

## Worked Example:

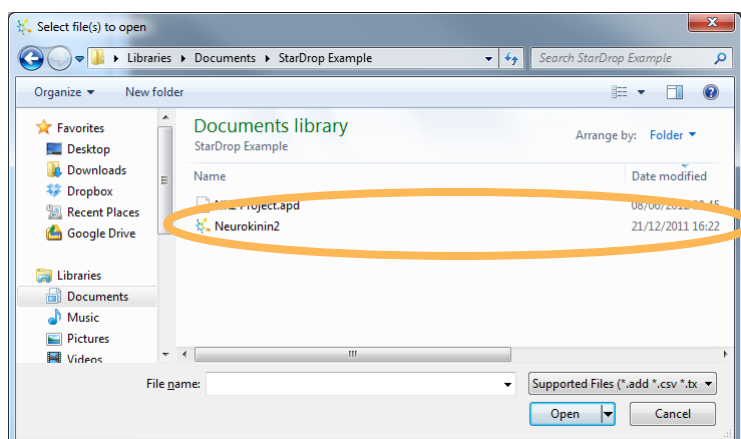
### Guiding Compound Selection and Design in Hit-to-Lead

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

- Start StarDrop from the Start menu.
- Open the file **Neurokinin2.add** by using the **File -> Open** menu.



- You will see a spreadsheet containing 191 structures and their measured affinities for Neurokinin 2 (in the column NK2 pKi).

	Structure	Compound ID	NK2 pKi	Series
1		50116745	9.523	Amide oxime
2		50110673	9.398	Amide oxime
3		50116733	9.398	Amide oxime
4		50110650	9.301	Amide oxime
5		50116727	9.301	Amide oxime
6		50116722	9.222	Amide oxime
7		50116082	9.222	Amide oxime
8		50116732	9.222	Amide oxime
9		50110655	9.222	Amide oxime

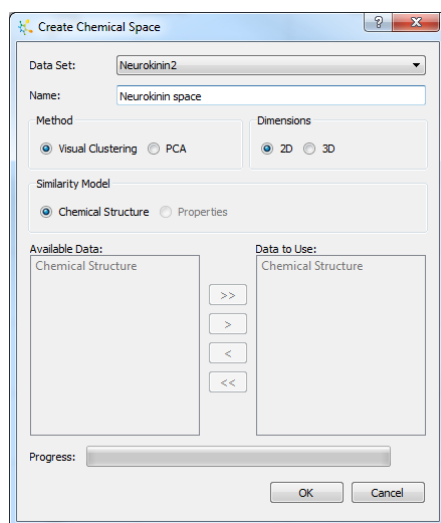
- To explore the chemical diversity of this library and the distribution of potency across it, we are going to first take a look at a chemical space plot of this data. Click on the **Visualisation** tab on the left.

- Click on the  button and then click **Create...** to create a new chemical space plot.

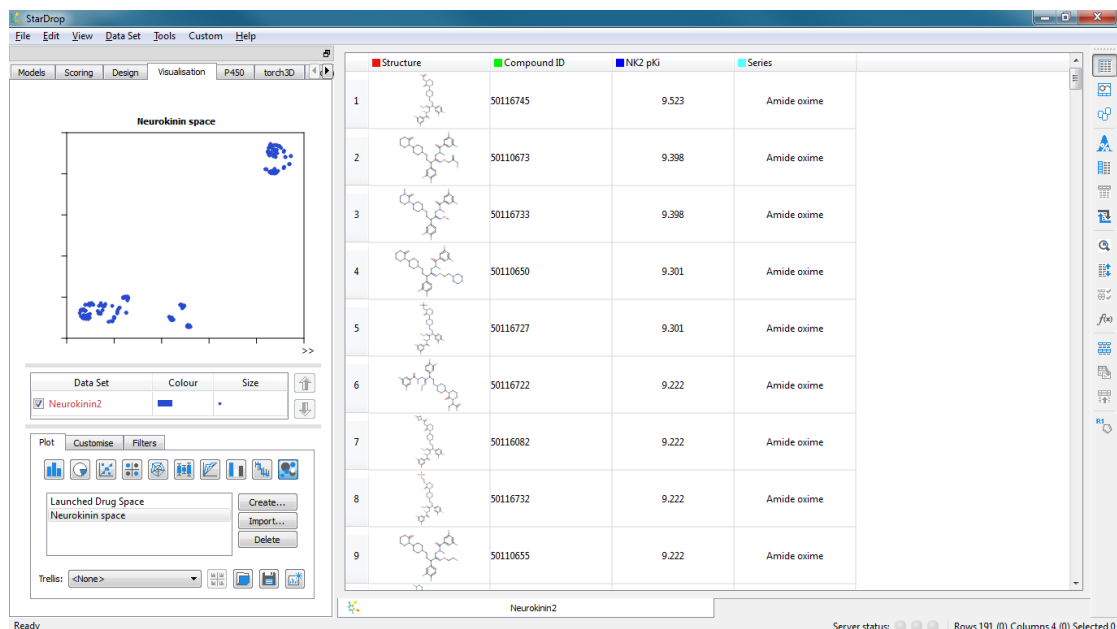
The screenshot shows the 'Visualisation' tab in StarDrop. The main area is a blank plot. Below the plot is a control panel with the following elements:

- Data Set:** A dropdown menu showing 'Neurokinin2' with a checkmark.
- Colour:** A color selection box.
- Size:** A size selection box.
- Plot:** A section with 'Customise' and 'Filters' tabs. It contains several icons for different plot types. The 'Chemical space plot' icon (a globe with a network) is highlighted by an orange arrow.
- Buttons:** 'Create...', 'Import...', and 'Delete' buttons are located below the plot icons. The 'Create...' button is highlighted by an orange arrow.
- Trellis:** A dropdown menu currently set to '<None>'.

- In the dialogue box that appears, type in a name for the chemical space plot you're about to create – in the example below we have called it **Neurokinin space**.

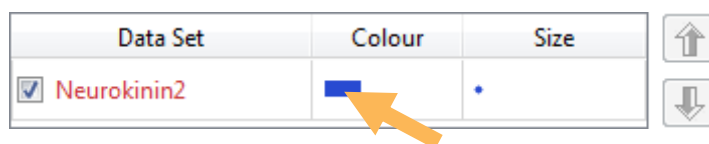


- We can build different types of chemical space plots based on chemical structure and/or property data, but in this case we're going to use the defaults, **Chemical Structure Only**, the **Visual Clustering** method and **2D**, as shown above.
- Click the **OK** button and wait until the chemical space has been created and the dialogue box disappears. When the process is complete a chemical space will be displayed in the **Visualisation** tab.

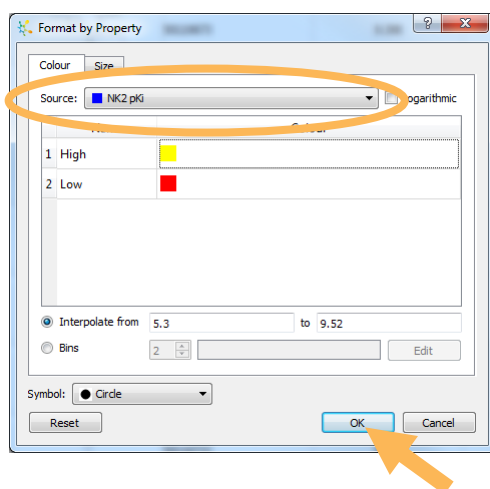


- Each point in the plot corresponds to one compound in the data set. Hovering the mouse pointer over a point will show the compound structure. Selecting a row in the data set will select the corresponding point in chemical space and vice versa.

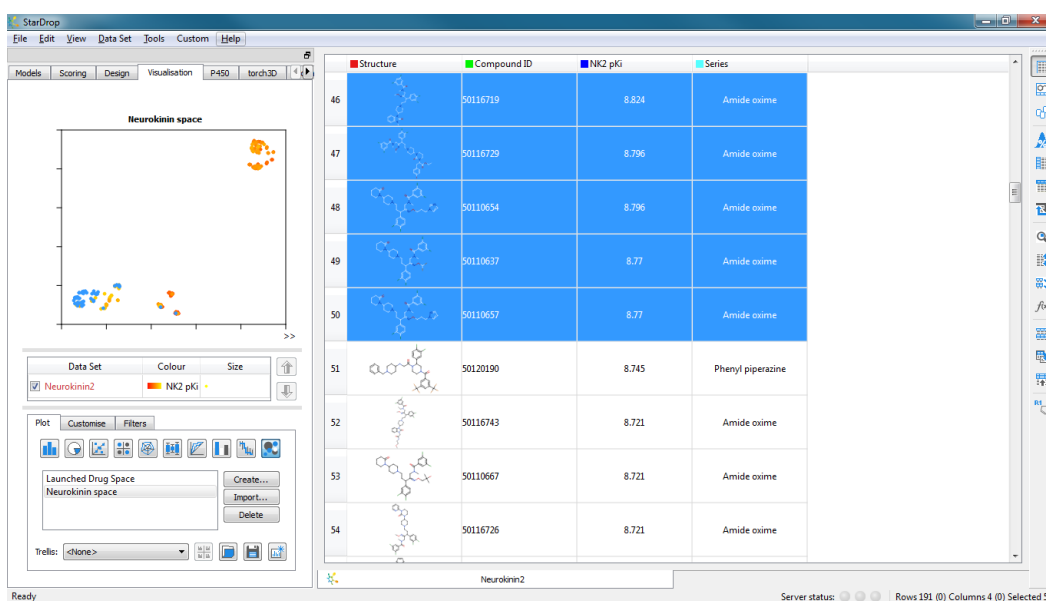
- To see the distribution of  $pK_i$  values across the chemical space, we need to change the key. Click on the block of colour next to 'Neurokinin2' in the key, as shown below, to bring up the **Format by Property** dialogue.



