Progenesis CoMet User Guide

Analysis workflow guidelines

for version 2.0

Contents

Introduction

This user guide takes you through a complete analysis of 25 LC-MS runs of metabolite containing samples for 4 biological groups and a Pool group (5 replicate runs per group) using the unique Progenesis CoMet 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 Progenesis 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 CoMet workflow, therefore if you are using your own data files please refer to Appendix 1 (page 45) 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. If you experience any problems or require assistance, please contact us at support@nonlinear.com

How can I analyse my own runs using CoMet?

You can freely explore the quality of your data using Data Import and then licence your own runs using this evaluation copy of Progenesis CoMet. Instructions on how to do this are included in a section at the end of the user guide document. Alternatively if you would like to arrange a demonstration in your own laboratory contact support@nonlinear.com and we will help you.

LC-MS Data used in this user guide

NLD would like to thank Dr Malin Olson at Institut für Veterinärpharmakologie und –toxikologie, Zurich Switzerland 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

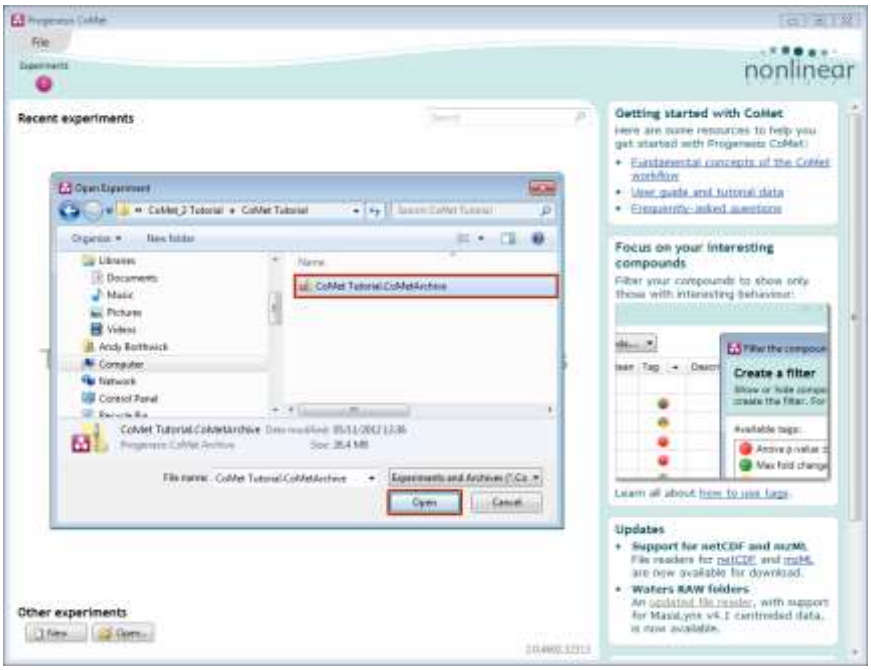
Progenesis CoMet adopts an intuitive **Workflow** approach to performing comparative LC-MS or GC-MS data analysis. The following user guide describes the various stages of this workflow (see below) focusing mainly on the stages from Alignment to Compound Statistics.

Restoring or starting a new CoMet Tutorial

If working with your original data files then refer to Appendix 1 page 45

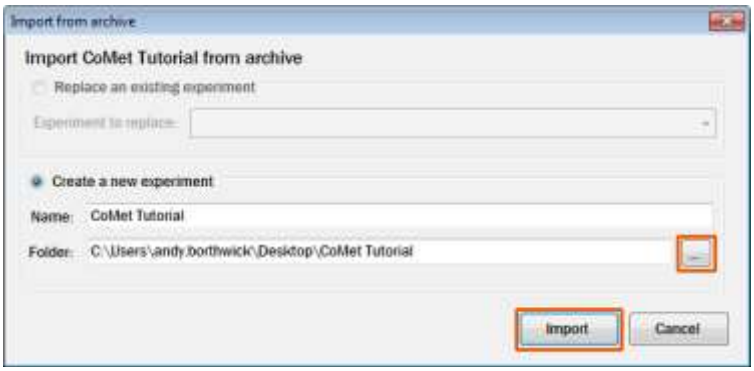
Open Progenesis CoMet and download the Compressed (.zip) Tutorial Archive file from the **'User guide and tutorial data**' link shown below, placing it in a **new folder** on your desktop. Before restoring the tutorial in the software **you must** first right click on the downloaded (.zip) file and extract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

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

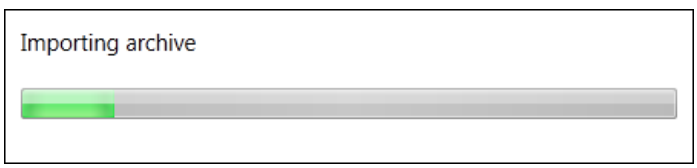


This opens the 'Import from archive' dialog.

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

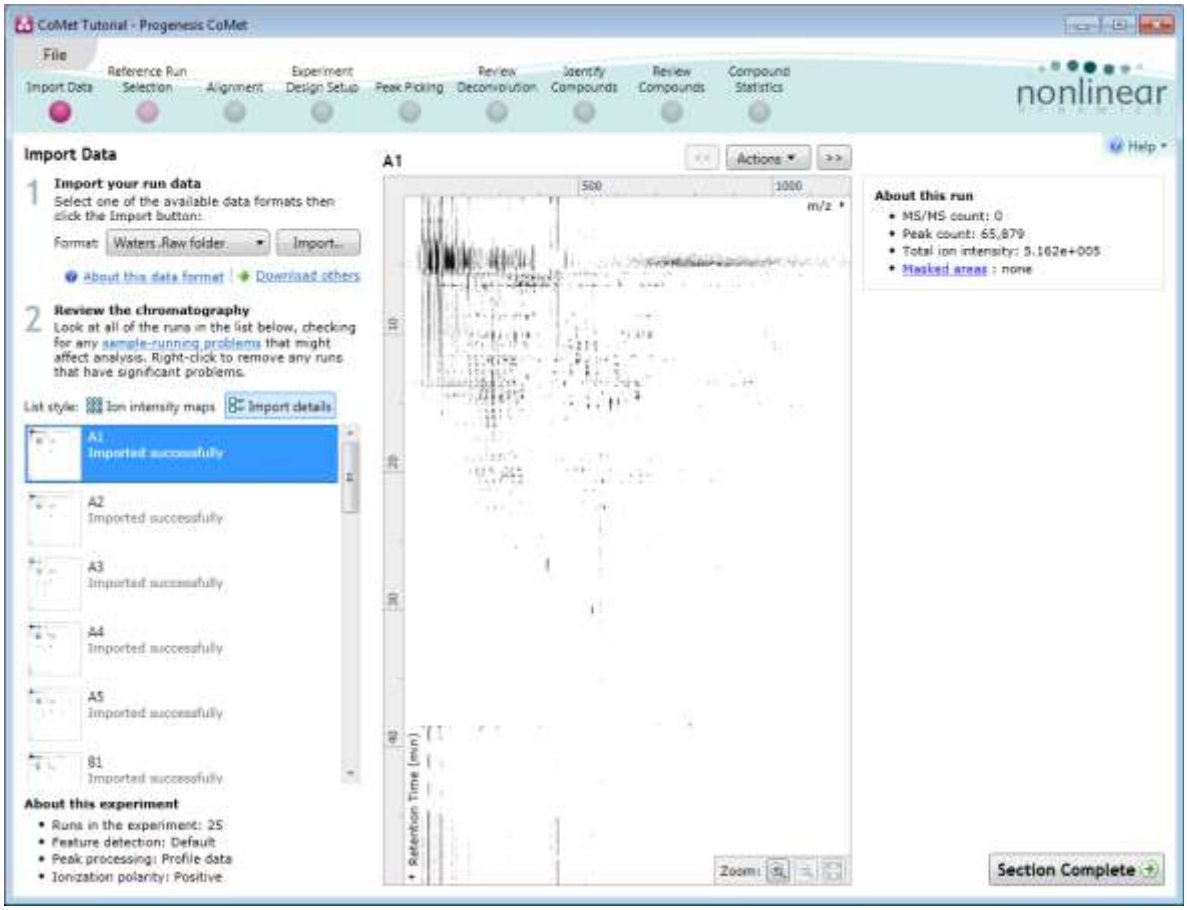


Then press **Import.**



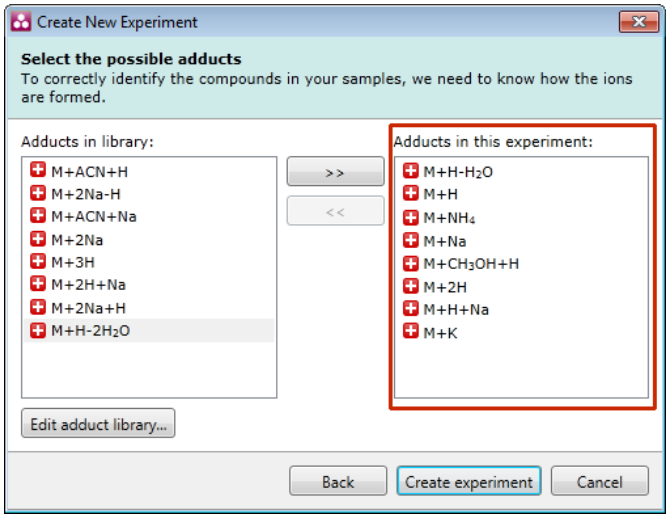
Note: use the **Replace an existing experiment** option if you want to over-write an existing version of the tutorial.

The tutorial will open at the Import Data stage with all the files listed.



The data file format for this tutorial was **Waters.Raw folder** captured from a **High resolution Mass Spectrometer** with the ionisation polarity set to **positive**. (As shown on the bottom left, **About this experiment** on **Import Data**, the first stage of the CoMet workflow.

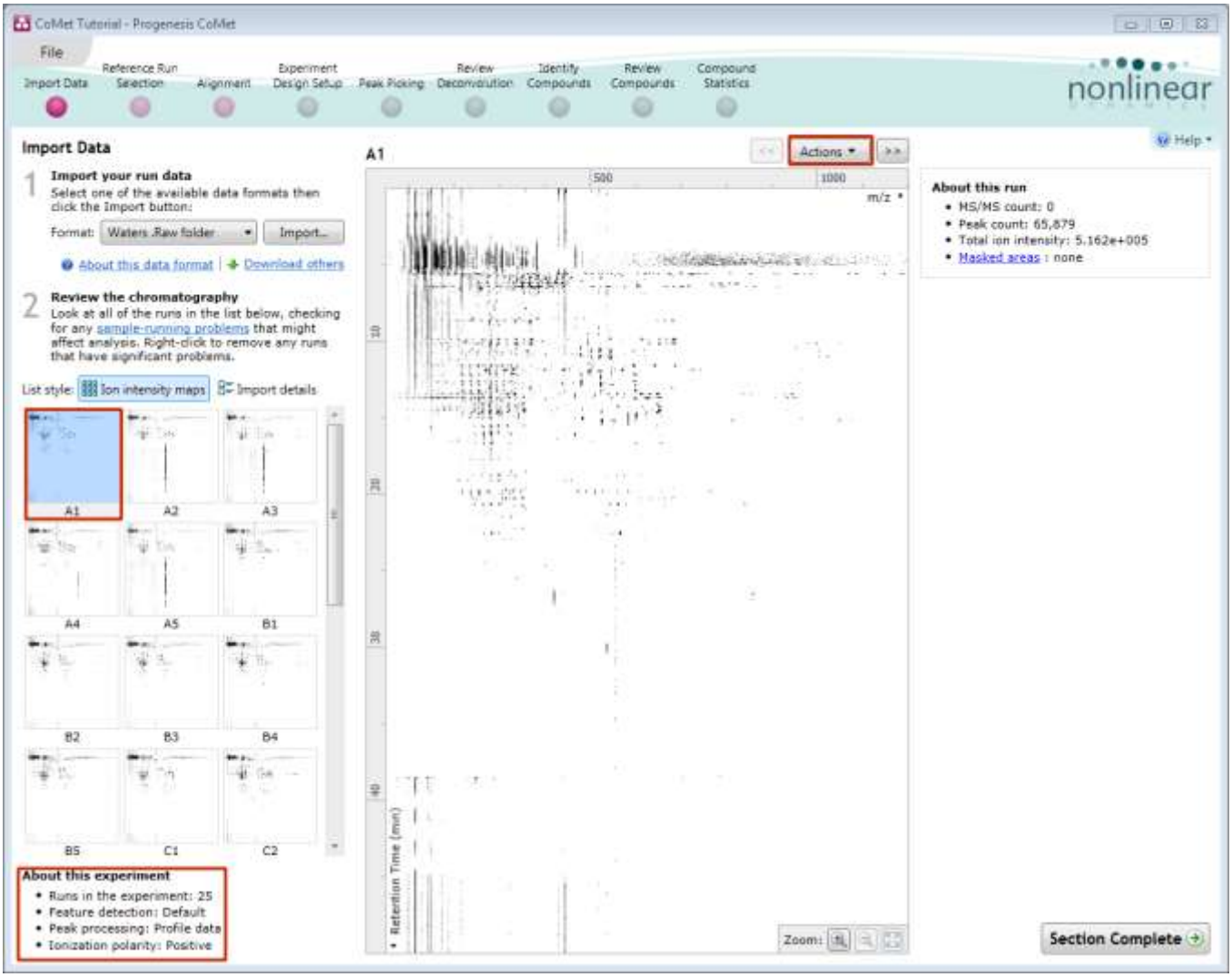
The list of possible adducts (shown below) was created when the experiment was originally created prior to data import. For additional details on adduct selection and setting up a New experiment refer to Appendix 1 (page 45).



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

Stage 1: Data import and QC review of CoMet data set

The CoMet 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. 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 experiment are displayed underneath the run list. This records the selections made when the experiment was created.

Note: the options for **'Peak processing, Feature detection Methods and Ionisation Polarity ',** selected when the experiment was created, are detailed in Appendix 1, page 45).

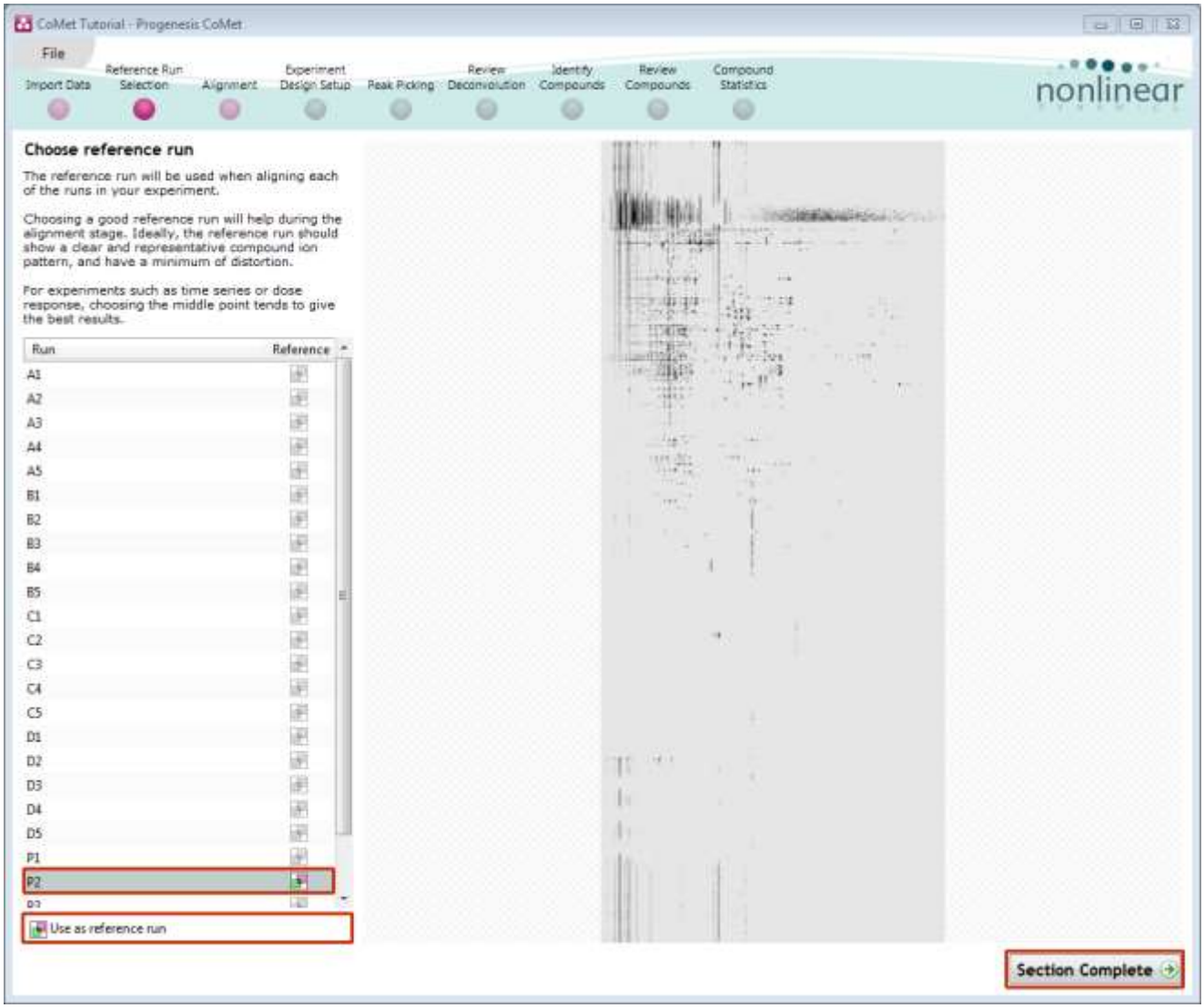
Tip: the **'Mask areas for peak picking'** facility (under **Actions**) allows you to exclude areas (usually early and/or late in the LC dimension (Retention Time) that appear excessively noisy due to capture of data during column regeneration (see Appendix 2, page 50). This is **NOT** used for this 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: Reference Run selection

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

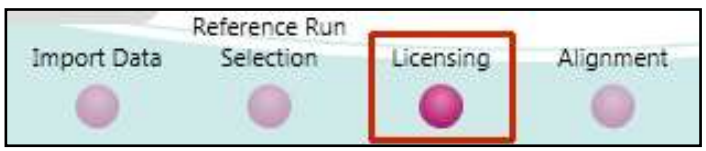


To select a Reference run either click on the run in the list and then click **Use as reference run** or double click on the run in the list. In this example **P2** is selected as the Reference Run

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

Stage 3: Licensing

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

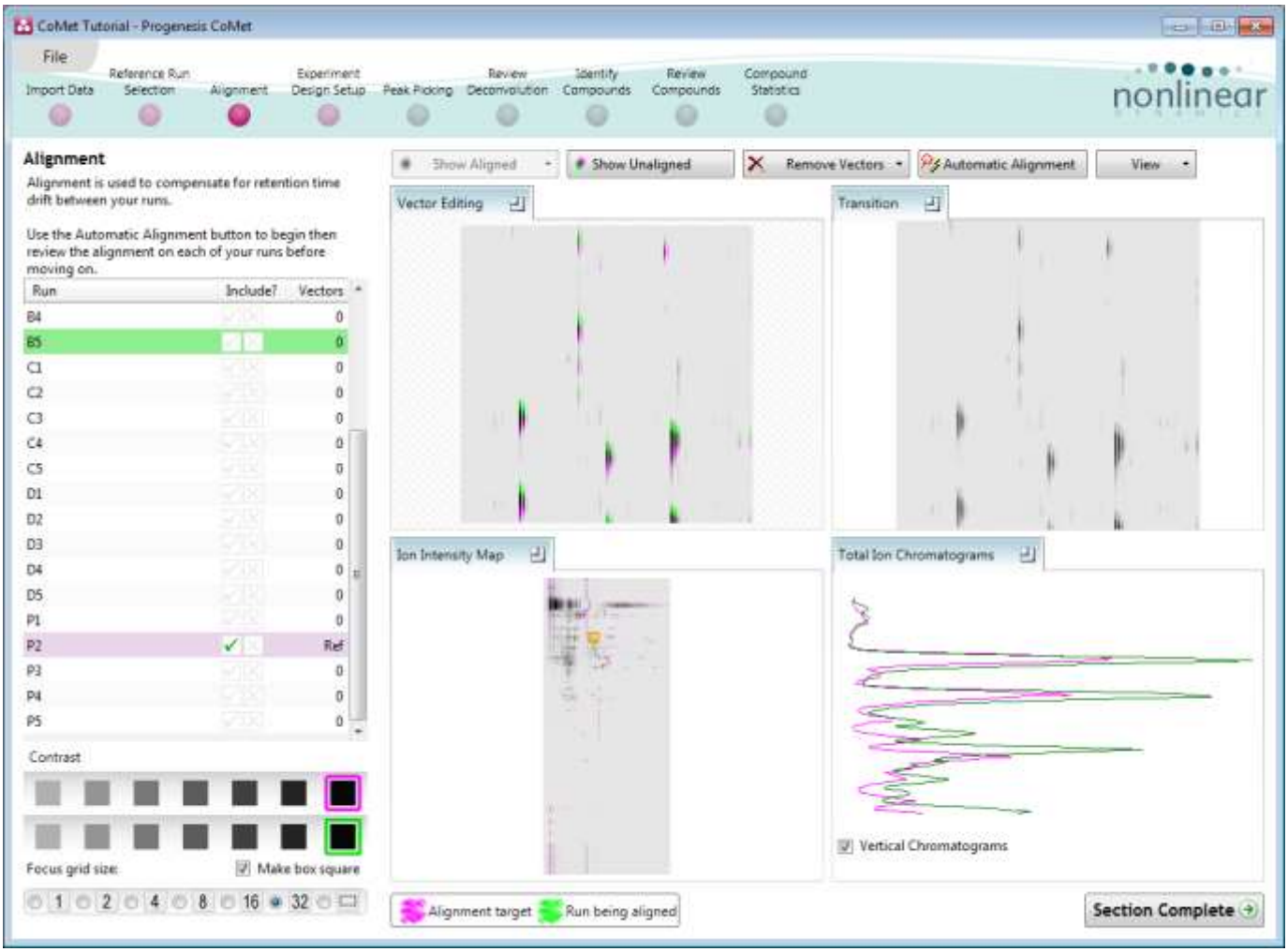


For details on how to use Licensing go to Appendix 3 (page 51)

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

Stage 4: Alignment

At this stage Progenesis CoMet Alignment opens displaying your data.

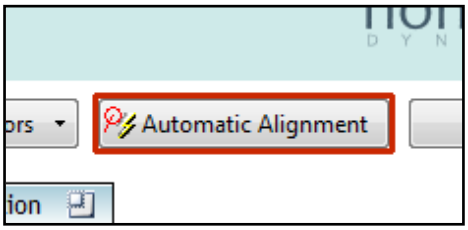


Generation of alignment vectors

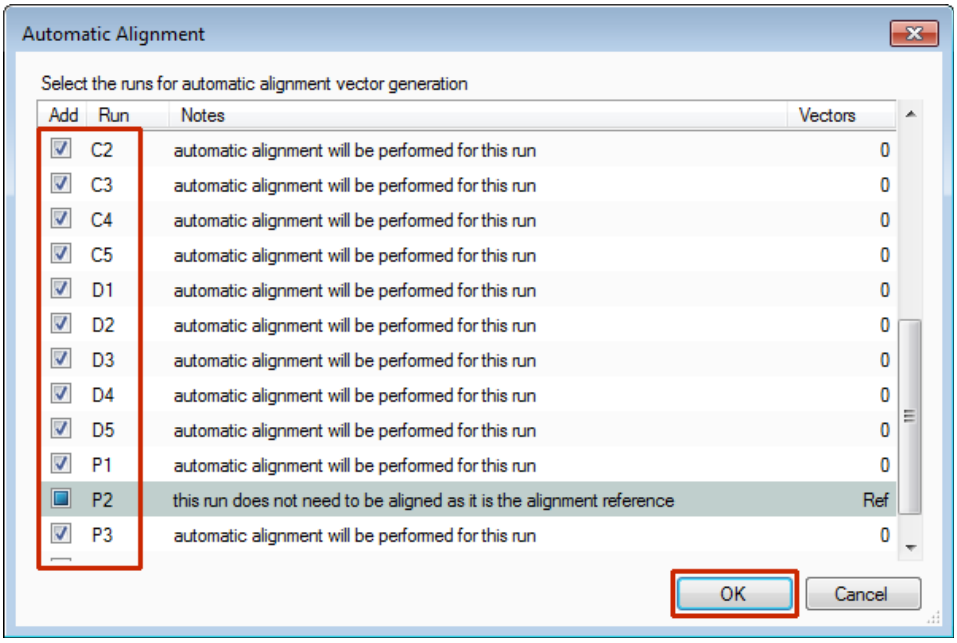
The alignment of the runs is required in the LC (retention time) direction, this is key to correcting for the variable elution of ‘compound ions’ during the chromatographic separation.

The Alignment algorithm will generate ‘Automatic’ vectors, in the retention time direction for each run, to enable the alignment of all the runs to the ‘Reference Run’.

Click **Automatic Alignment** on the top tool bar.



Select (tick) the runs you require to generate vectors for and click **OK**.



The following pages in this user guide explain in more detail the views and functions of the Alignment stage in the Progenesis CoMet, focusing on the Program layout.

These pages act as a useful guide and reference to the Alignment Stage that you can return to after having generated the Alignment vectors automatically.

Taking a detailed approach to alignment

In some cases, where the misalignment of the Retention Time is severe, using a combination of a ‘few’ manually placed vectors on each run and then using the Automatic vector wizard to generate the rest of the vectors for each run can give better results.

For more details on manual assistance of Alignment refer to Appendix 5 page 54

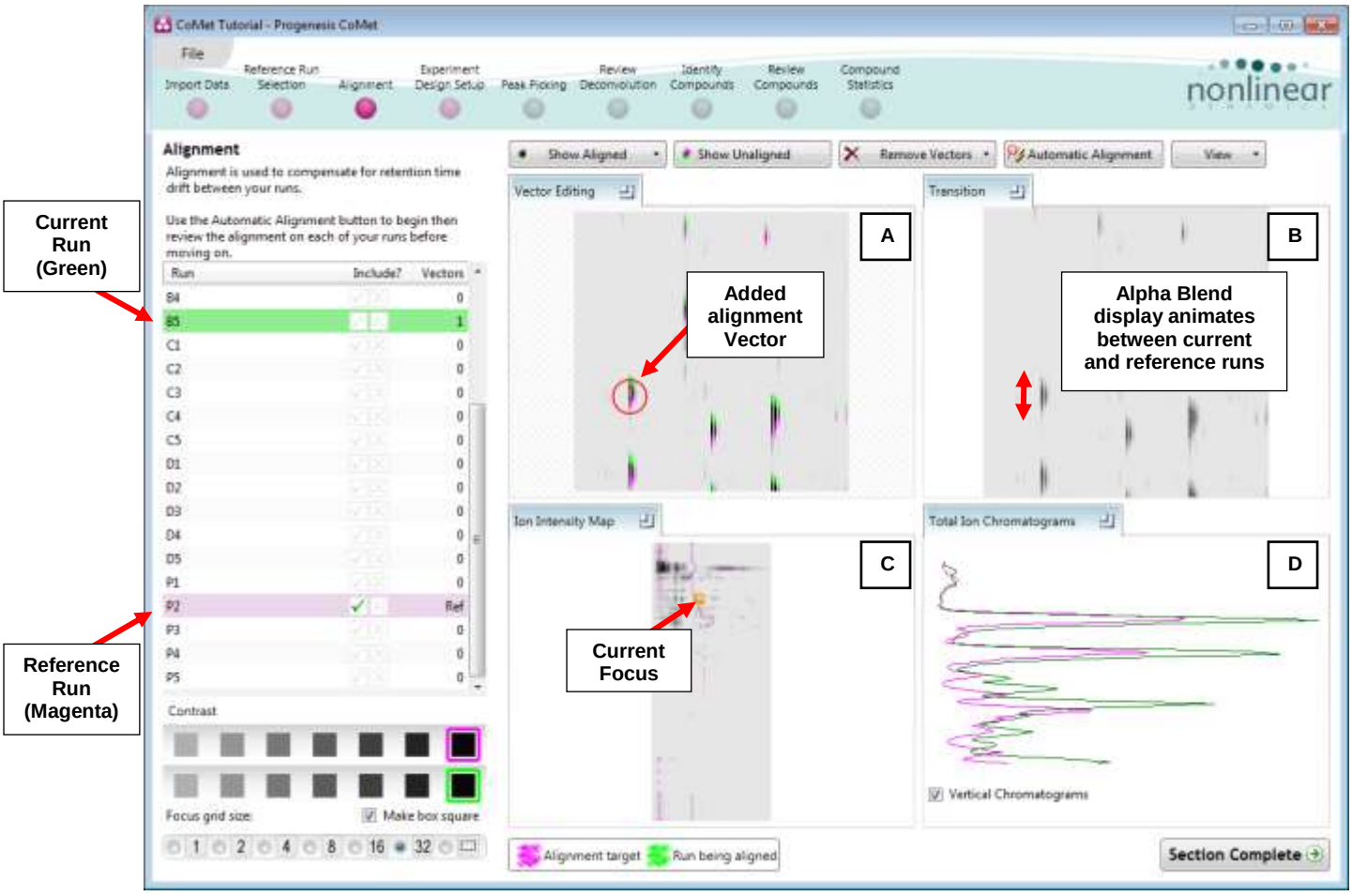
Layout of Alignment

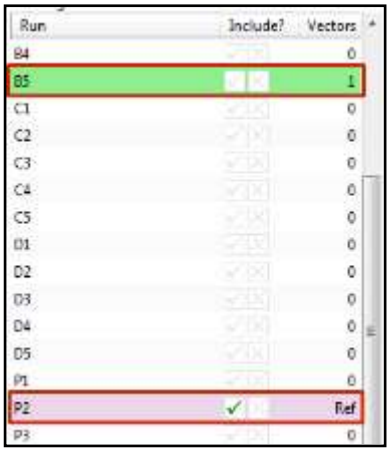
To familiarize you with Progenesis CoMet alignment, this section describes the various graphical features used in the alignment of the runs

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

- Click on the features shown in the current focus (orange rectangle) in Window C, this will update windows A, B and D as shown below.
- In window A **click and hold** the left mouse button on a green feature.
- If the green and magenta features (immediately above) have not aligned automatically then **drag** the green feature over the magenta feature and **release** the mouse button.
- The view will 'bounce' back and a red vector, starting in the green feature and finishing in the circled magenta feature will now appear as shown below in window A.

The experiment structure is displayed on the left of the screen in the **Run** panel.





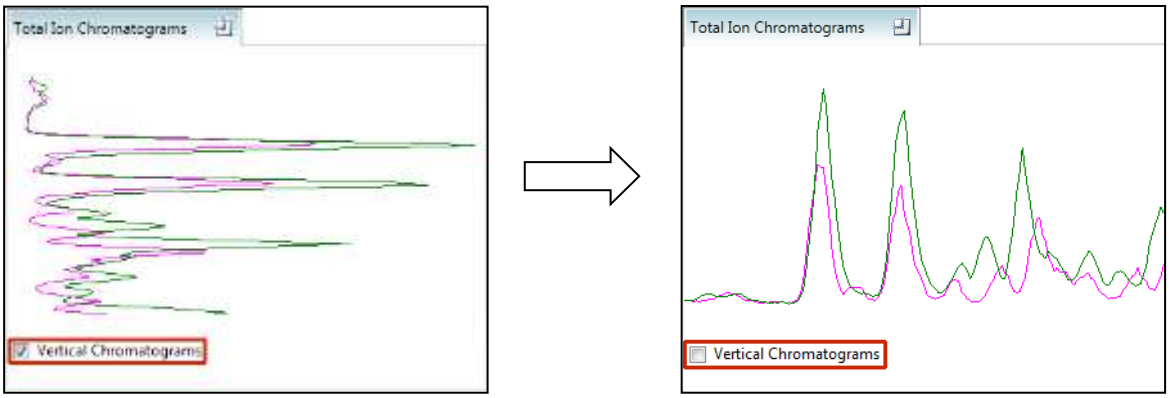
The **Runs:** panel shows the run that is currently being aligned in green, and the run it is being aligned to in magenta. This is the reference run you chose at the previous stage, in this case **P2.**

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 chosen reference run is displayed in magenta. This is where you place the alignment vectors.

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 will change to show a zoomed view of the area being aligned to help accurate placement.

Ion Intensity Map (Window C): shows the **focus** for the other windows. When you click on the view the orange 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 using the controls in the bottom left of the screen or by clicking and dragging out a new area with the mouse.

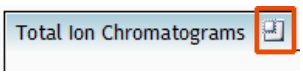
Total Ion Chromatograms (Window D): shows the current **total ion** chromatogram (green) overlaid on the Reference chromatogram (magenta). As the features are aligned in the **Vector Editing** view the chromatograms become aligned. The retention time range displayed is the vertical dimension of the Focus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



Note: the orientation of the TIC view can be changed according to individual preference

This view assists in the verification of the feature alignment .

