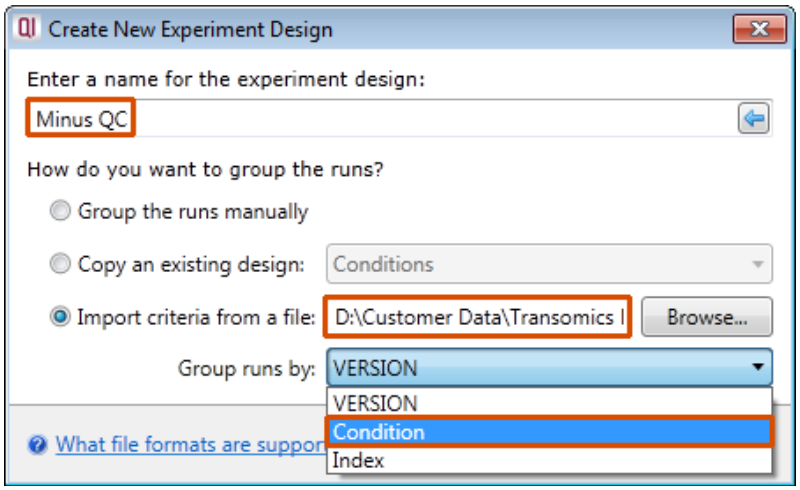
6. Repeat steps 1 to 5 until all the required runs are grouped into conditions.

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.



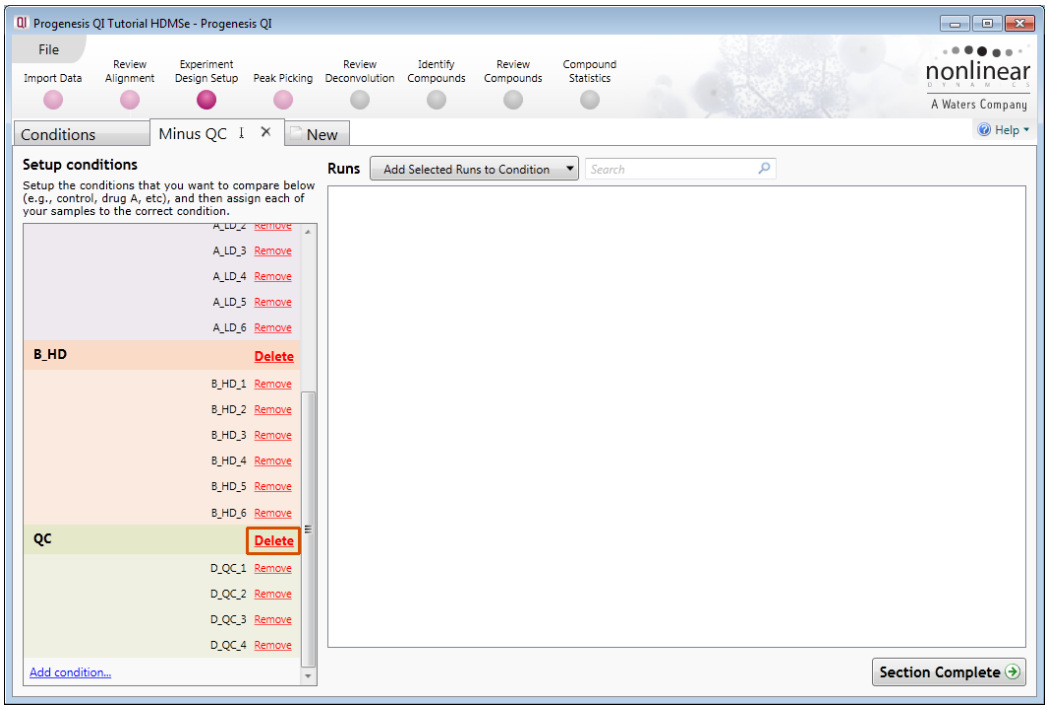
For this example there is a **Progenesis QI_HDMSe Tutorial.SPL** file available in the Experiment Archive you restored at the beginning of this tutorial exercise.

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 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**.



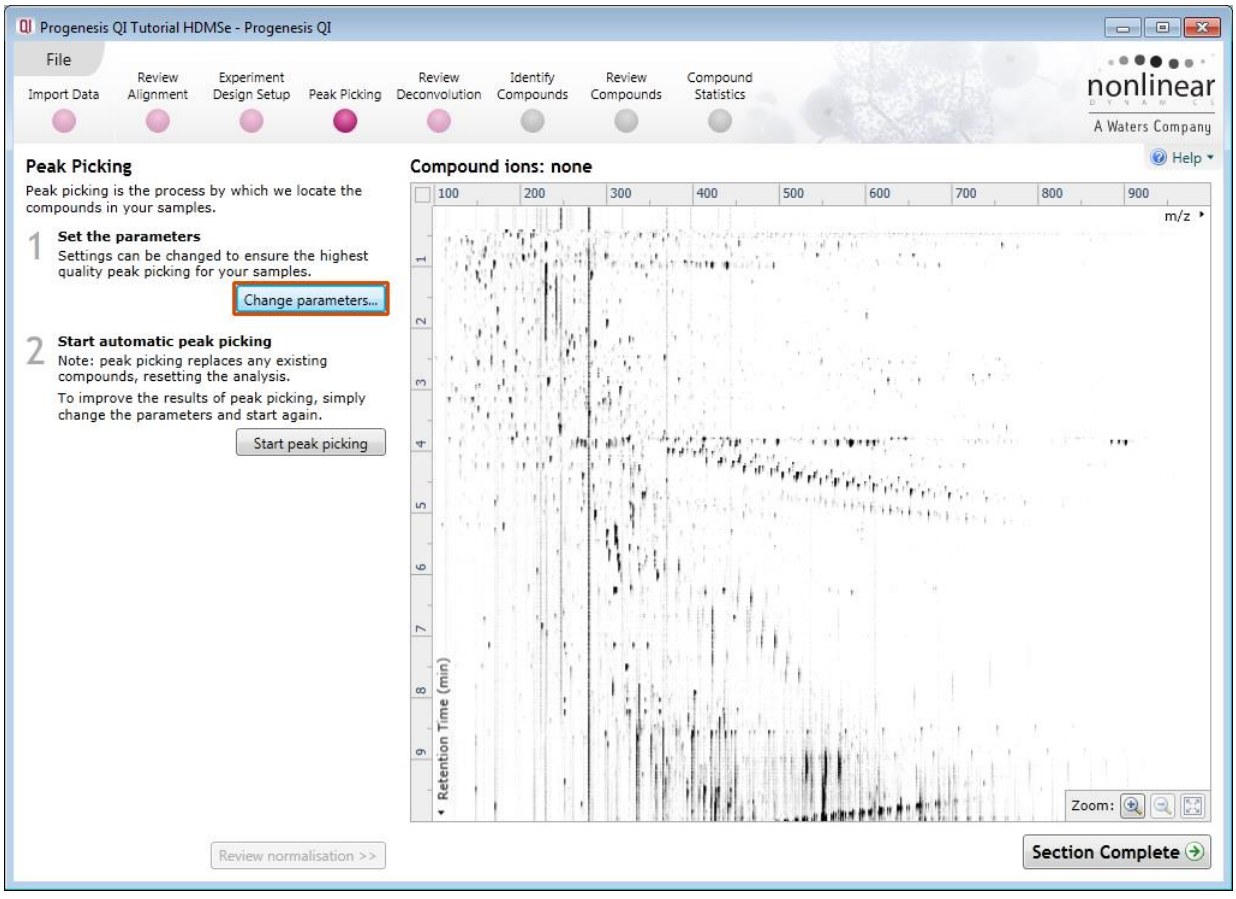
Note: On deleting a condition the runs will reappear in the Runs window.

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

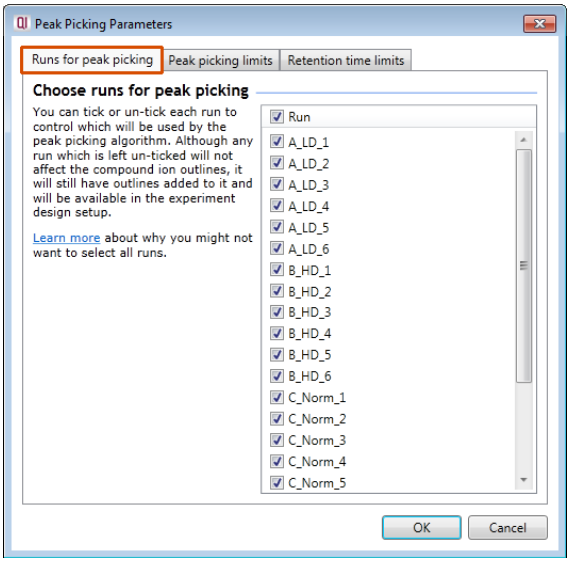
Having set up one or more Experimental Designs for your data the Peak Picking stage will open as shown below. Here you can easily define thresholds for the detection of peaks across all of your aligned runs.

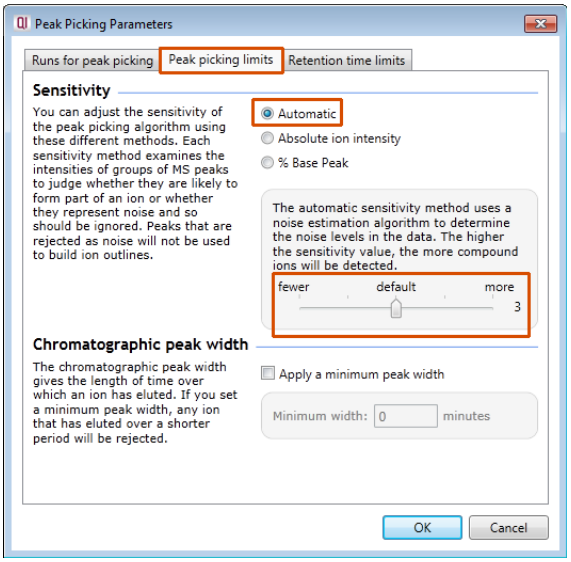


The ion intensity map displayed before Peak picking is for the Alignment Reference image. As the Peak Picking takes place this is replaced by the Aggregate map generated across all the aligned runs in your experiment.

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.

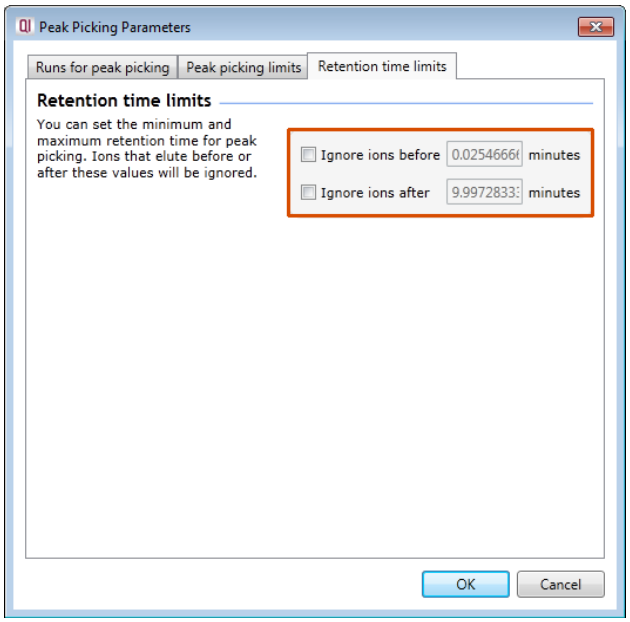
Tip: depending on run quality, a suggested minimum number of ticked runs should include at least one replicate of each experimental condition.

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.

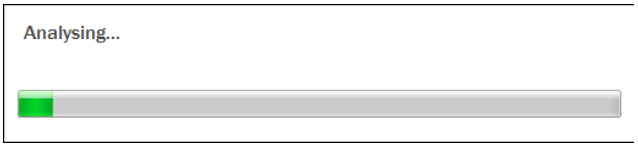
For the runs in this user guide we will use the default settings for the Automatic method and NOT apply a minimum peak width (as shown above).

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



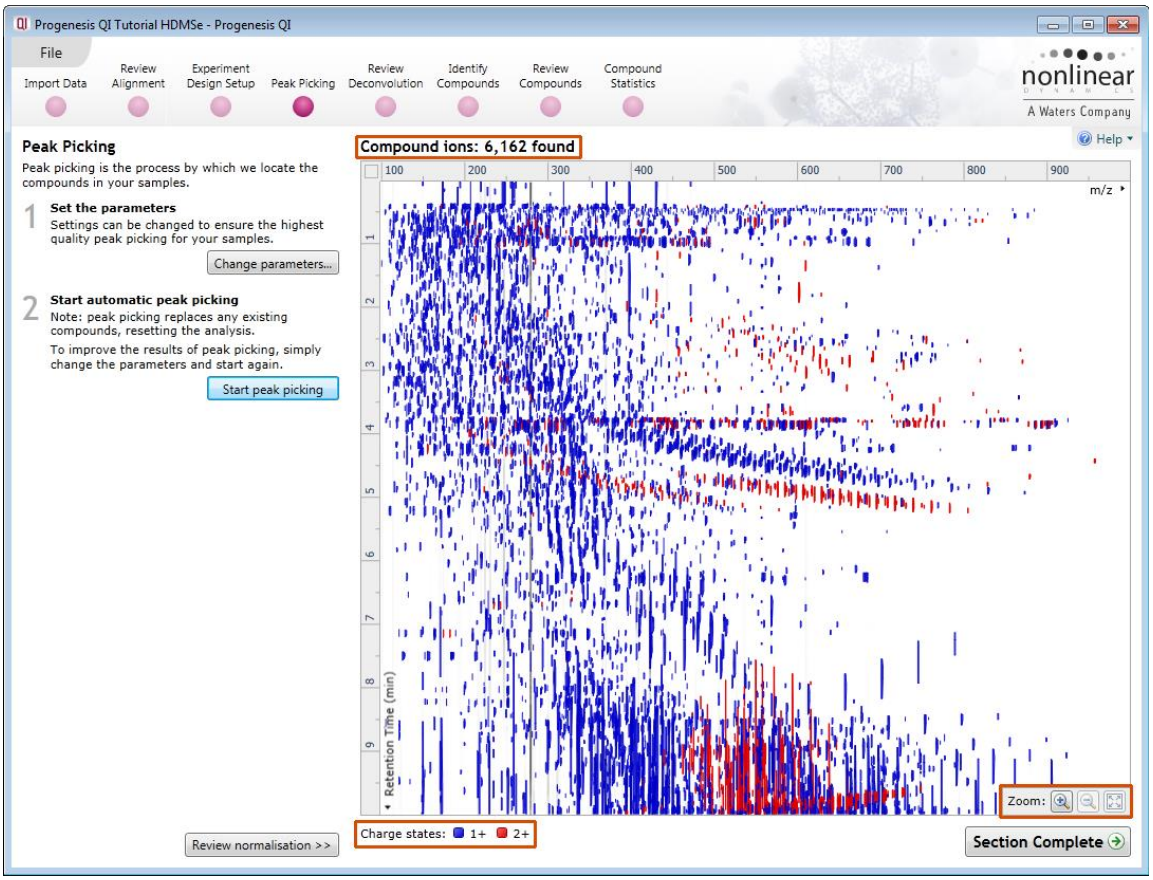
Click **OK** to close the Peak Picking Parameters dialog.

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



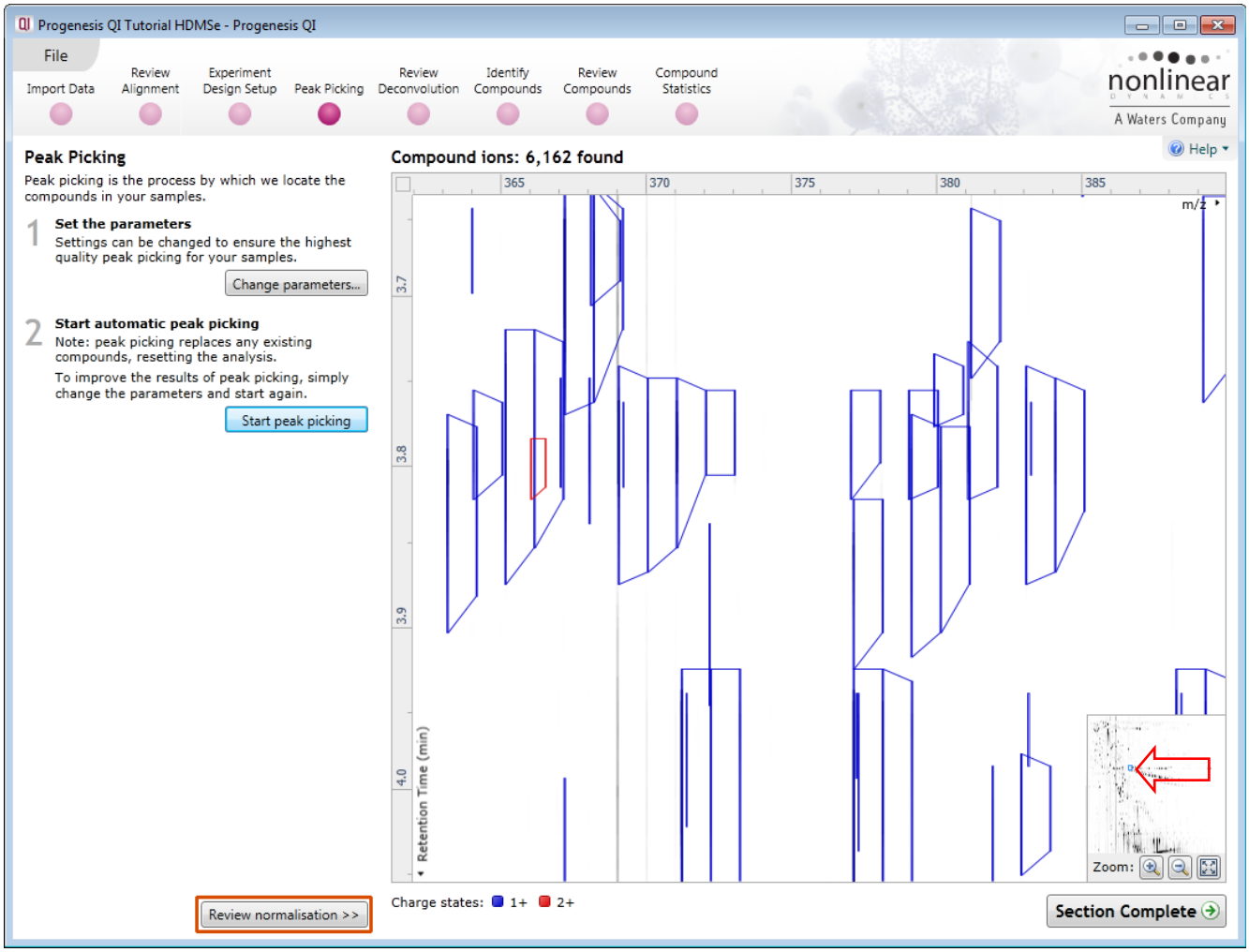
During the few minutes that the automatic analysis requires, a progress bar will appear telling you that it is Analysing.

Following automatic peak picking the detected peaks are displayed in a colour according to their charge state.



The actual number of peaks detected is recorded at the top of **the ion intensity map**.

To inspect the detection more closely drag out an area of interest with the **zoom tool**



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

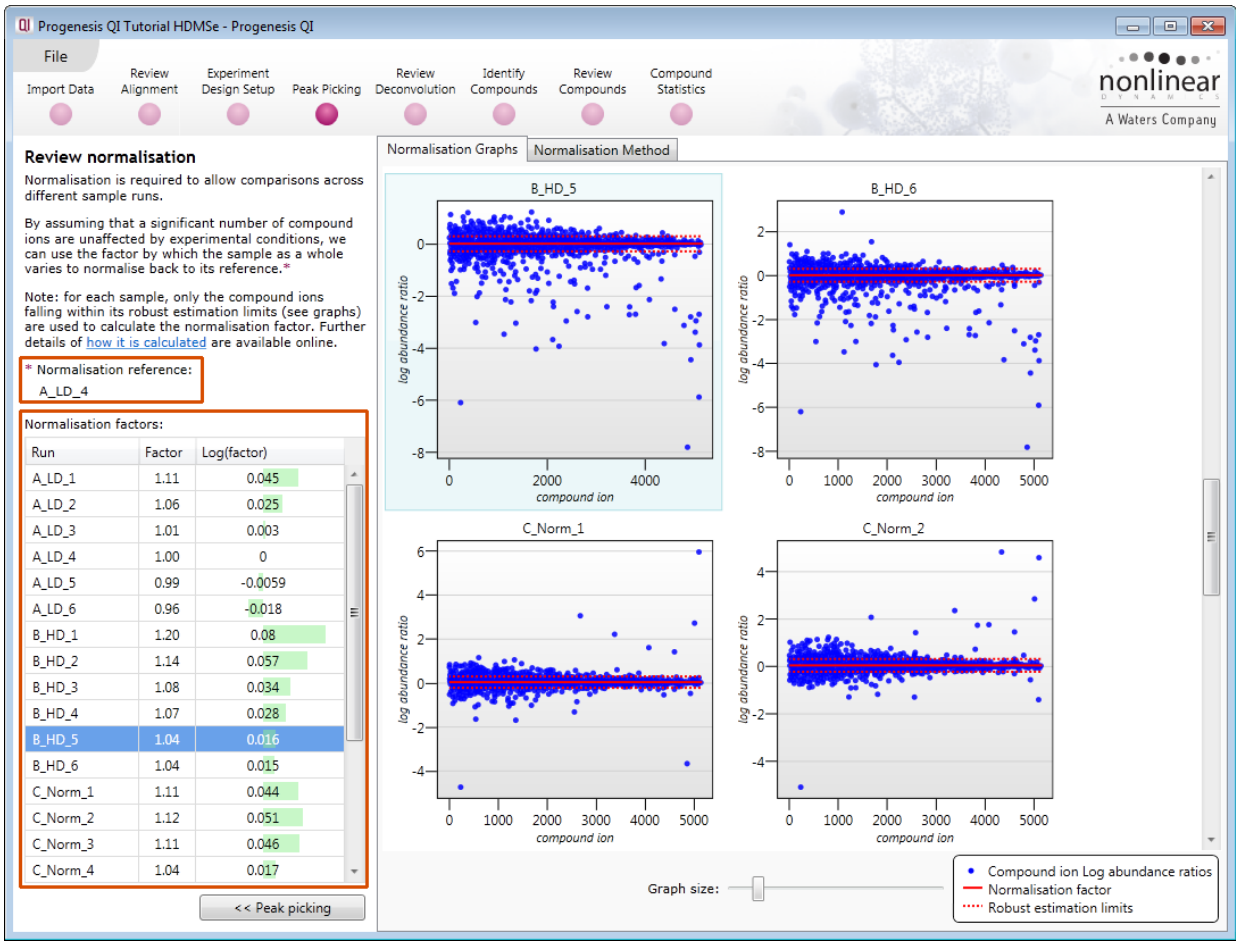
Stage 6B: Reviewing Normalisation

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.

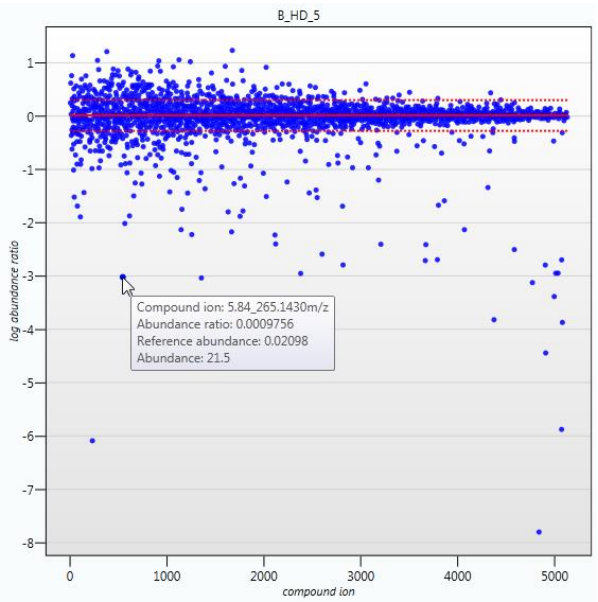


Calculation of Normalisation Factor:

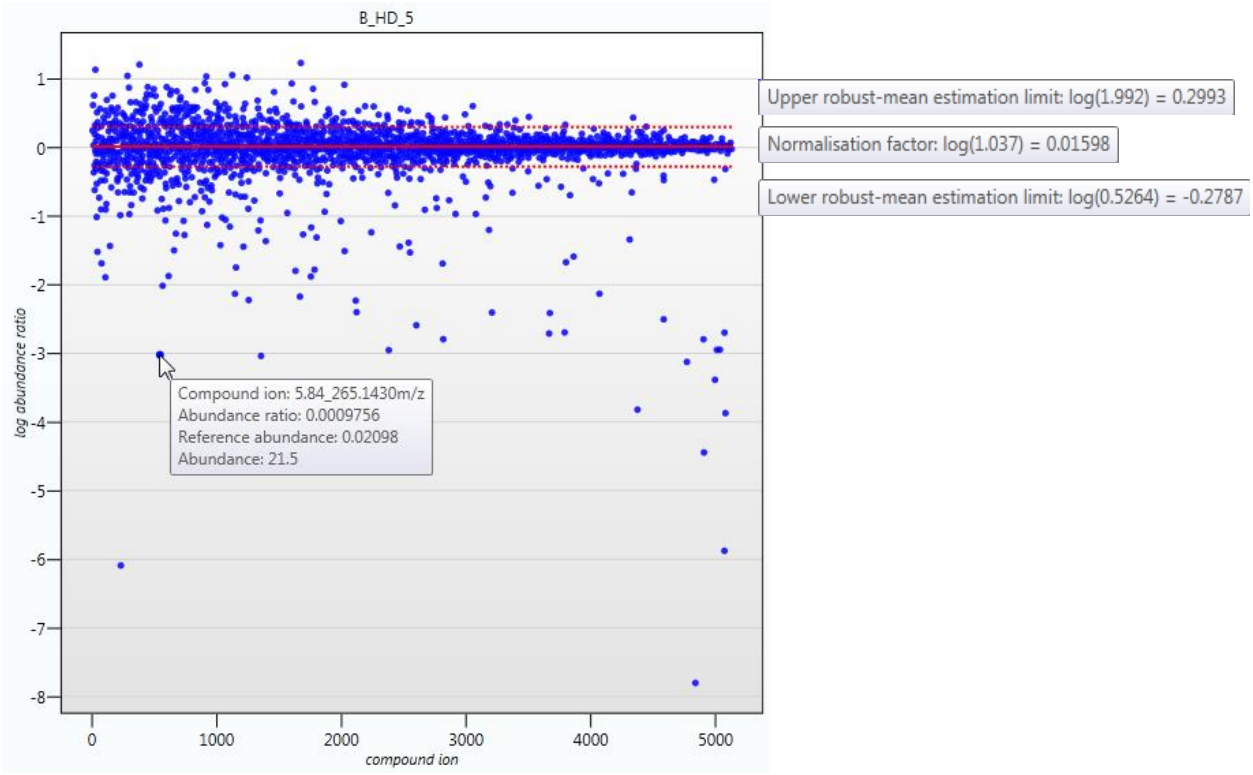
Progenesis QI will automatically select one of the runs that is 'least different' from all the other runs in the data set to be the 'Normalising reference'. The run used is shown above the table of Normalisation factors. (in this example it is A_LD_4)

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 features can be viewed as you hold the cursor over the dots on the plot.

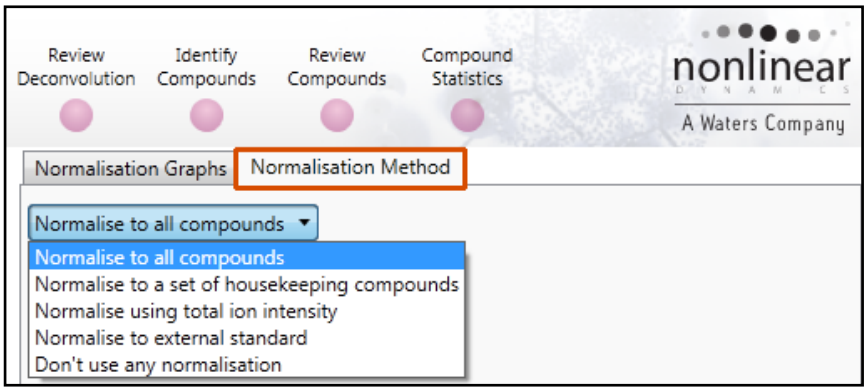


On the graph the compounds are shown ordered by ascending mean abundance. The normalisation factor is then calculated by finding the mean of the log abundance ratios of the compounds that fall within the 'robust estimated limits' (dotted red lines). Compounds outside these limits are considered to be 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



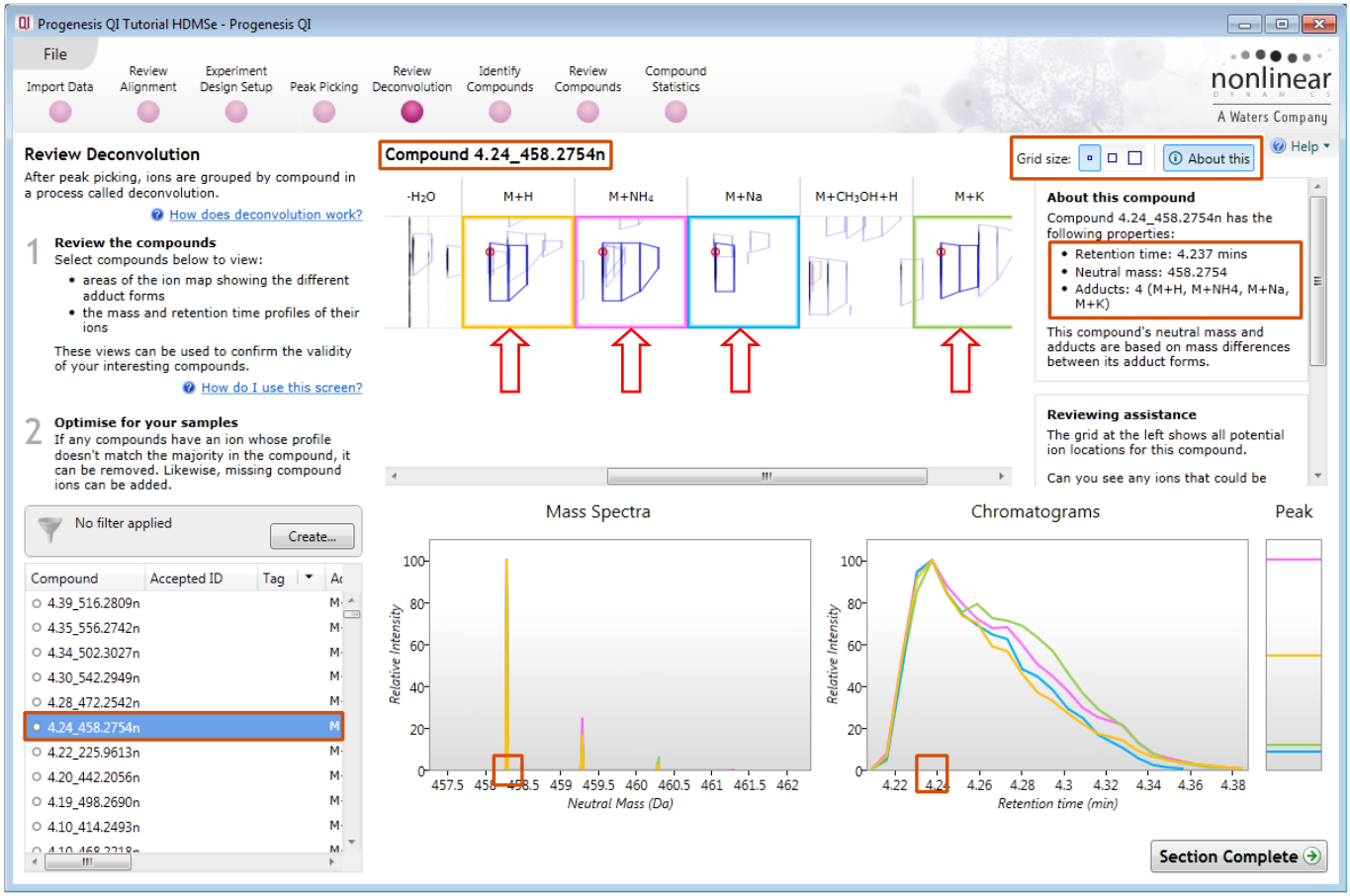
Note: once you have identified a set of 'housekeeping' compounds you can then apply the **Normalise to a set of housekeeping compounds** by using this option to locate and select the compounds.

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.



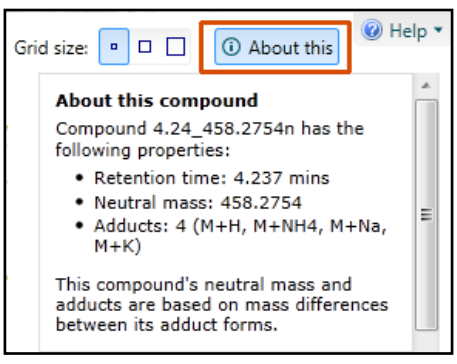
At this stage in the workflow you get the opportunity to review the process of Deconvolution for the Compounds. The table on the left displays the compounds ordered by the number of Adducts detected for each compound.

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 this example: **Compound 4.24_458.2754n**, 4 adducts have been detected: M+H, M+NH_4, M+Na and M+K (arrows) and displayed in the table.

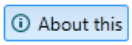
Tip: the '**About this**' panel, top right summarises the information for the current compound. If not displayed click on the **About this**.



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

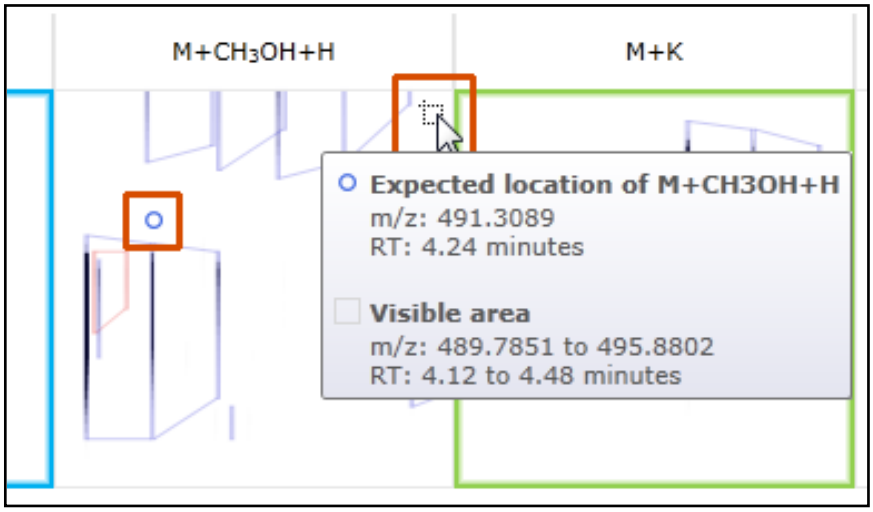
Use the Grid size to enlarge the montage view of the adducts.

