

## Worked Example:

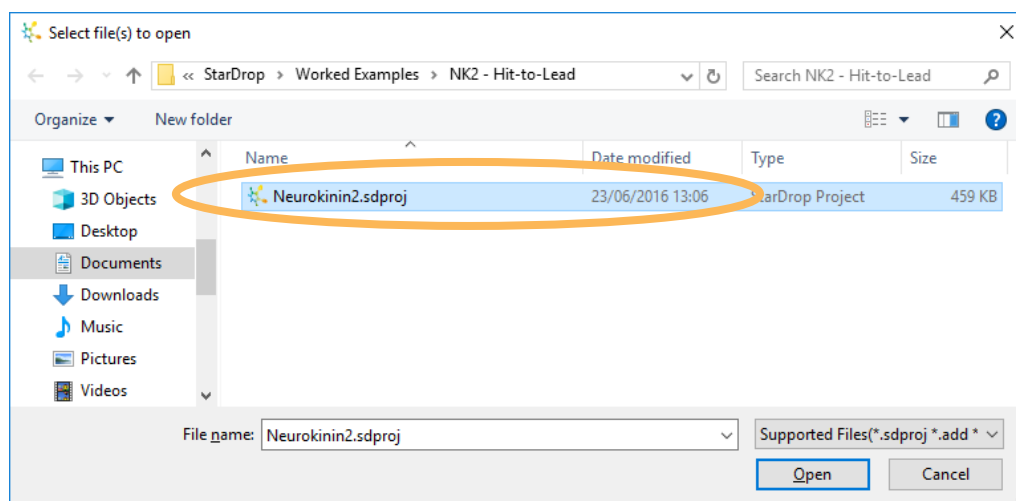
### Guiding Compound Selection and Design in Hit-to-Lead

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 not only be potent, but also have appropriate ADME properties to result in a high quality lead series. We will also use StarDrop to explore potential modification of one of the existing compounds to improve its properties.

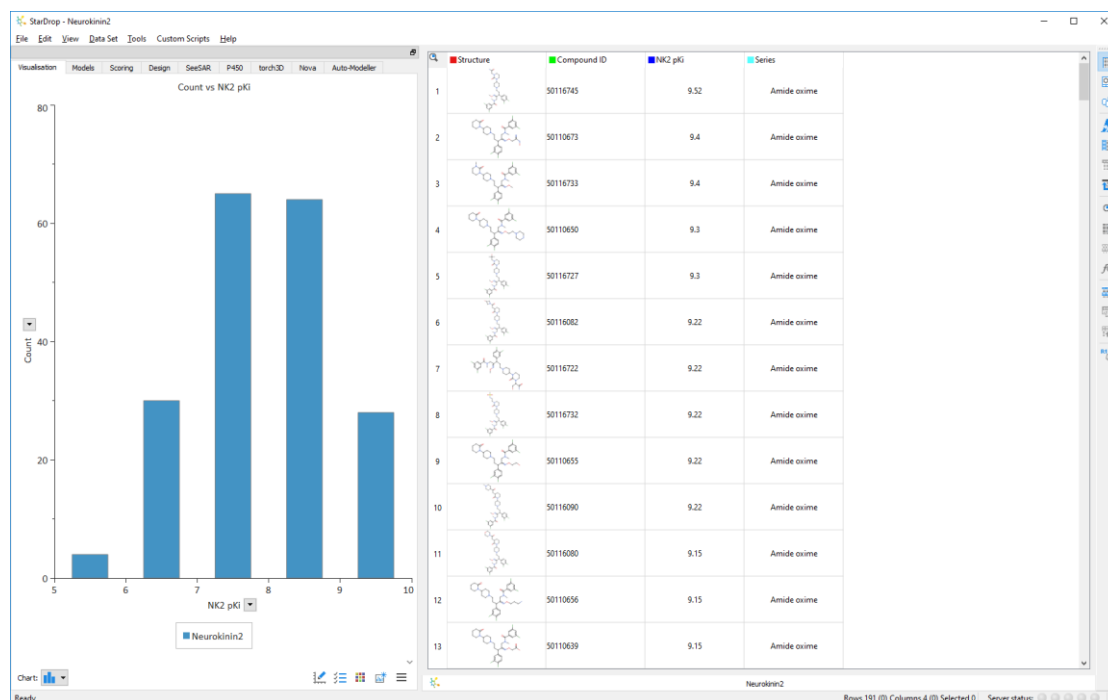
During this exercise we will use a variety of StarDrop's capabilities to explore the data in order to select and design compounds with a good balance of properties. Step-by-step instructions for all the features you will need to use in StarDrop are provided, along with screenshots and examples of the output you are likely to generate. If you have any questions, please feel free to contact [stardrop-support@optibrium.com](mailto:stardrop-support@optibrium.com).

### Exercise


- In StarDrop, open the file **Neurokinin2.sdproj** by selecting **Open** from the **File** menu.

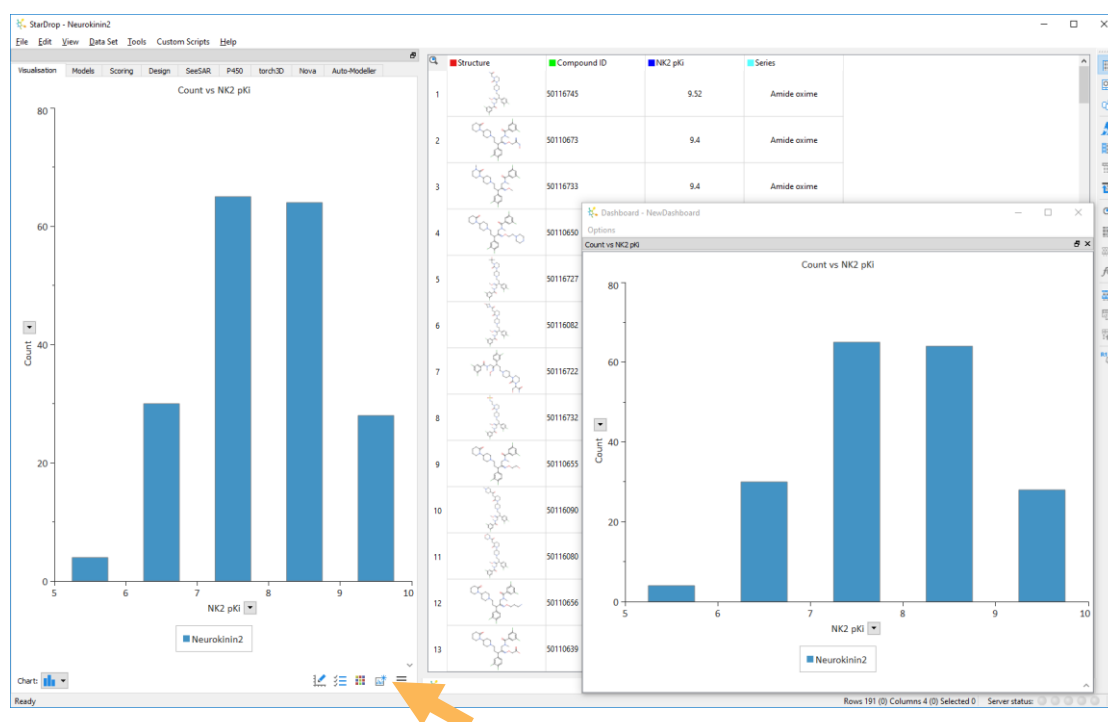


You will see a spreadsheet containing 191 structures and their measured affinities for Neurokinin 2 (in the column **NK2 pKi**) which are also shown in a histogram.

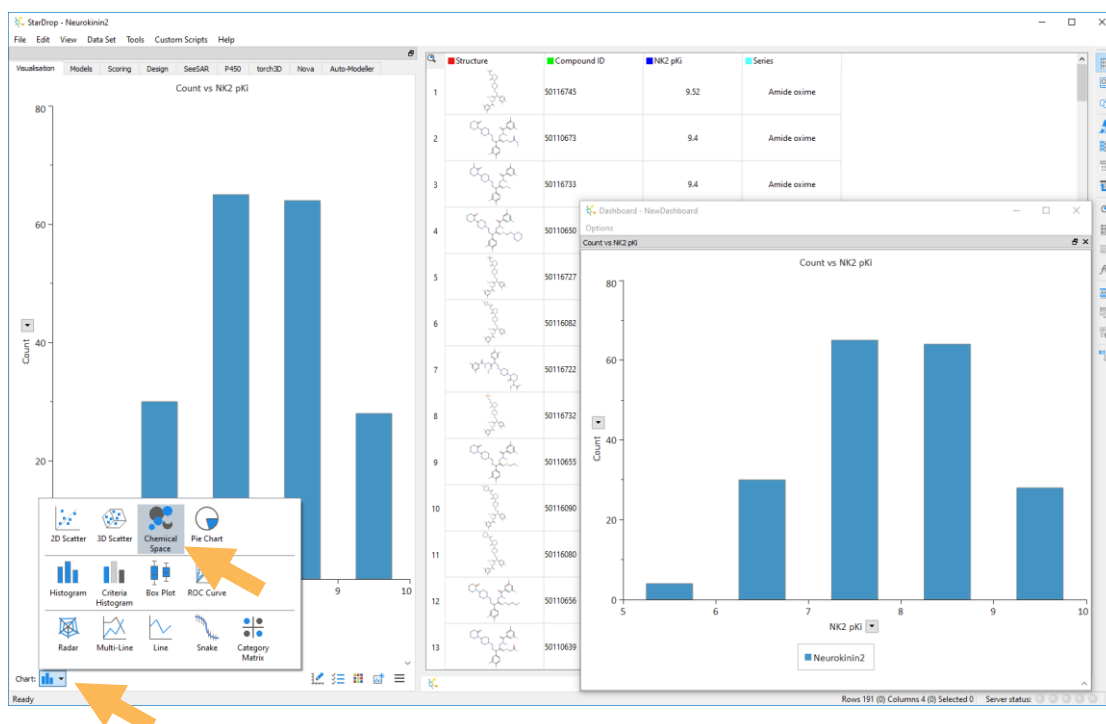


We can explore the distribution of potency across the chemical diversity of this library by creating a dashboard containing this histogram and a chemical space visualisation.

- 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.

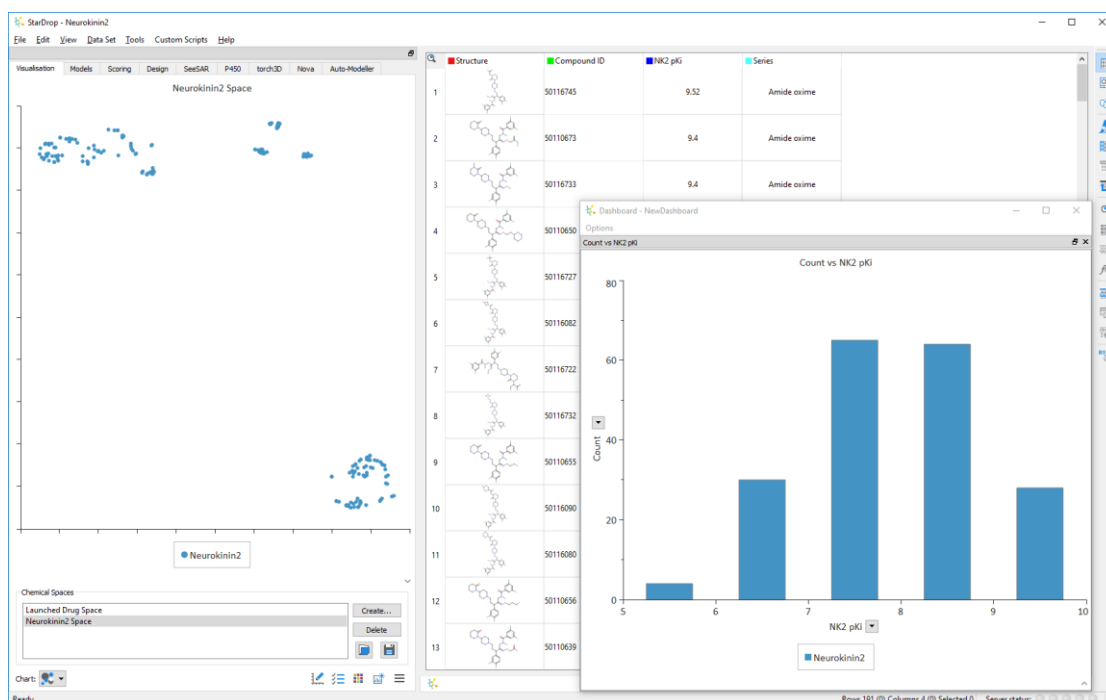


- Click on the **Create button** at the bottom of the **Visualisation area** to open the **Create Chemical Space dialogue** where we can define a new chemical space visualisation.
- Give the chemical space a name by typing in the **Name field**. In the example (right) we have called it **Neurokinin2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

The 'Create Chemical Space' dialog box is shown. It has a 'Data Set' dropdown set to 'Neurokinin2'. The 'Name' field contains 'Neurokinin2 Space'. Under 'Method', 'Visual Clustering' is selected. Under 'Dimensions', '2D' is selected. Under 'Similarity Model', 'Chemical Structure' is selected. There is a search bar with 'Enter search text' and a large text area below it. At the bottom, there are 'Select All', 'Clear', 'Progress' (with a progress bar), 'OK', and 'Cancel' buttons.