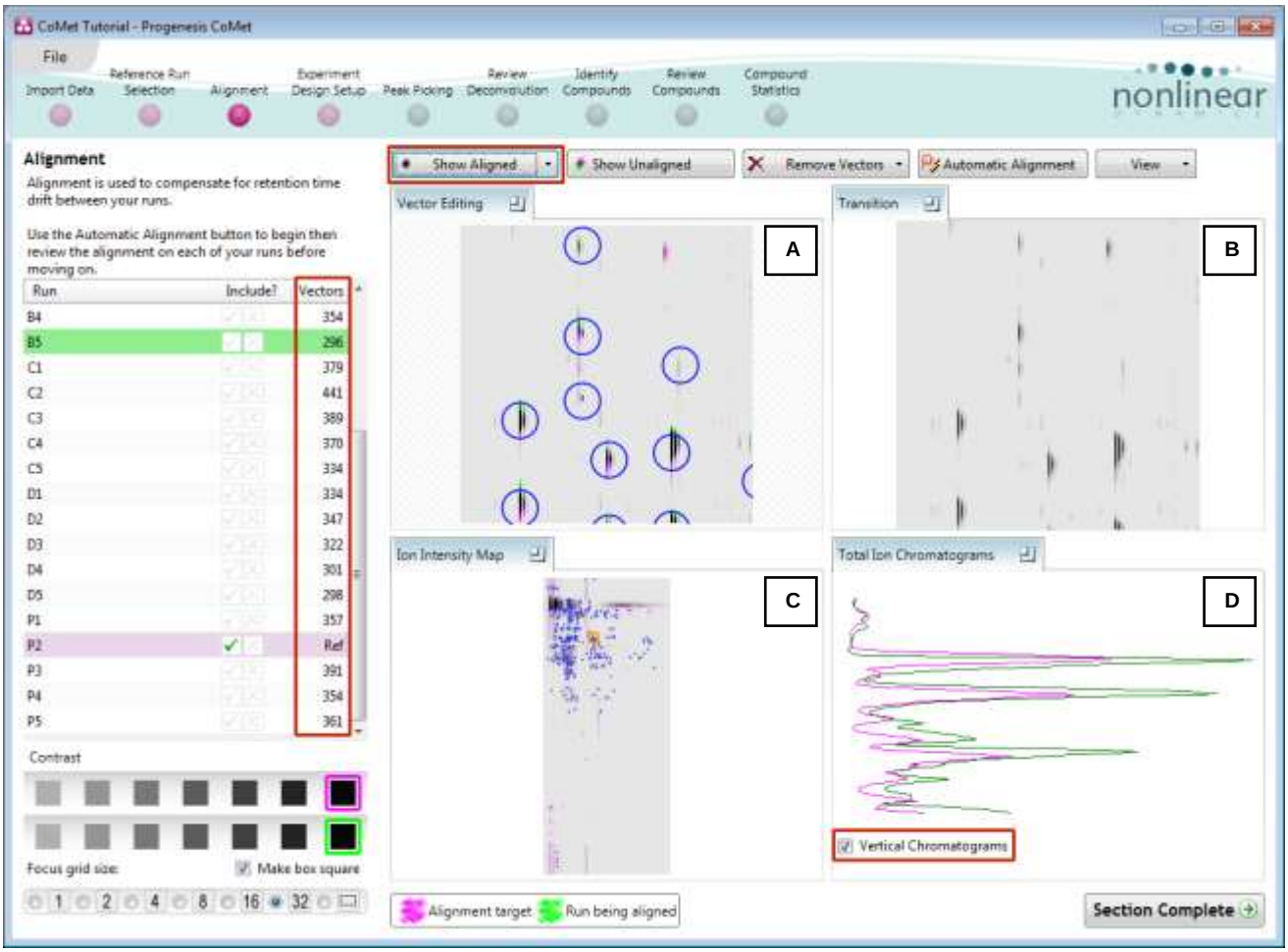
Note: the icon to the right of the 'Window' titles expands the view



Reviewing generation of alignment vectors

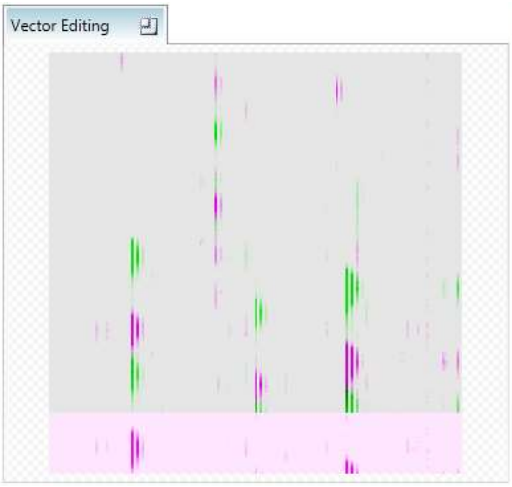
After applying **Automatic alignment** the number of vectors will be updated on the **Runs** panel and the vectors will appear (in blue) on the view.

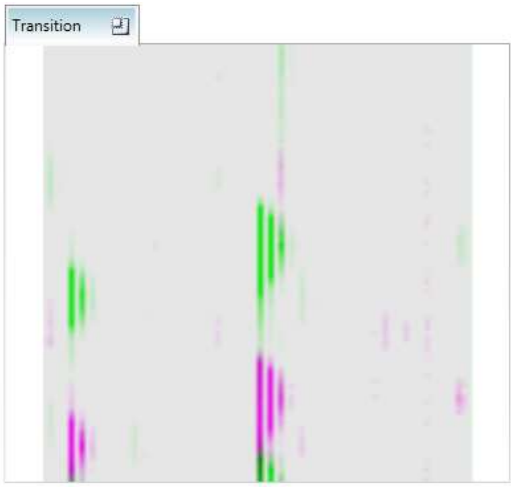
If the alignment has worked well then in Windows A and C the grid lines should show minimal distortion, Window B (Transition) will show features pulsing slightly but not moving up and down.

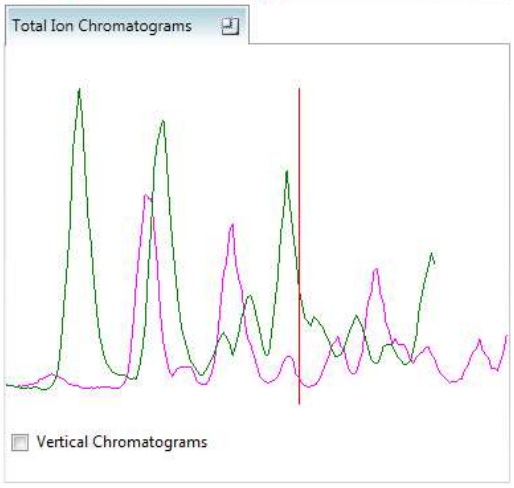


At this point, you should check the automatically placed (blue) vectors. This will be easier with a larger grid size. Make sure the grid size is set to 8 using the **'Focus grid size'** control at the bottom left of the window.

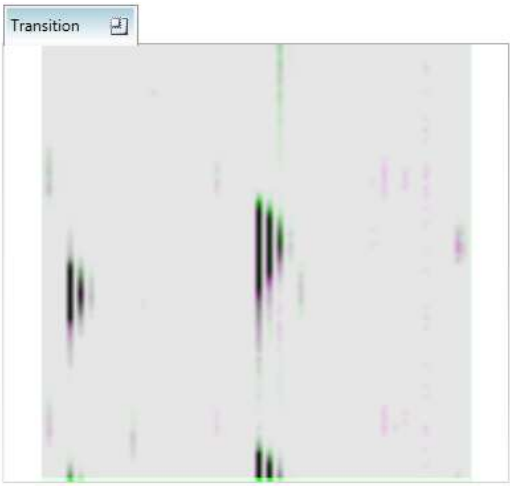
In each square, you can, if required edit the vectors to improve the run alignment (for more information refer to Appendix 5 (page 54).

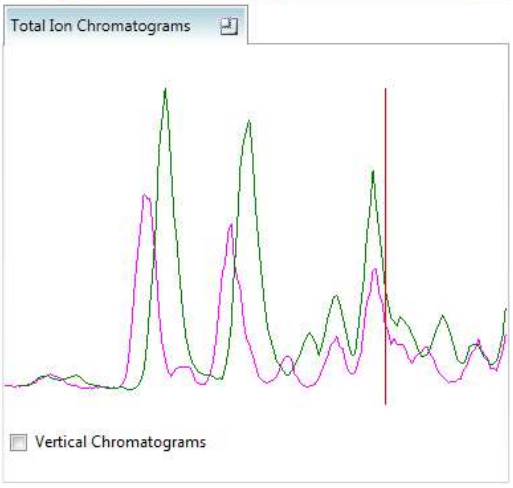






The Total Ion Chromatogram view allows you to further verify alignment of the two runs.



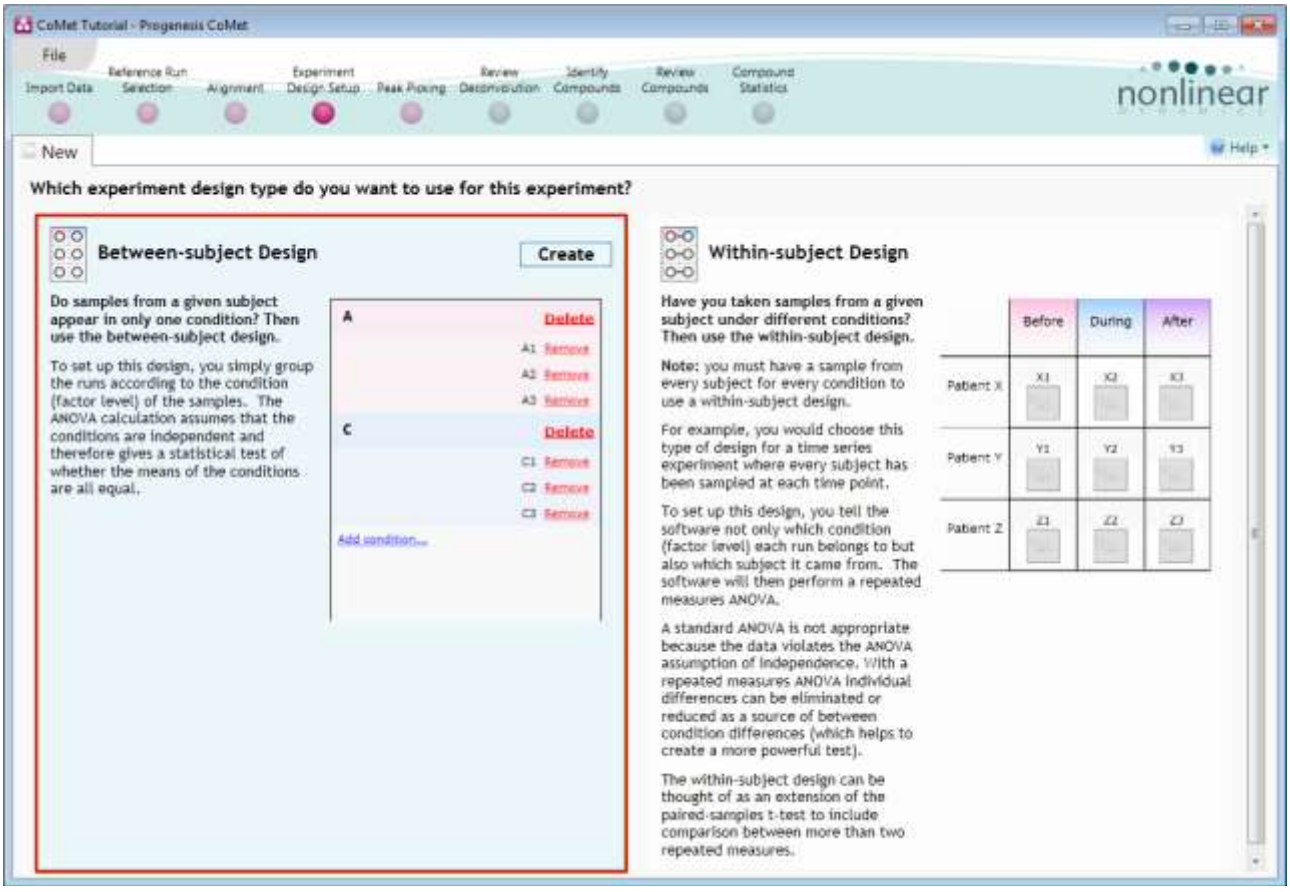


When **show aligned** is pressed you will see the corresponding effect on the alignment of the 'Total Ion Chromatograms'

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:



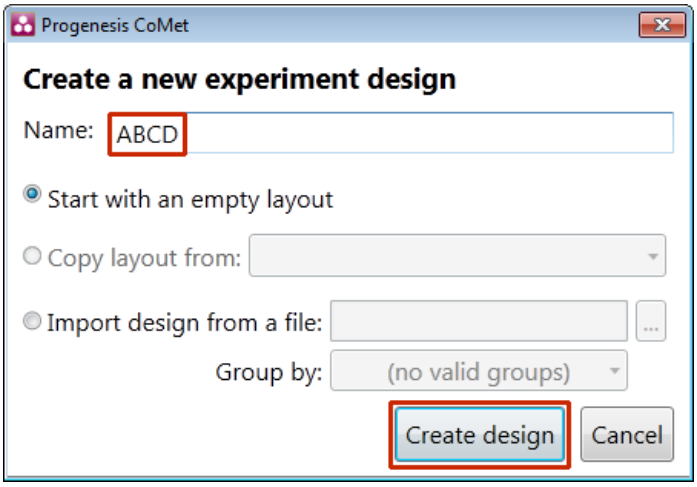
Between-subject design: here samples from any given subject appear in only one condition. (i.e. control versus various drug treatments).The ANOVA calculation assumes that the conditions are independent and applies the statistical test that assumes the means of the conditions are equal.

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.

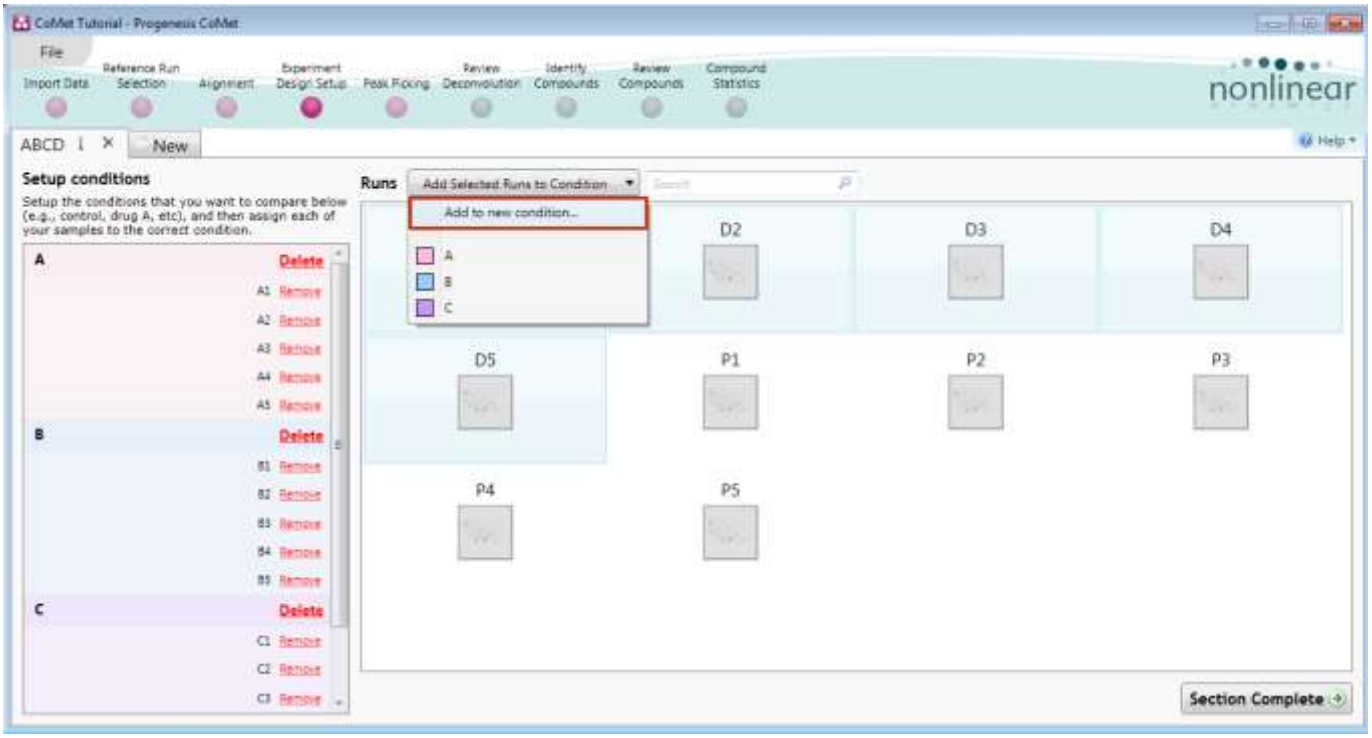
Additional information on how to apply the Within-subject Design is in Appendix 6 page 65

This experiment contains 4 conditions: A, B, C and D as well as Pools (P) 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.



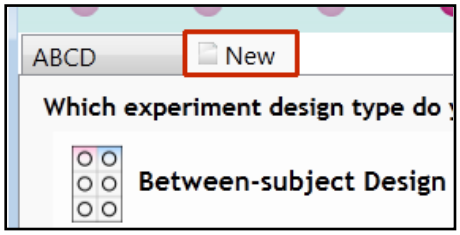
Give the new experimental design a name and then click **Create design**.



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. P) by over typing the default name.
6. Repeat steps 1 to 5 until all the runs are grouped into conditions.

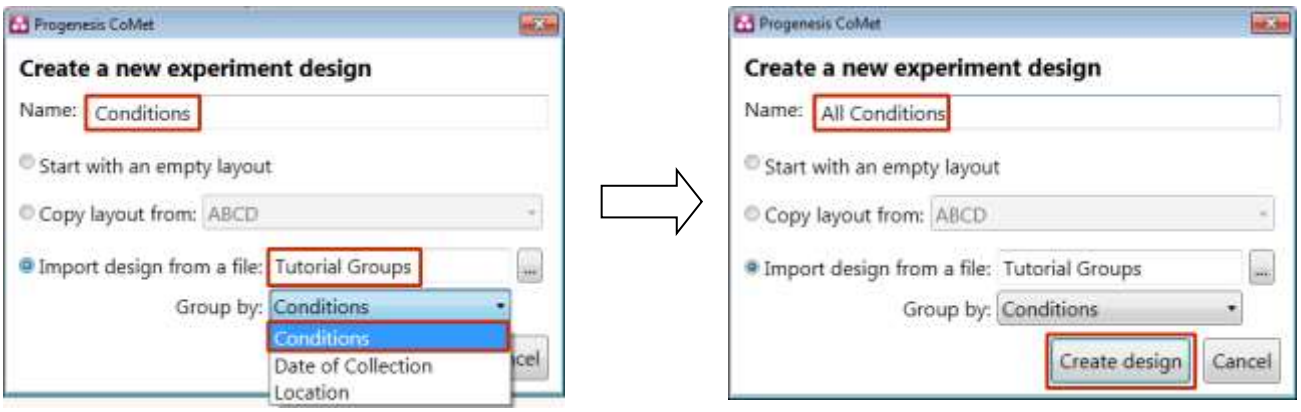
An alternative way to handling the grouping of this set and other larger (and more complex) experimental designs is to make use of sample tracking information that has been stored in a spread sheet at the time of sample collection and/or preparation.



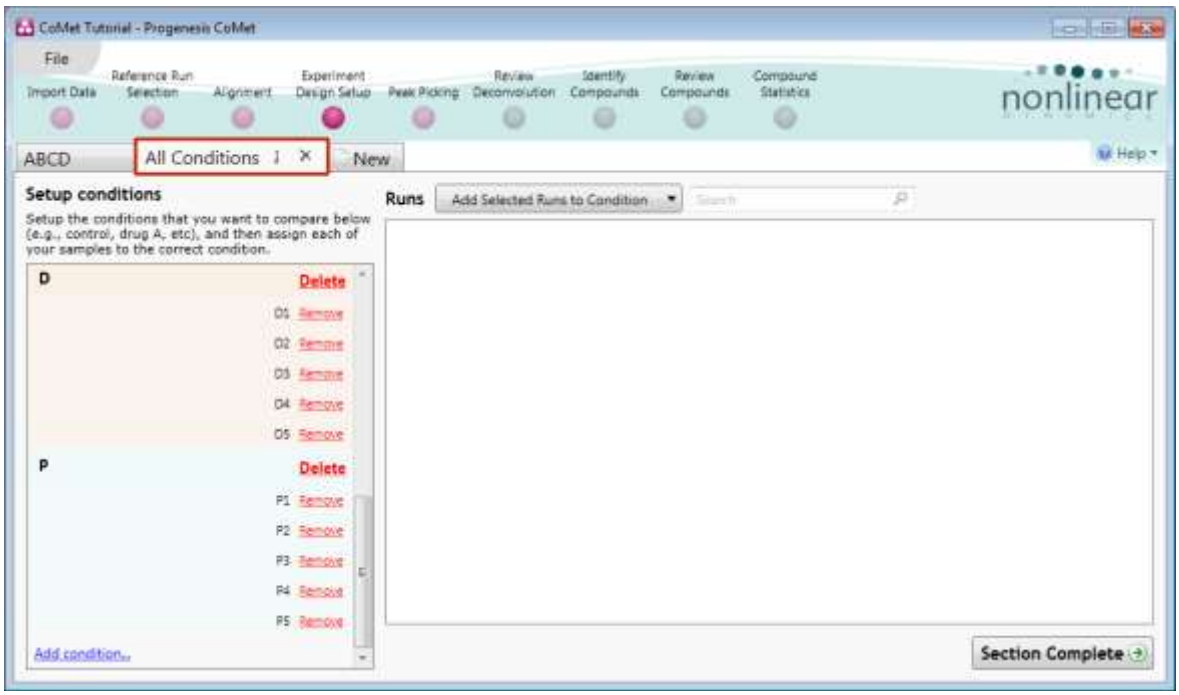
For this example there is a Tutorial Groups.csv file available in the Experiment Archive you restored at the beginning of this tutorial exercise.

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: Conditions.

Note: the design name will update to reflect your 'Group by' selection. You can rename the design as required.



When **Create design** the new design appears as the current Experimental Design.



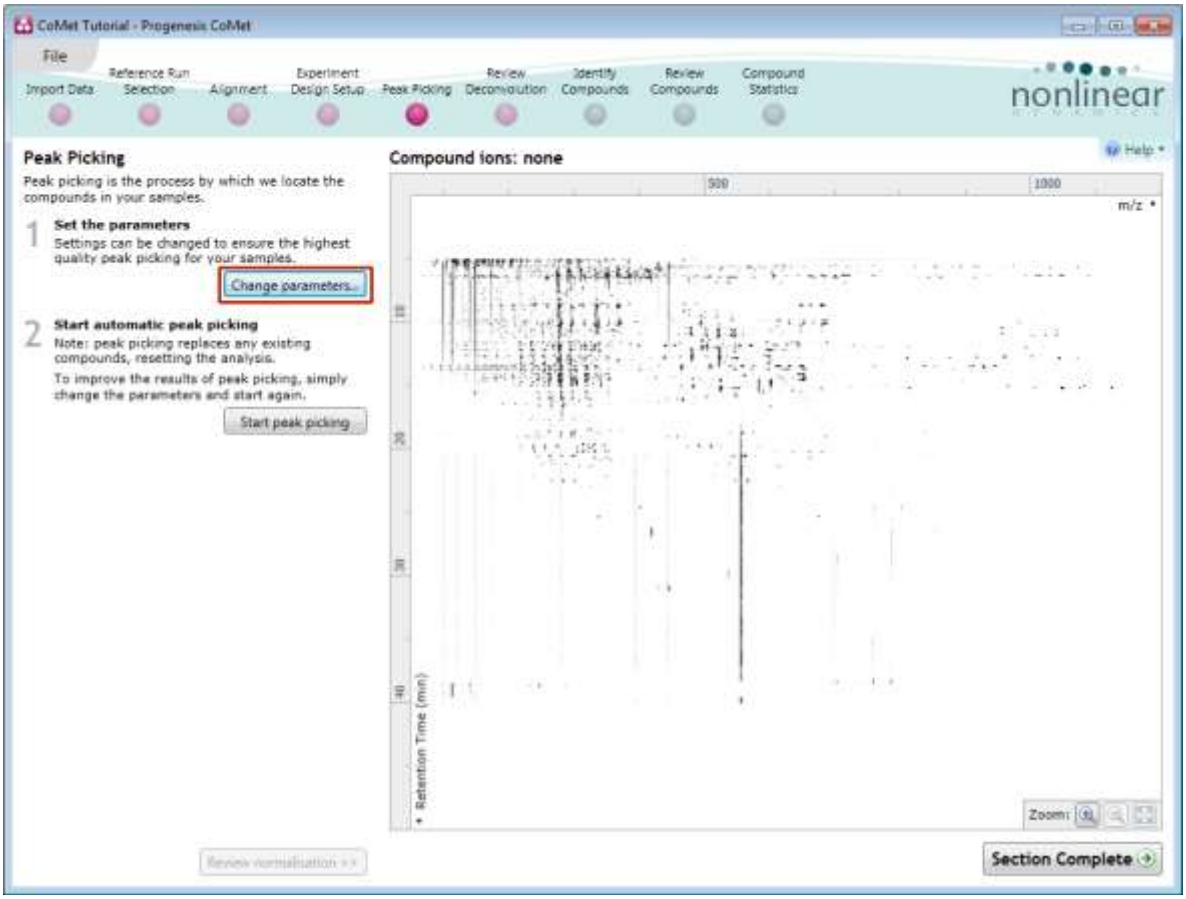
Note: you can use **Delete** on the Conditions panel to remove conditions that are not required in this particular design.

Note: both designs are available as separate tabs. Additional Experiment Designs can be added at any time in the analysis of your data

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

Stage 6: Peak Picking

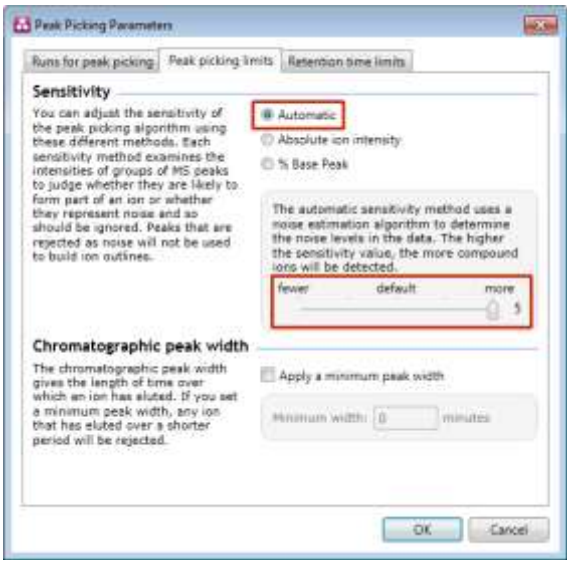
Having set up one or more Experiment Designs for your data the Peak Picking stage will open as shown below. Here you can easily define thresholds for the detection of peaks across all of your aligned runs.



The ion intensity map displayed before Peak picking is for the 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.

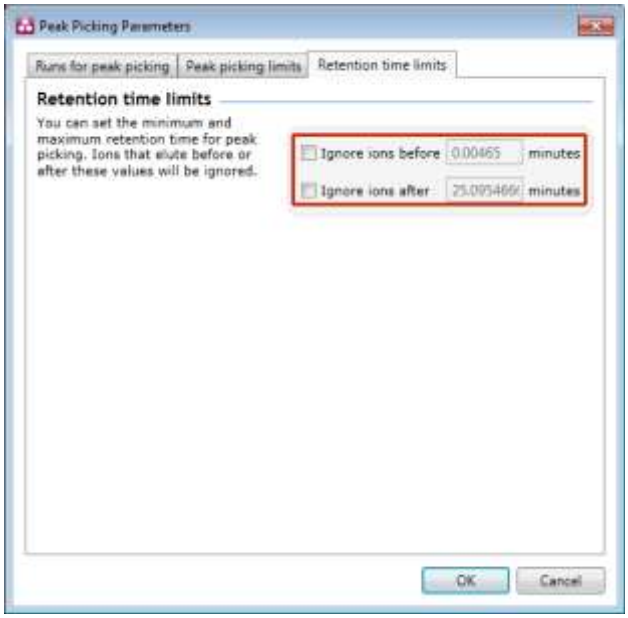
Tip: depending on run quality, a suggested minimum number of ticked runs should include at least one replicate of each experimental condition.

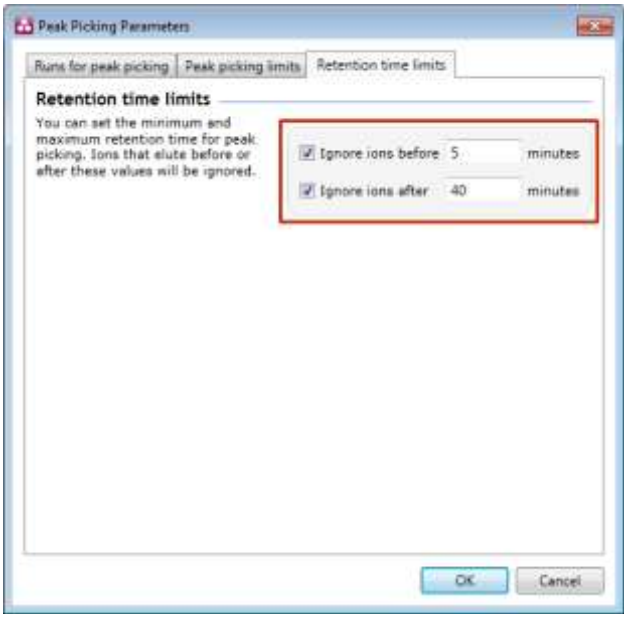
The sensitivity of the peak picking can be controlled by adjusting settings under the **Peak picking limits** tab.

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. Enter values of 5 and 40 min and tick the boxes as shown below.

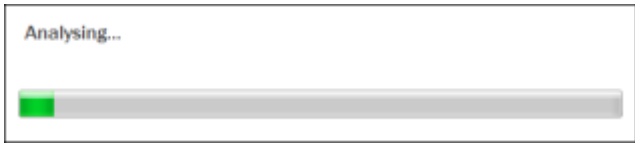




Click **OK** to close the Peak Picking Parameters dialog.

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

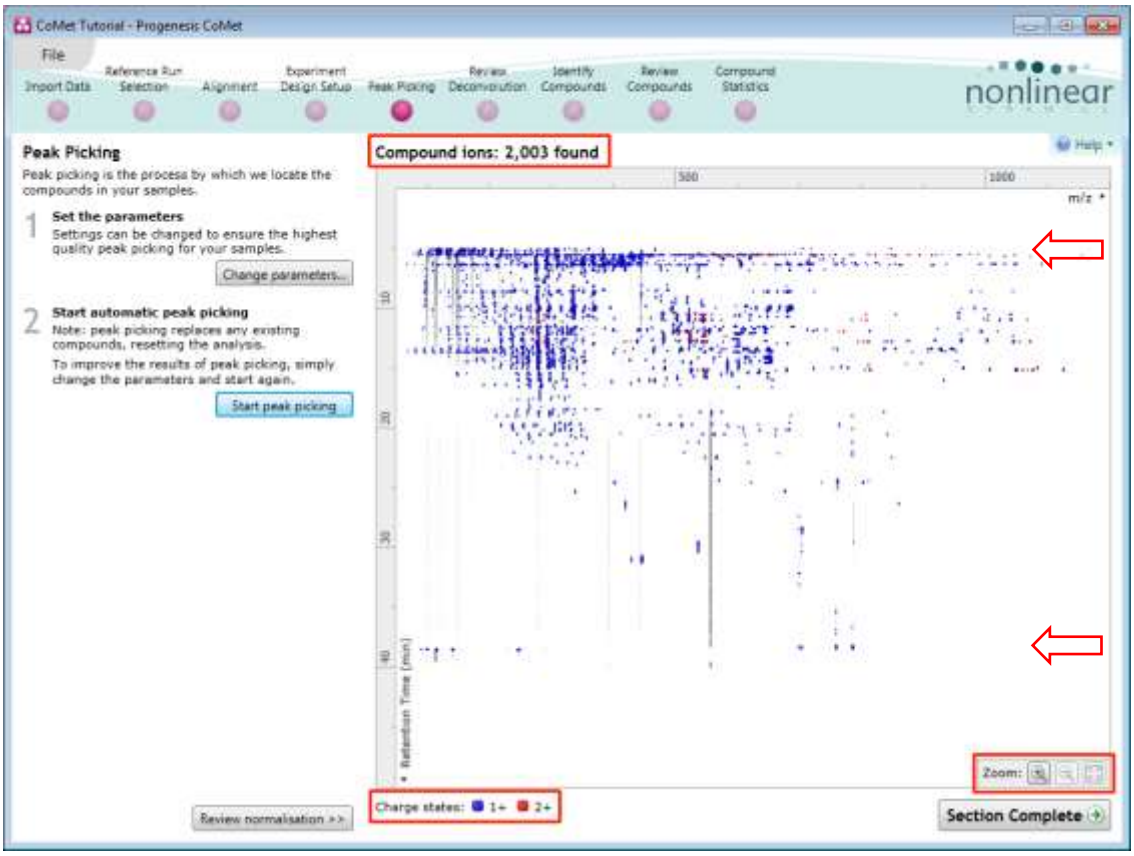
During the few minutes that the automatic analysis requires, a progress bar will appear telling you that it is Analysing.