Using **Compound 4.24_458.2754n** 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 

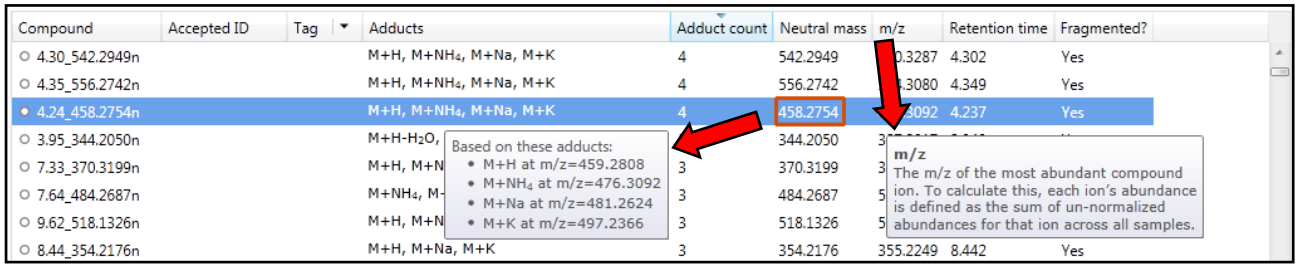
(A) Montage view showing location of detected adducts for compound

(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.

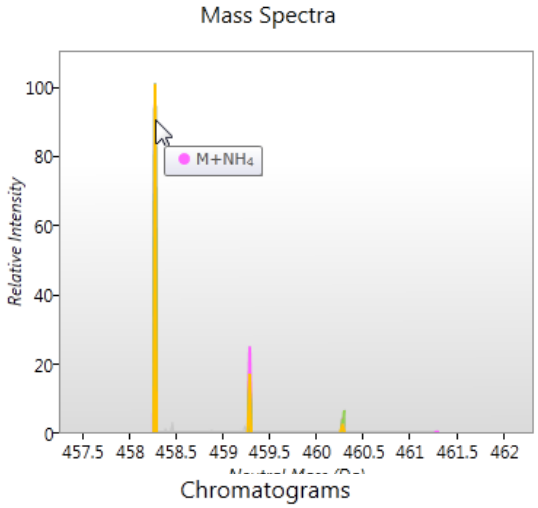


Note: hover the cursor over the icon on the top right of the panel and you will get information about the expected location and the area displayed.

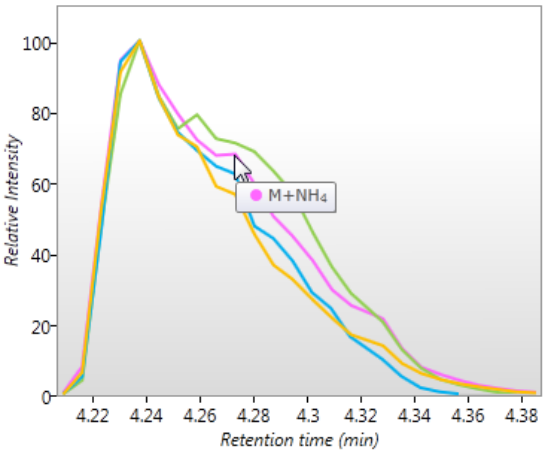
(C) The table displays the information known about the current compound with regards to the number of adducts, identity of the adducts and where deconvolution has been successful the Compound's Neutral mass is displayed. Note: when you hover the cursor over the **Neutral mass** for the current compound 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.



(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.



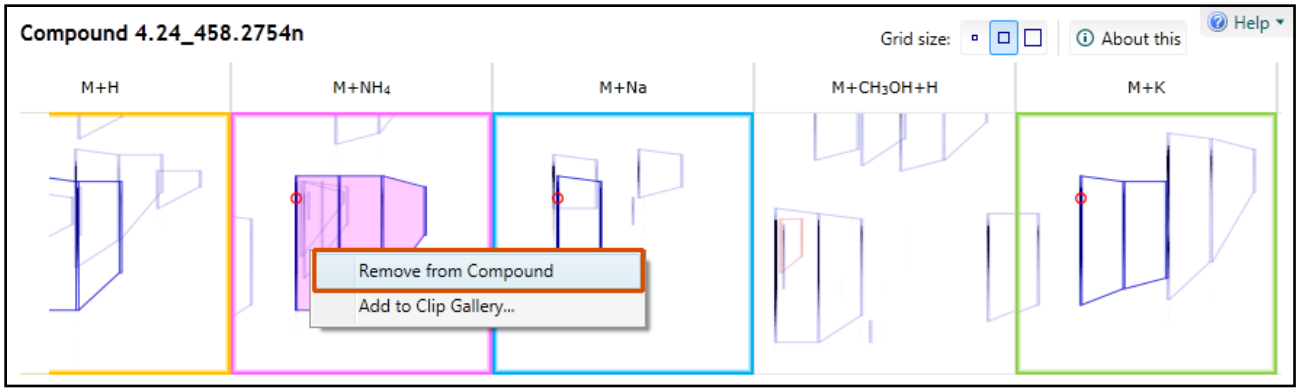
(F) This panel shows the actual Peak heights for the detected adducts of the Compound, colour coded by adduct.

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 dependant 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_458.2754n**, 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 indicating that the adducts of this compound have been edited is created.

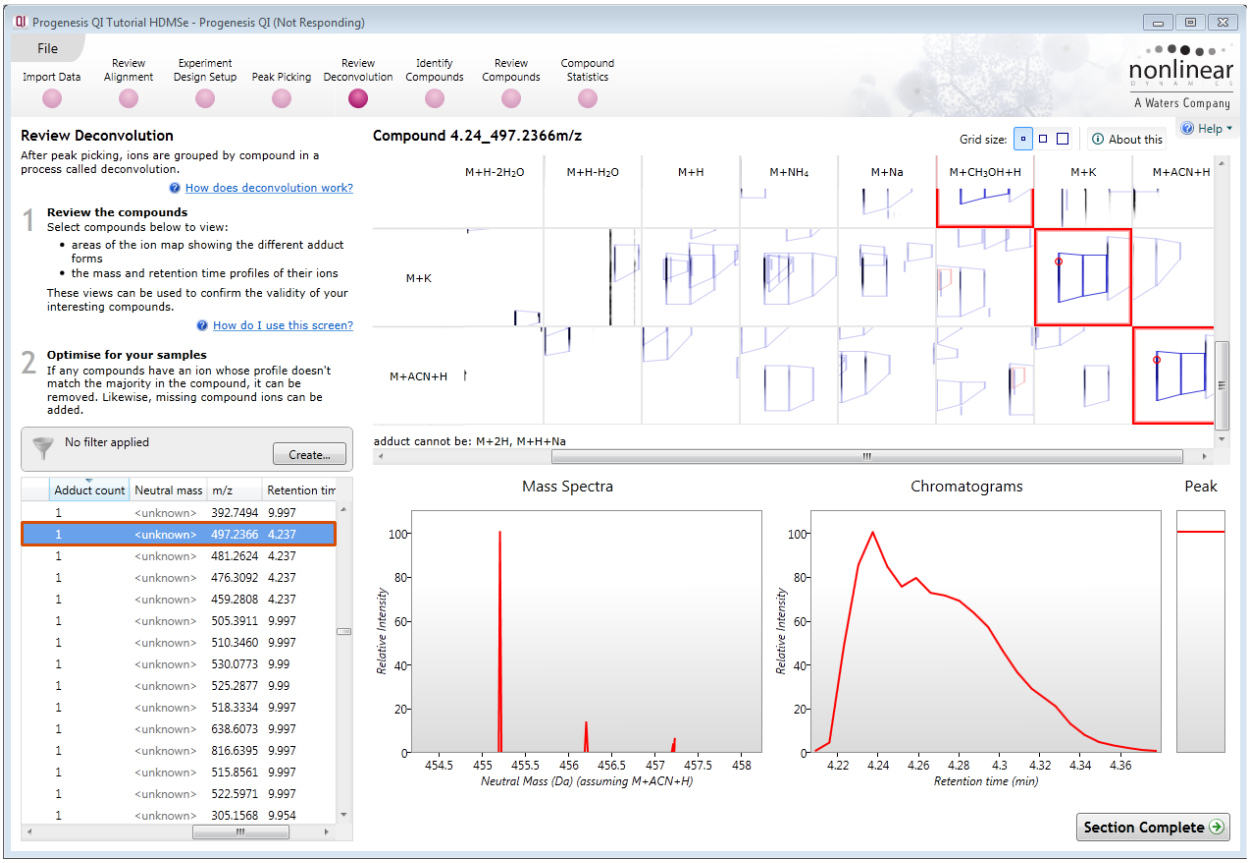
To add an adduct to a compound right click on a peak in the appropriate panel for the adduct and click **Add to compound,** again the table of compounds will update to reflect the change.

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.

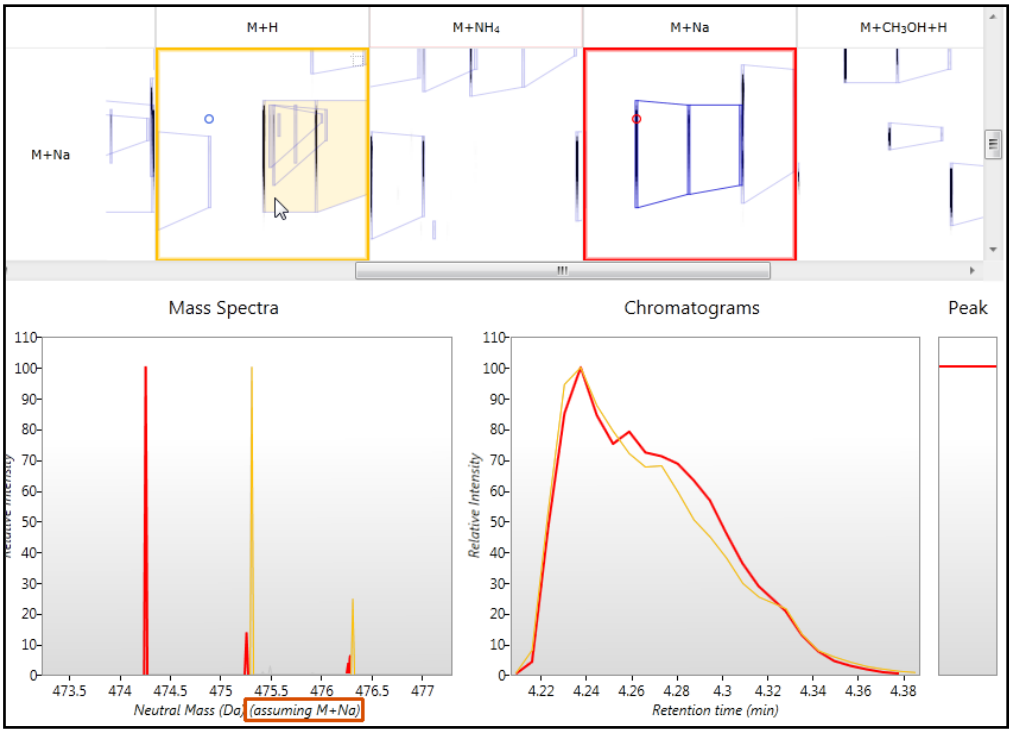
Try removing all the adducts from **Compound 4.24_458.2754n** to leave just the M+K.

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.

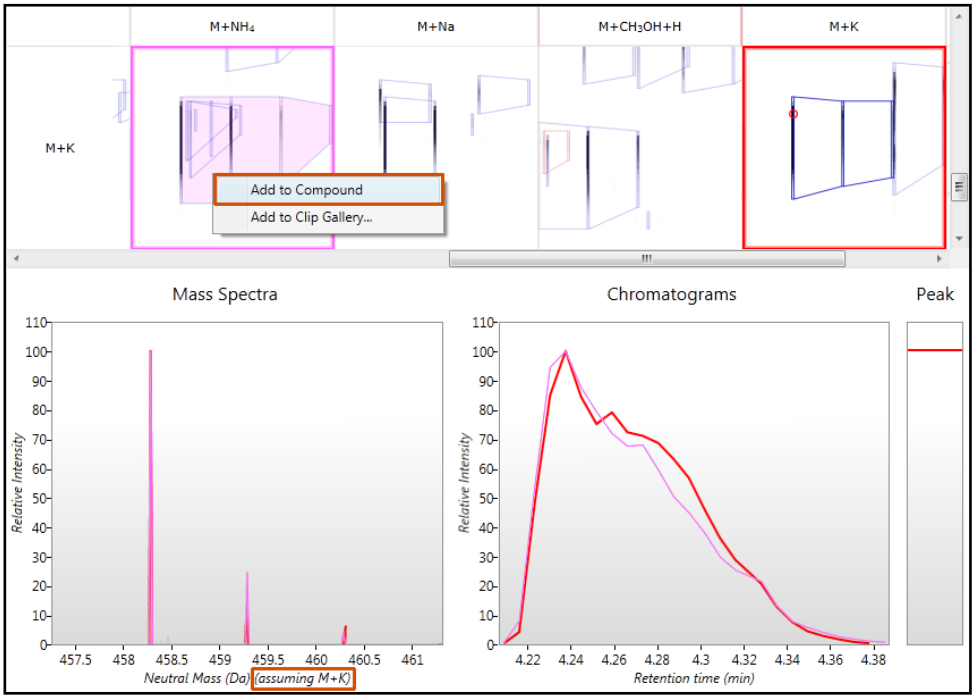
Note: because the peak is a charge state +1 this excludes H+2H and M+H+Na as being the compound's sole adduct on the basis of charge.



As an example to explore the posibility that **Compound 4.24_458.2754n'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_458.2754n**. This time when you hover over the cursor over the M+NH_4 adduct both the m/z on the mass spectra and RTon 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 by right clicking on it. As you add the second adduct the table will update to indicate that the compound has 2 adducts, a **Neutral mass** is calculated and the matrix is replaced by a single row of adducts which can be explored for the presence of additional adducts in this case an additional 2 adducts can also be added.

Exploring the expected location for the second adduct of a compound

After Peak detection 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.

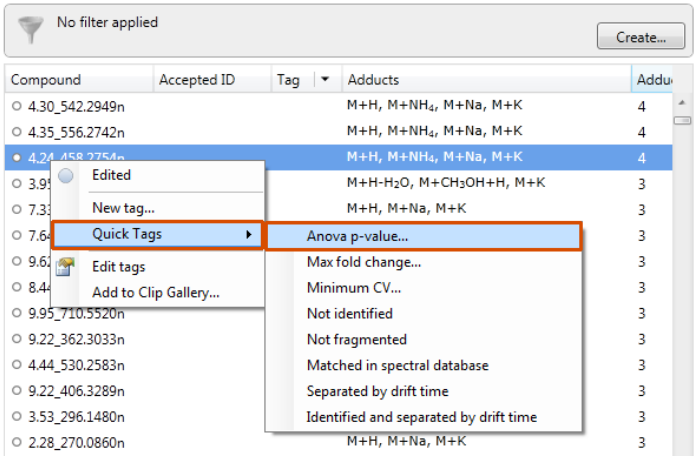
By adding the second adduct the montage and table will update to reflect the addition of the adduct to the compound and an Neutral mass will be assigned based on the 2 adducts. The matrix collapses to the single row displaying the locations of the two detected adducts.

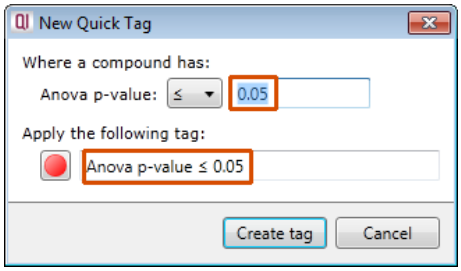
Using 'Tagging' to filter data for identification

In order to identify the deconvoluted 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.

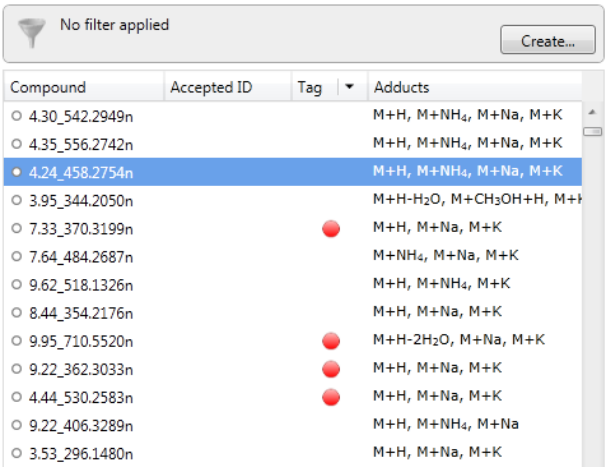
As an example of the use of tags we will set out to identify the compounds that are associated with a specific population of detected features. This population of features are required to show a significant (Anova p value < 0.05), 2 fold or greater increase in abundance for one of the conditions (A_LD, 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…**



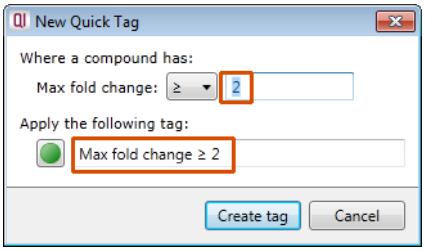


Accept the default value (≤ 0.05) and the offered name.



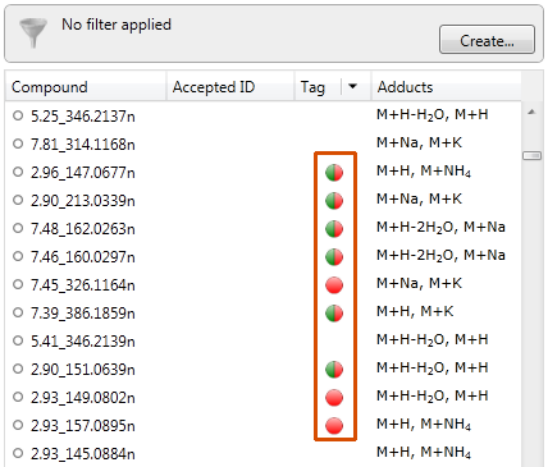
On pressing **OK** a tag appears in the table against all the compounds with an Anova p-value ≤ 0.05

To add a second **'Quick Tag'** for those features 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).



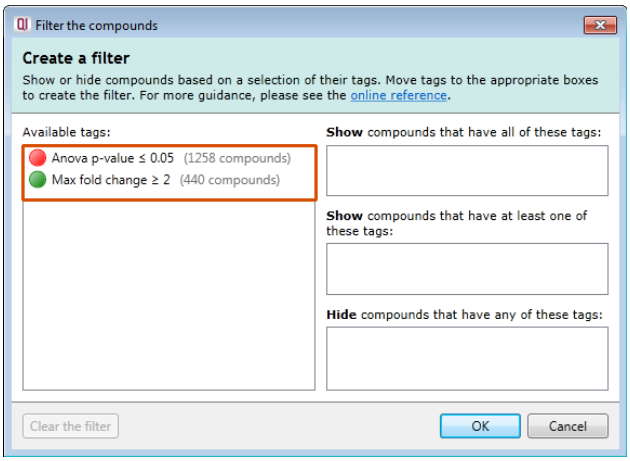
You will be offered a tag named **Max fold change ≥ 2**.

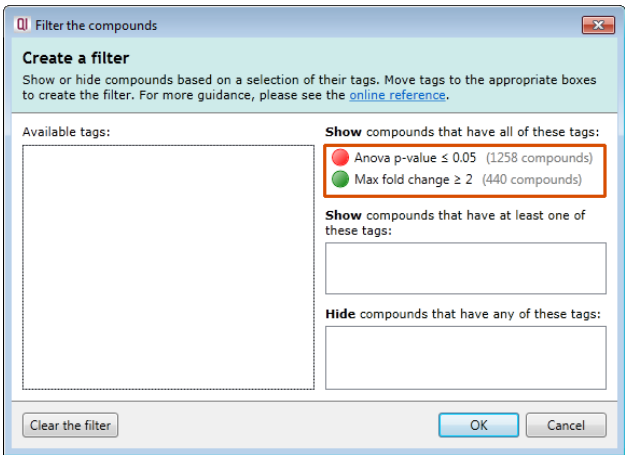
On pressing **Create tag** a second tag appears in the table against all the features with a Max fold change ≥ 2.



Now to get the table to display **only** those features which satisfy the required criteria, i.e. those features 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** features with all of these tags.

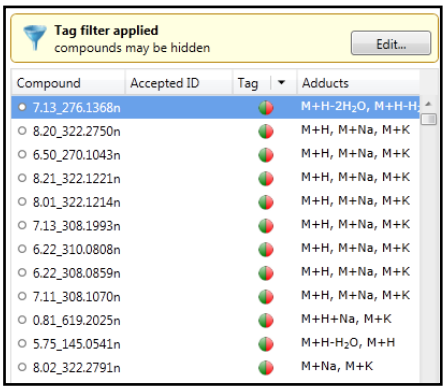




On clicking **OK** the table now displays only those features with the 2 tags.

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.



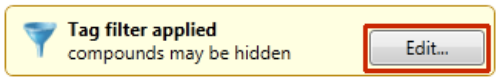
Stage 8: Identify Compounds

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

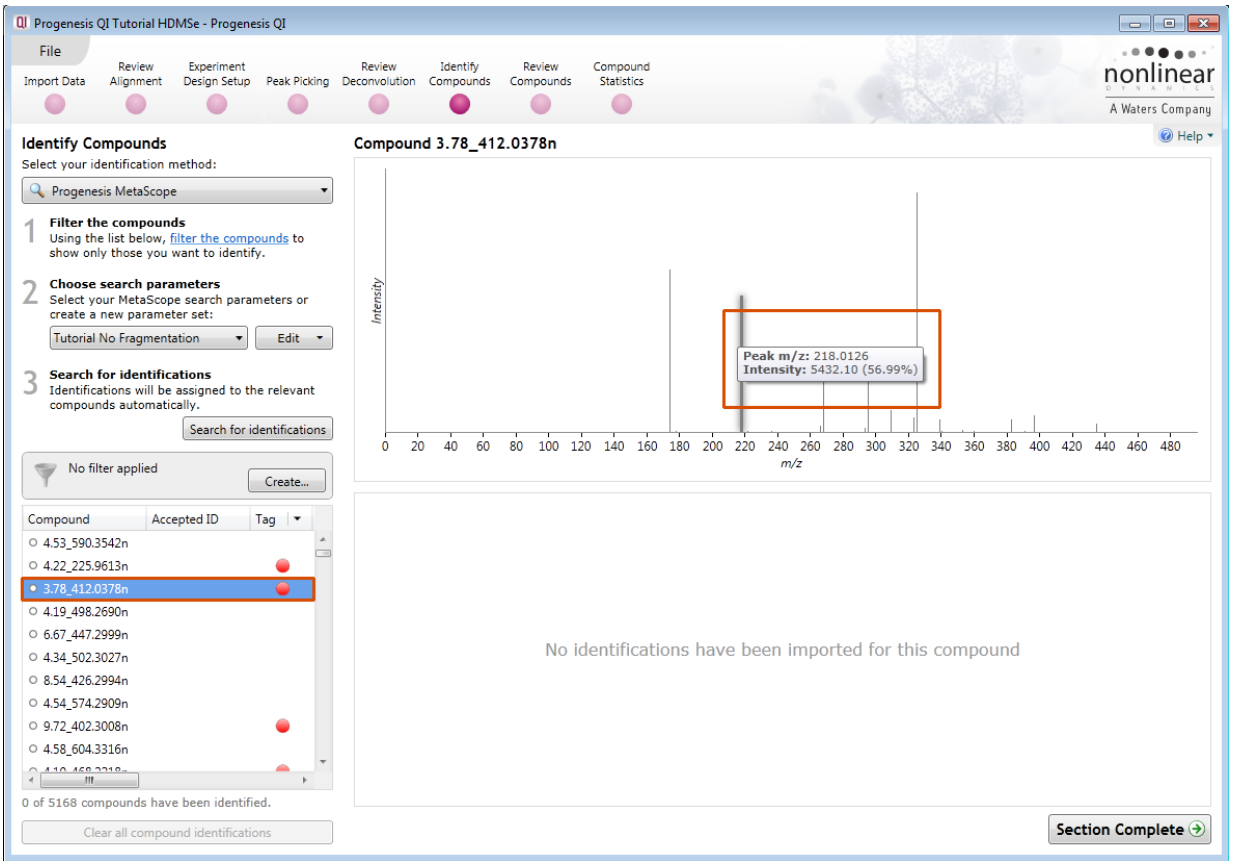
Note: that the filter you applied at the previous stage is currently active.

This means that when you export the feature data to a spread sheet, to use when searching compound identifications, only the features currently displayed in the table will be exported. In this case, only the features that display a significant 2 fold or greater increase in Mean abundance will be exported.

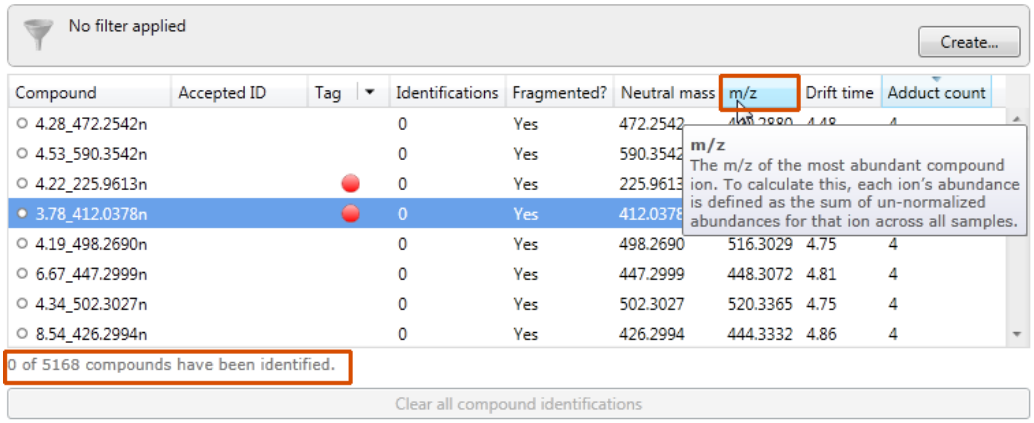
To display all the features in the table click **Edit** and then click on **Clear the filter.**



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



which corresponds to **Compound 3.78_412.0378n** which has a neutral mass of 412.0378 as indicated in the table.



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

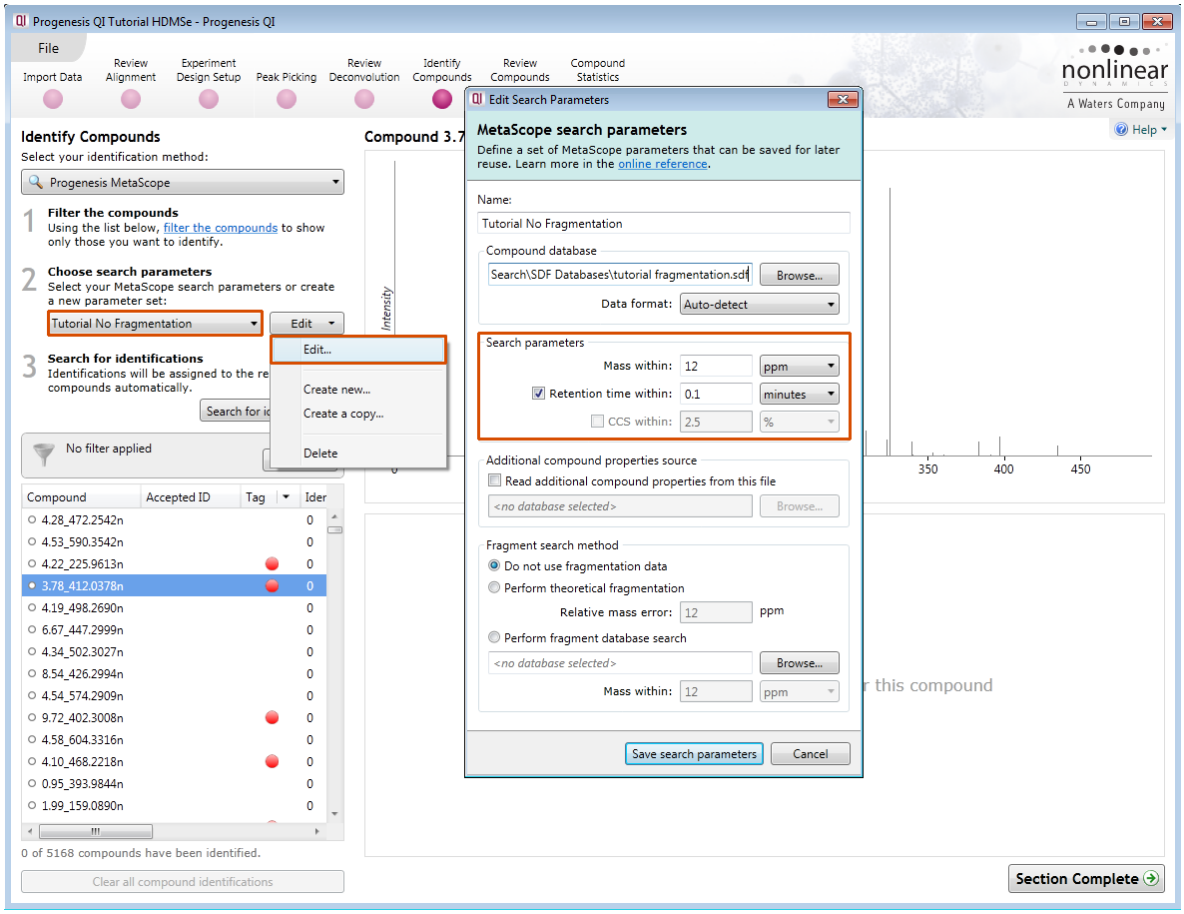
Currently there are 5168 compounds of which none have been identified.

Identification of compounds is performed 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, Collisional Cross Section and Theoretical Fragmentation depending on the format and content of the database.

In this example we will use **Progenesis Metascope,** setting the Mass Threshold to be within 12ppm in this case you can use the **Tutorial No Frgmentation** search parameters as shown above

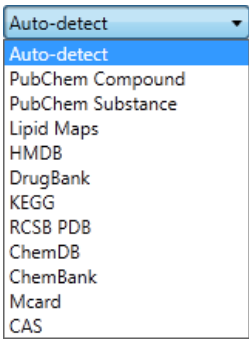
Browse to locate the database file.

Location of example SDF files after install:

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

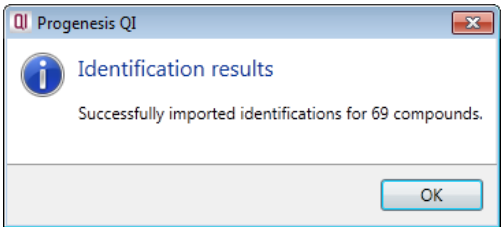
Select: Tutorial Fragmentation.sdf

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

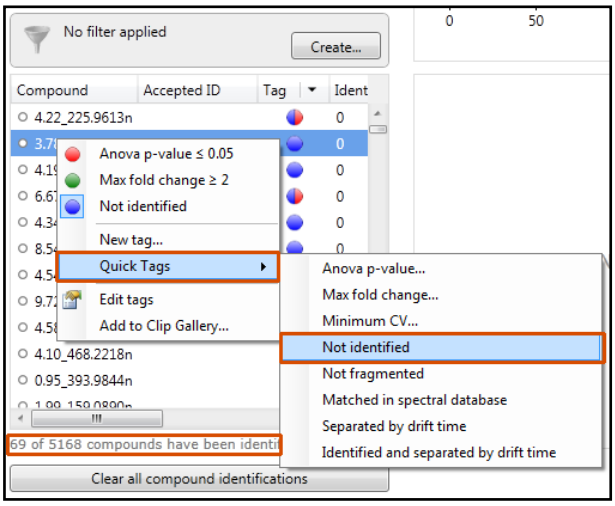


Save the Search Parameters and click **Search for identifications.**

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

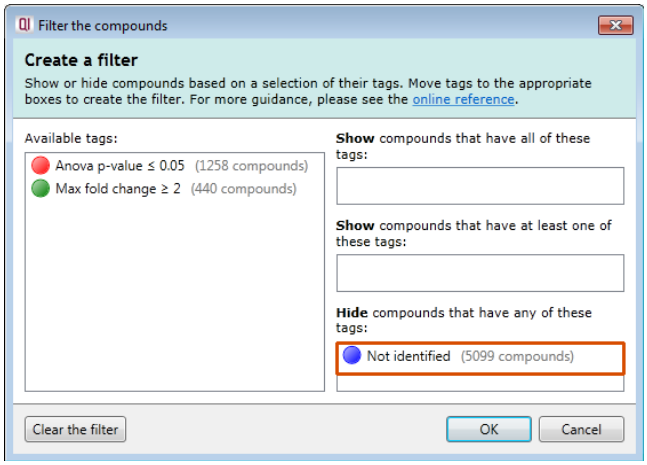


To focus only on the identified compounds, hide all the unidentified compounds. To do this you must first tag all the unidentified compounds. Right click on the table and select the Quick Tag for **Not Identified.**



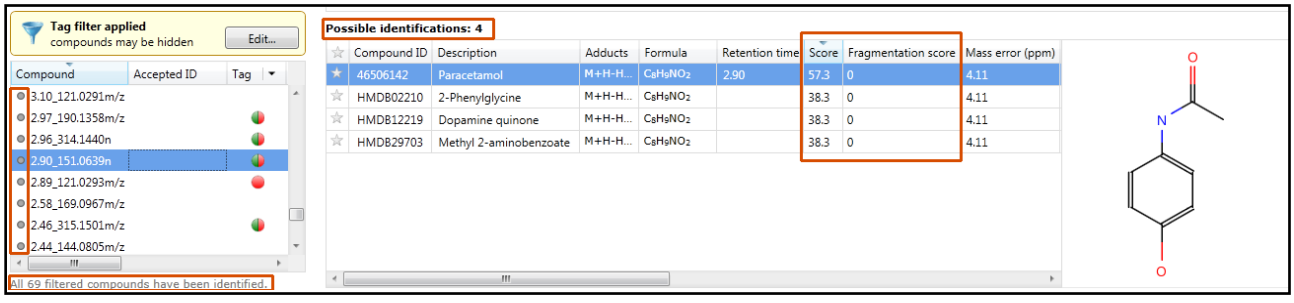
Then create a filter that hides these compounds by selecting Create and dragging the **Not Identified** tag on to the **Hide** panel.