- In the resulting dialogue, change the **Source** for the colour to **NK2 pKi** and click **OK**.




- The compounds with the highest  $pK_i$ s will be highlighted yellow, the lowest red.
- You will notice that a large proportion of the compounds with the highest  $pK_i$  values come from one of the three regions where compounds are clustered. To confirm this, select the first 50 rows (the data are already sorted) by clicking on row 1, holding down the **SHIFT** key and then scrolling to and selecting row 50.

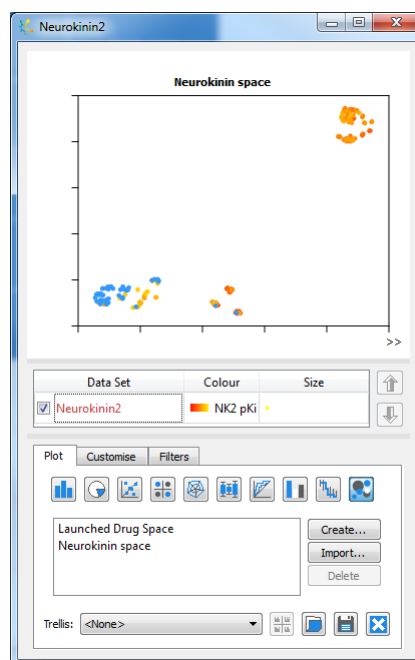


This may be the best region in which to focus for selecting compounds, but first we should consider the other properties that are important in a high quality lead.

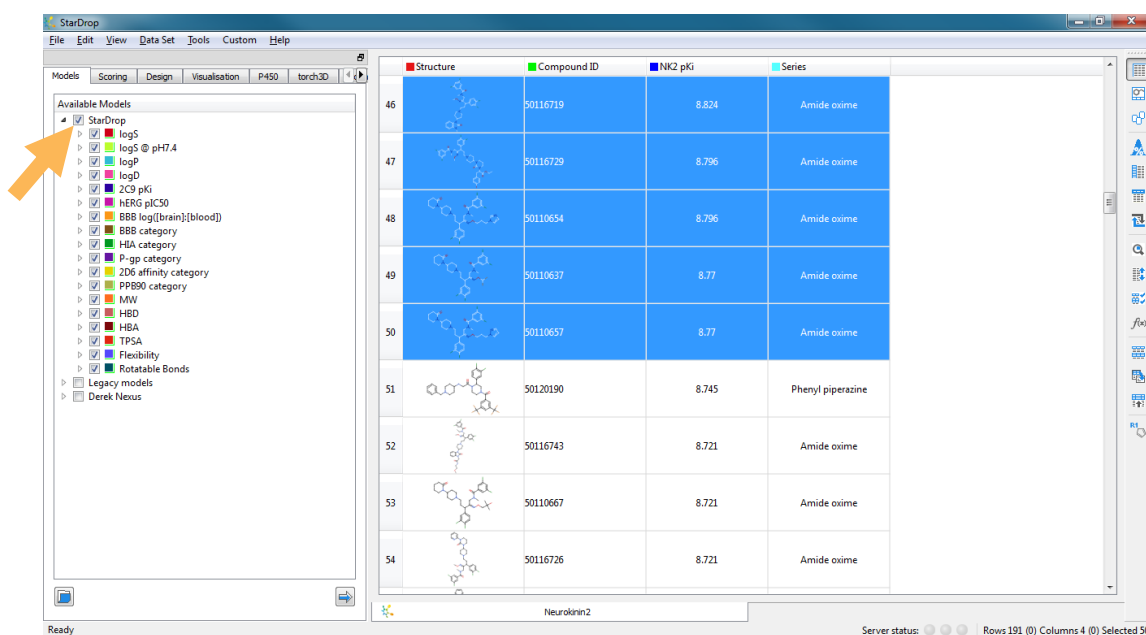
- We're going to keep this plot to use later, so










click on the  button in the **Visualisation** tab to create a separate window containing this plot.


If you wish, you can minimise this window to keep it out of the way for now.

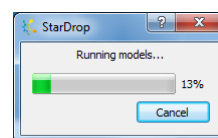


- We have no further measured data at this stage, so we're going to generate some predictions of ADME and physicochemical properties. Click on the **Models** tab.
- To run all of the StarDrop models, click the tick box next to the word StarDrop (this selects all of the models under this branch).



	Structure	Compound ID	NK2 pKi	Series
46		50116719	8.824	Amide oxime
47		50116729	8.796	Amide oxime
48		50110654	8.796	Amide oxime
49		50110637	8.77	Amide oxime
50		50110657	8.77	Amide oxime
51		50120190	8.745	Phenyl piperazine
52		50116743	8.721	Amide oxime
53		50110667	8.721	Amide oxime
54		50116726	8.721	Amide oxime


- Click the  button on the **Models** tab and 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.

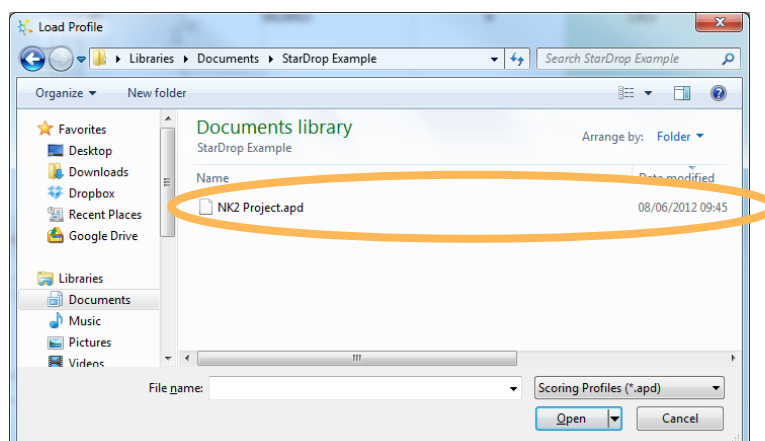


- Scroll across the data set to see the numerical and classification predictions that have been made. Due to the volume and complexity of the data, it is challenging to find the compounds which have the best balance of properties. So, we're going to use StarDrop's scoring tool which will make it easy to assess all of this information together.
- Click on the **Scoring** tab.

Structure	Compound ID	NK2 pKi	Series	logS	logS @ pH7.4
	50116745	9.523	Amide oxime	1.334	2.826
	50110673	9.398	Amide oxime	1.466	2.975
	50116733	9.398	Amide oxime	1.121	3.068
	50110650	9.301	Amide oxime	2.836	3.255
	50116727	9.301	Amide oxime	1.71	2.801
	50116722	9.222	Amide oxime	1.927	2.969
	50116082	9.222	Amide oxime	2.473	2.826
	50116732	9.222	Amide oxime	1.603	3.205
	50110655	9.222	Amide oxime	1.782	2.505

- A scoring profile enables you to define a set of criteria that is important for your project. StarDrop has some example profiles built in, but in this case we're going to load a profile designed for this Neurokinin 2 project.


- Click the  button to open a new scoring profile and load the profile called **NK2 Project.apd**.



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

The screenshot shows the StarDrop interface with a scoring profile for 'NK2 Project'. The profile includes properties like NK2 pKi, logS, HIA category, logP, hERG pIC50, 2C9 pKi, 2D6 affinity category, PPB90 category, BBB category, and BBB log([brain]:[blo...]. The table below shows the results for 9 compounds:

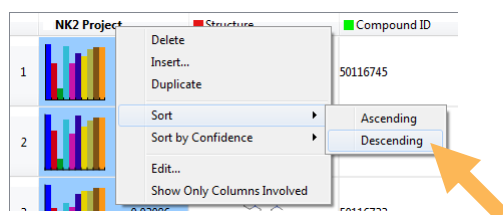
Structure	Compound ID	NK2 pKi	Series	logS	logS @ pH7.4
	50116745	9.523	Amide oxime	1.334	2.826
	50110673	9.398	Amide oxime	1.466	2.975
	50116733	9.398	Amide oxime	1.121	3.068
	50110650	9.301	Amide oxime	2.836	3.255
	50116727	9.301	Amide oxime	1.71	2.801
	50116722	9.222	Amide oxime	1.927	2.969
	50116082	9.222	Amide oxime	2.473	2.826
	50116732	9.222	Amide oxime	1.603	3.205
	50110655	9.222	Amide oxime	1.782	2.505

- Run the scoring by clicking the  button in the **Scoring** tab.
- A new column will be added to the data set containing a score for each compound, taking into account the property criteria and their relative importance in the profile, as well as the uncertainty in the underlying experimental and predicted data. The score is a value between 0 and 1, representing the likelihood of success of the compound against the profile we have defined.