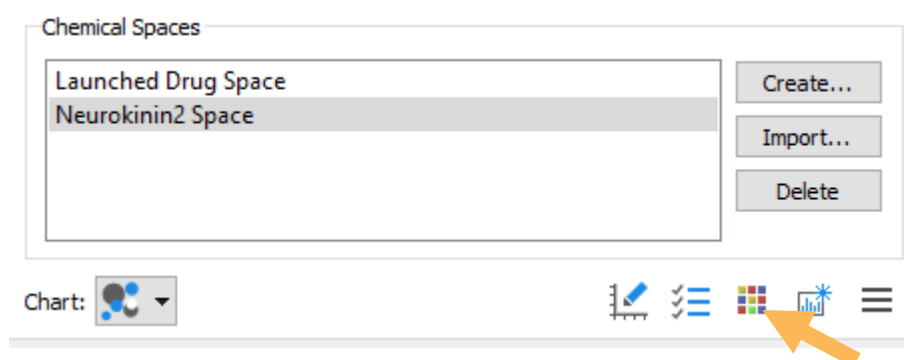
It is possible to build different types of chemical space plots based on chemical structure or property data, but in this case, we're going to use the defaults: **Chemical Structure**, the **Visual Clustering** method and **2D**, as shown (above).

- 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 will show the compound structure. Selecting a row in the data set will select the corresponding point in chemical space and vice versa. To see the distribution of  $pK_i$  values across the chemical space, we need to change the formatting.

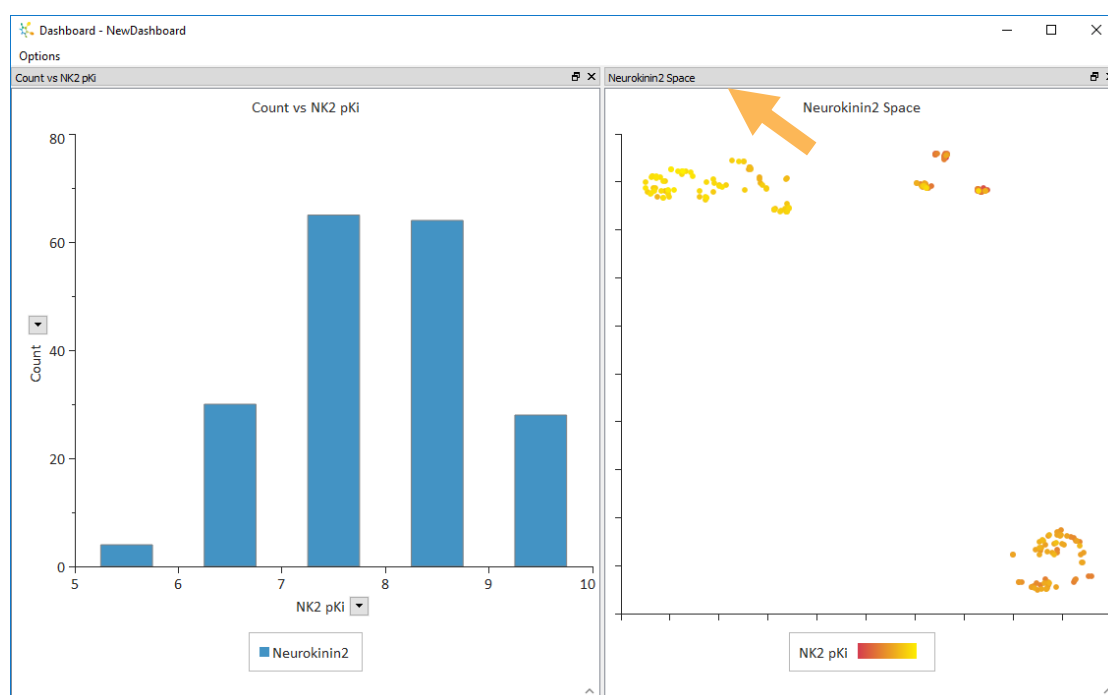
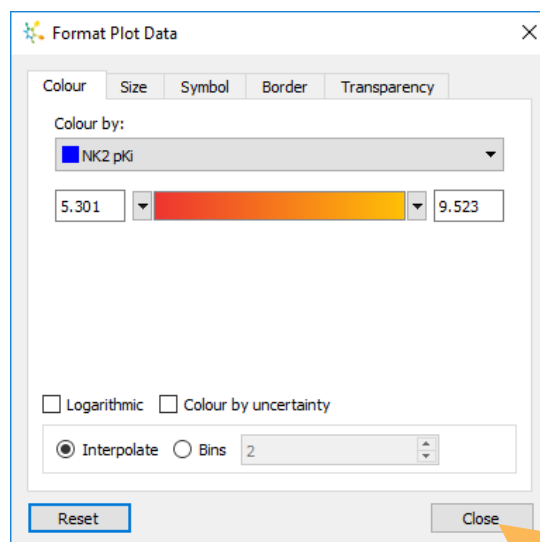
- Click on the **Format button**  at the bottom of the **Visualisation area** to bring up the **Format Plot Data** dialogue.



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

The compounds with the highest  $pK_i$  values will be highlighted yellow, the lowest red but you can change these by clicking on the colour bar. Note that the plot will update to show your changes as you modify the formatting options.

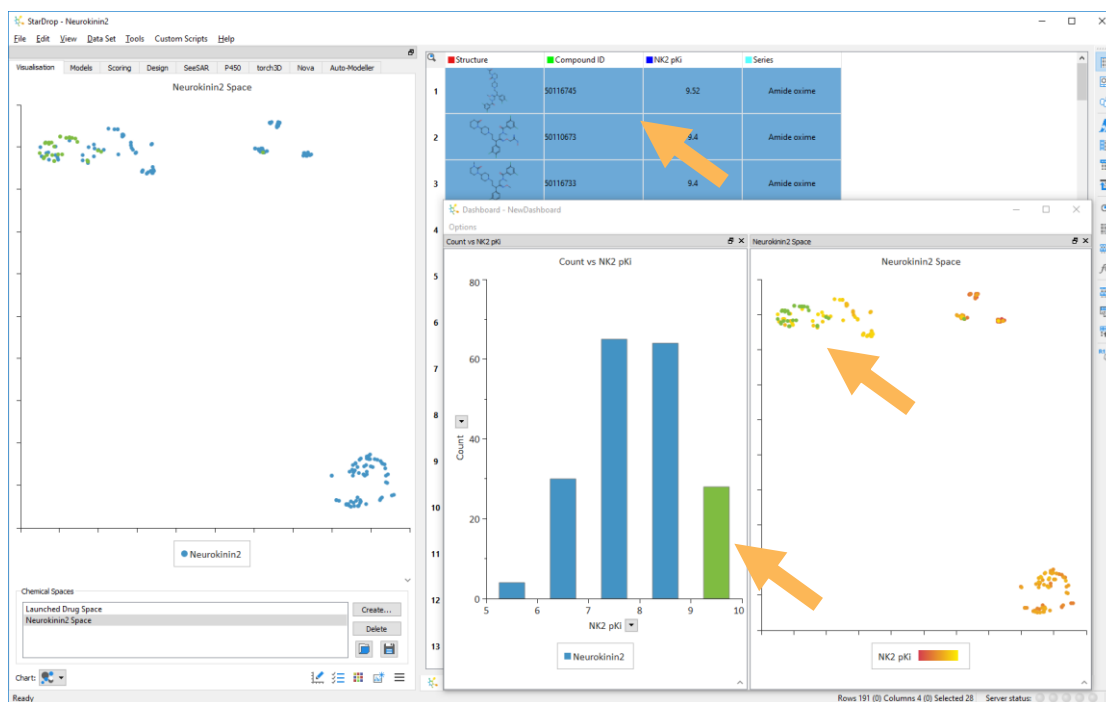
- Click the **Close button** to close the formatting dialogue.
- Add this into the dashboard by clicking the **Detach button**  again.



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 will display all the controls you need to change or format your chart.

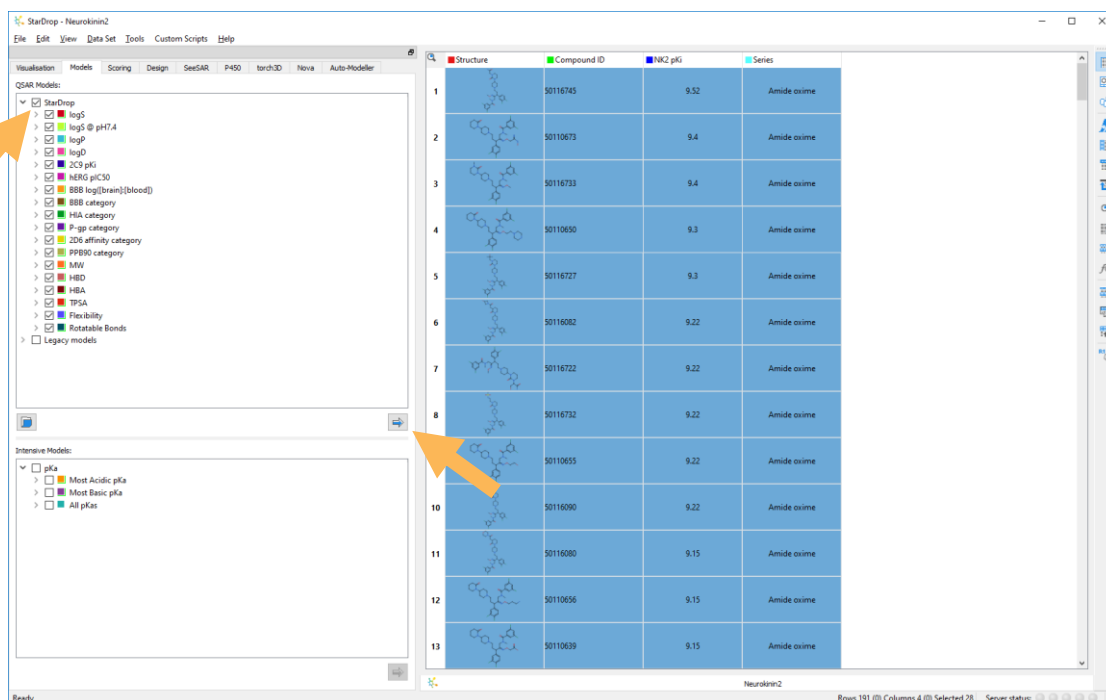
You will notice that a large proportion of the compounds with the highest  $pK_i$  values (yellow) come from one of the clusters in the chemical space.


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

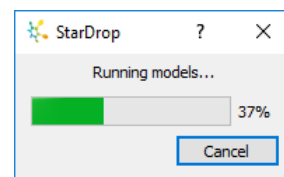


This will highlight the most potent compounds in the chemical space, as well as the data set. This may be the best region in which to focus for selecting compounds, but first we should consider the other properties that are important in a high-quality lead (you can minimise the dashboard window to keep it out of the way for now). We have no further measured data but can generate predictions of ADME and physicochemical properties.

- Click on the **Models** tab.
- To select all the StarDrop models, check the box next to the word StarDrop.

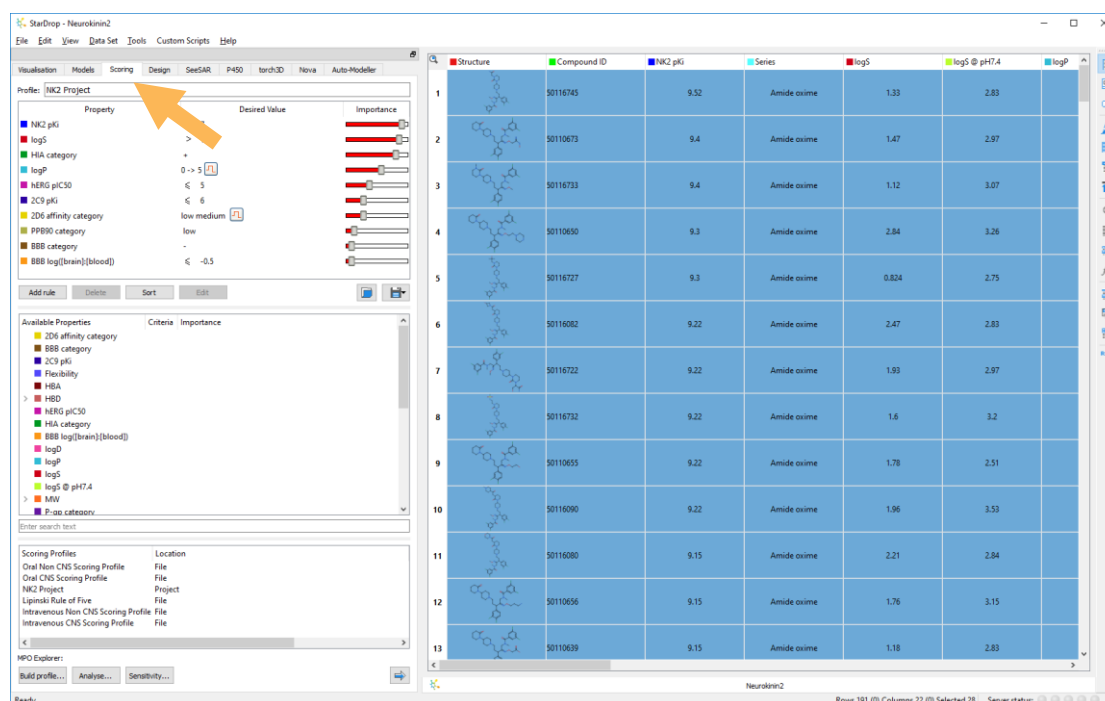


- Click the  button at the bottom of the **QSAR Models** area. A progress bar will be displayed while the predictions are calculated.