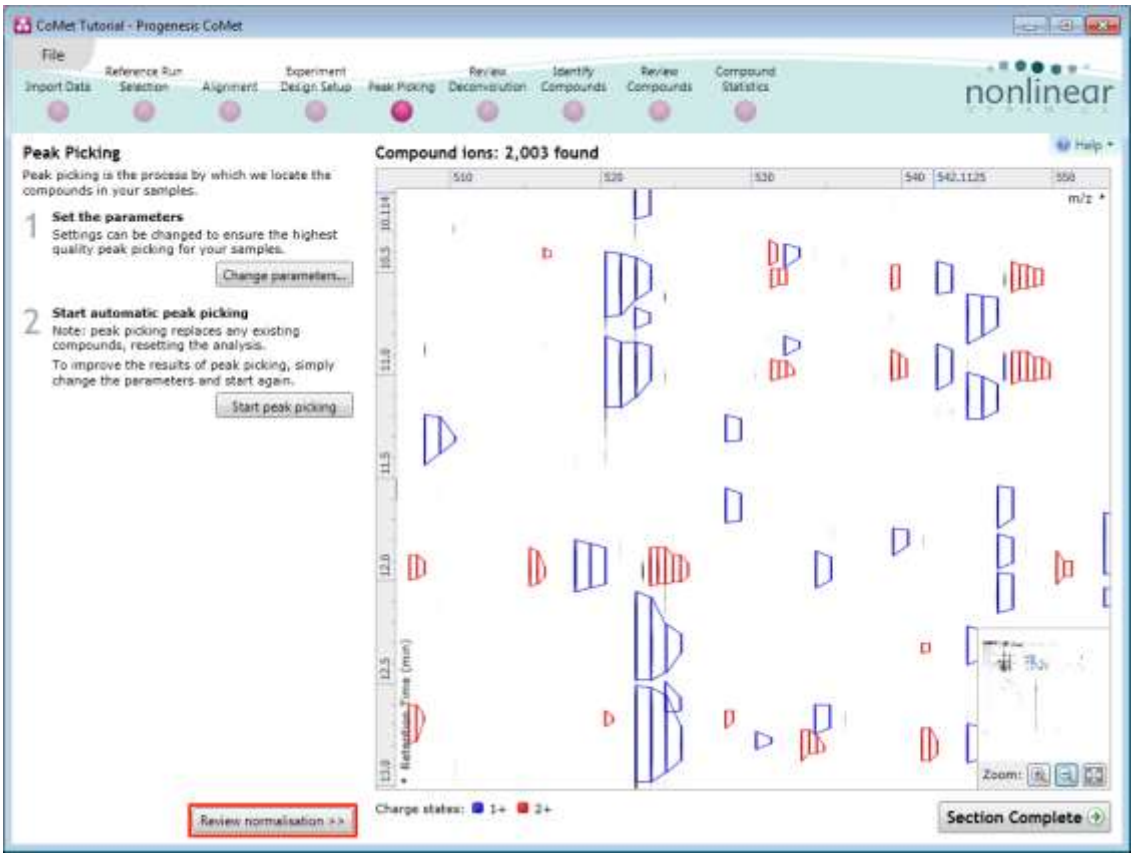
Following automatic peak picking the detected peaks are displayed in a colour according to their charge state.

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

Note: no peaks are detected before or after the 5 and 40 min retention time limits.



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.**

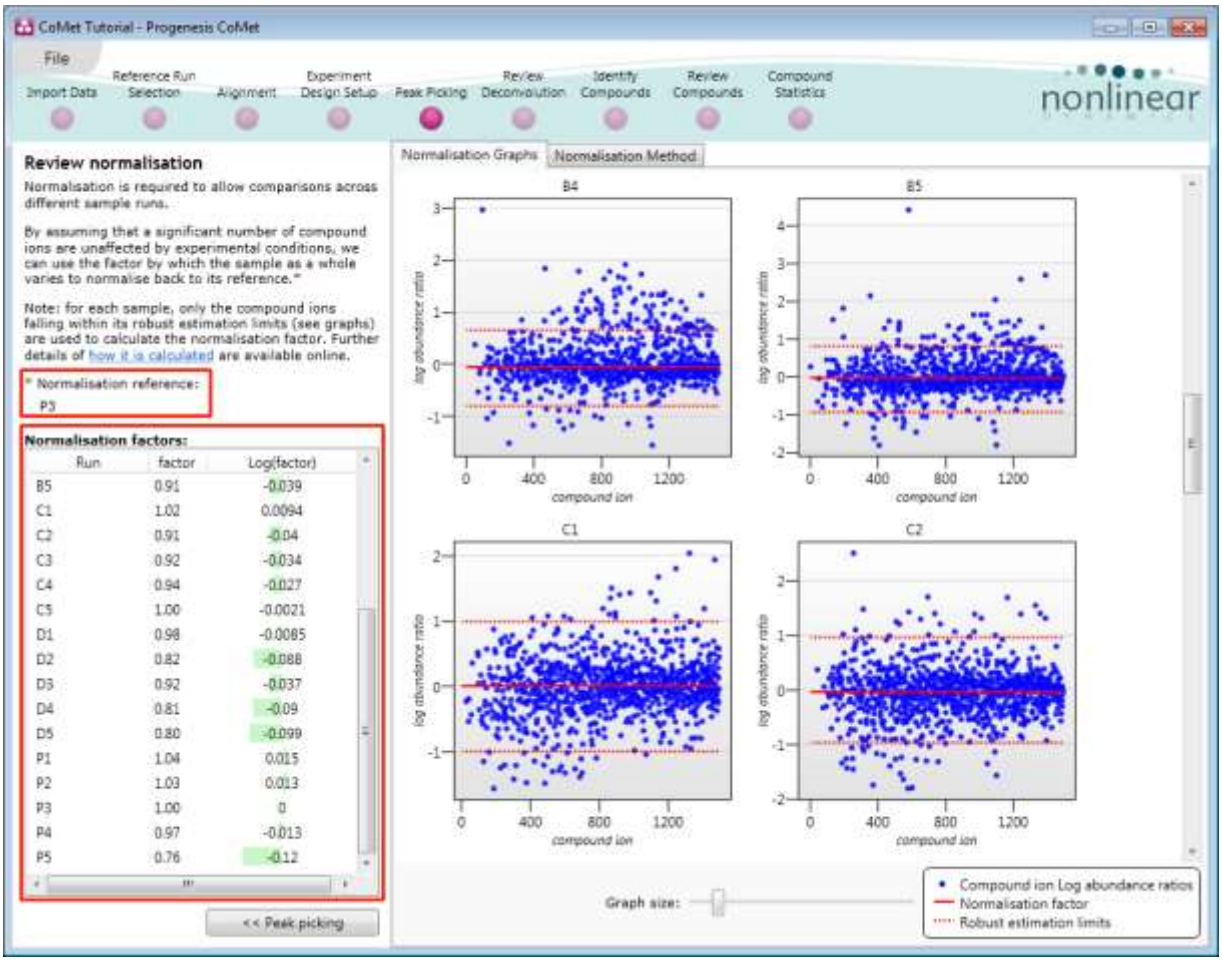
Stage 6B: Reviewing Normalisation

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.

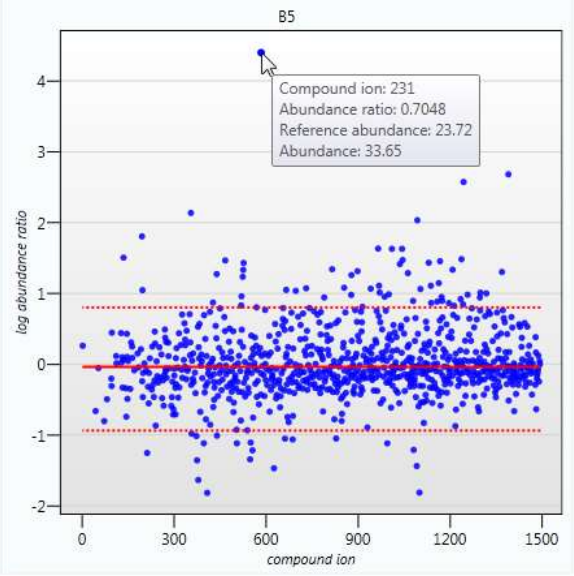


Calculation of Normalisation Factor:

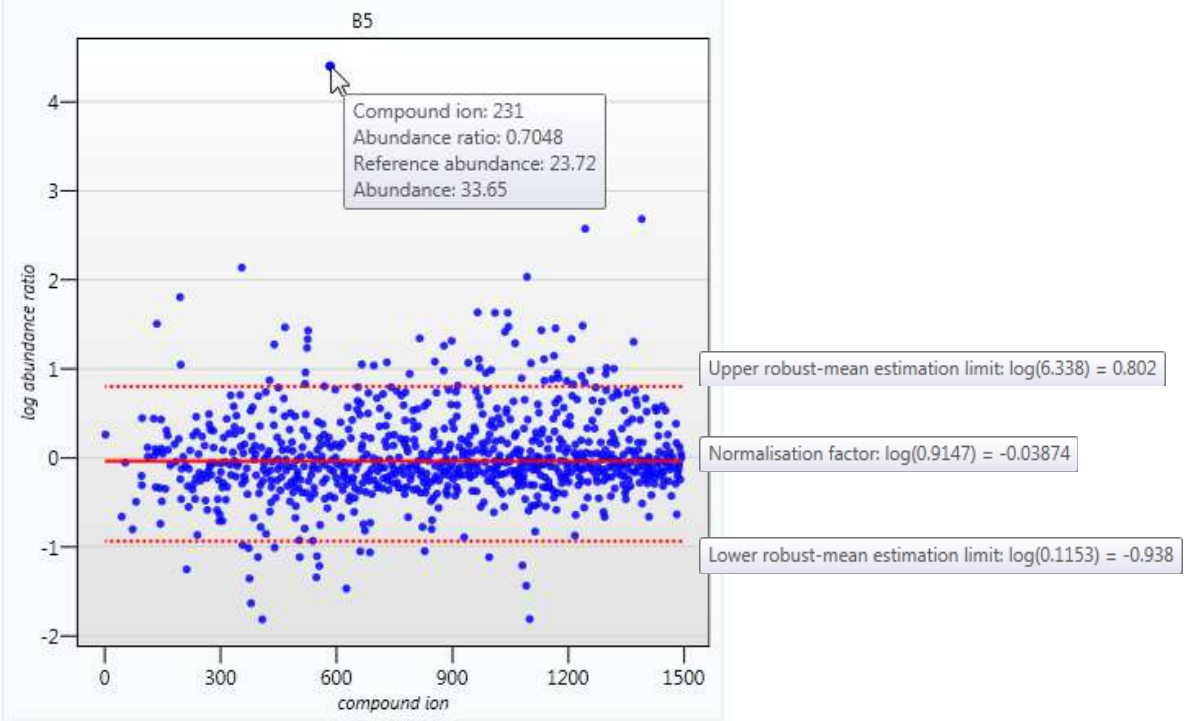
Progenesis CoMet will automatically select one of the runs that is 'least different' from all the other runs in the data set to be the 'Normalising reference'. The run used is shown above the table of Normalisation factors.

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

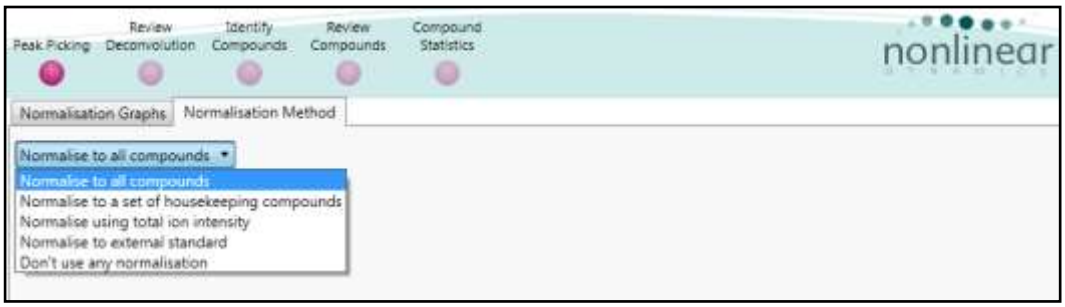
The details for individual features can be viewed as you hold the cursor over the dots on the plot.



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 normalisation



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



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

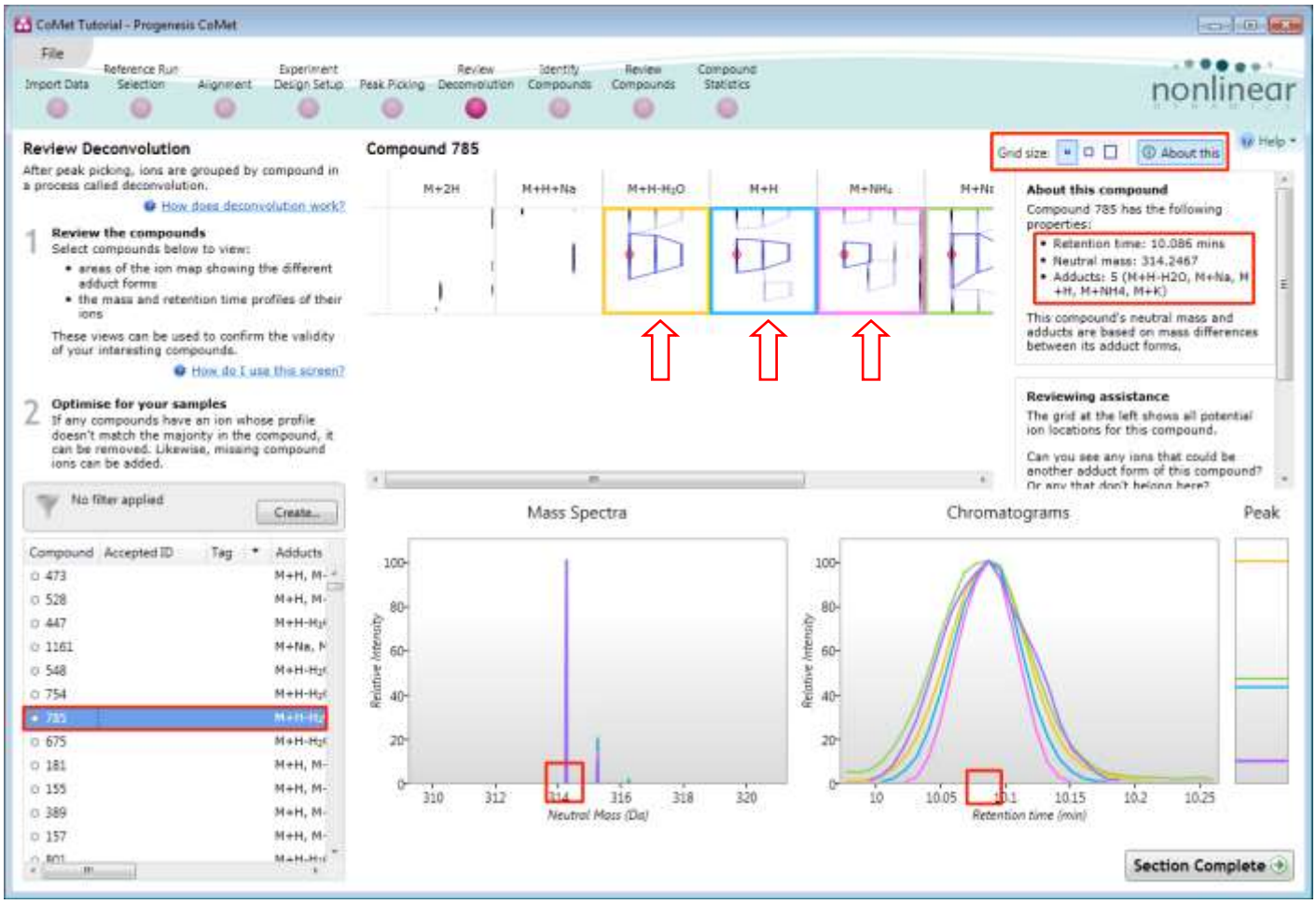
Note: there are 5 Normalisation modes

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.

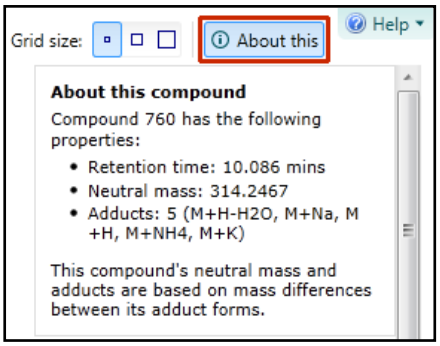


At this stage in the CoMet workflow you get the opportunity to review the process of Deconvolution for the Compounds. The table on the left displays the compounds ordered by the number of Adducts detected for each compound.

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 785**, 5 adducts have been detected: M+H-H_2O, M+H, M+NH_4 M+Na and M+K (arrows) and displayed in the table.



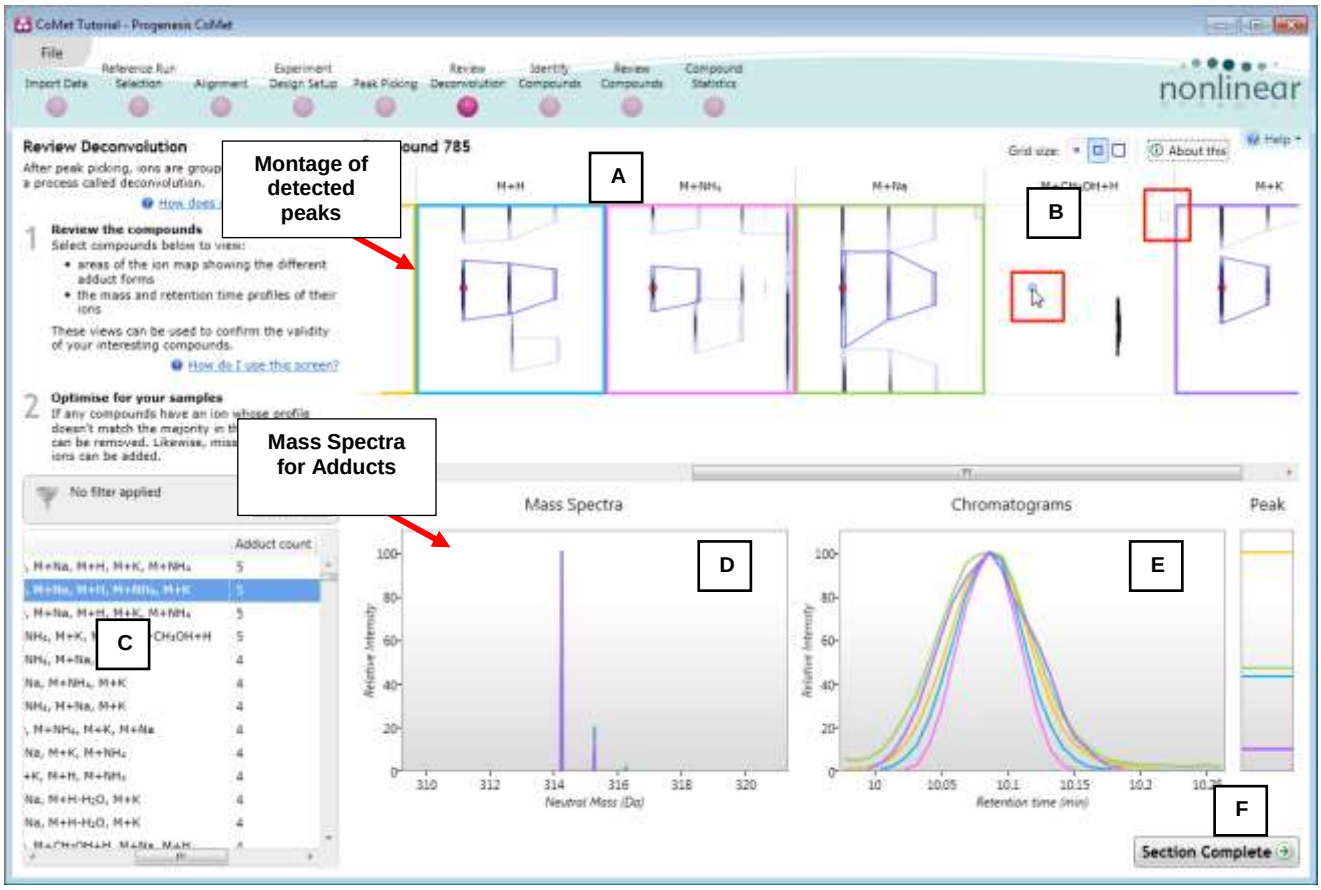
Tip: the '**About this**' panel, top right summarises the information for the current compound. If not displayed click on the **About this**.

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.

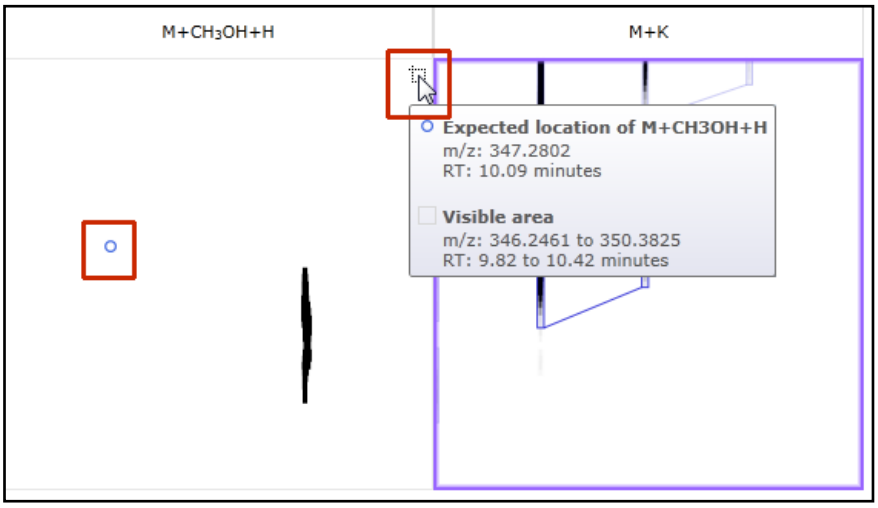
Using Compound 785 we can look at how the information is displayed on the multiple views following deconvolution.

Note: for clarity the 'About this' pane is hidden by clicking on the About this icon.

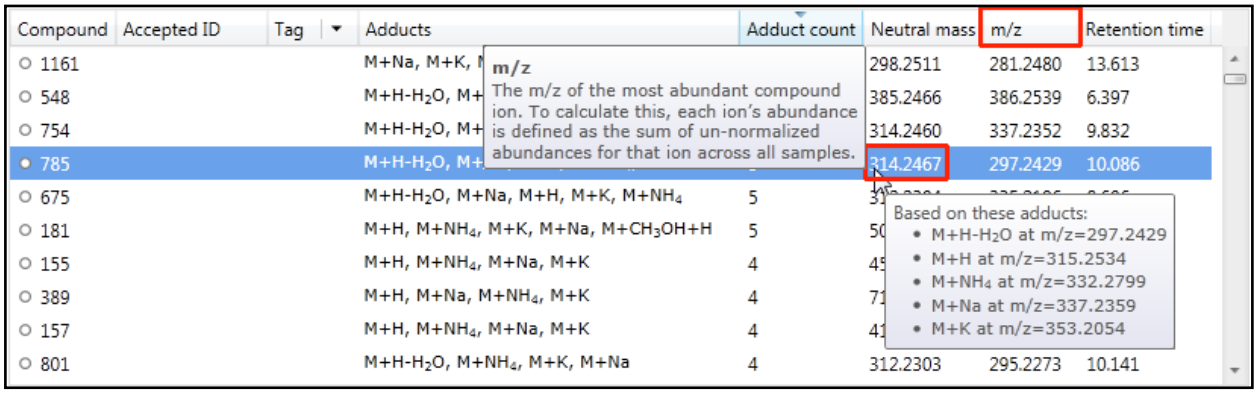


(A) Montage view showing location of detected adducts for compound

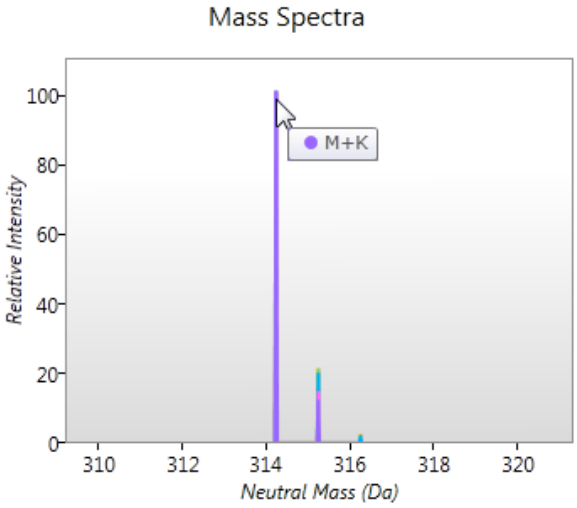
(B) Where an adduct has not been detected the expected location for the adduct is displayed on the relevant panel as a blue circle when you hover the cursor over the panel. In this case no ($M+CH_3OH+H$) adduct has been detected. **Note**: hover the cursor over the icon on the top right of the panel and you will get information about the expected location and the area displayed.



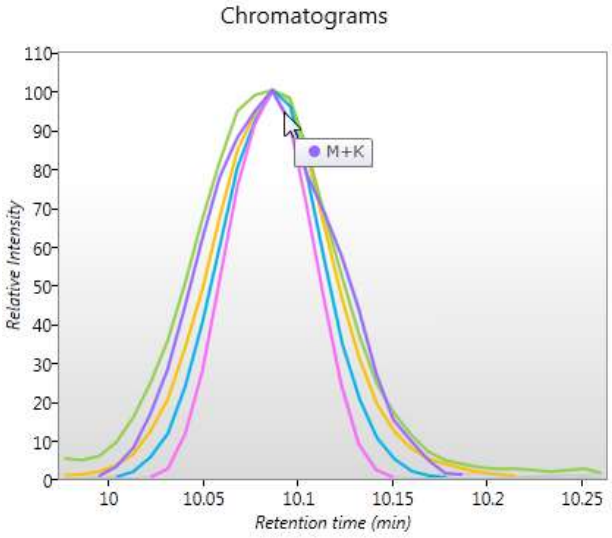
(C) The table displays the information known about the current compound with regards to the number of adducts, identity of the adducts and where deconvolution has been successful the Compound's Neutral mass is displayed. Note: when you hover the cursor over the **Neutral mass** for the current compound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



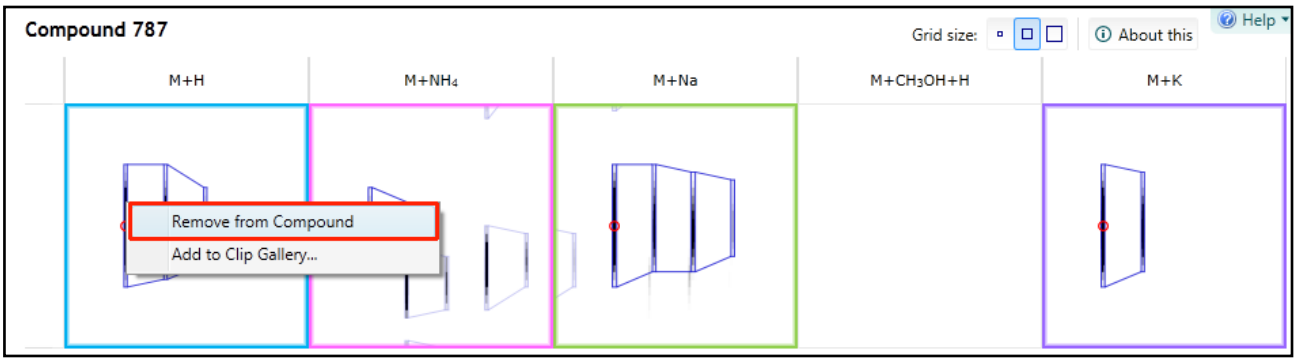
(F) This panel shows the actual Peak heights for the detected adducts of the Compound, colour coded by adduct.

Editing Compound adducts.

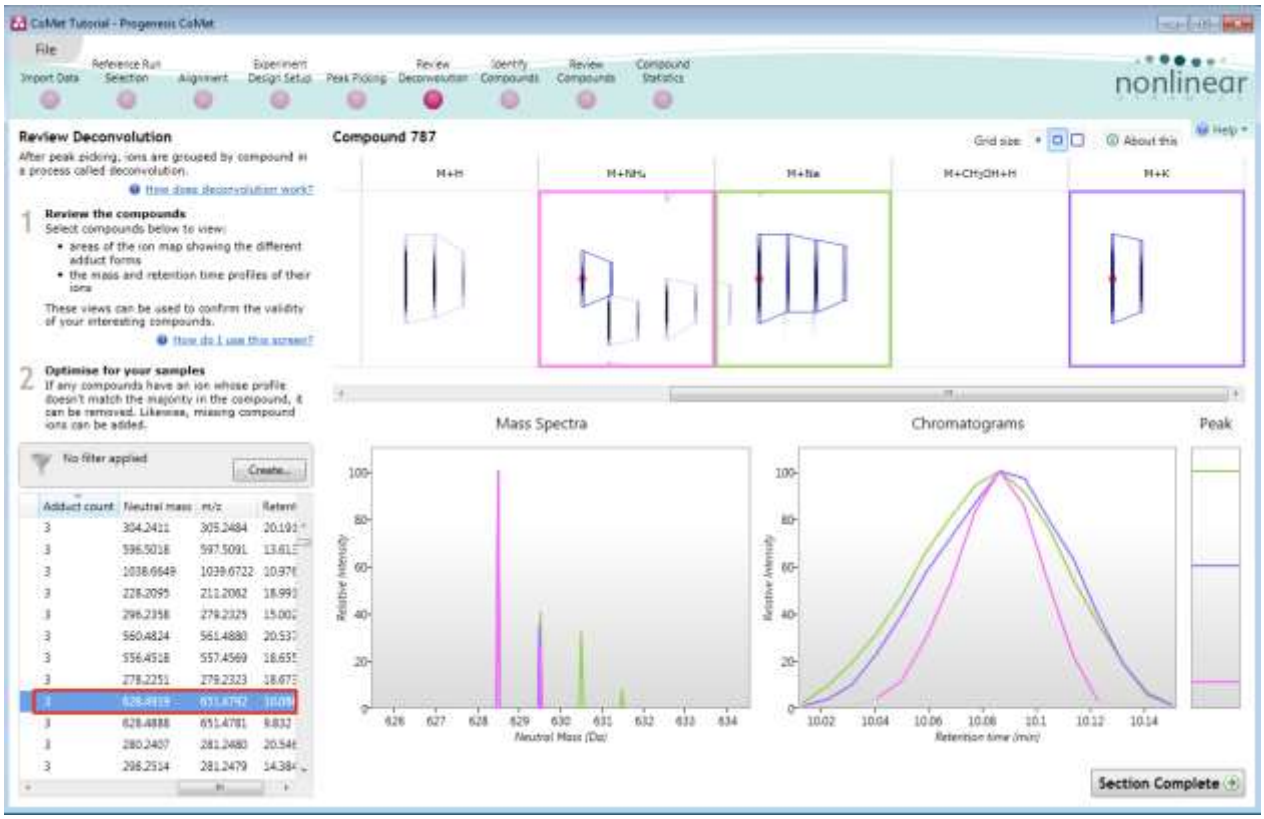
The number of adducts assigned to a compound by the process of deconvolution depends on the peak detection and the adducts selected at the beginning of the workflow when the experiment is created.

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+H adduct for compound 787 (which currently has 4 adducts), 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.

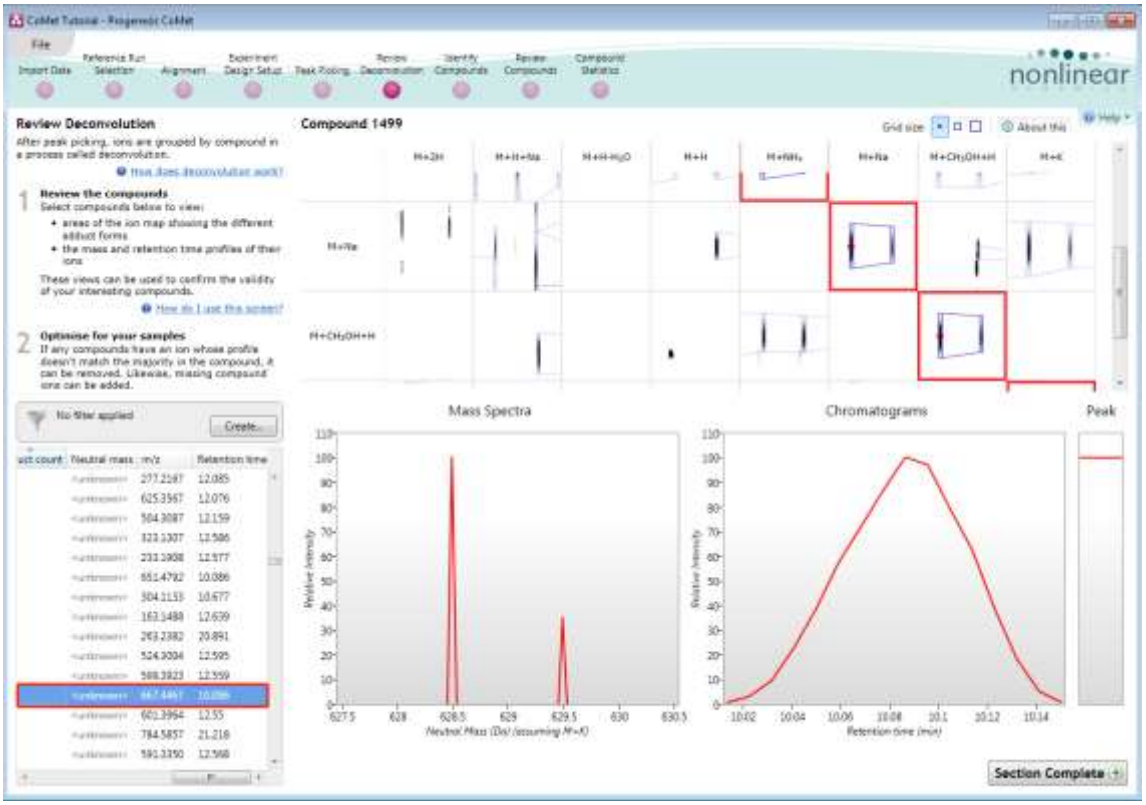
If the compound has only 2 adducts and you remove one of them then the montage view will change to display the full matrix of all possible peak locations for expected second adducts dependant on the m/z

values for these adducts. Also as there is only one adduct detected the Neutral mass for the compound can no longer be calculated.