The screenshot shows the StarDrop interface after scoring. A new column, 'NK2 Project', has been added to the table, containing scores for each compound. The scores range from 0.04407 to 0.06618.


Structure	Compound ID	NK2 pKi	Series	logS	NK2 Project
	50116745	9.523	Amide oxime	1.334	0.04803
	50110673	9.398	Amide oxime	1.466	0.0565
	50116733	9.398	Amide oxime	1.121	0.03086
	50110650	9.301	Amide oxime	2.836	0.06935
	50116727	9.301	Amide oxime	1.71	0.04439
	50116722	9.222	Amide oxime	1.927	0.06614
	50116082	9.222	Amide oxime	2.473	0.06618
	50116732	9.222	Amide oxime	1.603	0.05481
	50110655	9.222	Amide oxime	1.782	0.04407

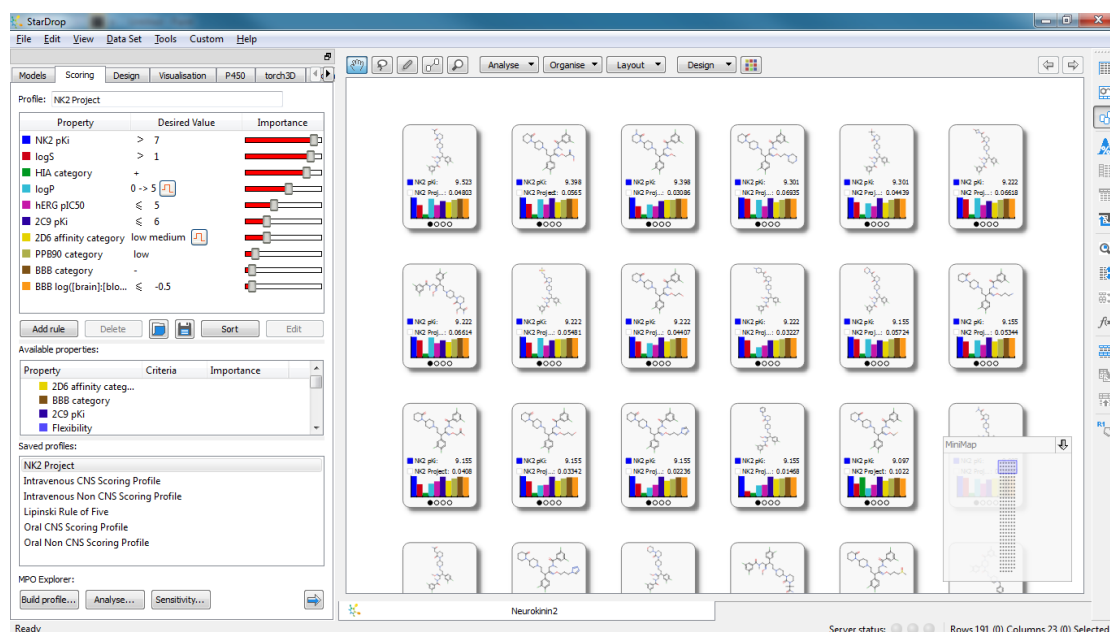
- To find the highest scoring compounds, right-click on the scoring column header and choose **Sort->Descending** to sort the data set from high to low score.



- 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 histograms with low bars indicate the 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.

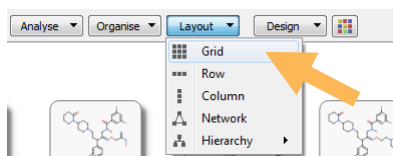
Having scored the compounds, we're now going to compare some compounds using StarDrop's Card View. This provides a convenient way to work with our compounds and data in the context of this project, by representing compounds on cards that can be moved, stacked and linked however we wish.


- Switch to **Card View** by clicking the  icon on the toolbar to the right. This will display 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.

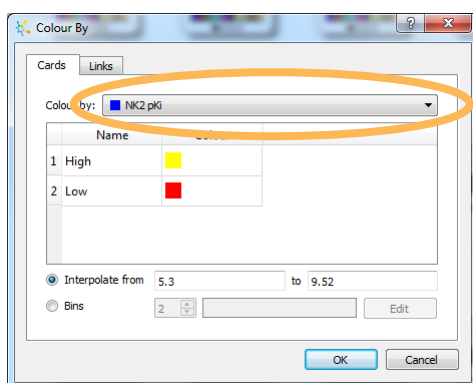




- Sort the grid in order of the score from top-left to bottom-right by choosing the **Grid** option from the **Layout** menu at the top of the screen:

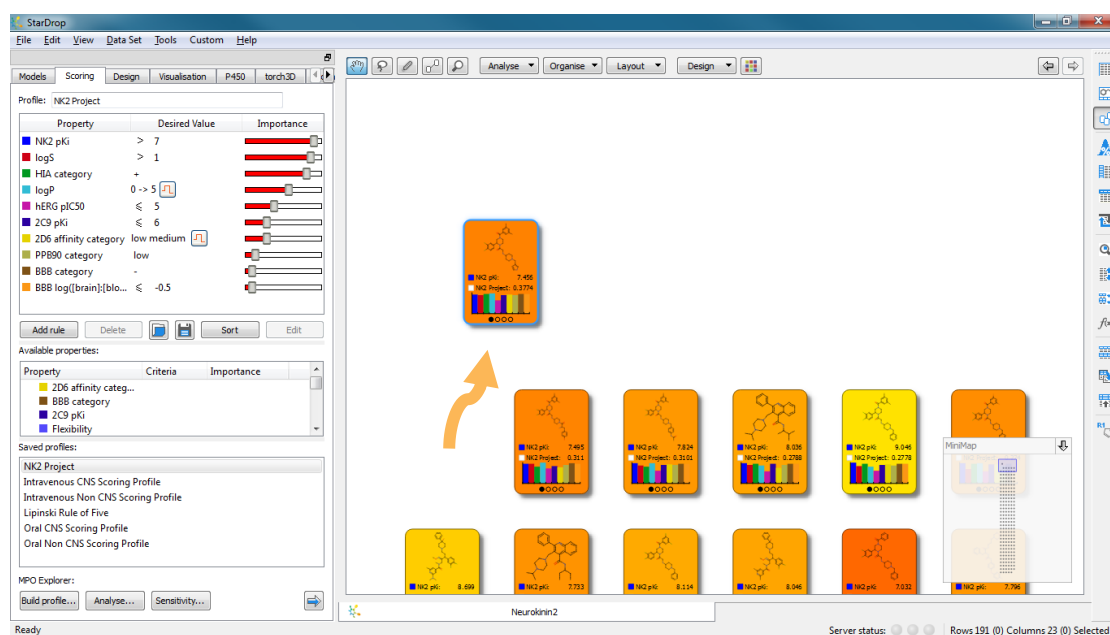


- We can now colour the cards to highlight the most potent, by clicking on the  icon and in the **Colour by** dialogue box choose to colour the points by **NK2 pKi**:



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

- Click on the highest scoring compound and drag it to the top. The display will move to make more space.



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

The screenshot shows the StarDrop software interface. On the left, there is a 'Profile' section for 'NK2 Project' with various properties and sliders. The main area displays a grid of compound cards, each showing a chemical structure, a bar chart, and numerical values for 'NK2 pKi' and 'NK2 Project'. An orange arrow indicates a drag-and-drop action from a card in the top row to a card in the middle row.

- 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.

The screenshot shows the StarDrop software interface with two zoomed-in compound cards. The left card is orange and displays the following data: NK2 pKi: 7.456, NK2 Project: 0.3774. The right card is yellow and displays the following data: NK2 pKi: 9.046, NK2 Project: 0.2778. Both cards show a chemical structure and a bar chart.

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 compounds in the library. We can investigate this further:

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

StarDrop interface showing two cards with chemical structures and property values:

Property	Value
logP:	3.971
logS:	2.332
hERG pIC50:	5.93

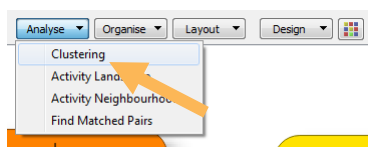
  

Property	Value
logP:	4.779
logS:	1.951
hERG pIC50:	6.679

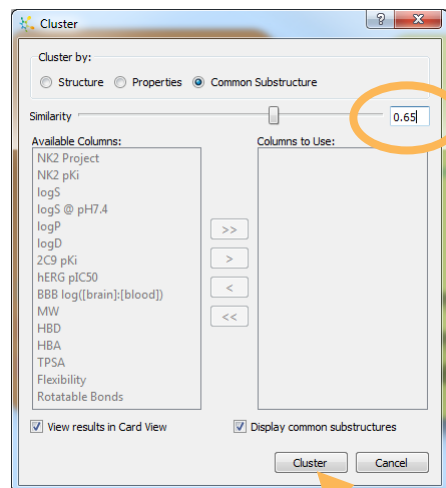
Here we can clearly see that the compound on the right has higher logP and hERG pIC50, which increase the risk for this compound and hence result in a lower score.