When this process is complete you will 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 and so 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.



	Structure	Compound ID	NK2 pKi	Series	logS	logS @ pH7.4	logP
1		50116745	9.52	Amide oxime	1.33	2.83	
2		50116673	9.4	Amide oxime	1.47	2.97	
3		50116733	9.4	Amide oxime	1.12	3.07	
4		50116650	9.3	Amide oxime	2.84	3.26	
5		50116727	9.3	Amide oxime	0.824	2.75	
6		50116682	9.22	Amide oxime	2.47	2.83	
7		50116722	9.22	Amide oxime	1.93	2.97	
8		50116732	9.22	Amide oxime	1.6	3.2	
9		50116655	9.22	Amide oxime	1.78	2.51	
10		50116690	9.22	Amide oxime	1.96	3.53	
11		50116680	9.15	Amide oxime	2.21	2.84	
12		50116656	9.15	Amide oxime	1.76	3.15	
13		50116639	9.15	Amide oxime	1.38	2.83	

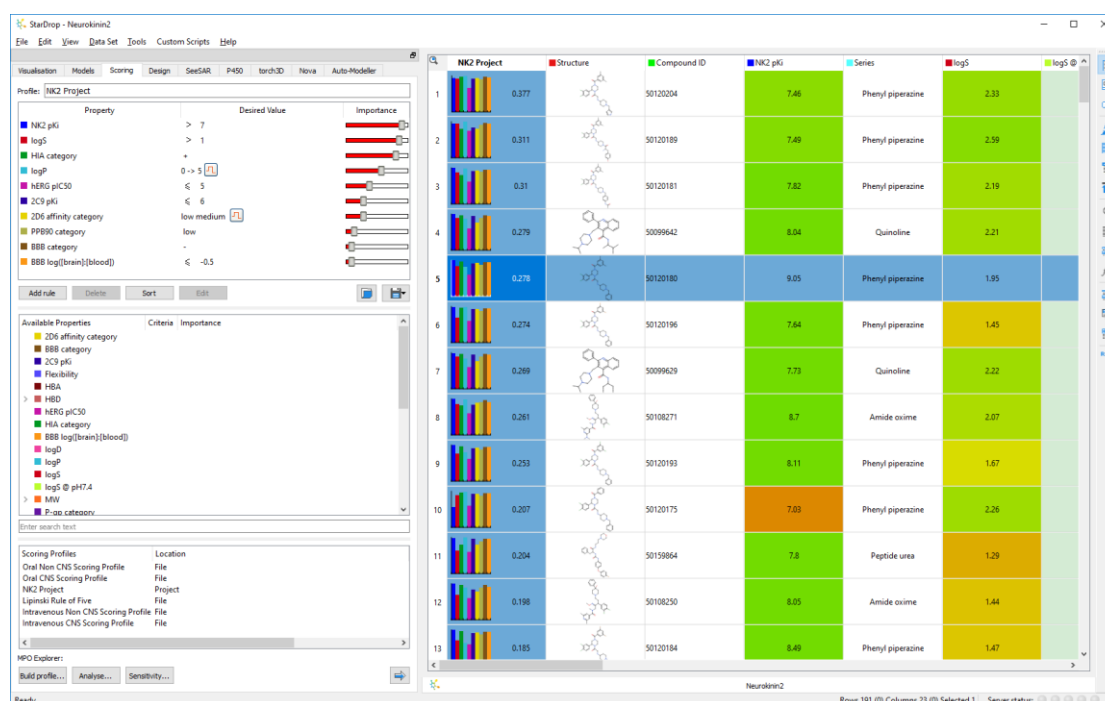
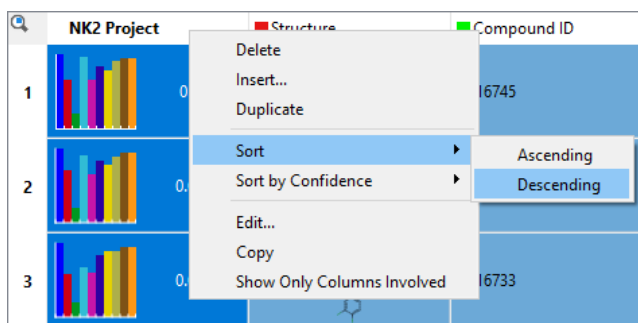
A scoring profile enables you to define a set of criteria that are important for your project. StarDrop provides some example profiles and, as part of the project file, we have already loaded a profile designed for this Neurokinin 2 project.

A scoring profile contains a series of properties along with criteria describing desired values and their relative importance. In this profile we are looking for compounds with good affinity for Neurokinin 2 and which are suitable for a peripheral target.

- Run 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, taking into account each property criteria, 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.

- 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 score.




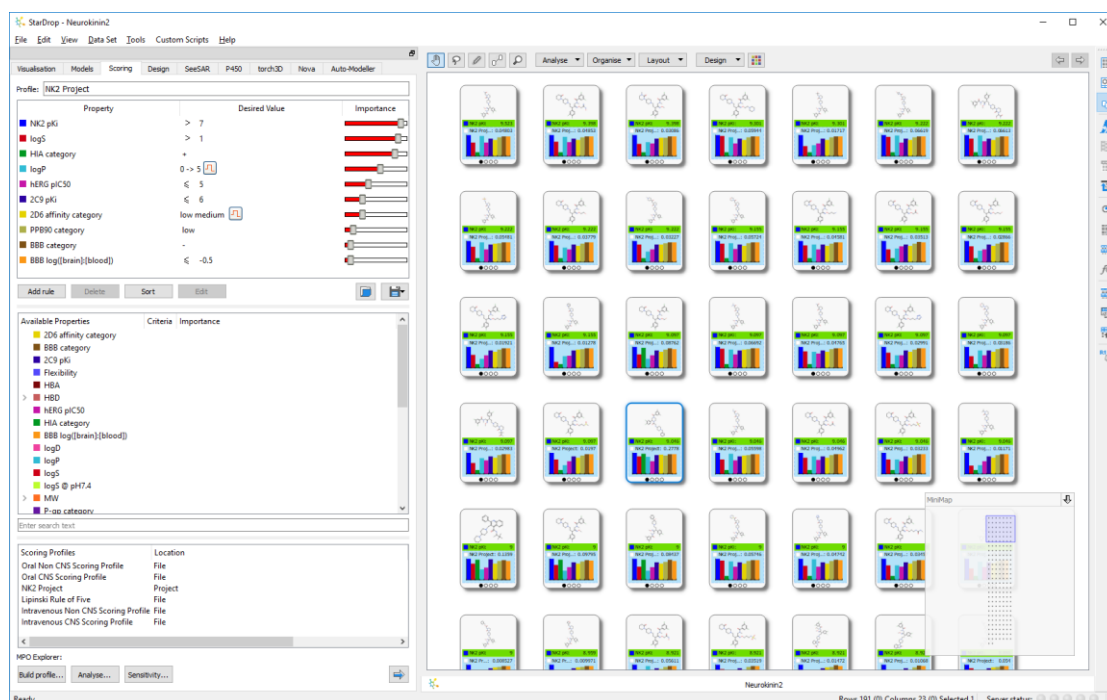
The histogram in each cell gives a quick overview of the impact of each property on the score for a compound. As you scroll down the data set, and the scores become lower, the low bars in the histograms indicate properties that have not met the requirements defined in the project profile, taking into account the confidence in the data and importance of each property.

Having scored the compounds, we're now going to compare some compounds using StarDrop's Card View. This provides a convenient way to work with our compounds and data

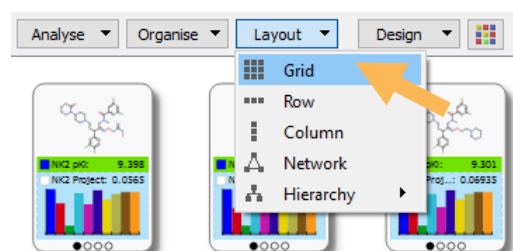



in the context of this project, by representing compounds on cards that can be moved, stacked and linked however we wish.

- Switch to **Card View** by clicking the **Card View**  button in the toolbar to the right. This will display 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 in which we are interested.

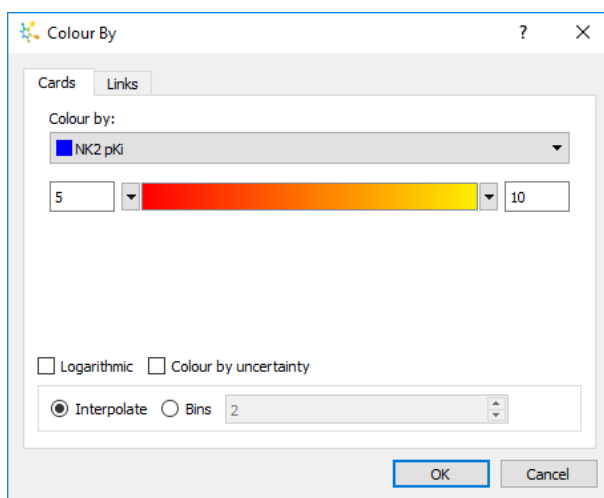


- Sort the grid in order of the score from top-left to bottom-right by choosing the **Grid** option from the **Layout** menu at the top of Card View.

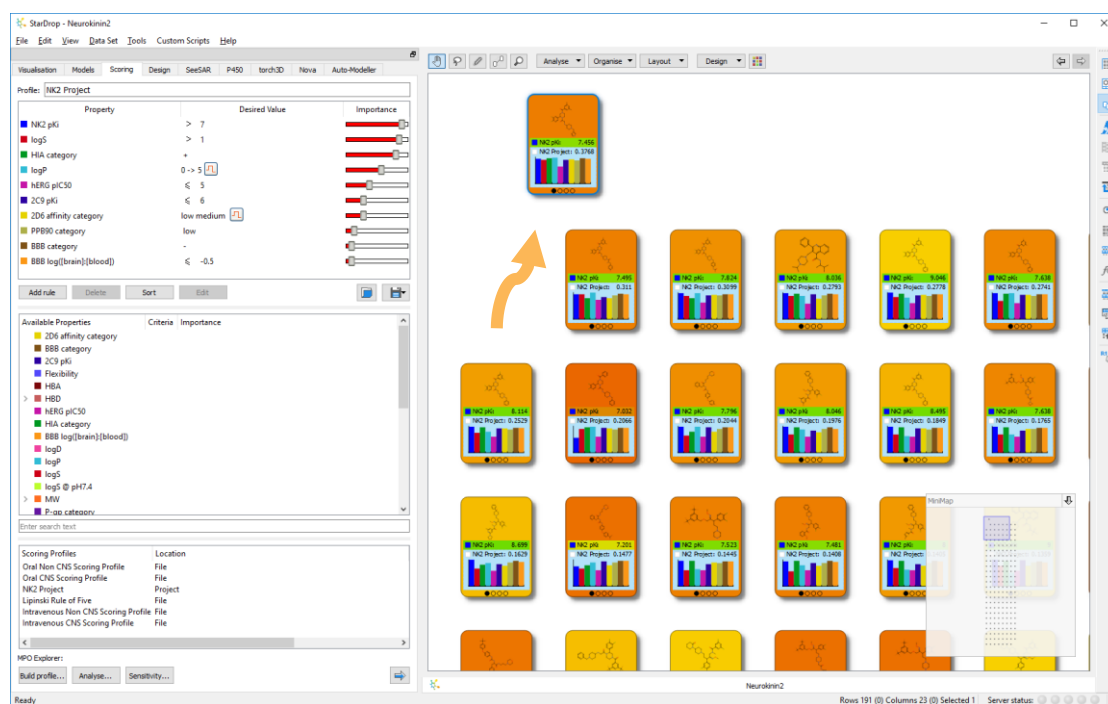


- We can now colour the cards to highlight the most potent, by clicking on the **Format** button  and in the **Colour By** dialogue choose to colour the cards by **NK2 pKi**.