Click **OK** to apply the filter.

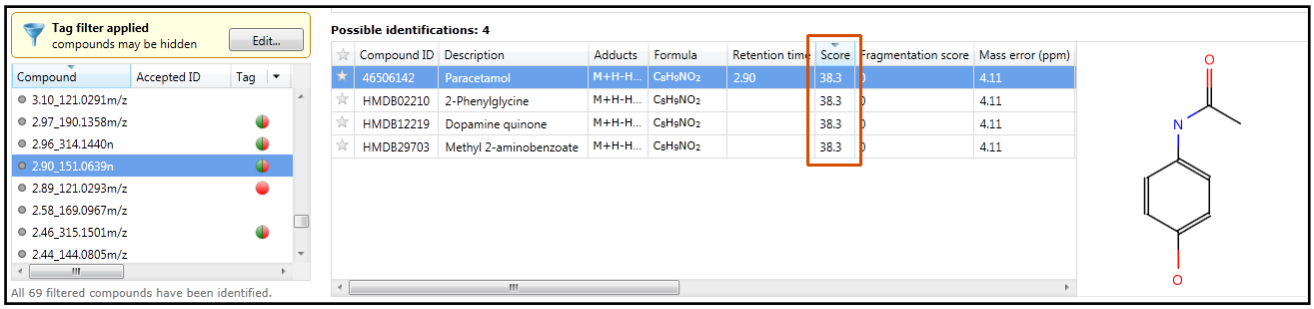


Having imported the search results you can see there are a number of possible identifications for the identified compounds.

The solid grey icon to the left of the Compound column indicates an identification has been made for this compound. With the available Retention time data for Paracetomol this would indicate that Paracetamol is a stronger match.

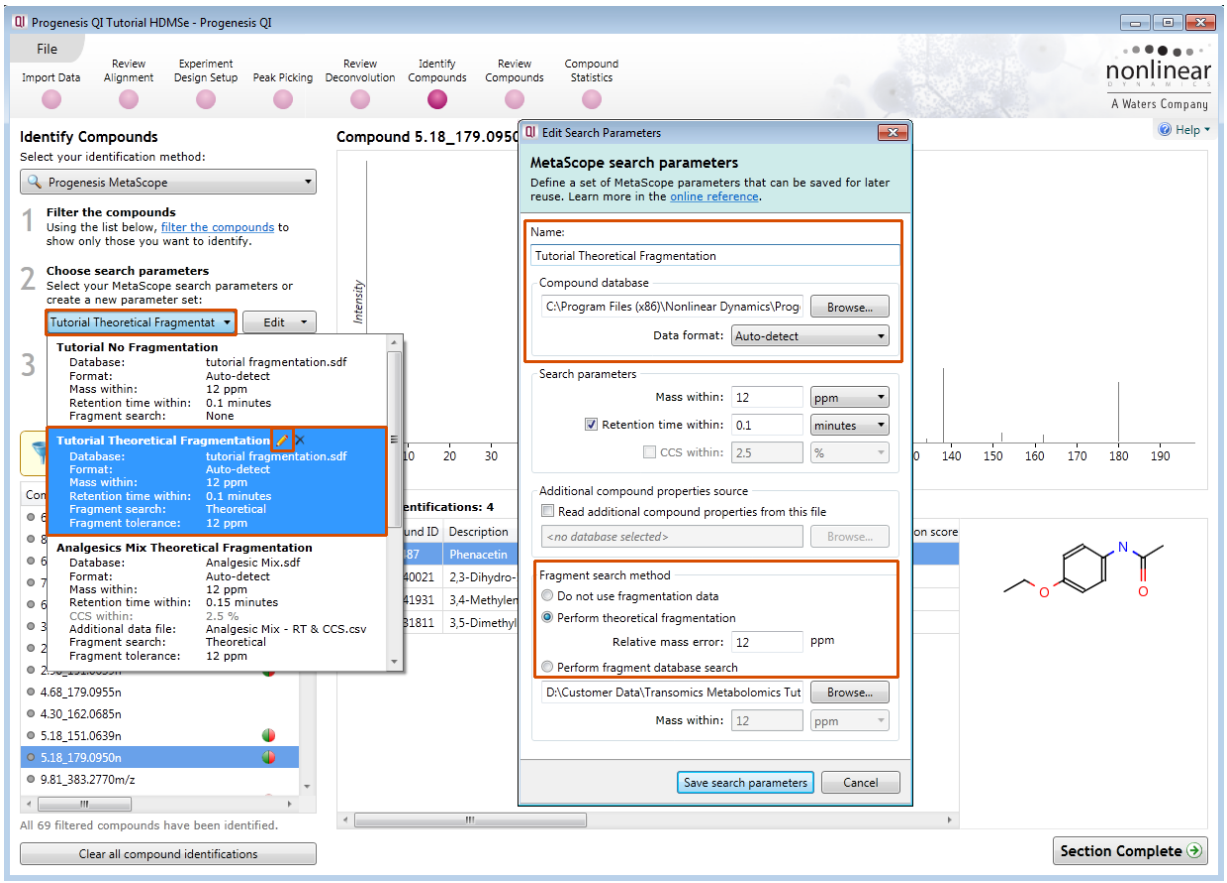


Note: when the Retention time data is **not** included as part of the search scoring then there is no difference in the scoring for the 4 identification candidates in the table to the right.



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.

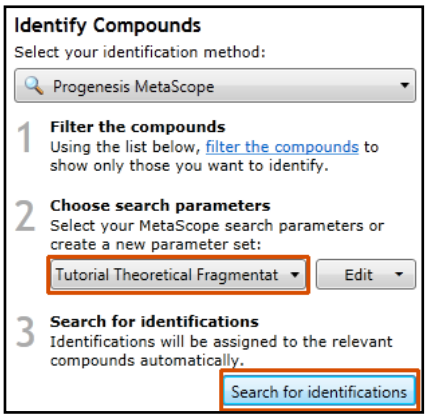


To perform theoretical fragmentation using the **tutorial fragmentation .sdf** as the source of the candidate compounds select the second Search Parameter Set (Tutorial Theoretical Fragmentation) from the drop down

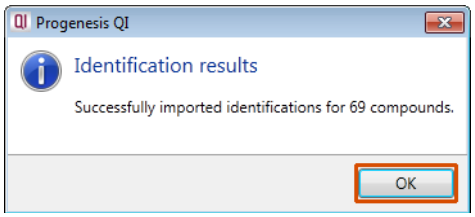
Note: you can review/edit the parameter set, either by clicking on the 'pencil' icon in the drop down or using the **Edit** option to the right of the drop down.

Ensure the **Perform theoretical fragmentation** from the Fragment search method is ticked and then save search parameters.

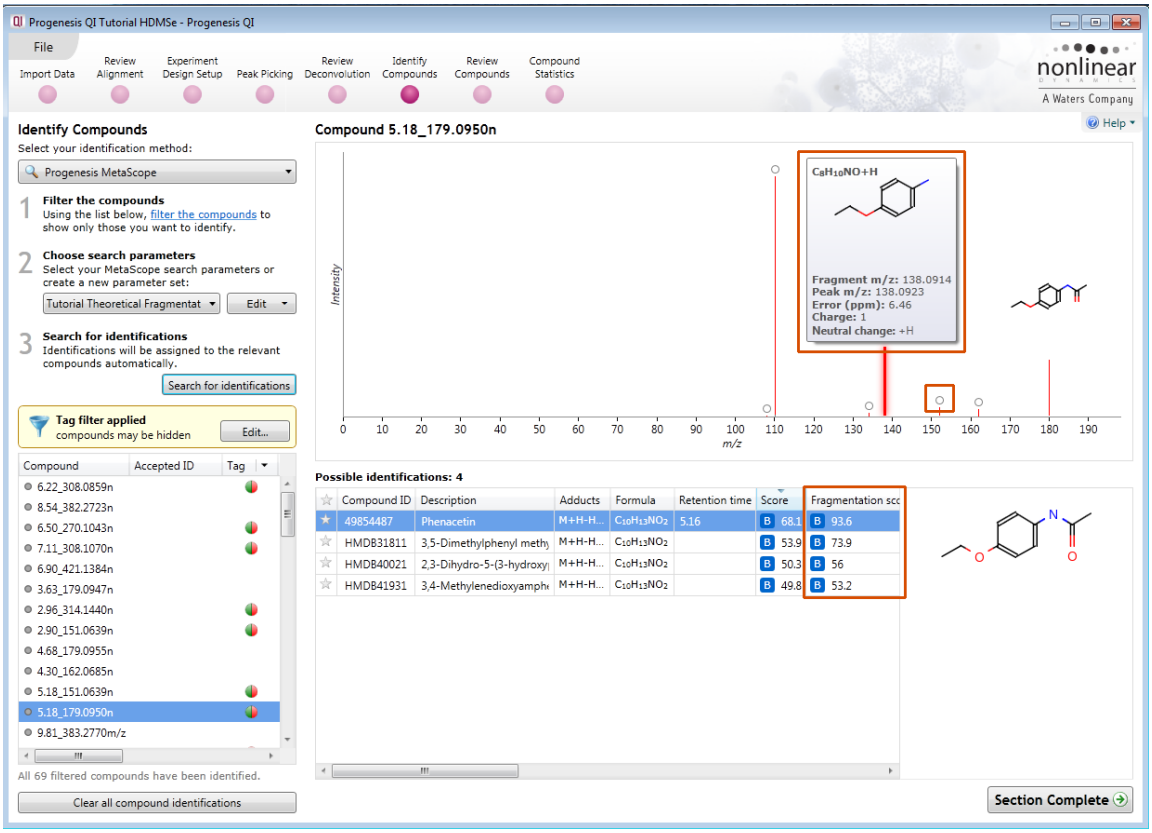
To perform the Theoretical Fragmentation search click on **Search for identifications**



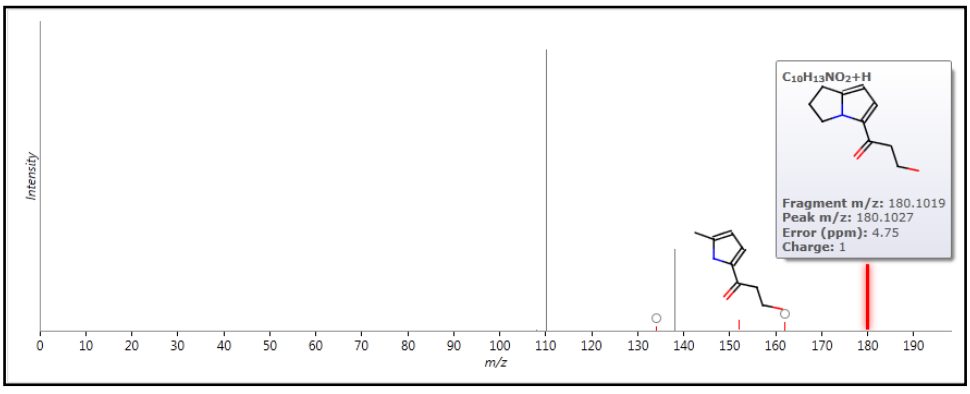
As before search results will be returned



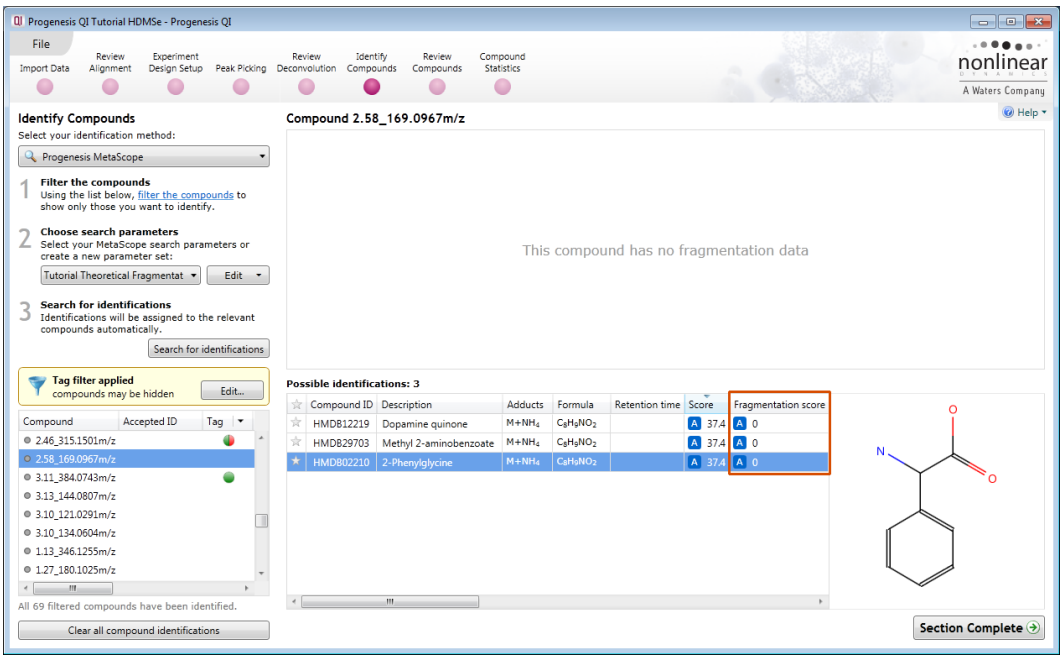
Where fragments have been measured for a compound and a match is made with the theoretical fragments from each candidate compound a solid coloured line appears on the Fragmentation graph. This is accompanied with either an image of the fragment displayed (if there is room) or an open symbol at the top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



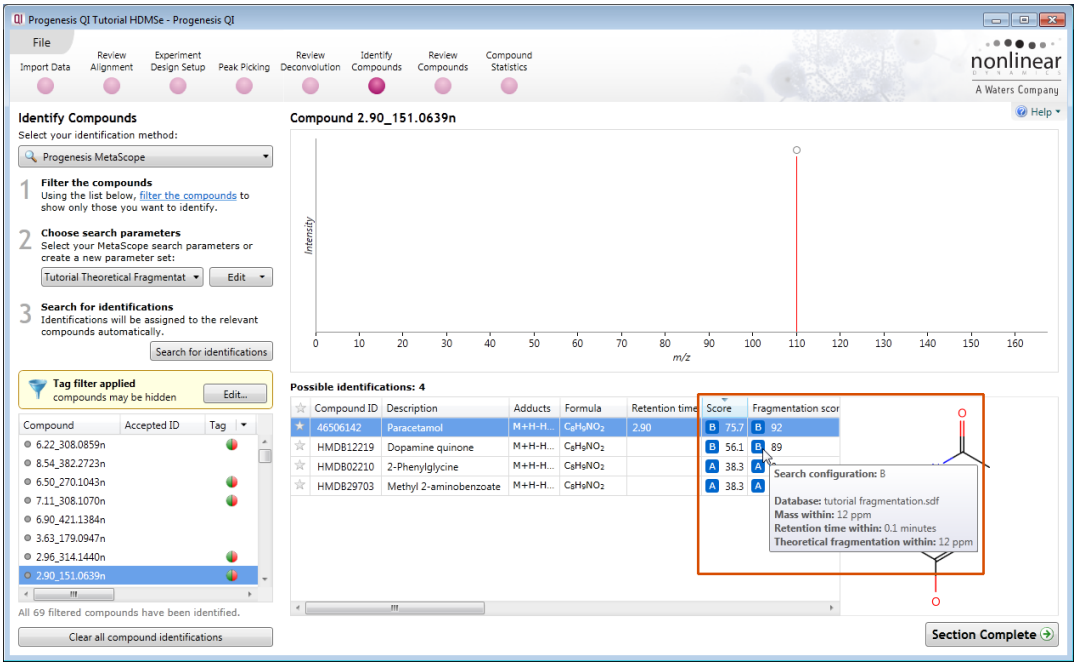
The fragmentation score for each possible identity appears in the table. When you click on one of the possible identifications with a lower fragmentation score the Fragmentation graph displays which fragments contribute to that score for example the **3rd ranked compound** Fragmentation matches (see below).



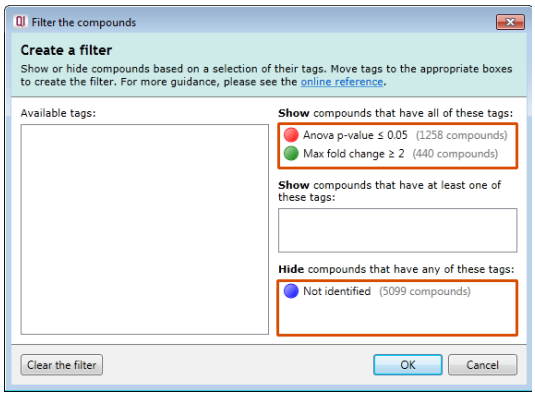
Note: a fragmentation score of 0 indicates either that **no** match was achieved between the theoretical fragments (of a candidate compound) and the measured fragments or in fact the compound has no fragmentation data.



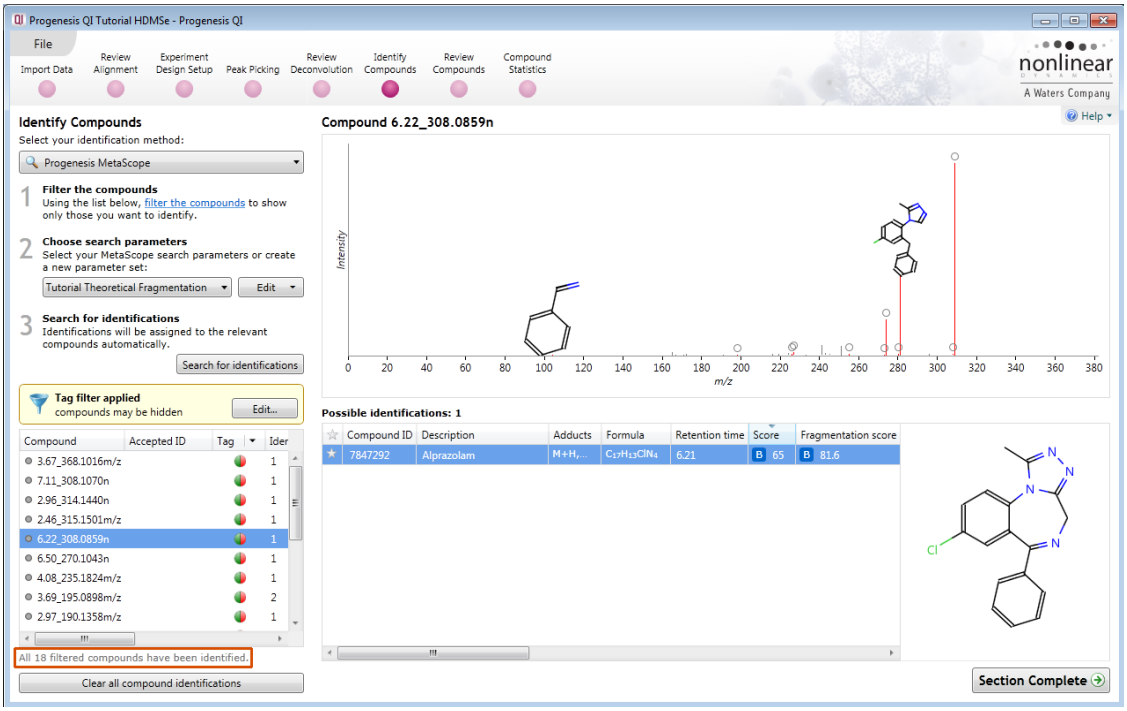
Note: where multiple searches have been performed the search yielding the highest overall score for each candidate is displayed. Also the search parameters giving rise to these scores are displayed as tooltips when the cursor is held over the Search Parameter indicator.



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

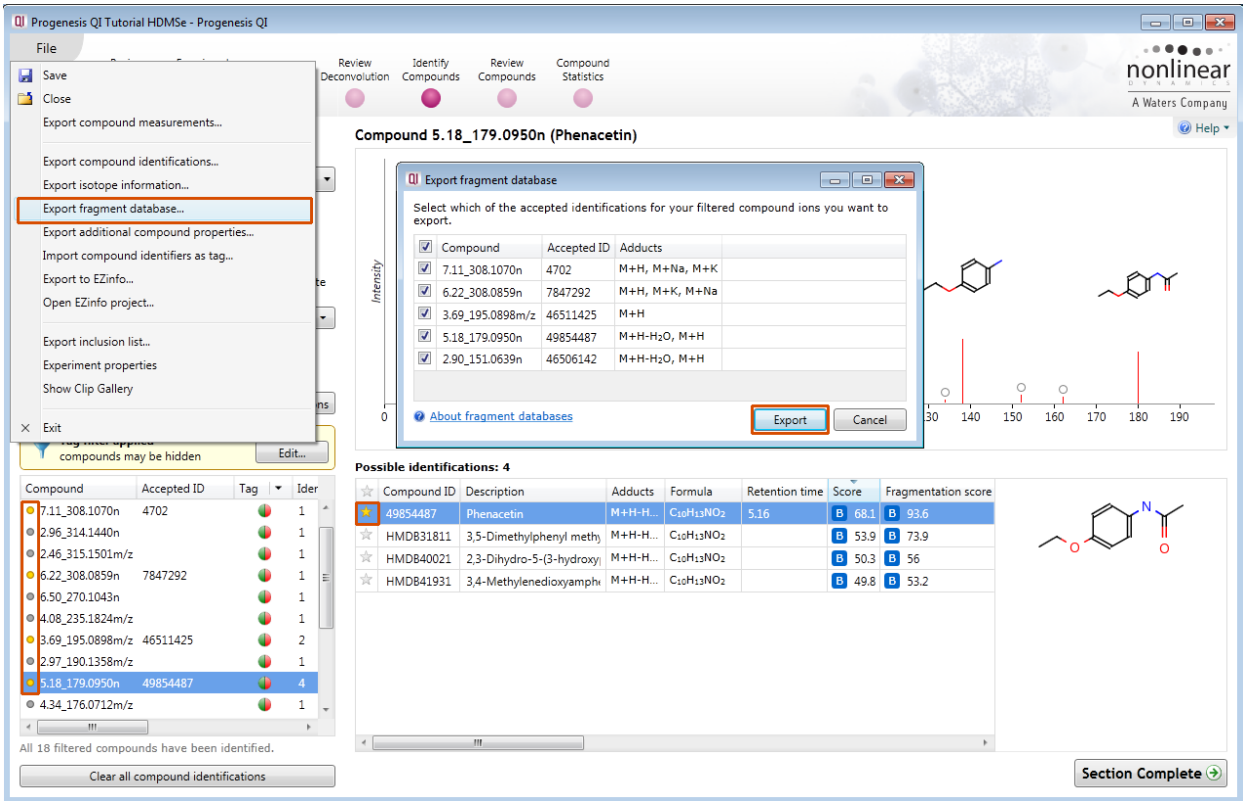


The table now displays only information for 18 identified compounds.



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.

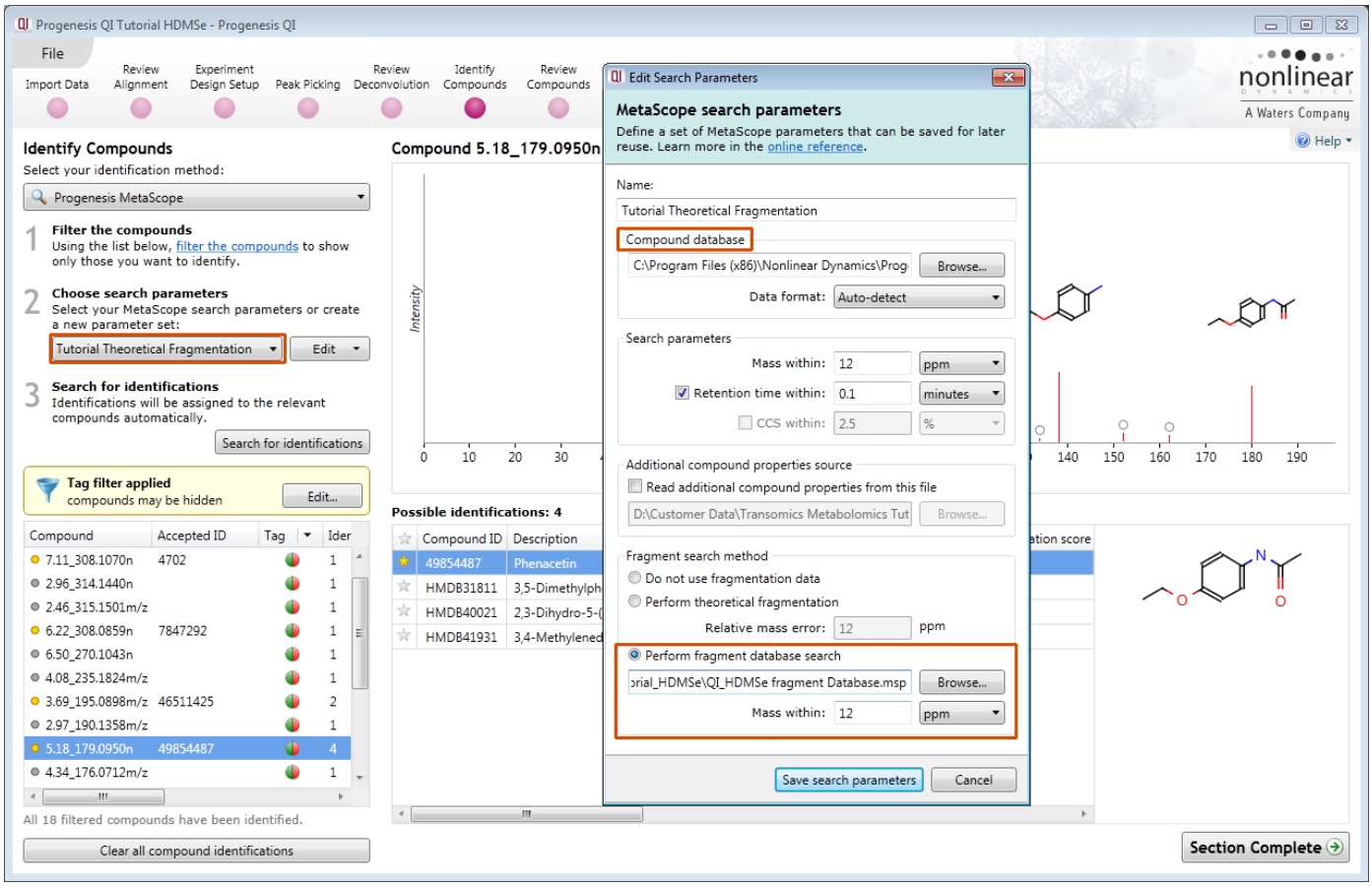


To accept the identity of a compound click on the star icon, it will turn yellow. Do this for a number of compounds displaying an acceptable fragmentation score.

Then select **Export fragment database…** from the file menu

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 HDMSe_.zip)



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

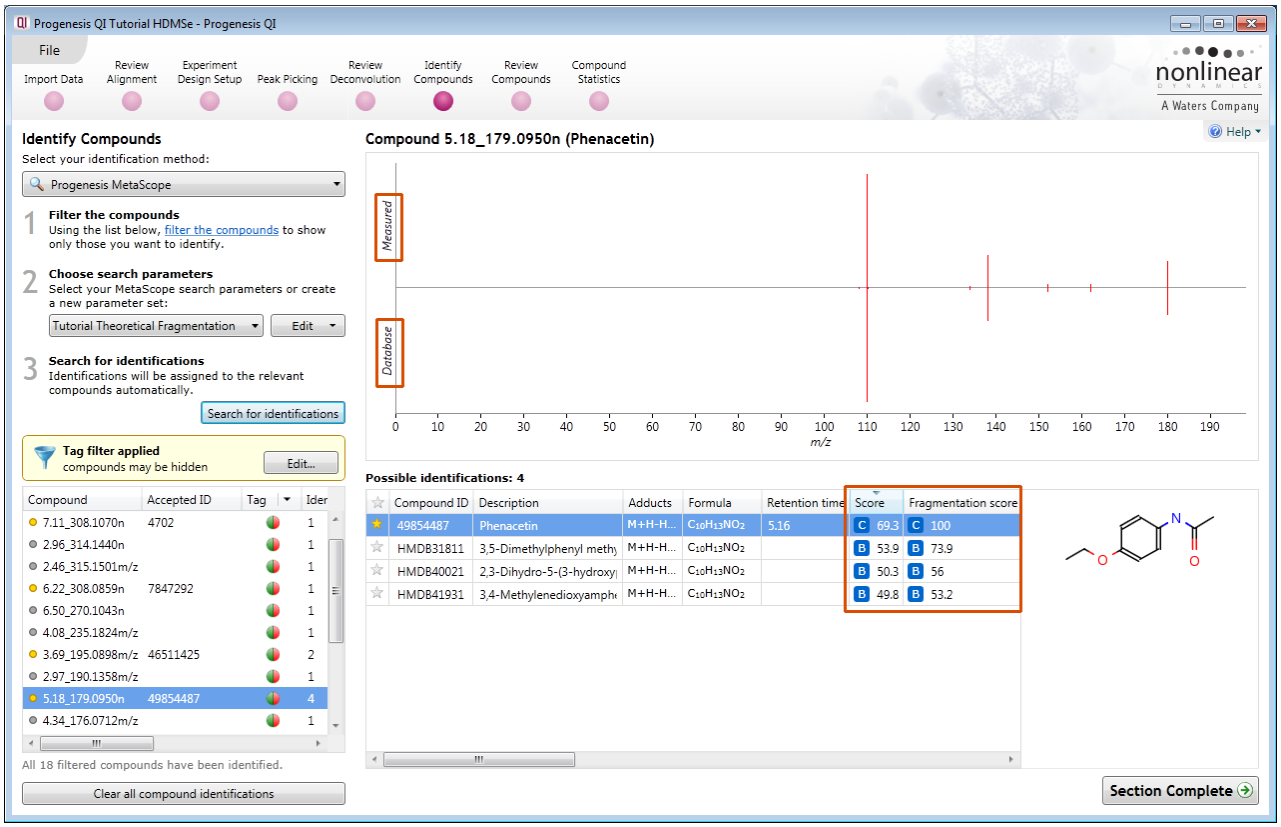
Now as an example re-perform the Tutorial Theoretical Fragmentation but this time instead of selecting the **Perform theoretical fragmentation** option select the **Perform fragment database search** option. Then locate the .msp file you have just created.

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

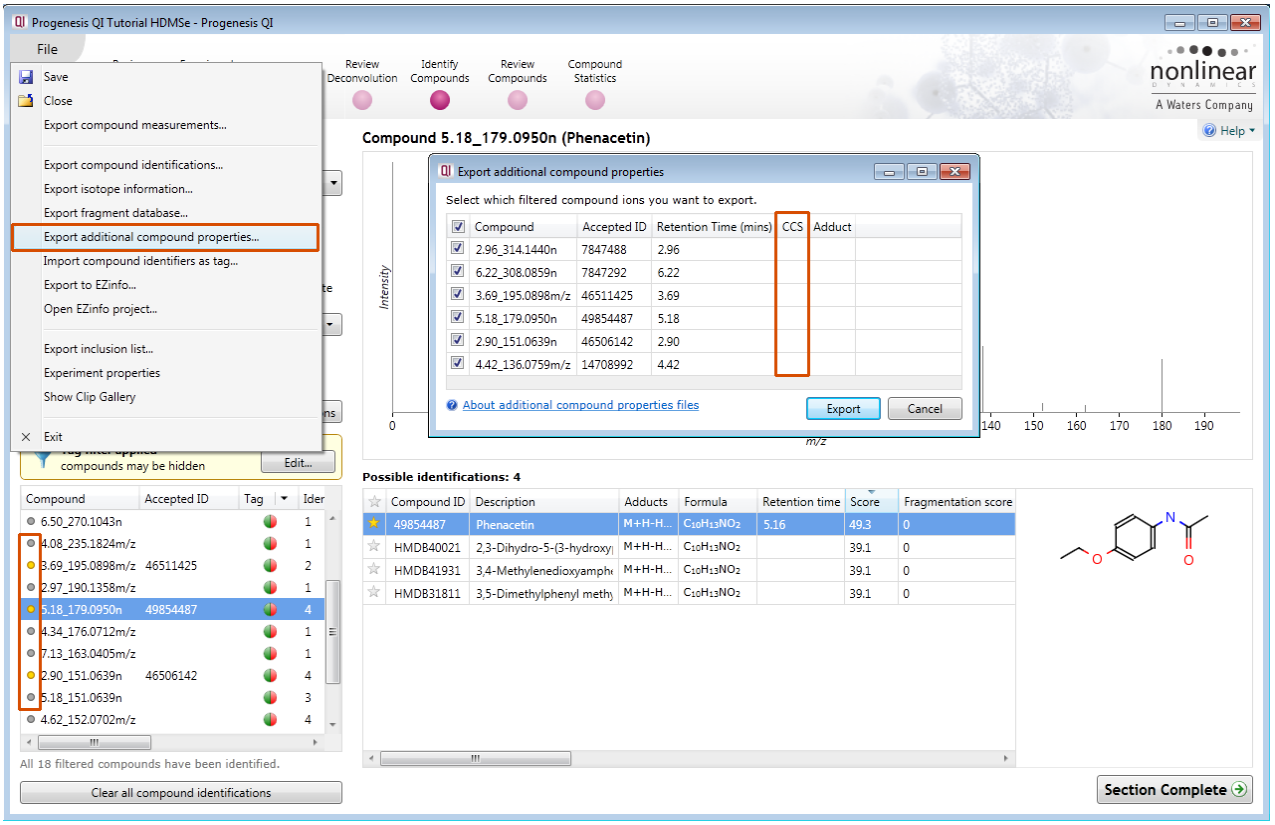
Note: you must use the same Compound database that was used when the original msp file was generated

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.

The current search with the highest score is designated C in this example.