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.

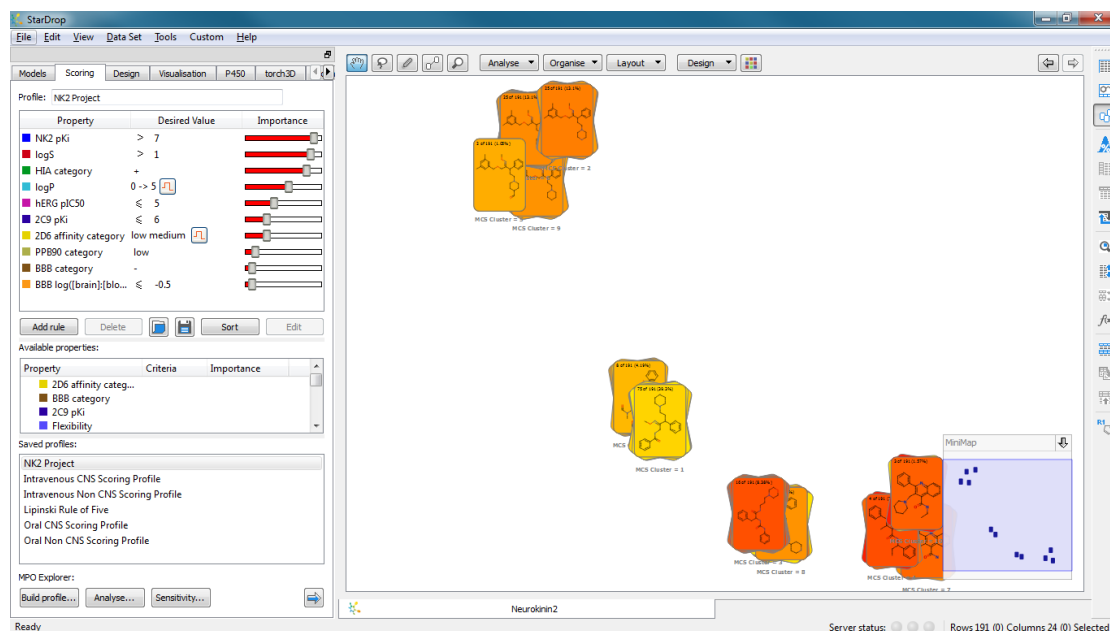
- From the **Analyse** menu at the top of the Card View, select **Clustering**.



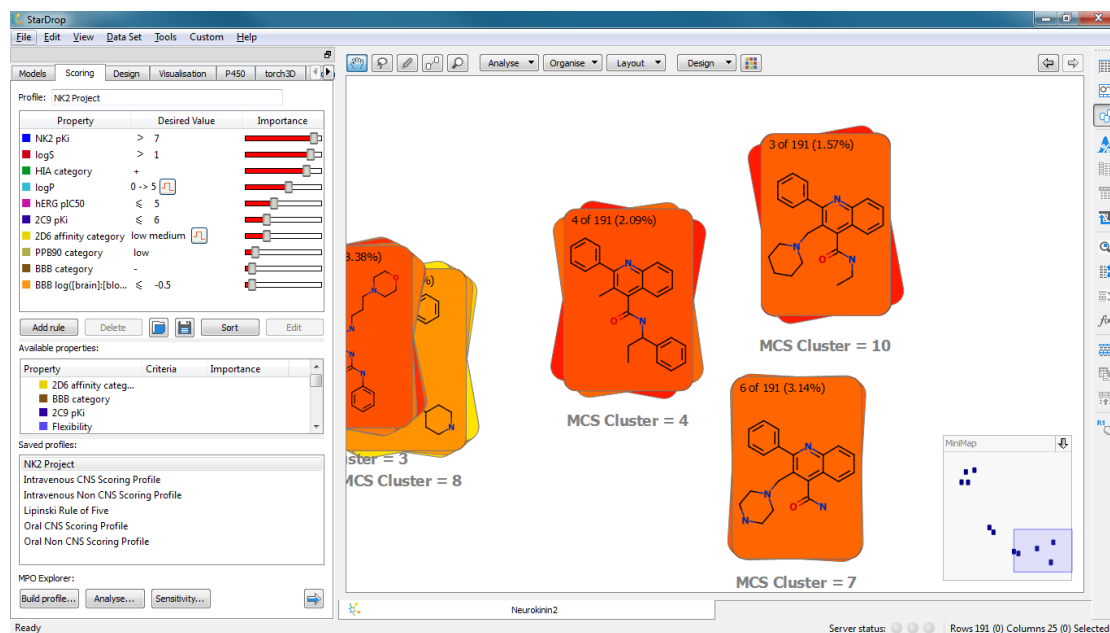
- In the **Cluster** dialogue box, choose the **Common Substructure** method and set the **Similarity** to 0.65, as shown below, and click **Cluster**.



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.

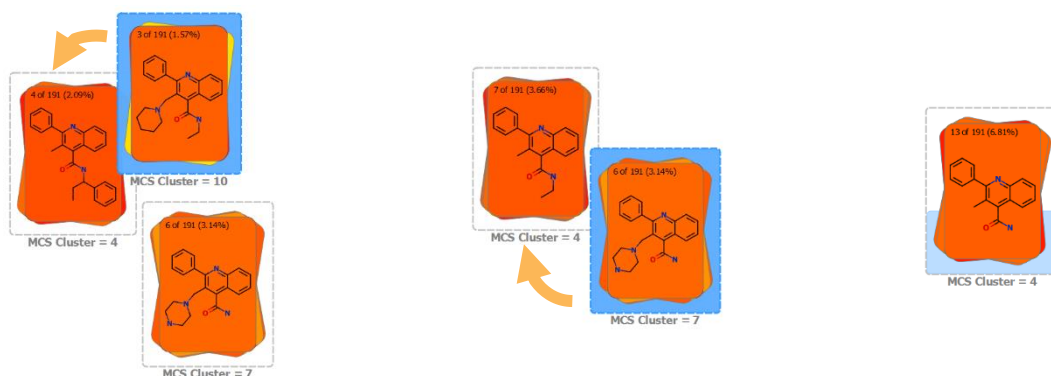


- Zoom in on the stacks corresponding to clusters 4, 7 and 10, which will be close together (to the bottom-right or top-right of the layout) and drag the background to centre these in the display. You can move the stacks to separate them slightly by clicking and dragging.

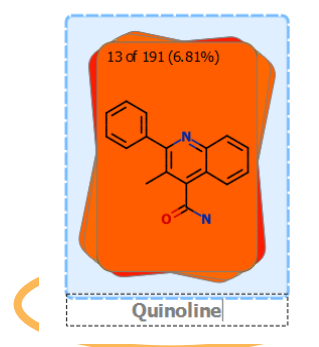


Here you can see that, although the compounds are in different stacks, this is an artefact of the cut-off assigned to the clusters and these groups all share the same quinolone scaffold and could be considered a single series.

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

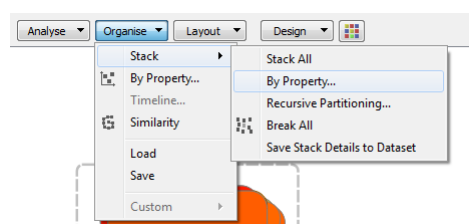


- We can then assign a name to this series by clicking and editing the label under the stack:

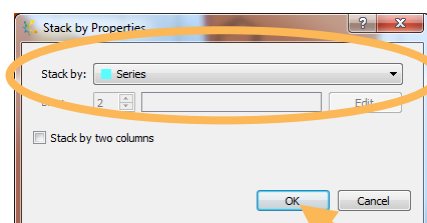


We could examine the other stacks and repeat this process, where necessary, to refine the clustering results. However, to save time, we have saved the series definitions in the data set and can use this to stack all of the compounds.

