StarDrop™

Version 8

StarDrop™ Worked Example: Guiding compound selection and design in Hit-to-Lead

optibrium™

Contents

Contents	2
Overview	3
Data import and visualisation	3
Predicting ADME properties and scoring compounds using MPO	8
Using CardView to visualise important relationships	11
Identifying high-quality chemistries for progression	19
Identifying activity cliffs to elucidate SAR	21
Interactive design	25
Conclusion	27

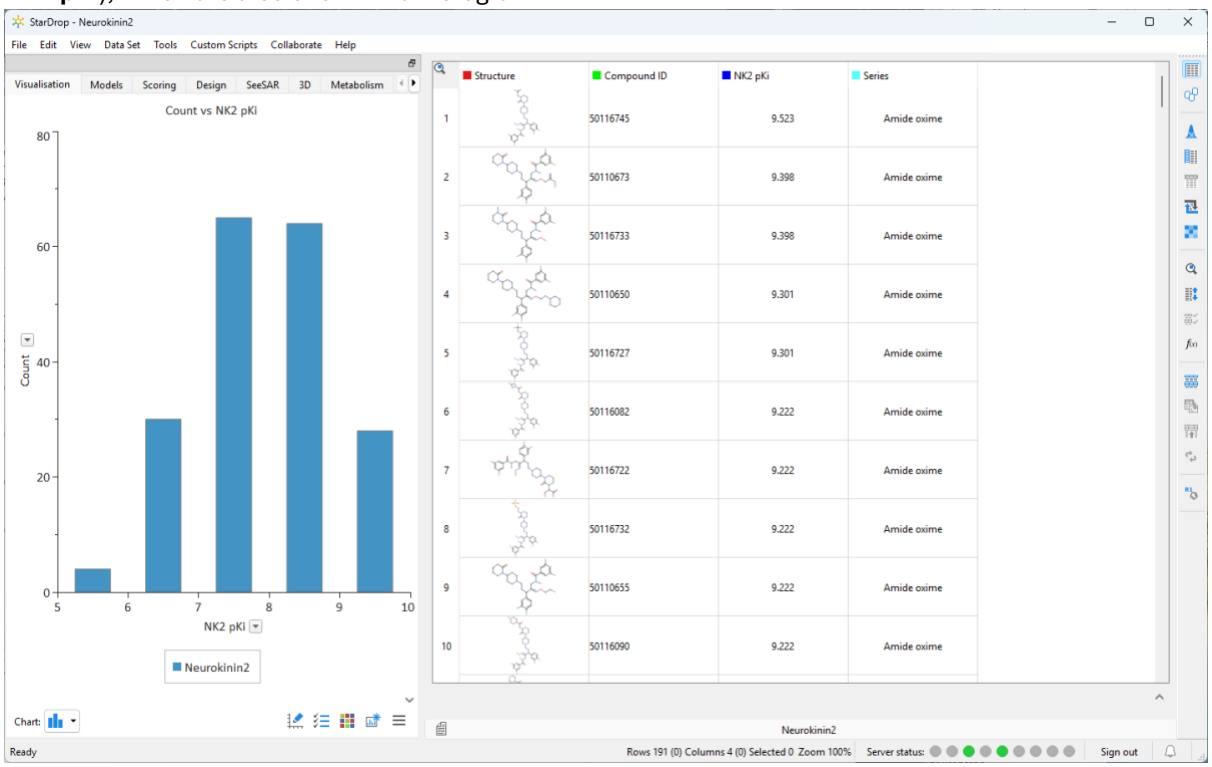
Overview

This example explores some of the challenges typically encountered in a hit-to-lead project. The objective, in this case, is to identify one or more high-quality chemistries for progression to detailed *in vitro* and *in vivo* studies, based on initial screening data for potency; ideally, the compounds chosen for progression should be not only potent but also have appropriate ADME properties to result in a high-quality lead series. We will also use StarDrop™ to consider a modification to one of the existing compounds to improve its properties.
During this exercise, we will use a variety of StarDrop's capabilities to explore the data to select and design compounds with a good balance of properties. Step-by-step instructions for all the features you will need to use in StarDrop are provided, along with screenshots and examples of the output you are likely to generate. If you have any questions, please feel free to contact support@optibrium.com.

Data import and visualisation

- In StarDrop, select **Open** from the **File** menu, find the StarDrop file **Neurokinin2.sdproj** and select it to open it.

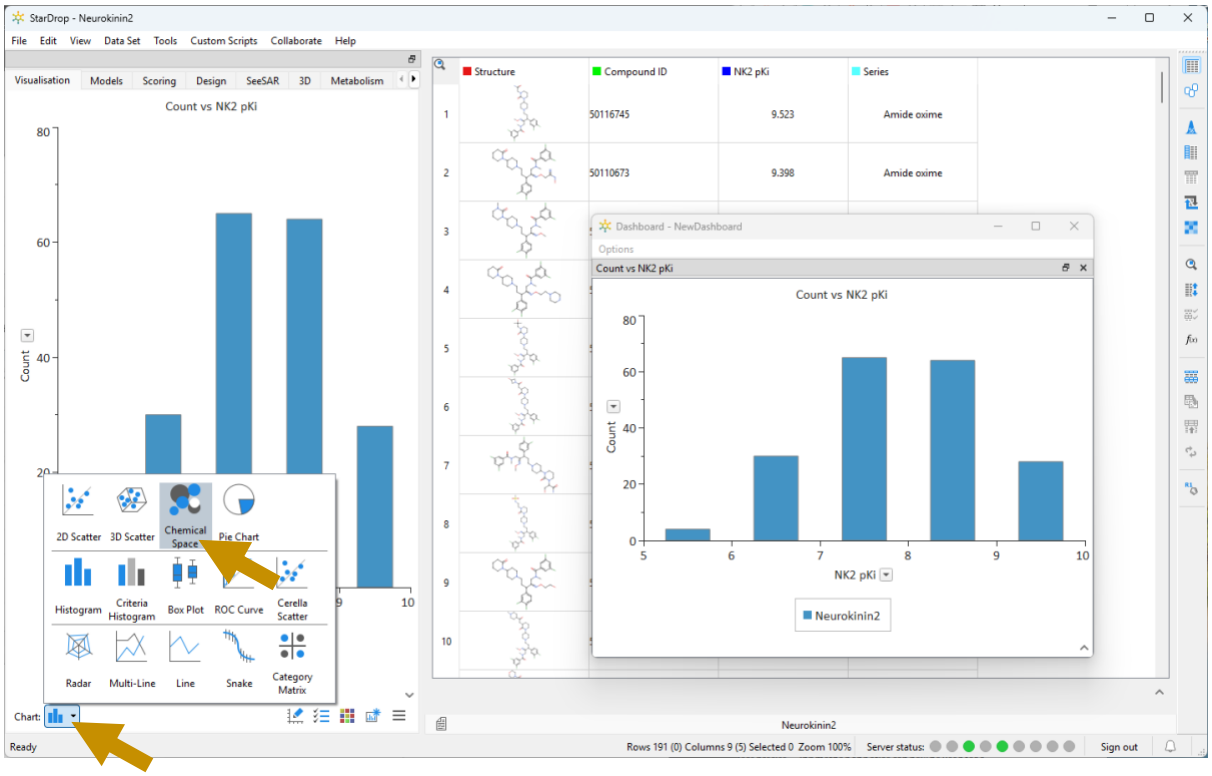
You can see a spreadsheet containing 191 structures and their measured affinities for Neurokinin 2 (in the column **NK2 pKi**), which are also shown in a histogram.



We can explore the distribution of potency across this library's chemical diversity by creating a dashboard containing this histogram and a chemical space visualisation.

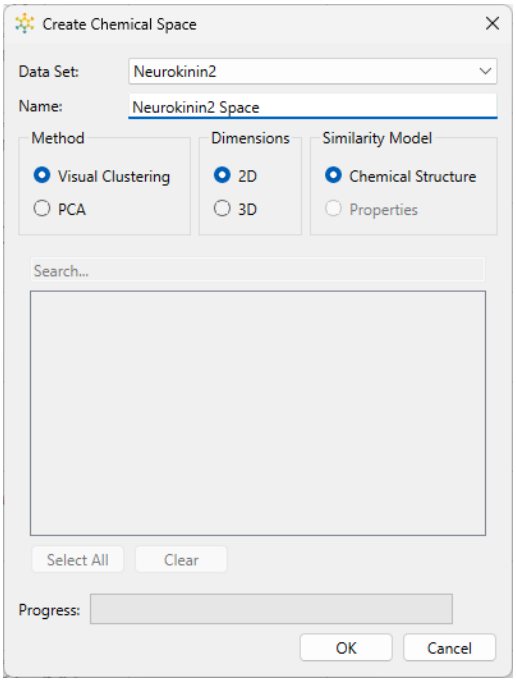
- At the bottom of the **Visualisation** area, click the **Detach** button to copy the histogram into a new dashboard.

- Now, click on the **Chart** button at the bottom and select **Chemical Space** from the menu.

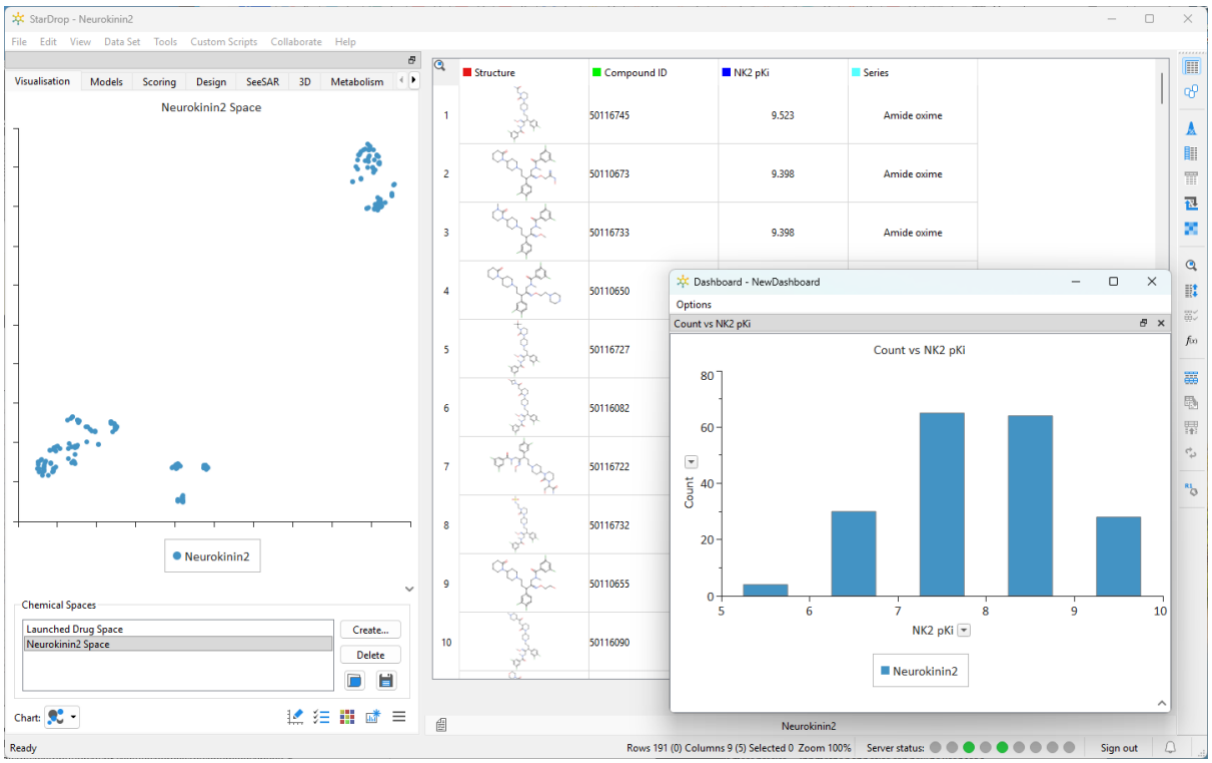


- Click on the **Create** button at the bottom of the **Visualisation** area to open the **Create Chemical Space** dialog, where we can define a new chemical space visualisation.
- Give the chemical space a name by typing in the **Name** field. In the example (right), we have called it **Neurokinin2 Space**.

It is possible to build different types of chemical space plots based on chemical structure or property data, but in this case, we will use the defaults: **Chemical Structure**, the **Visual Clustering** method, and **2D**, as shown (right).

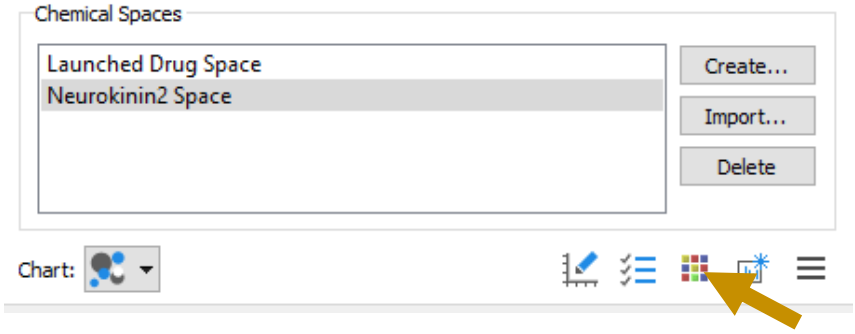


- Click the **OK** button and wait until the chemical space has been created and is displayed in the **Visualisation** area.

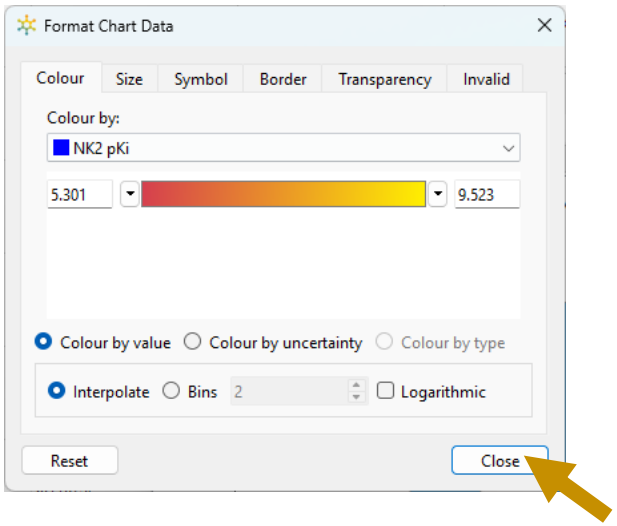


Each point in the chart corresponds to one compound in the data set. Hovering the mouse over a point shows the compound structure. Selecting a row in the data set selects the corresponding point in chemical space and vice versa. We need to change the formatting to see the distribution of pKi values across the chemical space.

- Click on the **Format** button at the bottom of the **Visualisation** area to bring up the **Format Plot Data** dialog.