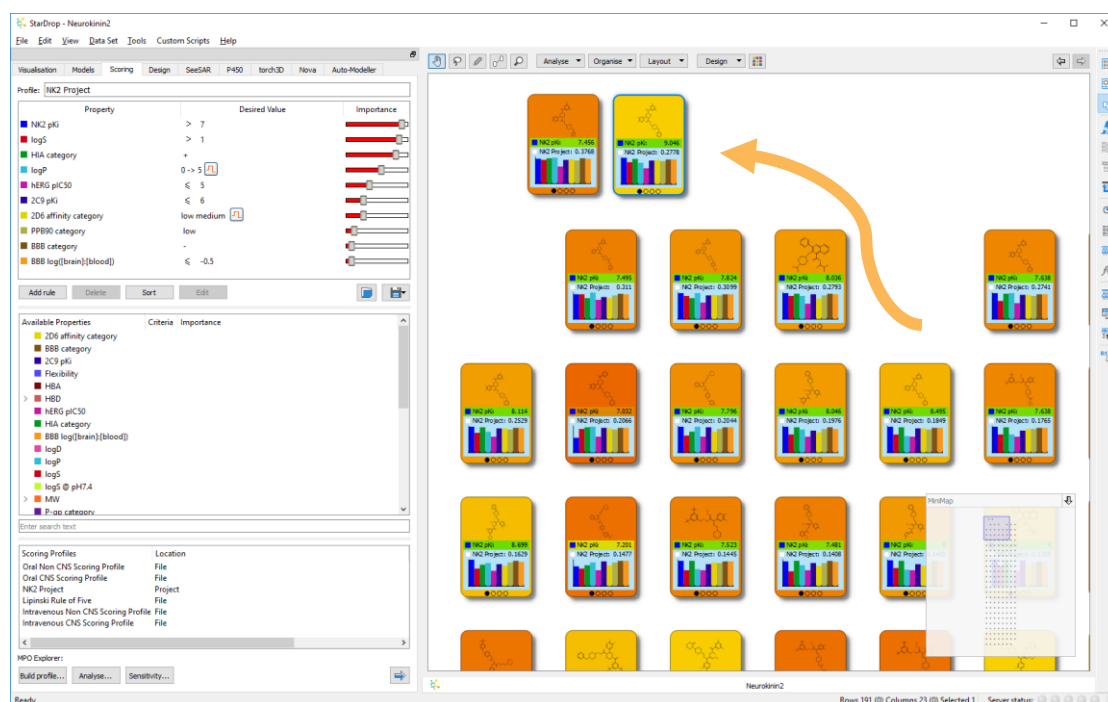
We can easily compare two compounds by placing their corresponding cards side-by-side.



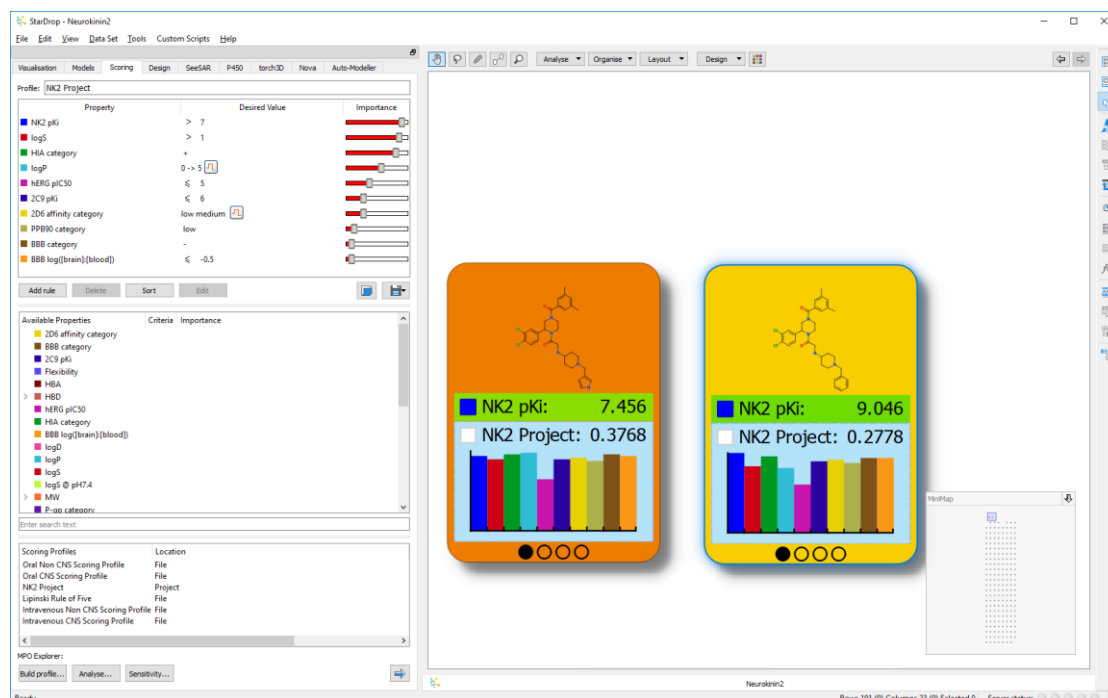
- Click on the highest scoring compound and drag it to the top. You can make space by dragging the background if necessary.



- Now select the fifth compound, which is one of the most potent, and drag it to lie beside the first.



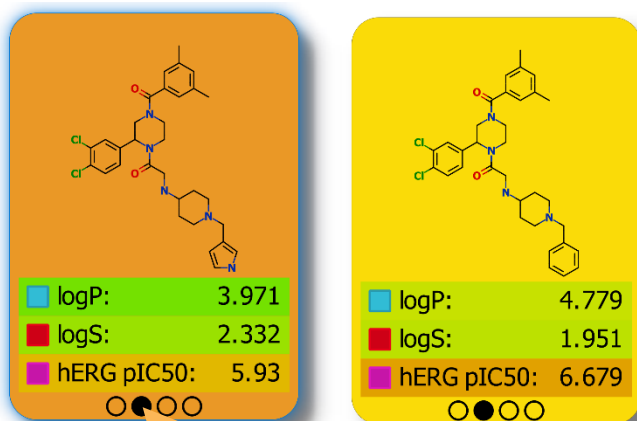
- You can now zoom in to these cards to look at them in more detail by pointing at the cards and using the mouse wheel (or if you don't have a mouse, by holding down the **CTRL** key and pressing the = key). Dragging the background will also help you to position the cards conveniently.



Here we can see that the compound on left has a higher score, even though the compound on the right is one of the most potent. We can investigate this further:

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

Here we can clearly see that the compound on the right has higher values for logP and hERG pIC<sub>50</sub>, which increase the risk for this compound and hence result in a lower score.

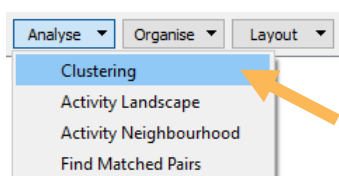


- Click on the fourth circle now to go to the notes page and then double-click on **Add notes** to add your own text which will be saved in a **Notes** column in the data set.

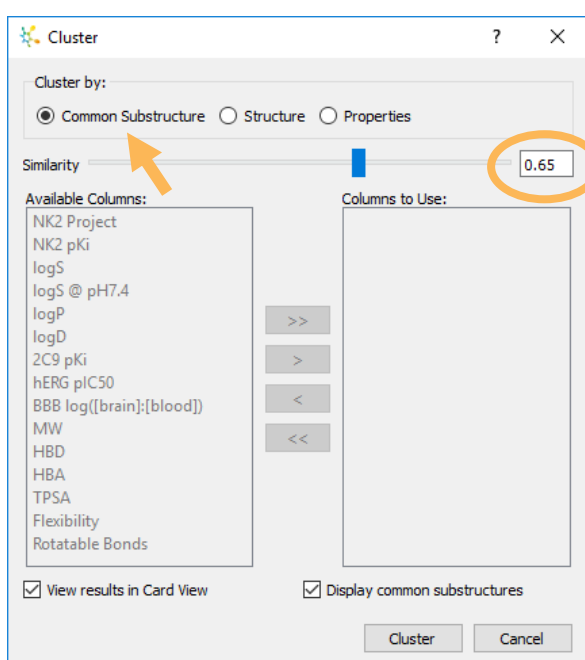


Having compared individual compounds, we would also like to explore how the properties vary across the different chemical series in this library. Clustering provides a convenient way to group compounds by chemical similarity.

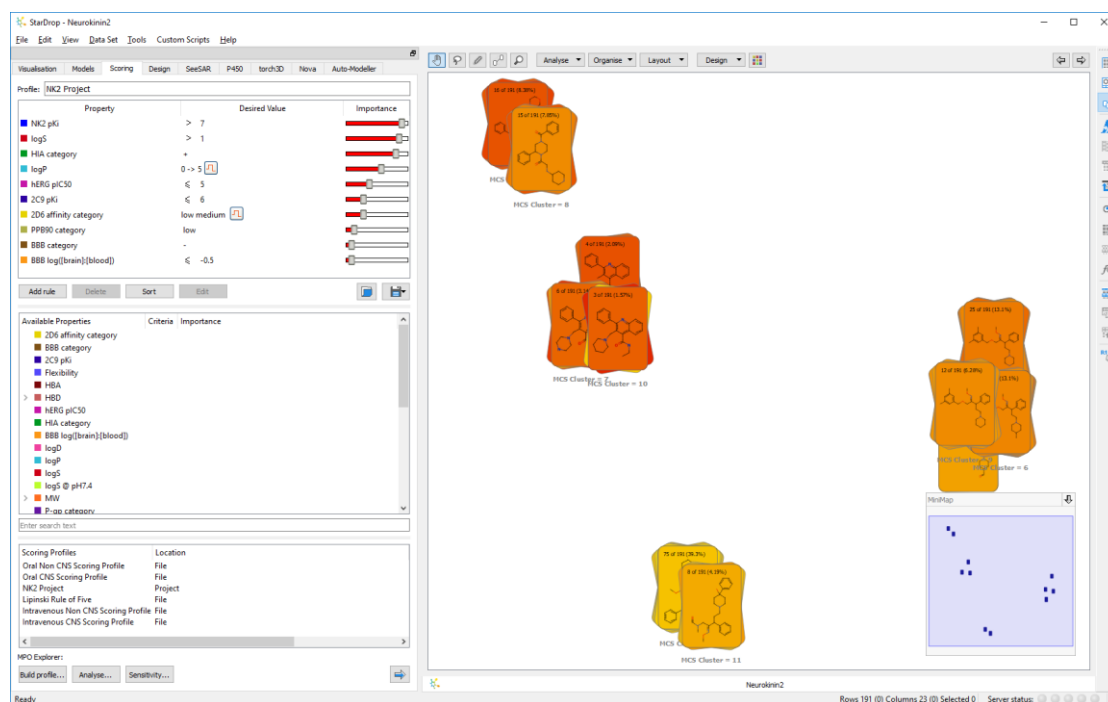
- From the **Analyse** menu at the top of the Card View, select **Clustering**.



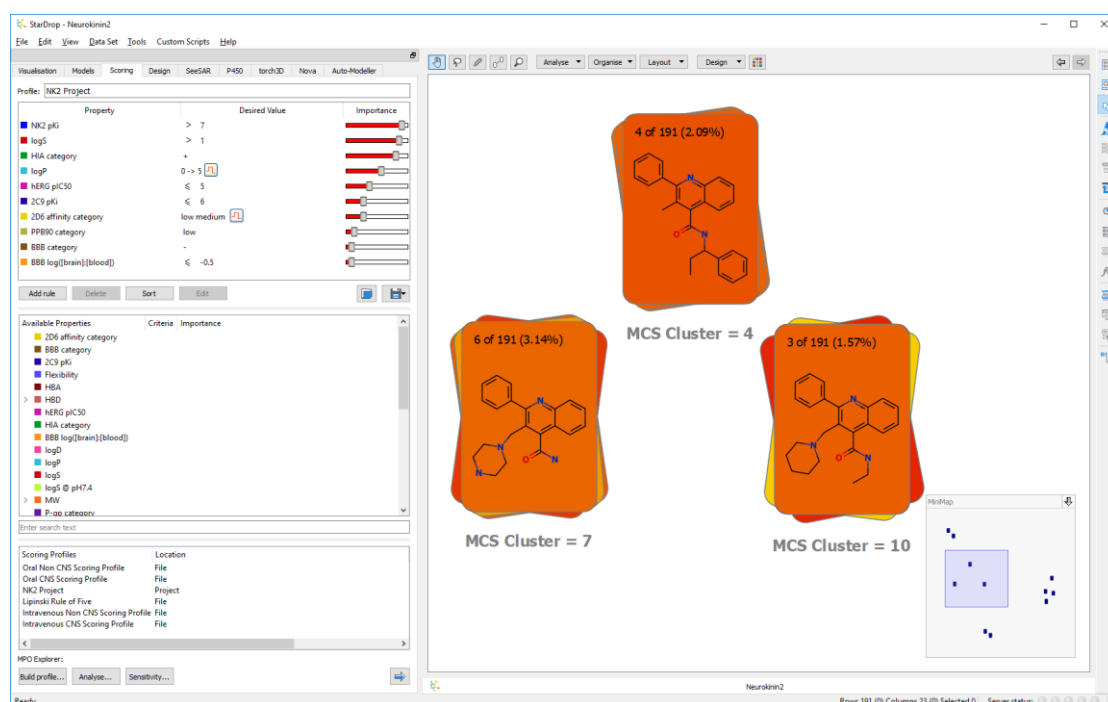
- In the **Cluster** dialogue, we will use the default **Common Substructure** method. Set the **Similarity** to **0.65**, as shown right, and click the **Cluster** button.