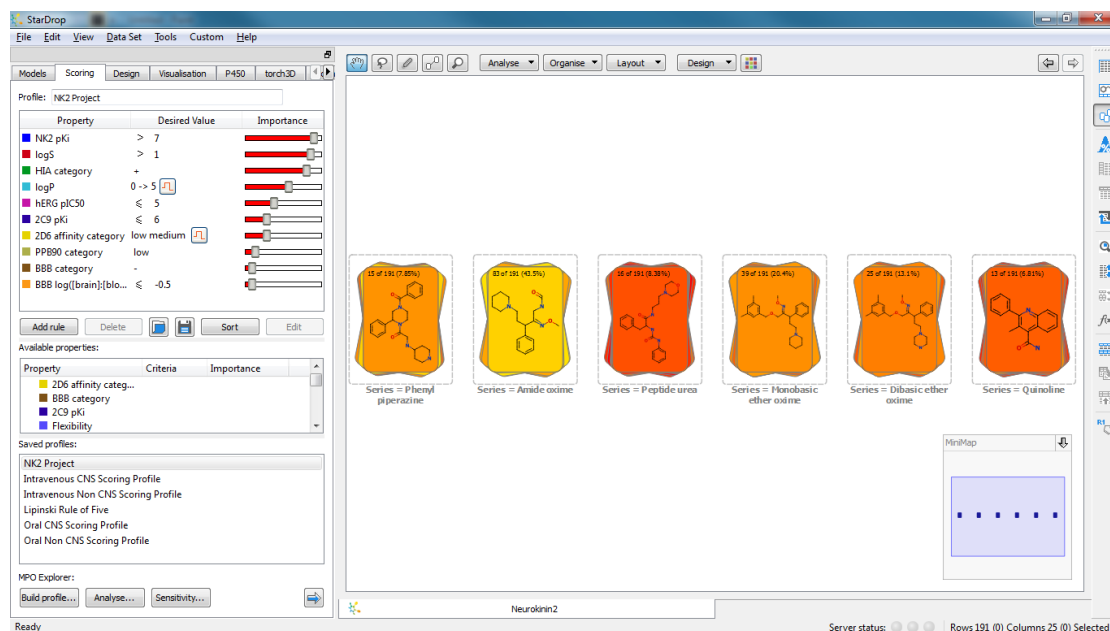
- From the **Organise** menu at the top of Card View, select the **Stack -> By property...** menu option:



- In the **Stack By Properties** dialogue box choose **Series** from the **Stack by:** drop-down menu and click **OK**.

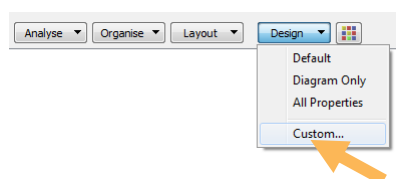


Here we can see that there are 6 series, of which the Amide oxime series has the highest average potency (it is the most yellow).

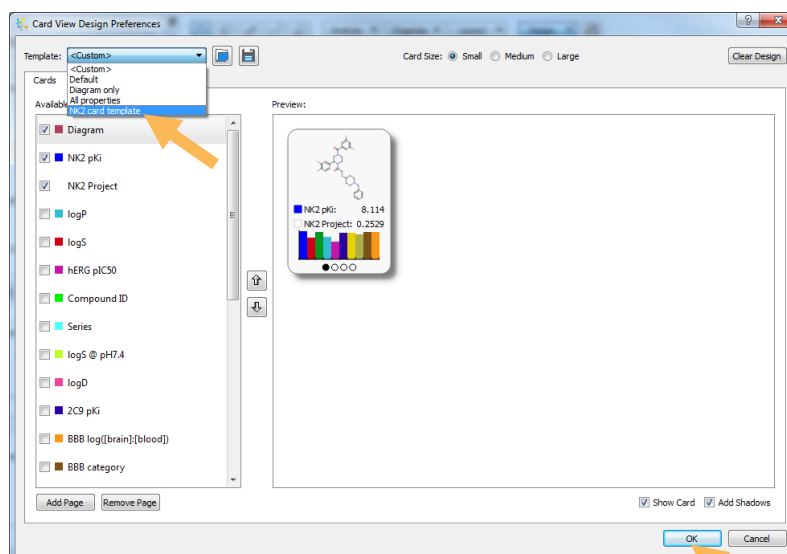


We can display more information on a stack. To illustrate this we can choose a new stack design.

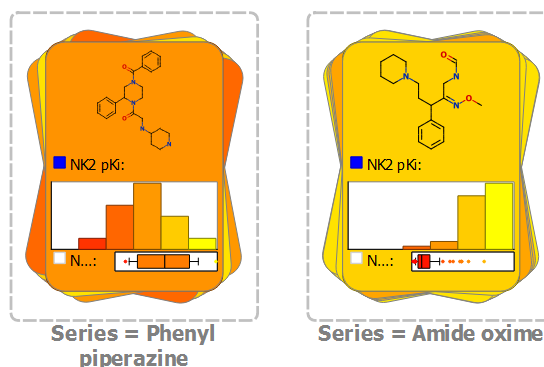
- Select the **Custom** menu option from the **Design** menu at the top of Card View.



- In the **Card View Design Preferences** dialogue box, you can choose many options for card and stack designs, but in this case we are going to choose a pre-defined template. From the **Template:** drop-down menu, choose **NK2 card template** and click **OK**:



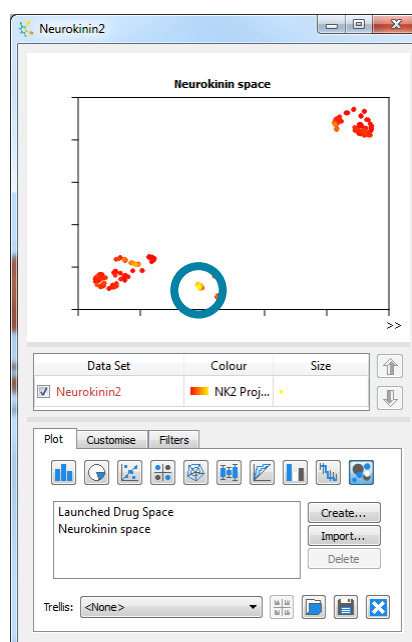
The resulting stacks show a histogram of the distribution of  $pK_i$  values of the compounds within each stack and, below this, a box plot showing the distribution of scores. This makes it straightforward to compare different series, for example the Amide oxime series has an excellent distribution of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



large majority of 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.

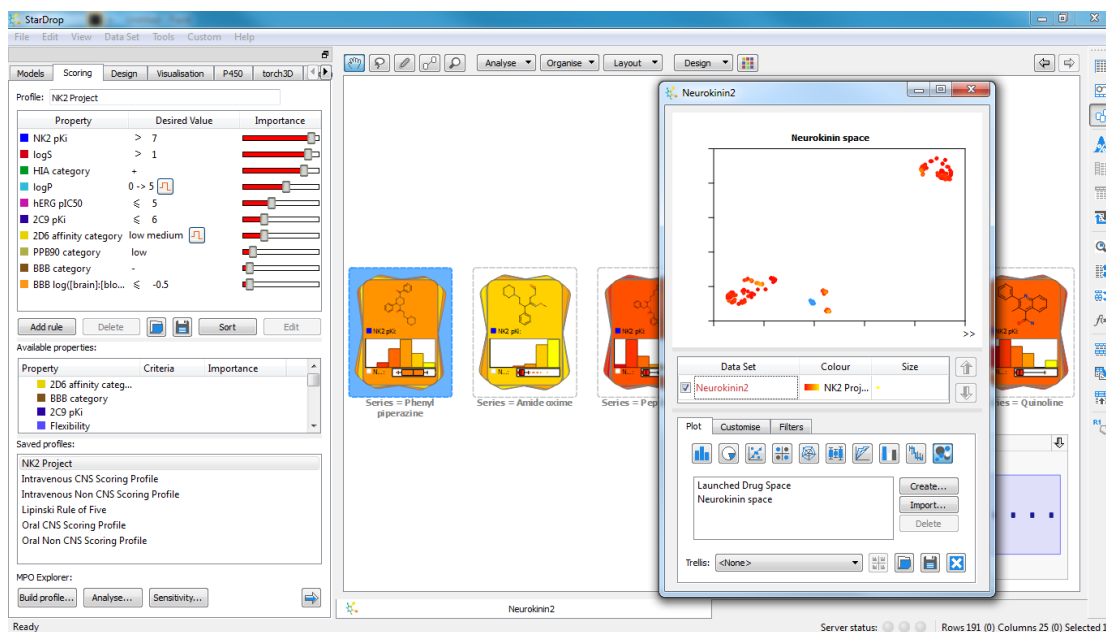
We can explore the distribution of scores across the data set as a whole by returning to the chemical space plot we created earlier.

- In the chemical space plot window click on the block of colour next to 'Neurokinin2' in the key in this window to bring up the **Format by Property** dialogue again and, as we did before, change the **Source** property to be the **NK2 Project** values. This way the points will be coloured based upon the overall scores of the compounds, not only the  $pK_i$  values.