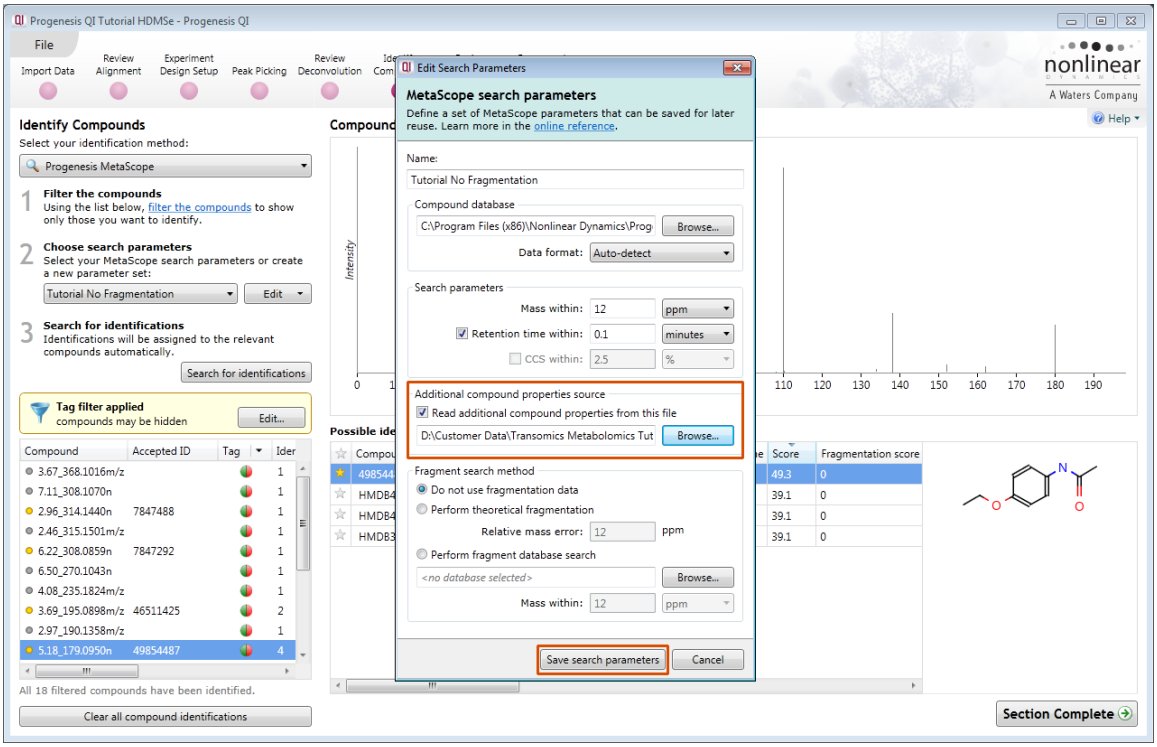


You can also manage and search against additional compound properties. The advantage of this is that you can create a database of compound data that retains information on an identified compound for m/z, retention time. Export and save this Compound properties database

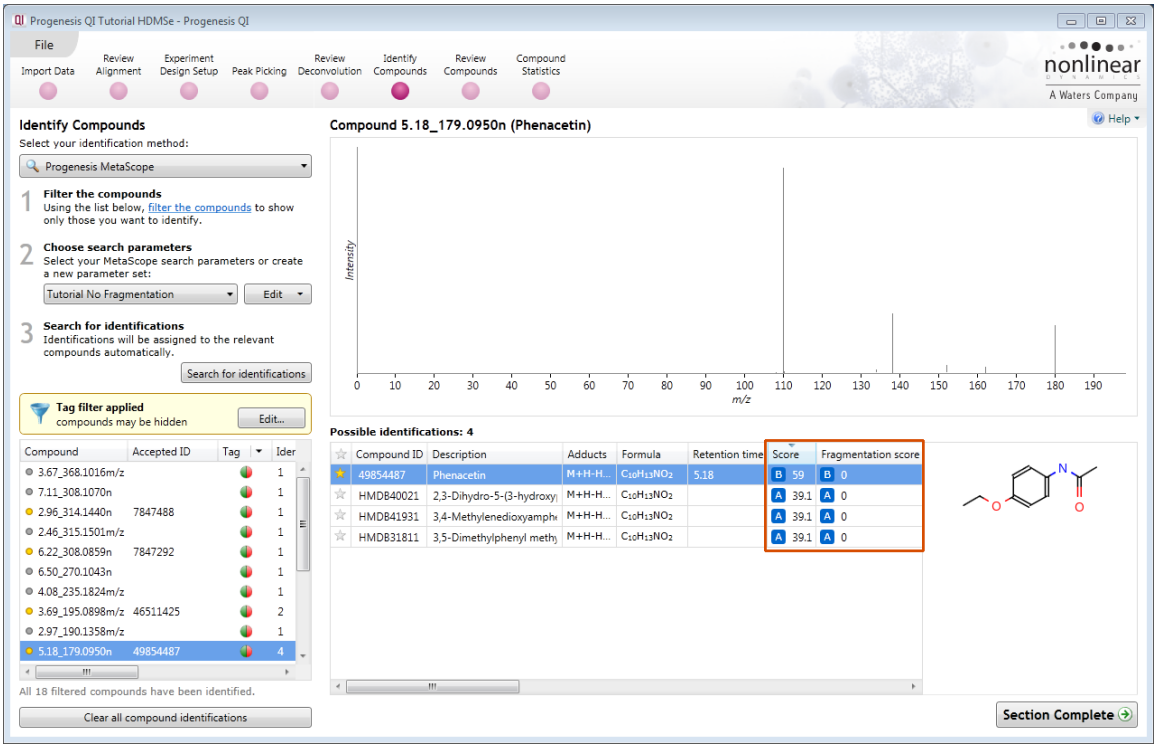


Note: if the data format contains Drift time data and it has been calibrated then the value for a compounds Collisional Cross Section (CCS) may also be stored as an additional compound property and used in the process of compound identification.

Now re-perform the Tutorial no Fragmentation search with the **Additional compound properties file** selected as the source.



When you re-perform the search, as expected, a higher overall score is achieved using the additional data.



Note: the saved **Additional Compound Properties** file is now the source of the theoretical information while the Compounds database file still provides the structure details.

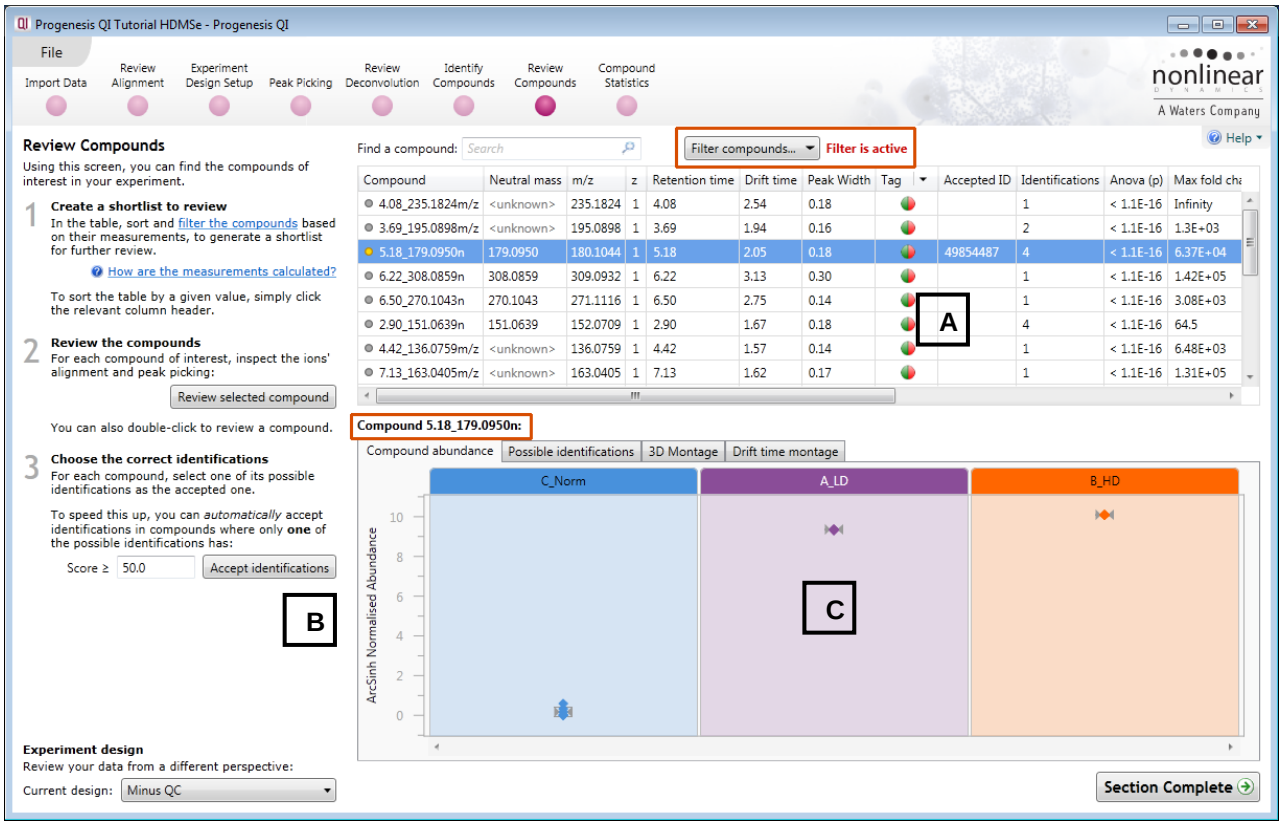
Each matching component (where available) contributes 20% to the overall score: Mass error, Retention time, Isotope similarity, Fragmentation score and Collisional Cross-Section (CCS).

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

Stage 9: Review Compounds

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

Note: you may have already set a Tag filter in the previous section that limits the table to displaying only those compounds that show a significant difference between groups and have one or more candidate identities.

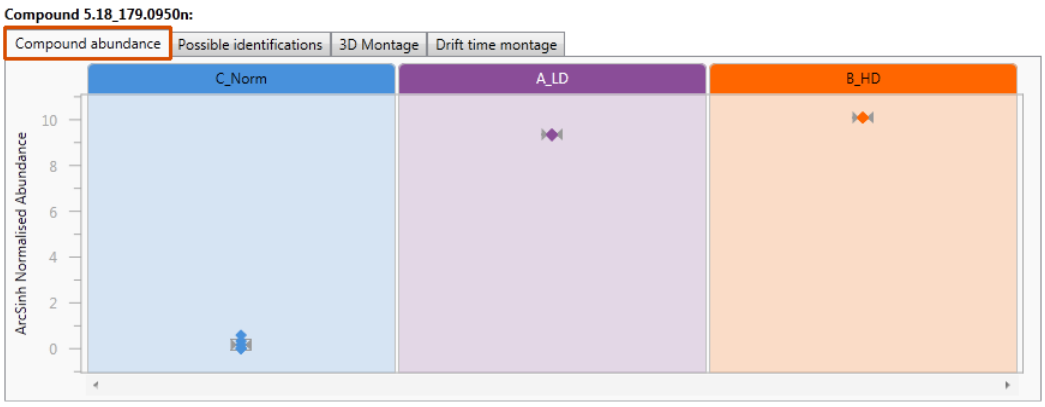


Window A displays the main table of Compounds with identifications as well as those that remain **unknown** after the identifications have been imported.

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

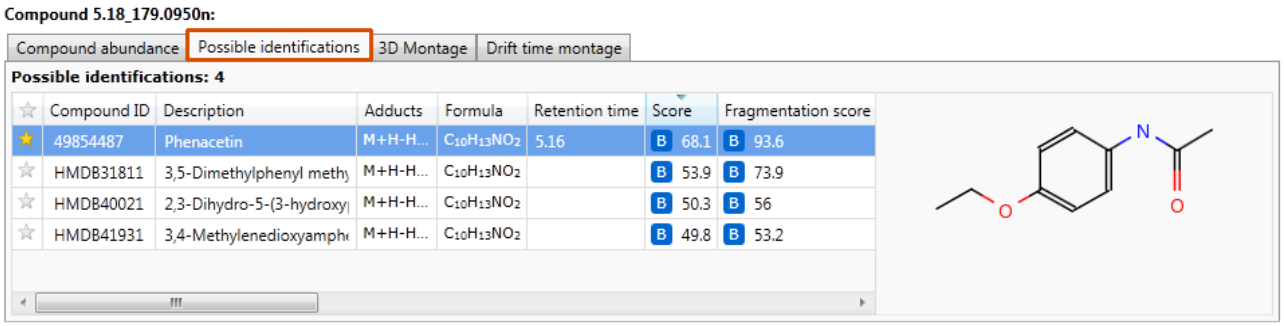
Window C displays either: a **Compound abundance** plot, list of **Possible identifications, 3D Montage**, **Drift Time Montage (for data collected with drift time)**, for the current compound highlighted in Window A.

Using **Compound 5.18_179.0950n** as an example the 4 views for Window C show:

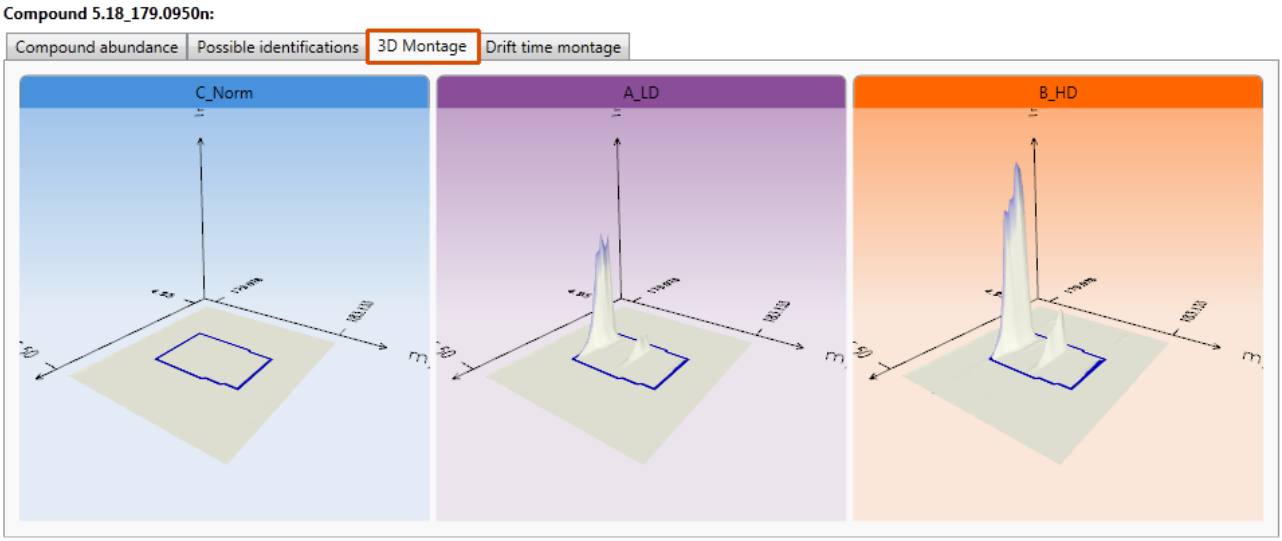


The **Compound Abundance** displaying normalised values for each run, a mean value and 3 standard deviations.

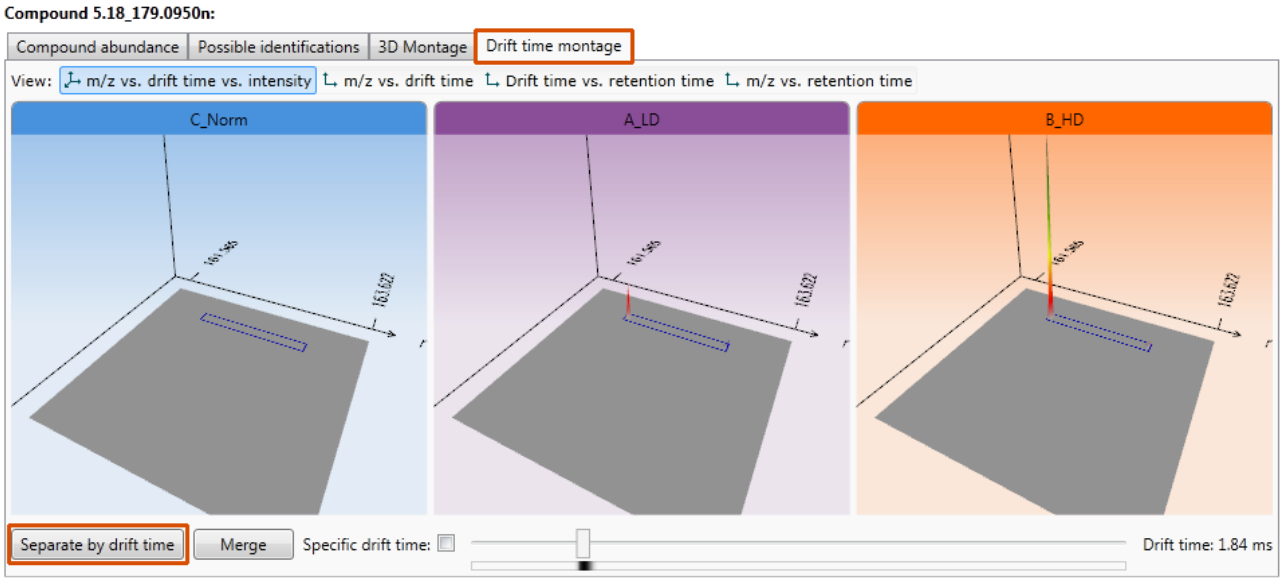
Possible identifications for current compound with Overall Score based on, Mass error, Retention Time (if available), Isotope Similarity based on the comparison of the measured isotope distribution for the compound vs the expected based on the compound formula, Fragmentation Score and CCS if available.



The 3D view based on the first run in each group.



The Drift Time Montage tab is only shown for HDMSe data

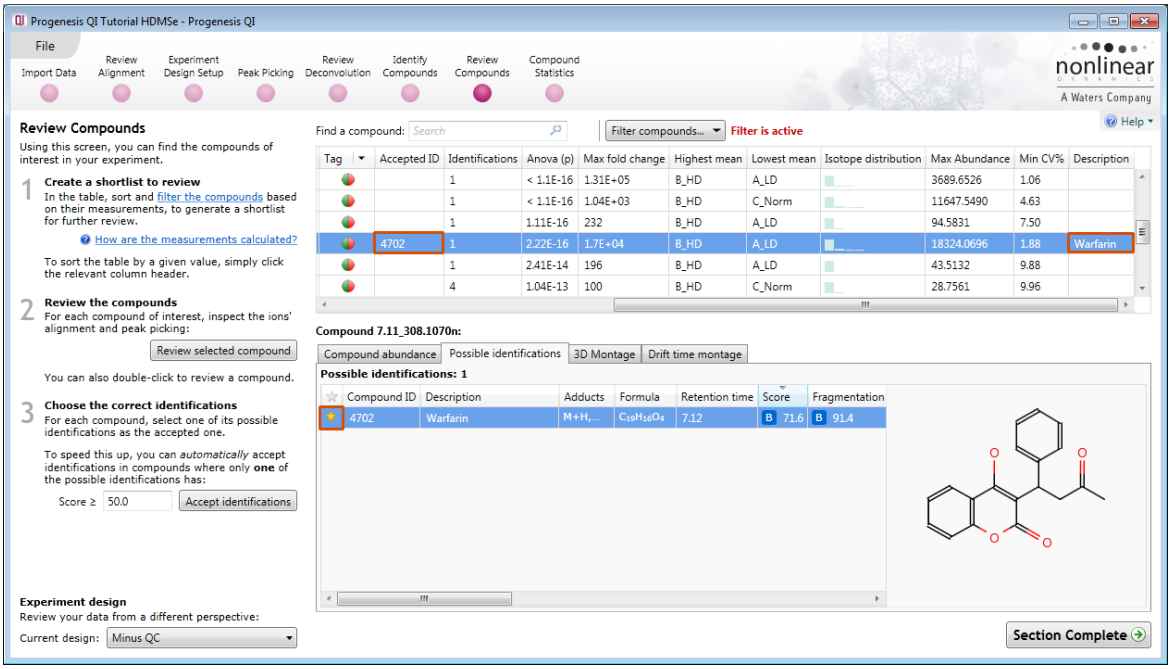


Setting a Compound's identity

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

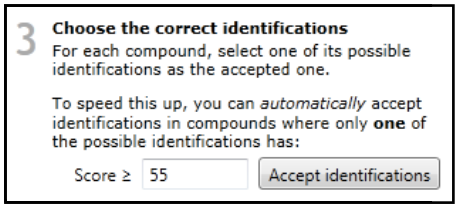
For example to accept **Warfarin** as the identity for **Compound 7.11_308.1070n** click on the 'star'.

This updates the compounds 'Accepted ID' and Description.

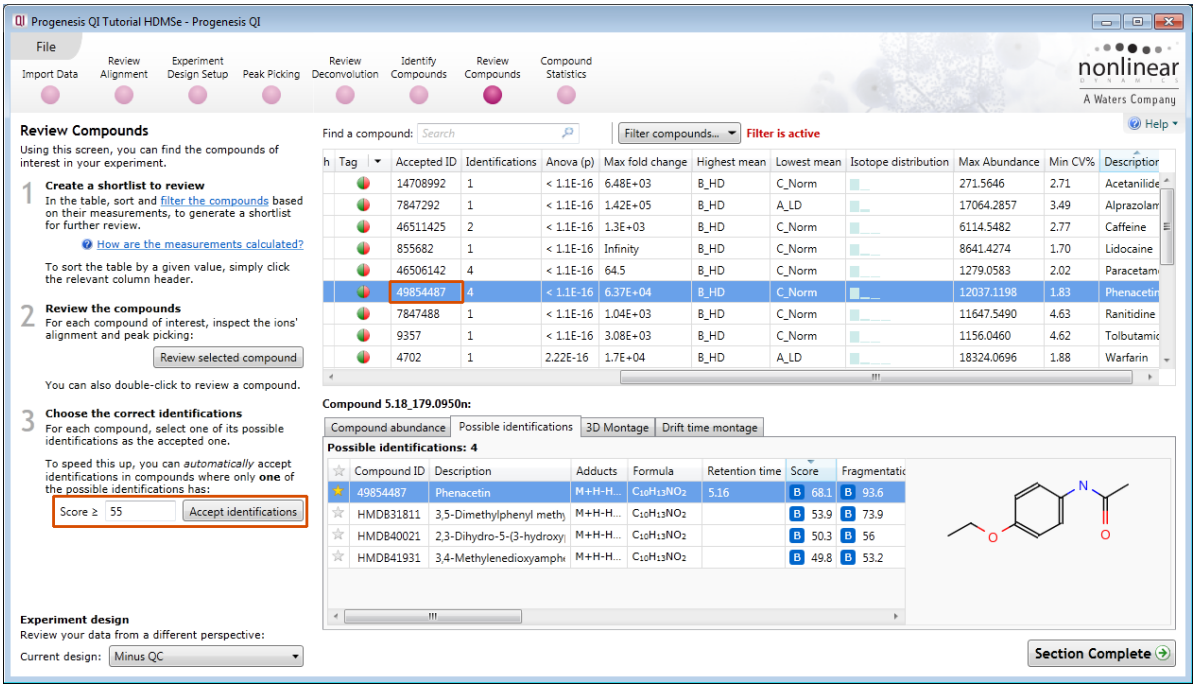


Rather than review all the possible compound identifications you can set a threshold for the score. This allows the automatic acceptance of compound identity where **only** one of the possible identifications is greater than the defined score.

As an example set the threshold as **55** and then order on **Accepted ID**.



Note: accepting a score threshold causes the Accepted ID and Description fields to be populated when there is a single ID with a score ≥ 55 for the compound.



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.

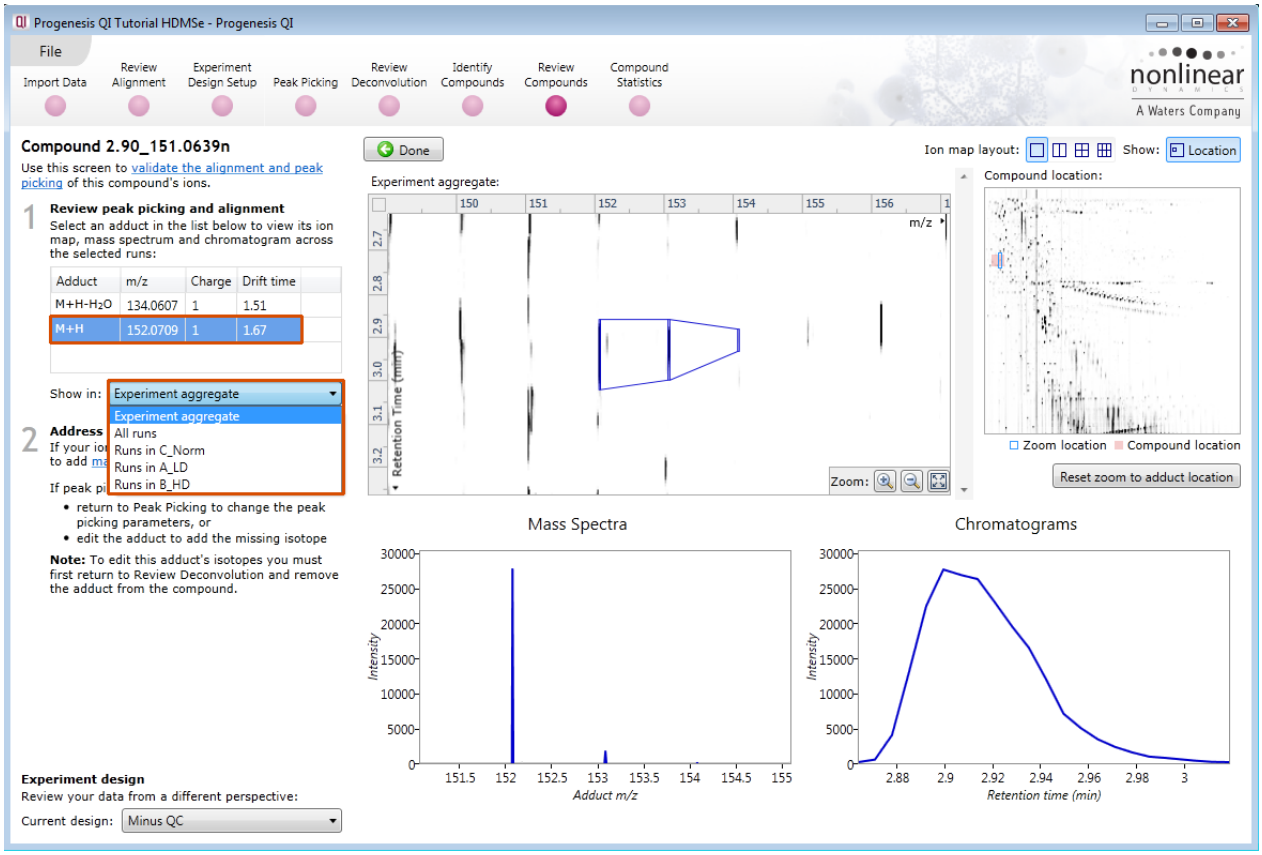
In this view:

- 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 identified as Paracetamol:

As the **Compound Validation** view opens you can select which adduct of the compound to review first.



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. The aggregate pattern of detection can be edited to add a missing isotope using the **Edit adduct** facility.

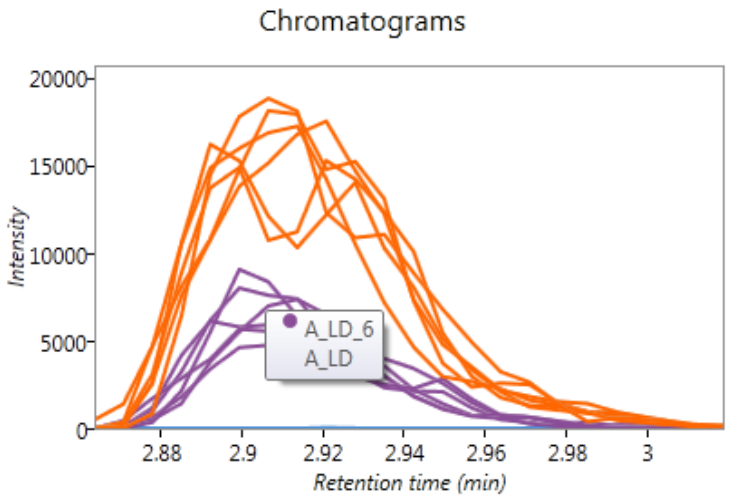
Note: details on Editing adducts is available in Appendix 5 (page 74)

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.

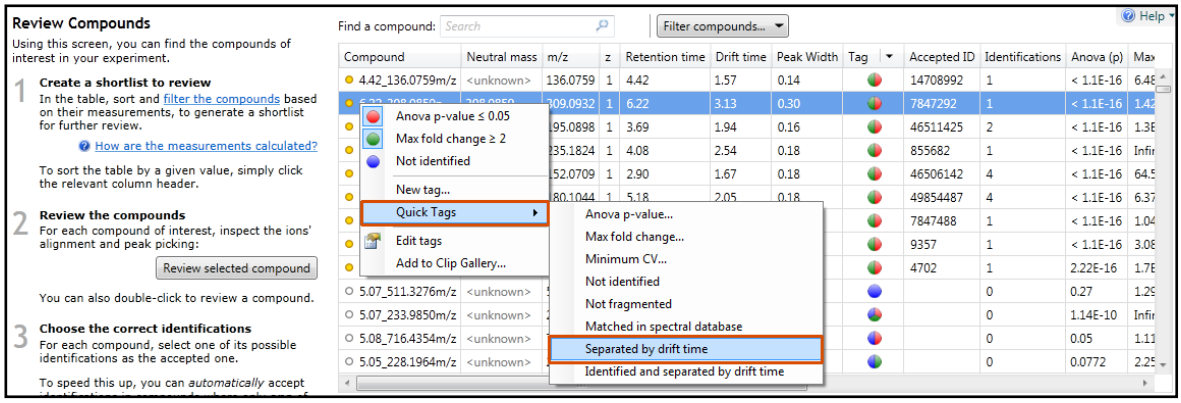
Note: that when the **All runs** are selected the mass spectra and chromatograms are shown for all the runs in the experiment.

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.



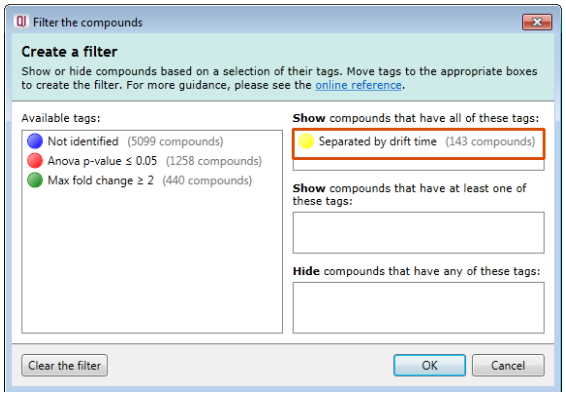
Compounds separated by Drift Time

To view compounds separated by drift time right click on the Review Compounds Table and create a Quick Tag for those compounds Separated by drift time

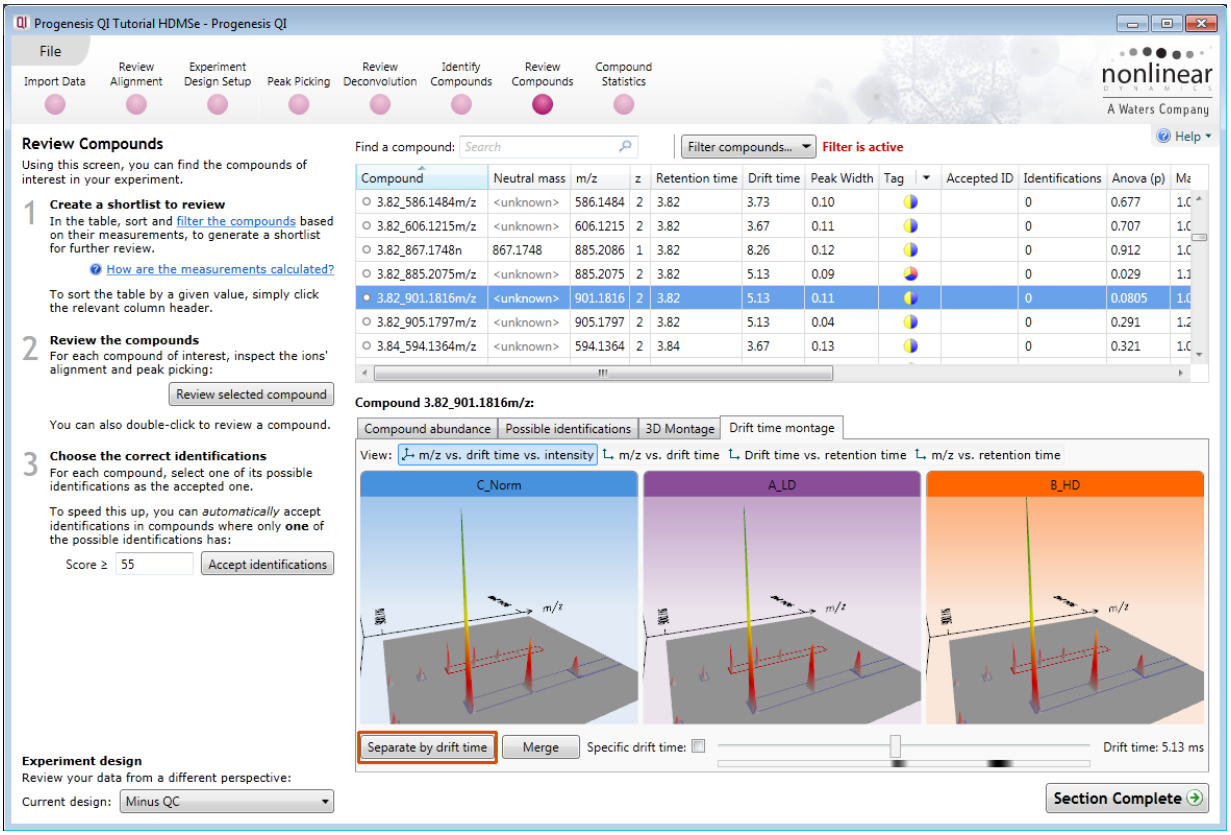


Once you have created the tag then filter the table to show only those compounds Separated by Drift Time.

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

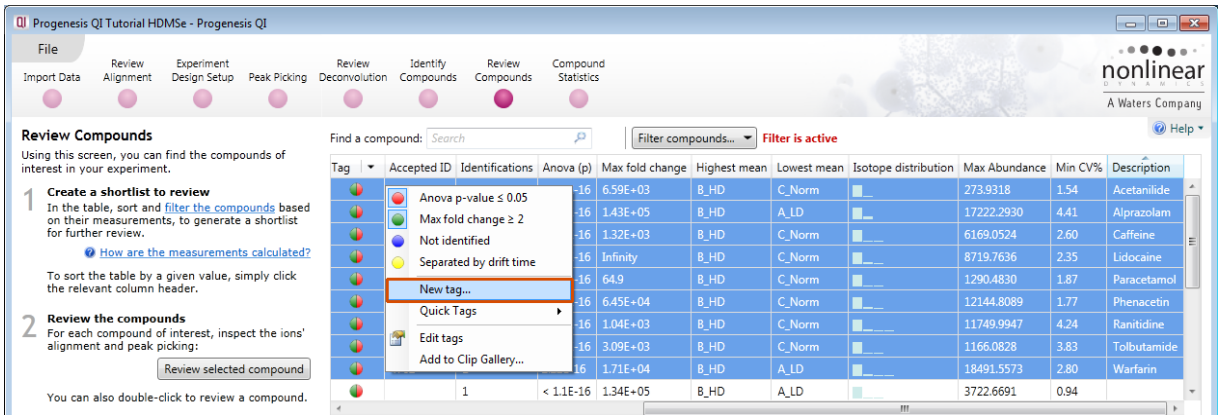


The table now shows the only the compounds that are separated by drift time.