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

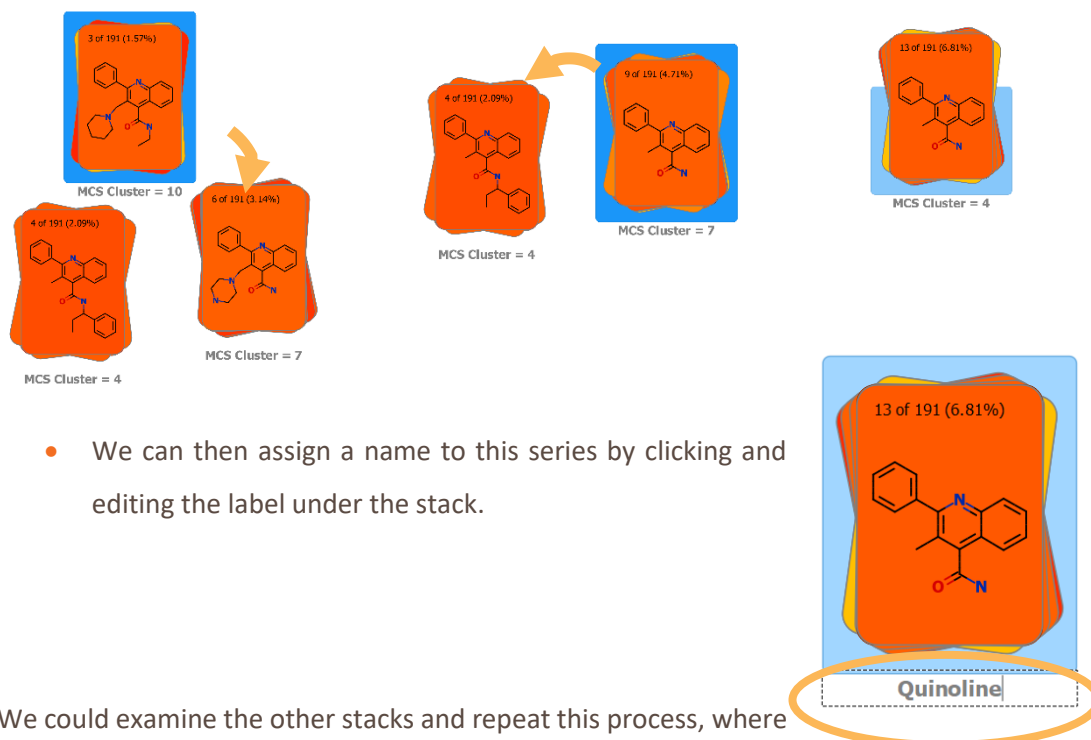


- Zoom in on the stacks corresponding to clusters 4, 7 and 10, which will be close together and drag the background to centre these in the display. You can move the stacks to separate them slightly by clicking and dragging.



Here you can see that, although the compounds are in different stacks, 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.

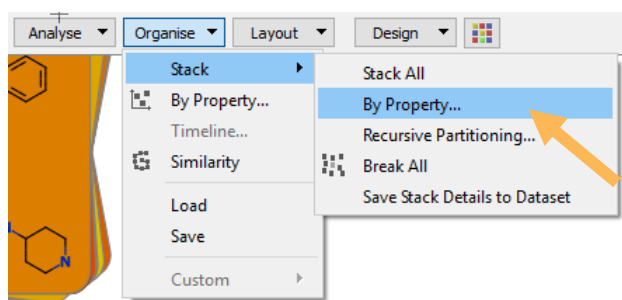
- We can easily combine these by dragging one stack on top of another.



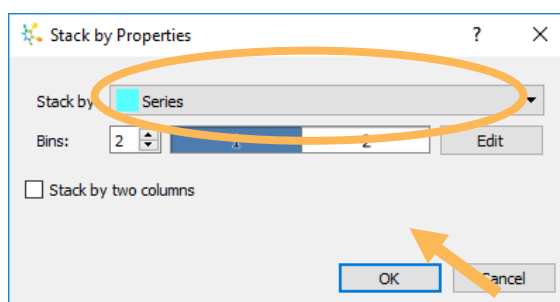
- We can then assign a name to this series by clicking and editing the label under the stack.

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 of the compounds.

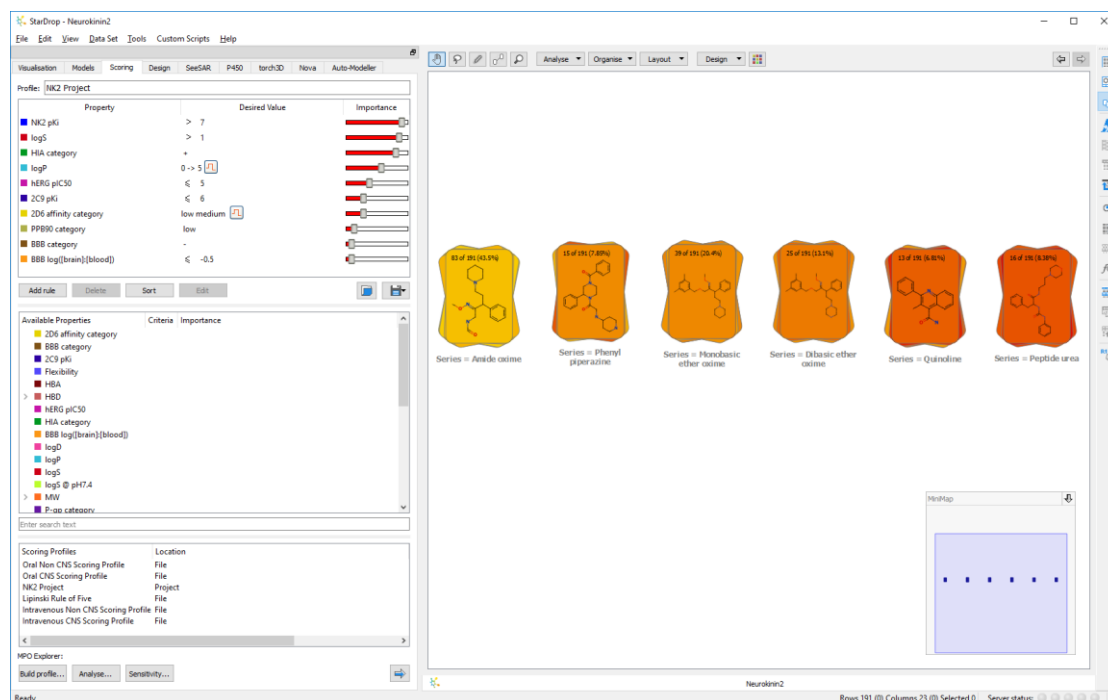
- From the **Organise** menu at the top of Card View, select **By property** from the **Stack** menu.



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

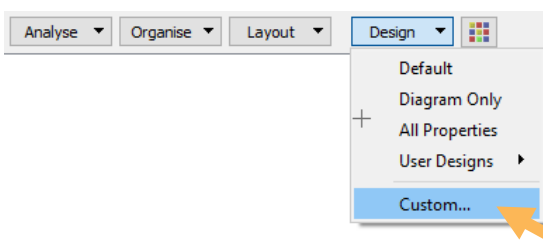


Here we can see that there are 6 series, of which the Amide oxime series has the highest average potency (it is the most yellow).



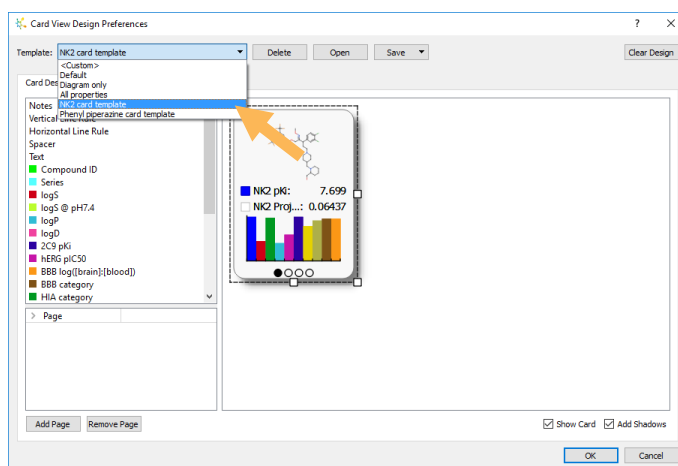
We can display more information on a stack. To illustrate this, we can choose a new stack design.

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



In the **Card View Design Preferences** dialogue, you can change card sizes, choose which properties should be displayed and how the properties should be arranged, but in this case, we are going to choose a pre-defined template saved as part of this project.

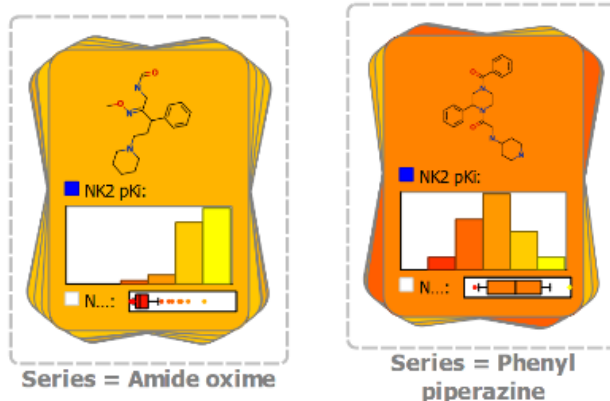
- Select **NK2 card template** and click the **OK** button.



**Note:** You can also select previously saved designs directly


from Card View's **Design menu** under **User Designs**.

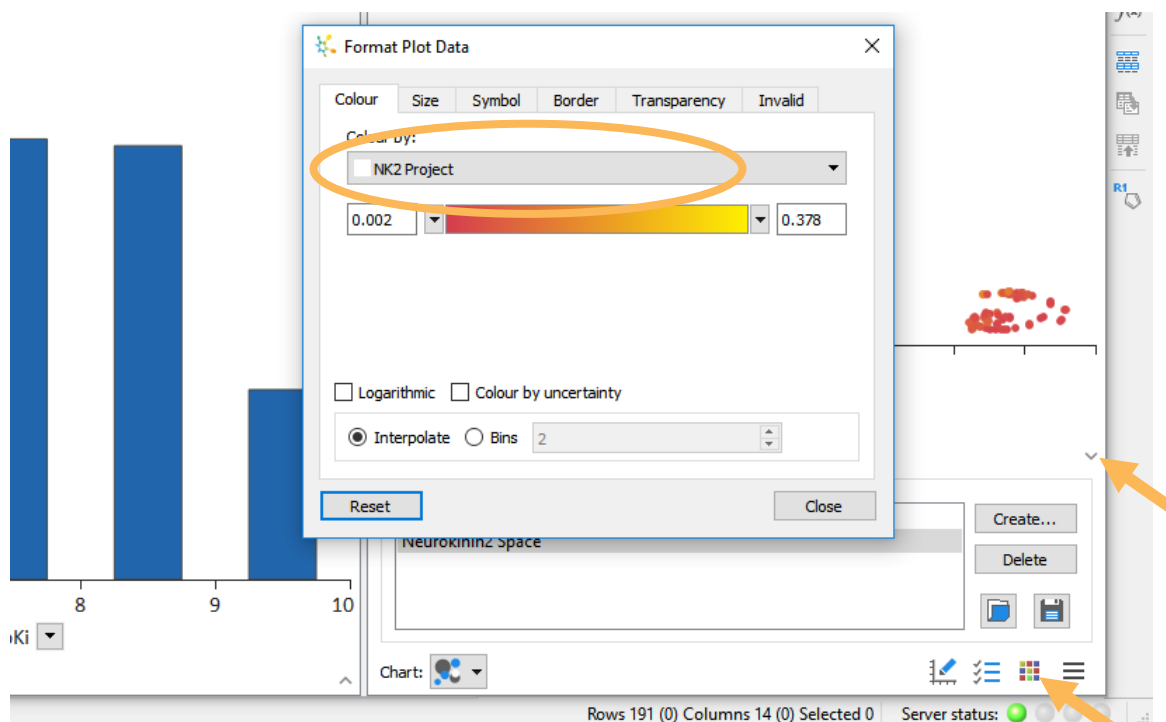
The resulting stacks show a histogram of the distribution of  $pK_i$  values of the compounds within each stack and, below this, a box plot showing the distribution of scores. This makes it straightforward to compare different series, for example the Amide oxime series has an excellent distribution of 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.