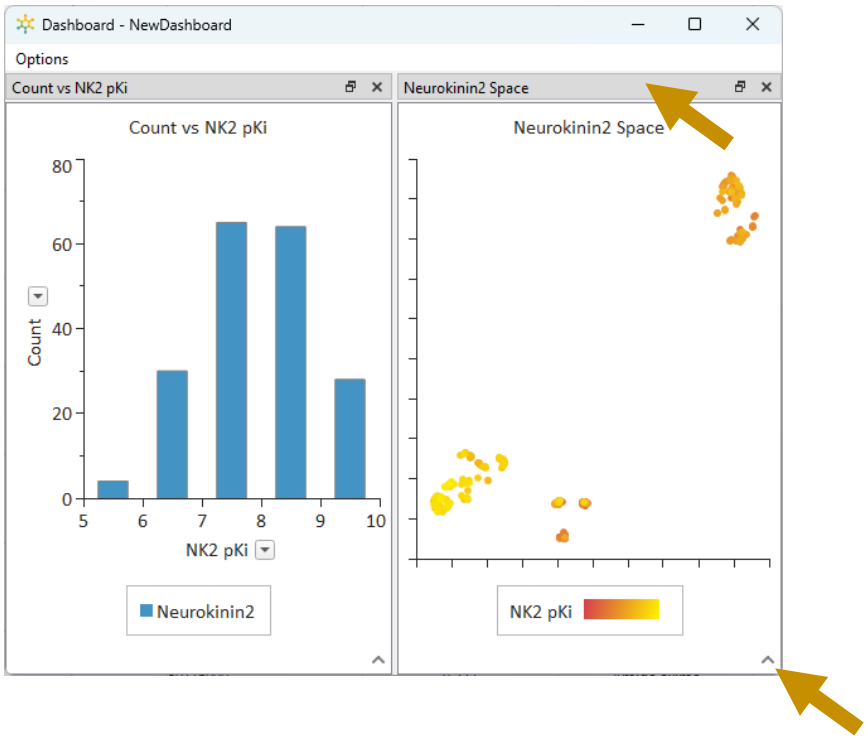


- Use the drop-down menu to **Colour by** the property **NK2 pKi**.



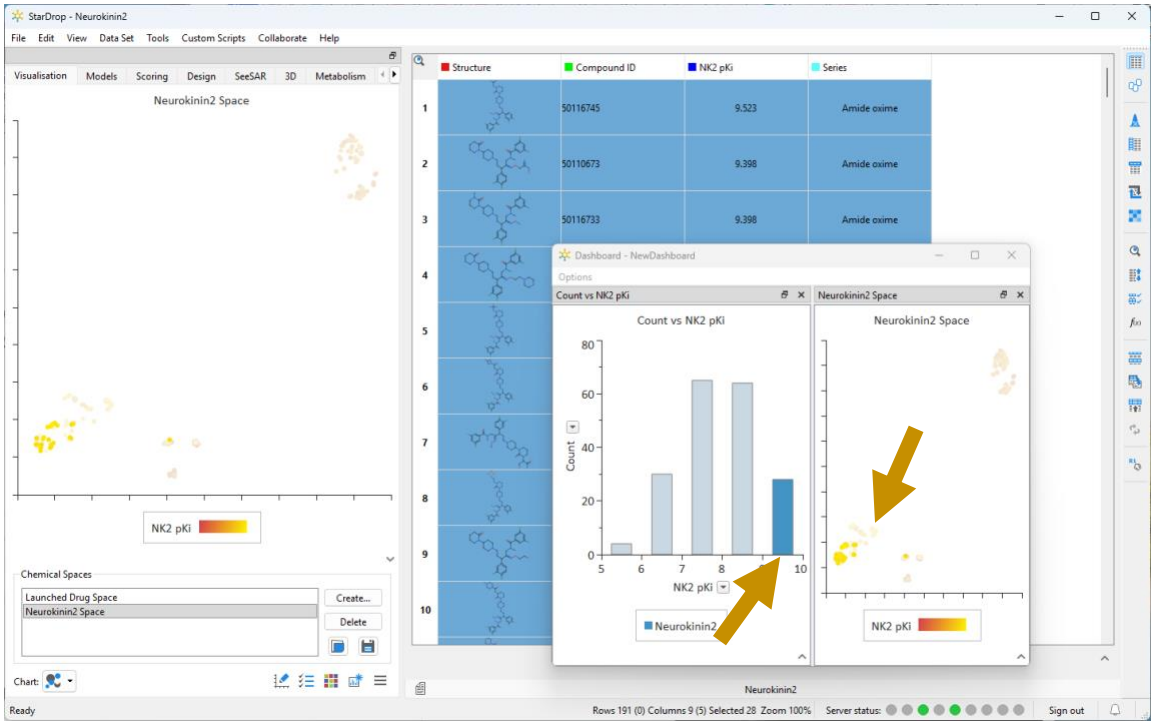
The compounds with the highest pK_i values are highlighted yellow and the lowest red, but you can change these by clicking on the colour bar. The plot updates to show your changes as you modify the formatting options. Click the Close button to close the formatting dialog.

- Add the chemical space to the dashboard by clicking the **Detach** button .



You can resize the dashboard window and drag the charts and the spacers between them to reorganise the dashboard. Clicking the arrow in the bottom corner of a chart displays all the controls you need to change or format your chart. A large proportion of the compounds with the highest pK_i values (yellow) come from one of the chemical space clusters.

- To confirm this, click on the right-hand histogram bar, which represents the highest pK_i values.

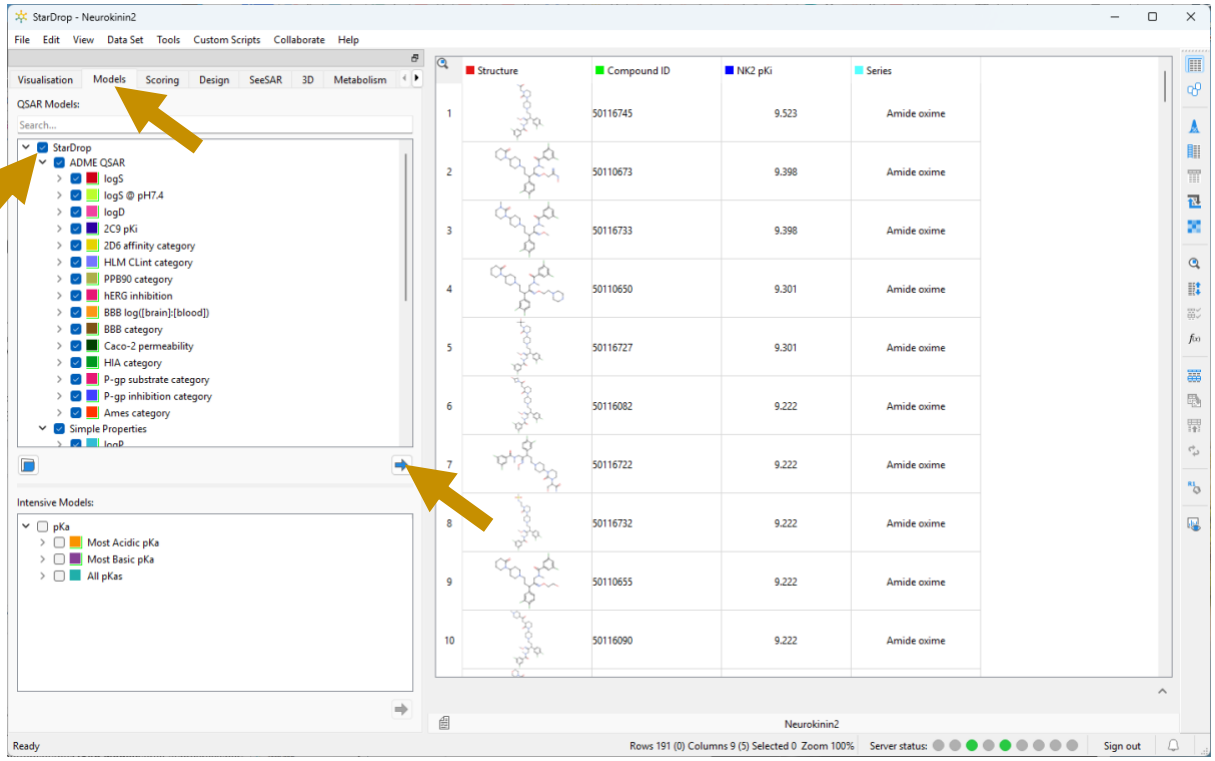


This highlights the most potent compounds in the chemical space, as well as in the data set. This may be the best region in which to focus for selecting compounds; however, first, we should consider the other properties that are important in a high-quality lead. We have no further measured data, but can generate predictions of ADME and physicochemical properties.

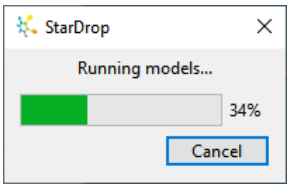
Predicting ADME properties and scoring compounds using MPO

We will use StarDrop's ADME models to generate predicted values for properties of interest for our project. These predictions can help us to prioritise compounds with the best chance of success on project objective and help us focus our resources only on the most important experiments to move the project forward.

- Click on the **Models** tab (you can minimise the dashboard window to keep it out of the way for now).
- To select all the StarDrop models, check the box next to the word **StarDrop**.

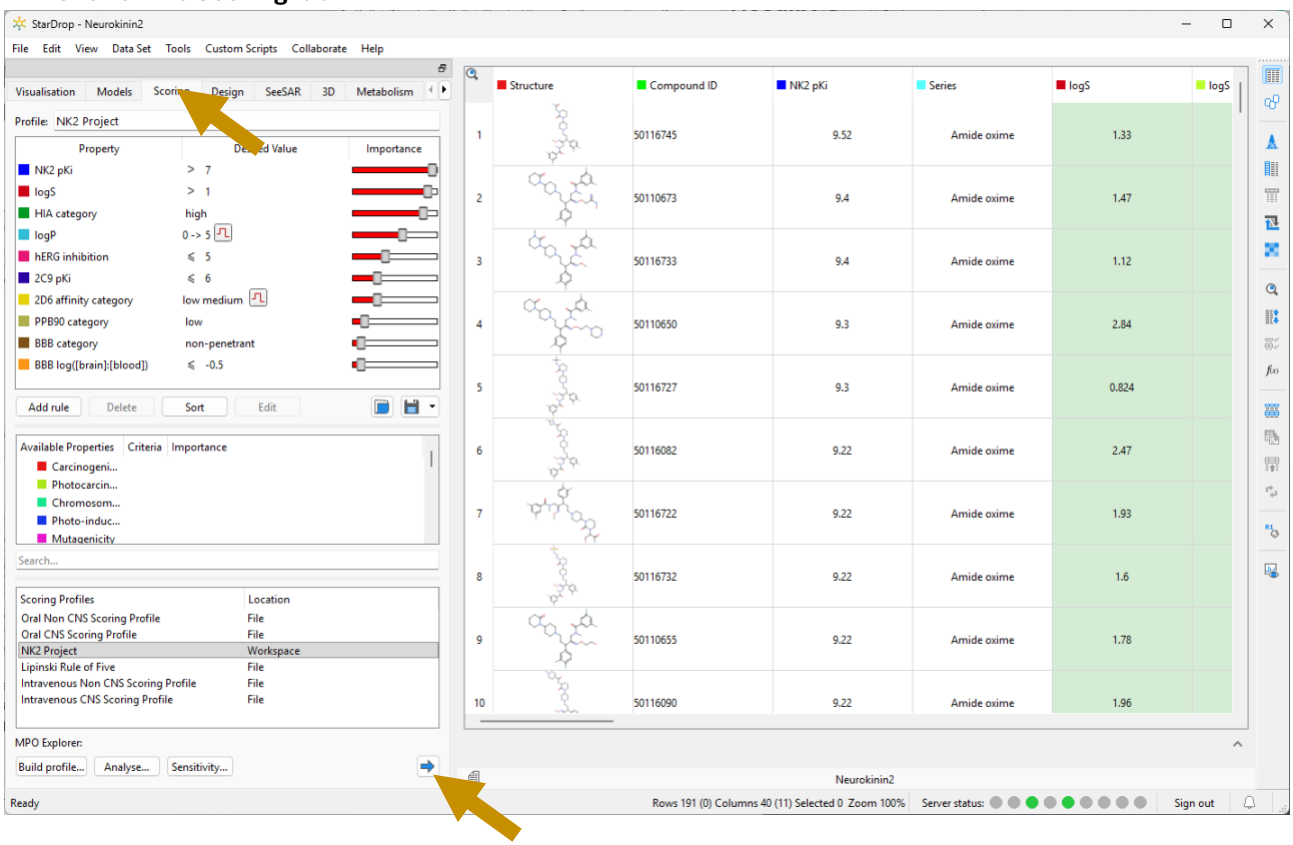


- Click the  button at the bottom of the **QSAR Models** area. A progress bar is displayed while the predictions are calculated.



When this process is complete, you can see that a new column has been added to the data set for each property calculated. Due to the volume and complexity of the data, it is challenging to find the compounds which have the best balance of properties; therefore, we're going to use StarDrop's approach to multi-parameter optimisation (called Probabilistic Scoring), which makes it easy to assess all this information.

- Click on the **Scoring** tab.



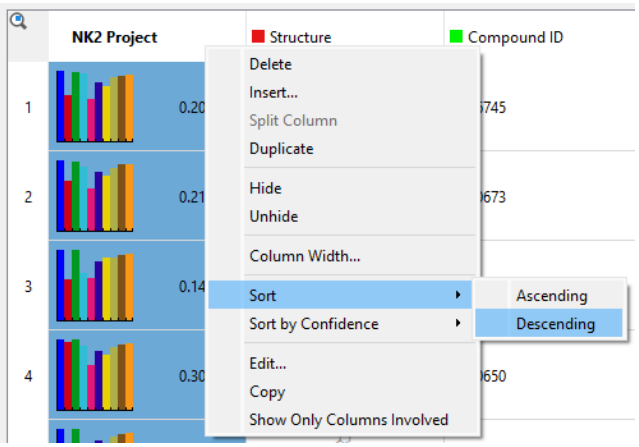
A scoring profile enables you to define a set of criteria that are important for your project. StarDrop provides some example profiles, and as part of the workspace file, we have already loaded a profile designed for this Neurokinin 2 project. It contains a series of properties along with criteria describing desired values and their relative importance.

In this profile, we are looking for compounds with a good affinity for Neurokinin 2 and suitable for a peripheral target.