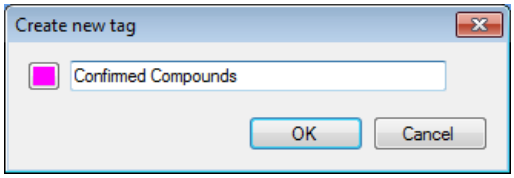


Exporting compound data

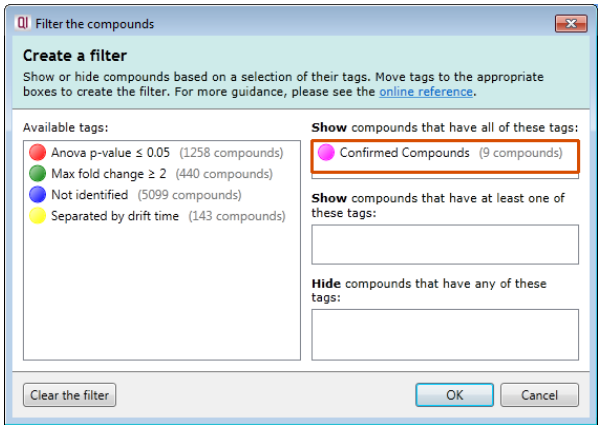
Compound data can be exported in a csv file format. You can either export the **compound measurements** or the **compound identifications.** As an example of Data export first clear any existing filters and then order the table on **Description** and highlight all the compounds which have a description.



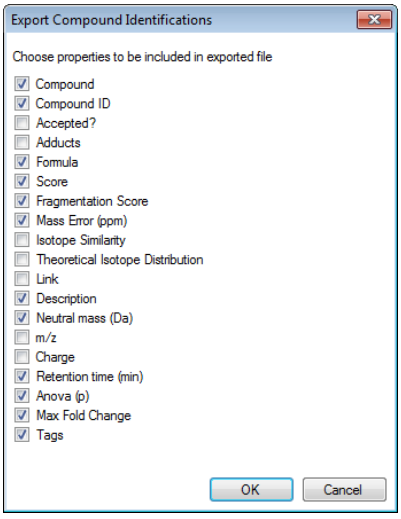
Then right click on the highlighted compounds and create a **New tag…** called **'Confirmed Compounds'**



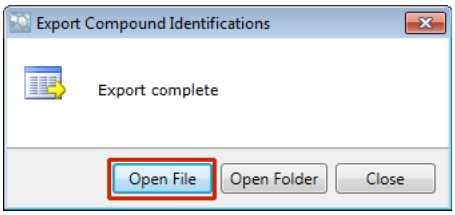
Click on Filter Compounds then drag the new tag onto the Show panel of the **Filter compounds** dialog and click **OK**.



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:

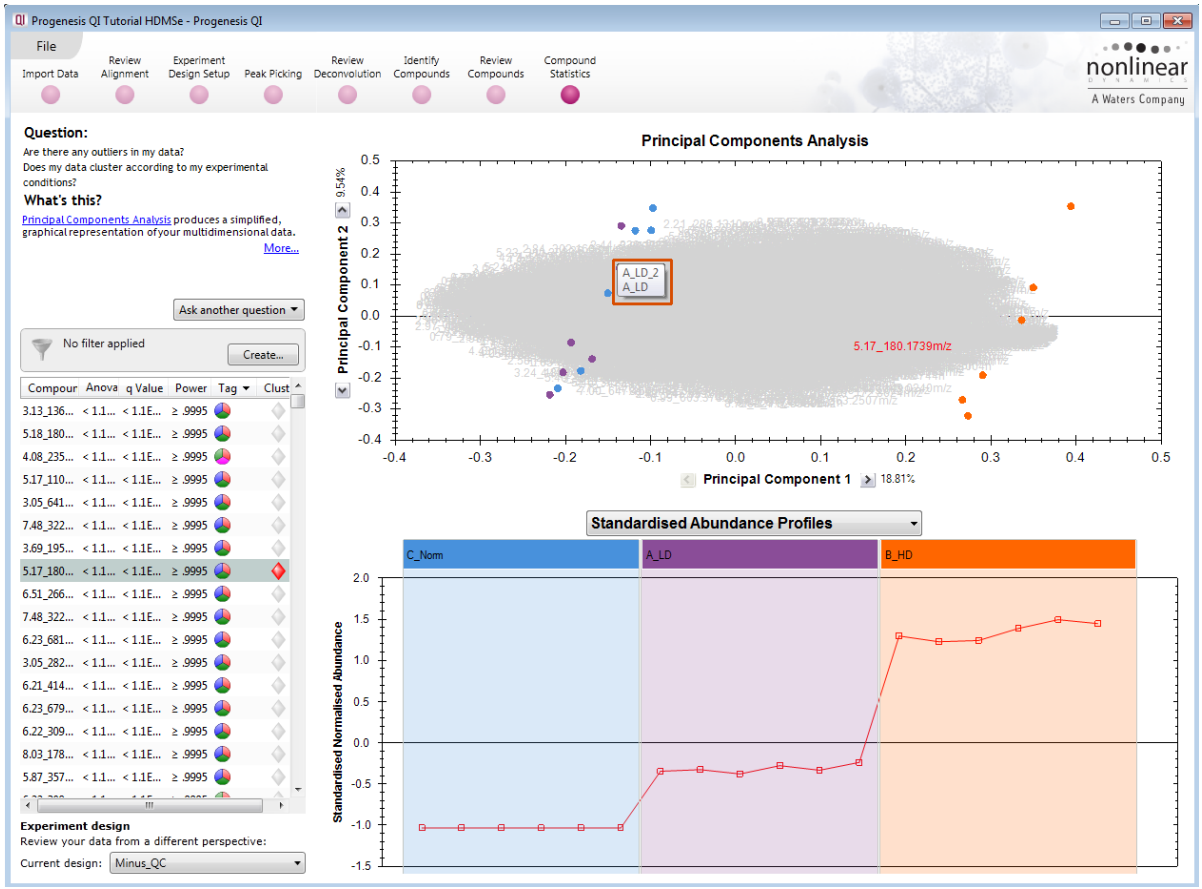


Stage 10: Compound Statistics

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

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

As an example we will start by examining the PCA for all of the 5168 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.

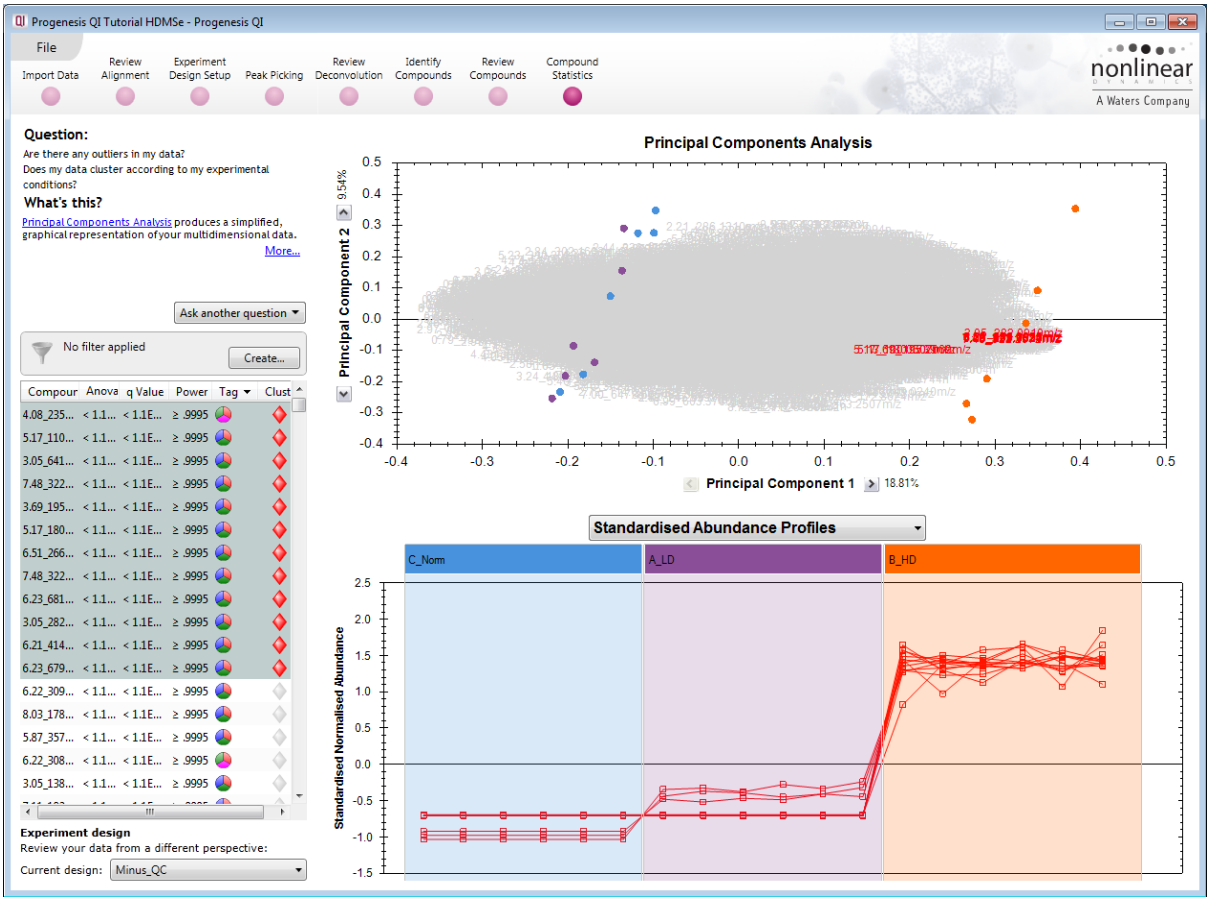
Principal Component Analysis (PCA)

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

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

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

PCA can be used to determine whether there are any outliers in the data and also look at how well the samples group. The groupings that can be observed on the 2D PCA plot can be compared to your experimental conditions and conclusions can be drawn regarding possible outliers in your data. Selecting 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.

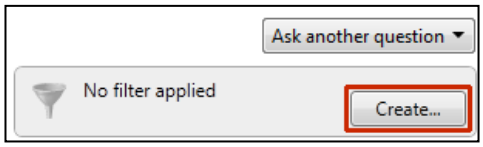
Also, for a given power of 80% we can determine how many samples are required to ensure we find a difference if it actually exists.

Note: Power analysis is discussed in Appendix 6 (page 79)

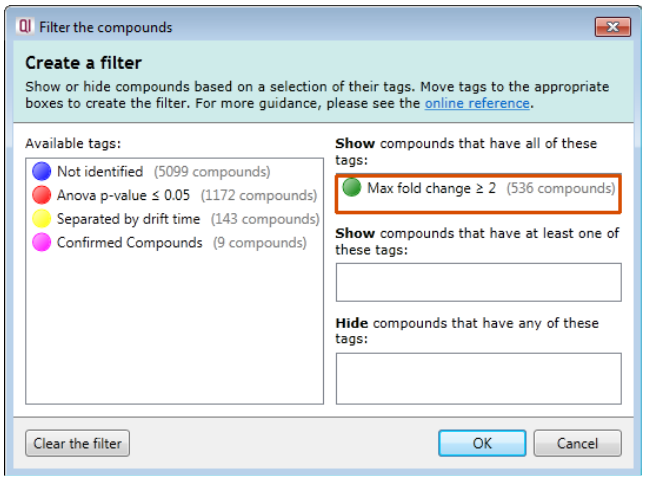
Correlation Analysis

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**

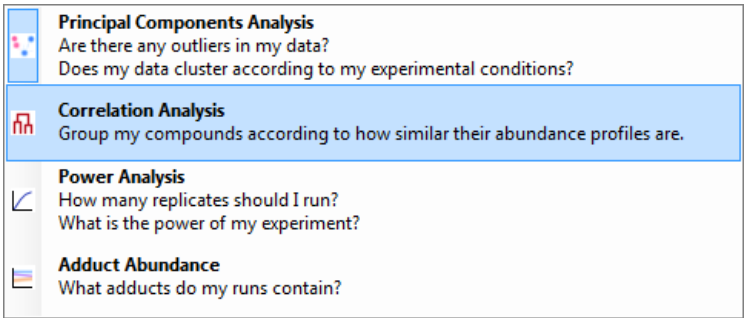


Set up the filter as shown below

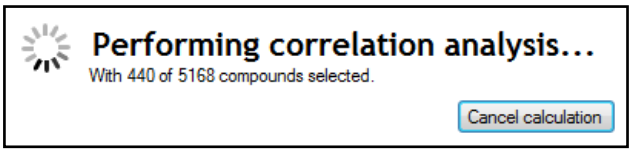


On pressing OK the PCA will recalculate.

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’



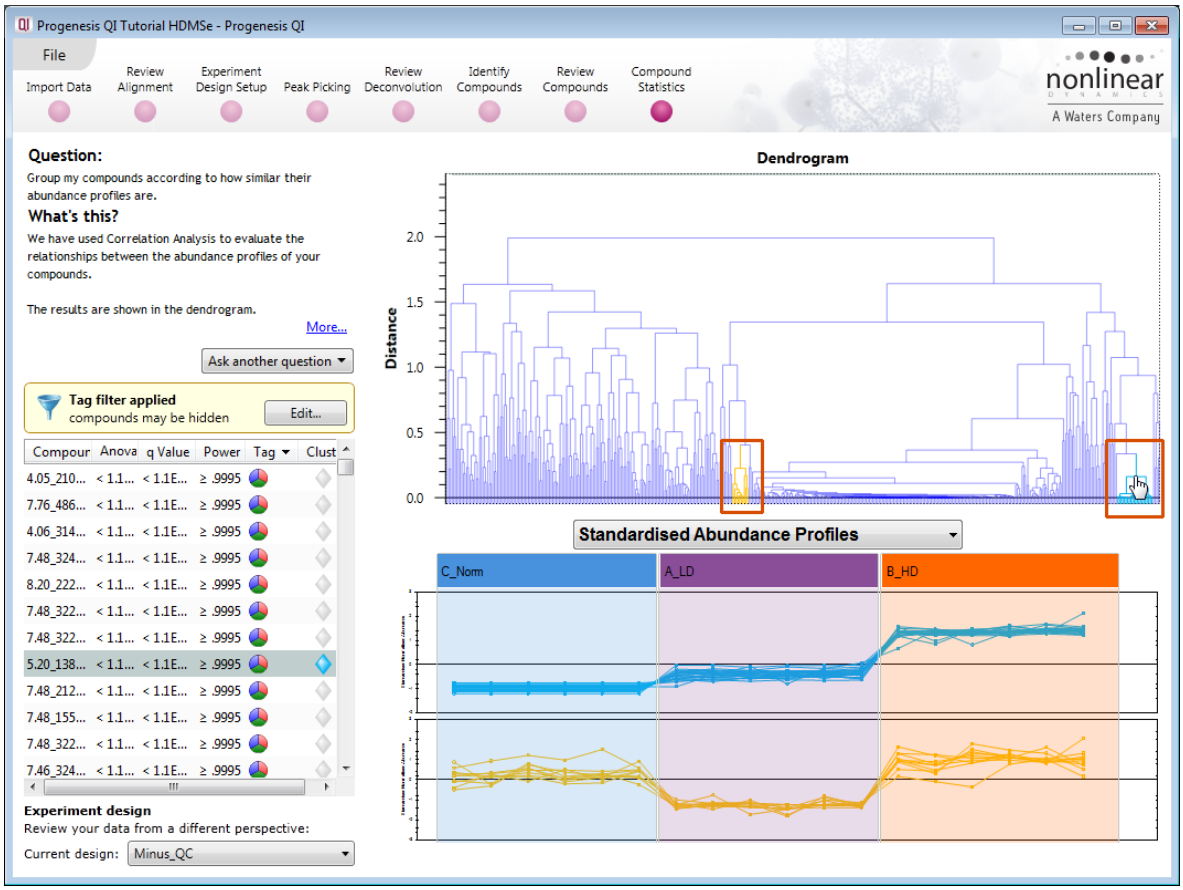
This time the statistically based question(s) being asked is:

‘Group my (selected) features according to how similar their expression profiles are’

The question is answered by:

‘Using Correlation analysis to evaluate the relationships between the (selected) features’ 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.

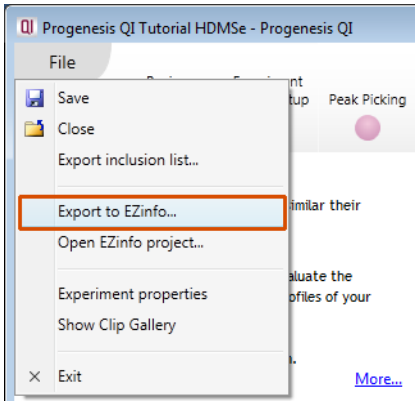


Correlation Analysis enables the grouping of compounds together according to how similar their abundance profiles are.

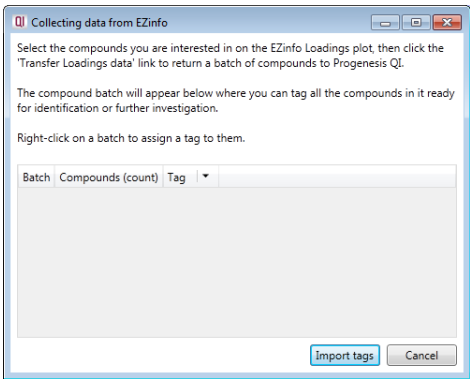
Clicking on a branch on the Dendrogram selects the compounds on the table. You can then tag this group, of 'potentially' related compounds, by right clicking and creating a 'New tag' for them. Then use the tag to focus on them at the **Review Compounds** stage.

Exporting to EZinfo

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

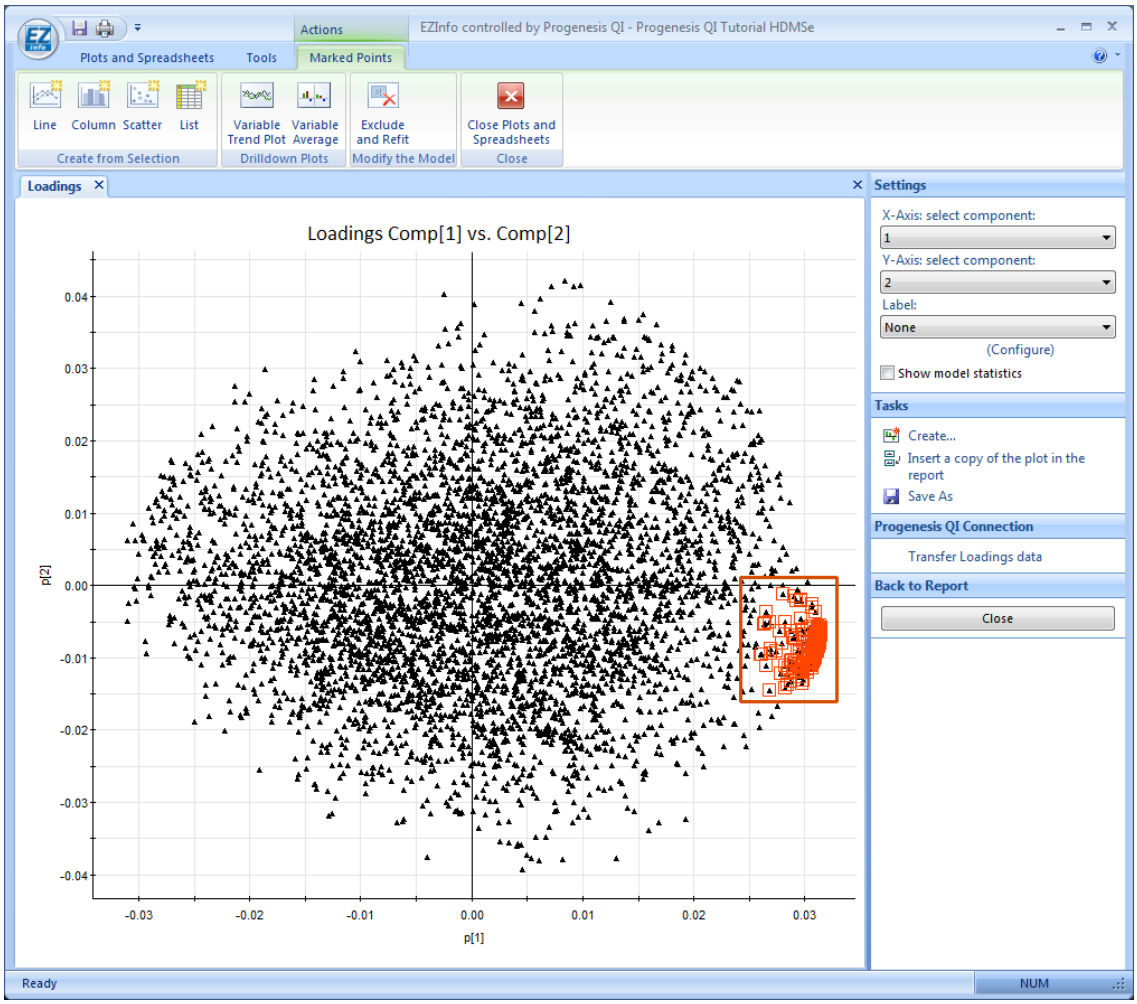


Save the EZinfo project in the same folder as your Progenesis QI experiment



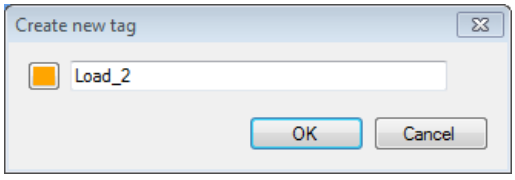
A dialog for **Collecting data from EZinfo** opens in Progenesis QI and EZinfo opens displaying the exported data in EZinfo. Change the to the loadings plot :

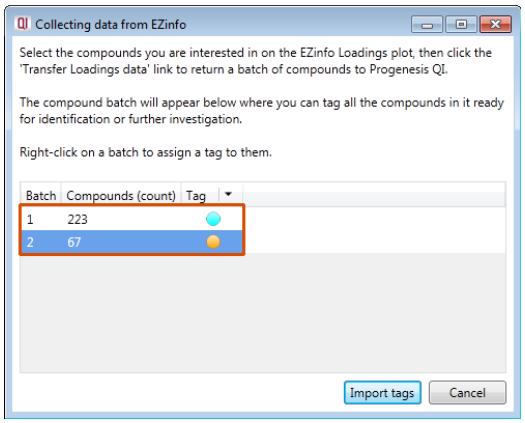
Select a set of points and click on **Transfer Loadings data**. Repeat this process for other sets of data.



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

Select and right click on each batch and create a **New tag…**

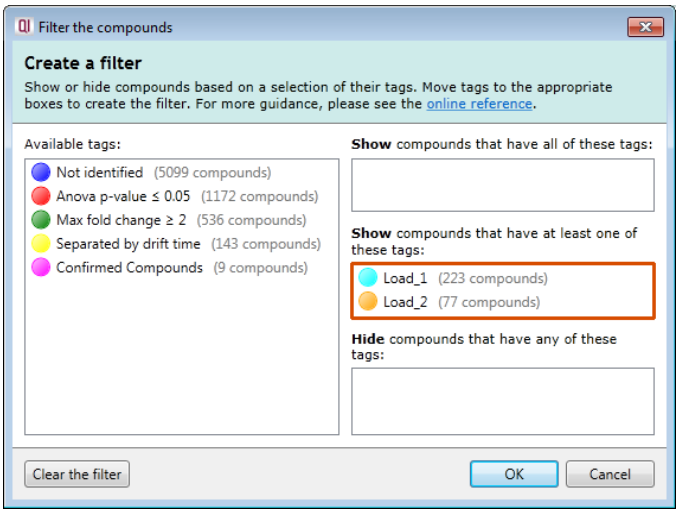




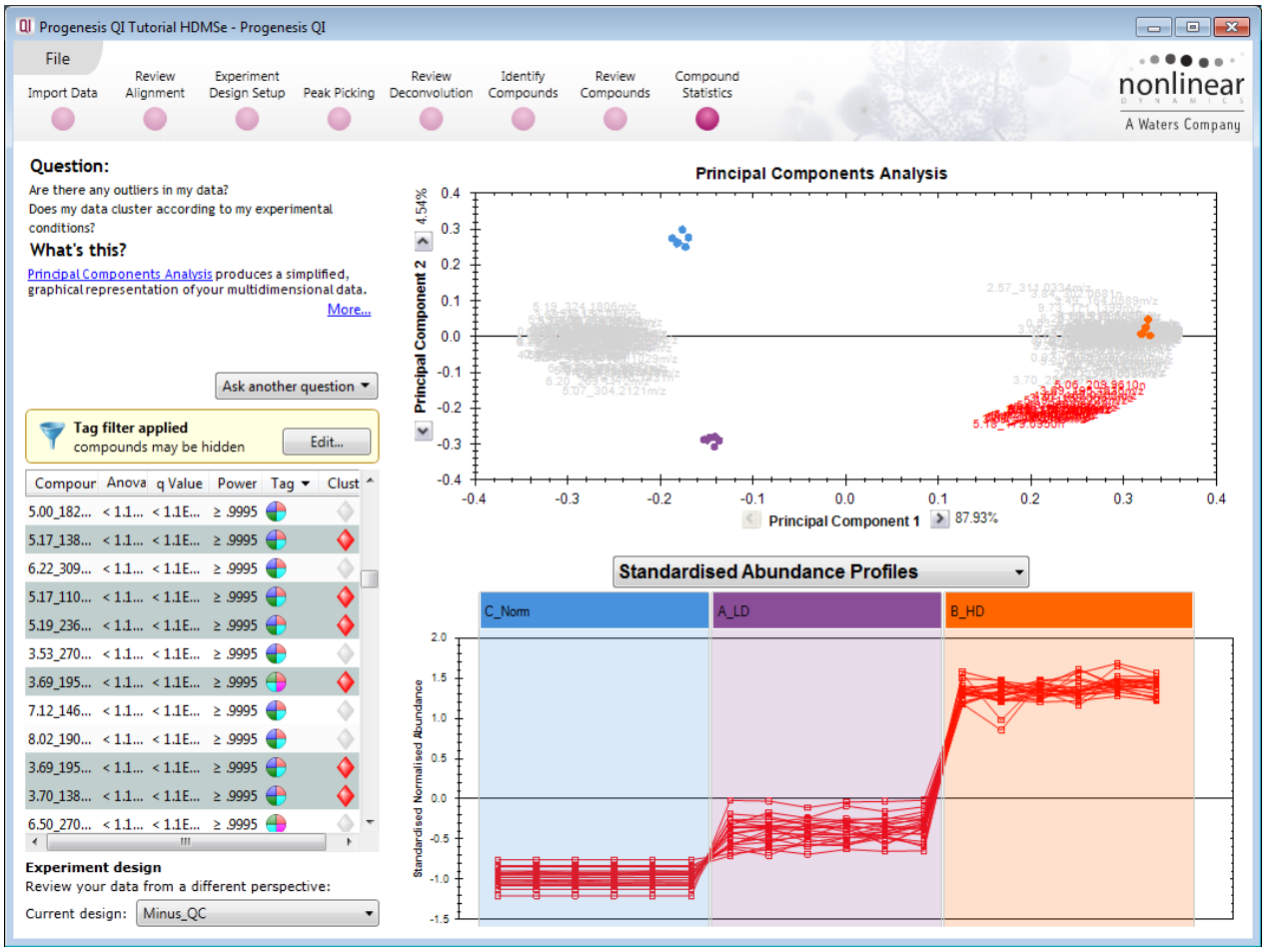
Click **Import tags**

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

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



Click OK to create the new PCA.



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

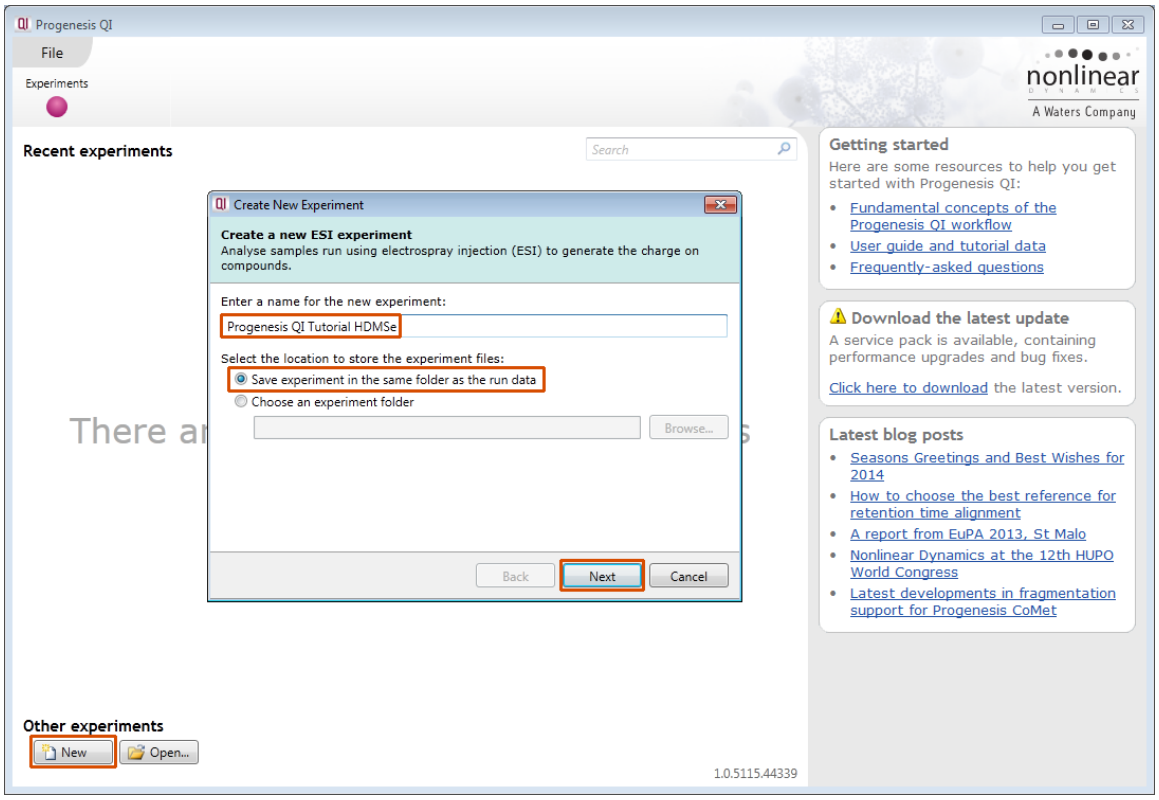
Congratulations!

This document has taken you through a complete analysis using Progenesis QI, from Alignment through Analysis to generating lists of interesting compounds using powerful Multivariate Statistical analysis of the data.

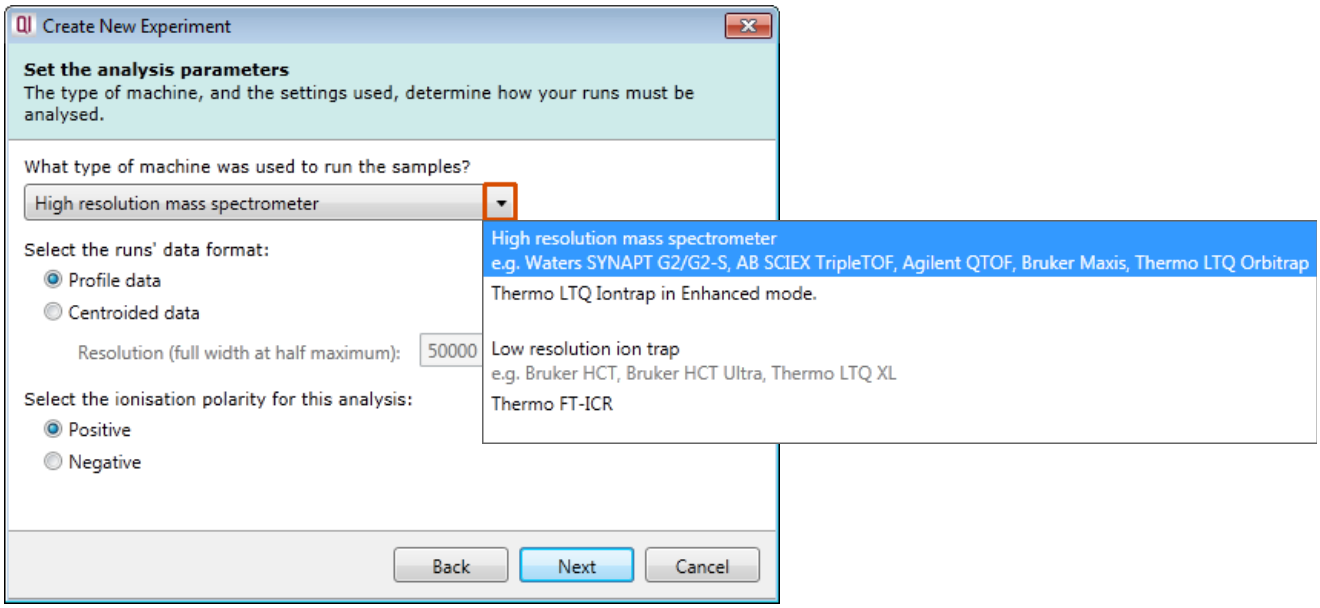
Appendix 1: Stage 1 Data Import and QC review of data set

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**.

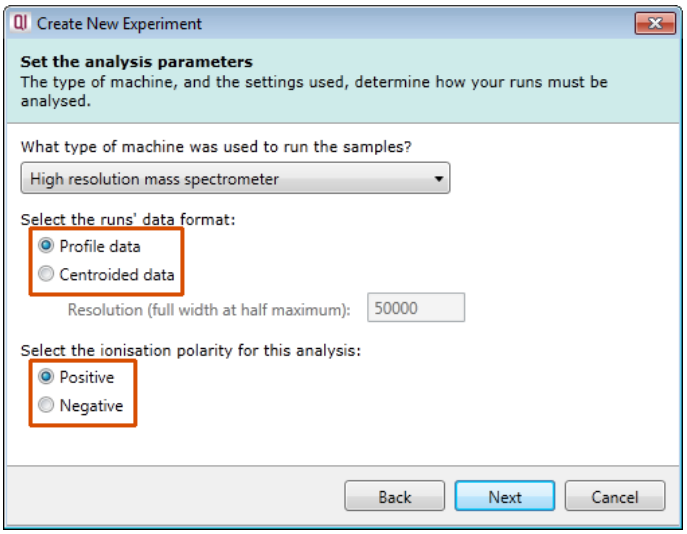


Now select the machine type that was used to run the samples.