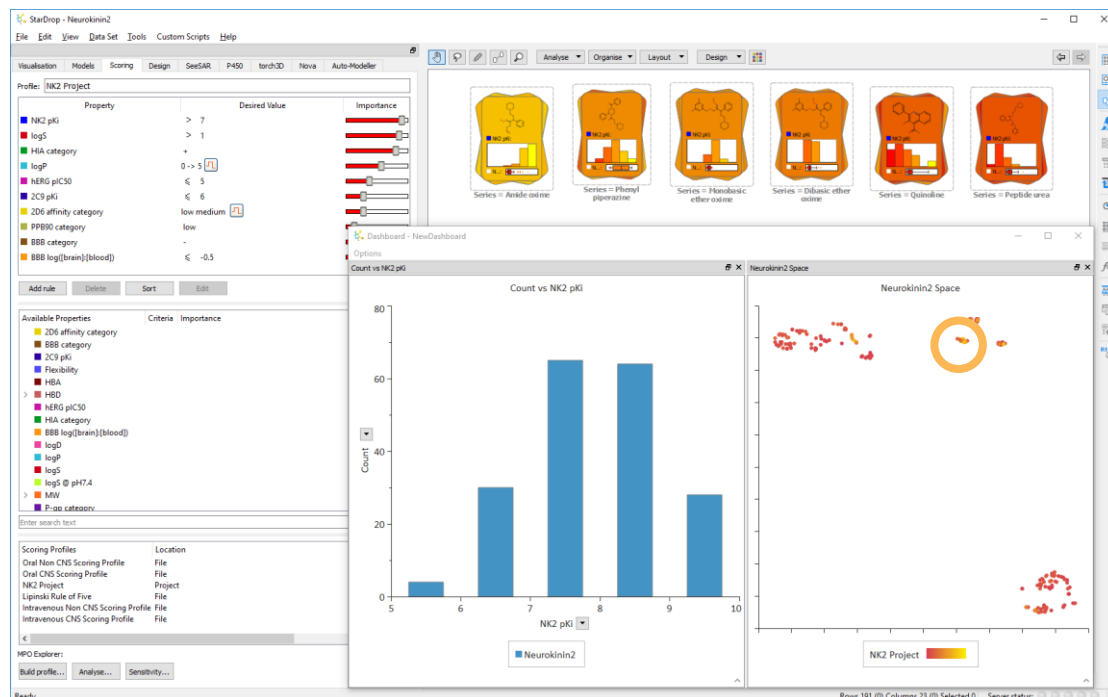
We can explore the distribution of scores across the whole data set by returning to the dashboard we created earlier.

- Click the arrow in the bottom corner to display the controls, click on the **Format button**  at the bottom to bring up the **Format Plot Data** dialogue again.



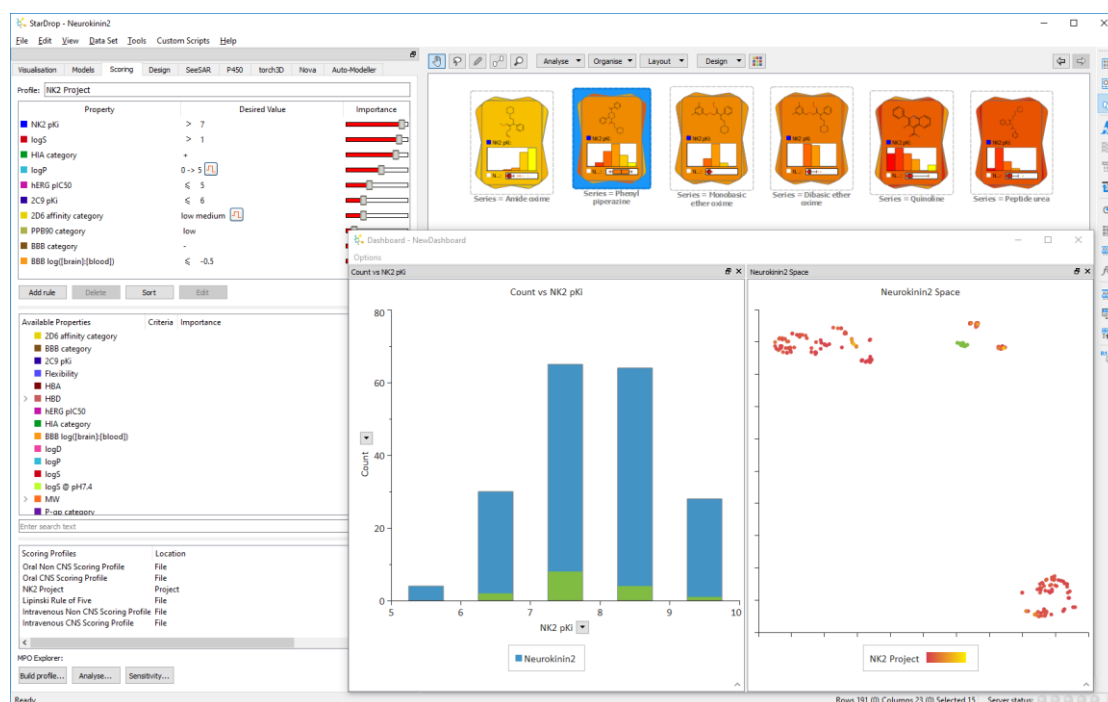


- Change the **Source** property to be the **NK2 Project** score values, before clicking the **Close** button.



The points are now coloured based upon the overall scores of the compounds, not just the pK<sub>i</sub> values. We can see that the pattern has changed dramatically from that which we could see when looking only at potency. There is just one small cluster of yellow points (highlighted by the blue circle) as a 'hot spot' of high scoring compounds.

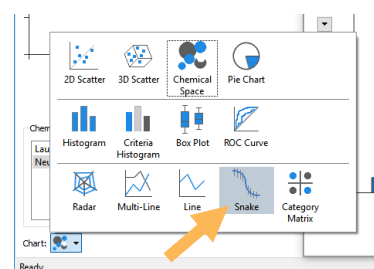
- Click with the left mouse button and draw around this yellow cluster in the chemical space. You will see that the stack corresponding to the Phenyl piperazine series is highlighted, confirming that the hot spot corresponds to this series.

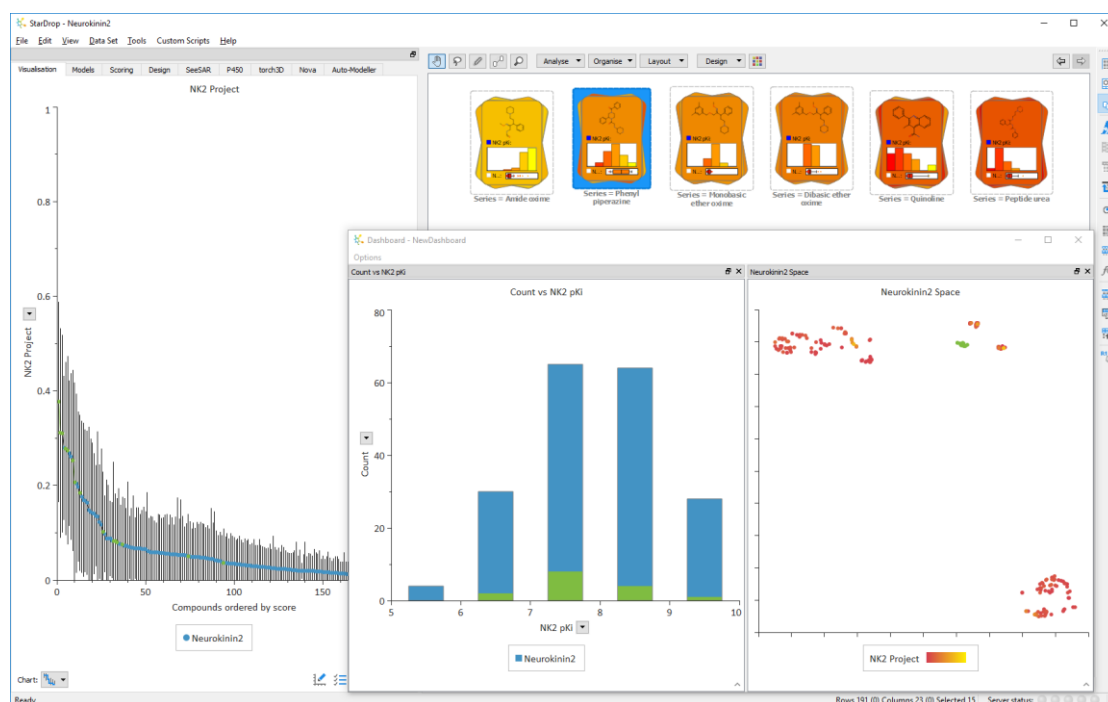


Although this appears to be the best series, we should consider the impact of the uncertainty in the underlying data on our ability to confidently distinguish between compounds. Therefore, we will use another type of chart to explore this.

- Go to the **Visualisation** area again.
- Now open the **Chart** menu at the bottom and choose a **Snake** plot (as shown right).

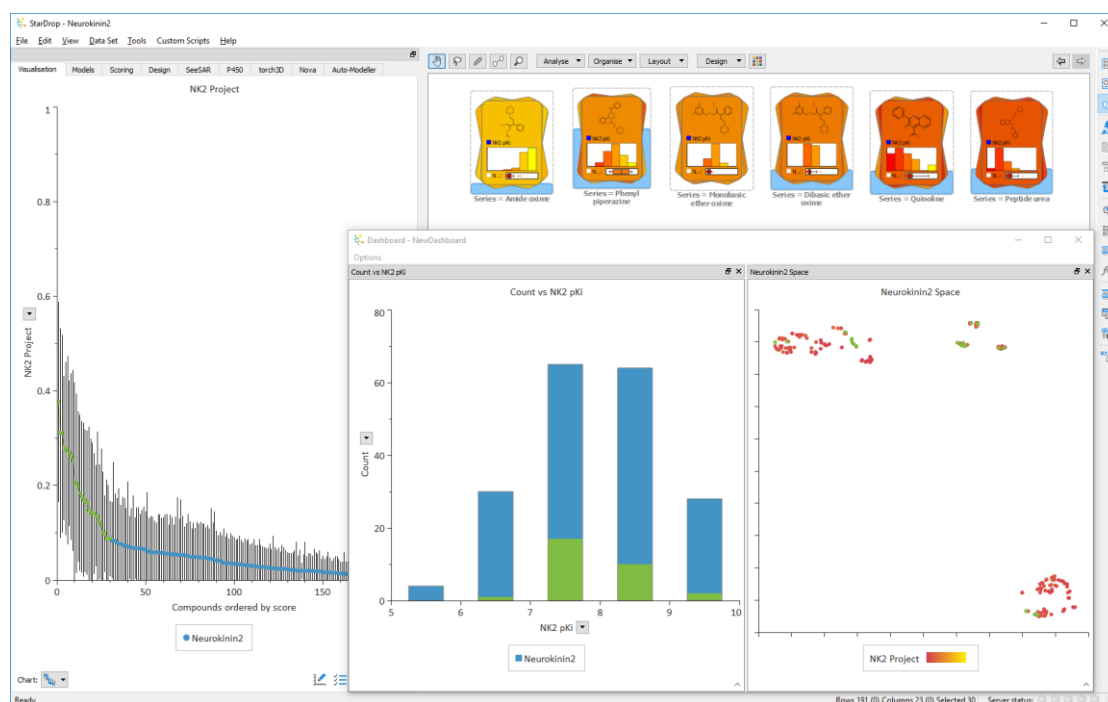
A 'snake plot' shows the scores (on the y-axis) for all compounds in order from highest score to lowest score (along the x-axis). The overall uncertainty in each score, due to the uncertainty in the underlying data, is also displayed as an error bar around each point.





The top 30 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 30 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.

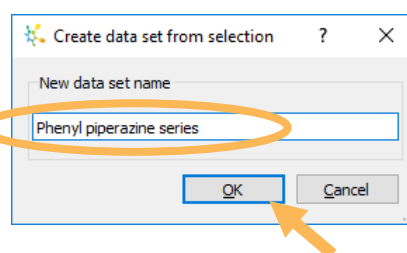
- Select the top 30 compounds in the snake plot by drawing a ring round them with the mouse.