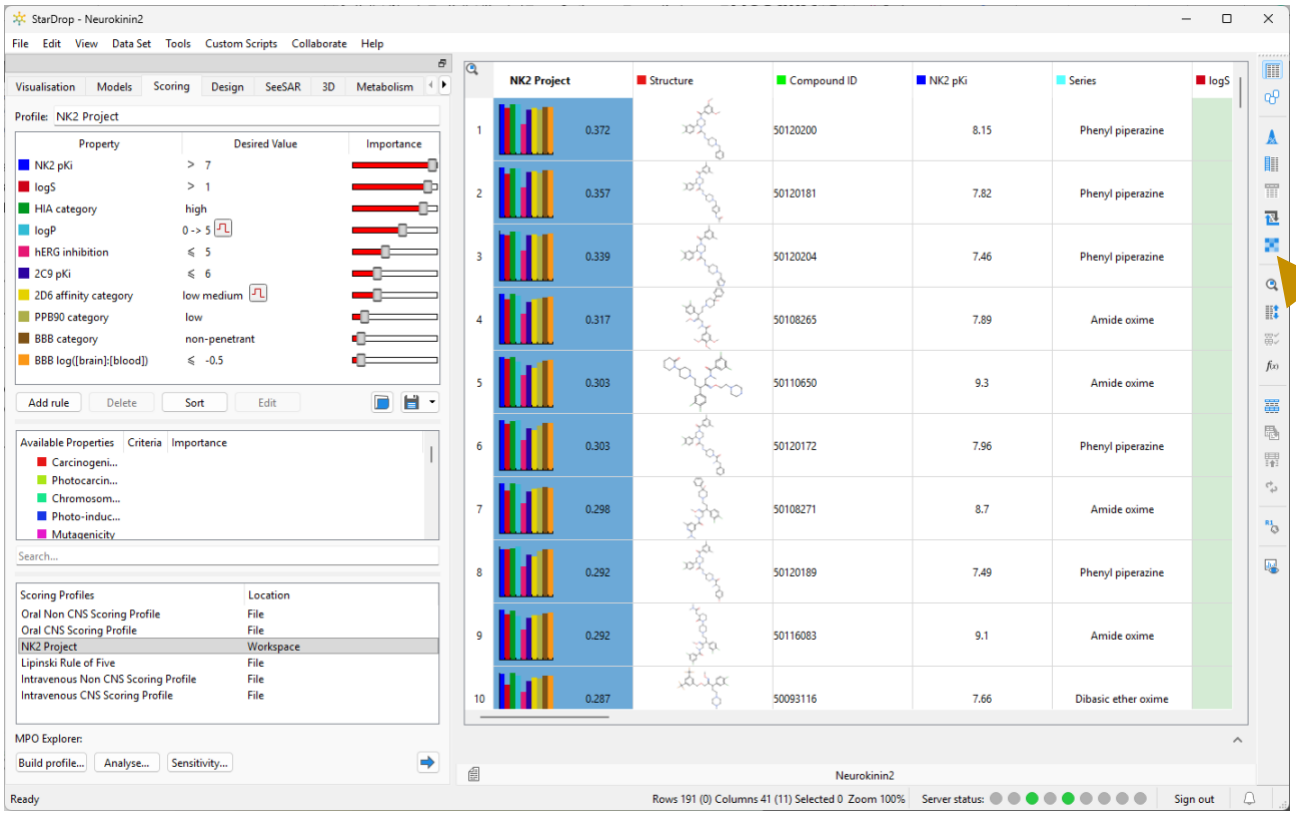
- Run the scoring by clicking the  button at the bottom of the **Scoring** area.

A new column is added to the data set containing a score for each compound, considering each property criterion, its relative importance in the profile, and the uncertainty in the underlying experimental and predicted data. The score is a value between 0 and 1, representing the likelihood of the compound meeting all the criteria in the profile. The histogram in each cell gives a quick overview of each property's impact on the compound's score.

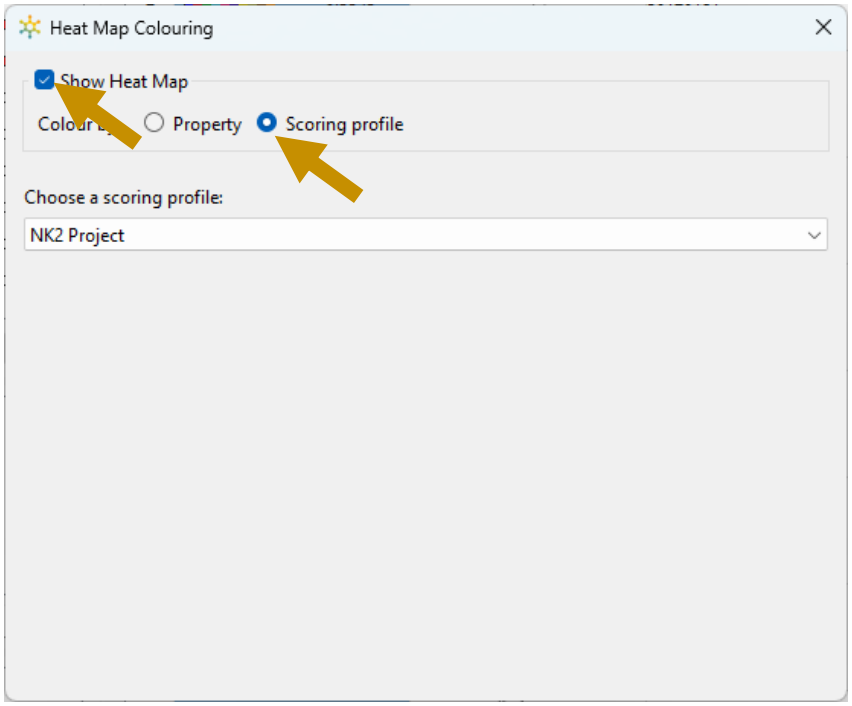
- To find the highest-scoring compounds, right-click on the scoring column header and choose **Descending** from the **Sort** menu to sort the data set from high to low.

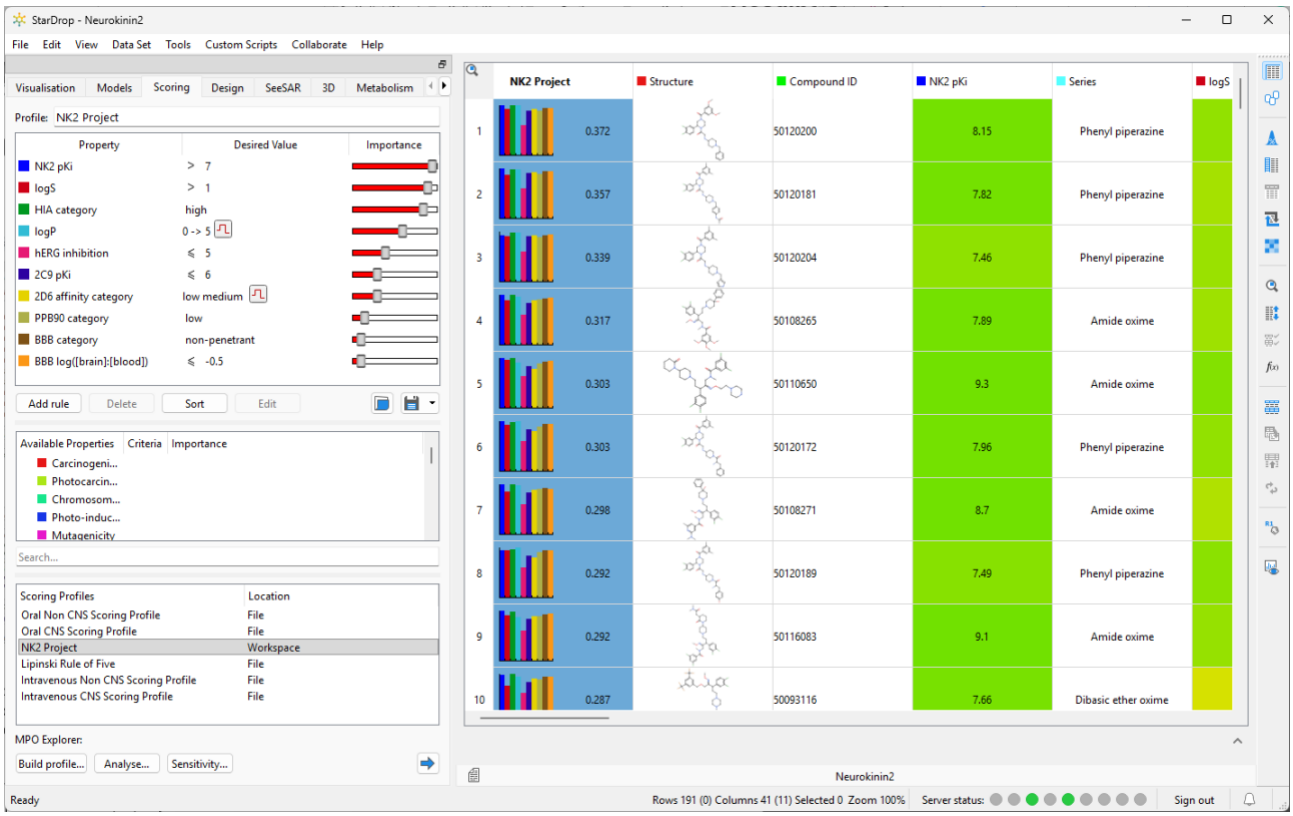


- Colour the data table by the Scoring Profile by clicking the **Heat Map** icon  .



- In the **Heat Map Colouring** window, tick **Show Heat Map** and select **Scoring Profile**. Close the window.





Each value in the data table will be coloured to show its contribution to the overall score for the selected scoring profile, taking into account the associated uncertainty.

Note: You can customise the colours and properties of the Heat Map using the **Colour Settings** in the **General Preferences** under the **File** menu.

As you scroll down the data set through decreasing scores, the low bars in the histograms indicate properties that have not met the requirements defined in the project profile, considering the confidence in the data and the importance of each property.

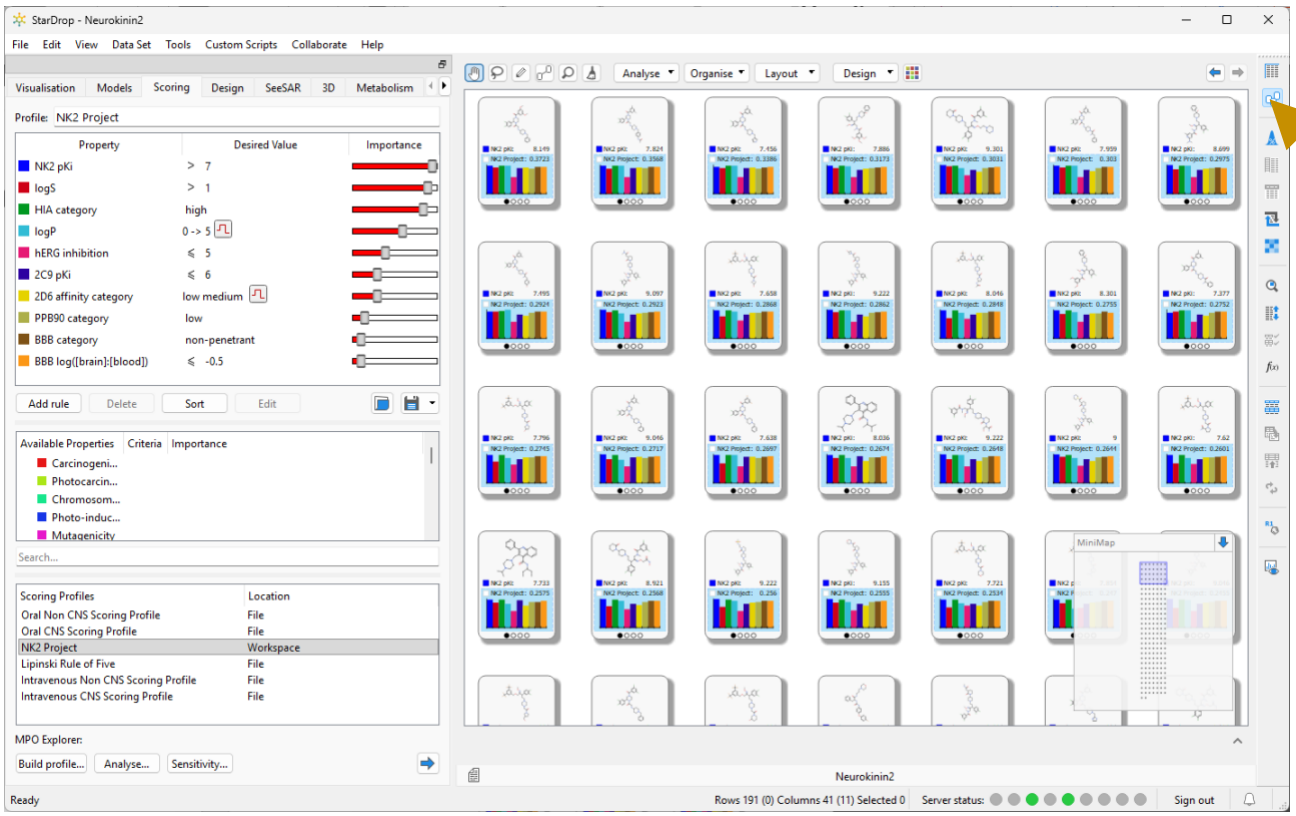
- Click the **Heat Map** icon and untick **Show Heat Map**.

Having scored the compounds, we're now going to compare some compounds using StarDrop's Card View®.

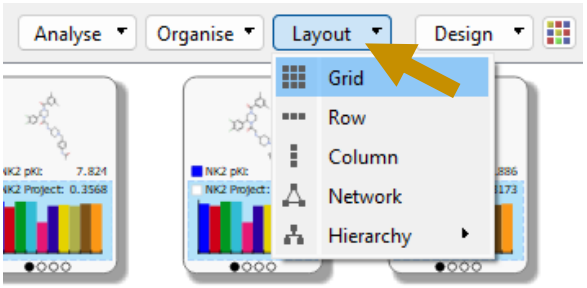
Using CardView to visualise important relationships

StarDrop's Card View® provides a convenient way to work with our compounds and data by representing compounds on cards that can be moved, stacked and linked however we wish.