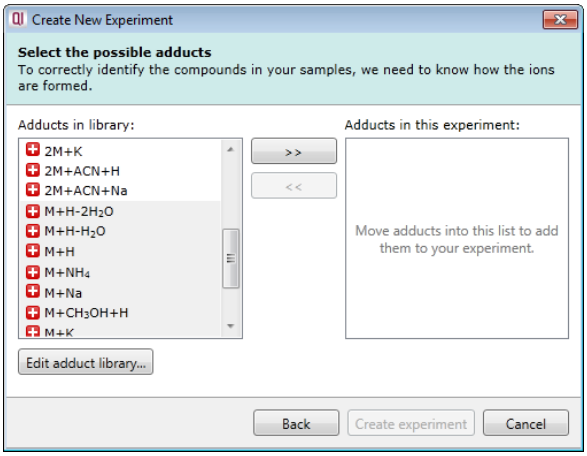


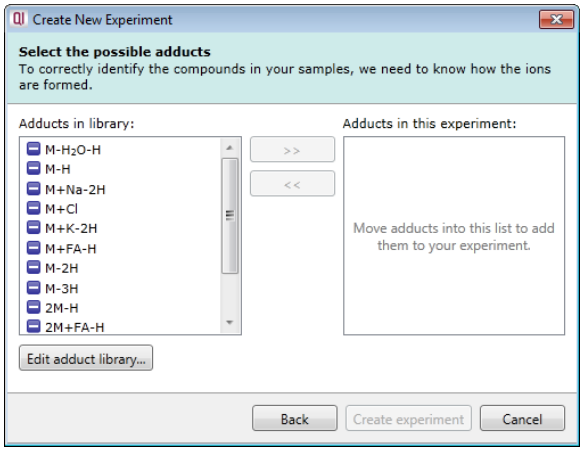
Note: if you have converted or captured the data as centroided then select Centroided data and enter the Resolution (full width at half maximum) for the MS machine used.

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

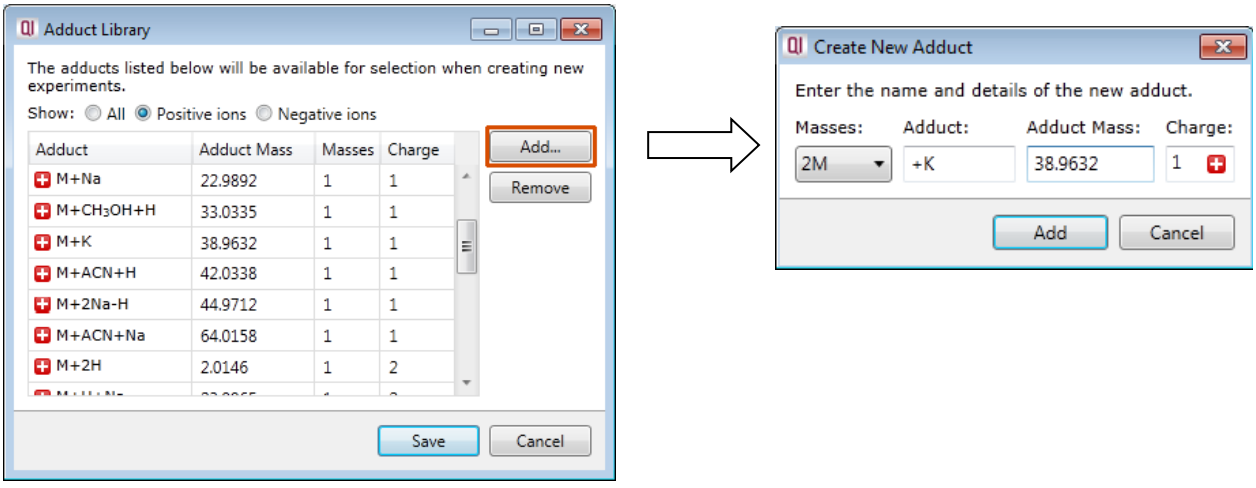


Finally select the ionisation polarity, either **positive** or **negative**, this will determine which list of possible adducts is available on the next page.

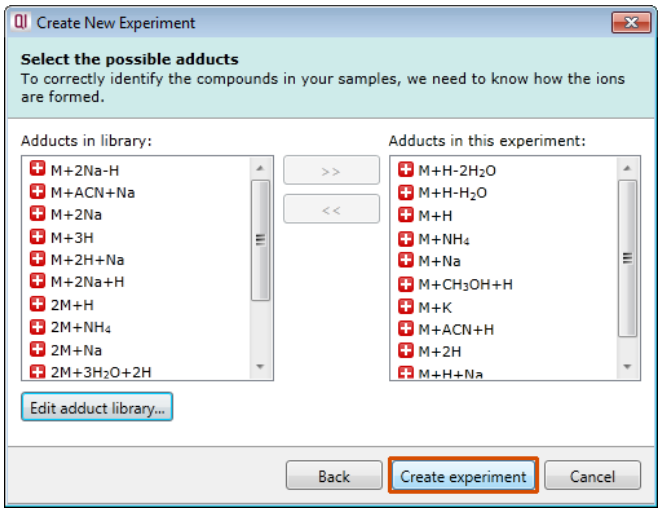




To add an adduct to the library select **Edit adduct library…** and then click **Add**. Now provide the Mass (monomer, dimer or trimer), Adduct, Adduct Mass and Charge of the new adduct and then add it to the library.



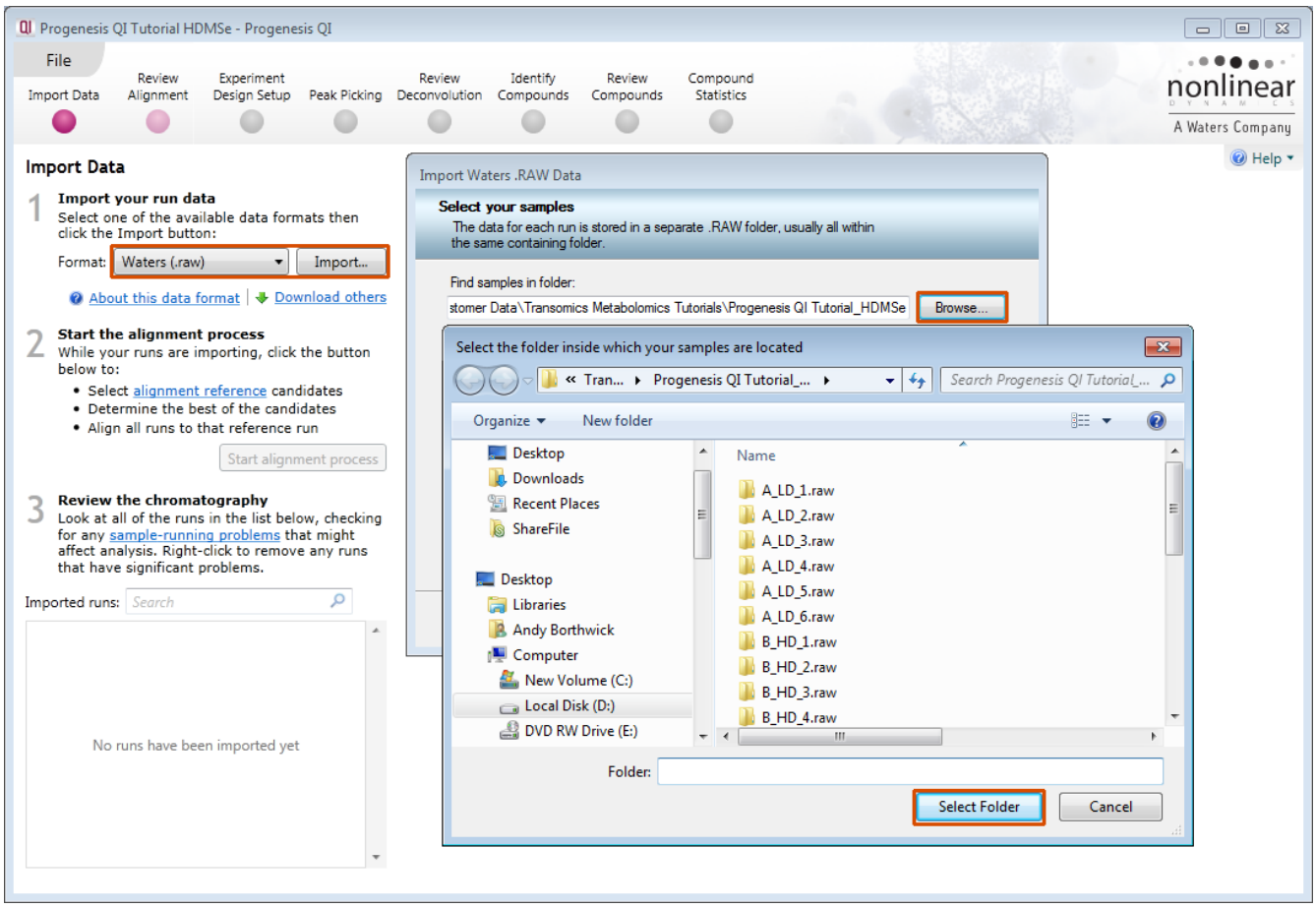
Now select the expected adducts based on you knowledge of your experimental conditions. In this example the polarity was positive and a basic set of Adducts was selected as shown on the right panel below.



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

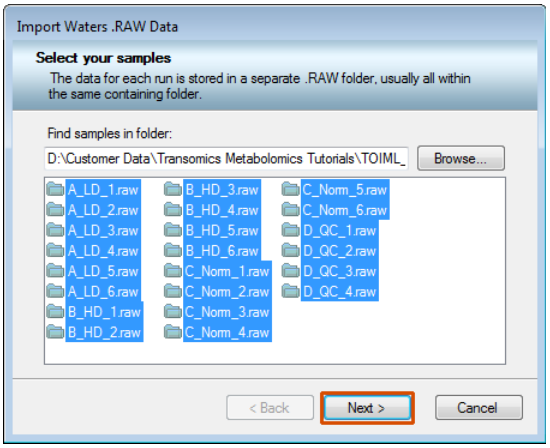
Select the 'Import Data file format', in this example they are **Waters (.raw)** files

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).

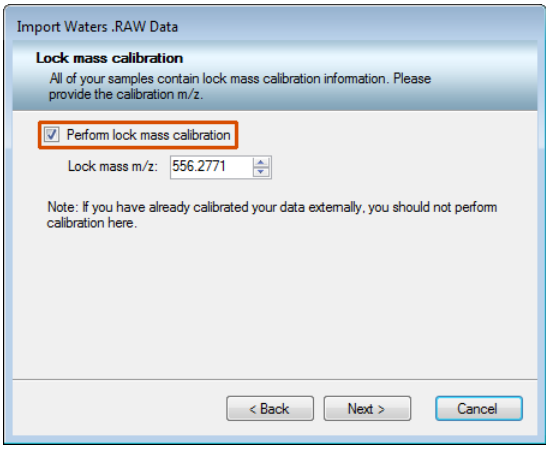
Select the required **.raw** folders for your experiment and click **Next.**



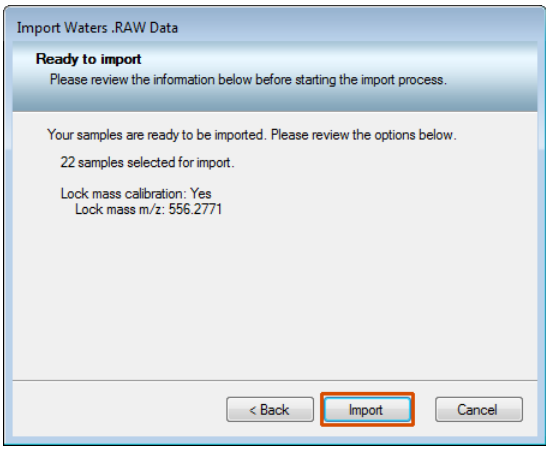
The folders will be examined and the Dialog will ask you for Lock Mass Calibration details.

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

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



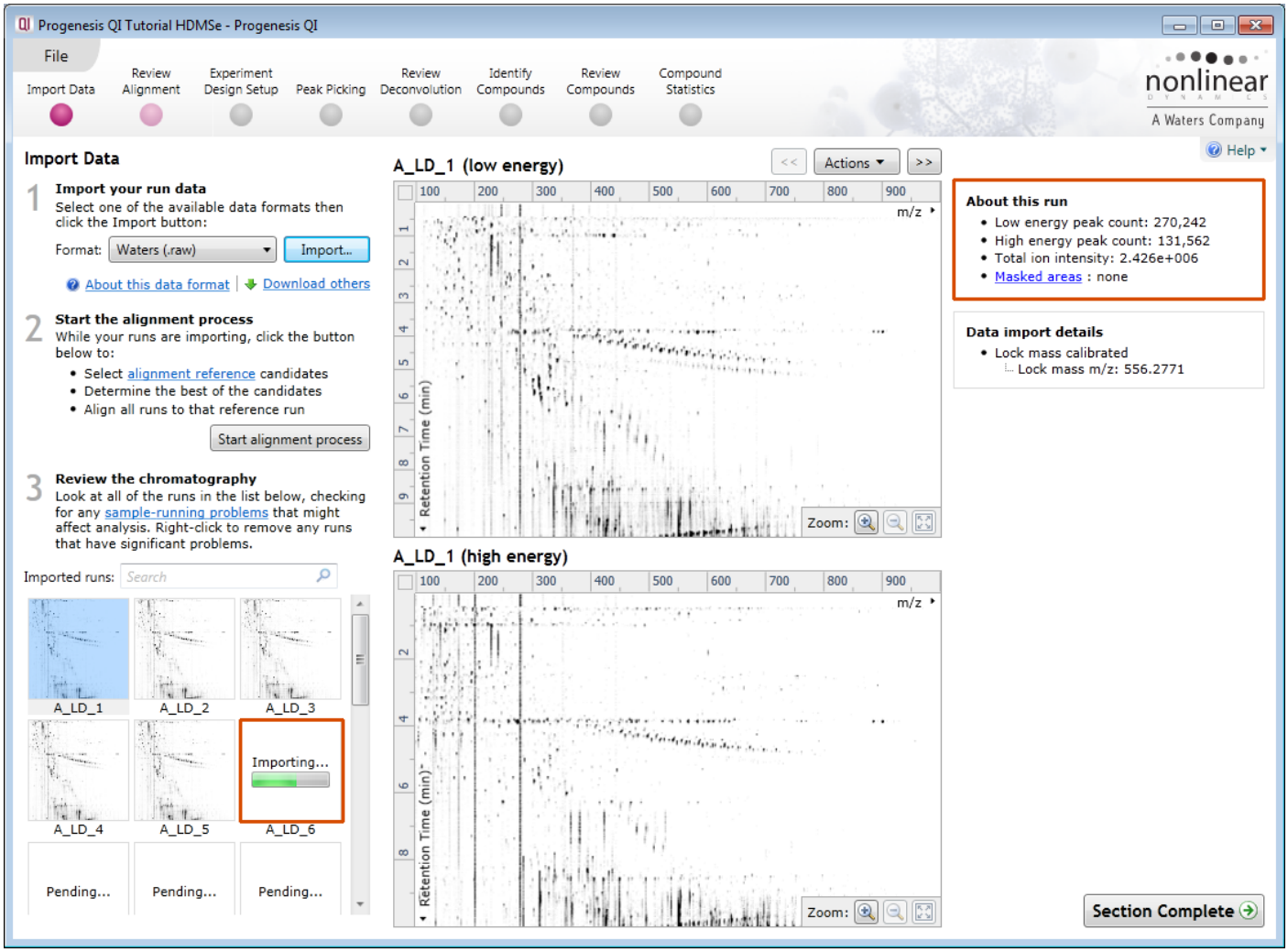
Finally a summary page appears. Click **Import**



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 you will be warned if any of the data files have been 'centroided' during the data acquisition and conversion process.



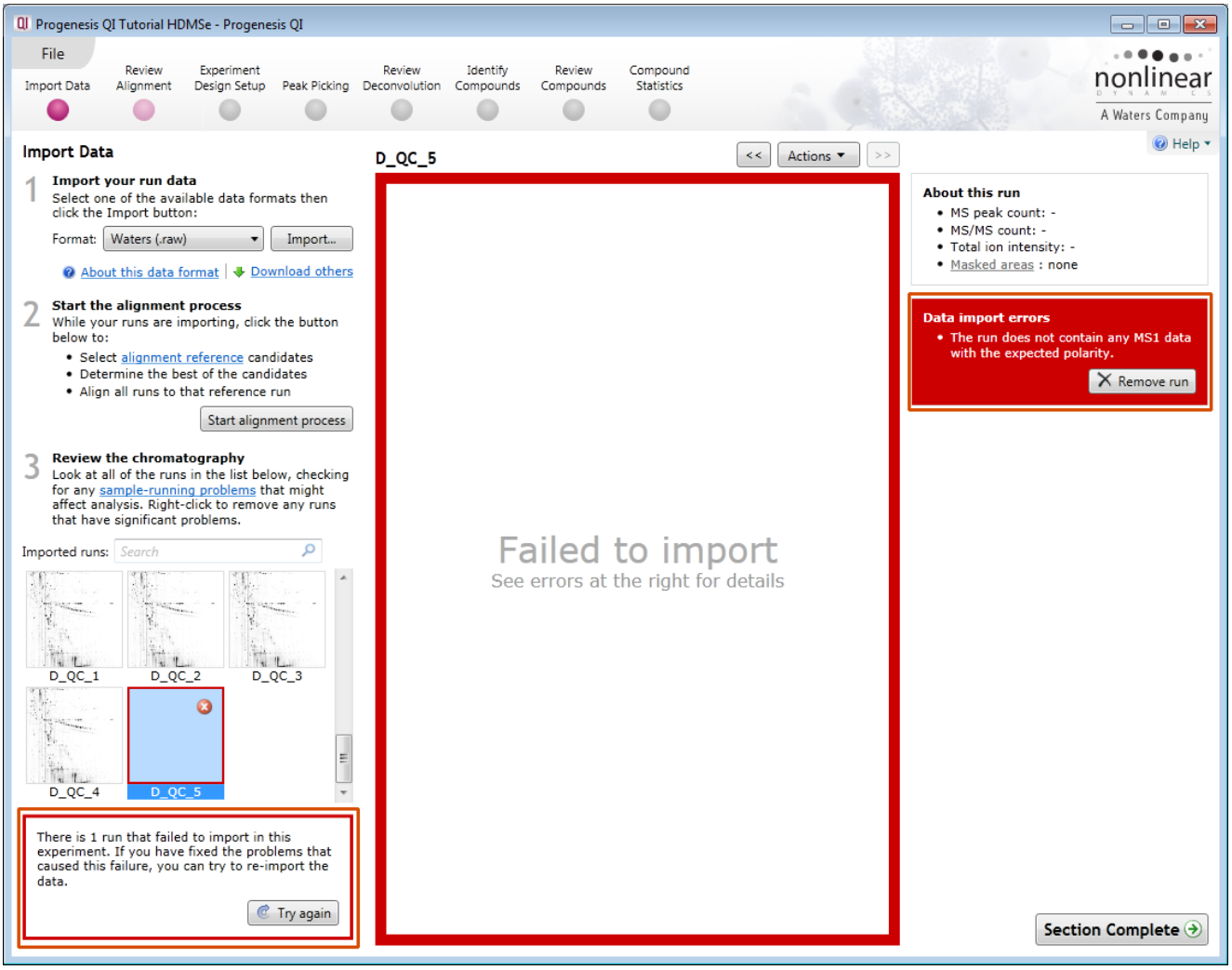
Note: as each data file is loaded the progress is reported in the **Import Data** list. The dialog below the run reports on the QC of the imported Data files. In this case 'No problems found' with this data file.

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

Note: as the loading process starts you can also start the automatic alignment before the loading has completed. This is a 2 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.

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

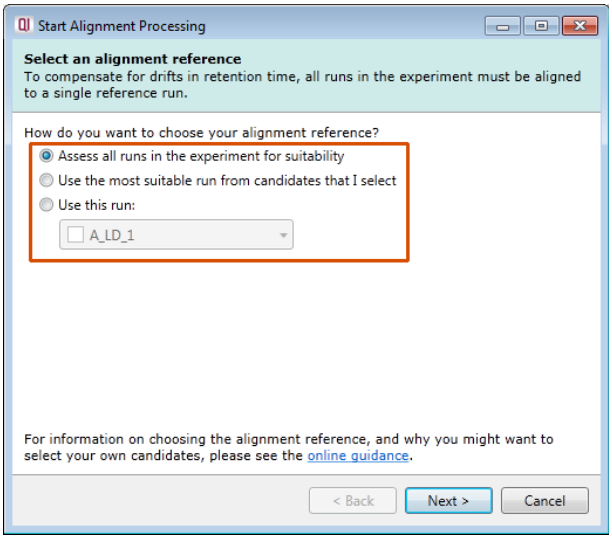
An example of a failed import is shown below



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 alignment process** to start the automatic alignment of your runs.

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

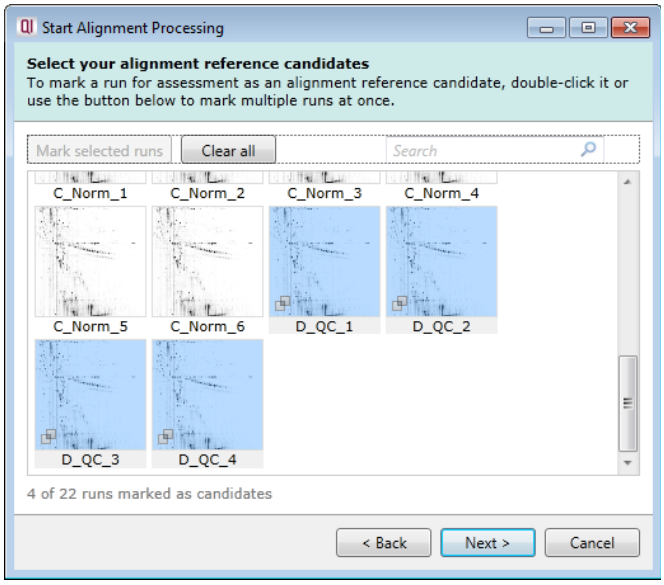


1. **Assess all runs in the experiment for suitability**
   This method compares every run in your experiment to every other run for similarity.
   The run with the greatest similarity to all other runs is chosen as the alignment reference.

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 can take a long time

2. **Use the most suitable run from candidates that I select**

This method asks you to choose a selection of reference candidates, and the automatic algorithm chooses the best reference from these runs.



When you have some prior knowledge of your runs suitability as references:

runs from pooled samples (QC pools for example)

runs for one of your experimental conditions will contain the largest set of common compounds

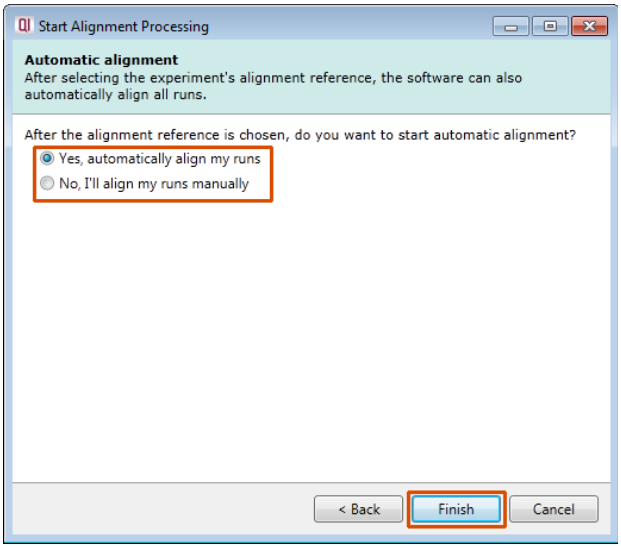
3. **Use this run**

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:

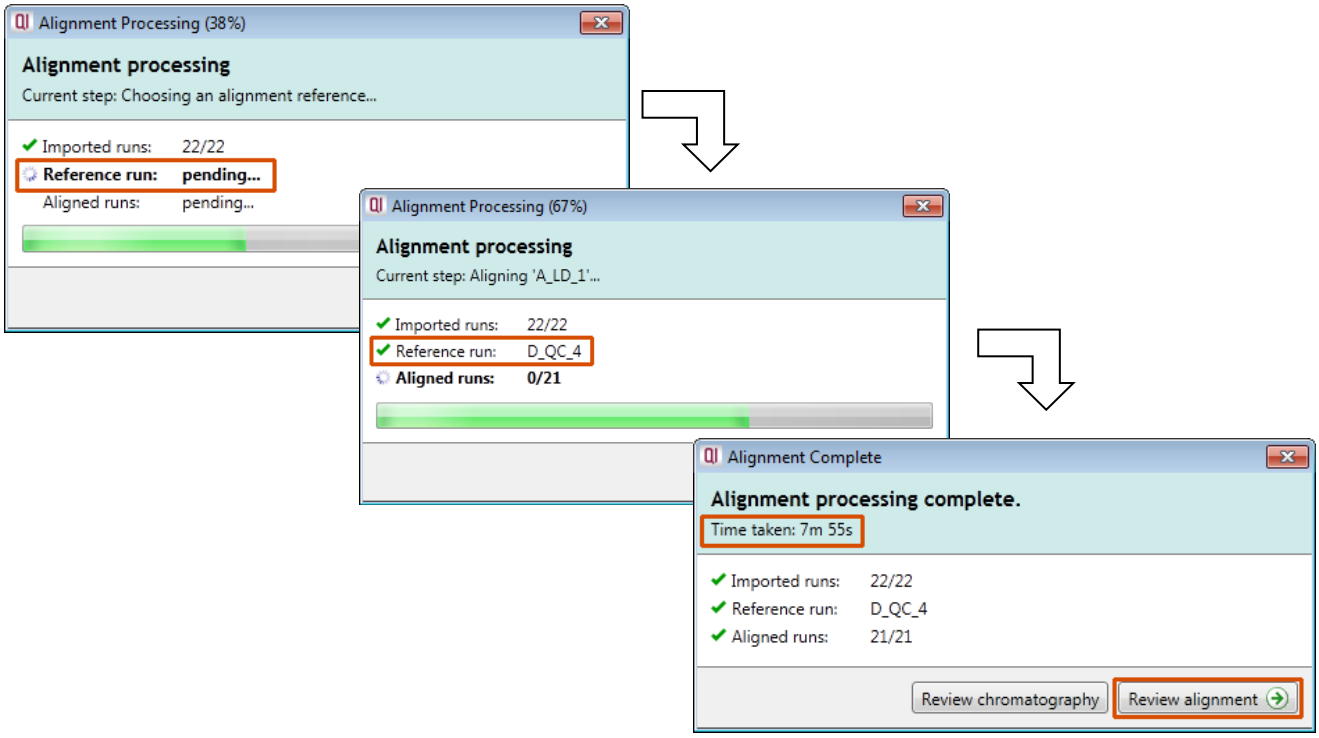
- If you choose a pending run which subsequently fails to load, alignment will not be performed.
- If you choose a run before it fully loads, and it turns out to have chromatography issues, alignment will be negatively affected (for this reason we recommend that you let your reference run fully load and assess it's chromatography before loading further runs).

Once you have selected how to handle the choice of Reference run you will now be asked if you want to align your runs automatically or manually.



Select automatically and click finish.

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



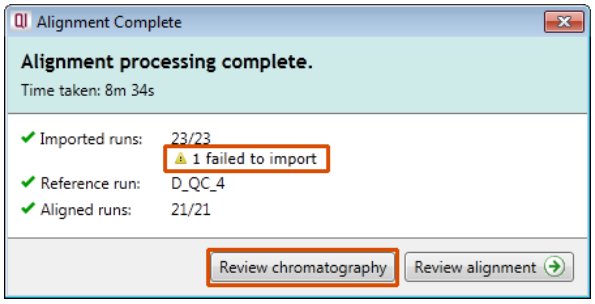
Once the Reference run has been chosen the automatic alignment is performed. As the whole process proceeds you get information on what stage has been performed and also the % of the process that has been completed.

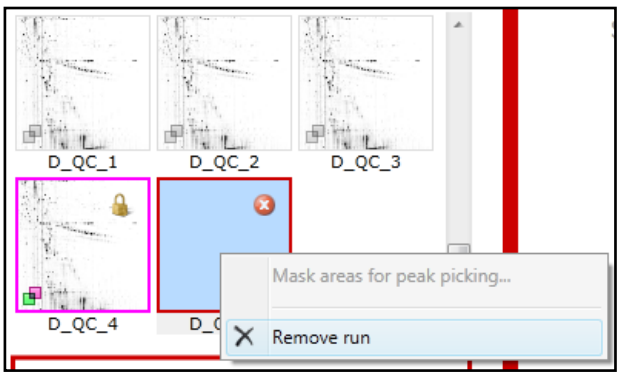
Note: At this stage you have the option to Review the Chromatography or go straight to the review of the Automatic Alignment of your data.

Review Chromatography

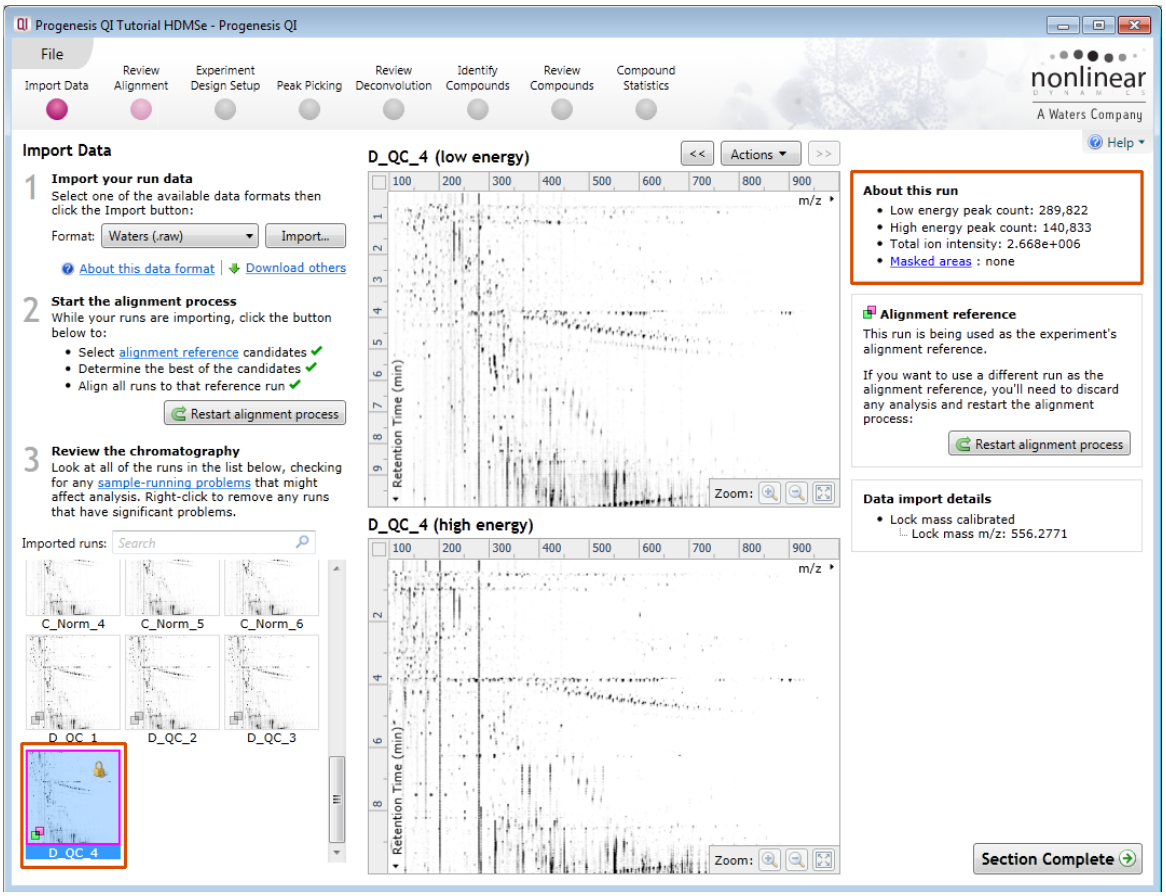
Each data file appears as a 2D representation of the run. At this stage you will be warned if any of the data files have failed to import. The reasons for import failure(s) include: run is centroid when importer is expecting a profile (continuum) run, the run has a different polarity to the other runs etc..

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





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



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

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

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

Now move to the next stage in the workflow (page 11 in this user guide) by clicking **Section Complete.**

Appendix 2: Licensing runs (Stage 3)

When setting up a **New experiment** if you are evaluating Progenesis QI for proteomics with unlicensed runs then the licensing page will open after **Import Data section**.

If you already have a programmed dongle attached to your machine then the License Runs page will not appear.

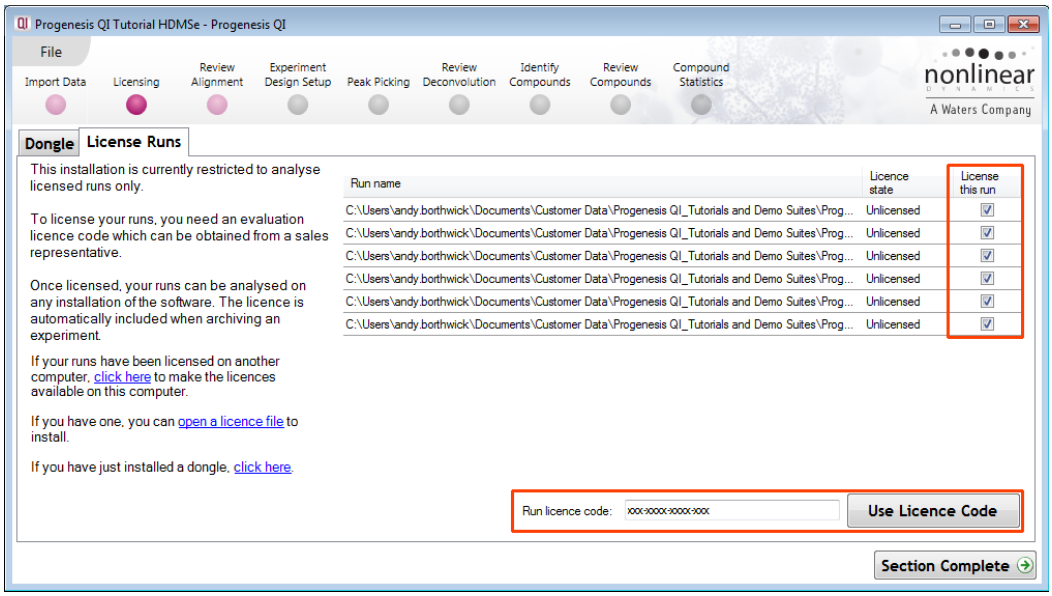
To use this page to License your Runs you must first either obtain an **'Evaluation' Licence Code from a Sales Person or purchase a licence code directly**.