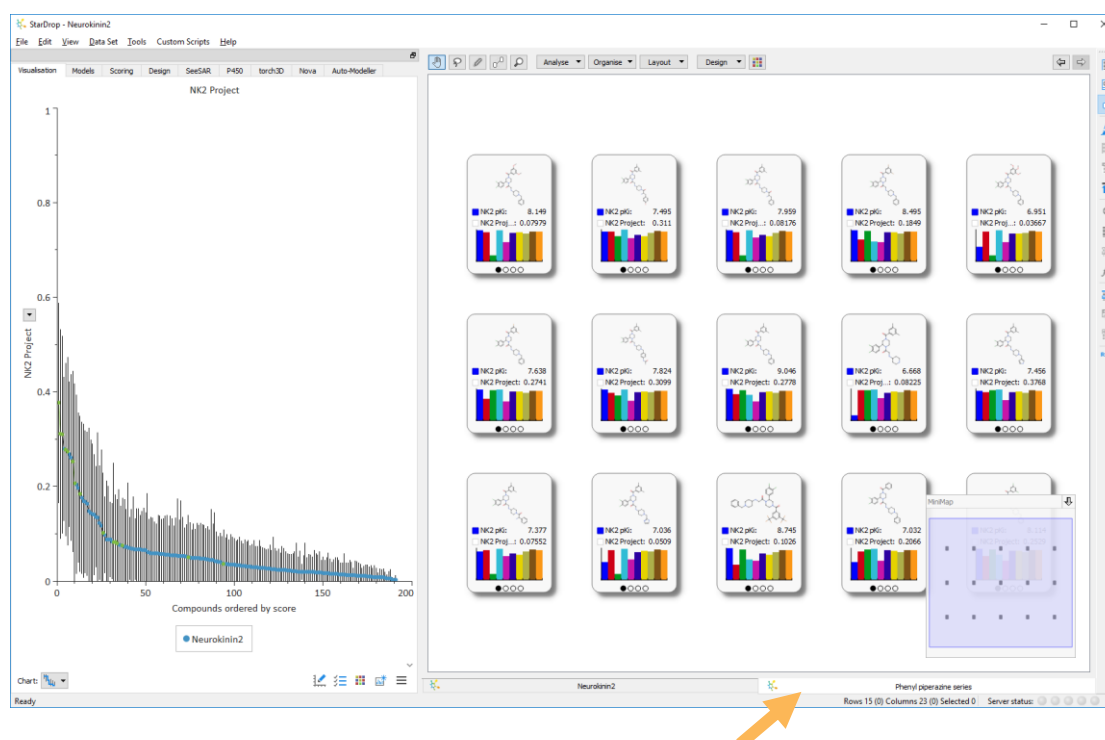
Notice that this selection includes compounds from different regions of the chemical space and a small number of compounds from every stack except the Monobasic ether oxime series. This suggests that some of these chemistries cannot be rejected with confidence. Therefore, it may be more appropriate to sample a small number of compounds from some of 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.


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.

- Hide 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** dialogue, give the new data set a name, e.g. "Phenyl piperazine series" and click the **OK** button.

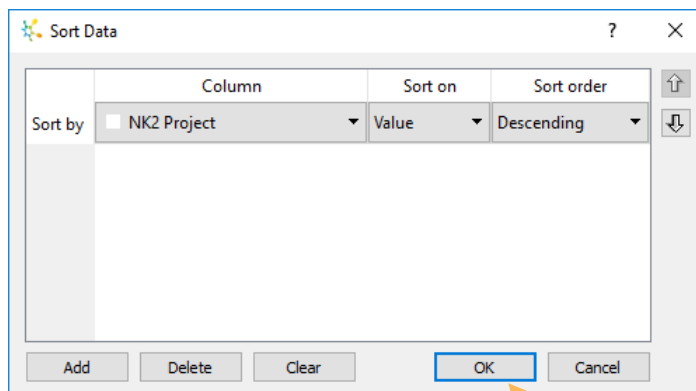


This will copy the 15 compounds from this series into a new data set which appears in a new window (you can change back to the full data set by clicking on the tab at the bottom)



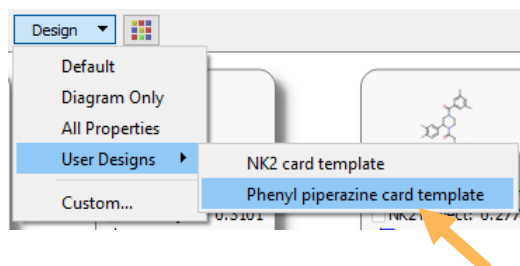
- We can sort the cards in order of descending score using the sort dialogue by clicking on the **Sort button**  in the tool-bar on the right.

- The score column **NK2 Project** is the first one in the data set and will already be chosen so simply click the **OK button**.



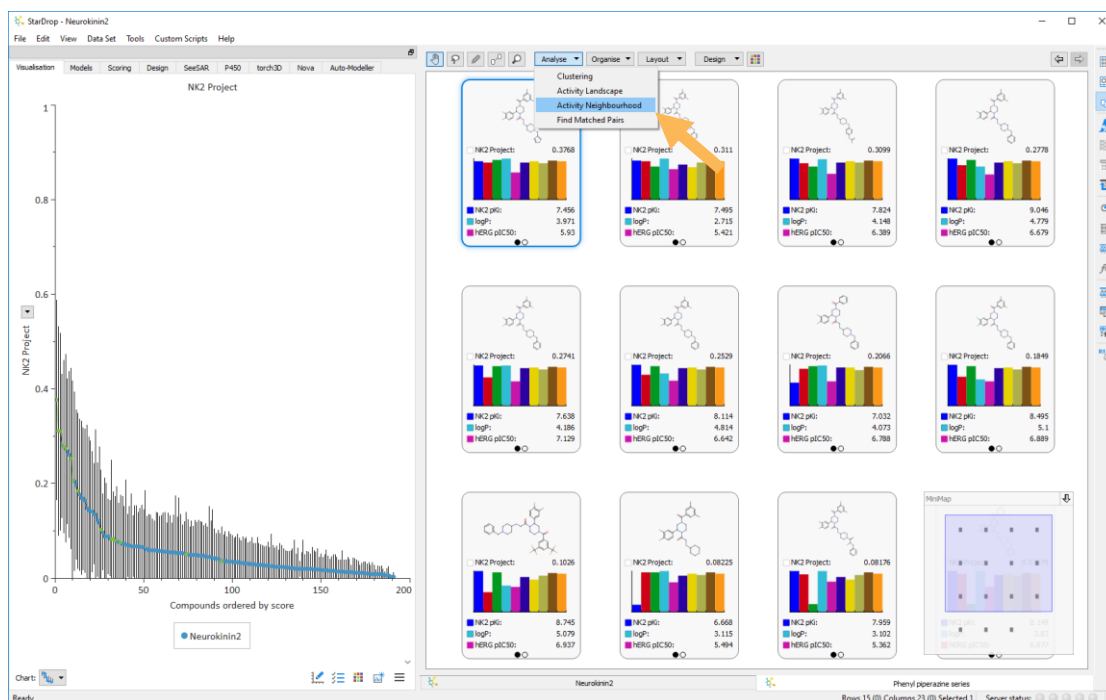
We'd like to look at these compounds in more detail, so we'll load a new card template that displays more data on each page. When loading a saved card template that is already part of a project it is not necessary to open the **Card View Design Preferences** dialogue because the template will be available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

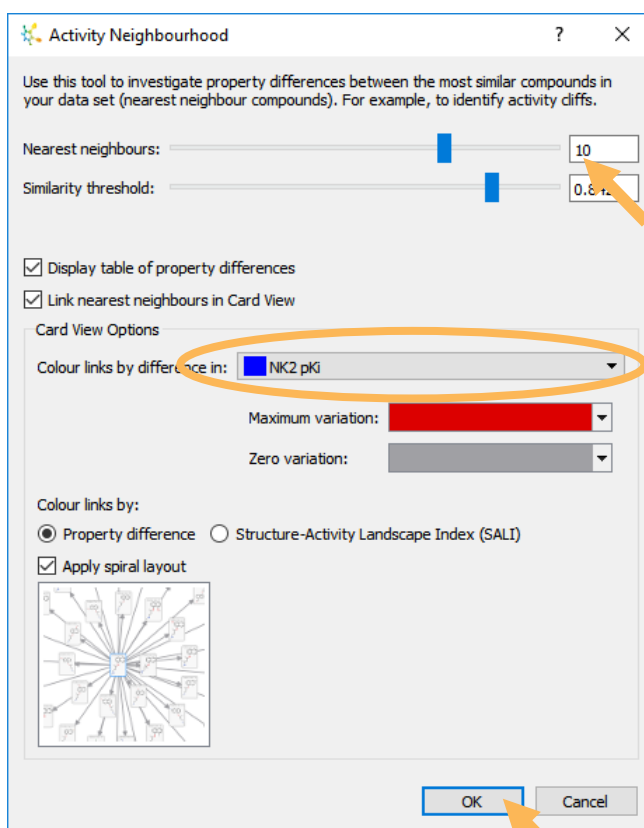
- Select the card in the top-left of the grid (this is the highest scoring compound in this series). Now select **Activity Neighbourhood** from the **Analyse** menu at the top of Card View.



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.

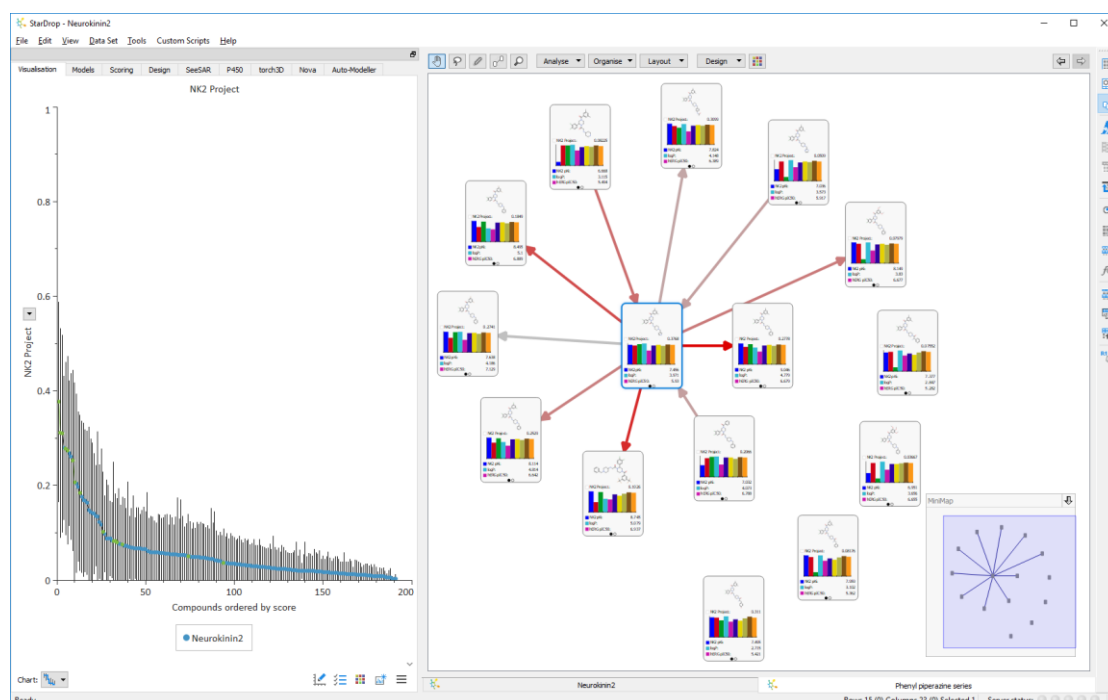
- In the **Activity Neighbourhood** dialogue, keep the default number of nearest neighbours (10) and choose to colour the links by the difference in **NK2 pKi**. Then click the **OK** button.

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



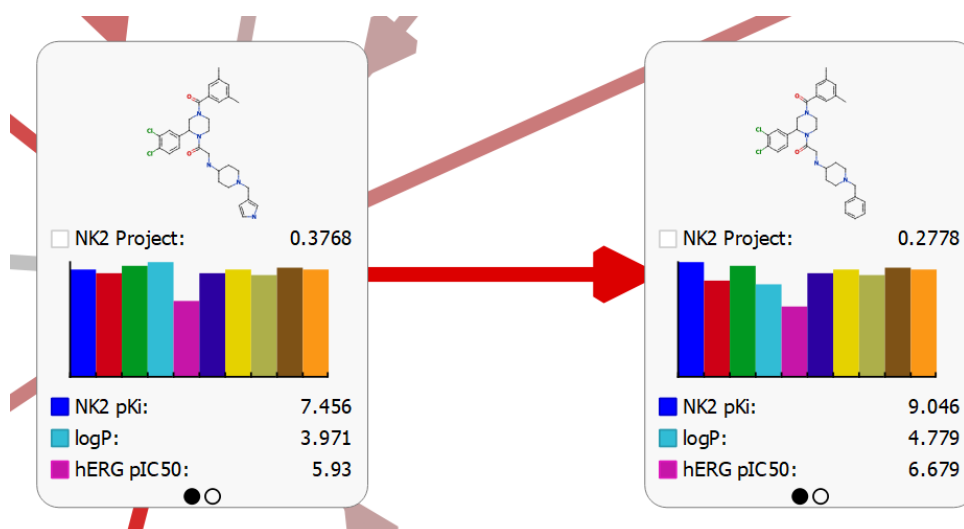
indicates the direction of *increase* of the selected property, in this case NK2 pKi. The colour indicates the magnitude of the change, in this case from red being the largest to grey being

the smallest. Therefore, 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 will also be produced, but we won't use it in this example, so you can close it if you wish.