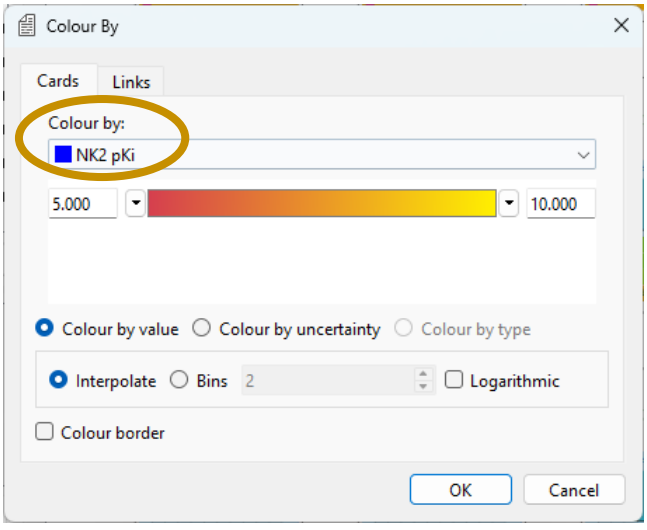
- Switch to **Card View** by clicking the **Card View** button in the toolbar to the right. This displays the data set as a grid of cards. You can choose which properties you would like to see on a card, but in this case we have already created a card design that shows the scoring profile and some of the properties we are interested in.
- Go to **Design**, **User Designs**, and select the NK2 card template.



- Sort the grid by the score from top-left to bottom-right by choosing the **Grid** option from the **Layout** menu at the top of Card View.

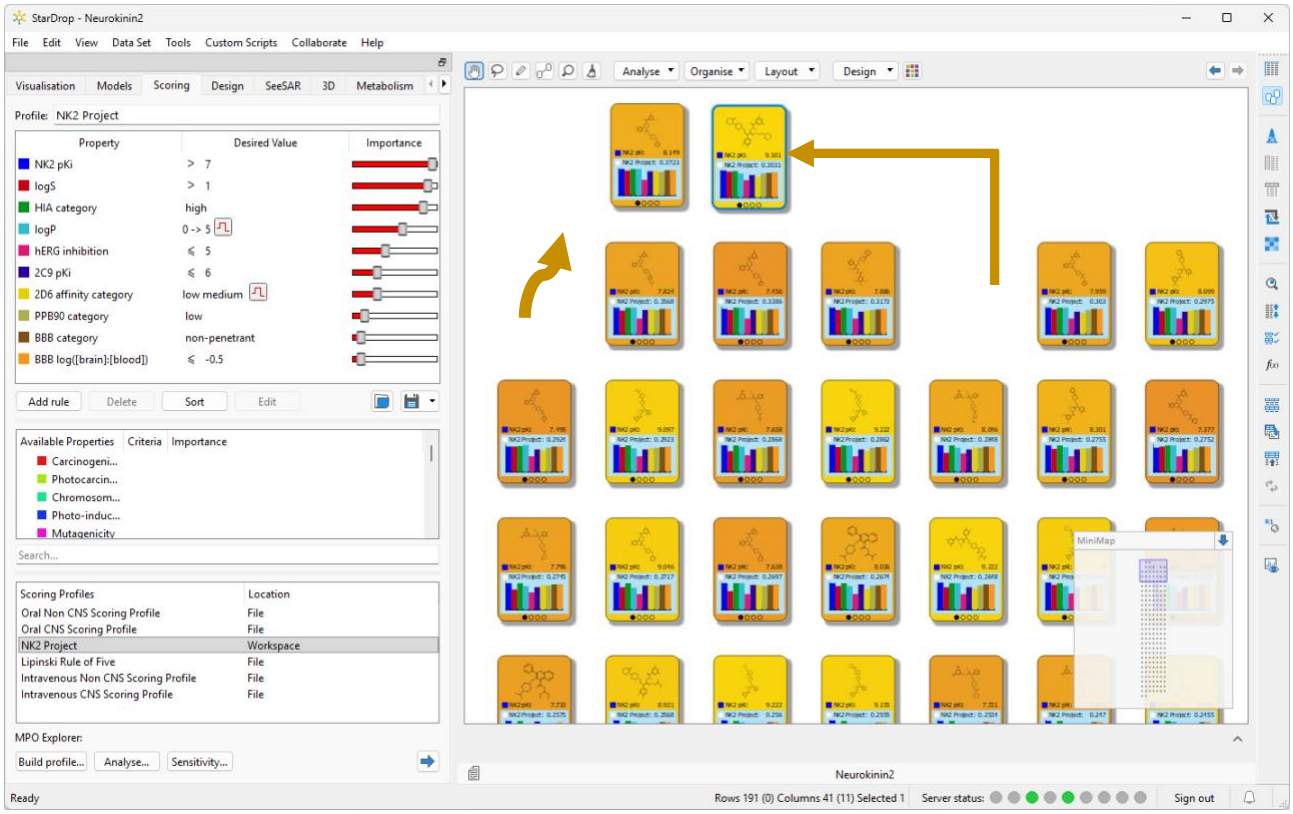


- Colour the cards to highlight the most potent by clicking on the **Format** button  and in the **Colour By** dialog, choosing to colour the cards by **NK2 pKi**.



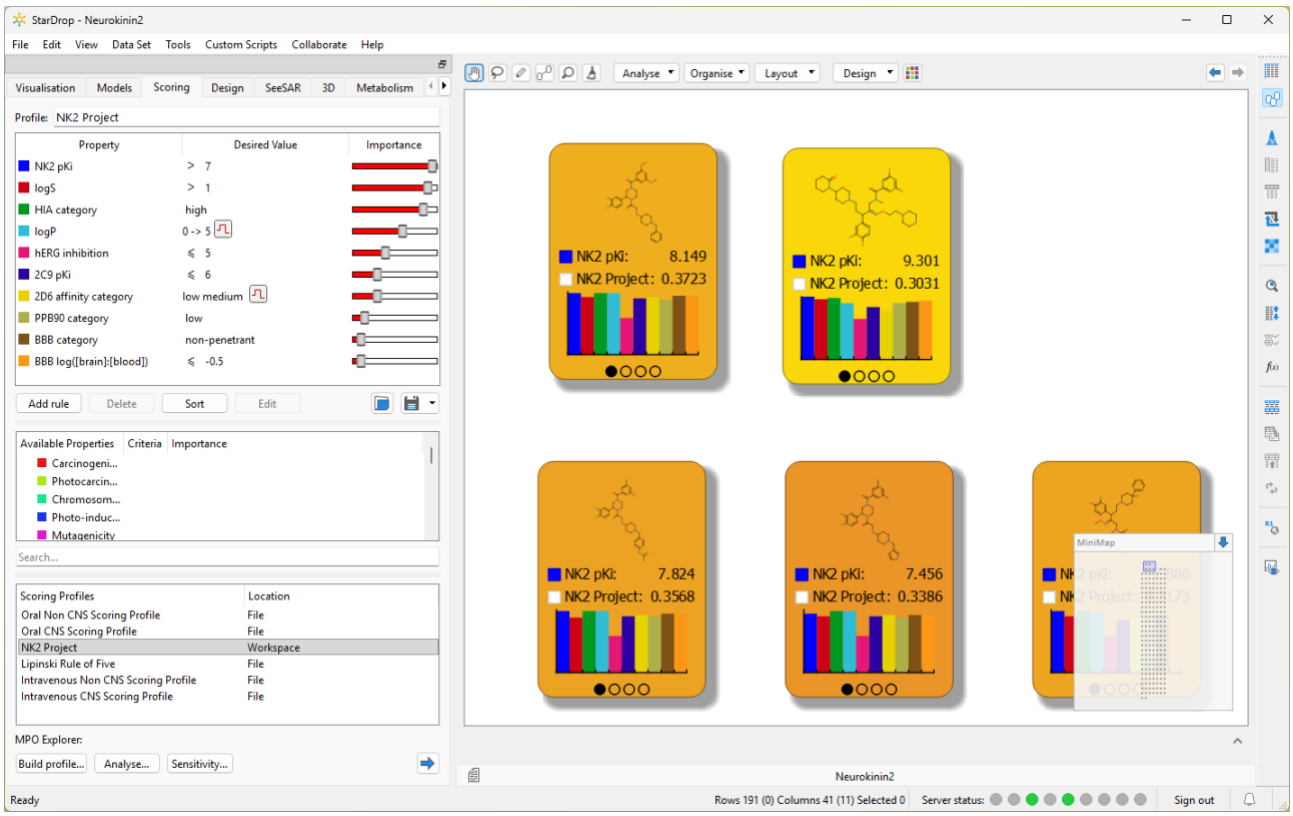
We can easily compare two compounds by placing their corresponding cards side by side.

- Click on the highest-scoring compound and drag it to the top. You can make space by dragging the background if necessary.
- Now select the fifth compound, which is one of the most potent, and drag it to lie beside the first.



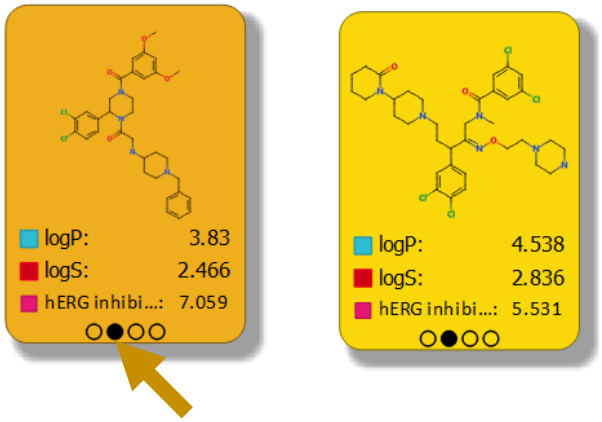
- Zoom in on these cards to look at them in more detail by pointing at the cards and using the mouse wheel (or see navigation hint below). Dragging the background will also help you to position the cards conveniently.

Navigation hint: You can click and drag on white space to move around the Card View layout. Zoom in and out by scrolling with the mouse wheel, using pinch gestures on the trackpad or touch screen or using **Ctrl** and **+** keys and **Ctrl** and **-** keys to zoom in and out. A **MiniMap** in the bottom right-hand corner provides easy navigation around the Card View interface by clicking any location within the MiniMap. You can open and close the MiniMap using the blue arrow at the top corner of the MiniMap box. You can also use the **Move Back** or **Move Forward** buttons  at the top right of the card layout to reset a view following any changes you may have made. For more details about using Card View, please see Section 5 of the User Guide.



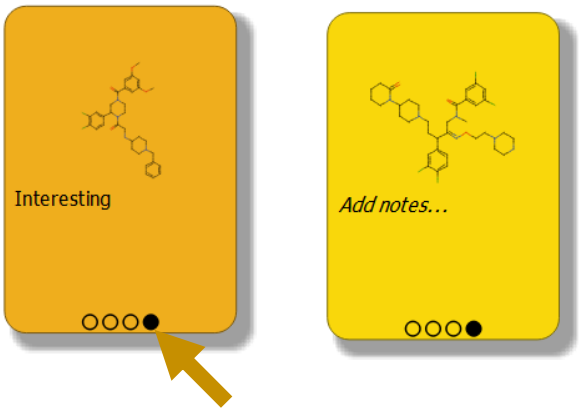
Here, we can see that the compound on the left has a higher score, even though the one on the right is among the most potent. We note by comparing the histograms from the scoring profile that the compound on the left has a better balance of ADME properties. We can investigate this further:

- Click on the second circle at the bottom of a card to change to the second 'page'.



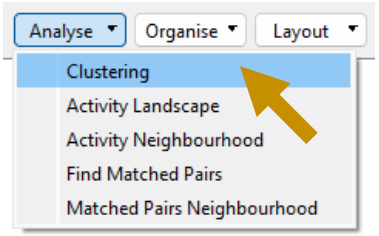
Here we can see that both compounds have a hERG liability, and the compound on the right also has a high logP, reducing its score based on the project goals.

- Click on the fourth circle now to go to the notes page and then double-click on ***Add notes...*** to add your own text, which will be saved in a **Notes** column in the data set.

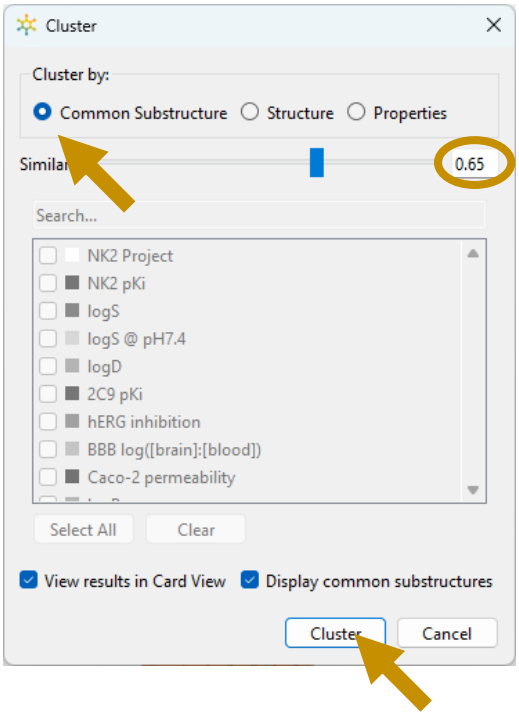


Having compared individual compounds, we would also like to explore how the properties vary across the different chemical series in this library. Clustering provides a convenient way to group compounds by chemical similarity.

- From the **Analyse** menu at the top of Card View, select **Clustering**.

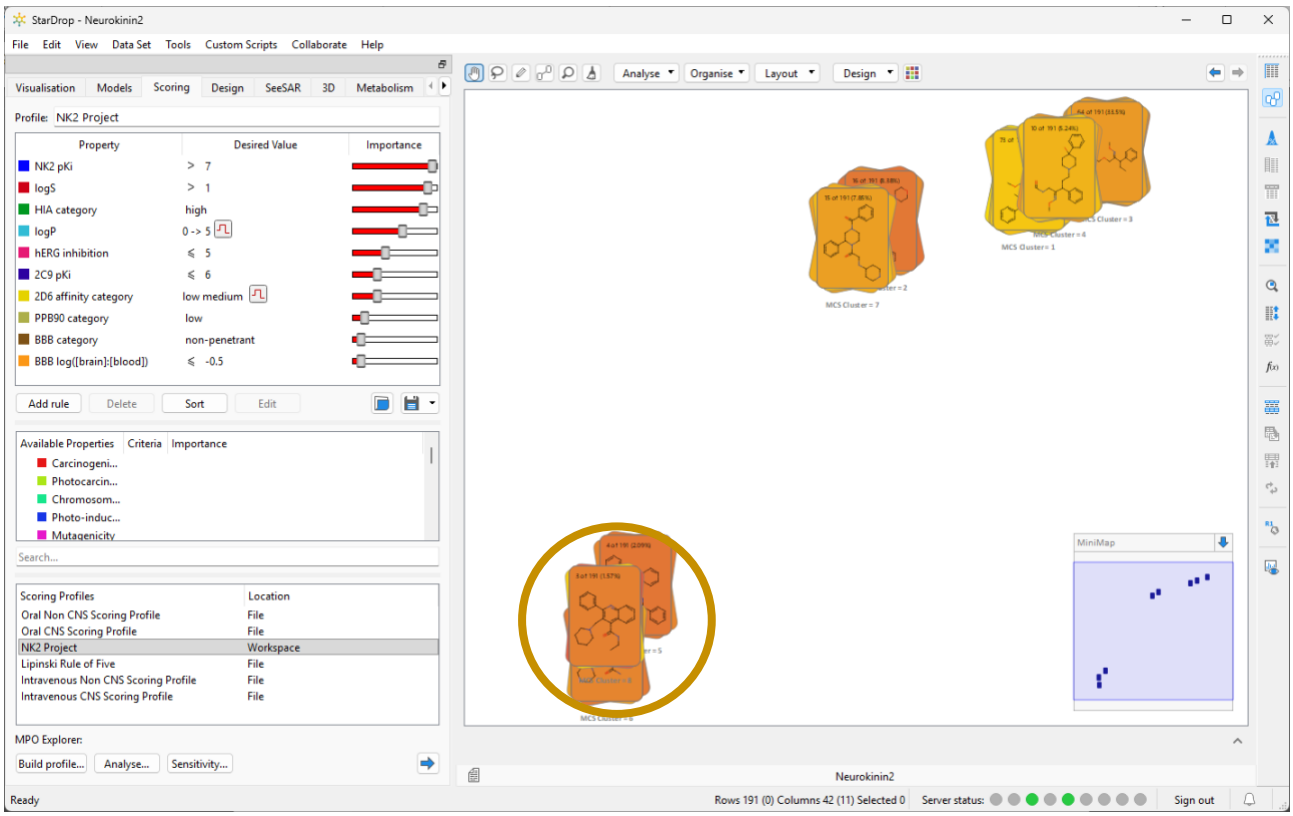


- In the **Cluster** dialog, use the default **Common Substructure** method. Set the **Similarity** to **0.65**, as shown right, and click the **Cluster** button.