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

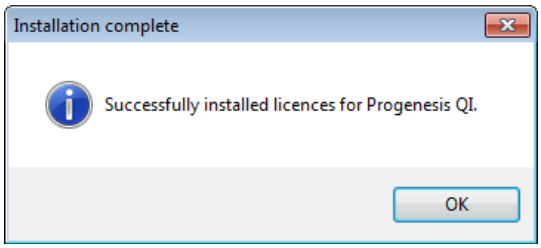
The runs in your experiment will be listed as shown below.

To activate license(s) for the selected runs enter the code in the space provided and click **Use Licence code**.

Note: you will need an internet connection to use this method.



A message confirming successful installation of your licences will appear.



Click **OK**, the view will update and Alignment, the next stage in the workflow, will open with the licensed files.

Appendix 3: Manual assistance of Alignment

Approach to alignment

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.

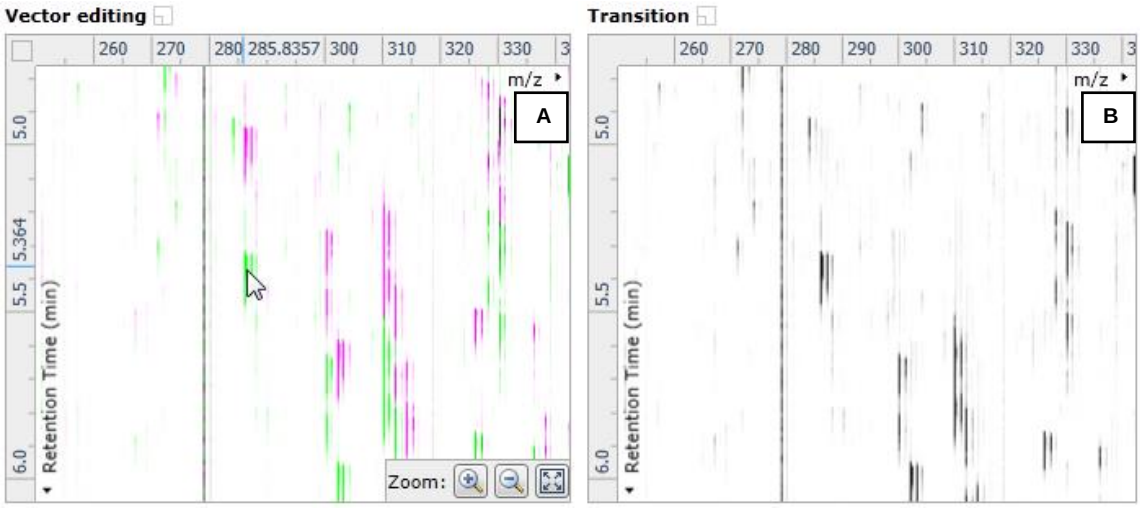
Note: when you click anywhere in the Ion Intensity Map this will **reposition the focus** and update the other views accordingly.

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

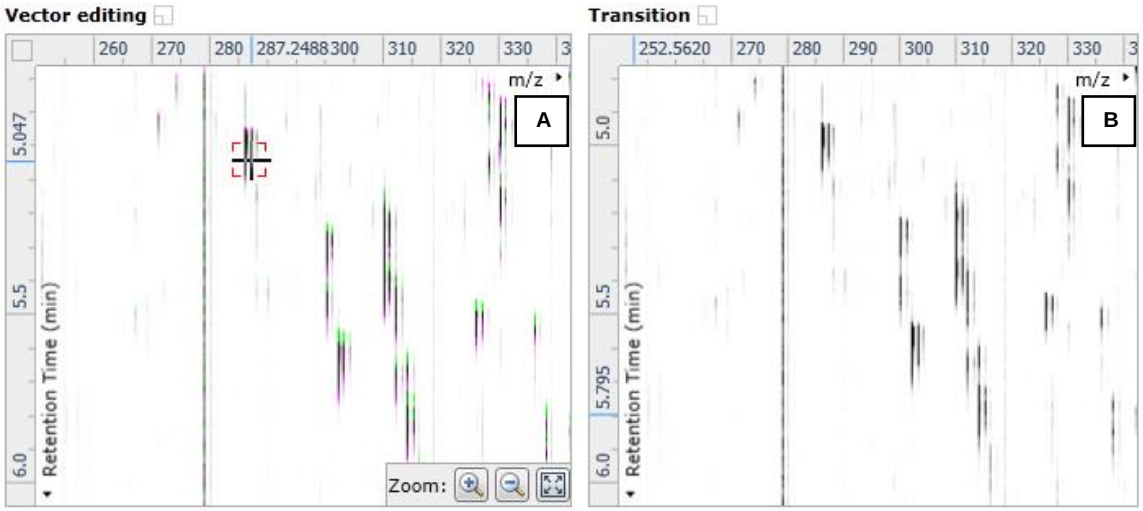
Note: the Ion Intensity Map gives you a colour metric, visually scoring of the current alignment and an overall score is placed next to the Vectors column in the table. With each additional vector this score will update to reflect the overall quality of the alignment. The colour coding on the Ion intensity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

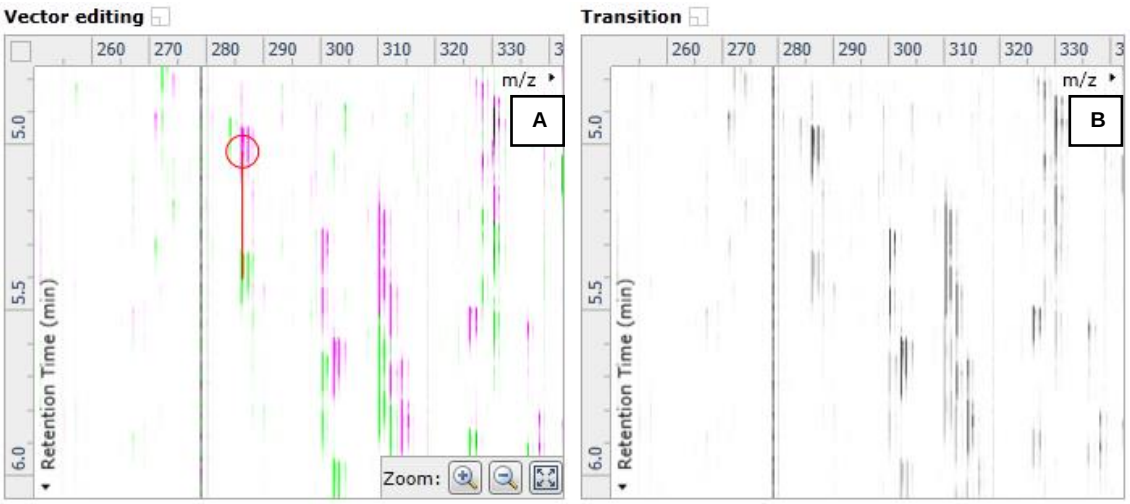


5. As you are holding down the left mouse button (depending on the severity of the misalignment), the alignment vector will automatically find the correct lock. If not, drag the green feature over the corresponding magenta feature of the reference run. The red box will appear as shown below indicating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



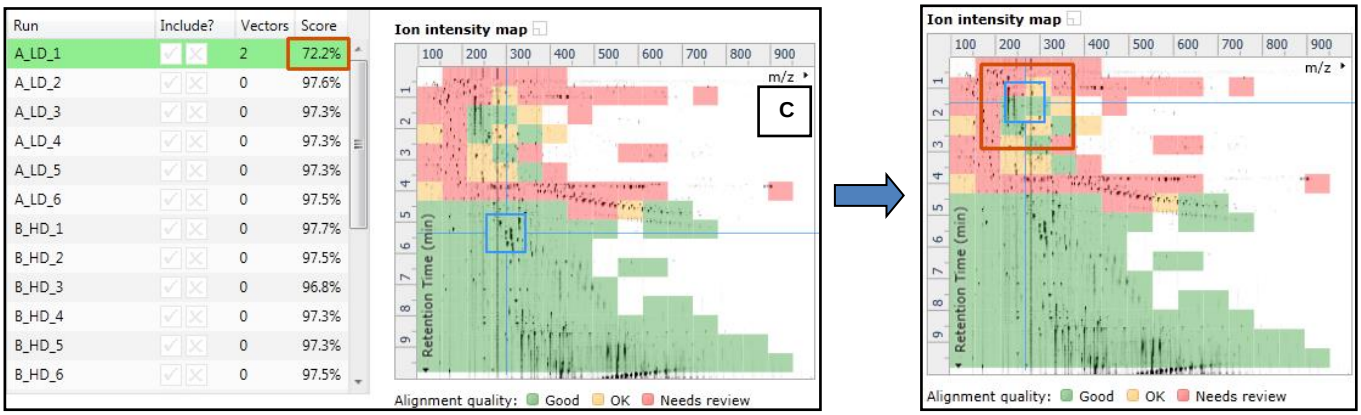
Tip: while holding down the mouse button hold down the Alt button. This will allow smooth movement of the cursor as the Alt key allows you to override the 'automatic alignment' performed as you depress the mouse button.

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.

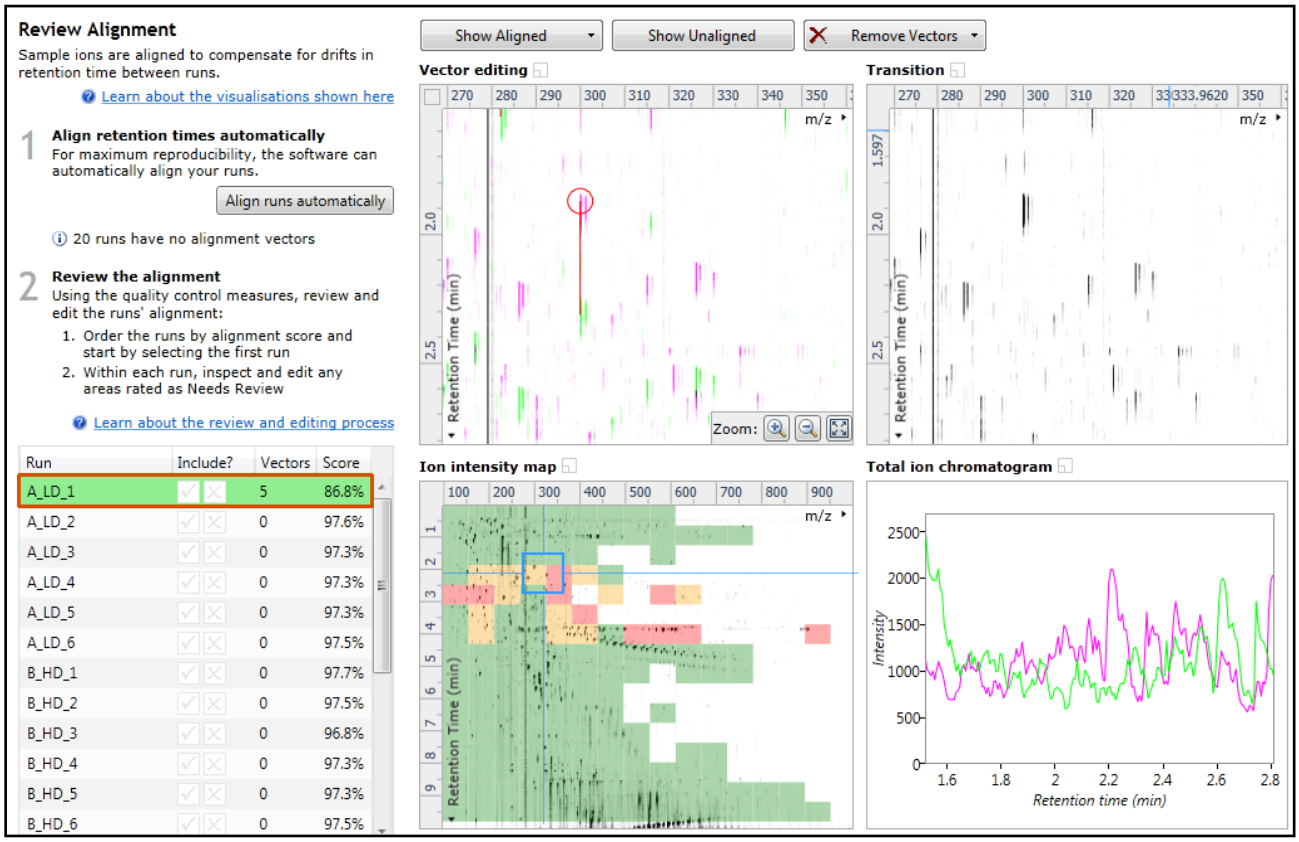


8. Adding an additional vector will improve the alignment further as shown below.

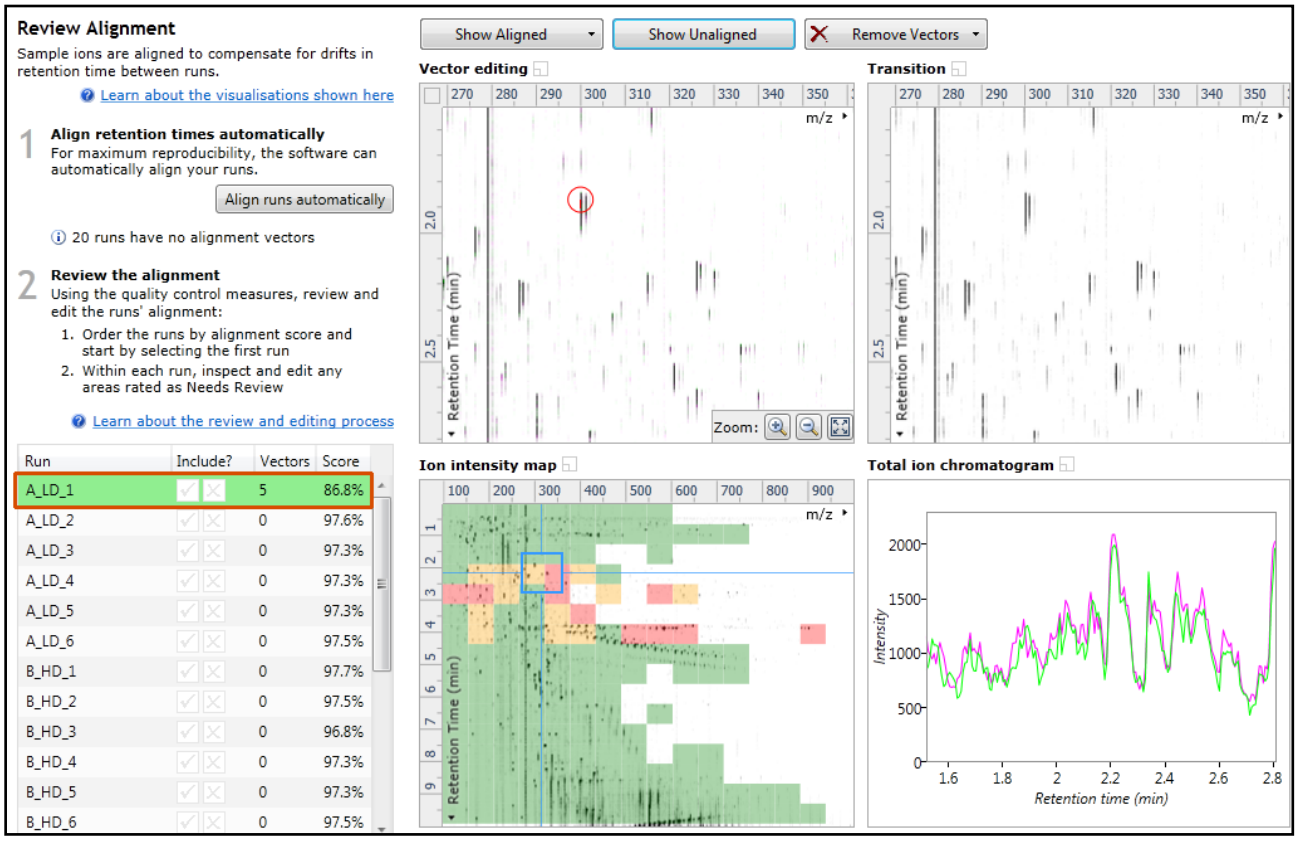
9. Repeat this process moving the focus from top to bottom on the **Ion Intensity Map** view the number you add is dependent on the length of gradient used and the severity of the misalignment. In many cases a single vector per alignment is all that is required to improve the performance of the Automatic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



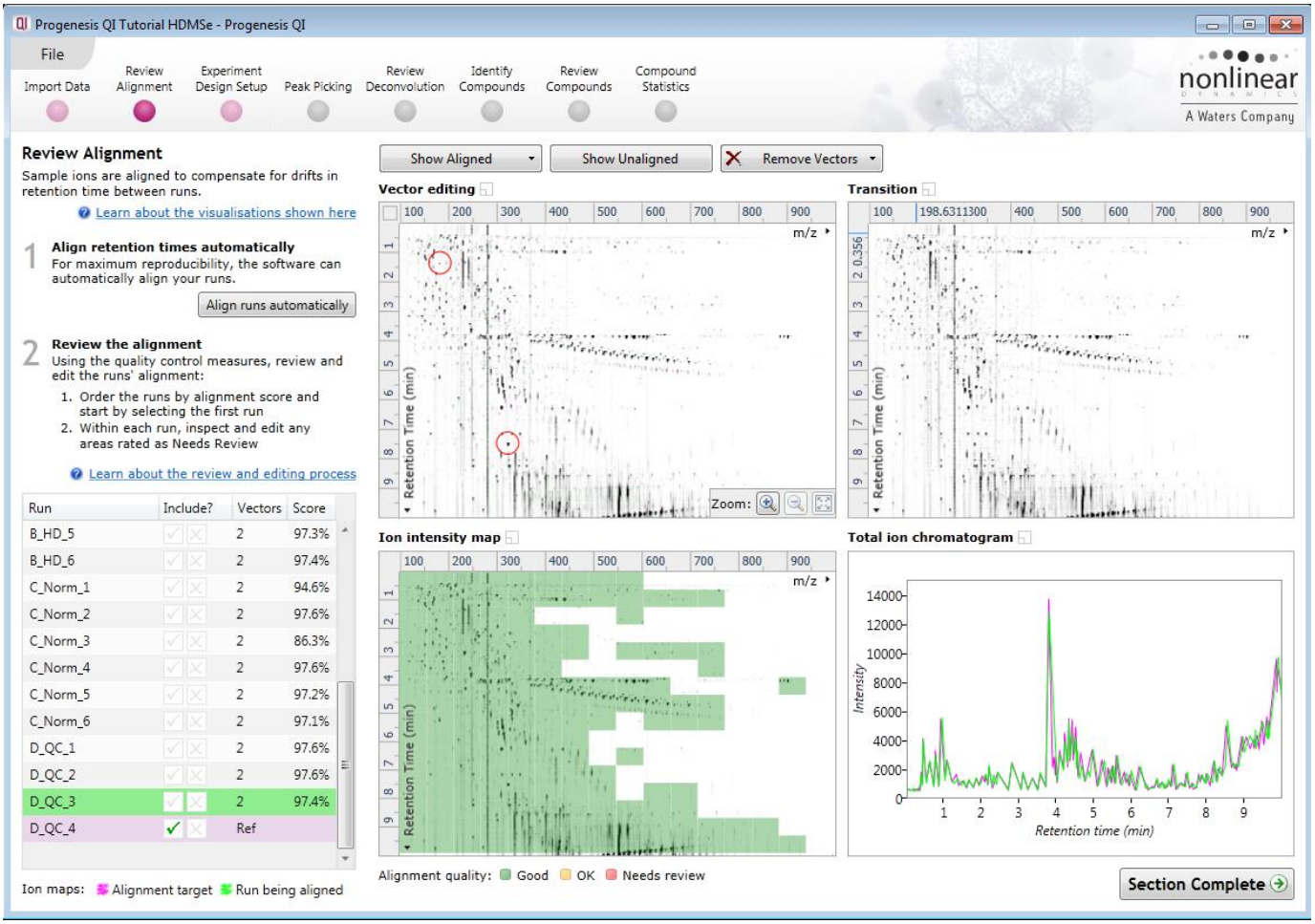
The number of manual vectors that you add at this stage is dependent on the misalignment between the current run and the Reference run. The effect of adding the manual vectors can be seen when you press show aligned



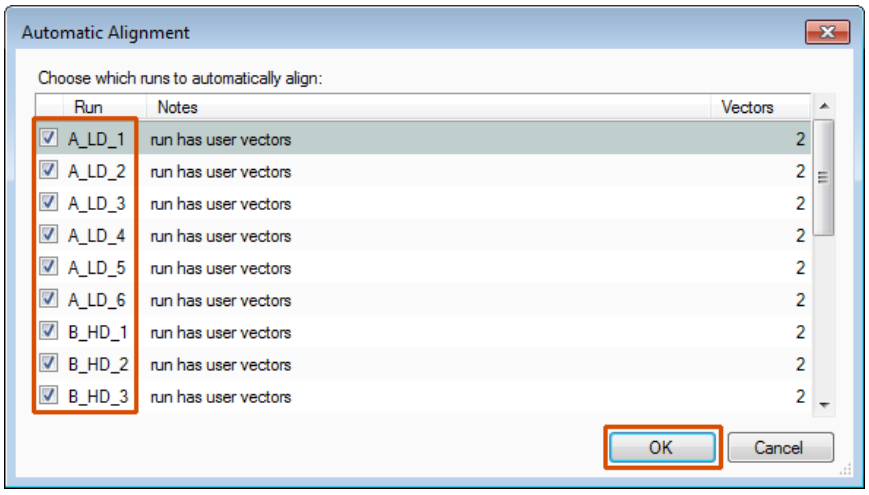
10. Repeat this process for all the runs to be aligned.

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.



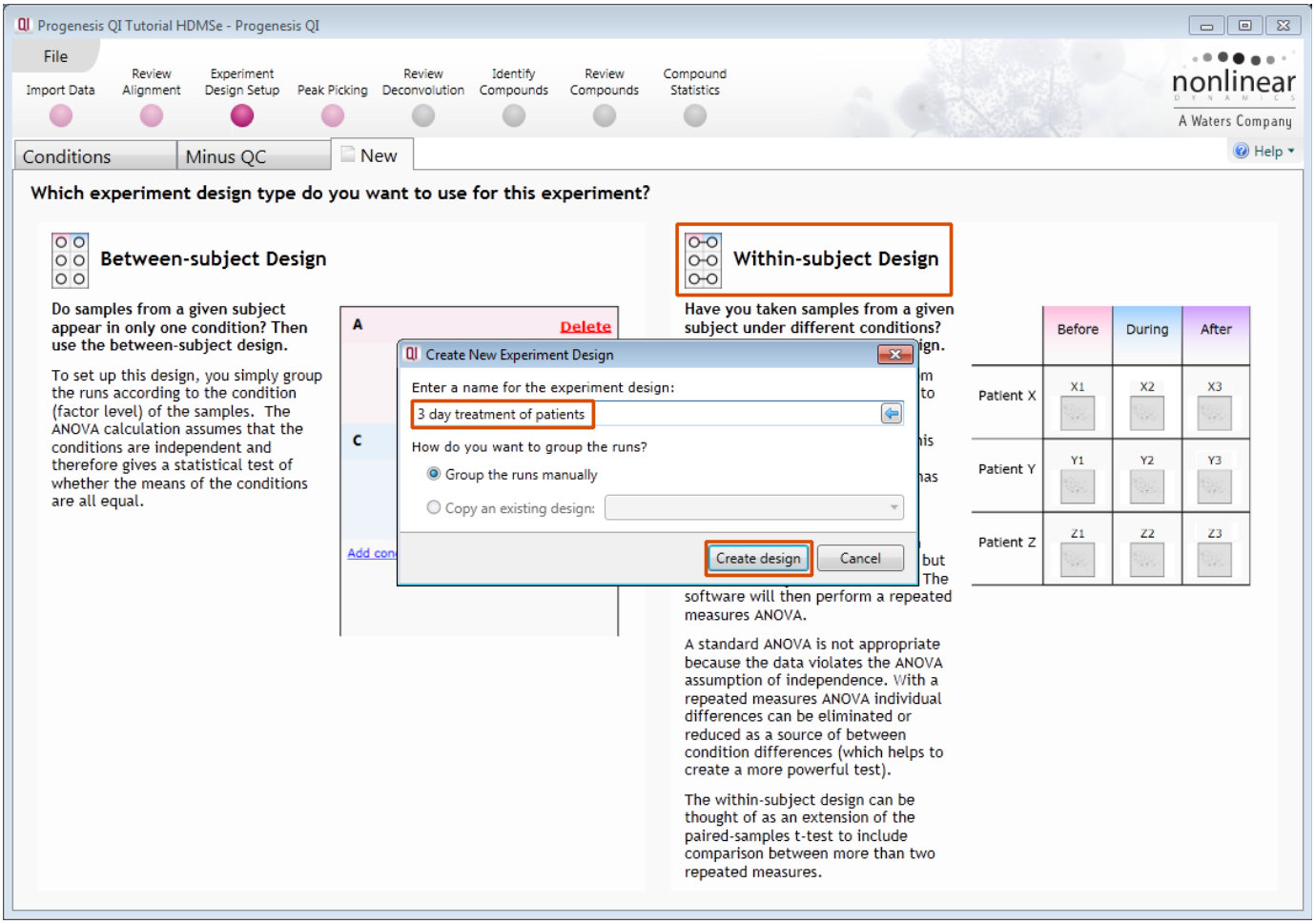
Note: the tick box next to the **Run** name controls whether automatic vectors will be generated for each run.

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 4: Within-subject Design

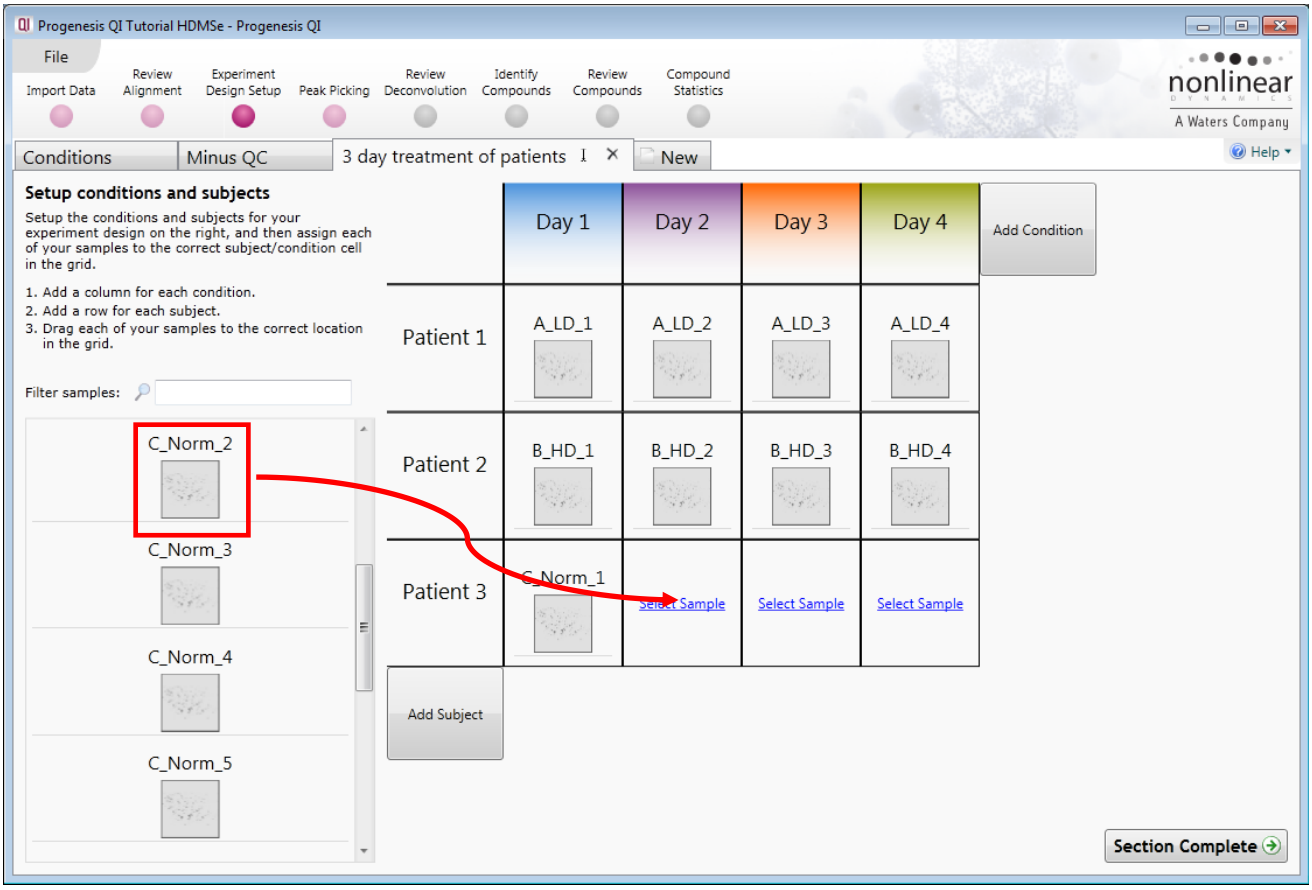
To create a **Within-subject Design** for your data set select this option on the **Experiment Design Setup** page and enter the name of the design.

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.

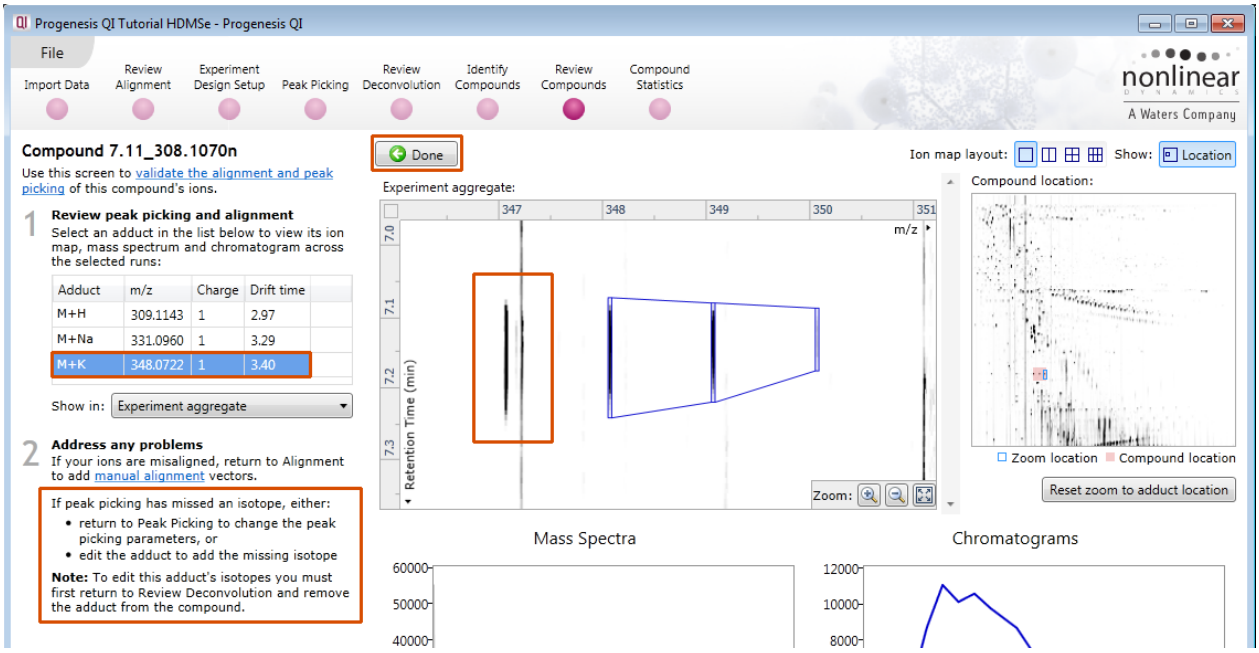


You can create additional Experimental Designs using the New tab

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

Appendix 5: Editing an adduct to add a missing isotope

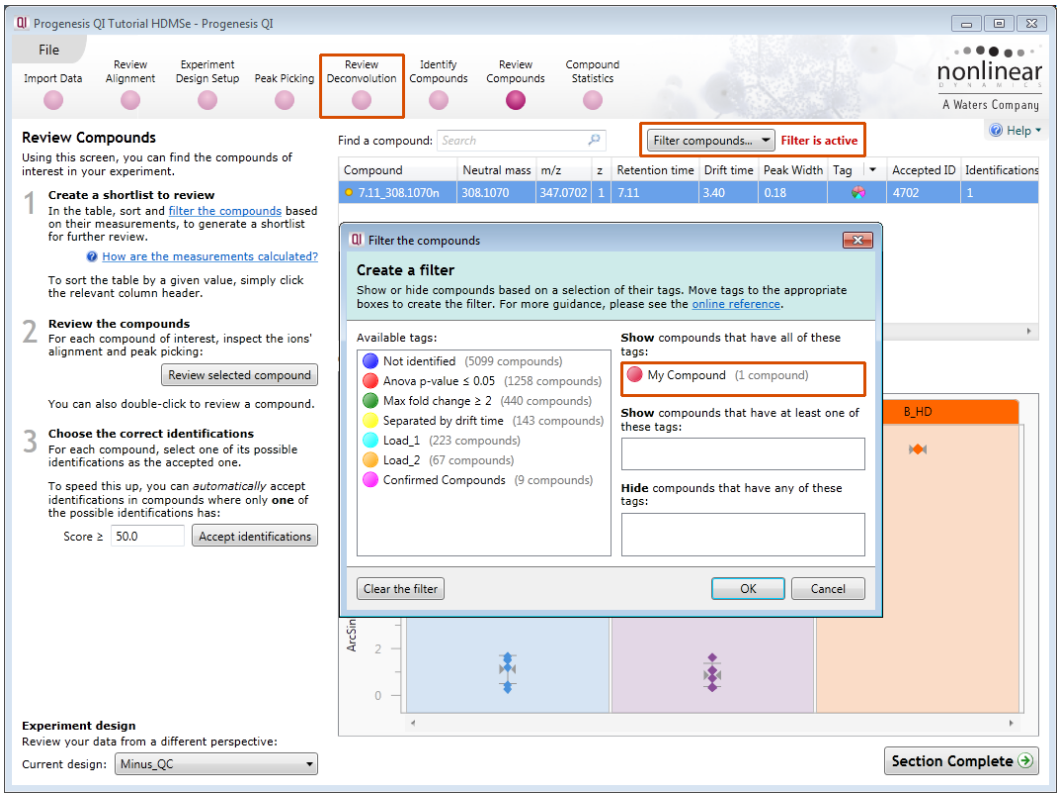
When using the Compound Validation view you find an example where the Peak picking has missed an isotope for an adduct ion as shown below. As shown below the monoisotopic peak for the M+K has not been detected.