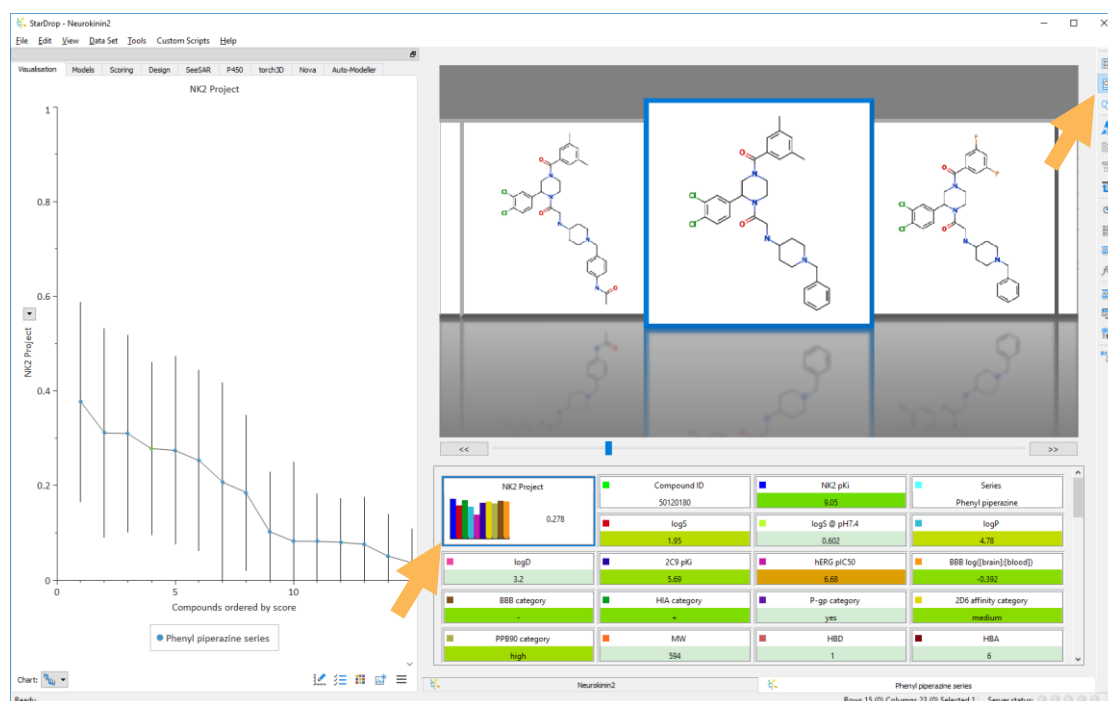
- Zoom in to the central compound to compare it with its closest neighbour (using the middle mouse wheel or **CTRL** and **=** keys).

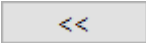
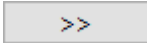


Here we can see that replacement of the pyrole with a phenyl has resulted in a 1.5 log-unit improvement in potency (from 25 nM to <1 nM). However, the overall score is reduced because of an increase in logP and hERG pIC<sub>50</sub>, as we saw earlier.

We might wish to explore alternative approaches to improving these properties while retaining potency against the target. To do this, we can use **Molecule View** and the interactive **Design tool with Glowing Molecule**.

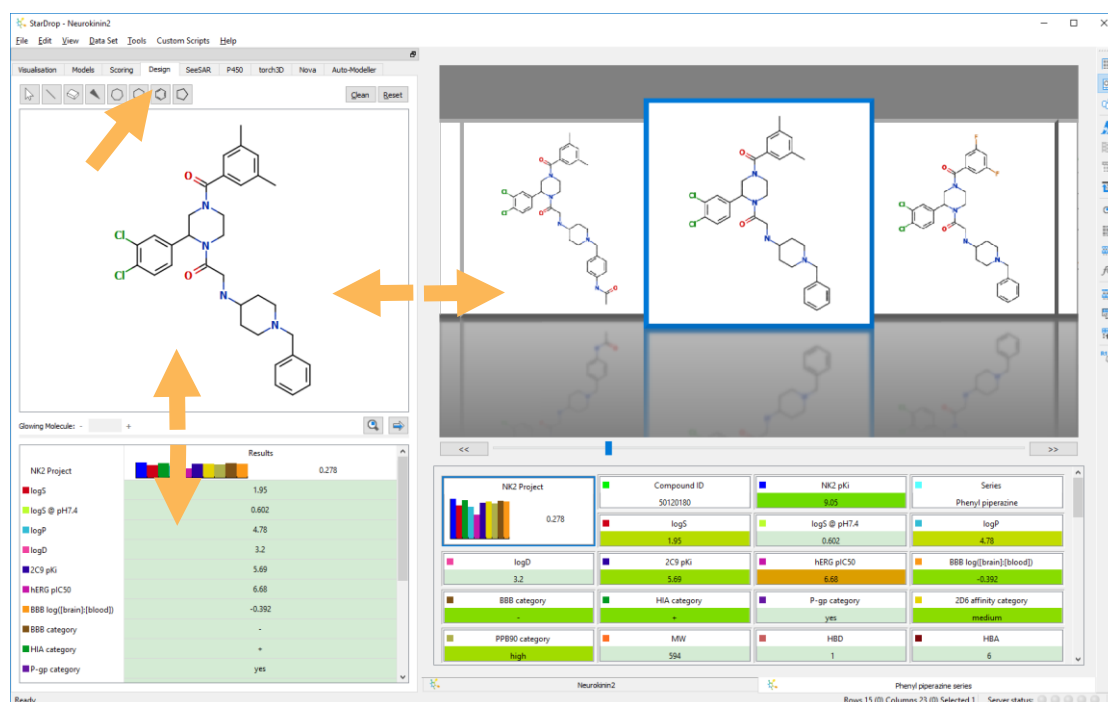
- Select the compound on the right and click the **Molecule View button**  on the right-hand tool-bar to switch into Molecule View.
- Click on the score cell to show a heat map indicating the contribution of each property to the overall score.



This shows an overview of all of the properties of the compound at the top. You can scroll through the compounds using the  and  buttons.

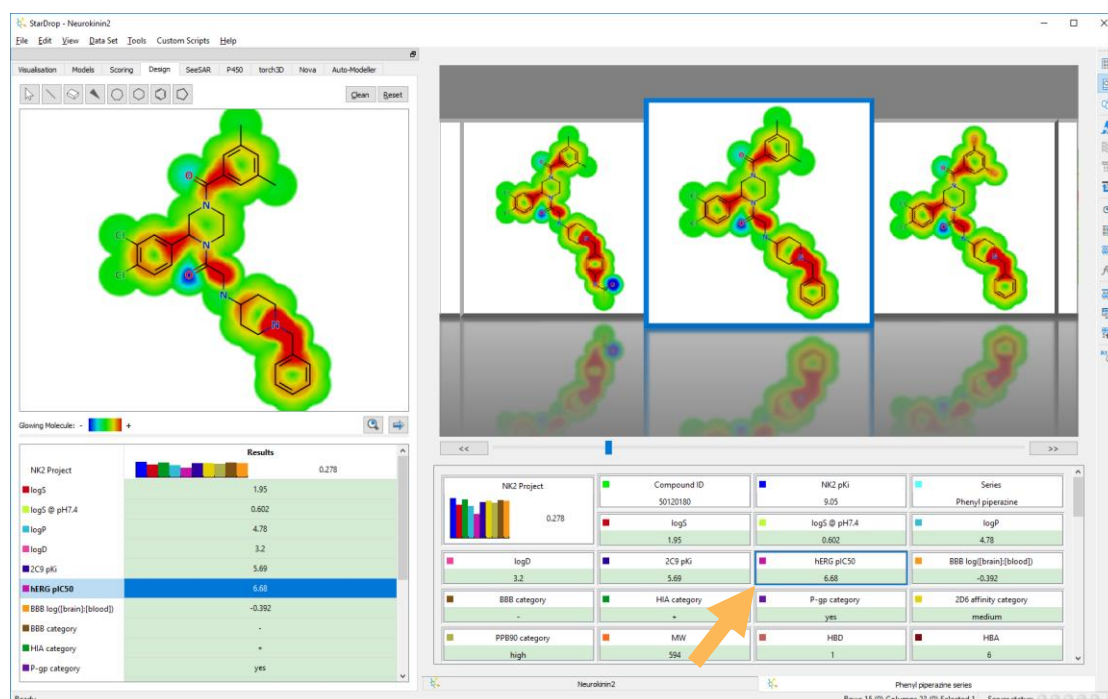
- Switch to the **Design area** to see the selected compound in the interactive designer. If necessary, you can zoom out to show the entire molecule using the mouse wheel.






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.

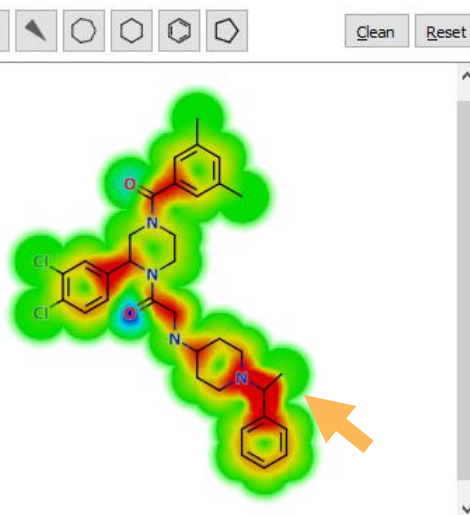
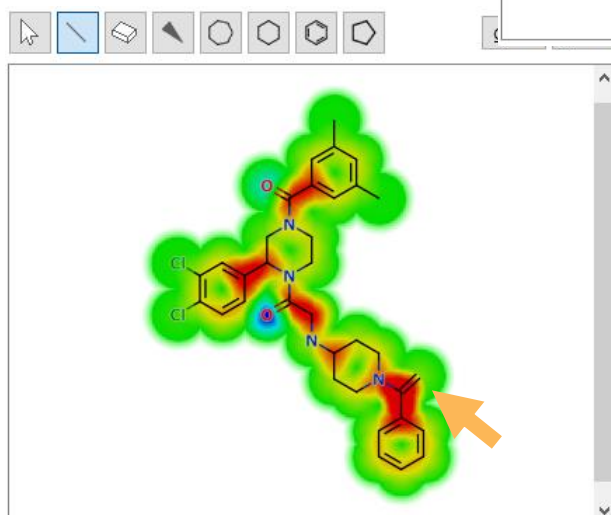
- Click on the **hERG pIC50** value in the summary table in Molecule View.



This highlights each molecule, using StarDrop's Glowing Molecule, to indicate which regions are having the greatest influence on the predicted value for hERG pIC<sub>50</sub>. 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

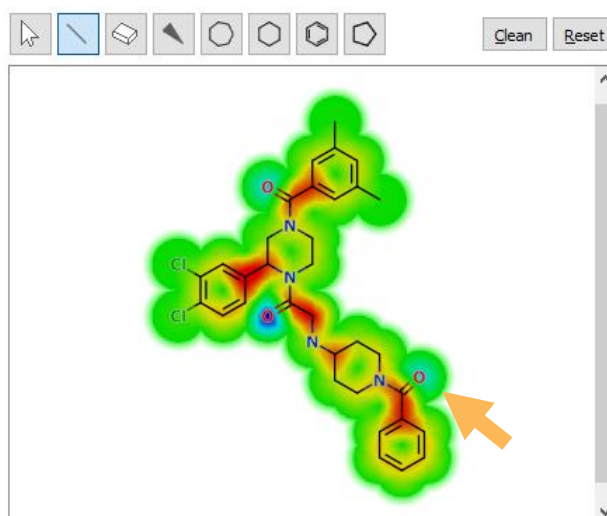
one that is highlighted blue, bearing in mind that the substitutions might be beneficial or detrimental to NK2 potency. Note that the basic piperidine nitrogen and adjacent linker is highlighted in red. We will change this amine to an amide and explore the impact of this change.

- Select the **bond tool** , 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).



- Now 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 is having an effect that is decreasing the predicted hERG  $pIC_{50}$  and that these changes have not-only reduced the predicted hERG affinity, but have also reduced the logP.

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 will not cover the Auto-Modeller in this exercise, but if you would like to see a demonstration please contact [stardrop-support@optibrium.com](mailto:stardrop-support@optibrium.com).

This example illustrated how we are able to explore the chemical space of our project to look for compounds and series which have a good overall balance of properties and investigate the SAR within an individual series. We have also seen how we are able to focus in on individual compounds and use the Glowing Molecule to help design compounds which better meet our project needs.