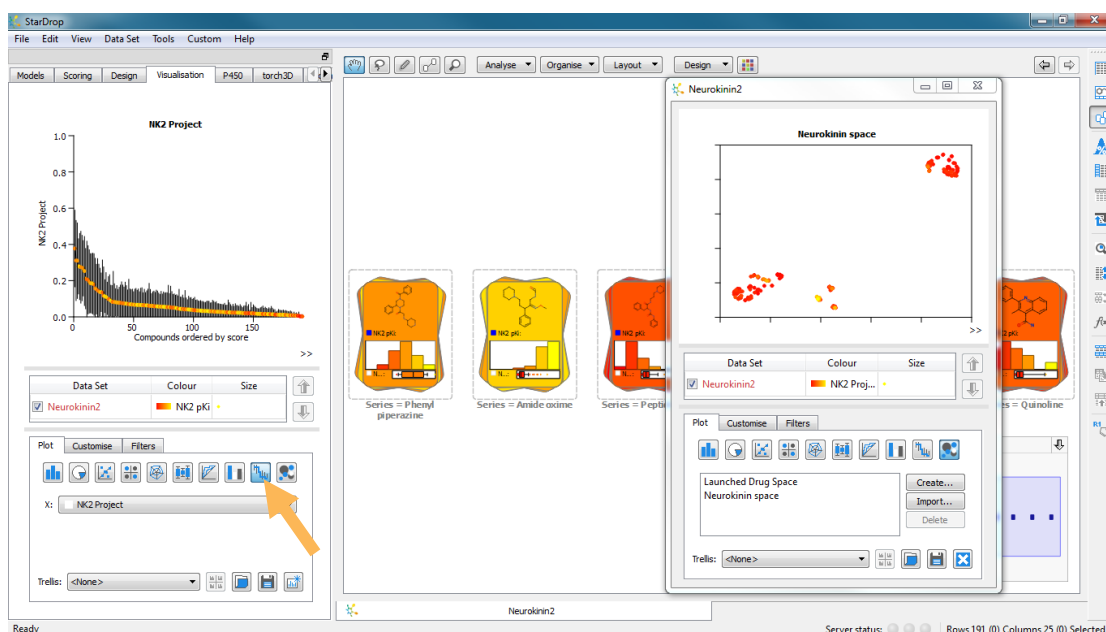
Here we can see that the pattern has changed dramatically from that we could see when looking only at potency and only a small cluster is highlighted in yellow as a 'hot spot' of high scoring compounds.

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



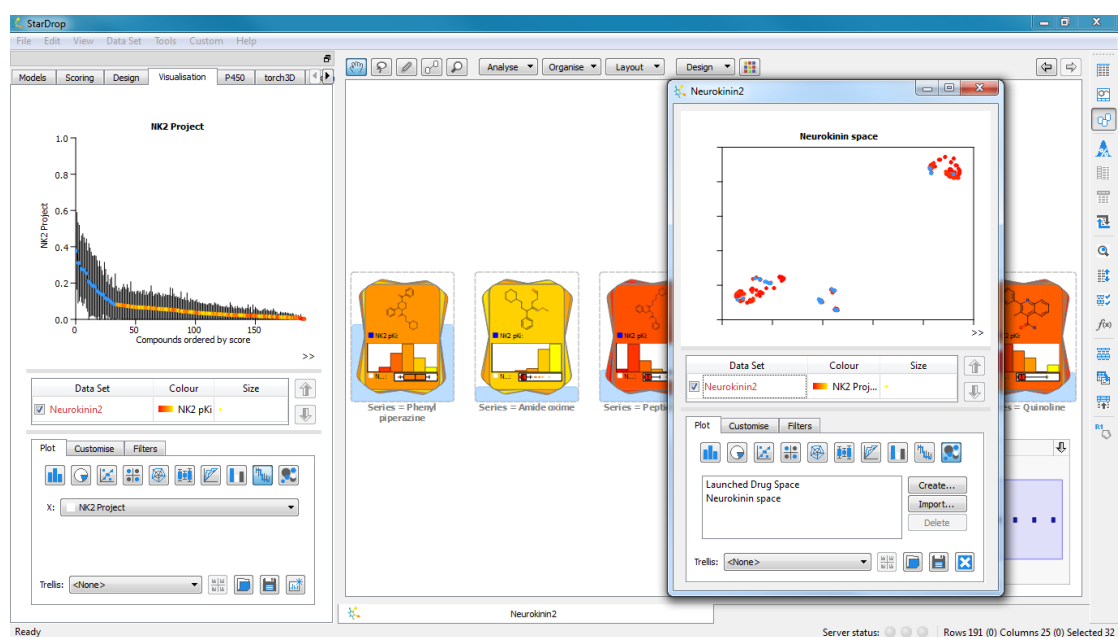
Although this appears to be the best series, we should consider the impact of the uncertainty in the underlying data on our ability to confidently distinguish between compounds. Therefore, we will use another type of plot to explore this.

- Click on the **Visualisation** tab again.
- Now select the  button on the **Visualisation** tab to change the plot type. 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.






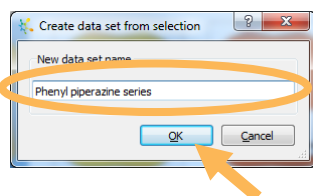
- Looking at the 'snake plot' we can see that the top ~30 compounds cannot be confidently distinguished from the top scoring compound (notice that the error bar for the top-scoring compound overlaps with the error bars of approximately the top 30 compounds). Therefore, we should consider selecting compounds from this range to further explore their properties and make a confident selection of potential lead series. Selecting the top 30 compounds in the snake plot (by drawing a ring round them with the mouse) highlights where these lie within the chemical space and within the stacks in Card View.



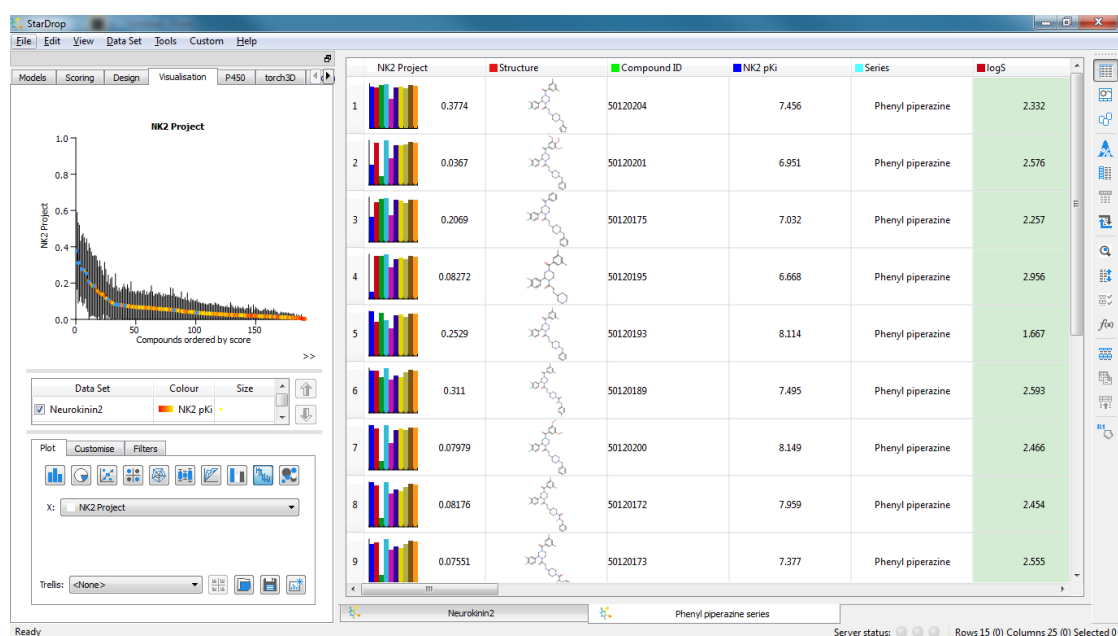
Notice that this selection includes compounds from different regions of the chemical space and a small number of compounds from each stack other than the Monobasic ether oxime series. This suggests that some of these chemistries cannot be rejected with confidence. Therefore, it may be more appropriate to sample a small number of compounds from some of these alternative chemistries to generate some experimental ADME data. These data will have greater confidence than predicted values and hence we will be able 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.

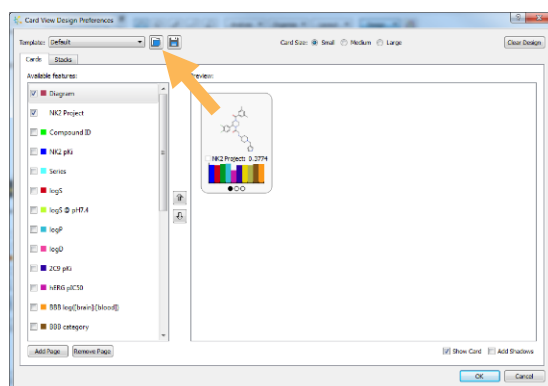
- To do this, select the stack corresponding to this series and select the **Create From Selection** menu item from the **Data Set** menu in the StarDrop menu bar or, equivalently, click the  button in the toolbar to the right.
- In the **Create data set from selection** dialogue box, give the new data set a name, e.g. “Phenyl piperazine series” and click **OK**.