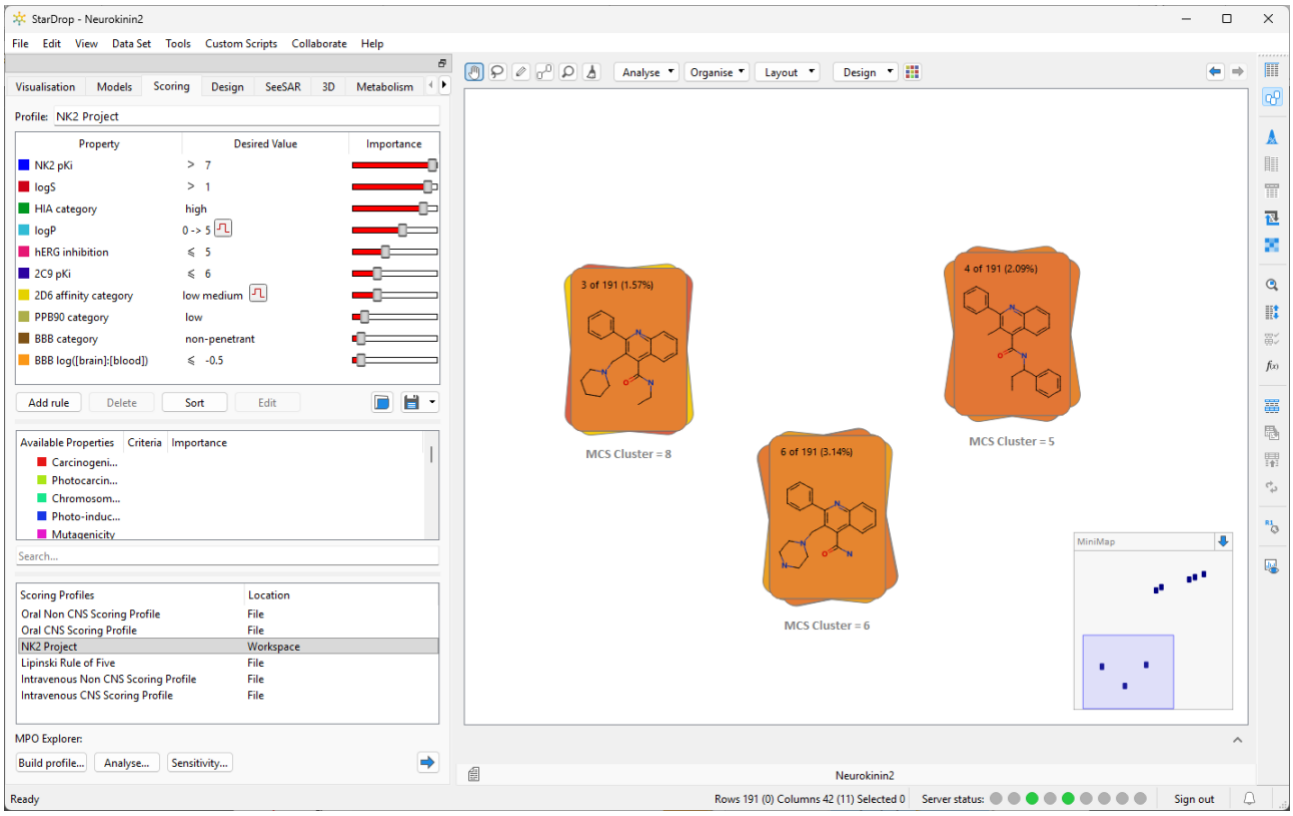


The compounds are grouped by common substructure to form 'stacks' containing multiple compounds. The stacks are positioned such that stacks with similar common substructures are close together.

Note: You may see the stacks laid out in a different orientation, but you should see the same clusters of stacks shown in the example below.

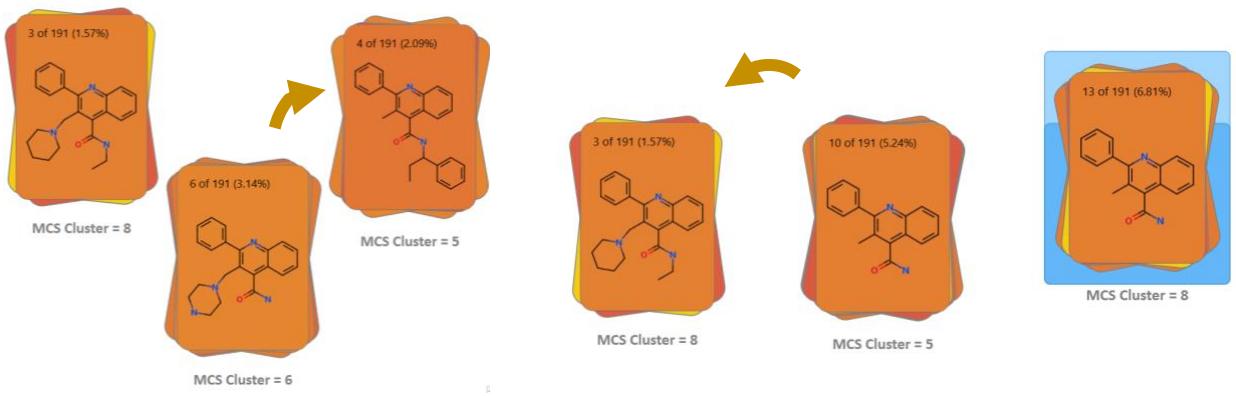


- Zoom in on the stacks corresponding to clusters 5, 6 and 8, which are close together and drag the background to centre these in the display. You can move the stacks to separate them slightly by clicking and dragging.



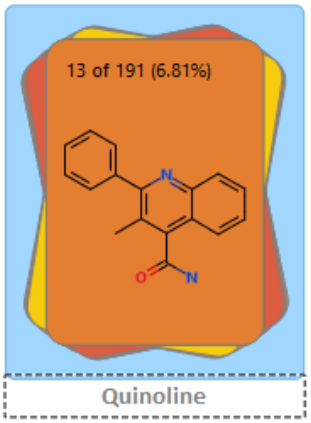
Although the compounds are in different stacks, we can see that this is an artefact of the cut-off assigned to the clusters, and these groups all share the same quinolone scaffold and could be considered a single series.

- Combine these stacks by dragging one stack on top of another.

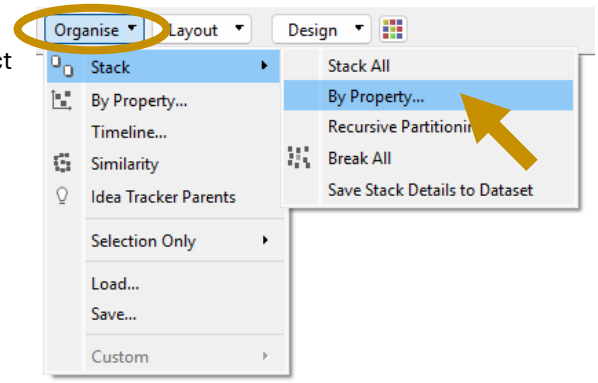


- Assign a name to this series by clicking and editing the label under the stack.

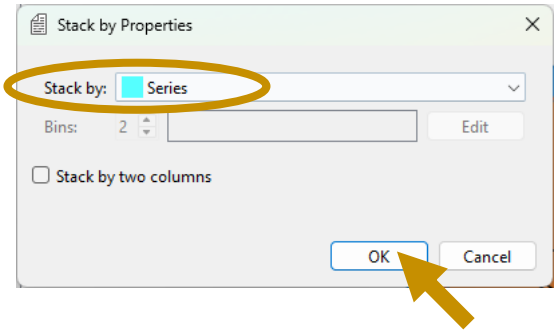
We could examine the other stacks and repeat this process, where necessary, to refine the clustering results. However, to save time, we have saved the series definitions in the data set and can use this to stack all the compounds.



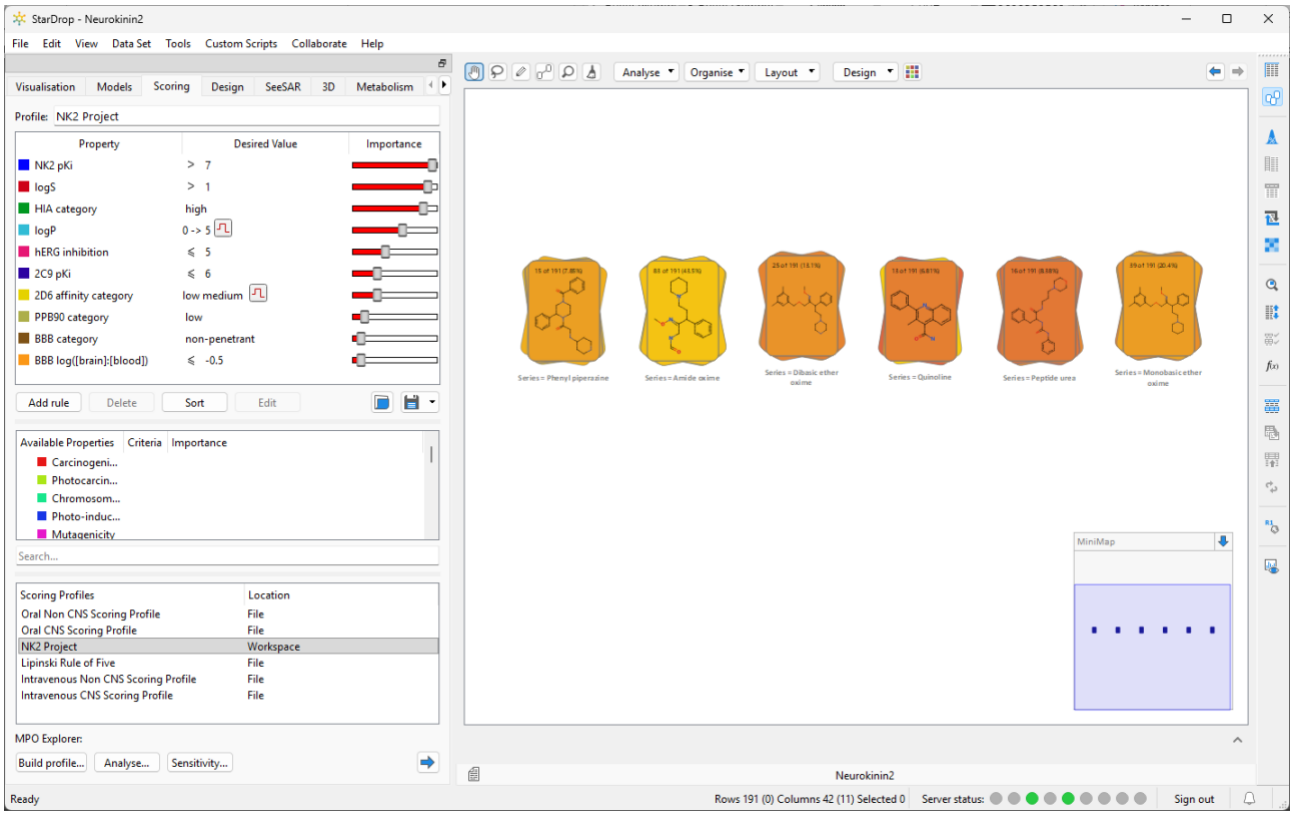
- From the **Organise** menu at the top of Card View, select **By Property** from the **Stack** menu.



- In the **Stack by Properties** dialog choose **Series** from the **Stack by** menu and click the **OK** button.

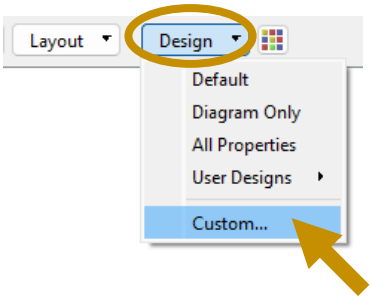


Here, we can see that there are 6 series, of which the Amide oxime series has the highest average potency (the most yellow).



We can display more information on a stack. To illustrate this, we can choose a new stack design.

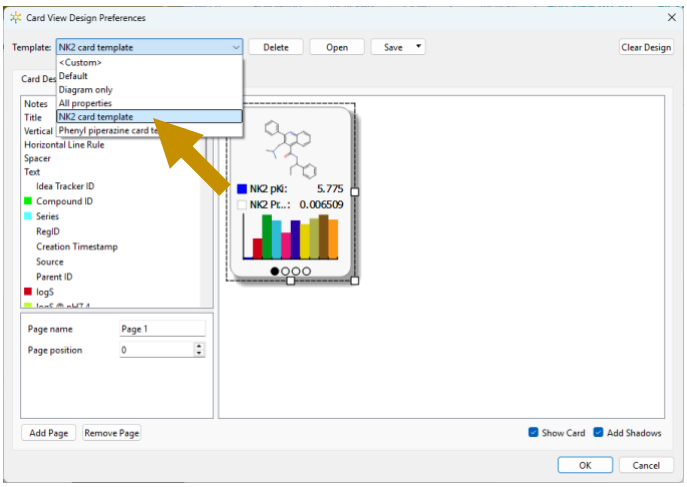
- Select **Custom** from the **Design** menu at the top of Card View.