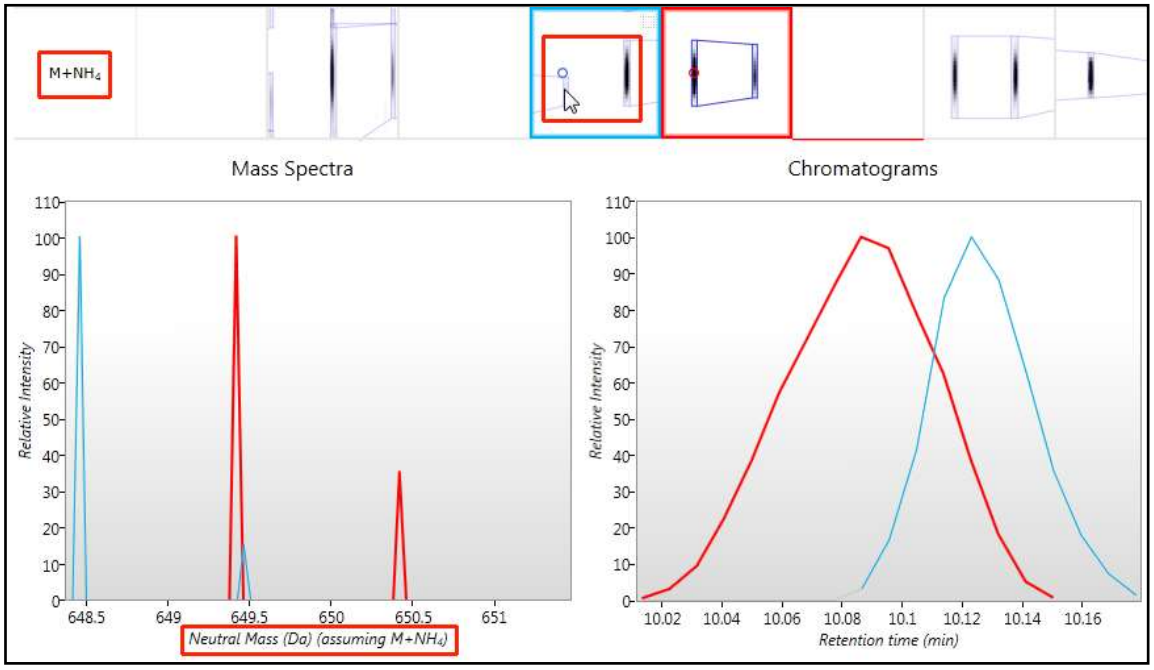
Try deleting the M+NH_4 and the M+Na adducts from compound 787

Note: when you do this the table updates and the Compound id is updated to Compound 1499.

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.

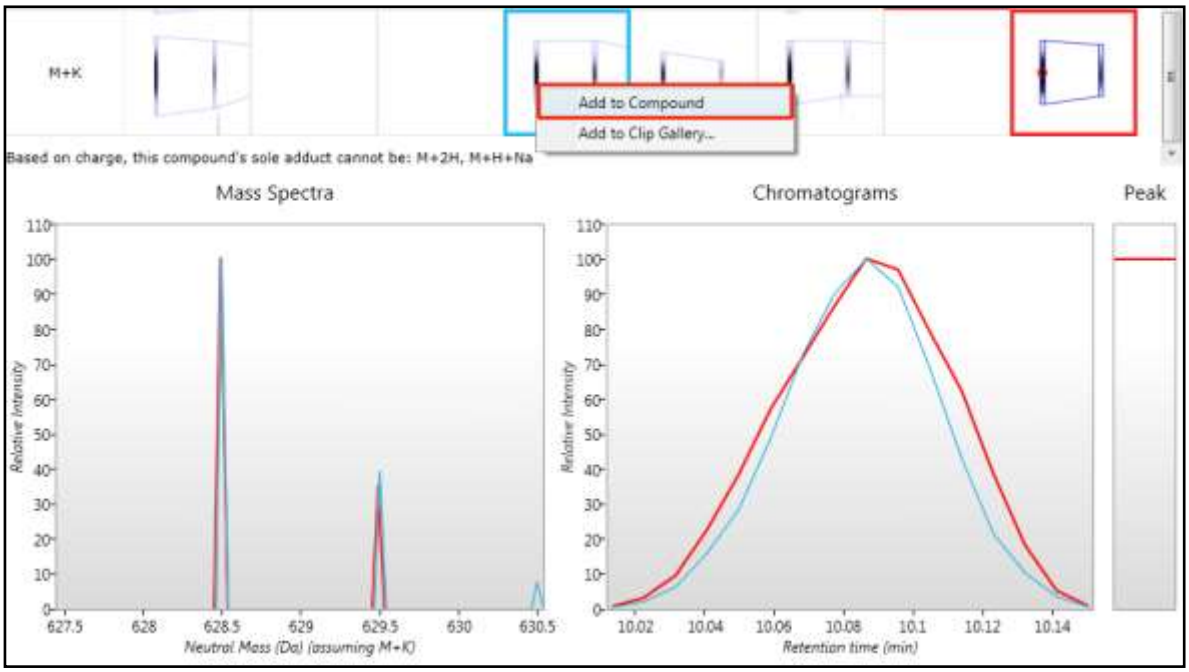


As an example to explore the posibility that Compound 1499's sole adduct is M+NH_4 try hovering over the detected peaks for the expected adducts. In this case when you hover over M+H as the possible second



adduct you can see that the peak elutes at a different retention time on the chromatogram, and the mass spectra does not coincide. On that basis the second compound adduct cannot be M+H when M+NH_4 is assumed to be the sole detected adduct.

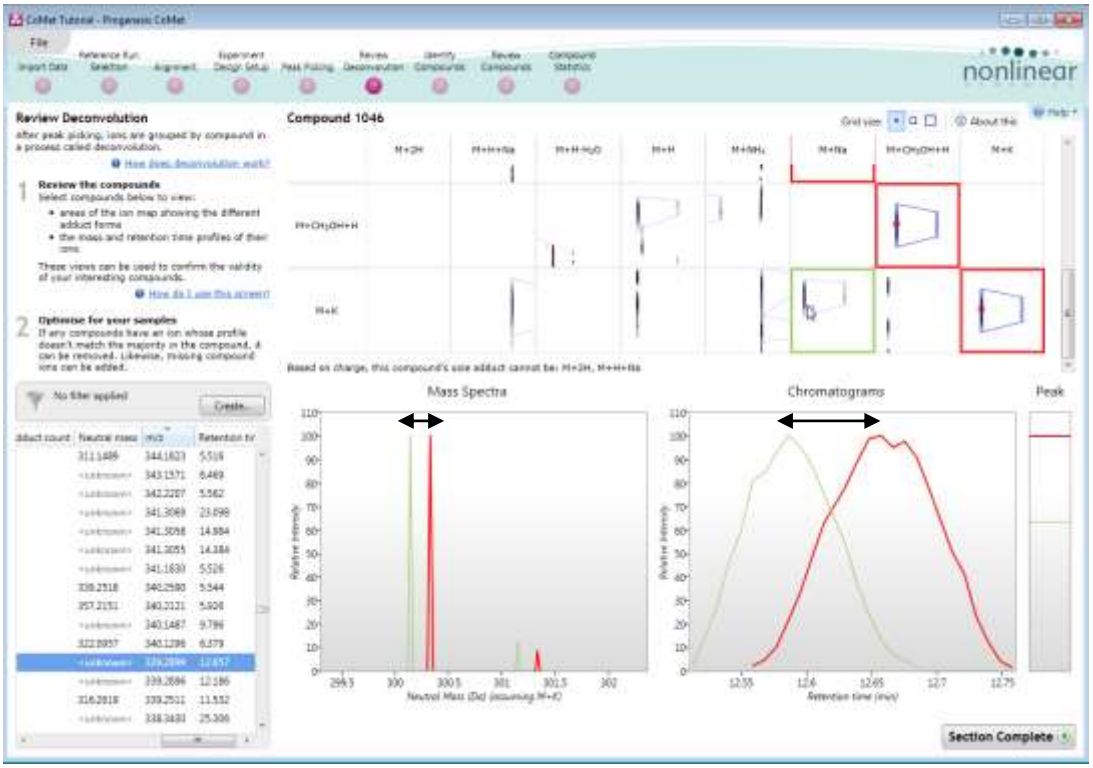
Now try M+K as the assummed sole adduct of Compound 1499. This time when you hover over the M+H adduct both the m/z on the mass spectra and RTon the chromatogram appear to coincide on the graphs with the corresponding location for the existing single adduct M+K.



To add the M+H 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 M+NH_4 and the M+Na 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.

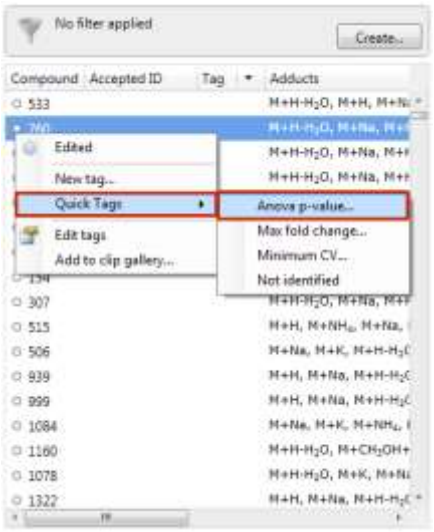
By adding the second adduct the montage and table will update to reflect the addition of the adduct to the compound and an Neutral mass will be assigned based on the 2 adducts. The matrix collapses to the single row displaying the locations of the two detected adducts.

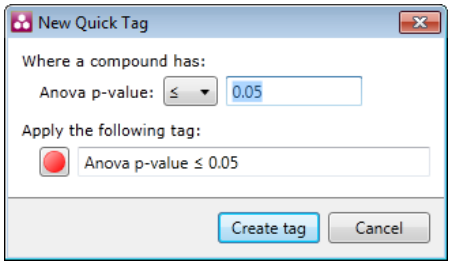
Using 'Tagging' to filter data for identification

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.

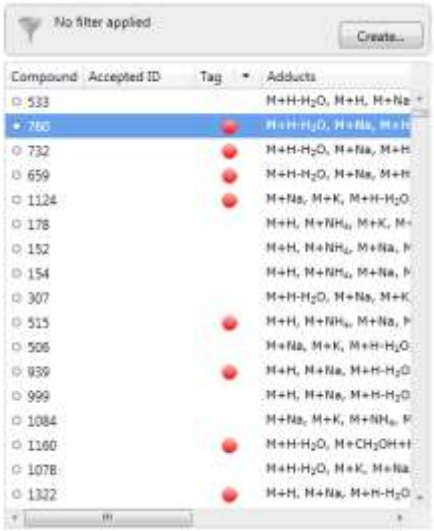
As an example of the use of tags we will set out to identify the compounds that are associated with a specific population of detected features. This population of features are required to show a significant (Anova p value < 0.05), 2 fold or greater increase in abundance for one of the conditions (A,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

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…**





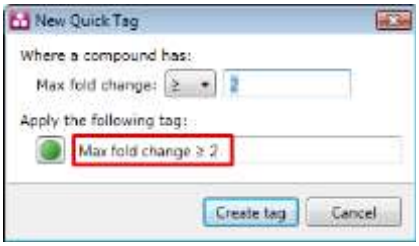
Accept the default value (≤ 0.05) and the offered name.



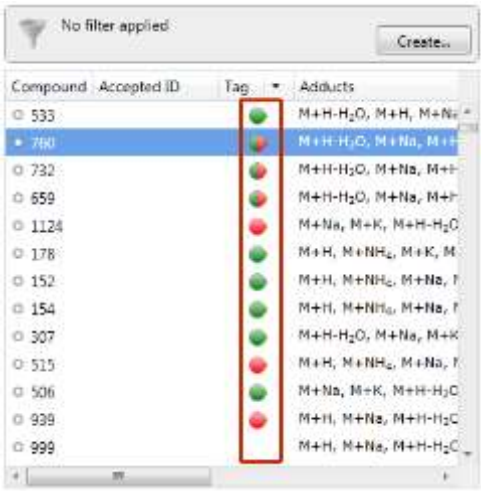
On pressing **OK** a tag appears in the table against all the compounds with an Anova p-value ≤ 0.05

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**.

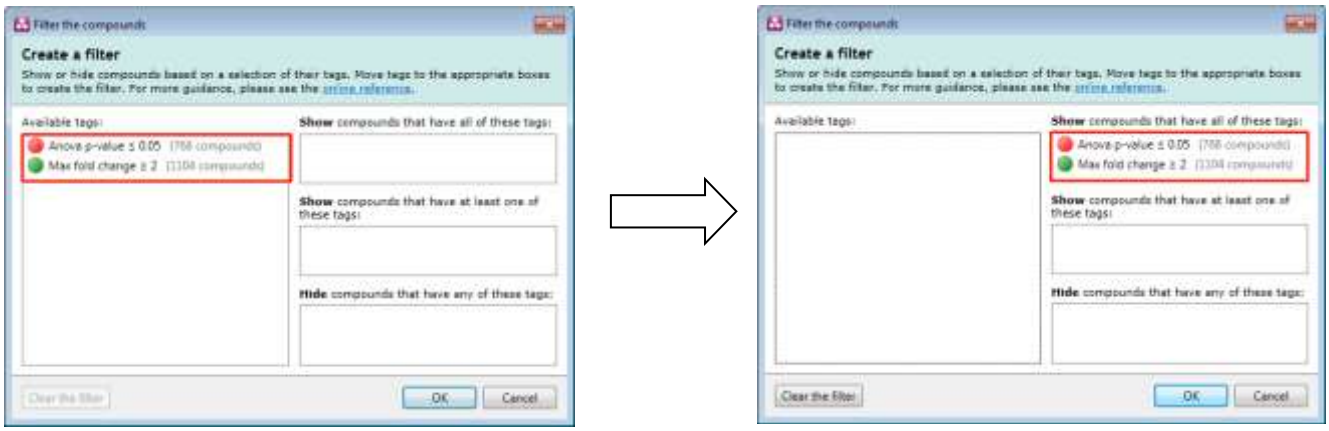


On pressing **OK** a second tag appears in the table against all the features with a Max fold change ≥ 2.



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.

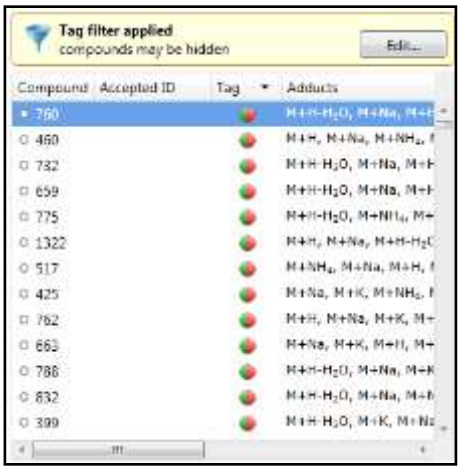
To set up a filter click **Create** to open the Filter dialog and drag the new tags on to the **Show** features with all of these tags.



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.

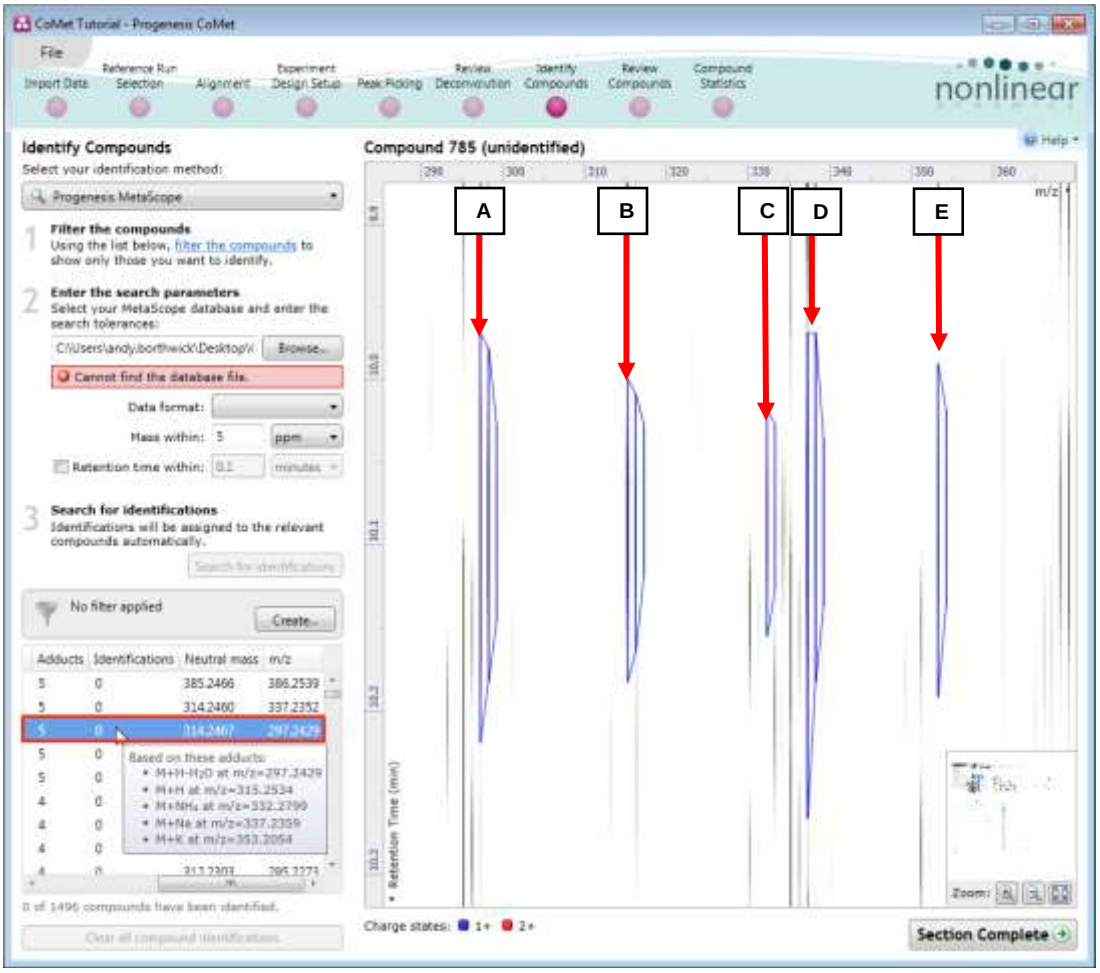
Click **Edit** and **Clear the Filter** then **OK** so that no Tag filter is applied.



Now 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. At this stage the view displays the current compound's adducts, in this example there are 6 adducts:



A. M+H-H_2O m/z = 297.2429

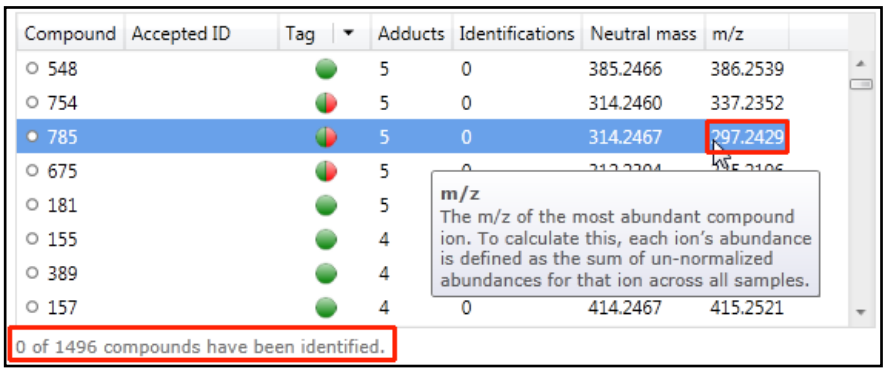
B. M+H m/z = 315.2534

C. M+NH_4 m/z = 332.2799

D. M+Na m/z = 337.2359

E. M+K m/z = 353.2054

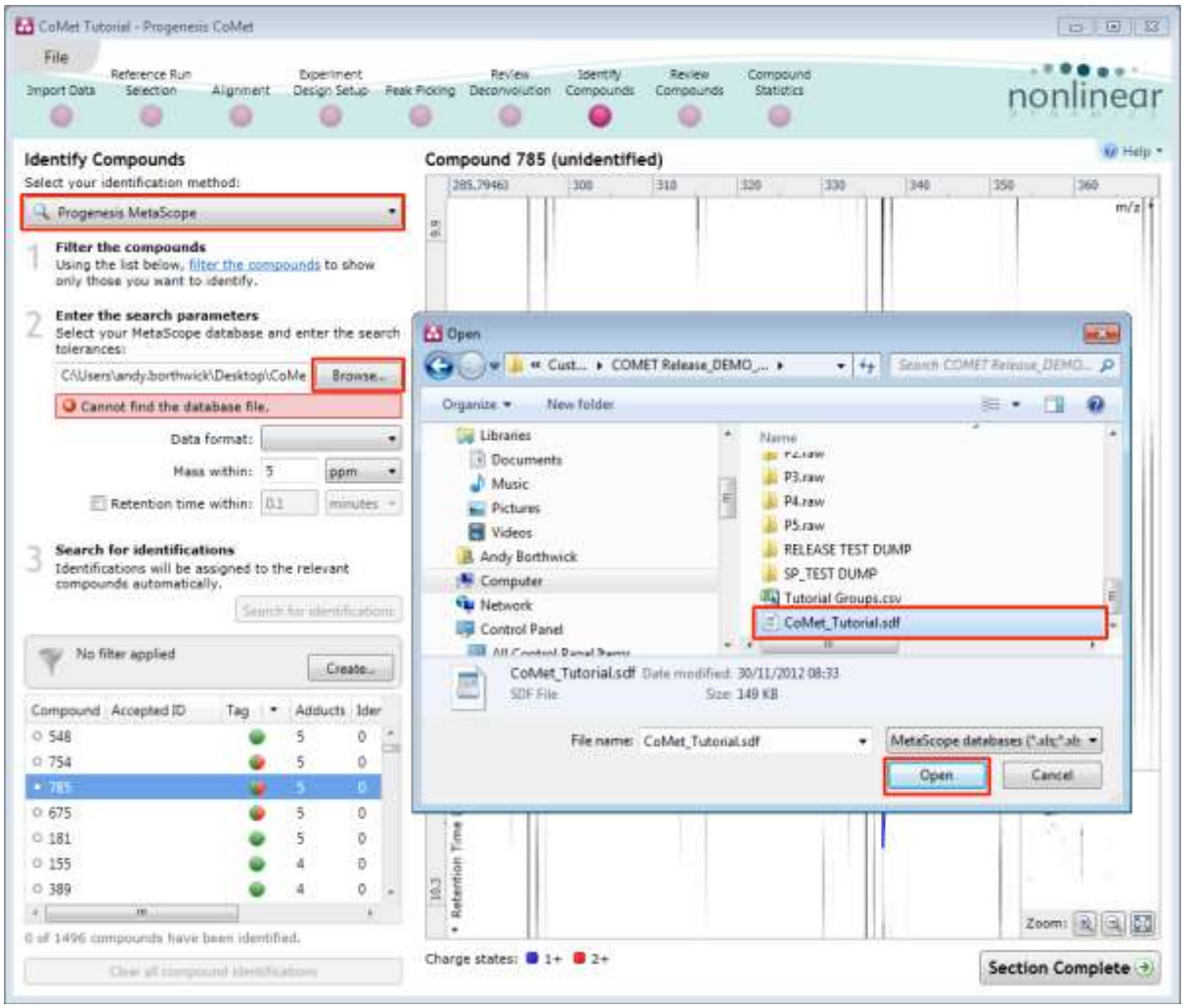
F. which correspond to compound 760 which has a neutral mass of 314.2467 as indicated in the table.



Note: the m/z displayed is for the most abundant compound ion (in this case the M+H-H_2O adduct).

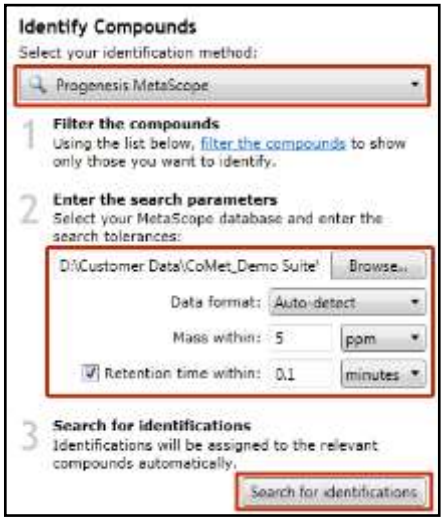
Currently there are 1496 compounds of which none have been identified.

To identify compounds 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 and Retention time as required and .SDF format files.



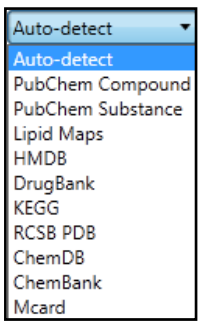
In this example we will use **CoMet_Tutorial.SDF,** a small database of compounds where you can set thresholds for Mass and Retention time as required.

Having selected the identification method as **Progenesis MetaScope**.



Browse to locate the database file which will be in the folder you extracted the CoMetTutorial.zip.

Select: Comet_Tutorial.sdf

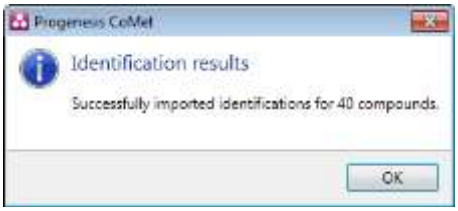


There are a number of Data Formats available including **Auto-detect** which we will use here.

Set the **Mass** threshold: 5ppm and the Retention Time to 0.1 min

To perform the search click **Search for identifications**.

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

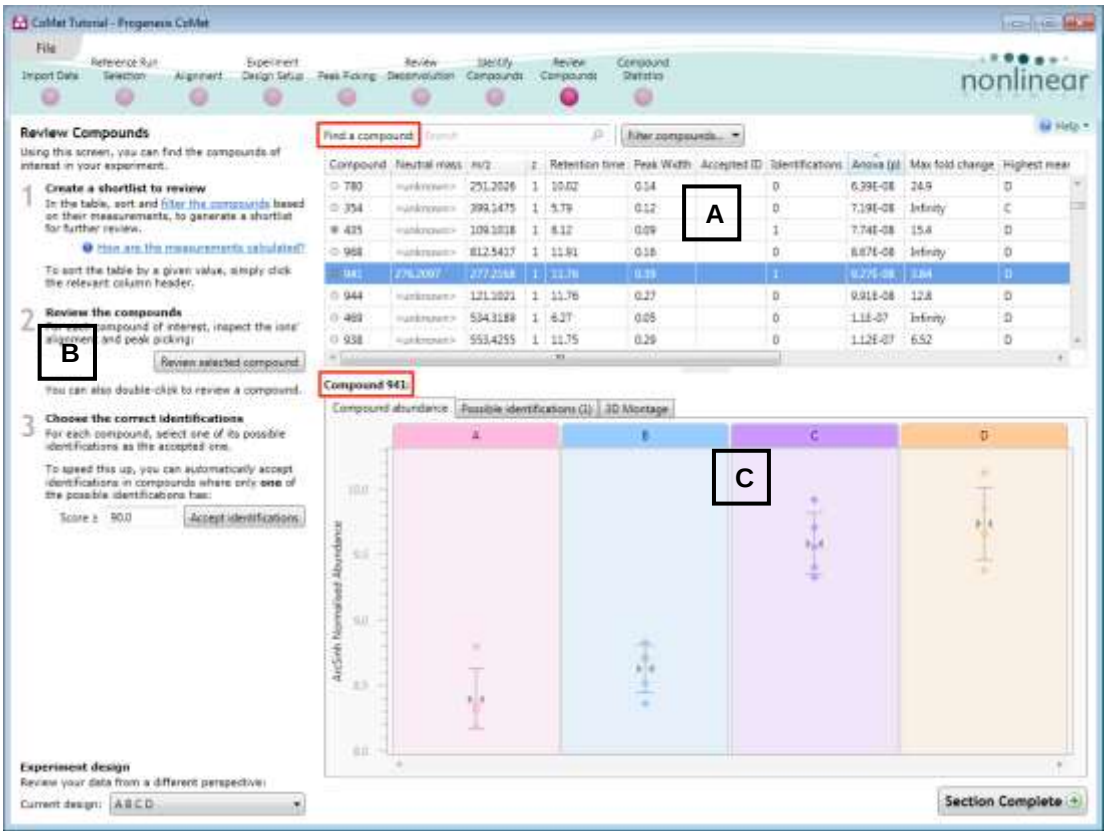


Also the Table will update the IDs column displaying the number of identifications imported for each Compound.

Having imported the search results click **Section Complete** to move to the **Review Compounds** stage.

Stage 9: Review Compounds

The Review Compounds opens allowing you to examine and confirm identities of the compounds.



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

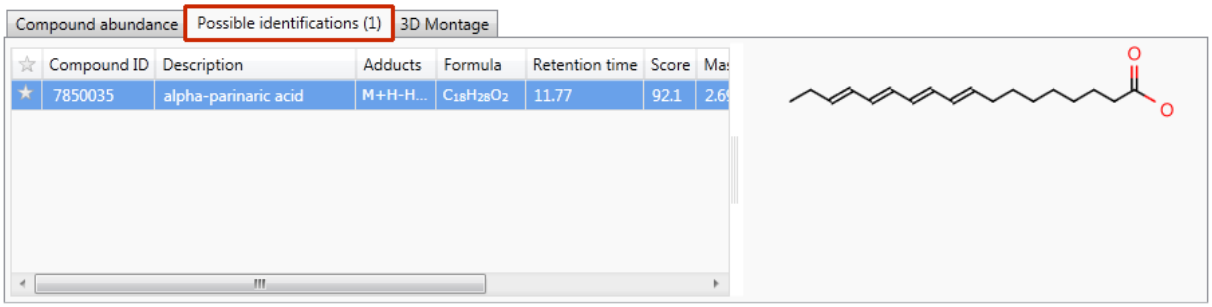
Window B Provides tools to set thresholds for the acceptance of identifications based on a score.

Window C displays the **Compound abundance** plot, a list of **Possible identifications** and a **3D Montage**, for the current compound highlighted in Window A.

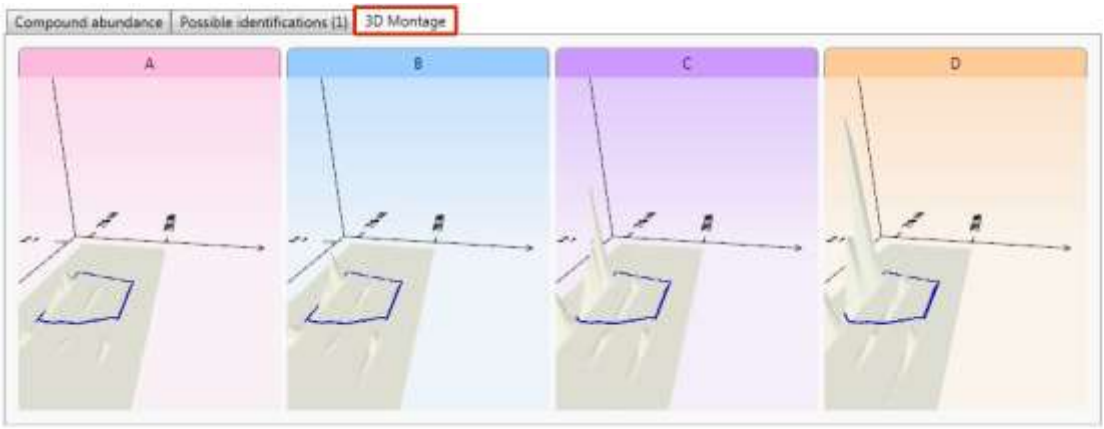
Using the Find a compound: explore **compound 941** in the 3 views for Window C:

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

Possible identifications for current compound, including Search Engine score, Mass error and isotope similarity based on the comparison of the measured isotope distribution for the compound vs the expected based on the compound formula. The current compound structure is also displayed.



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

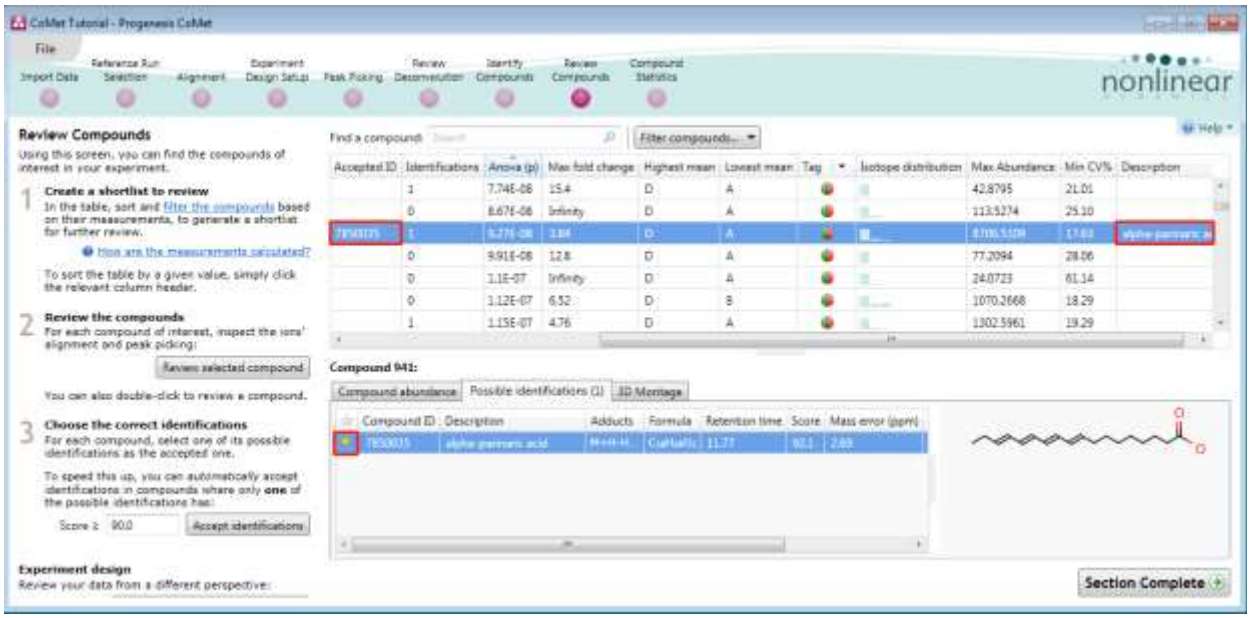


Setting a Compound's identity

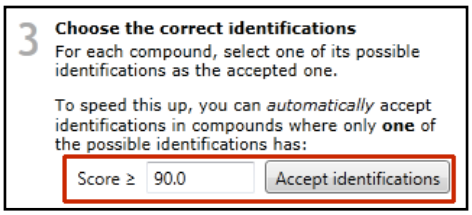
Based on your acceptance criteria, following review of these measures, you can choose to 'Accept' the identity of the compound by clicking on the **star** at the start of the row, or leave it greyed out.