For the purposes of explaining the processes involved in editing this adduct we have (on purpose) deconstructed the M+K adduct detection.

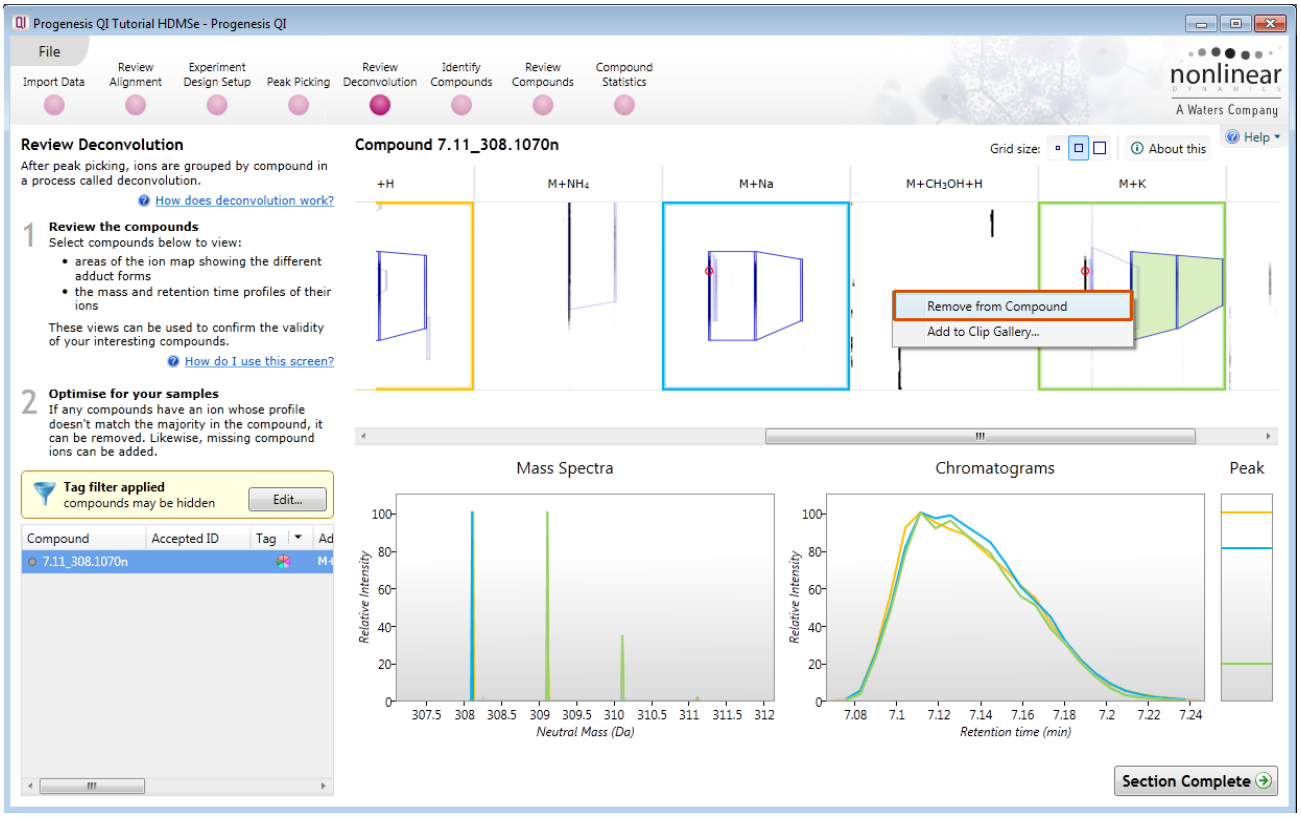
As the note on the screen suggests you must first return to the **Review Deconvolution** screen and remove the adduct from the compound.

Before doing this, Click **Done** and create a tag for the compound you are about to change. Then filter so that only this compound is displayed at the **Review Compounds** stage.

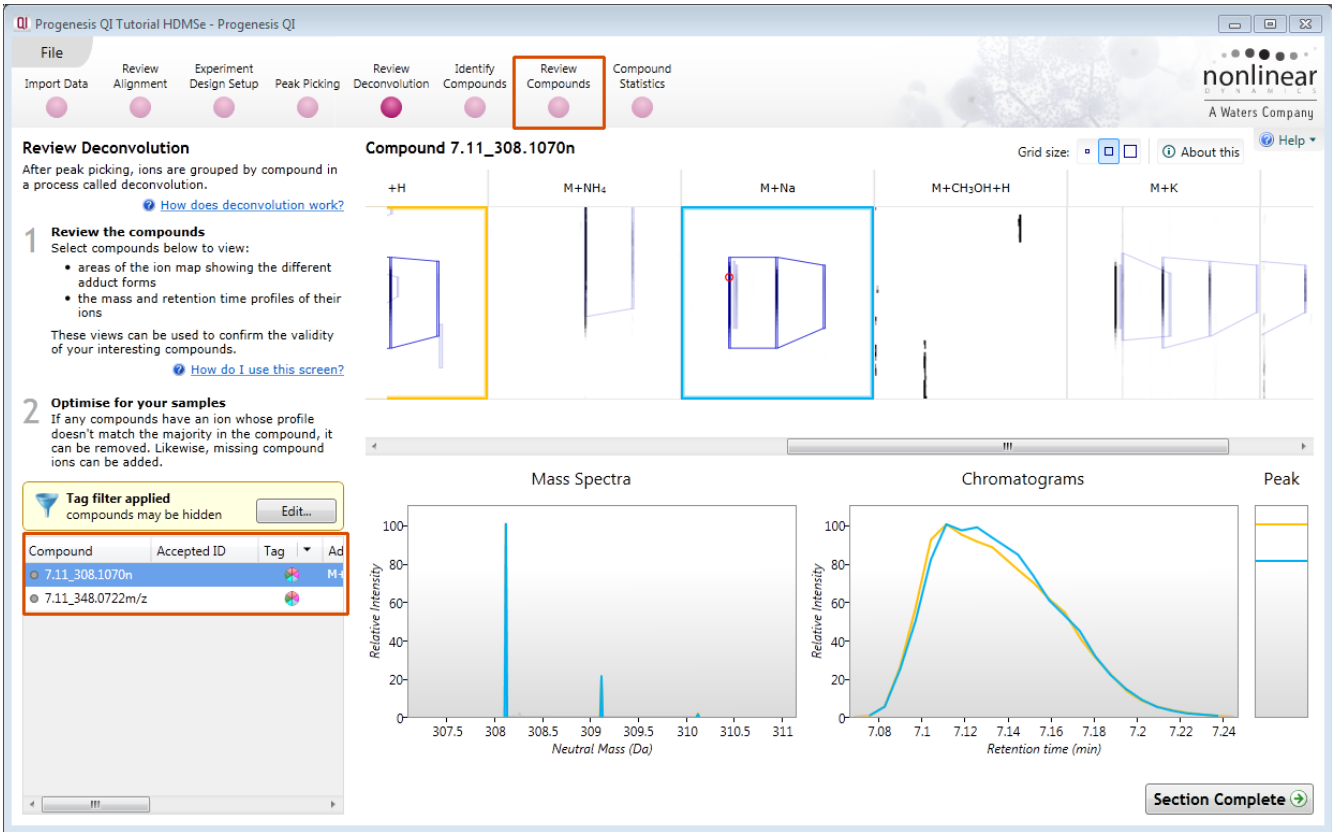


Now click **Review Deconvolution** on the workflow

In this view the 3 Adducts are displayed. Right click on the M+K adduct and select **Remove from Compound**.

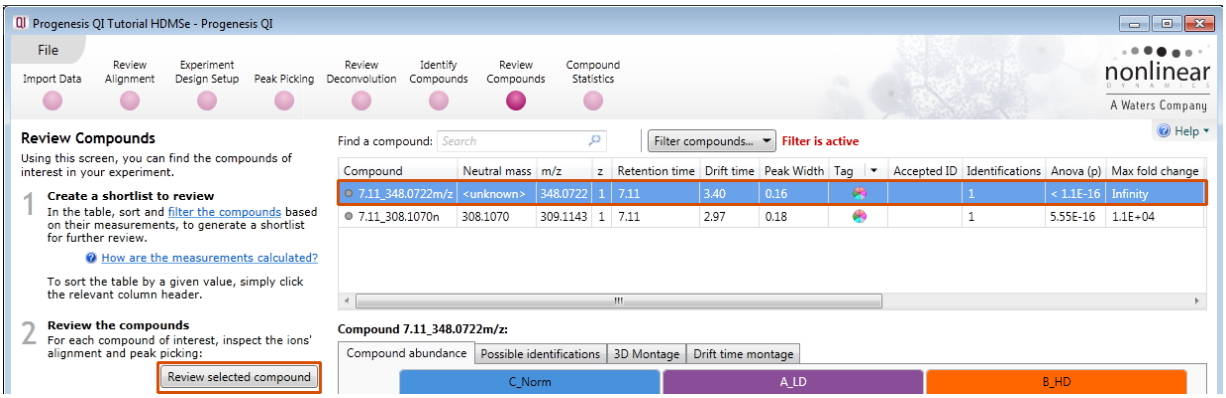


The table will update to show 2 compounds (both tagged). The second one is based on the single adduct M+K.

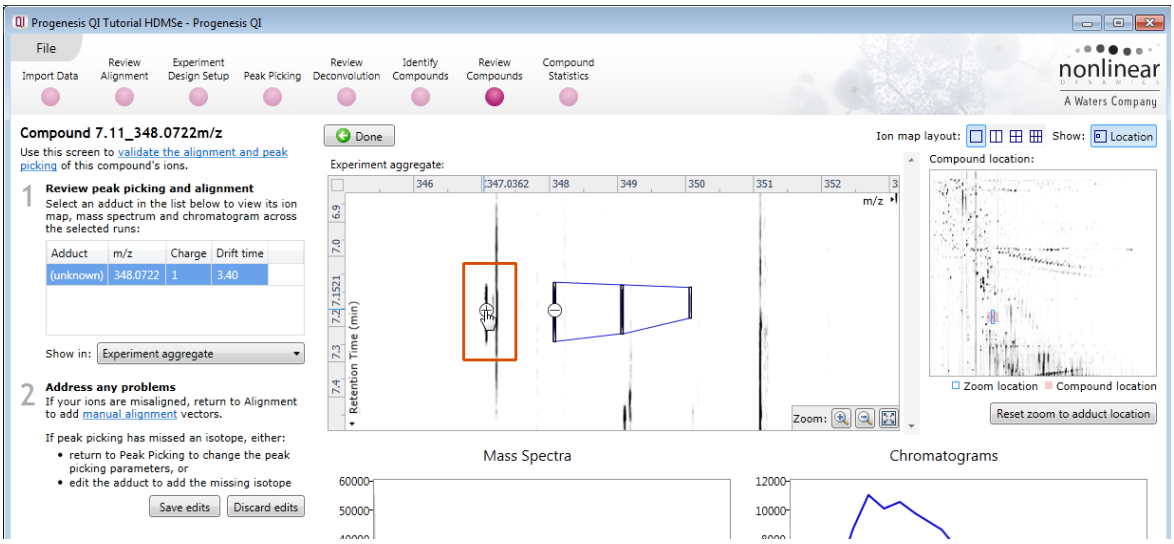


Now click on **Review Compounds.**

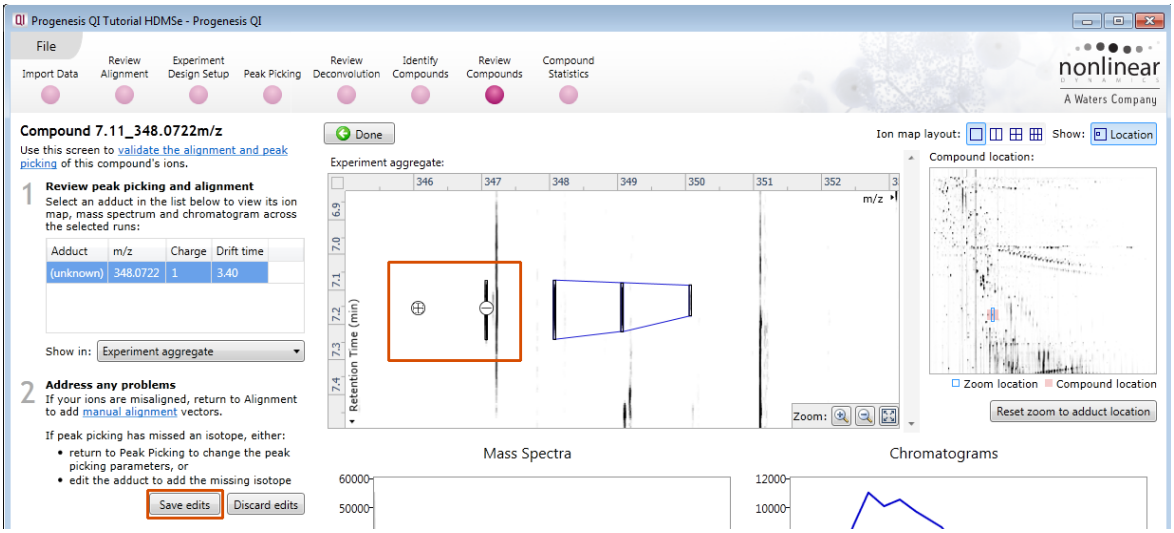
Select the compound with the ‘unknown’ Neutral Mass and click **Review selected compound**



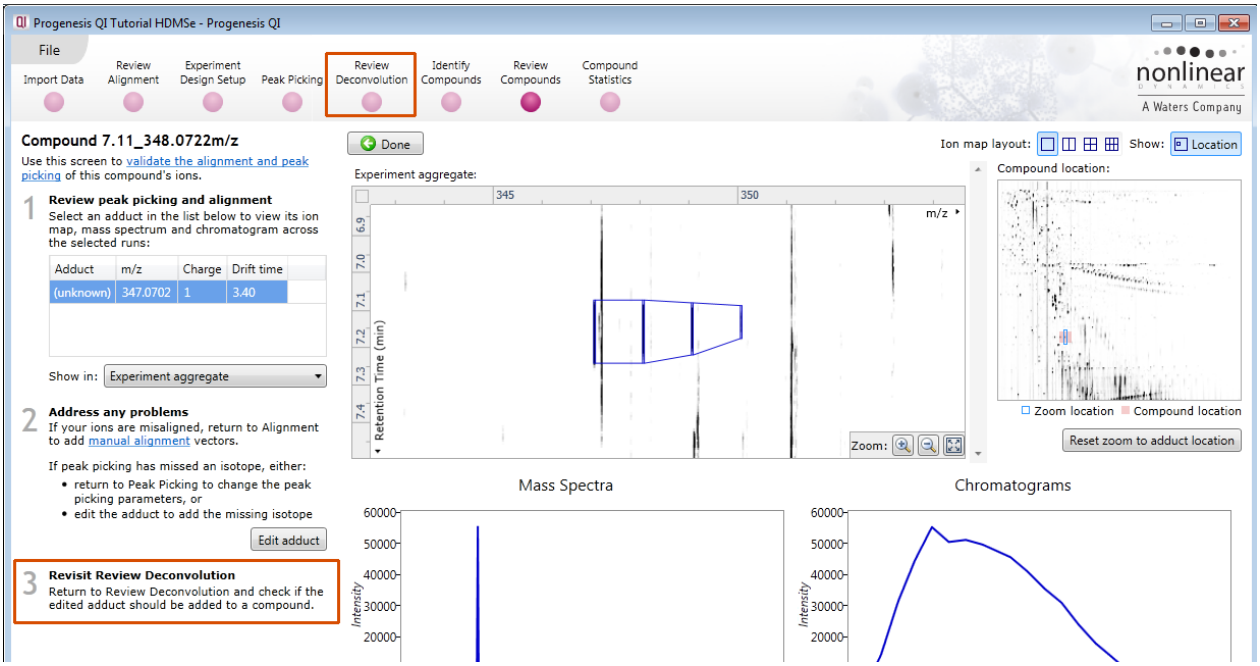
Now click on **Edit Adduct**. A **plus and minus** will appear on the experiment aggregate view. Click on the **plus** to add the missing isotope.



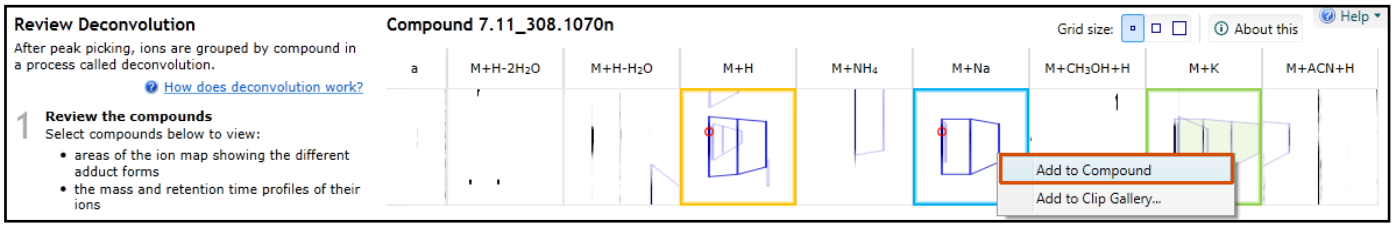
The view updates to show the minus sign now on the added isotope. Click on **Save edits** to accept the edit



The view will update to show the added isotope and a note will appear indicating that you should return to Review Deconvolution to check if the edited adduct should be added to a compound . Click on **Review Deconvolution** on the workflow.



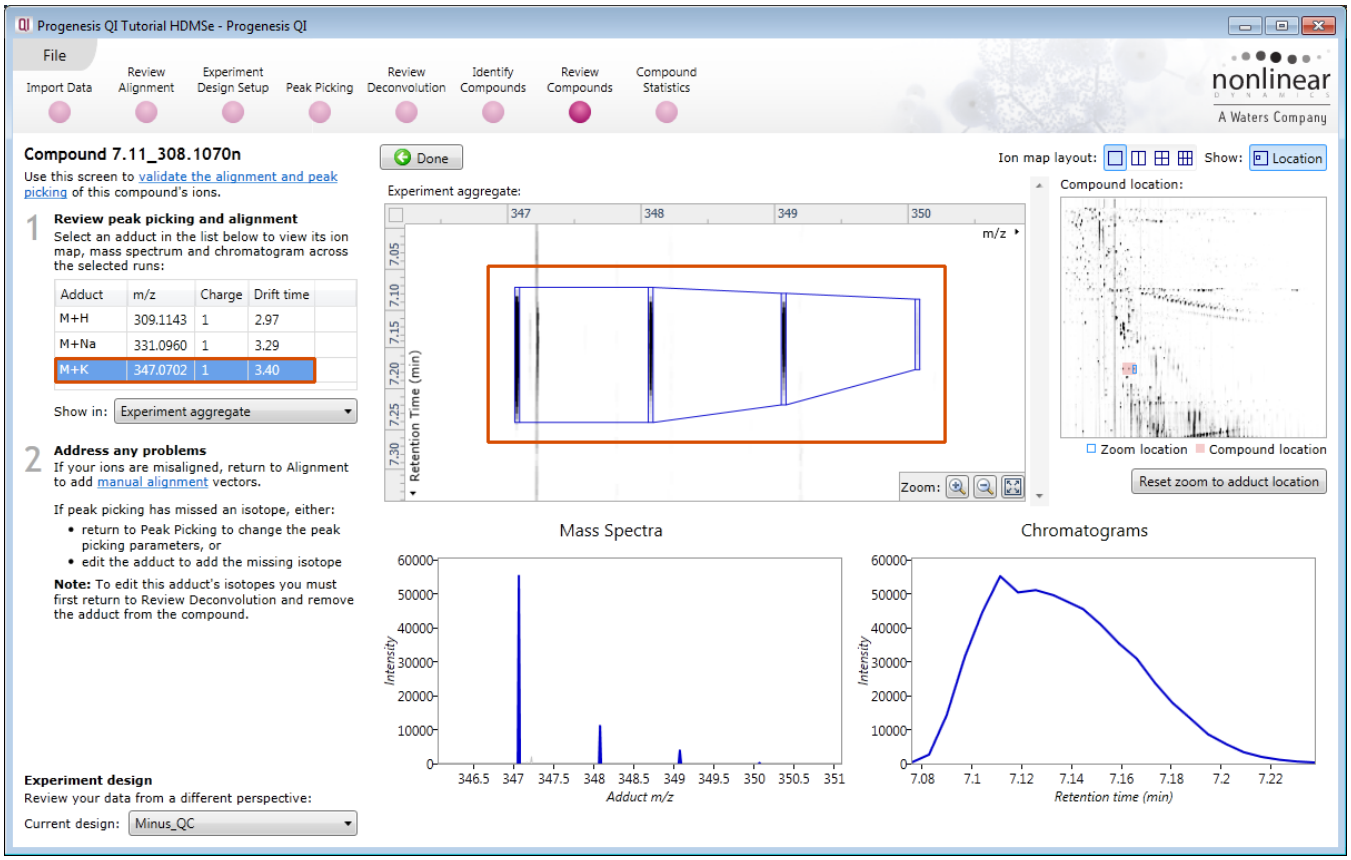
Right click on the edited adduct and select **Add to Compound**



The compound and its details will update to and the compound table will now show a single compound.

Now click on **Review Compounds** and click **Review selected compound** with the edited compound selected.

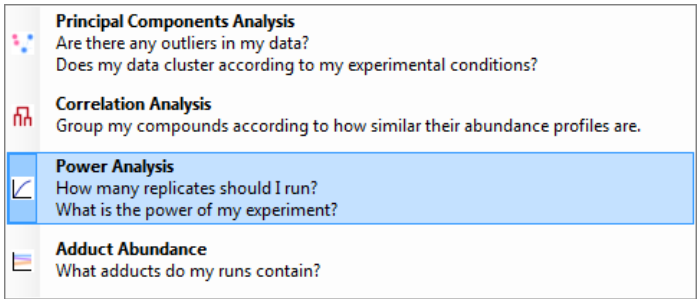
When you now click on the M+K adduct it contains 4 isotopes and its m/z has altered accordingly to reflect the addition of the isotope.



Appendix 6: 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.



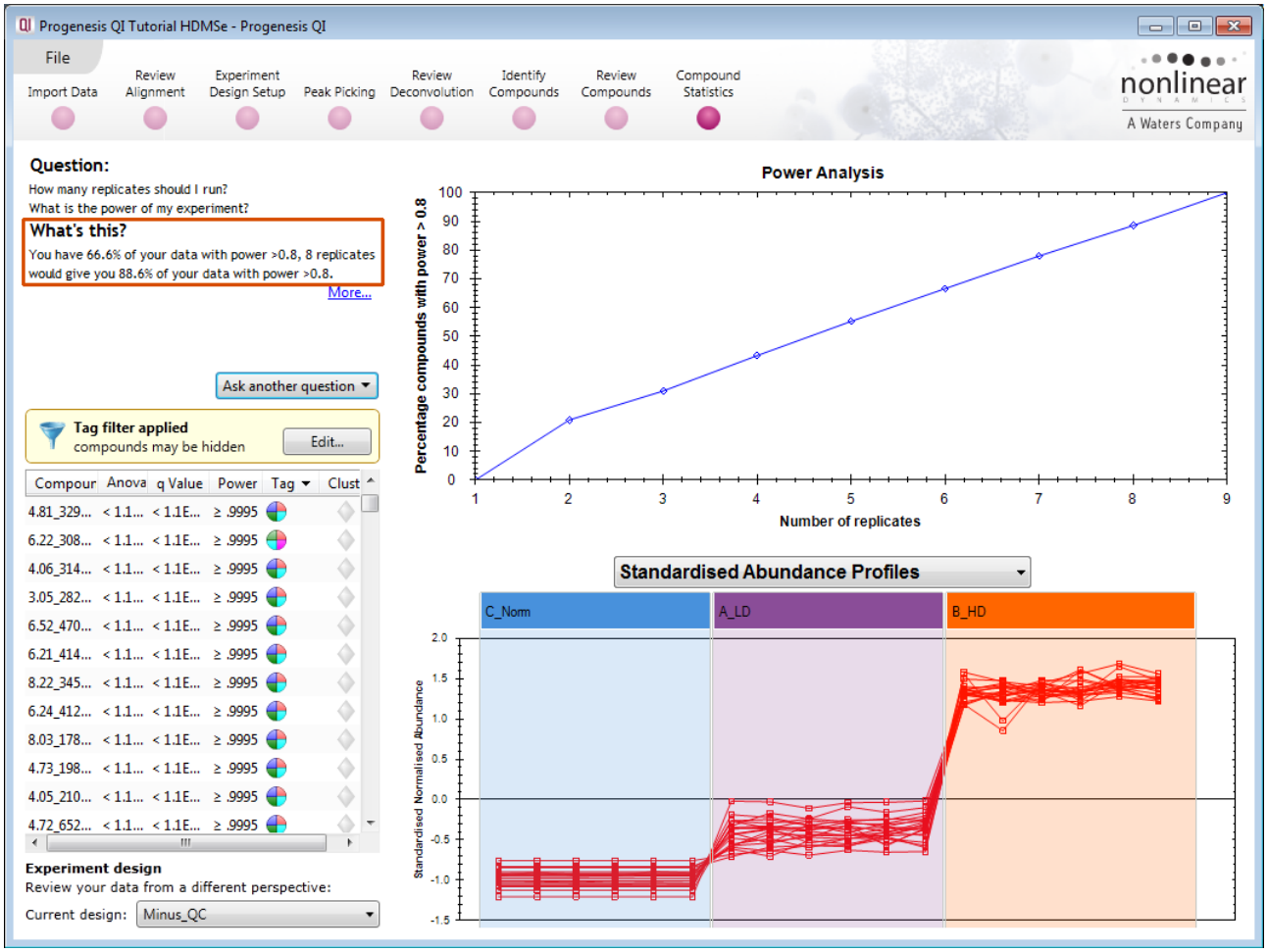
Select the option

'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 the **Significant p<0.05 compounds (1258),** as an example, view the power analysis.



This is displayed graphically showing that 66.6% of the 1258 compounds have a power of 80% or that 8 replicates would give you 88.6% of your data with power > 0.8.

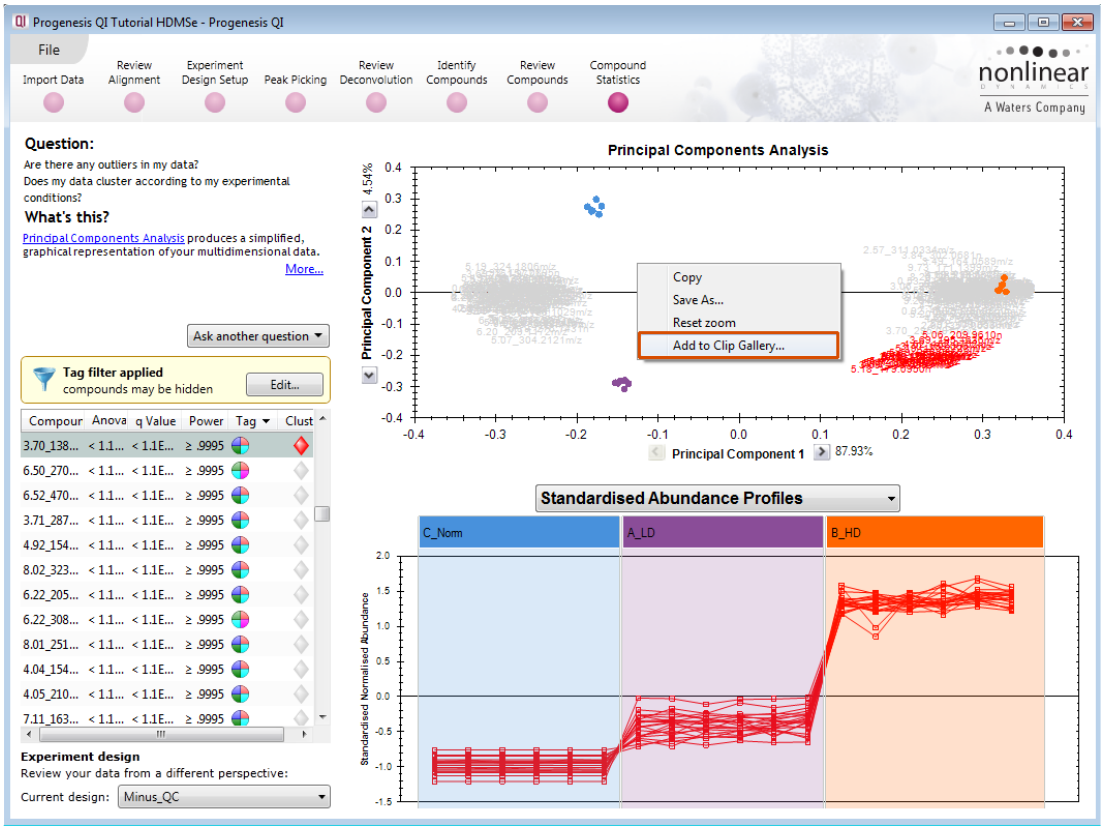
Appendix 7: Using Clip Gallery to Save and Export Pictures and Data

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

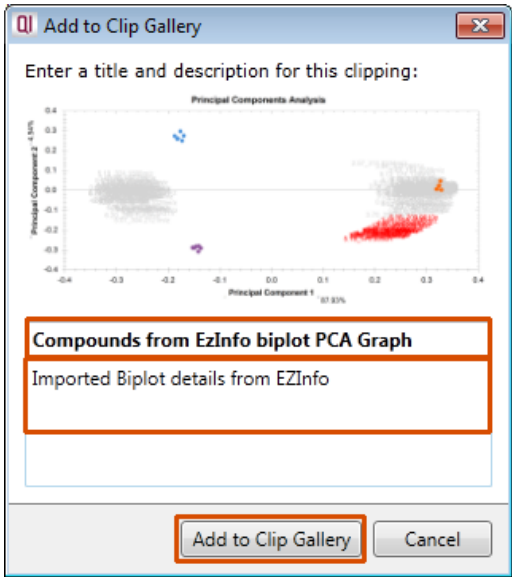
The saved images are retained as part of the experiment and are stored accordingly. This facility allows you to capture (high resolution) images that can be used in the development of specific reports and/or used as part of the process of publishing your experimental findings.

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….

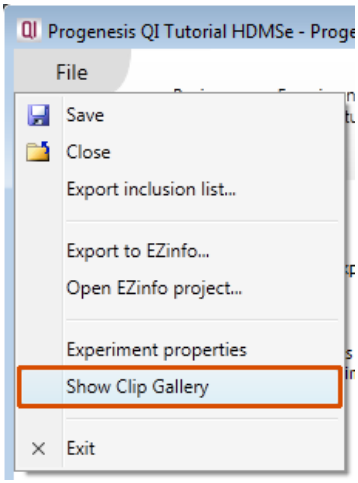


This will open a dialog displaying what is to be saved and allows you to alter the title and provides a description of the image for later reference.

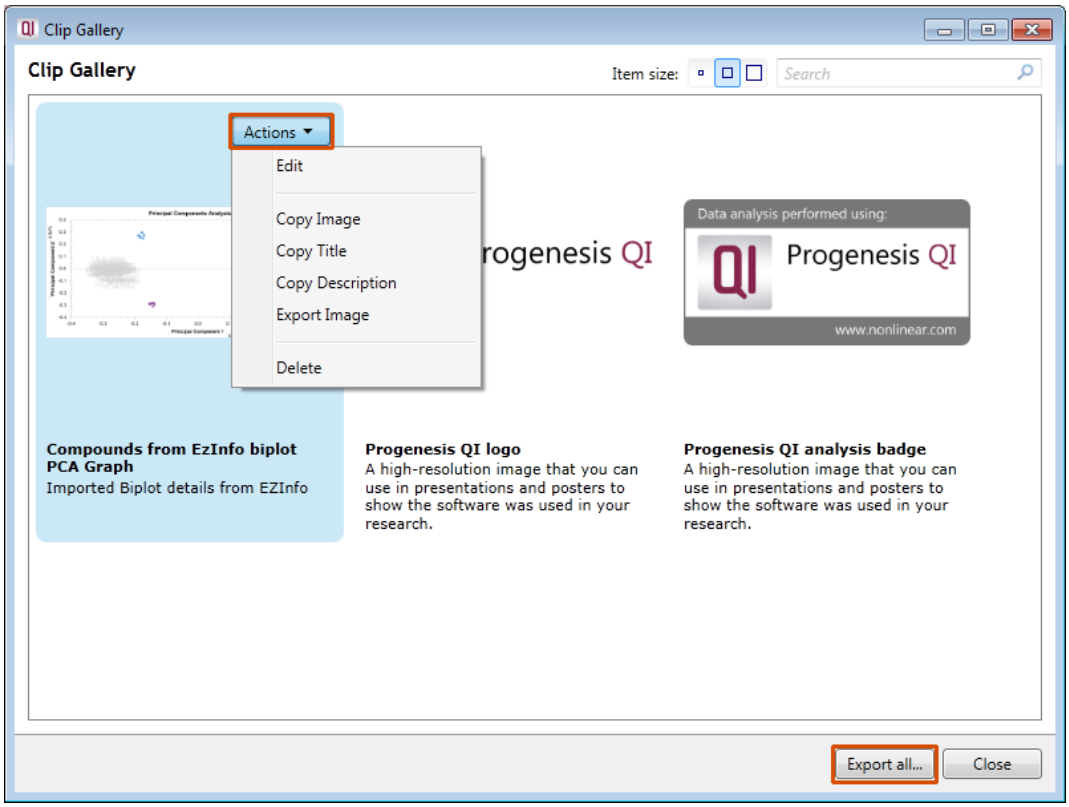


Enter details as required and click **Add to clip gallery**

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.

Appendix 8: Waters Machine Specifications and approx. timings

This appendix provides information on the time(s) taken at each stage and the total time taken to analyse a set of 22 (Phase 1) HDMSe runs on a Waters Demo Spec PC.

Machine Spec: Lenovo

Processor: Intel® Xeon® CPU 2.66GHz 12core X5650 @ 2.67GHz

RAM: 24.0 GB

System Type: 64-bit Operating System

File Folder Size: Each file folder (.RAW): 0.75 Gig

Analysis Stages:		**Per file**	**Total**	
Import Data:	Loading of Raw data per file	1min	22min	Total for 22 files
	(re-opening at Import Data)	5s		
Alignment:	Automatic alignment of data		6.3min	
	(re-opening at Alignment)		10s	
Peak Detection:	Automatic Detection of data		4.5min	
	(re-opening at Peak Detection)		10s	
Identify Compounds:	Performing .SDF Compound Search		5s	
	(re-opening at Identify Compounds)		10s	
Total Analysis Time Including Loading			35min	
Restoring Progenesis QI tutorial HDMSe experiment from archive			20s	