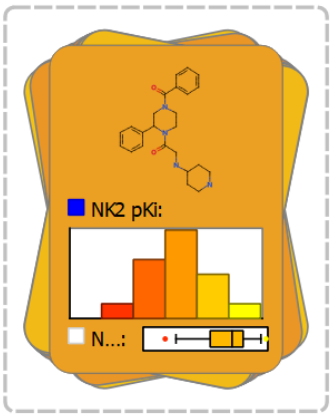
In the **Card View Design Preferences** dialog, you can change card sizes, choose which properties to display, and how they should be arranged. In this case, we are going to choose a predefined template saved as part of this workspace.

- Select **NK2 card template** and click the **OK** button.

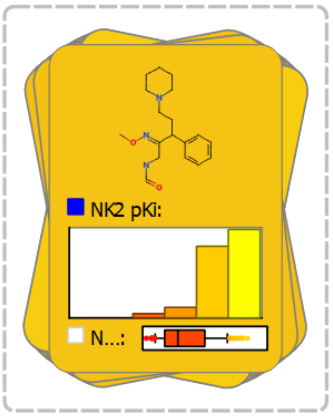
Note: You can also select previously saved designs directly from Card View's **Design** menu under **User Designs**.



The resulting stacks show a histogram of the distribution of pK_i values of the compounds within each stack and, below this, a box plot showing the distribution of scores. This makes it straightforward to compare different series. For example, the Amide oxime series has an excellent distribution of potency; however, most scores are very low due to poor results for other properties. In contrast, the Phenyl piperazine series has fewer potent compounds but a significant proportion of high-scoring compounds, suggesting that this would be a good series to investigate further.



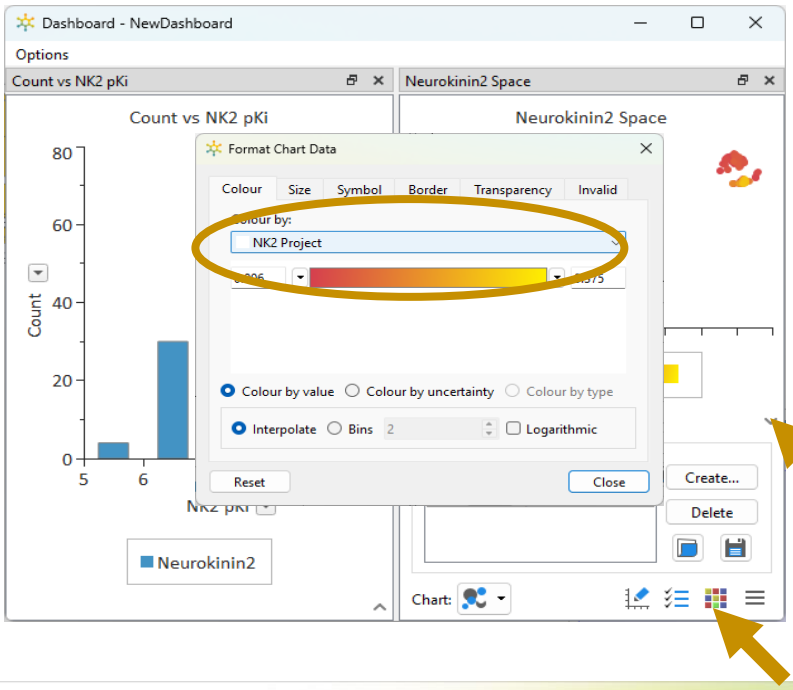
Series = Phenyl piperazine



Series = Amide oxime

Identifying high-quality chemistries for progression

We can explore the distribution of scores across the whole data set by returning to the dashboard we created earlier.



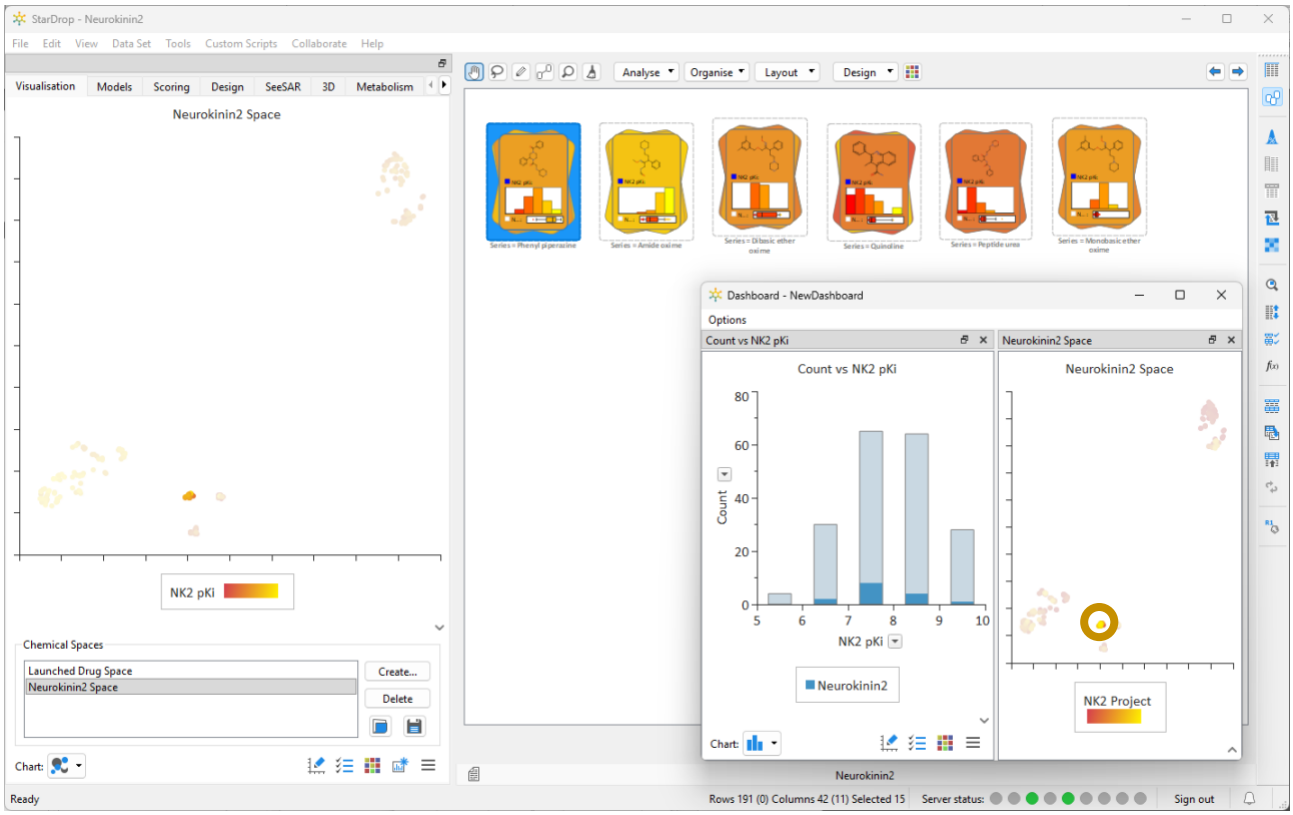
- Go to the **Visualisation** area again, click on the **More options** button at the bottom and choose **Open Dashboard**.

- Click the arrow in the bottom corner of the Chemical Space plot to display the controls and click on the **Format** button at the bottom to bring up the **Format Chart Data** dialog again.

- Change the **Colour by** property to **NK2 Project**, before clicking the **Close** button.

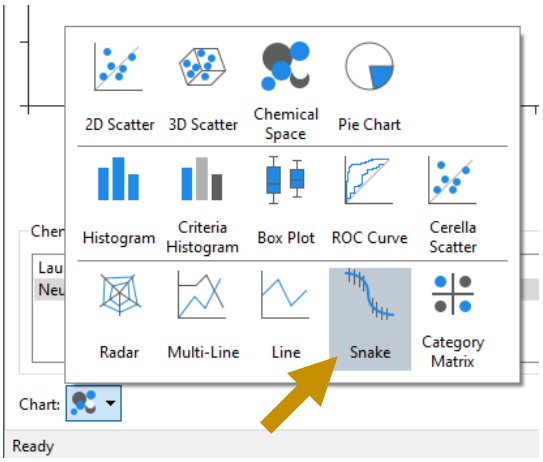
The points are now coloured by the compounds' overall scores, not just their pK_i values. We can see that the pattern has changed dramatically from what we saw when looking only at potency. There is just one small cluster of yellow points (highlighted by the orange circle) as a 'hot spot' of high-scoring compounds.

- Using the mouse, draw around this yellow cluster in the chemical space. You will see that the stack corresponding to the Phenyl piperazine series is highlighted, confirming that the hot spot corresponds to this series.



Although this appears to be the best series, we should consider the impact of the uncertainty in the underlying data on our ability to distinguish between compounds confidently. Therefore, we will use another type of chart to explore this.

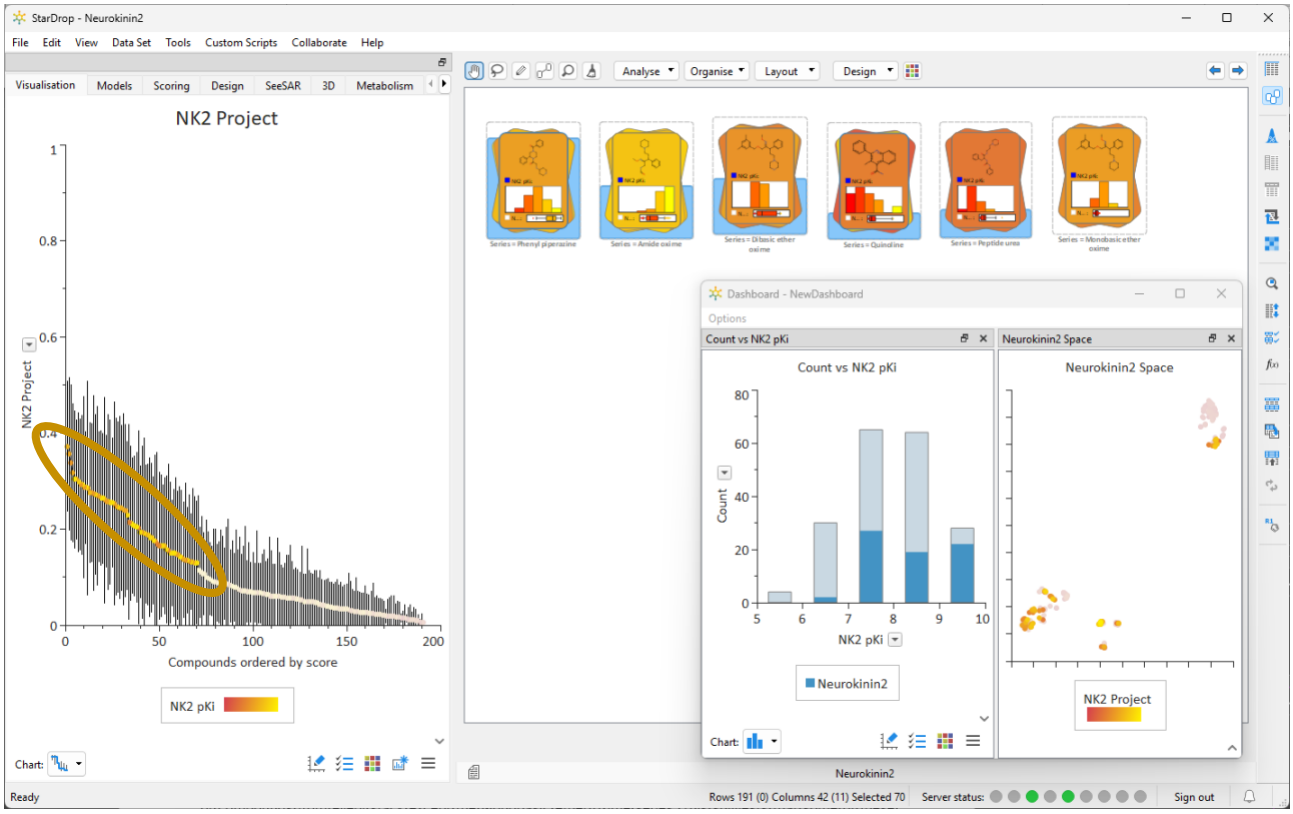
- At the bottom of the **Visualisation** area, open the **Chart** menu at the bottom and choose a **Snake** plot (as shown to the right).



A 'snake plot' shows the scores (on the y-axis) for all compounds in order from highest score to lowest score (along the x-axis). The overall uncertainty in each score, due to the uncertainty in the underlying data, is also displayed as an error bar around each point.

The top 70 compounds in the snake plot cannot be confidently distinguished from the top-scoring compound (notice that the error bar for the top-scoring compound overlaps with the error bars of approximately the top 70 compounds). Therefore, we should consider exploring the properties of the compounds in this range further so that we can make a confident selection of a potential lead series.

- Select the top 70 compounds in the snake plot by lassoing them with the mouse.

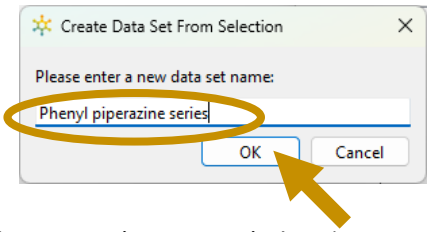


Notice that this selection includes compounds from different regions of the chemical space and a small number of compounds from every stack except the Monobasic ether oxime series. This suggests that some of these chemistries cannot be rejected with confidence. Therefore, it may be more appropriate to sample a small number of compounds from these alternative chemistries to generate some experimental ADME data. These data will have lower uncertainty than predicted values, enabling us to identify with greater confidence the chemistries that will yield a high-quality lead series.