As an example, click on the 'star' to accept **Alpha-parinaric acid** as the identity of compound 941.

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.

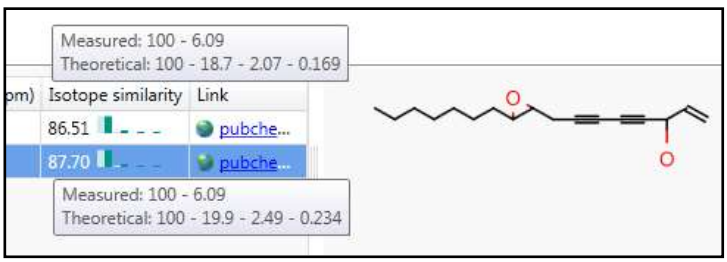


As an example use the default threshold of 90 and then order on **Accepted ID**

Now review a compound with multiple IDs and an entry in the Accepted ID column (i.e. compound 1091)

Note: no Accepted ID has been entered as both possible identifications have a score ≥ 90

Note: Use the cursor tooltips to compare the Measured with the Theoretical Isotope distributions for each compound.



You can use the comparative **Scores**, **Isotope Similarity** and **Structure** to decide on the Compounds identity. Click the Star to assign identity.

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:

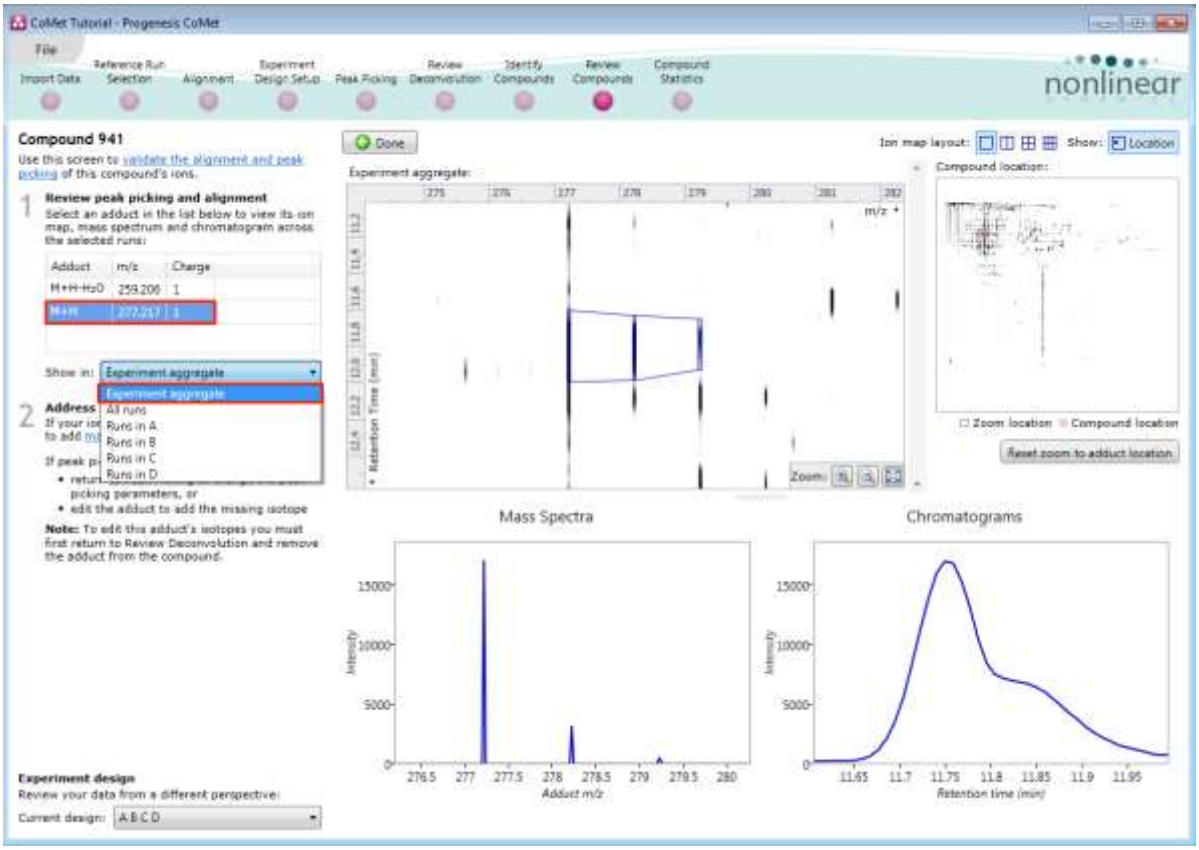
- A montage of ion maps helps to validate the ions' peak picking and alignment
- Graphs of mass spectra and chromatograms provide further confirmation of correct peak picking

To open this view either double click on the current compound in the **Review Compounds** screen or click on **Review selected compound.**

Try double clicking on compound **941** identified as **Alpha-parinaric acid.**

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.



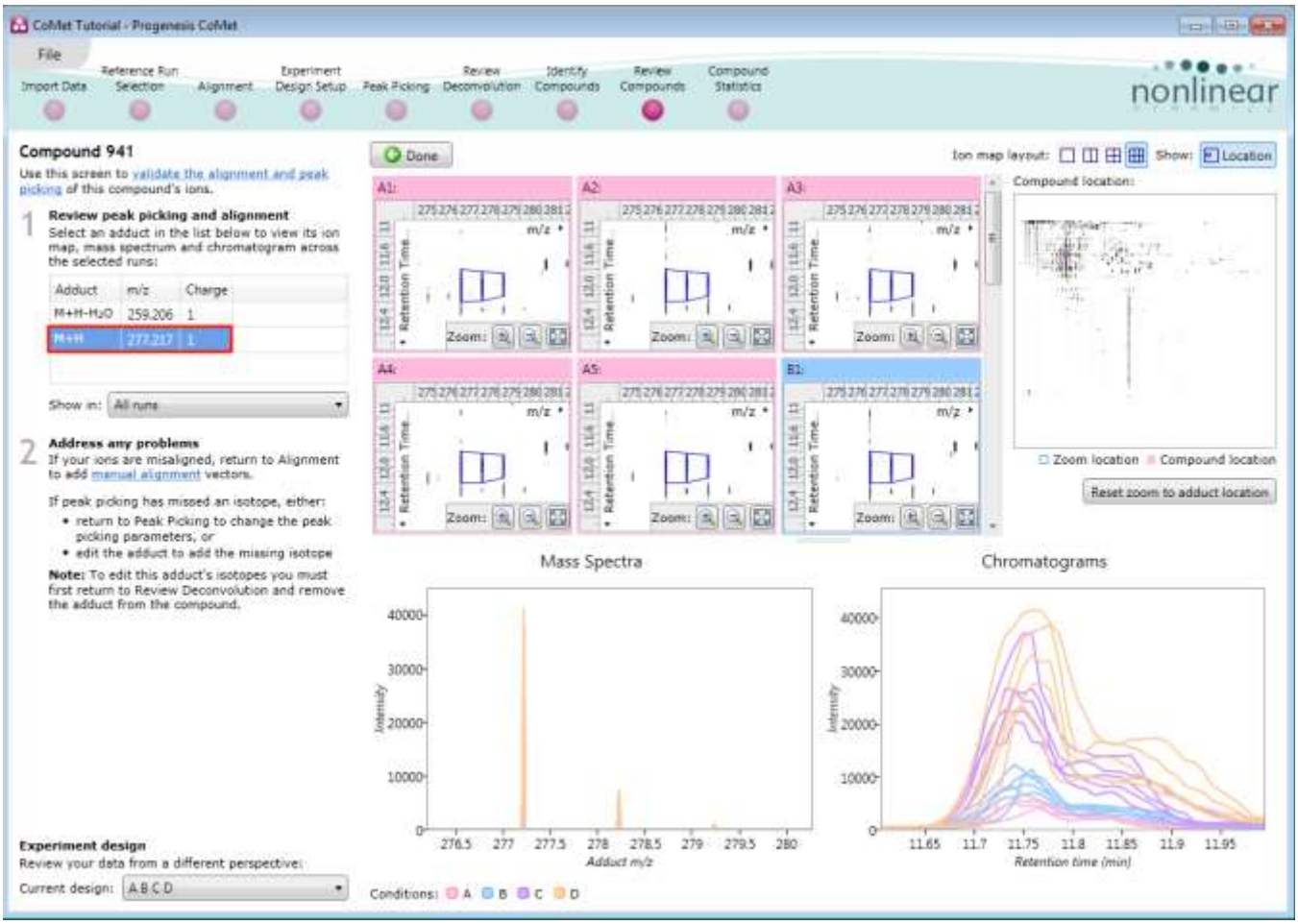
Below the list of adducts, a drop-down list lets you select whether you want to see the selected adduct:

- 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 he **Edit adduct** facility.

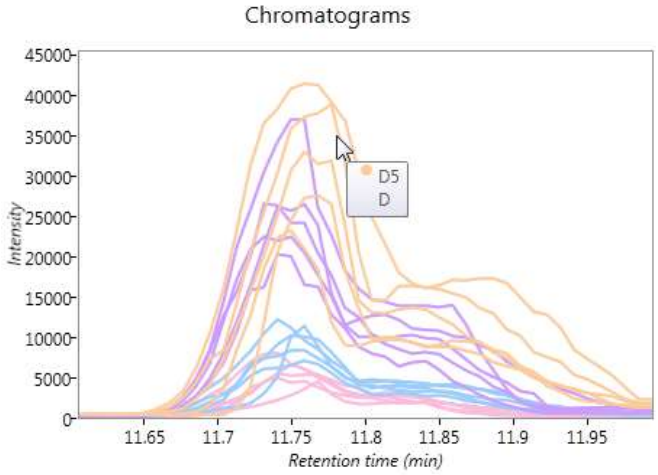
Note: More details on Editing adducts is available in Appendix 5 (page 59)

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.

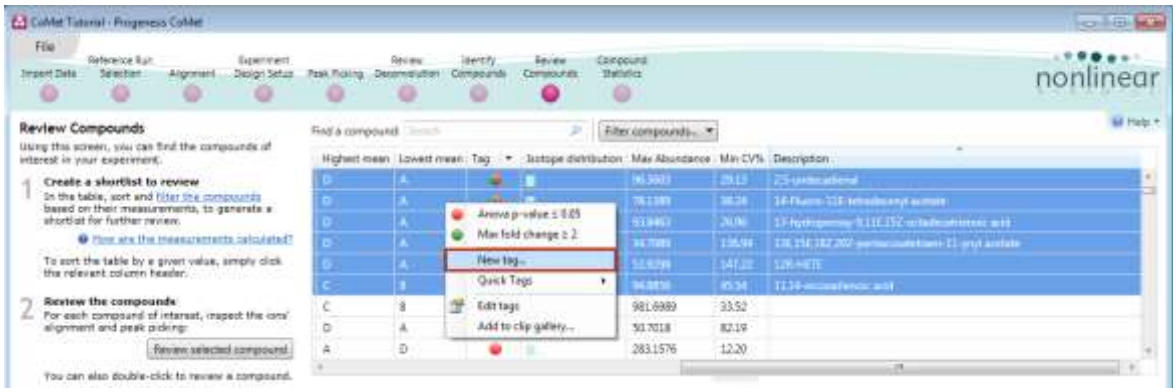
Note: that when the **All runs** is selected the mass spectra and chromatograms are shown for all the runs in the experiment.



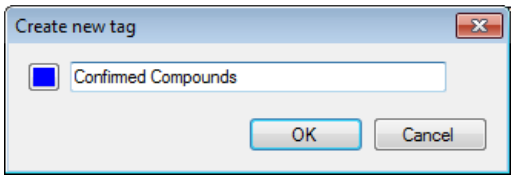
In the chromatograms view each line corresponds to a single run, coloured according to the experiment condition to which it belongs. If you hold the cursor over a line, the name of the run which generated it will appear. When showing all runs, these graphs give further validation of how the adduct form's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

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.



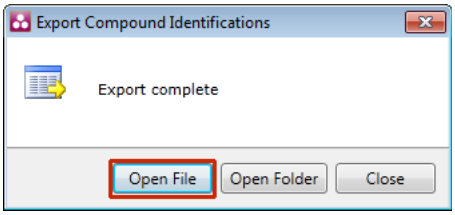
Then right click on the highlighted compounds and create a New tag called **'Confirmed Compounds'**

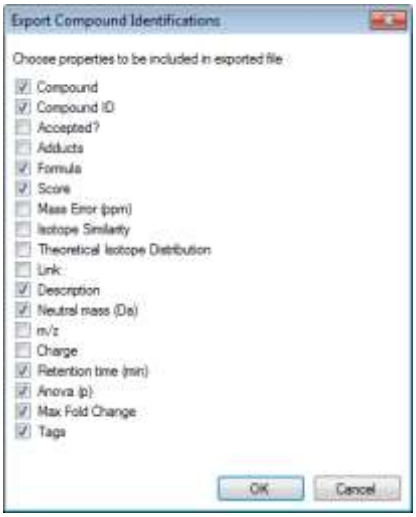


Drag the new tag onto the Show panel of the **Filter compounds** dialog and click **OK**.

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:



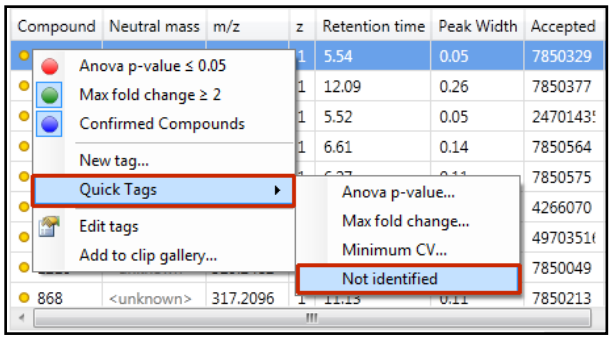


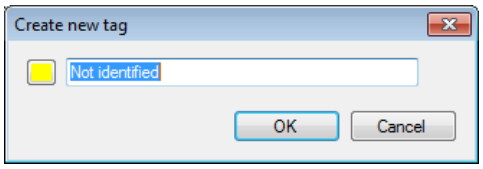
Additional Tags for Progenesis Stats

Tagging can be used to identify groups of compounds with similar properties. These properties can be determined from the ordering and highlighting of compounds in the table at the Review Compounds stage and/or using the **Quick Tags** options. Using combinations of these tags you can easily filter your data for export to external applications (Excel etc).

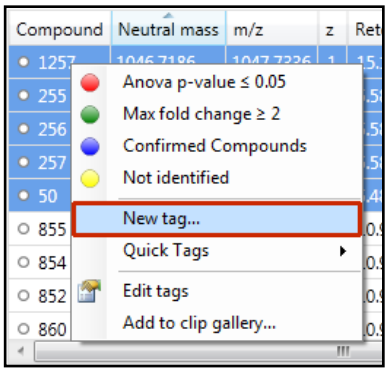
In the previous stage we created Quick tags for **Anova (p)** and **Max fold change**

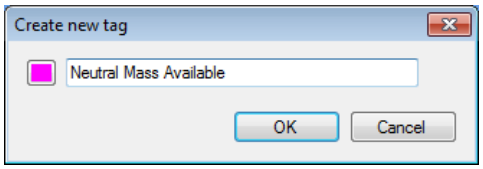
Here we will create 2 additional tags, first a Quick Tag for **Unidentified compounds** by right clicking on the table and selecting **Not identified** from the Quick Tags





Now order the table on **Neutral Mass** and highlight all the compounds which have a value for Neutral Mass. Right click on the highlighted area and select New Tag. Then call the tag **Neutral Mass Available.**





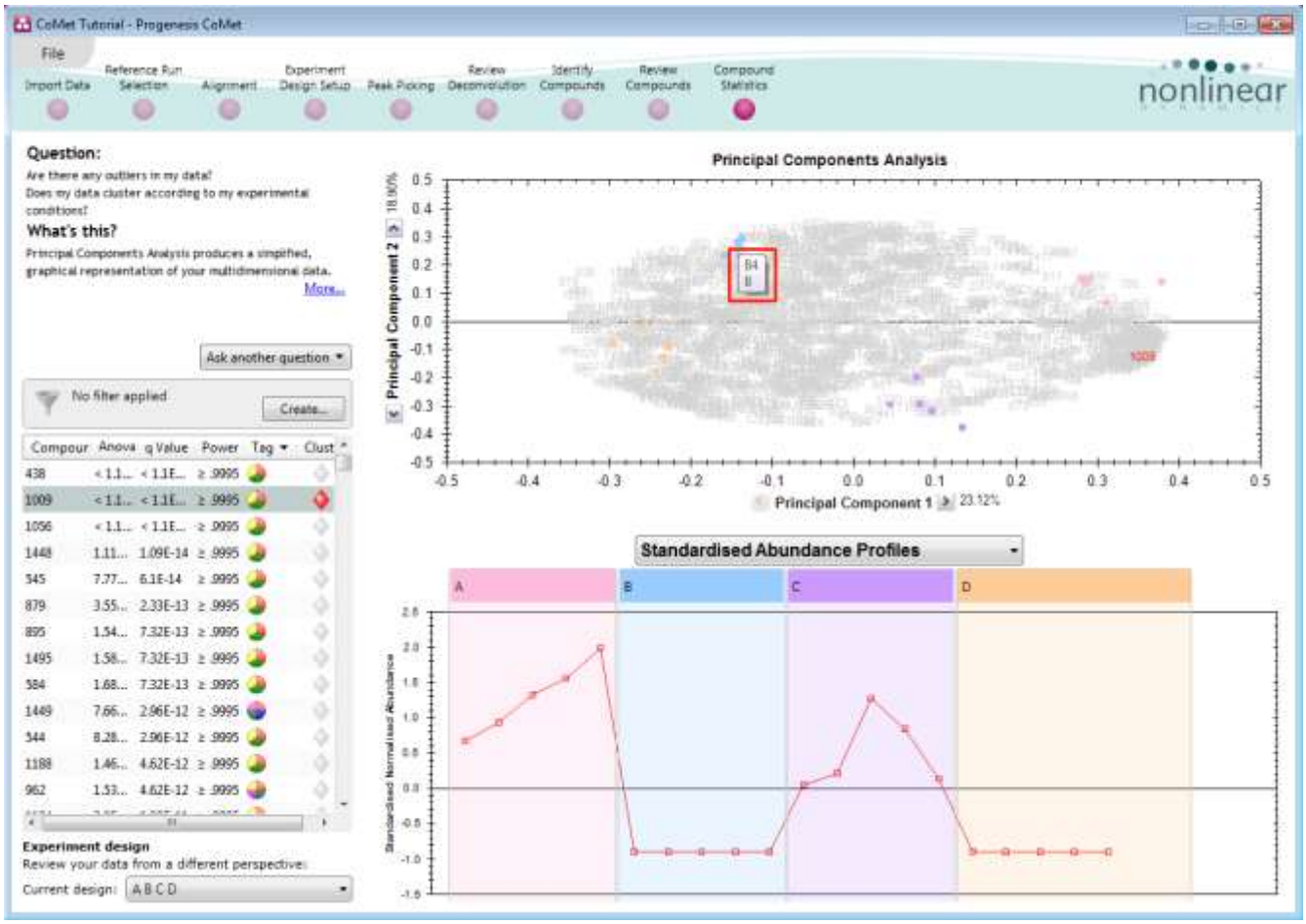
Having created these additional tags click **Section Complete** to move to the **Compound Statistics** stage.

Stage 10: Compound Statistics

The user guide now describes the functionality of the Multivariate Statistics. This section is only available if Progenesis Stats is licensed.

Progenesis 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 1496 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 **Progenesis 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 Stats 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.

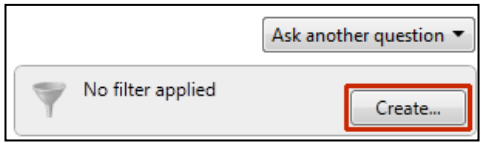
Also, for a given power of 80% we can determine how many samples are required to ensure we find a difference if it actually exists.

Note: Power analysis is discussed in Appendix 7 (page 67)

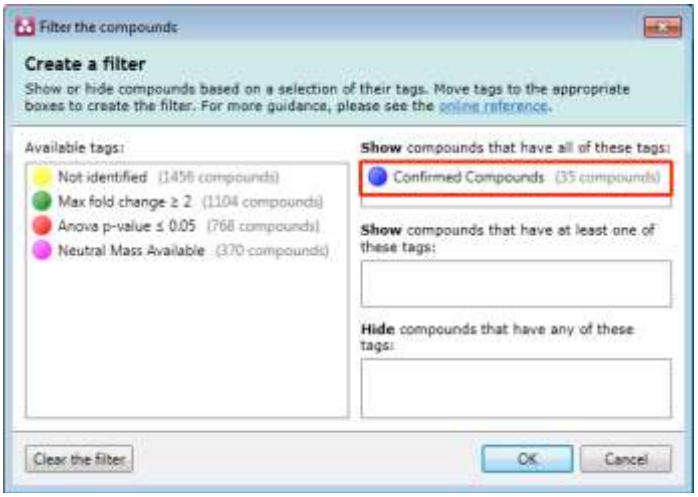
Correlation Analysis

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 Confirmed.

To filter the data click **Create**

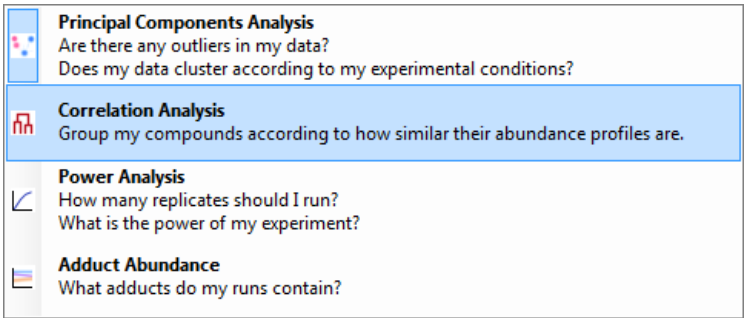


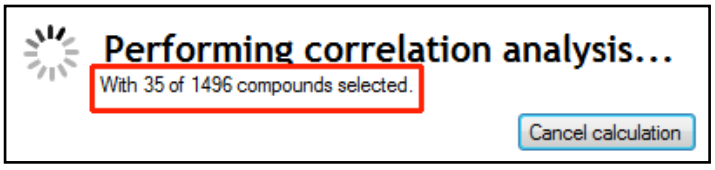
Set up the filter as shown below



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:

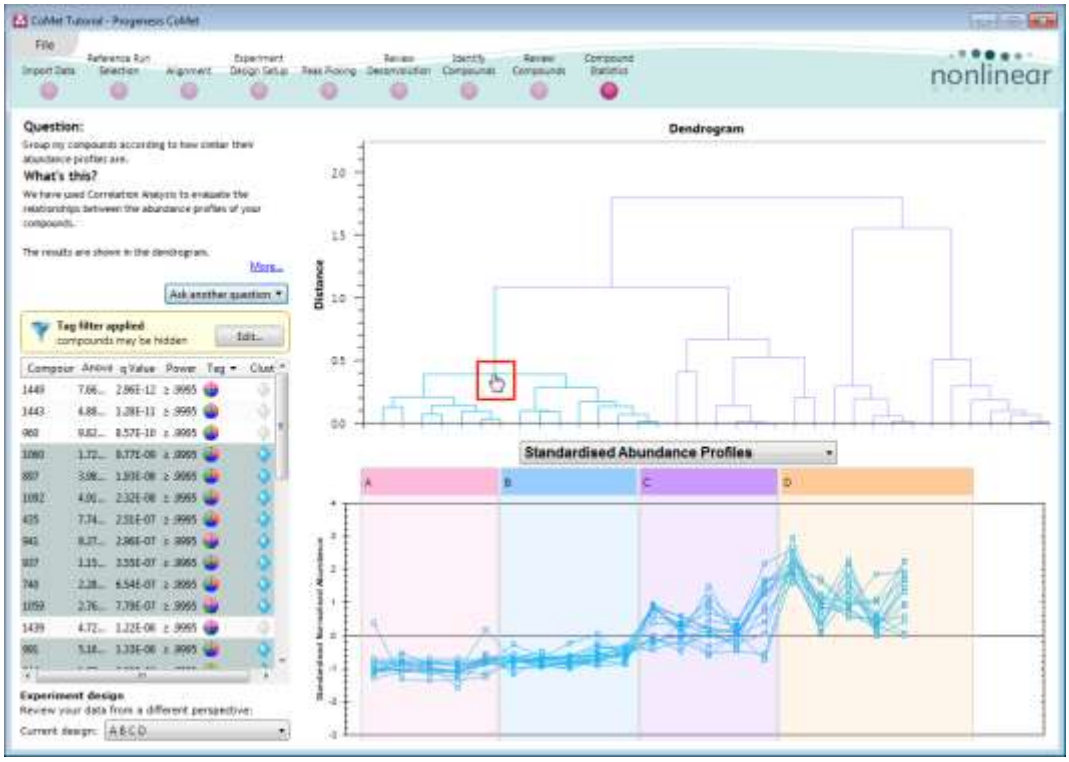
'Group my (selected) features according to how similar their expression profiles are'

The question is answered by:

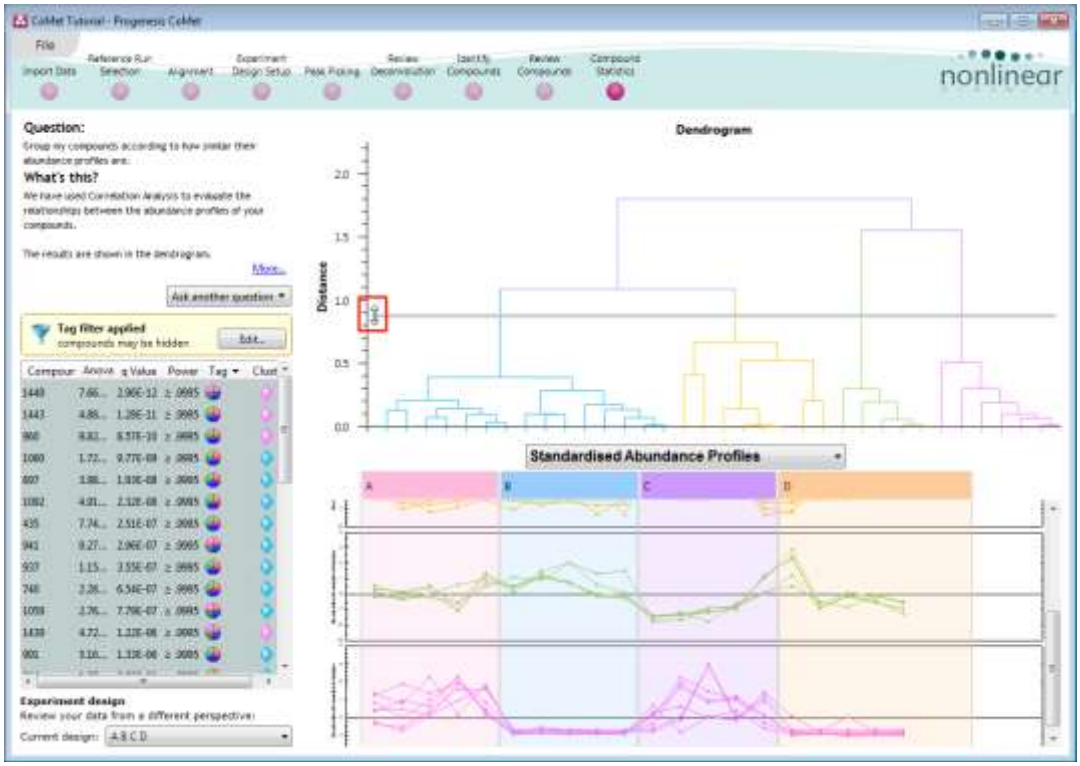
'Using Correlation analysis to evaluate the relationships between the (selected) features' expression profiles'.

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.



Finally use the slider (located at the top of the dendrogram) to highlight multiple branches by dragging it down.

Each highlighted branch will have an Abundance Profile coloured the same.

Congratulations!

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

Hopefully our example has shown you how this unique technology can deliver significant benefits with

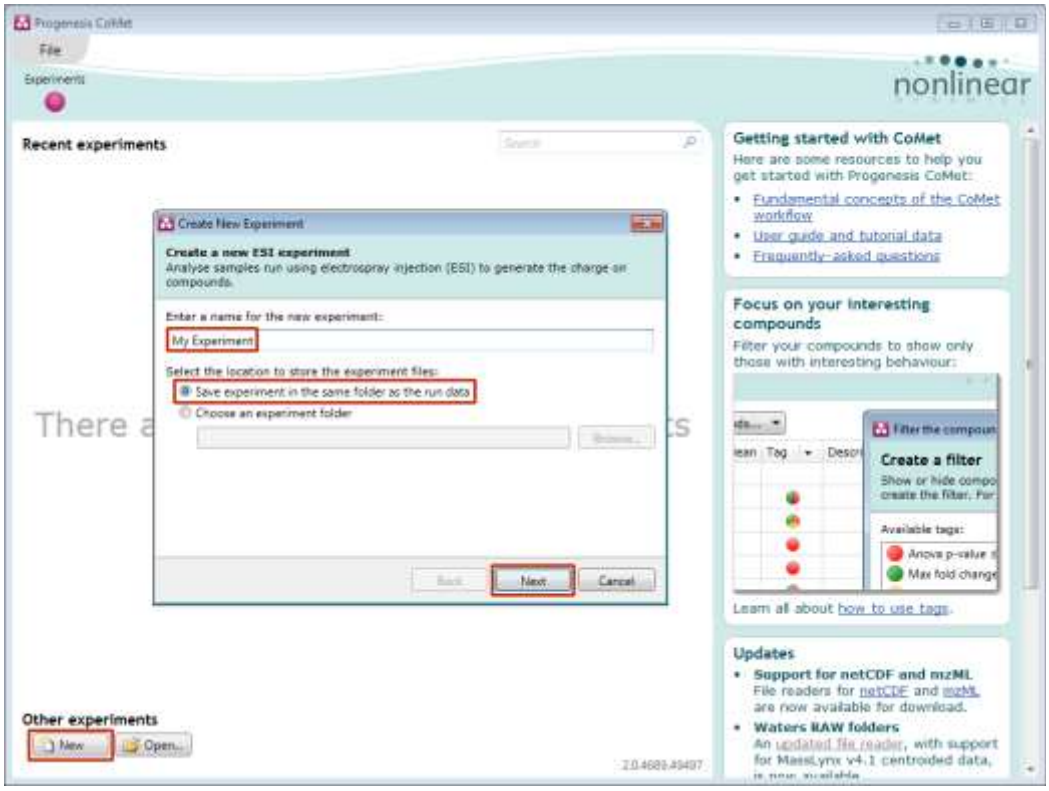
- Speed
- Objectivity
- Statistical Power

If you would like to see the benefits of running Progenesis CoMet using your own runs and explore the Progenesis CoMet workflow please go to Appendix 3: Licensing Runs (page 51).

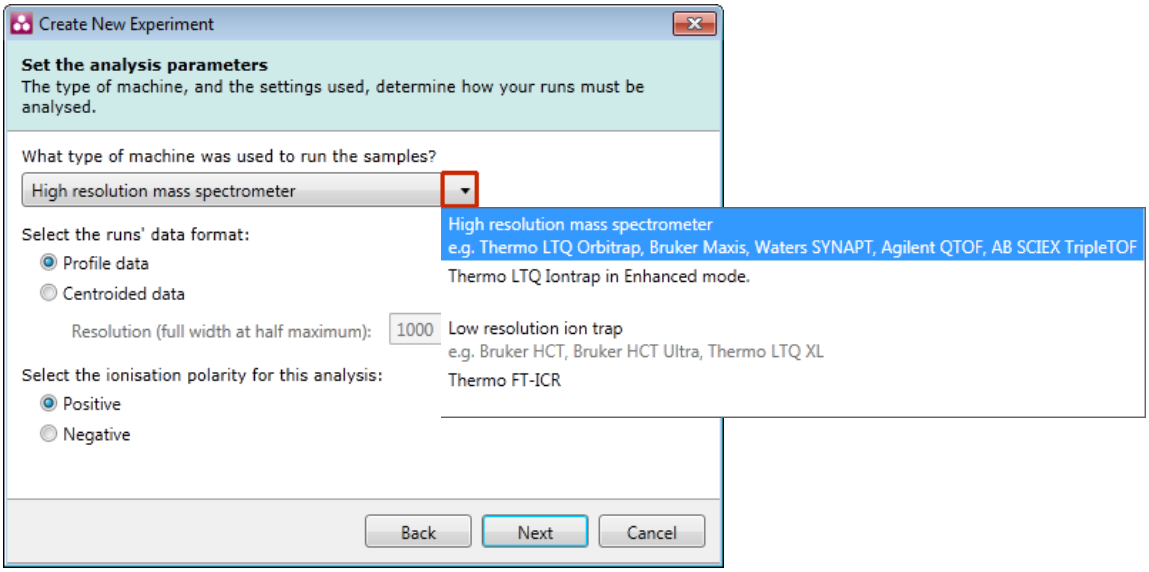
Appendix 1: Stage 1 Data Import and QC review of CoMet data set

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

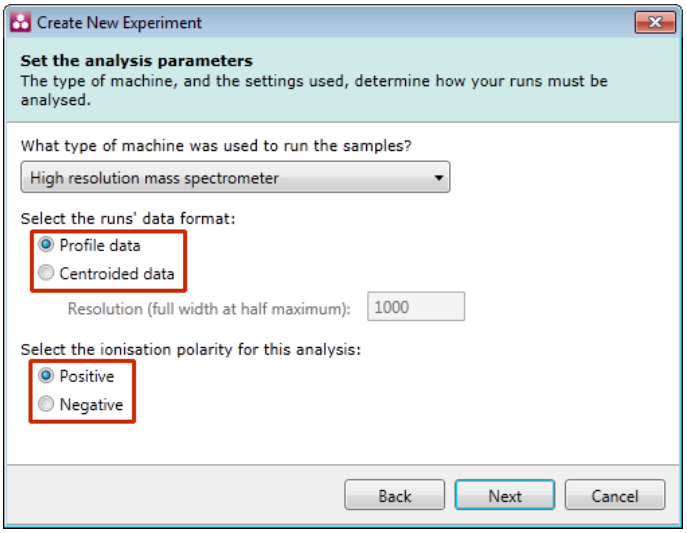
To create a new experiment with your files select **New** give your experiment a name and then select a location to store the experiment files. Click **Next**.