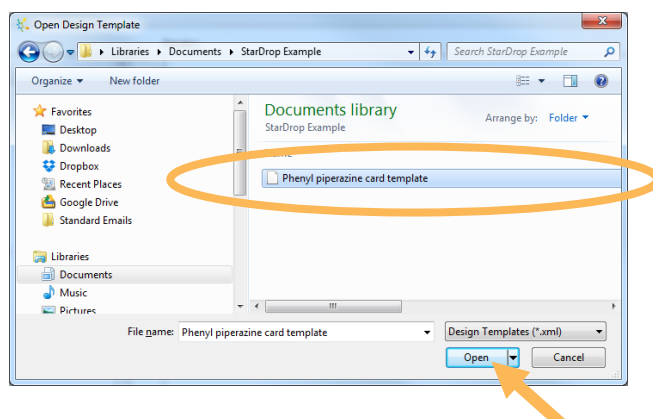
This will create a new data set containing only this series.



- Click on the Card View button, , on the toolbar to change to Card View for this data set.
- We'd like to look at these compounds in more detail, so we'll load a new template that displays more data on each page. To do this, again select the **Custom** option from the **Design** menu, as before and click the  button in the **Card View Design Preferences** dialogue box.



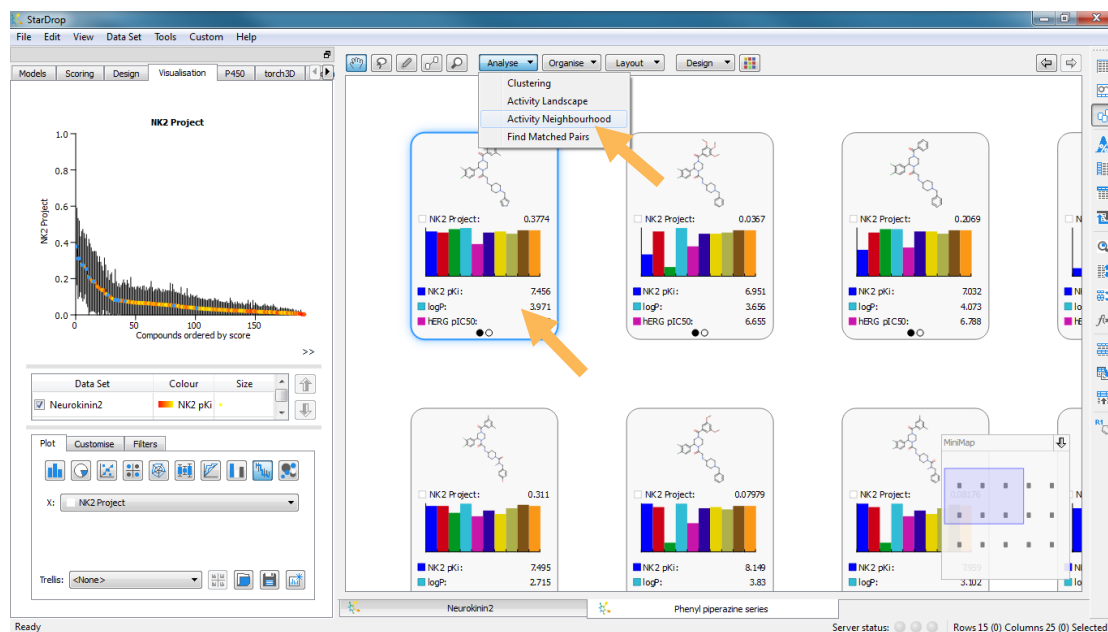
- Select the template **Phenyl piperazine card template** and click **Open**.



- Then click **OK** in to apply this new template.

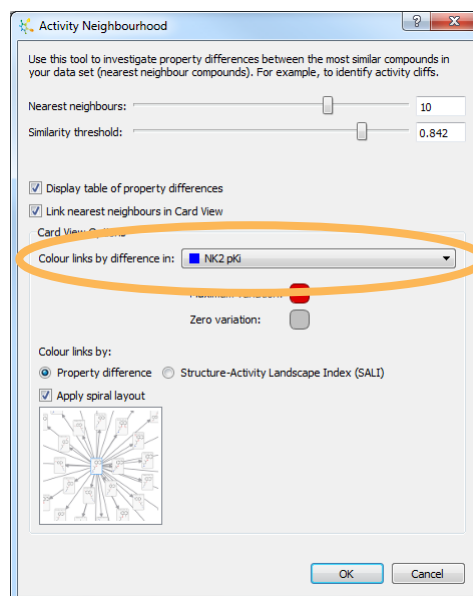
We will use the Activity Neighbourhood analysis to explore how small structural modifications can influence potency around the highest scoring compounds (this is also known as activity cliff detection).

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

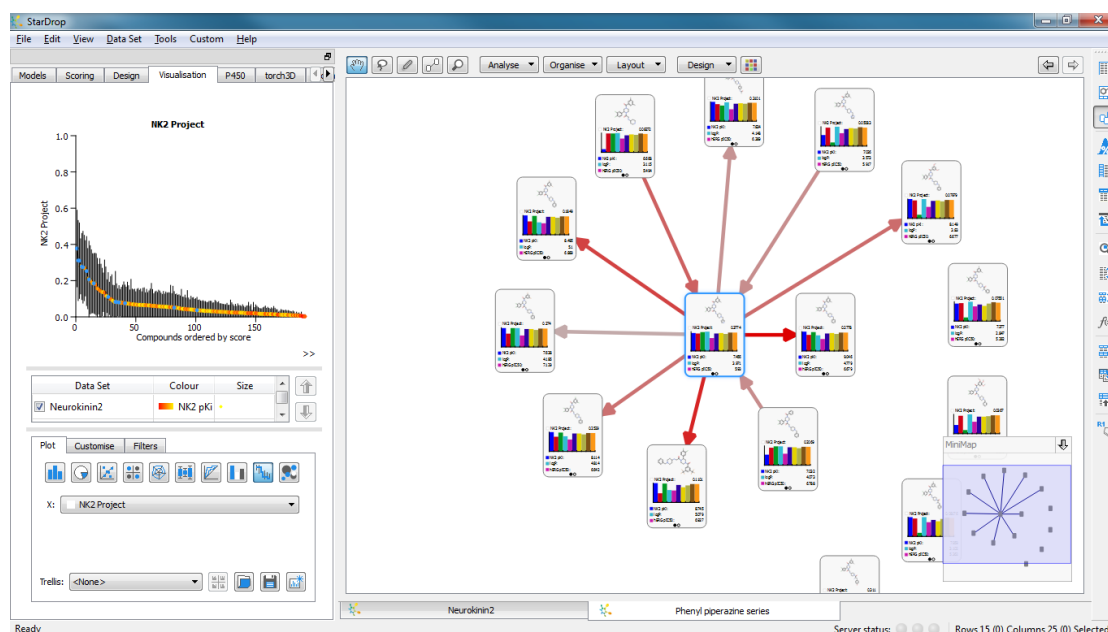


The Activity Neighbourhood tool compares each compound with the selected 'reference' compound and arranges these in order of structural similarity. The nearest neighbours are linked to the reference and that link can be coloured by the change in a selected property.

- In the **Activity Neighbourhood** dialogue box, use the default number of nearest neighbours (10) to link and choose to colour the links by the difference in **NK2 pKi**:

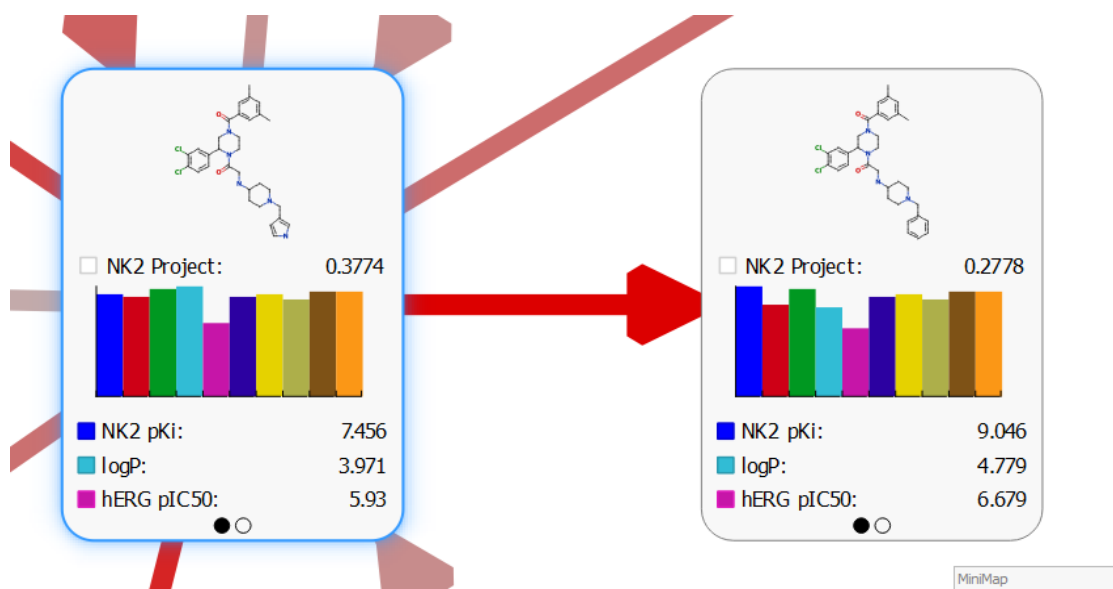


The resulting display shows the reference compound in the centre and the remaining compounds arranged in a spiral in order of decreasing similarity (increasing distance) from the reference. The 10 closest compounds are joined to the reference by a link; the arrow indicates the direction of *increase* 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 strong, 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.