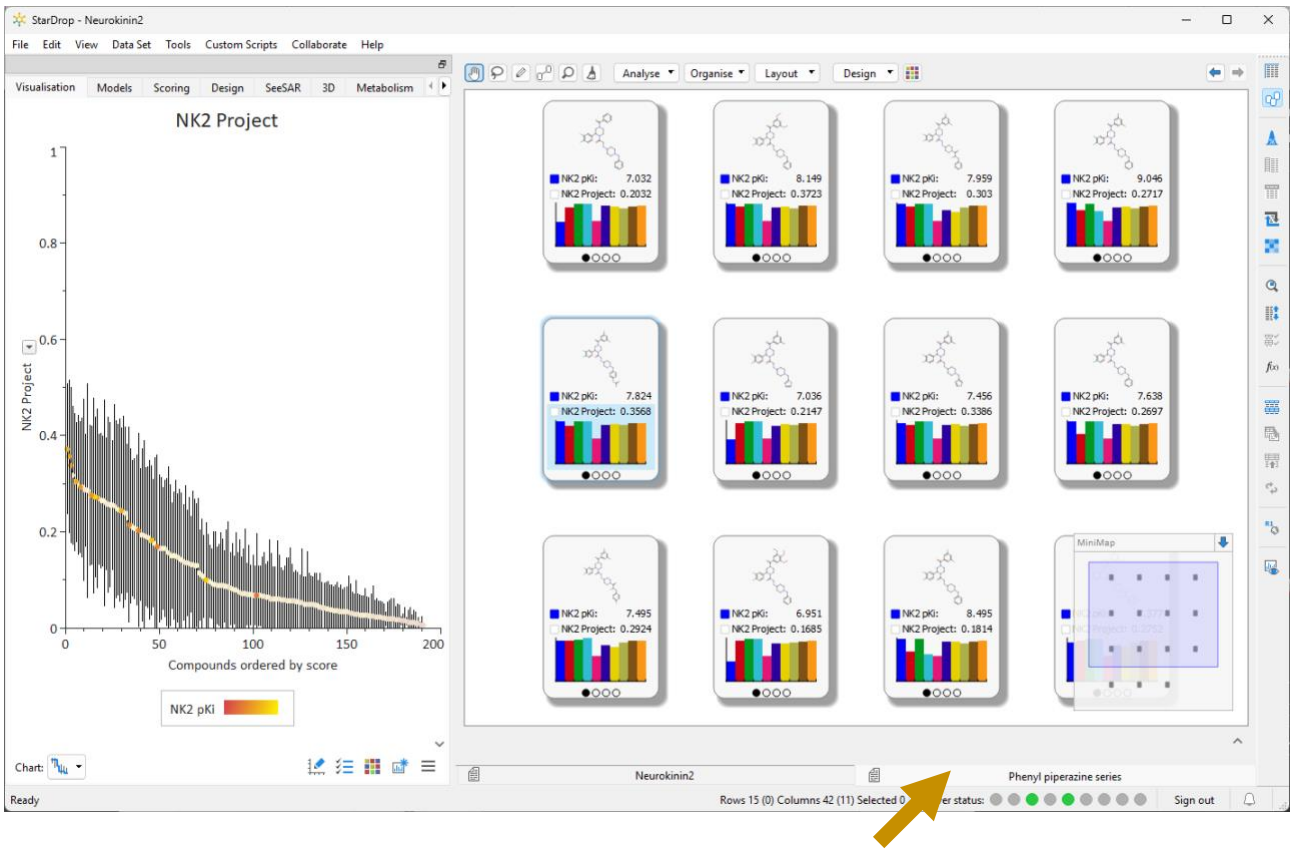
Identifying activity cliffs to elucidate SAR

Having looked at the data across this library, let's focus on the Phenyl piperazine series to see if we can identify any interesting structure-activity relationships with which to guide the design of further improvements within this series.

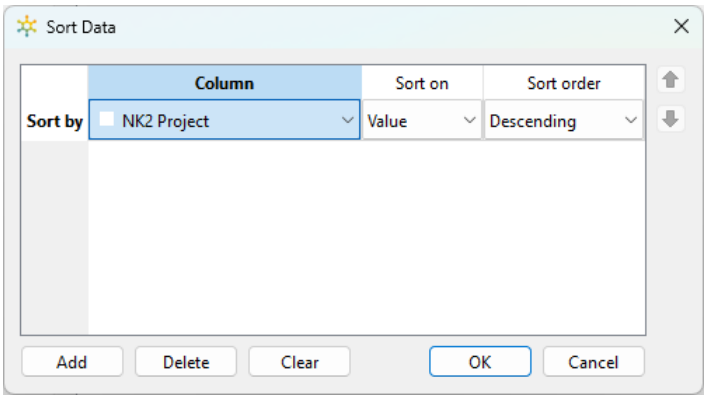
- Minimise the dashboard again.
- Select the stack corresponding to the Phenyl piperazine series, and from the **Data Set** menu, select **Create From Selection**.
- In the **Create Data Set From Selection** dialog, give the new data set a name, e.g., "Phenyl piperazine series", and click the **OK** button.



This copies the 15 compounds from this series into a new data set which appears in a new window (you can change back to the full data set by clicking on the tab at the bottom).

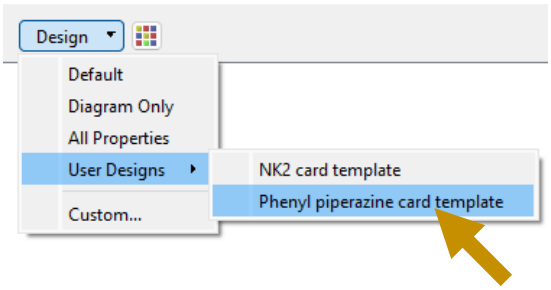


- Sort the cards in order of descending score using the Sort Data dialog by clicking on the **Sort** button in the toolbar on the right.
- The score column **NK2 Project** is the first one in the data set and will already be chosen, so click the **OK** button.



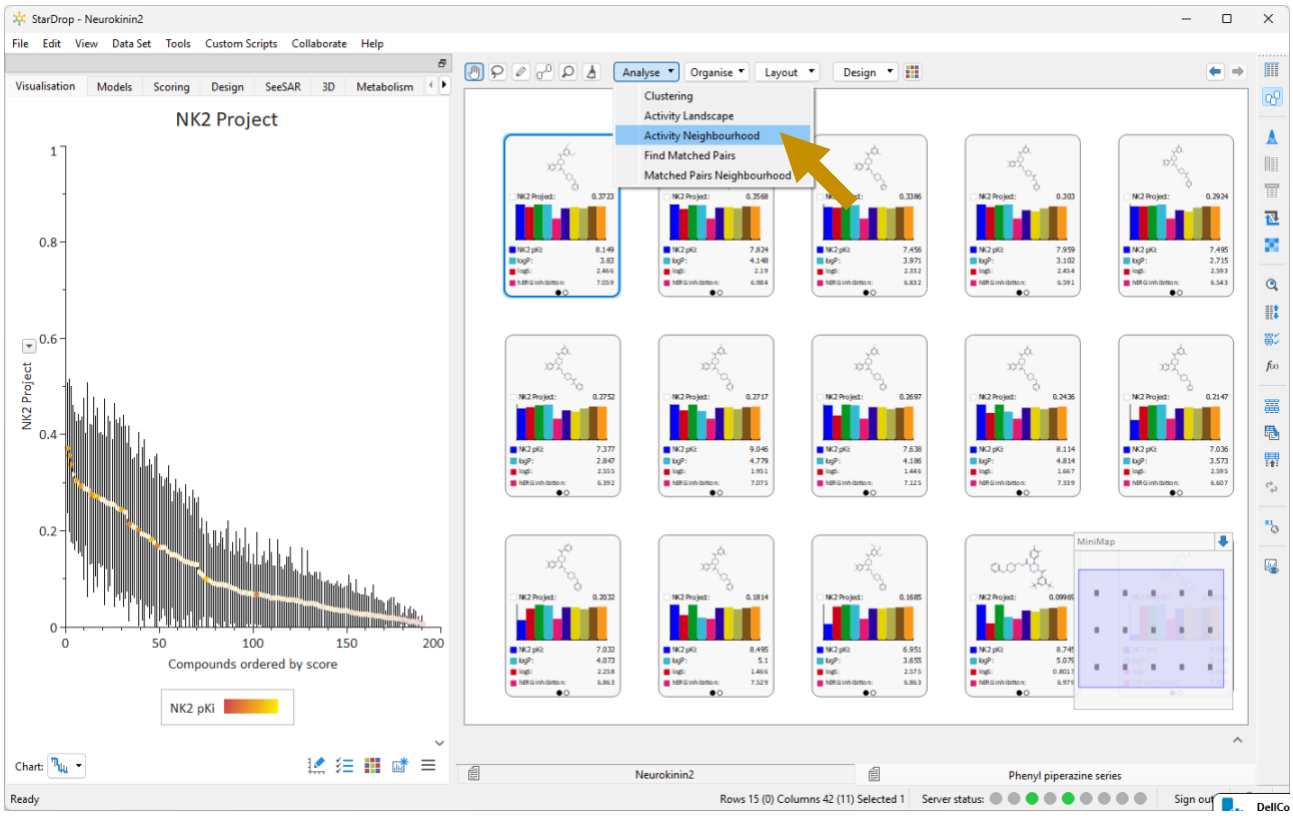
We'd like to look at these compounds in more detail, so we'll load a new card template that displays more data on each page. When loading a saved card template that is already part of a workspace, it is not necessary to open the **Card View Design Preferences** dialog because the template will be available in the **User Designs** menu.

- Select the **Design** menu, then **User Designs** and then **Phenyl piperazine card template**.



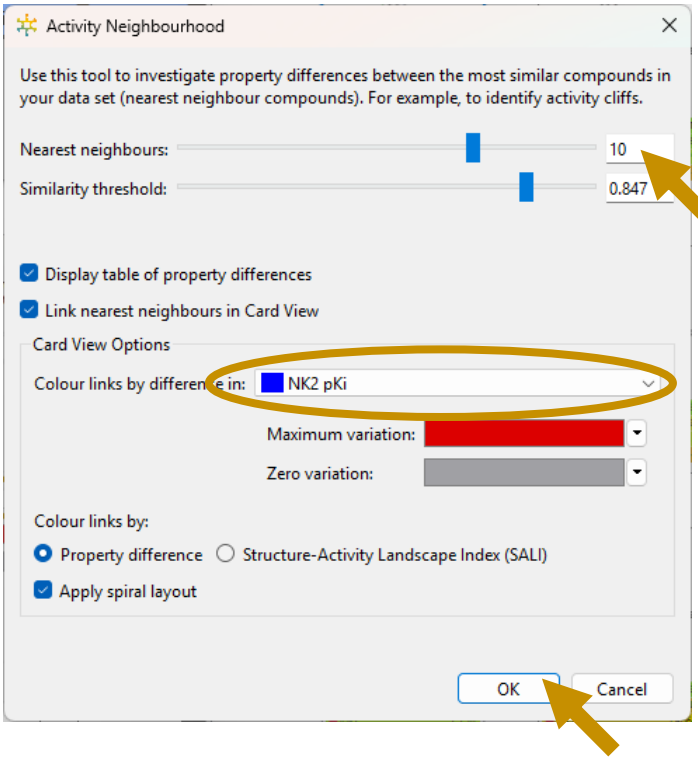
We will use the Activity Neighbourhood analysis to explore how small structural modifications can influence potency around the highest-scoring compounds (this is also known as activity cliff detection).

- Select the card in the top-left of the grid (this is the highest-scoring compound in this series).
- Now select **Activity Neighbourhood** from the **Analyse** menu at the top of Card View.

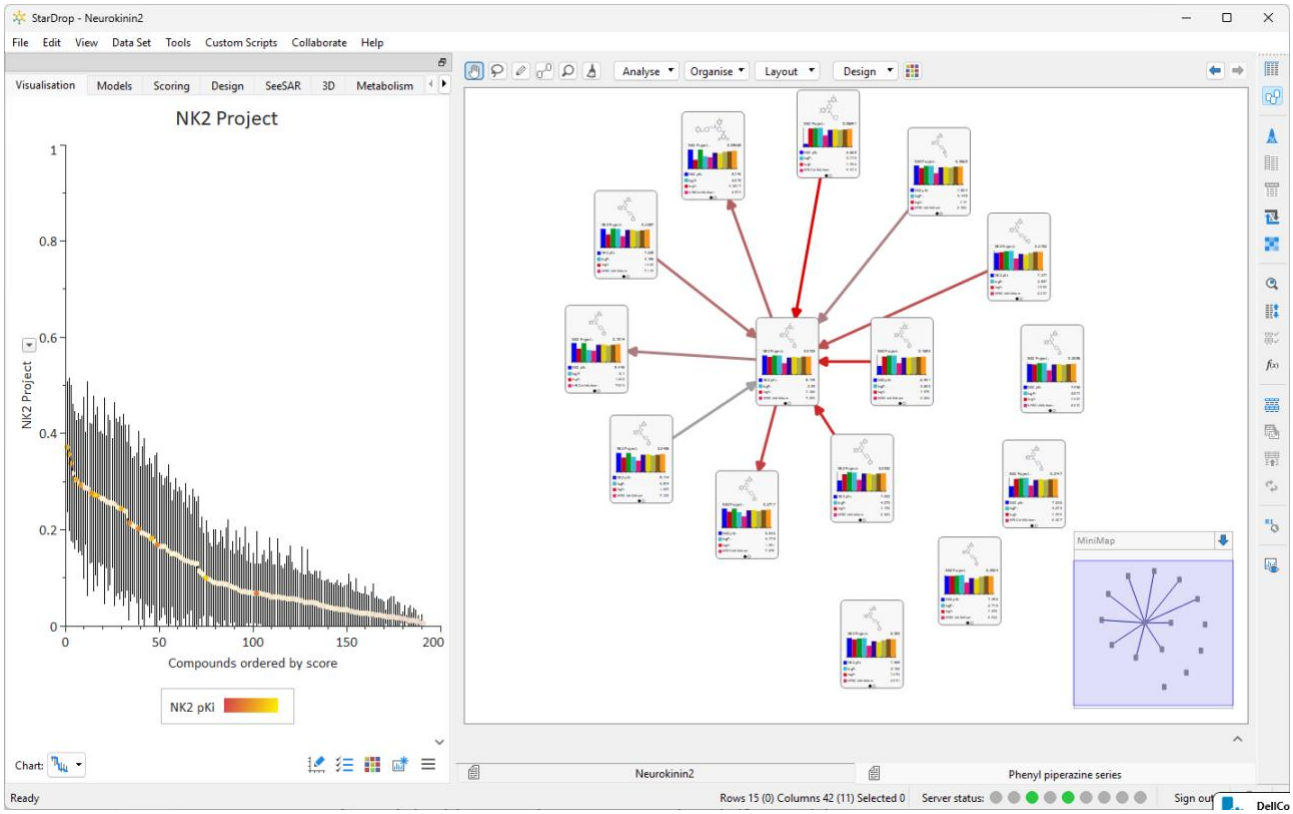


The Activity Neighbourhood analysis compares each compound with the selected 'reference' compound and arranges these in order of structural similarity. The nearest neighbours are linked to the reference, and that link can be coloured by the change in a selected property.

- In the **Activity Neighbourhood** dialog, keep the default number of nearest neighbours (10) and choose to colour the links by the difference in **NK2 pKi**. Then click the **OK** button.