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

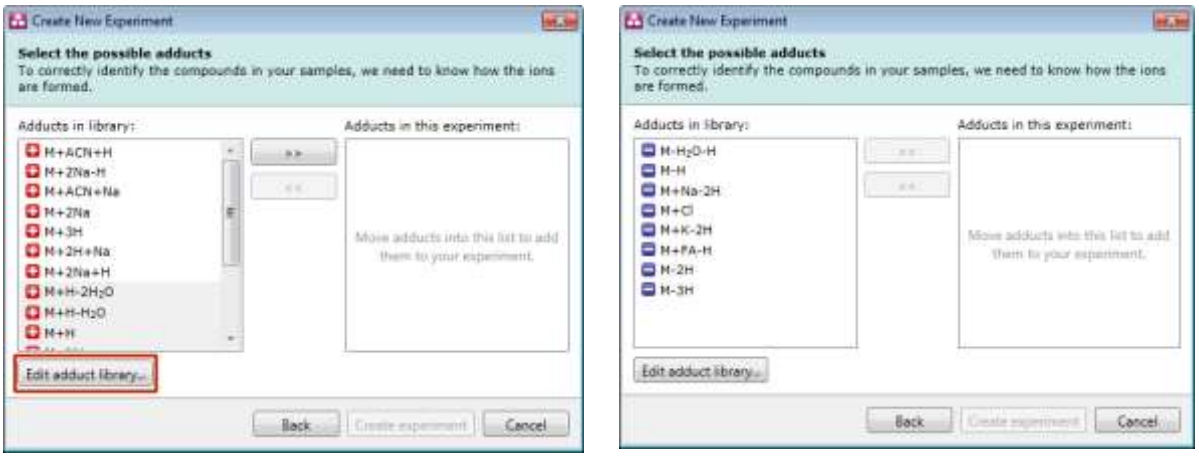


Note: if you have converted or captured the data as centroided then select Centroided data and enter the Resolution (full width at half maximum) for the MS machine used.

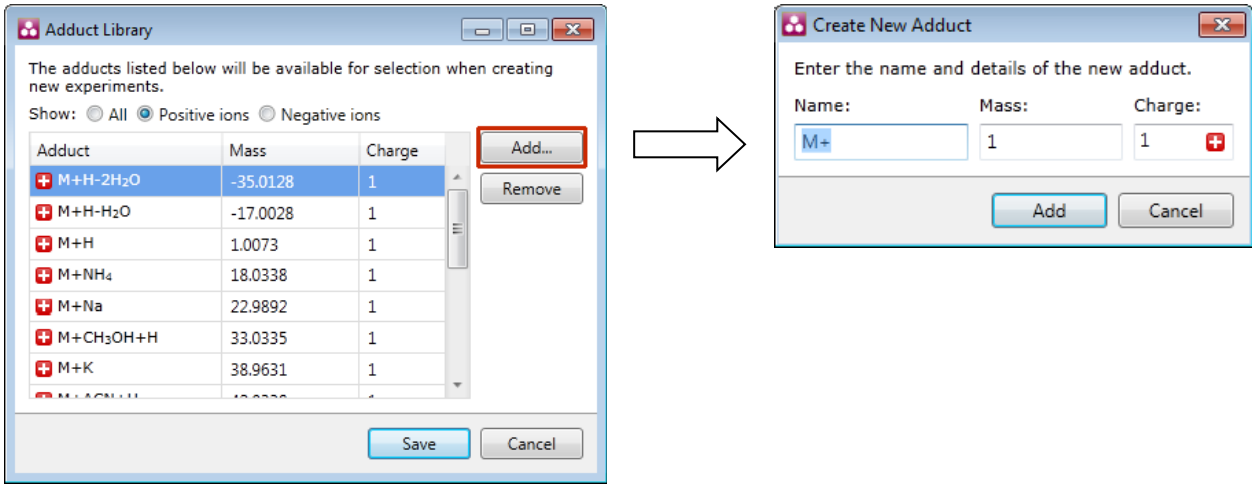


In this example the data was captured in profile mode.

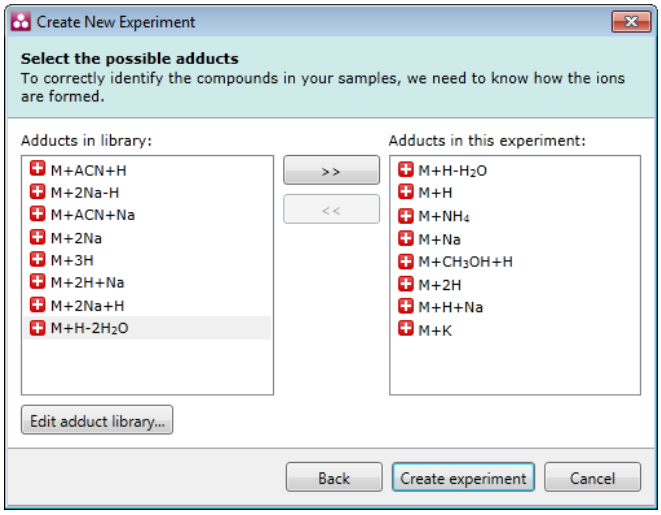
Finally select the ionisation polarity, either **positive** or **negative**, this will determine which list of possible adducts is available on the next page.



To add an adduct to the library select **Edit adduct library…** and then click **Add**. Now provide the Name, Mass and Charge of the new adduct and then add it to the library.



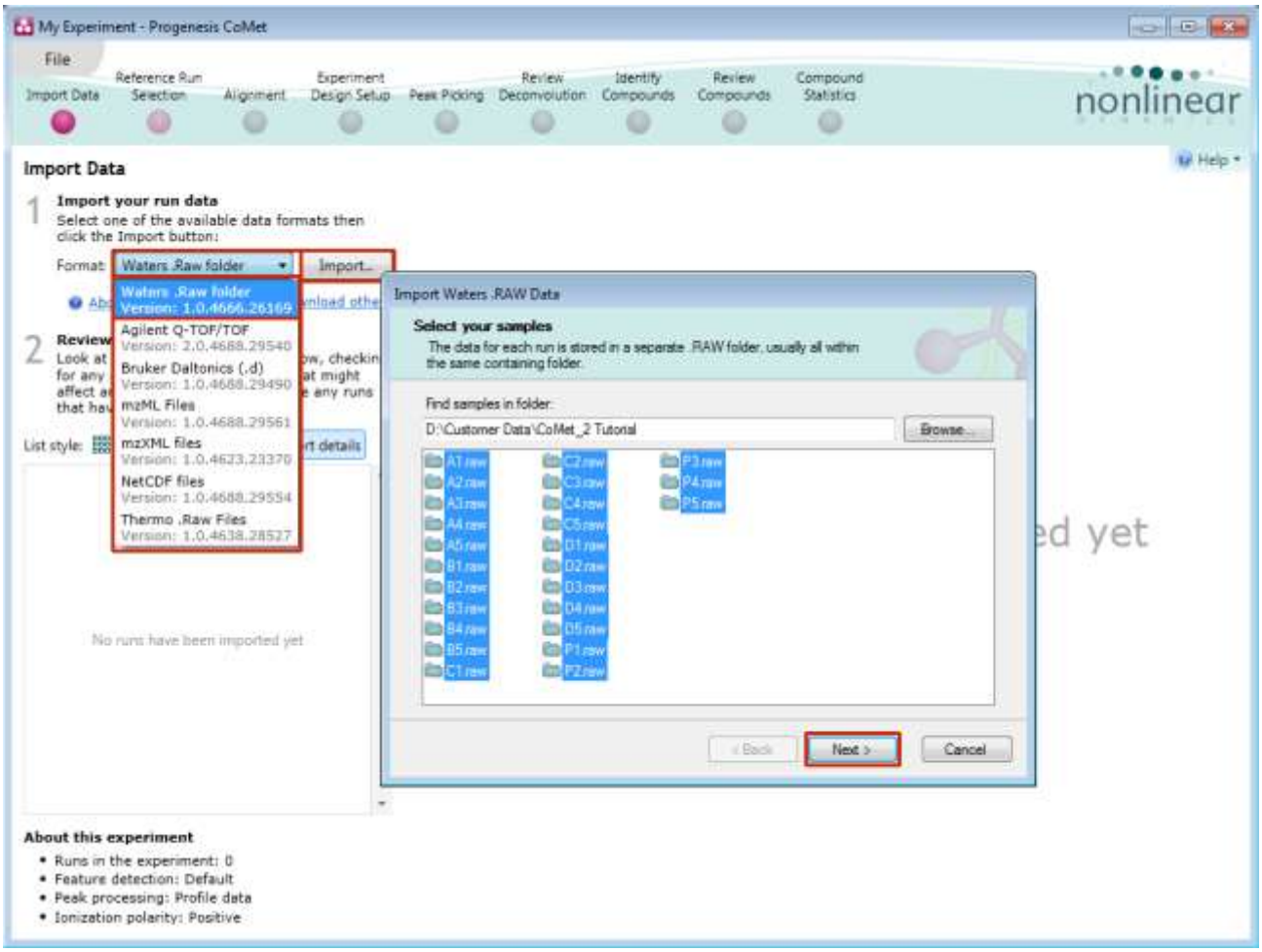
Now select the expected adducts based on you knowledge of your experimental conditions. In this example the polarity was positive and a basic set of Adducts were selected as shown on the right panel below.



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

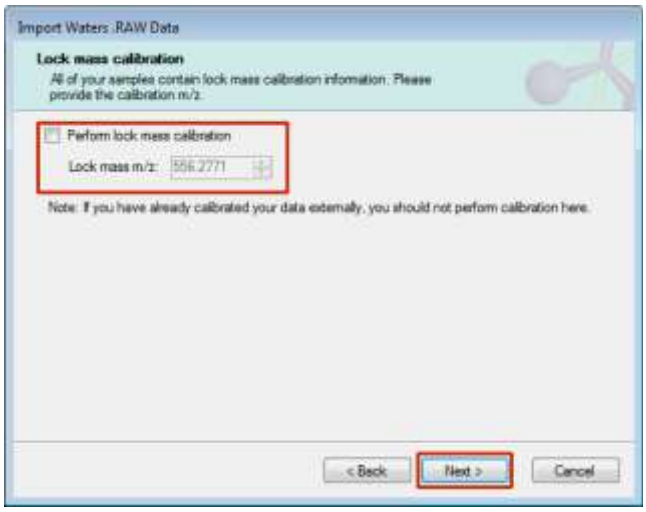
Select the 'Import Data file format', in this example they are **Waters.Raw folders**

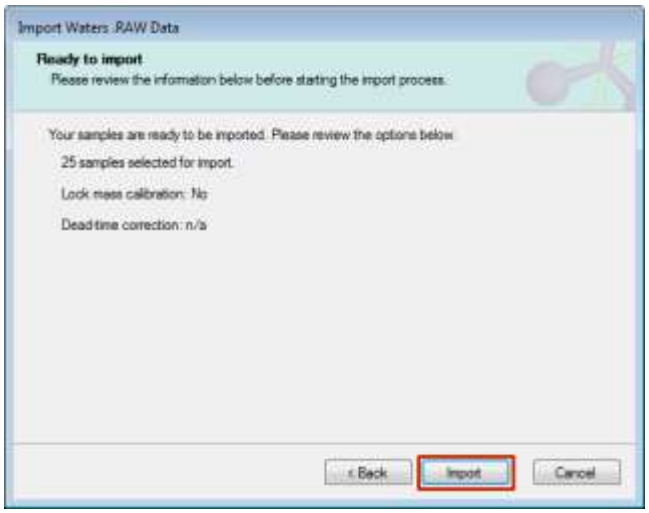
Then locate your data files using **Import...**



Locate and select all the Data files, in this example (A1 to P5).

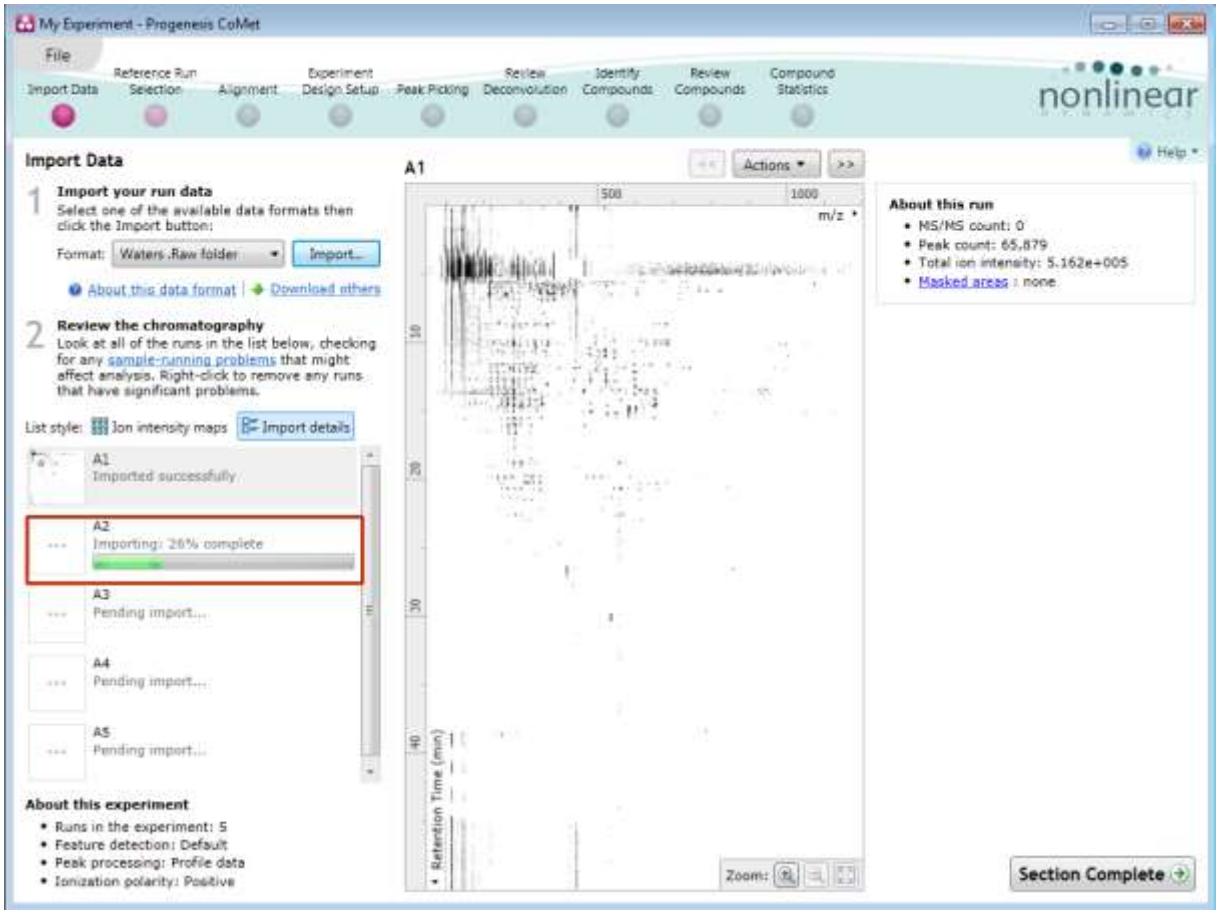
Set the Lockmass calibration (required for Waters data). Note in some cases the Lockmass calibration may have been performed, in which case leave the calibration option unticked.





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

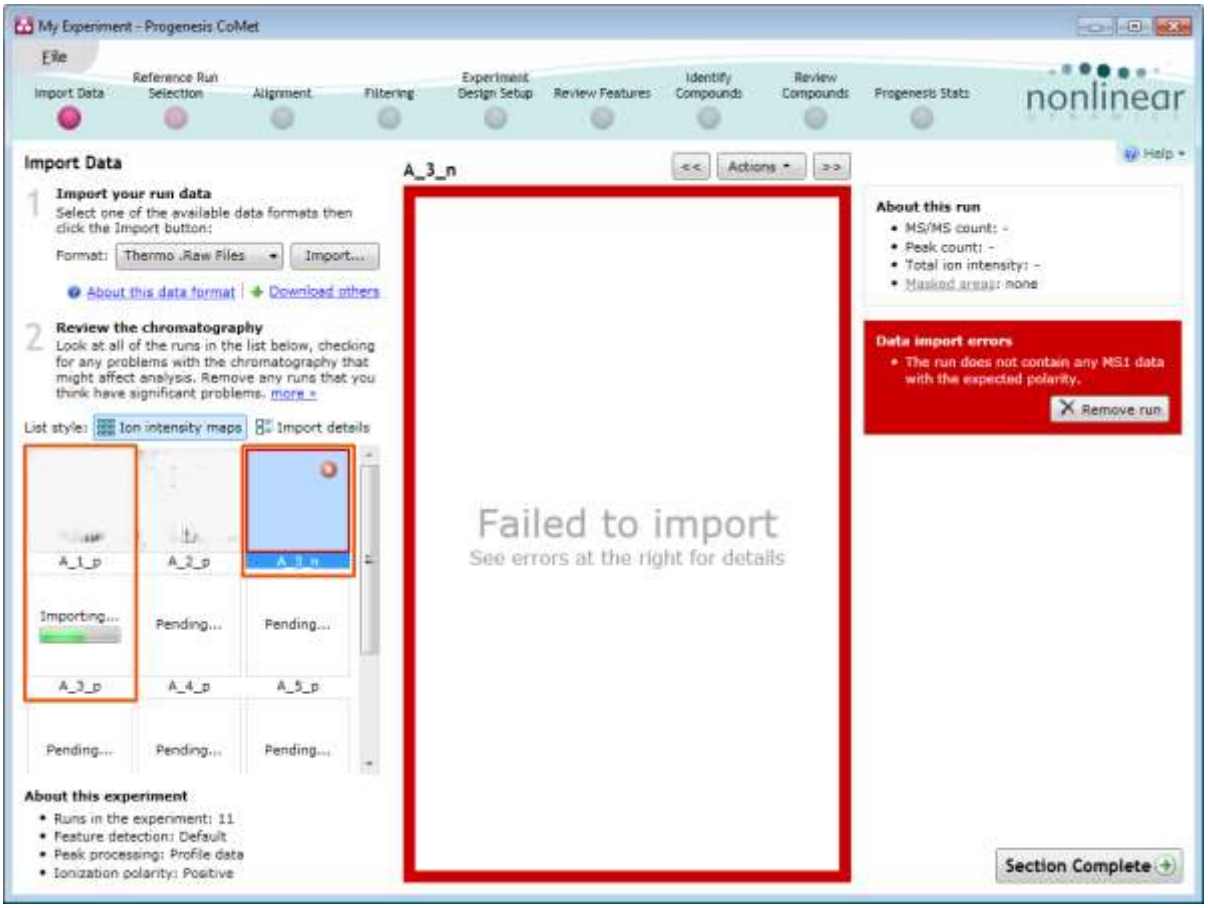
Progress bars report the importing of each file.



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 failed to be imported and a possible reason for the failure.



Note: as each data file is loaded the progress is reported in the **Import Data** list with a progress bar for the file currently being imported and a 2D representation of the successfully imported files.

The dialog to the right of the main panel provides details of the current file and if it has failed import it can be removed by clicking **Remove run** in the Data import errors panel. In this example it had the incorrect polarity.

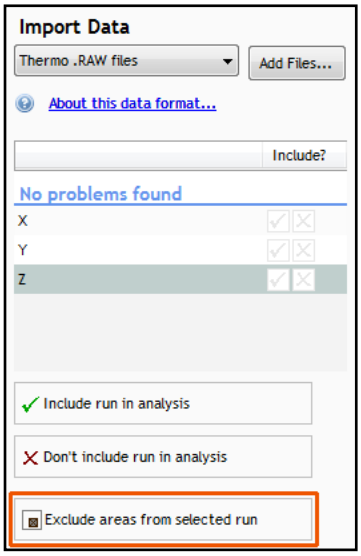
On the bottom left of the Import data list are the experiment details.

Now move to the next stage in the workflow (page 8 in this user guide) by clicking **Section Complete.**

Appendix 2: Stage 1 Data QC review and addition of exclusion areas

During the process of Data QC you may identify areas of the raw data for a particular run that appear 'noisy' yet still have identifiable 'isotopic patterns'.

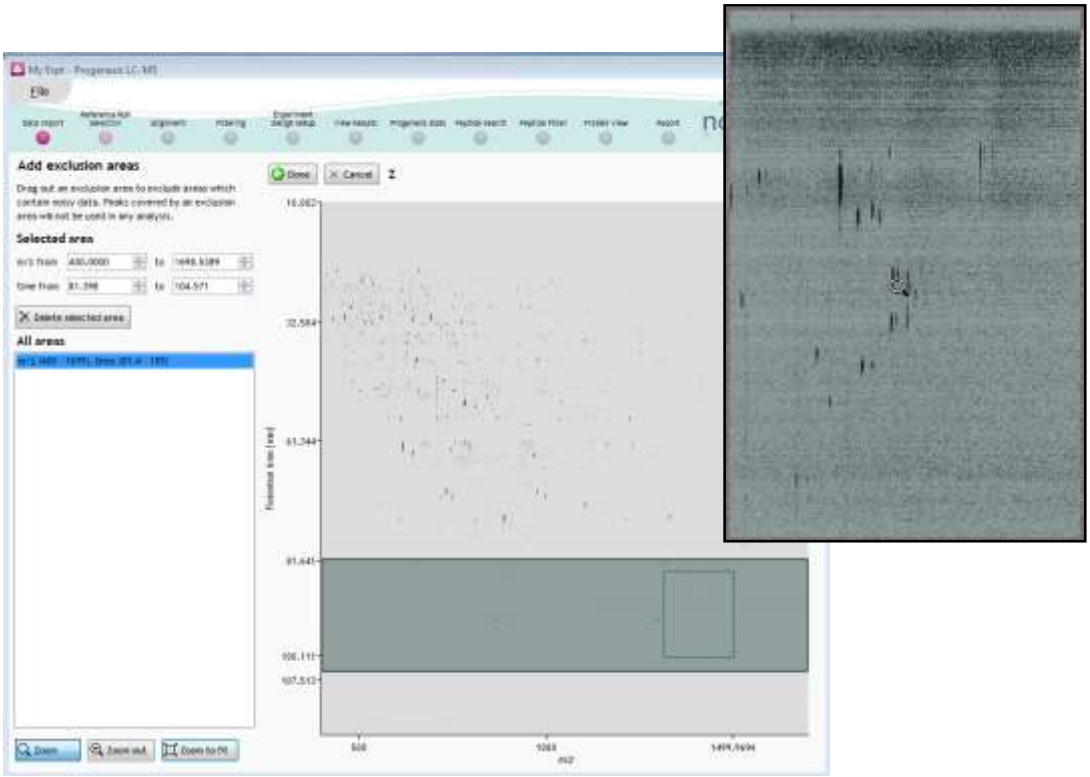
For example if the run is part of a 'replicate set' of runs it is possible to exclude such areas on the noisy run by applying a mask to the area. By doing so this area is excluded during the initial part of the detection process in order that it does not 'interfere' with the detection of the features in the replicate group.



To do this select Exclude areas from selected run on the bottom left of the screen.

Drag out an area over the noisy part of the run to create the mask.

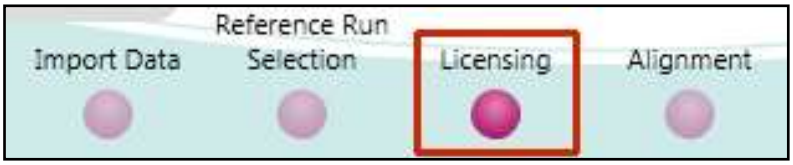
Note: if you now zoom into the masked area using the **Zoom** tool you will see the isotopic features in the noise.



Note: if the level of noise is high and affecting many of your runs a preferred approach would be to re-optimise the chromatography to improve the levels of noise in your data

Appendix 3: Licensing runs (Stage 3)

When setting up a **New experiment** if you are evaluating Progenesis CoMet with unlicensed runs then the licensing page will open after **Reference Run Selection**.



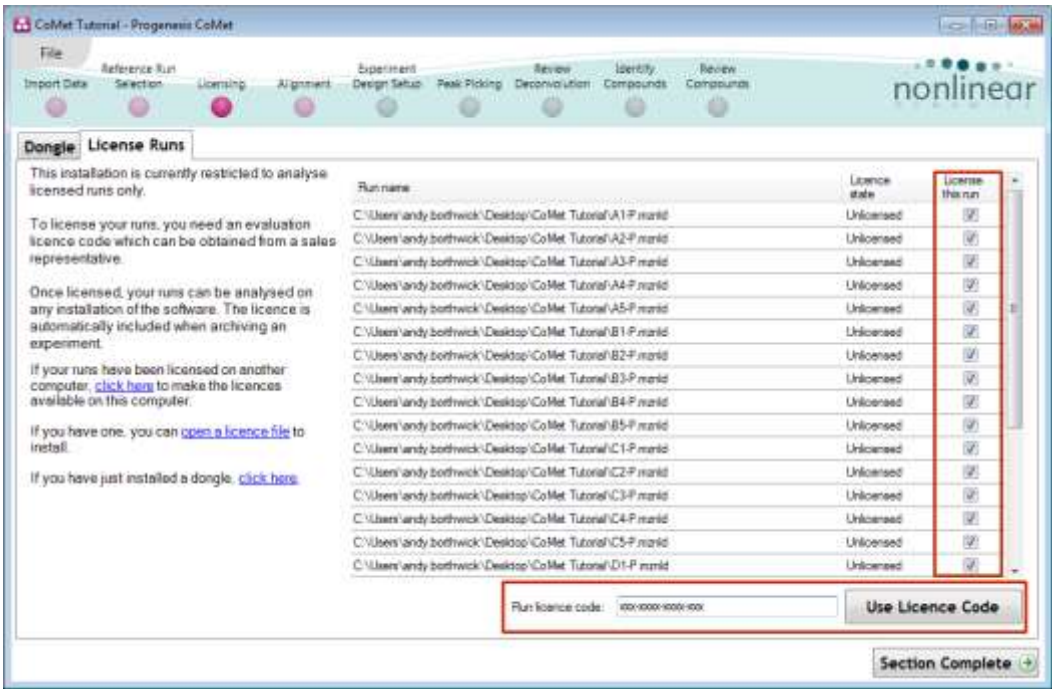
If you already have a programmed dongle attached to your machine then the License Runs page will not appear.

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

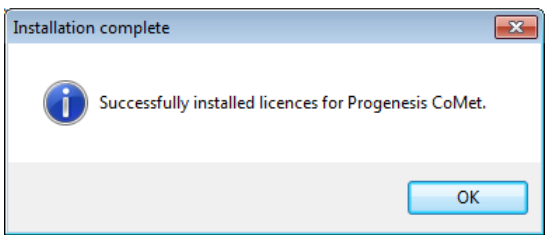
Each code will allow you to license a set number of runs.The runs in your experiment will be listed as shown below.

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

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



A message confirming successful installation of your licences will appear.

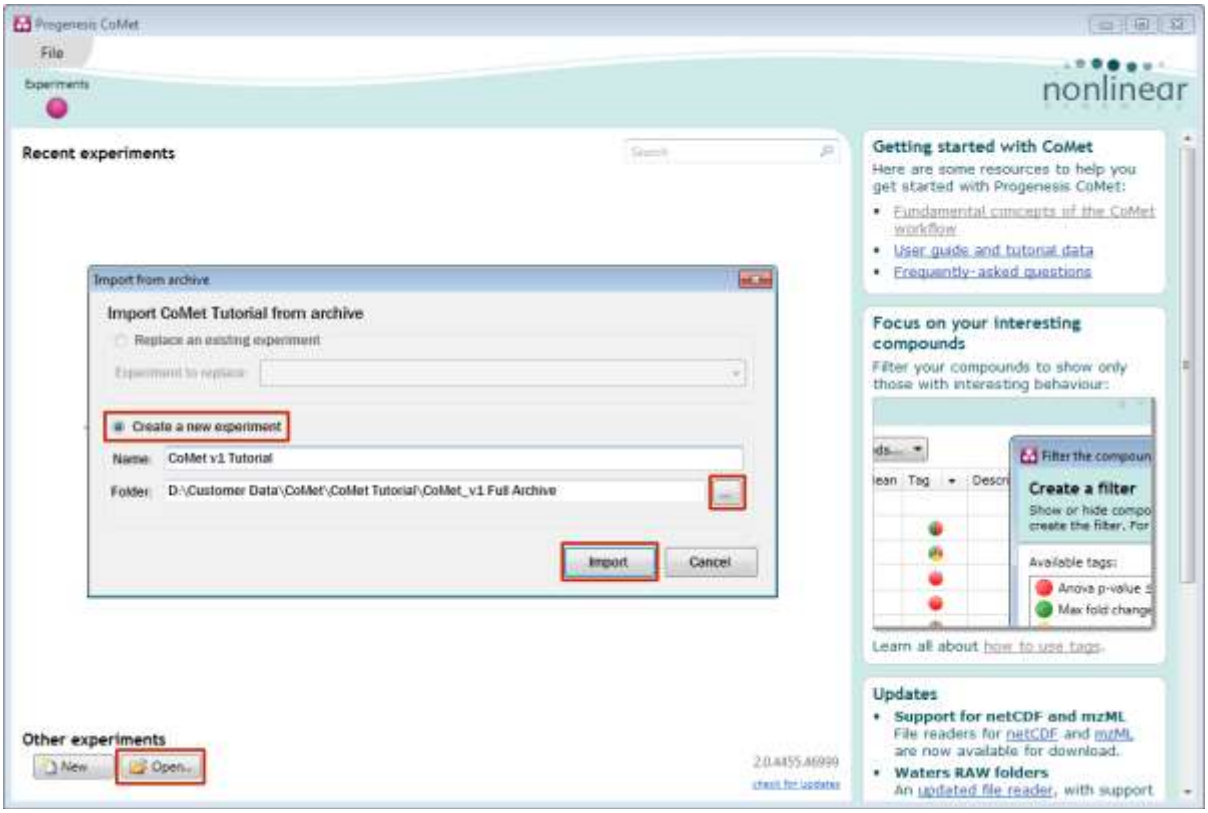


Click **OK**, the view will update, and Alignment, the next stage in the workflow, will open with the licensed files.

Appendix 4: Loading a CoMet version 1 experiment

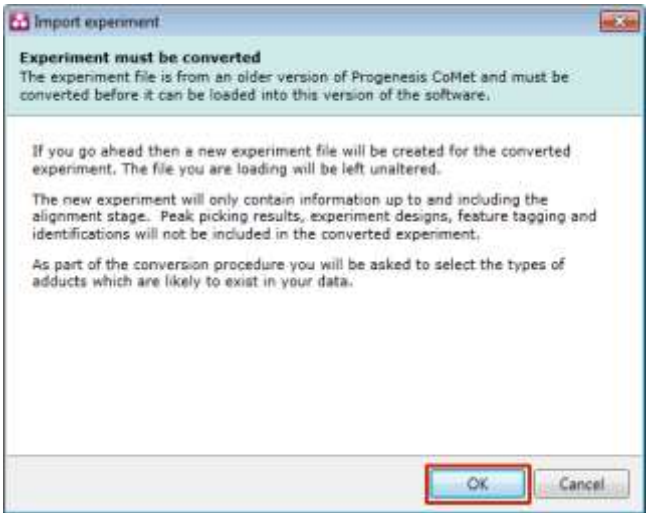
If you have performed analysis of a data set in CoMet version1 you can load this experiment either from restoring an archive or from a CoMet version1 open experiment.

First locate, using the **Open** dialog, the version 1 archive (or current CoMet v1 experiment) then select **Create a new experiment** locate a folder to restore the archive and select **Import** .

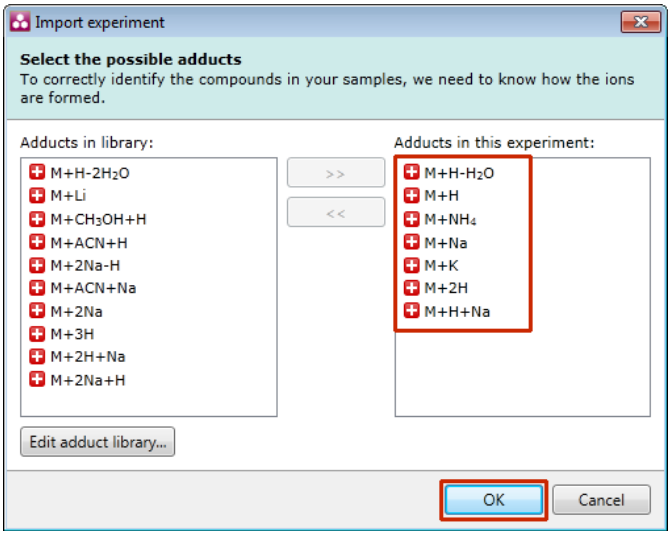


The archive will start to restore then a dialog will warn you that the experiment must be converted to version 2. It also tells you that the new (converted experiment) will open at the end of the Alignment stage. Peak Picking results etc.