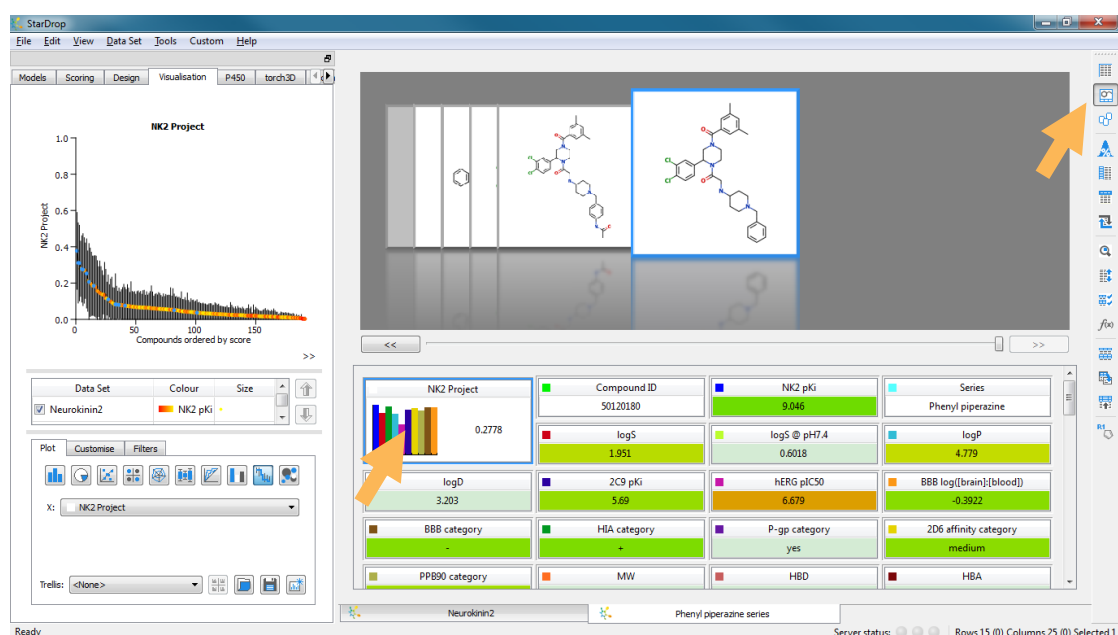
- Zoom in to these compounds (using the middle mouse wheel or **CTRL** and = keys).



Here we can see that replacement of the pyrole group 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.

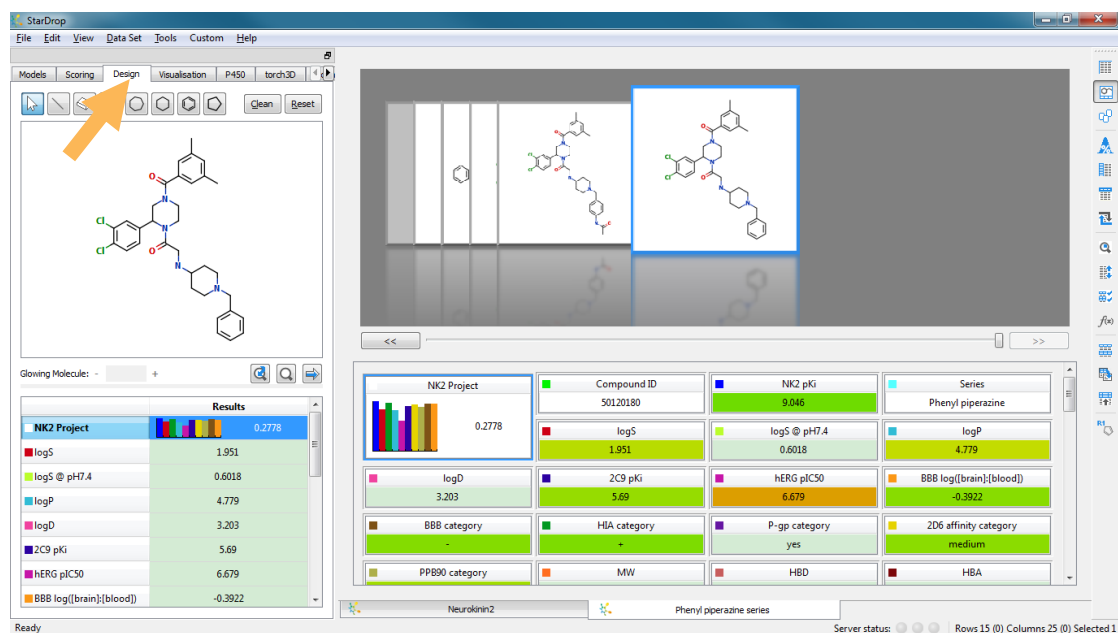
We might wish to explore alternative approaches to improving these properties while retaining potency against the target. To do this, we can use **Molecule View** and the interactive **Design** tool with **Glowing Molecule**.

- Select the compound on the right and click the  button on the right-hand tool-bar to switch into Molecule View.



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 and clicking on the score cell will show a heat map indicating the contribution of each property to the overall score.

- Switch to the **Design** tab and the selected compound will appear in the interactive designer. (If necessary you can zoom out in the designer to show the entire molecule using the middle mouse wheel or drag the bar between the drawing area and the summary table below the editor to make the drawing area taller and drag the bar between the Design tab and the data set to make it wider).




- Click on the **hERG pIC<sub>50</sub>** value in the summary table in Molecule View.

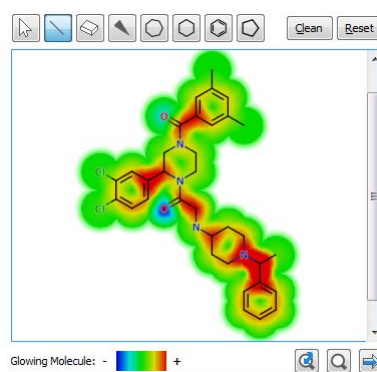
The screenshot shows the StarDrop software interface. On the left, a 3D model of a molecule is shown with a glowing surface, where red indicates regions that increase the hERG pIC<sub>50</sub> value and blue indicates regions that decrease it. On the right, a summary table displays various properties for the molecule. The 'hERG pIC<sub>50</sub>' value is highlighted in blue, and an orange arrow points to it.

Property	Value
NK2 Project	0.2778
logS	1.951
logS @ pH7.4	0.6018
logP	4.779
logD	3.203
2C9 pKi	5.69
<b>hERG pIC<sub>50</sub></b>	<b>6.679</b>
BBB log([brain]:[blood])	-0.3922
Compound ID	50120180
NK2 pKi	9.046
Series	Phenyl piperazine
BBB category	-
HIA category	+
ADME category	yes
PPB90 category	-
MW	-
HBD	-
HBA	-

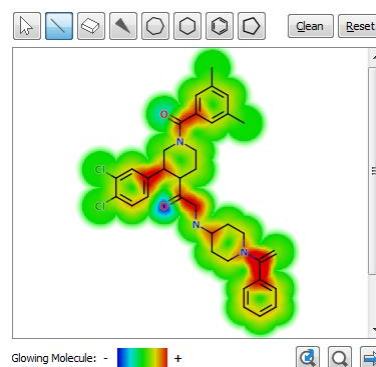
This highlights each molecule, using StarDrop's Glowing Molecule, to indicate which regions are having the greatest influence on the predicted value for hERG pIC<sub>50</sub>. Areas highlighted red tend to increase the property value and areas highlighted blue tend to decrease the property value. Therefore, to reduce hERG inhibition, we should remove a red region or add a blue group, 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.

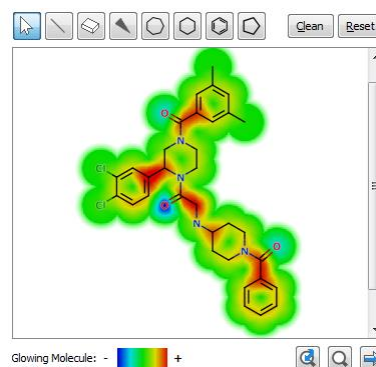
- Select the bond tool  and click on the carbon chain atom next to the amine and while the mouse is clicked, drag away and then release to draw a methyl group. Make sure you don't connect this to any other atom! (If you do, hover the mouse pointer over the drawing and press **CTRL** and **Z** to undo 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.



Notice that the amide O has a blue glow, indicating that it is having an effect to decrease the predicted hERG pIC<sub>50</sub> 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.

At the top of the summary table you will see a question mark against the score. This is because we do not yet have an experimental pK<sub>i</sub> value for the NK2 affinity. 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 with 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.