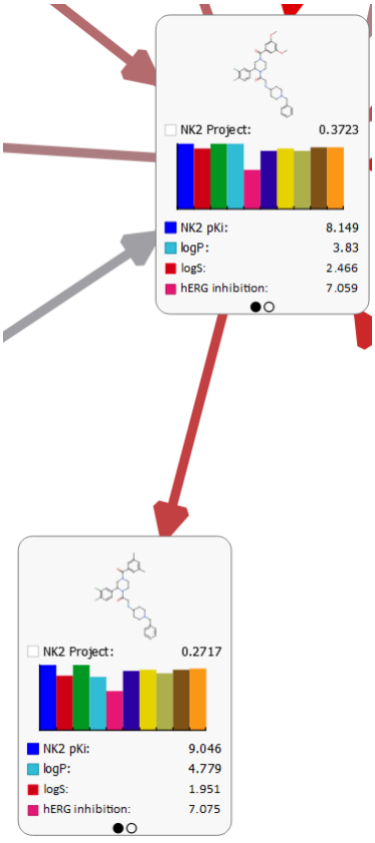
The resulting display shows the reference compound in the centre and the remaining compounds arranged in a spiral in order of decreasing similarity (increasing distance) from the reference. The 10 closest compounds are joined to the reference by a link; the arrow indicates the direction of *increase* of the selected property, in this case, NK2 pK_i. The colour indicates the magnitude of the change, in this case, from red being the largest to grey being zero change. A short link with a bright, red colour indicates an activity cliff, i.e., a small change in structure that gives rise to a large change in activity. One example of this can be seen between the reference compound and its closest neighbour.



Note that a table summarising the property changes for the nearest neighbours is also displayed, but we won't use it in this example, so you can close it if you wish.

- Zoom in to the central compound to compare it with its third closest neighbour (using the middle mouse wheel or **Ctrl** and **+** keys).

Here, we can see that replacing the two methoxy groups with methyl groups has resulted in almost a 1 log-unit improvement in potency (from about 7 nM to <1 nM). However, the overall score is reduced because of the poorer balance of ADME properties, as we saw earlier.

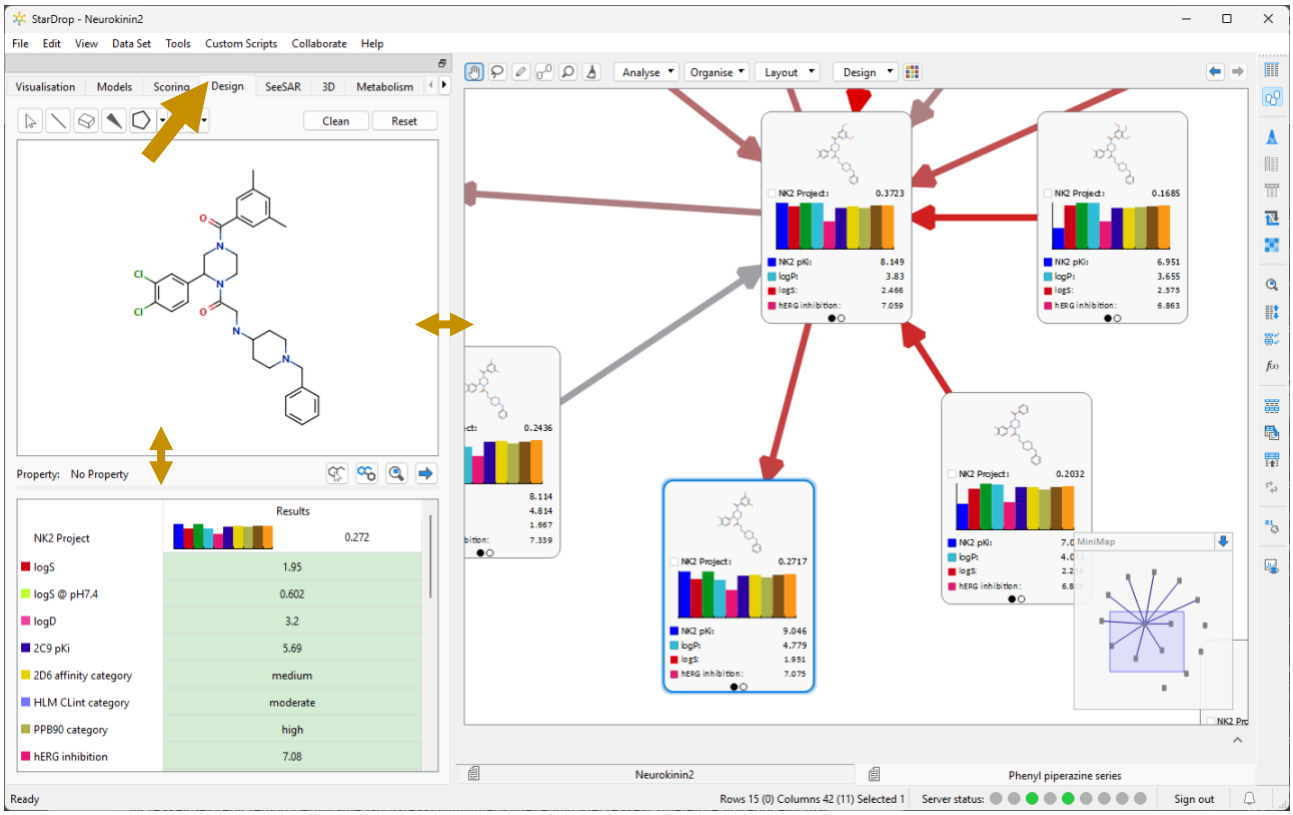


Note: the two compounds considered here are not matched pairs and would be missed by a matched pairs analysis because there is a change at two positions on the compound. The activity analysis complements matched pair analysis and picks up these compounds as being very similar to one another.

Interactive design

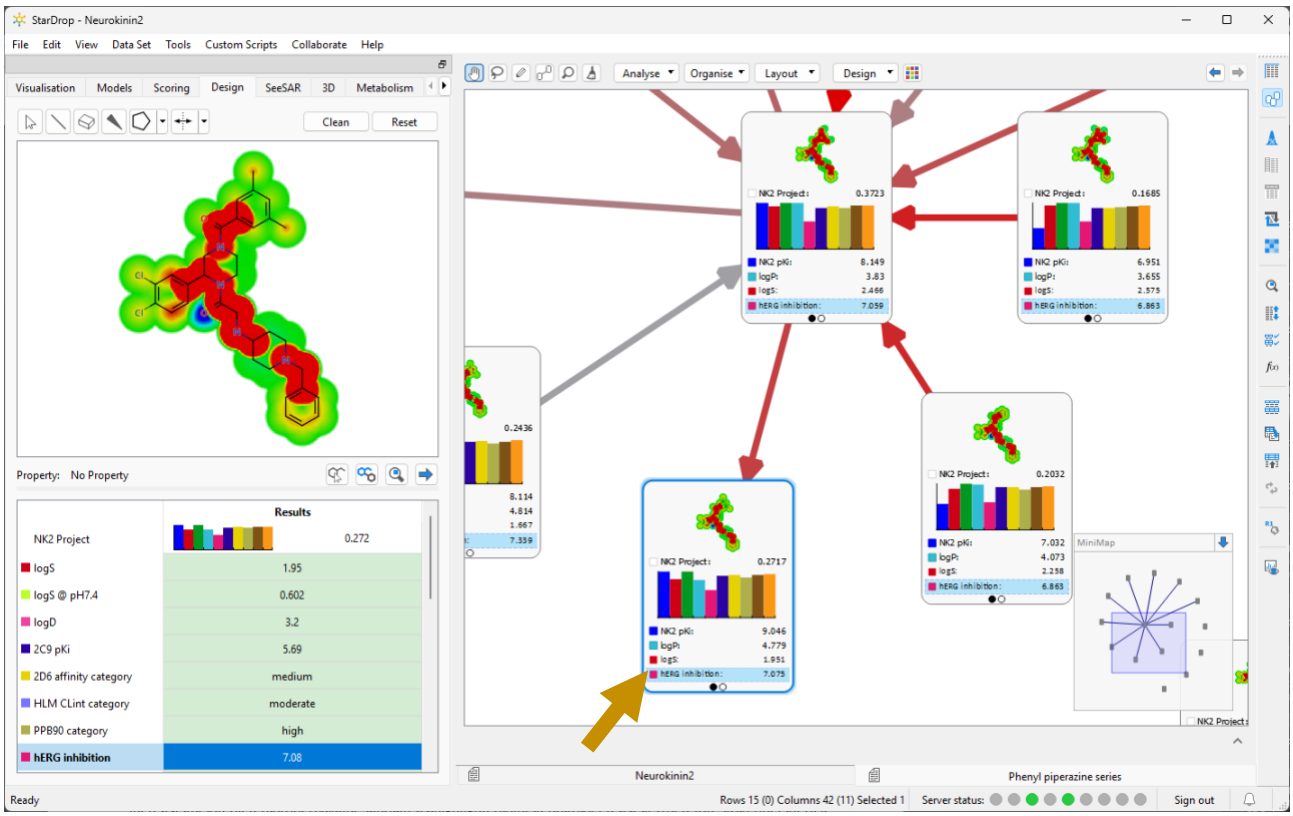
We might wish to explore alternative approaches to improving these properties while retaining potency against the target. We can use the interactive Design tool with the Glowing Molecule™ to do this.

- Click on the bottom compound of the pair (the one with the higher pKi) to select it. It will get a blue border.
- Switch to the **Design** area to see the selected compound in the interactive designer. If necessary, you can zoom out to show to entire molecule using the mouse wheel.



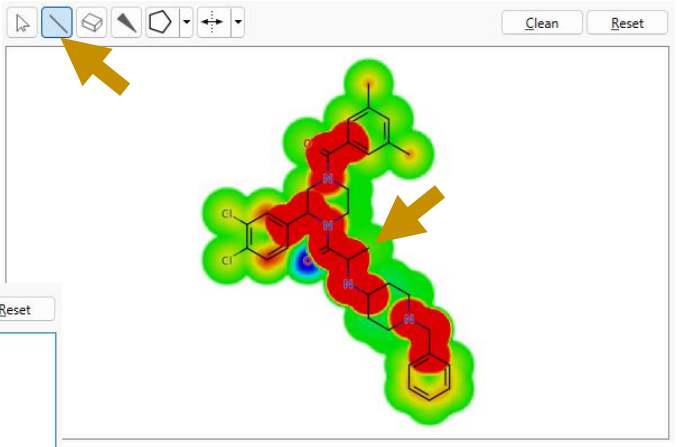
Note: You can drag the bar between the drawing area and the summary table or between the drawing area and the data set to resize the drawing area.

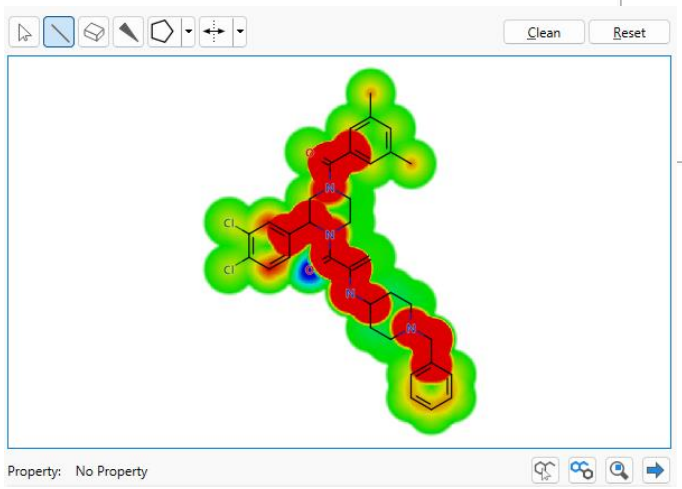
- Click on the **hERG inhibition** value in Card View to display the Glowing Molecule.



This highlights each molecule, using StarDrop's Glowing Molecule, to indicate which regions of the compound structure are having the greatest influence on the predicted value for hERG pKi. Areas highlighted red tend to increase the predicted property value, and areas highlighted blue tend to decrease it. Therefore, to reduce hERG inhibition, we should remove a group highlighted red or introduce one that is highlighted blue, bearing in mind that the substitutions might be beneficial or detrimental to NK2 potency. Note that the central linker region of the molecule is highlighted in red. We will change this amine to an amide and explore the impact of this change.

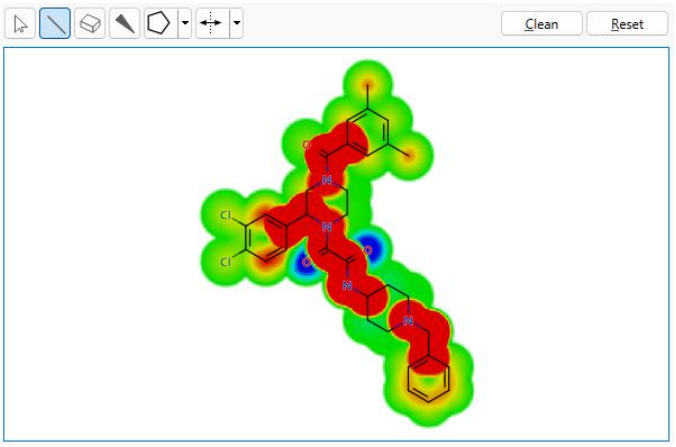
- Select the **bond** tool  and click and drag on the carbon next to the amine to draw a methyl group. Make sure you don't connect this to any other atom! (If you do, hover the mouse pointer over the drawing and press **Ctrl** and **Z** to undo it.)





- Click on the new bond you have drawn to make this a double bond.

- Finally, point the mouse at the carbon at the end of the double bond and press the **O** key to replace the Carbon with Oxygen and complete the amide.



While editing the compound, you will see that all the predicted properties in the summary table below the drawing area update as the molecule is being modified. Notice that the amide O has a blue glow, indicating that it has an effect that is decreasing the predicted hERG pKi.

This change has reduced not only the predicted hERG affinity but also the predicted logP. The predicted properties are shown in the table below the sketched molecule in the design area. The hERG pKi has decreased from 7.075 to 6.815, while the logP has decreased from 4.779 to 4.336 for this molecule.

Feel free to explore the effects of other changes on the predicted properties of these compounds. To add this, or any other compound you design, into your data set, simply click the  button in the **Design** area.

At the top of the summary table, you will see that the score has decreased. This is because we do not yet have an experimental pK_i value for the NK2 affinity for this new compound. To generate predictions of NK2 affinity, we can use the StarDrop Auto-Modeller™ to build a new model of the data we have. We have not discussed the Auto-Modeller as part of this exercise, but if you would like more information, please contact support@optibrium.com.

Conclusion

This example illustrated how we can explore the chemical space of our workspace to identify compounds and series with a good overall balance of properties, and to investigate SAR within an individual series. We have also seen how we can focus on individual compounds and use the Glowing Molecule to help design compounds that better meet our project needs.

If you have any questions about this example or would like help applying these techniques in your work, please contact support@optibrium.com. Feel free to explore our online community at https://optibrium.com/resources/ for additional tutorials and videos.