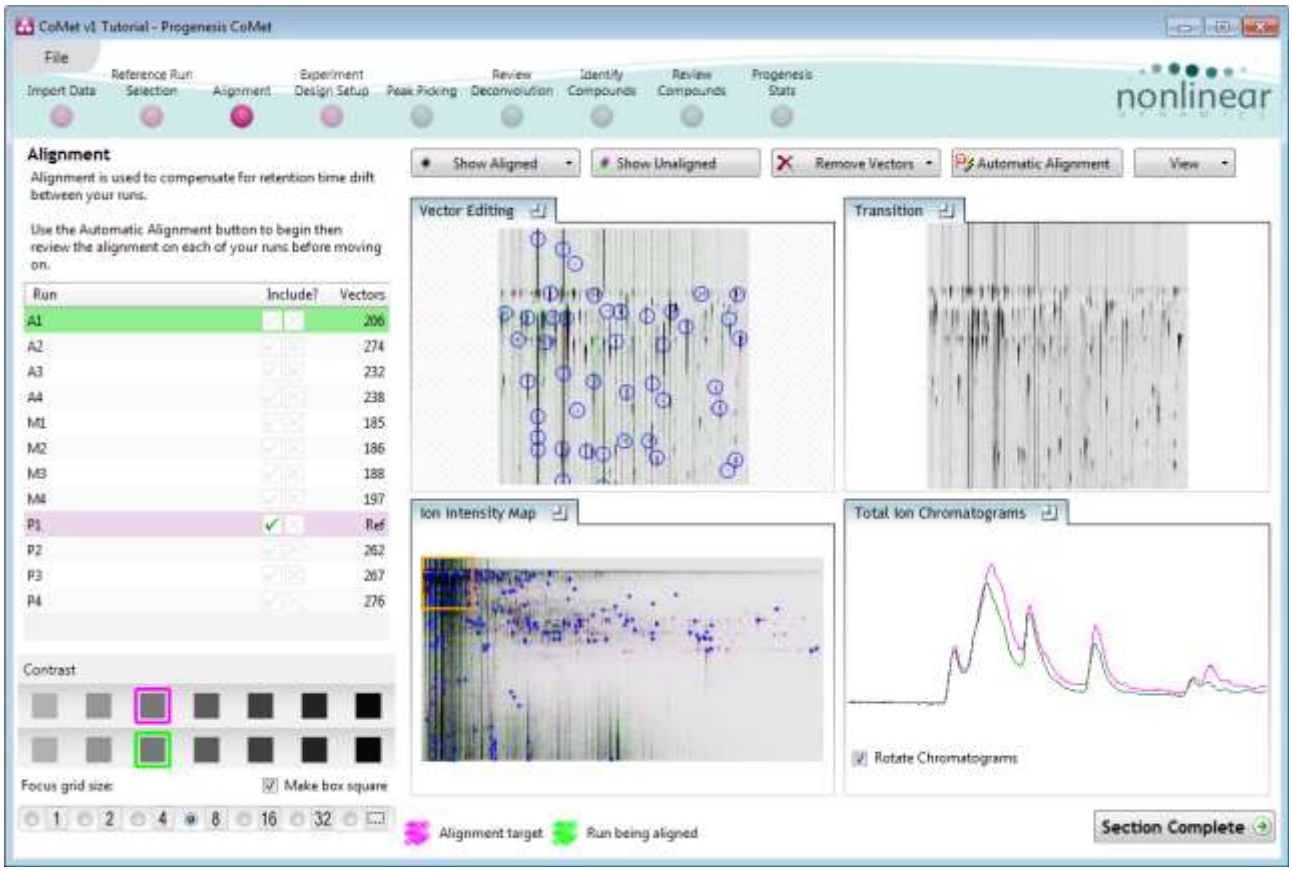
To continue click **OK**



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 were selected as shown on the right panel below.



Click **OK** to open the converted experiment at the completed alignment stage.



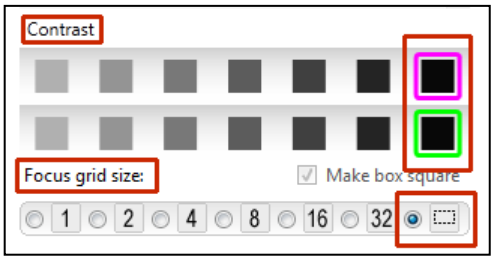
Now continue with the analysis of your data in the CoMet version 2 workflow.

Appendix 5: Manual assistance of Alignment

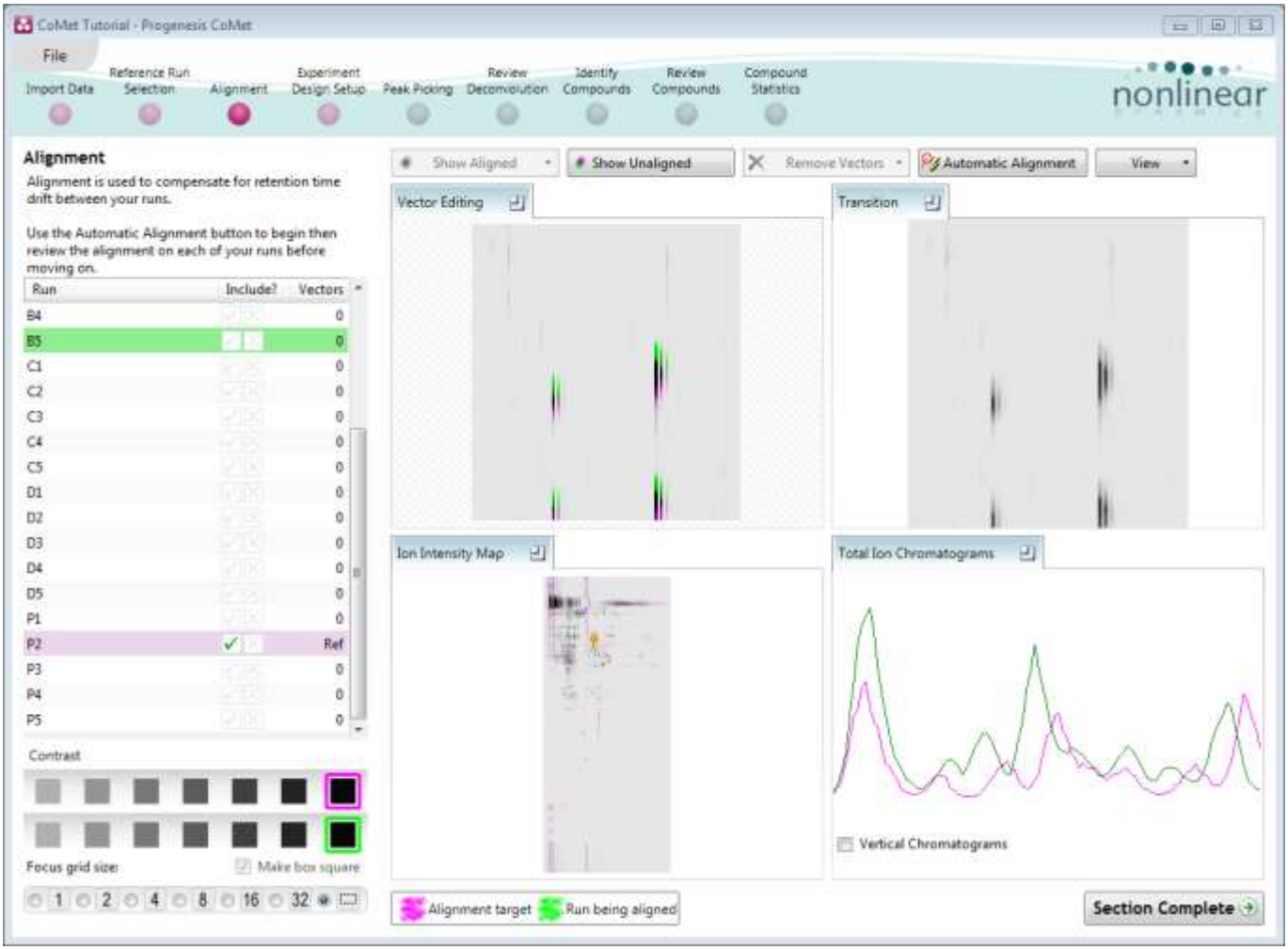
Approach to alignment

To place manual alignment vectors on a run (B5 in this example):

1. Click on Run B5 in the **Runs** panel, this will be highlighted in green and the reference run (P2) 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 ensure that the size of the focus area is set to **16 or 32, or in this example use a custom size** for the Focus grid size on the bottom left of the screen.



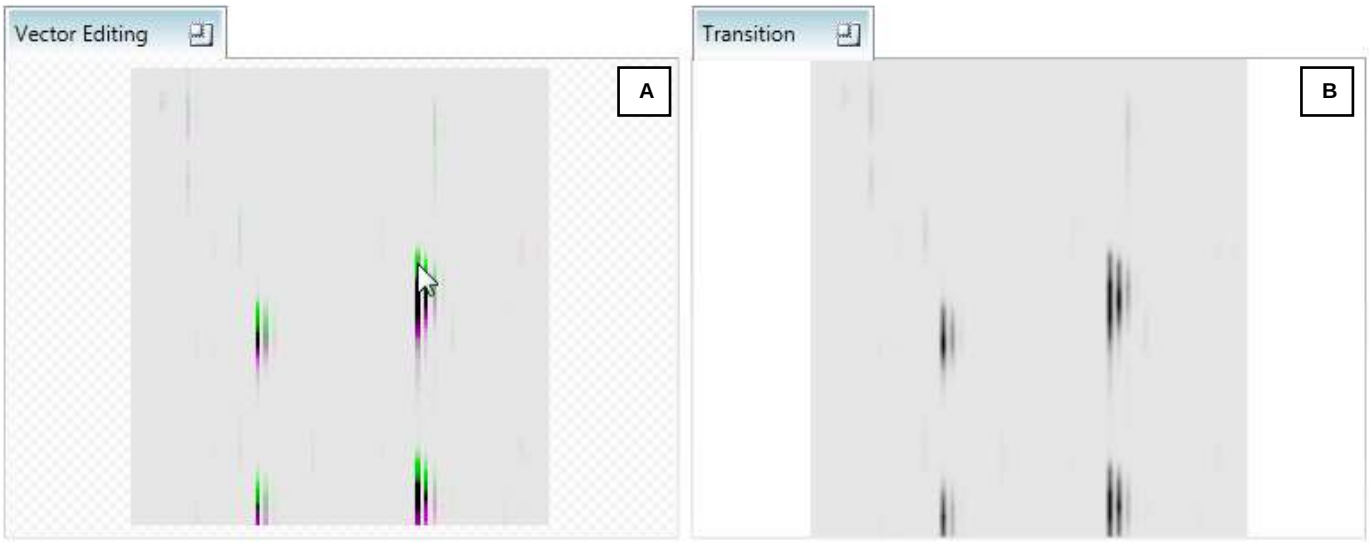
Click on an area (see below) in the **Ion Intensity Map** window to refocus all the windows. Adjust Contrast as required



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

Note: The **Total Ion Chromatograms** window 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.

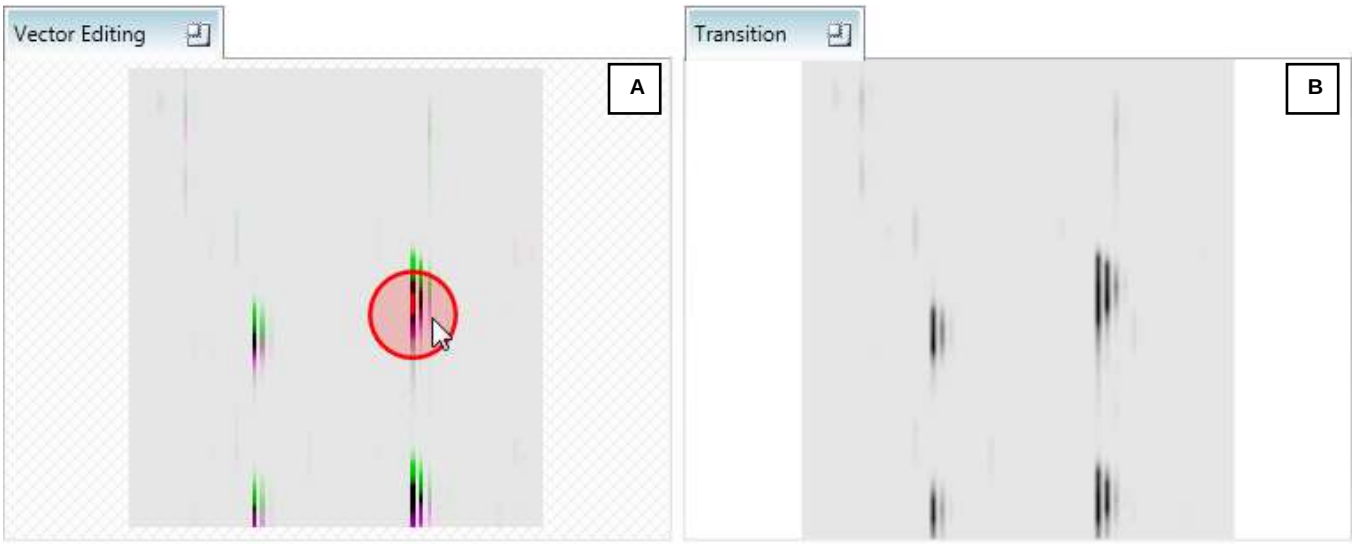
4. Click and hold on a green feature in Window A as shown below.



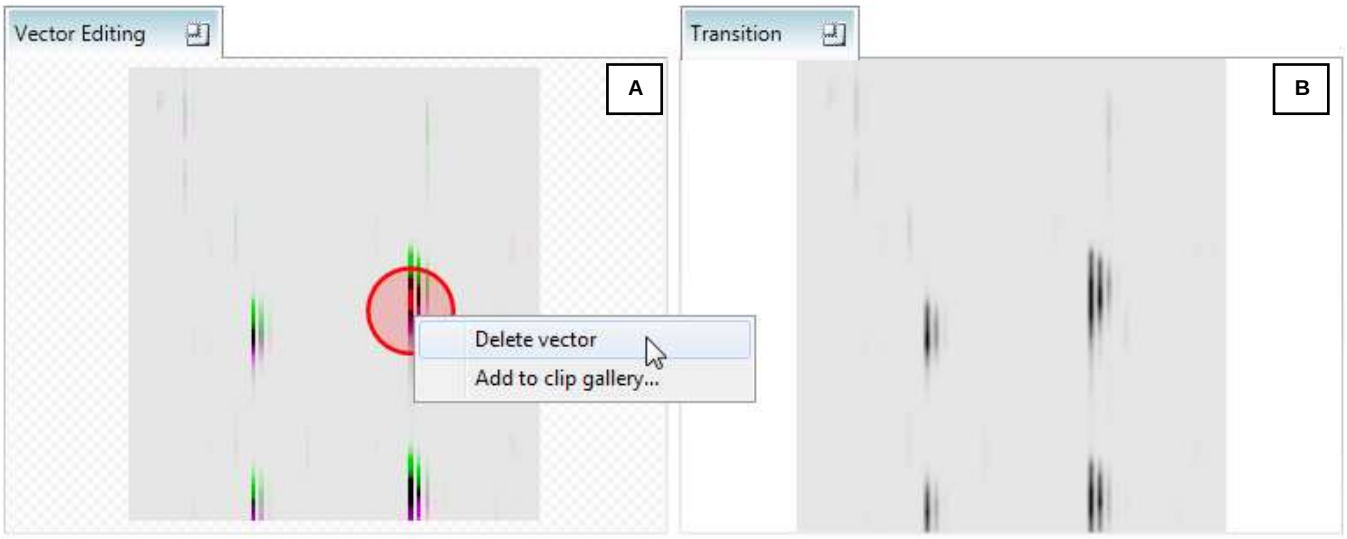
5. As you are holding down the left mouse button drag the green feature over the corresponding magenta feature of the reference run. The red circle will appear as shown below indicating that a positional lock has been found for the overlapping features.

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

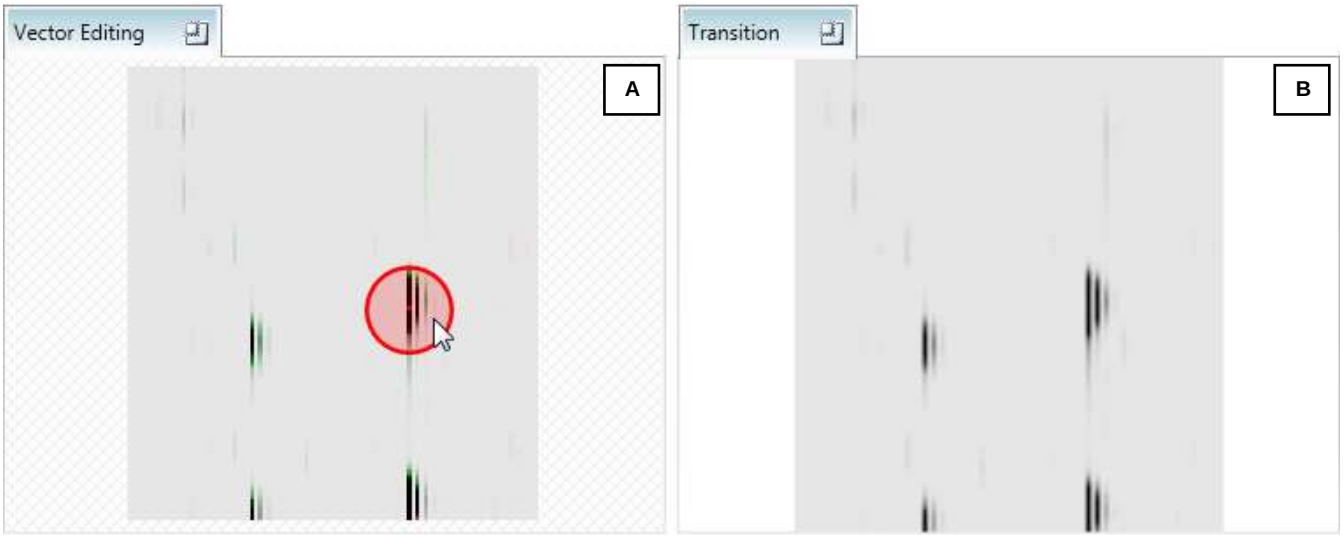
6. On releasing the left mouse button the view will 'bounce' back and a red vector, starting in the green feature and finishing in the magenta feature will appear.



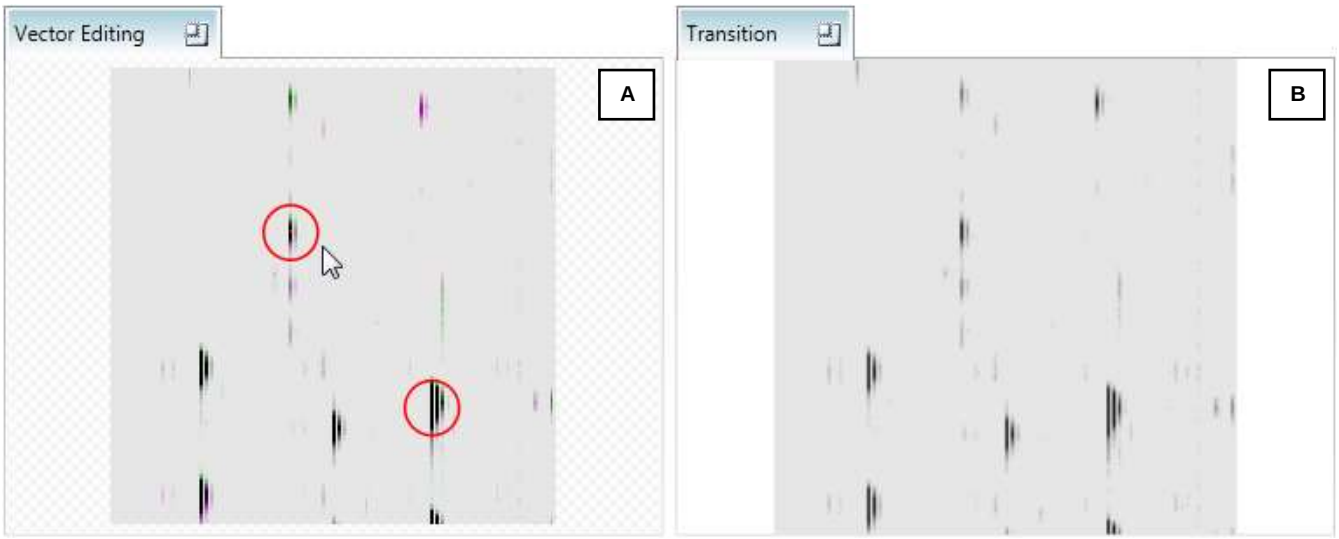
Note: an incorrectly placed vector is removed by right clicking on it in the **Vector Editing** window



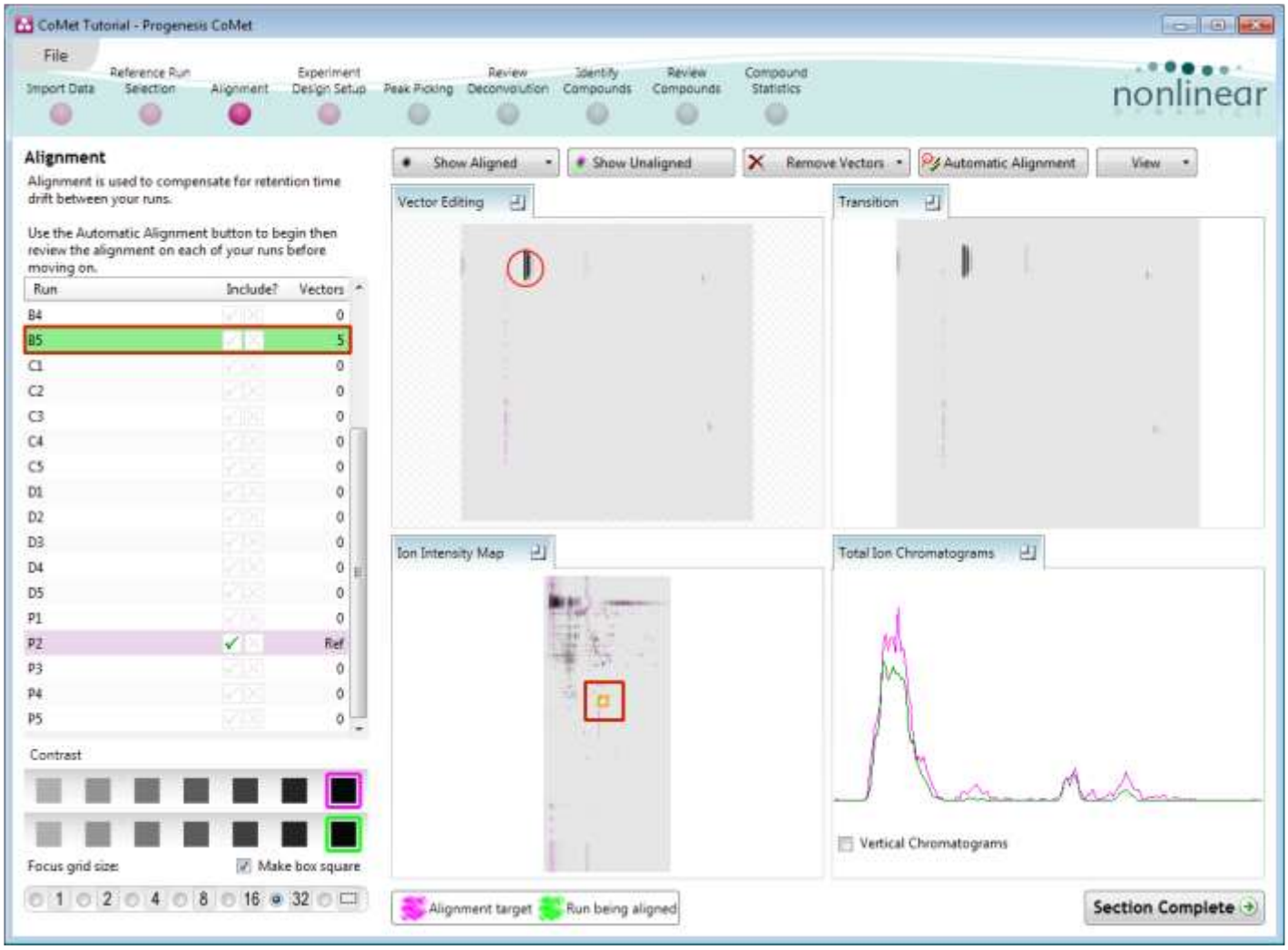
7. Now click **Show Aligned** on the top tool bar to see the effect of adding a single vector.



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



9. Repeat this process moving the focus from top to bottom on the **Ion Intensity Map** view



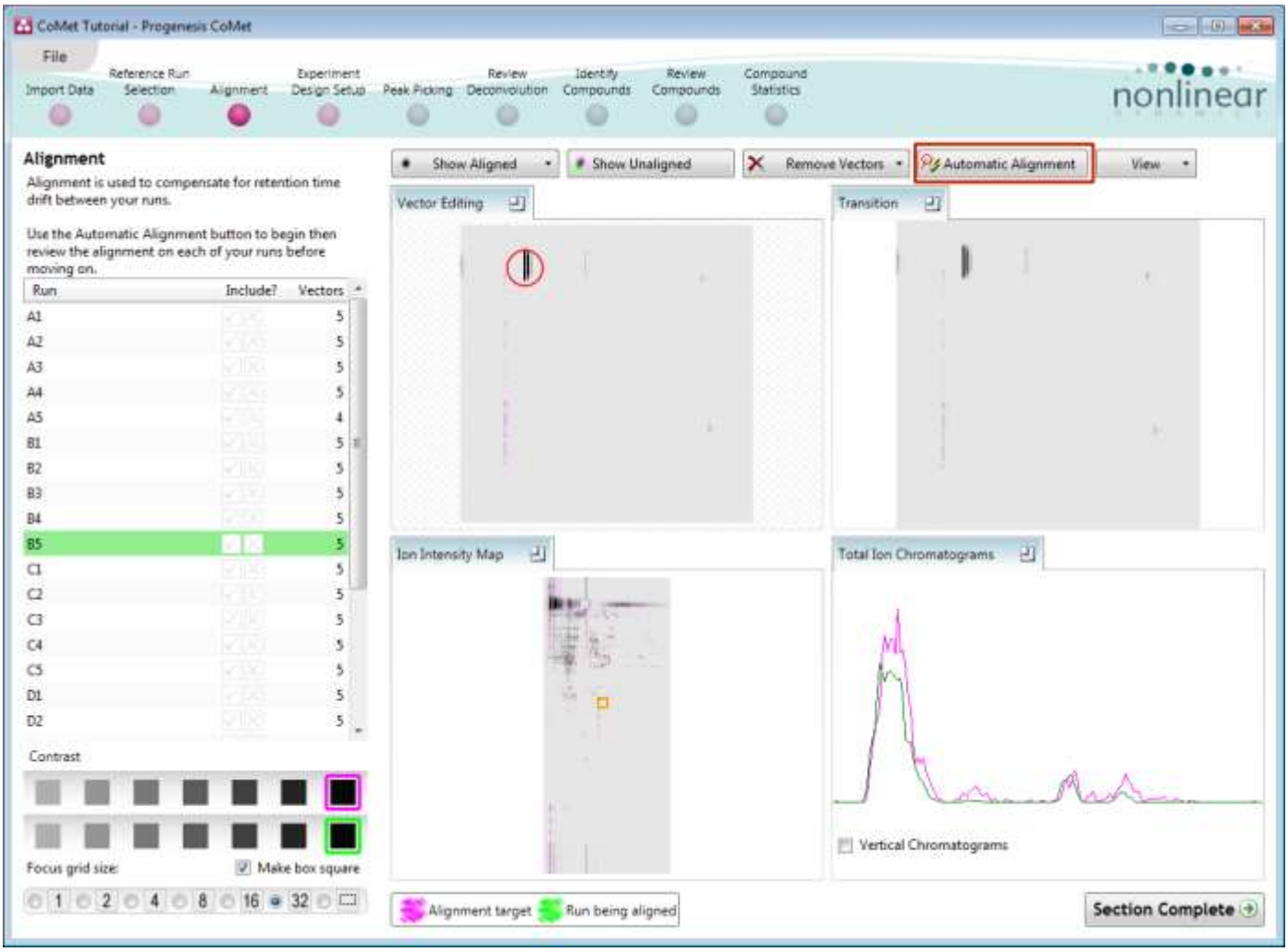
Note: the number of vectors you add is recorded in the **Runs** table

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

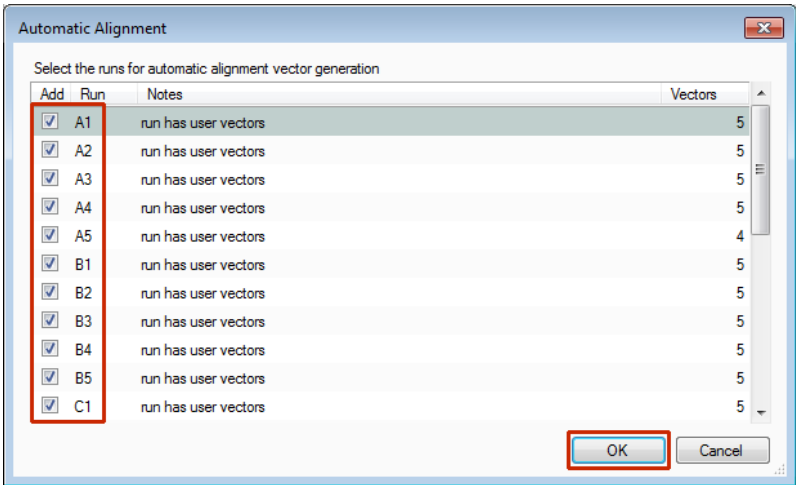
The number of manual vectors that you add at this stage is dependent on the misalignment between the current run and the Reference run. In many cases this may require only 1 or2 manual vectors

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.



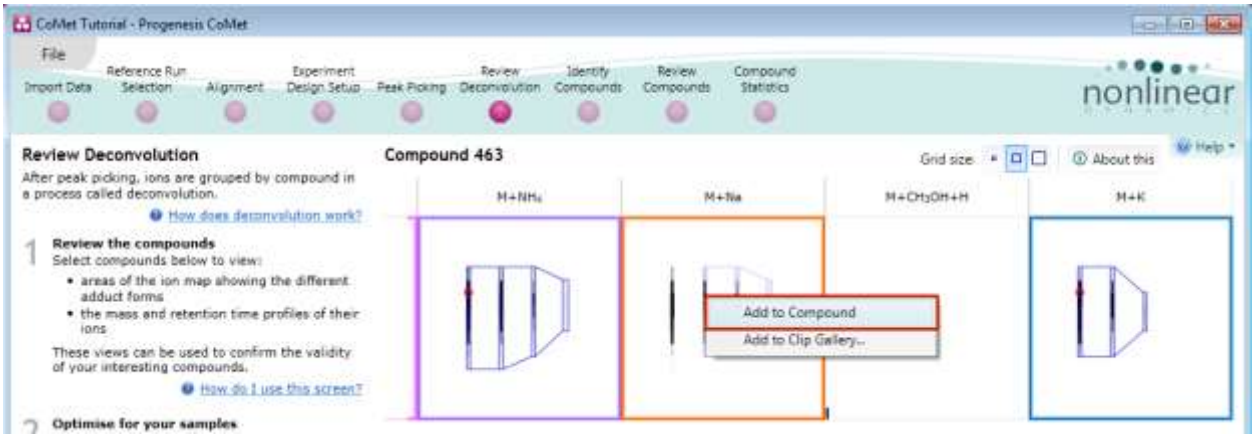
Note: the tick boxes next to the **Run** name control whether vectors will be generated for each run.

Appendix 5: Editing an adduct to add a missing isotope

When using the Review Deconvolution or Compound Validation views you find an example where the Peak picking has missed an isotope for an adduct ion (as shown below). As a result the adduct has not been included as part of the compound measurement. As shown below the monoisotopic peak for the **M+Na** has not been detected hence the deconvolution has not included this as an adduct of Compound 463.

You can combine the tools in both **Review Deconvolution** and **Compound Validation** (accessed through the **Review Compound** stage) to edit and add the edited adduct to the compound.

First right click on the detected ion in the **M+Na** panel for Compound 463 and select **Add to Compound**

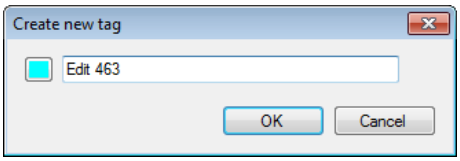


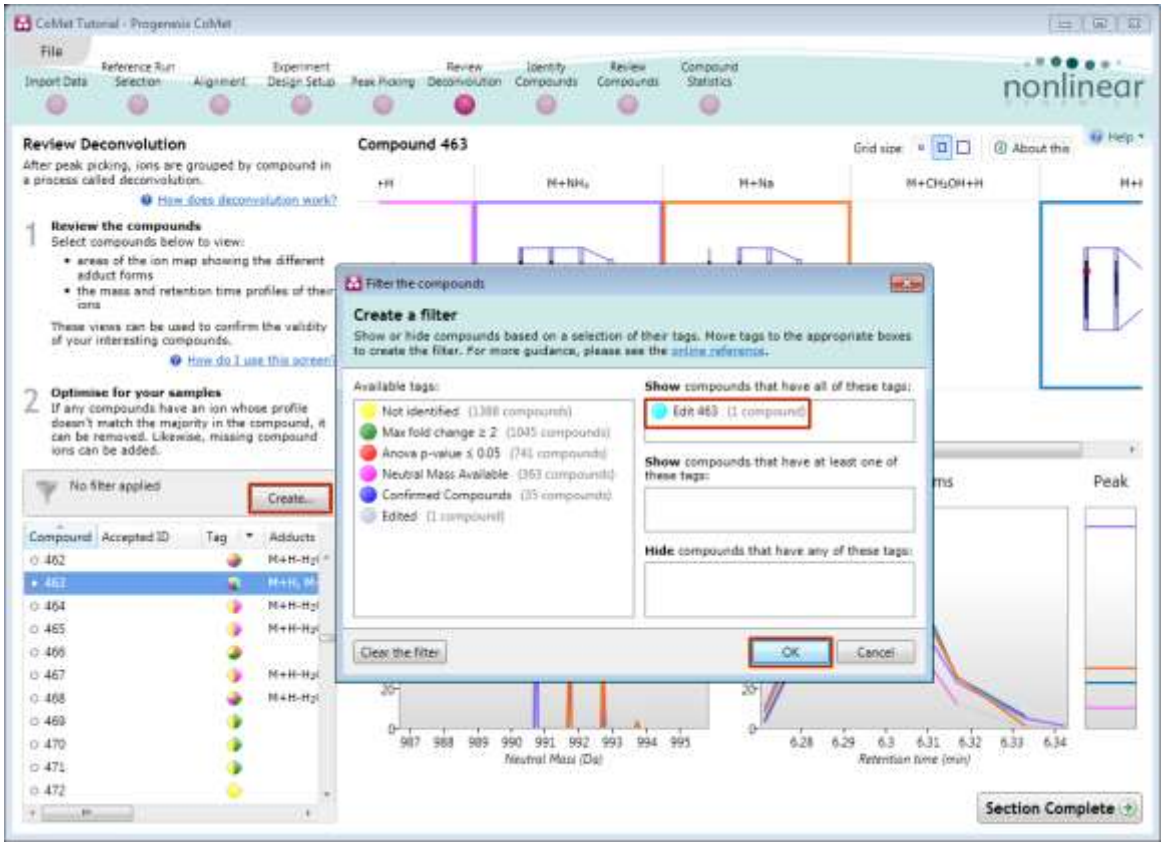
As you do this the detected ion is added as an adduct to Compound 463. In the table 463 the Edited tag is added to compound 463 and the adduct count is increased by one.

Right click on Compound 463 in the table and select New Tag… Call the tag Edit 463.

Now filter the table so that it is only displaying Compound 463

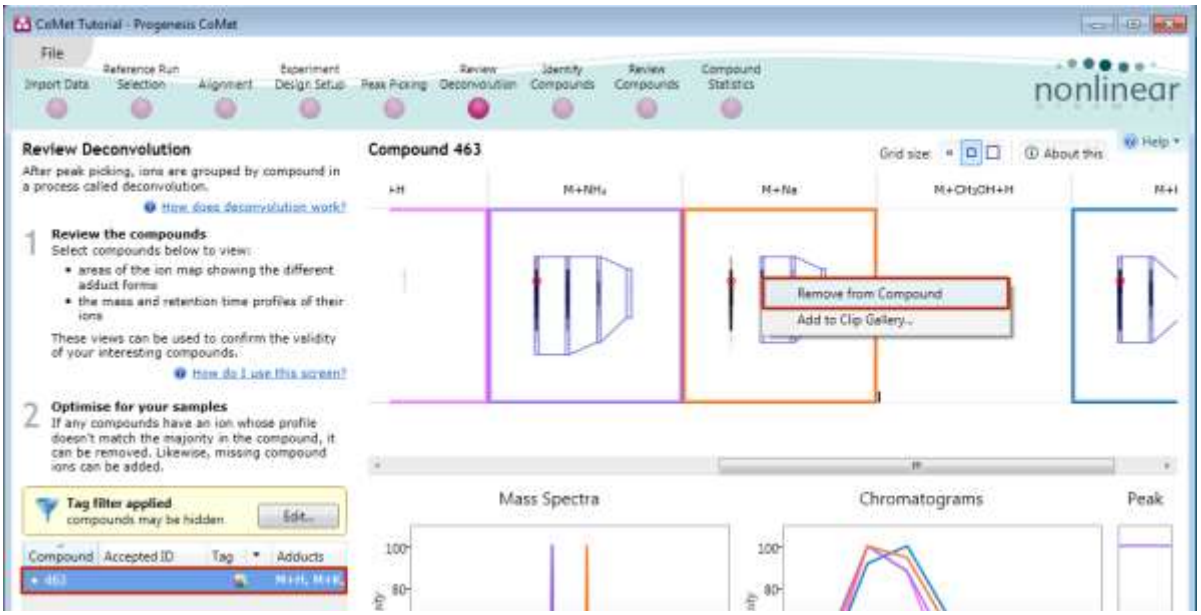
To do this click on **Create** and drag the new tag onto the Show panel and click OK.



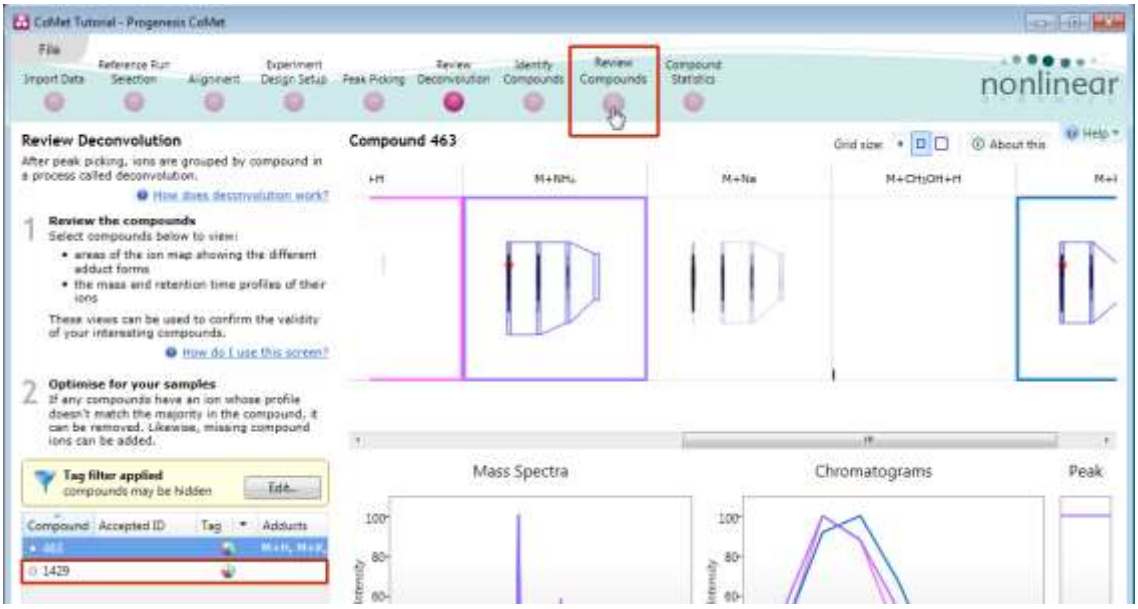


The table will now display details for a single compound with the new tag.

Now right click on the **M+Na** adduct and select **Remove from Compound**

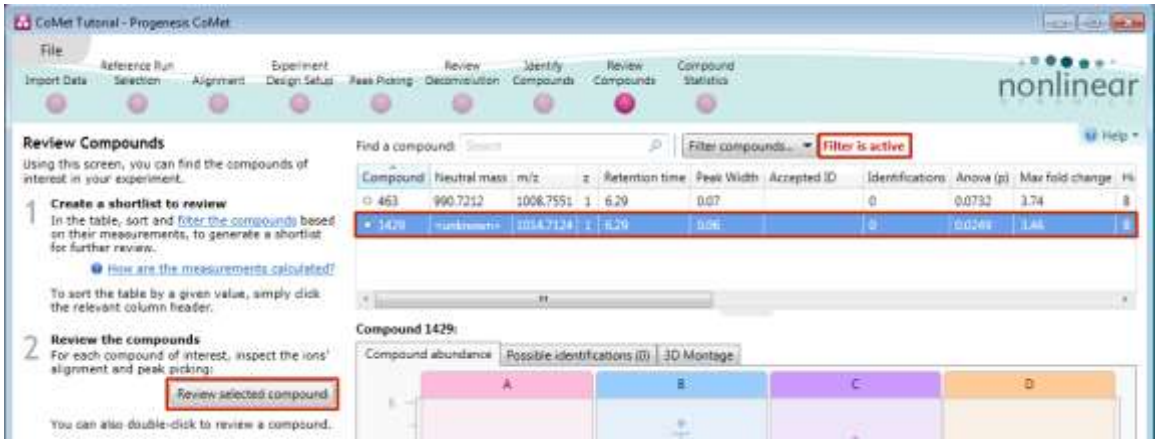


The table will now display 2 tagged compounds the second one corresponding to the **M+H** adduct that has been removed from compound 463 (in this example it has been given the id: 1429

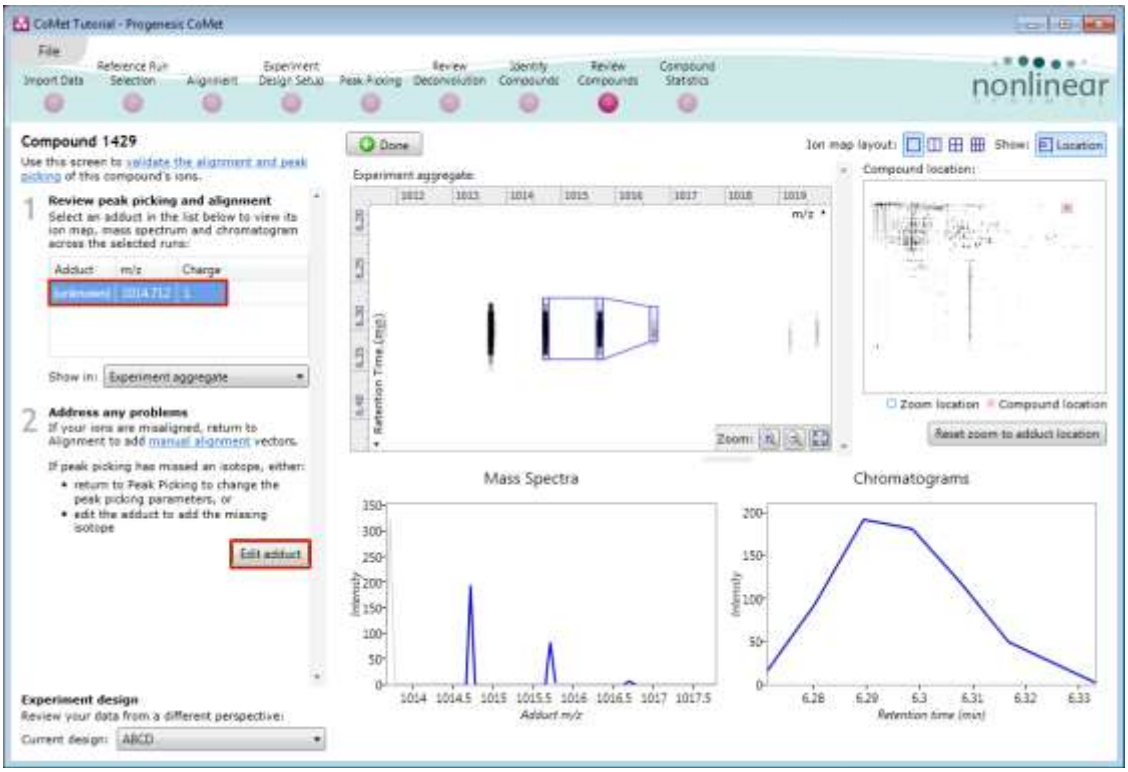


Use the Workflow to move to **Review Compounds**.

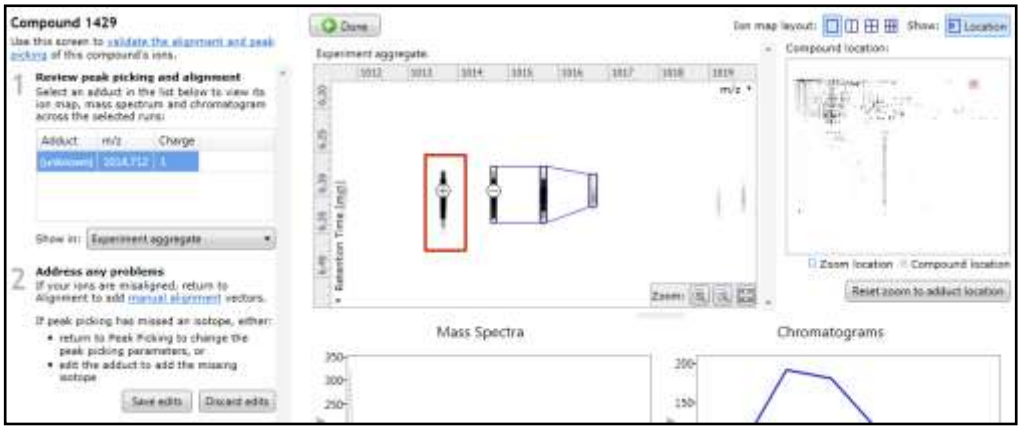
Note: the filter remains active as you move to this stage in the workflow making it easy to focus on the adduct you are going to edit



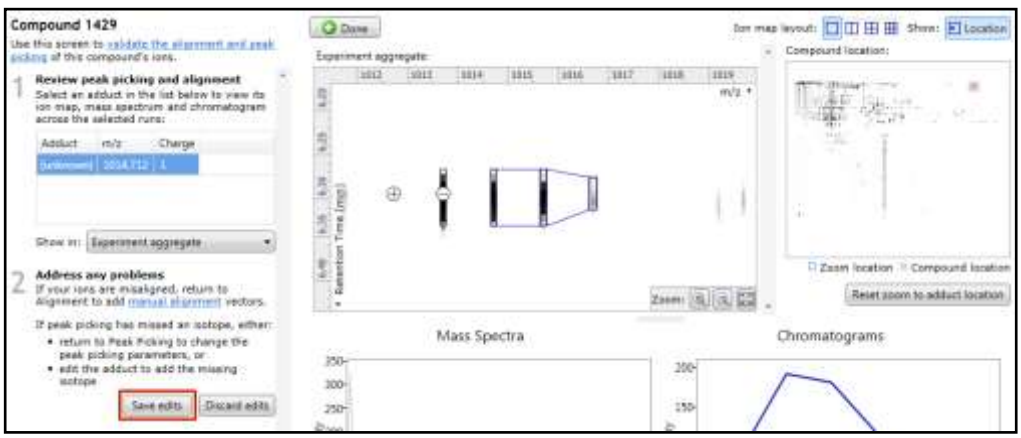
Select the compound 1429 and click **Review selected compound** this will open the Compound Review stage focused on 1429.



Click on **Edit adduct** and then click on the **plus** sign that appears over the missing isotope

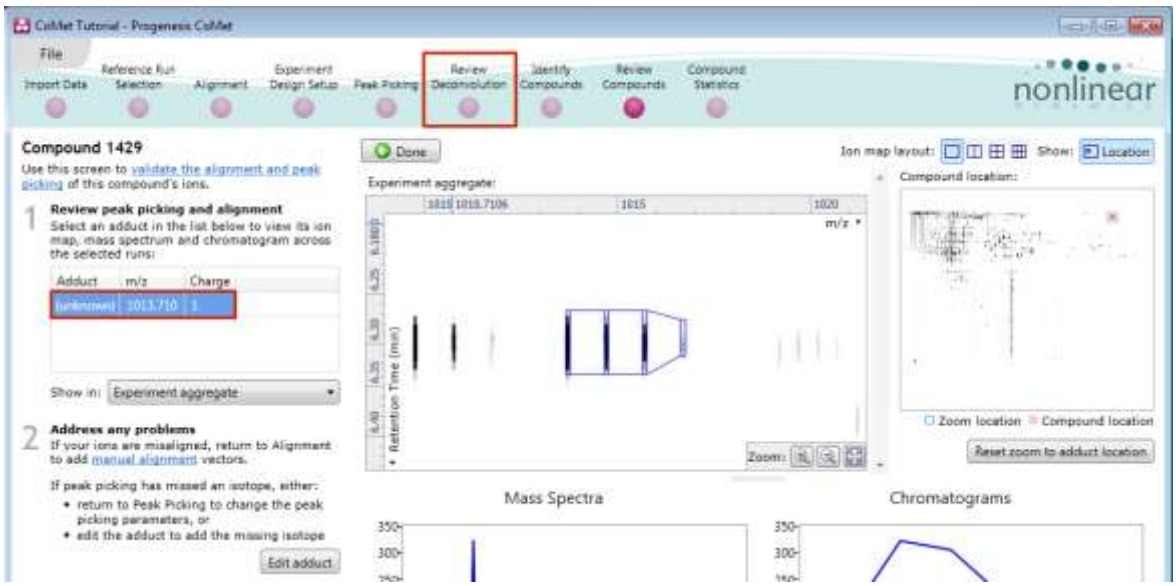


The view updates to show the minus sign now on the added isotope.



Click on **Save edits** to accept the edit

The m/z for the edited adduct updates to 1013.710 as the missing isotope is added.



Click on **Review Deconvolution** on the workflow where the 2 compounds are still displayed as a result of the active tag. Click on compound 463 and now in the **M+Na** pane right click and select **Add to Compound** to add the edited adduct to compound 463.

Note: following the addition of the missing isotope the Mass spectra for the **M+Na** adduct now coincides with the other adduct s of Compound 463 as compared to the unedited form on page 59.

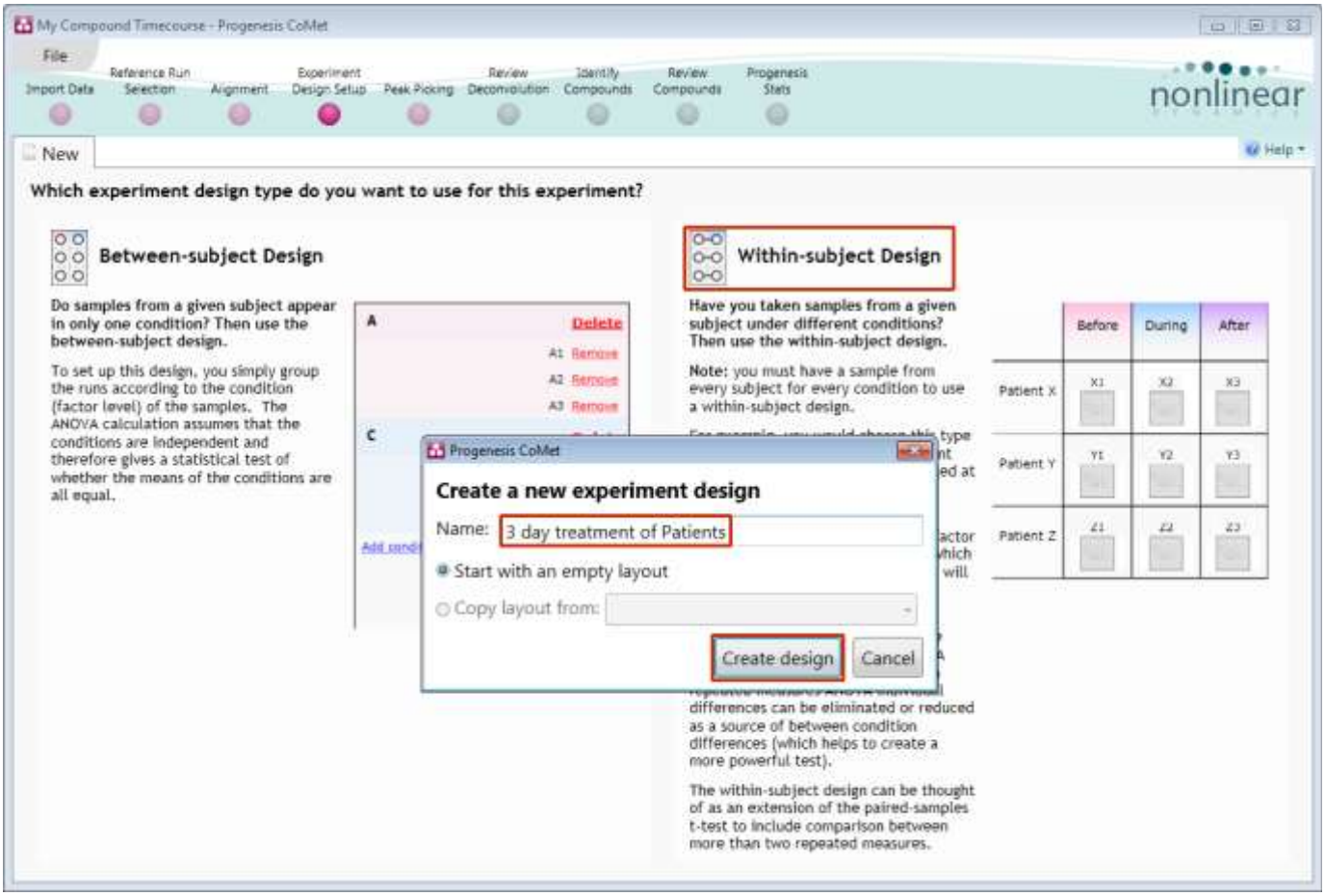
As you add the edited adduct compound 1429 is collapsed into 463 leaving this as the only tagged form in the table.

Finally click on **Edit** and **Clear the filter**.

Appendix 6: Within-subject Design

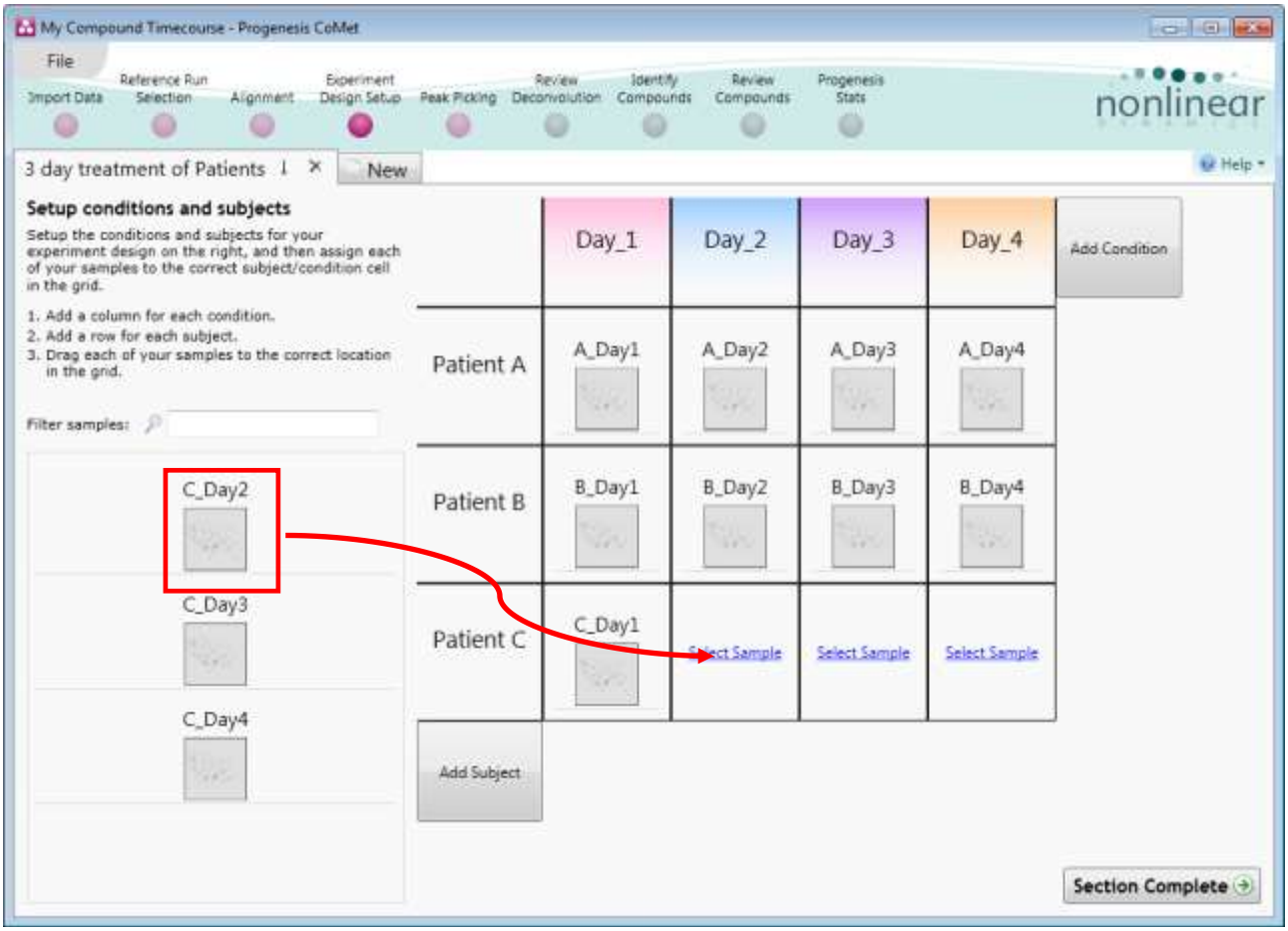
To create a **Within-subject Design** for your data set select this option on the **Experiment Design Setup** page and enter the name of the design.

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.



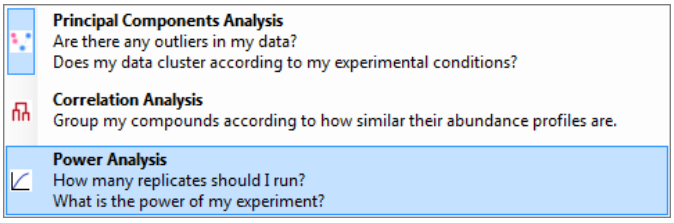
You can create additional Experimental Designs using the New tab

All of these Experimental Designs are available at all the following stages in the CoMet workflow

Appendix 7: 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 Progenesis 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 Progenesis Stats screen. A selection of 3 tools will appear in the form of questions.



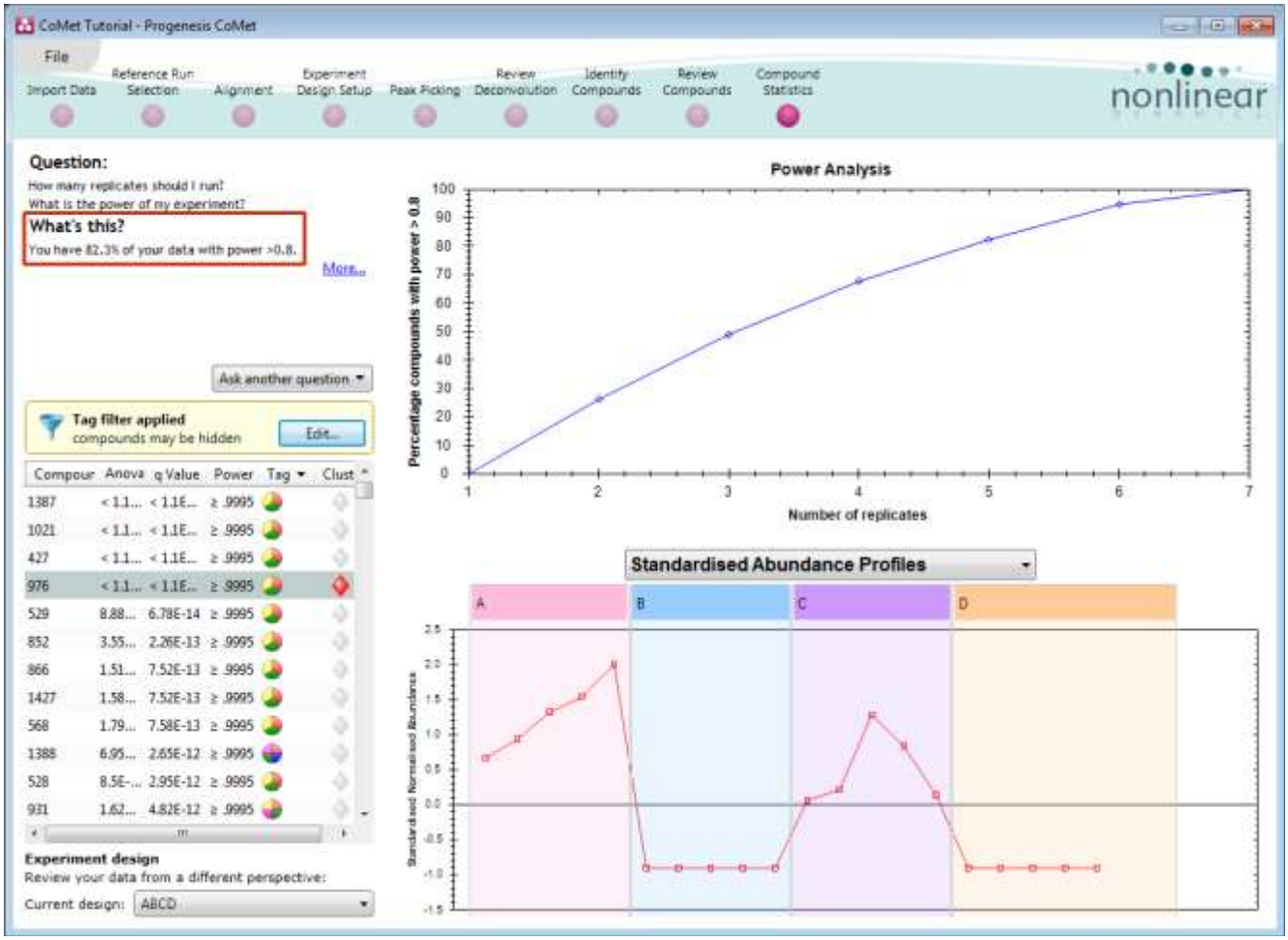
Select the option

'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 (741),** as an example, view the power analysis.



This is displayed graphically showing that 82.3% of your data with power > 0.8.

Appendix 8: Using Clip Gallery to Save and Export Pictures and Data

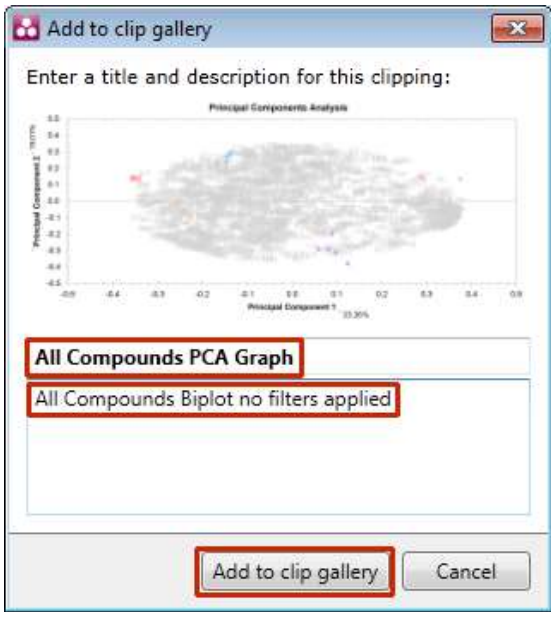
At every stage of the CoMet workflow the images and data tables can be Added to the Clip Gallery.

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 ClipGallery, at the **Compound Statistics** view, displaying the PCA plot.

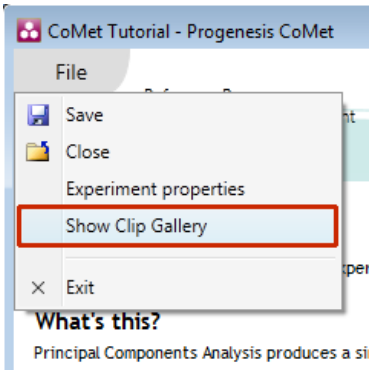
Right click on the **Biplot** View and select Add to clip gallery….

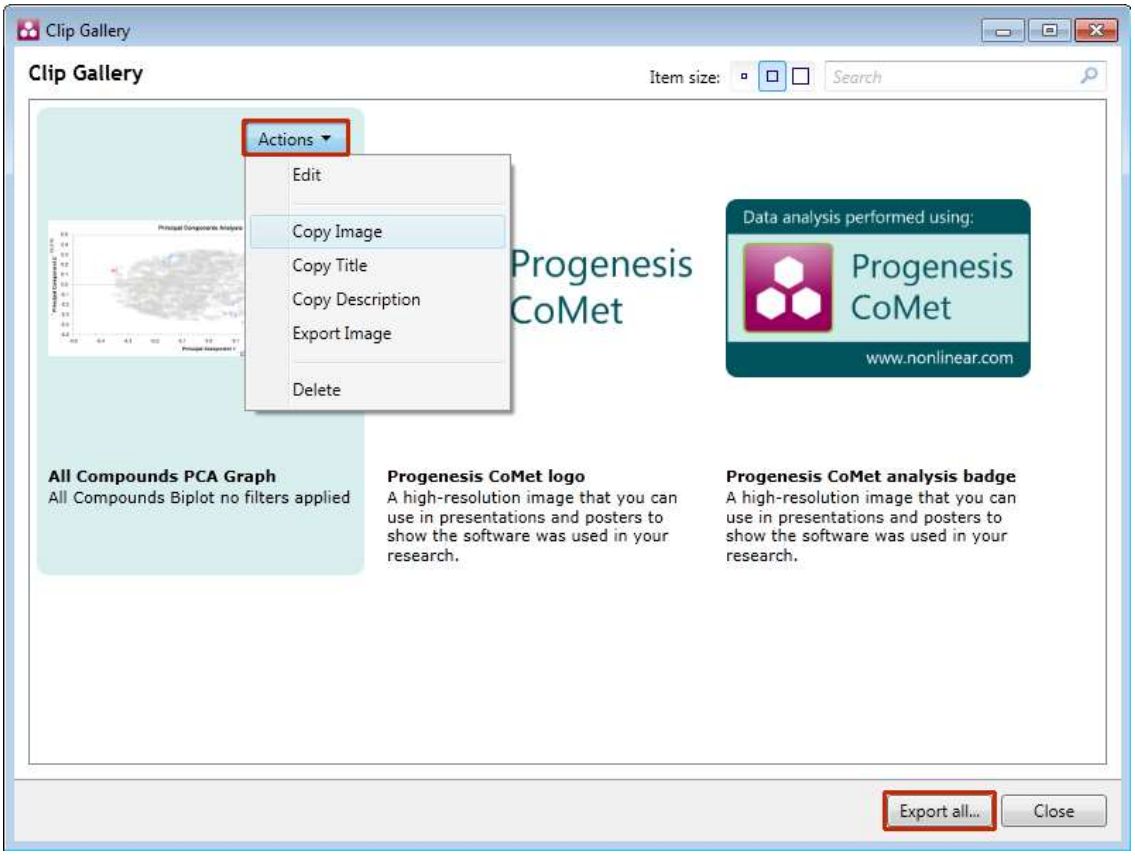
This will open a dialog displaying what is to be saved and allows you to alter the title and provides a description of the image for later reference.



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.