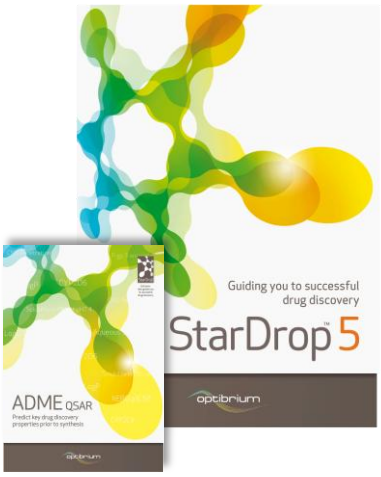


Worked Example:

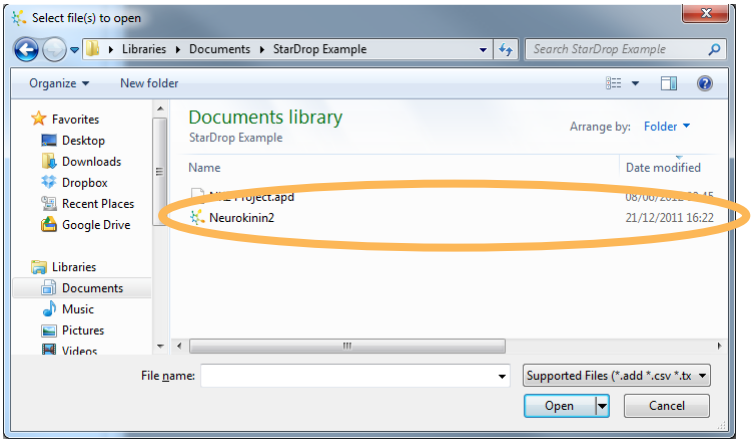
Guiding Selection and Design in Hit-to-Lead

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.

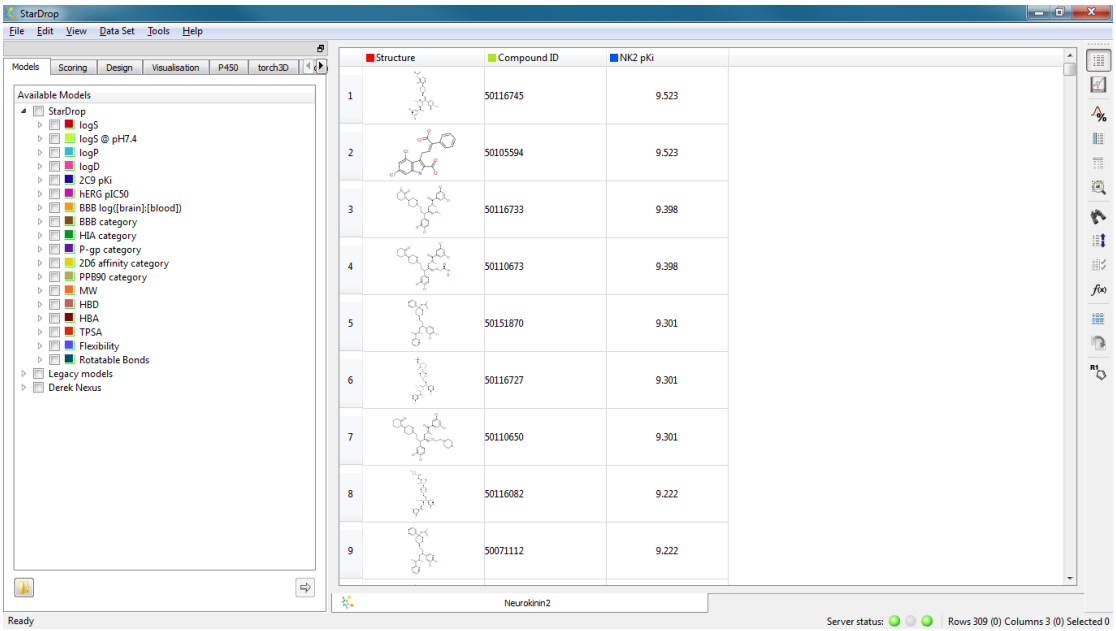
During this exercise we will use a variety of StarDrop's capabilities to explore the data in order to evaluate and select molecules with a good balance of properties. Step-by-step instructions for all the features you will need to use in StarDrop are provided, along with screenshots and examples of the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

Exercise

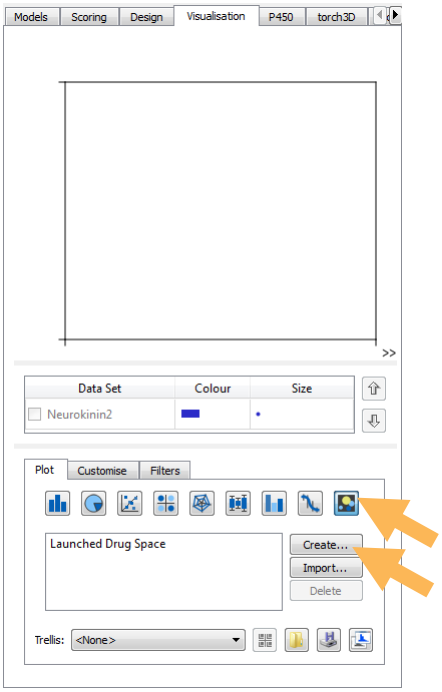
- Start StarDrop from the Start menu.
- Open the file **Neurokinin2.add** by using the **File -> Open** menu.



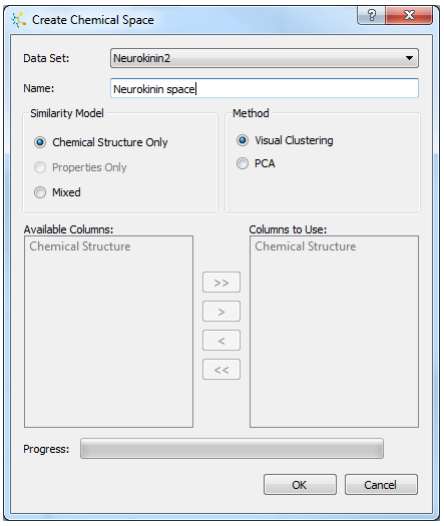
- You will see a spreadsheet containing 309 structures and their measured affinities for Neurokinin 2 (in the column NK2 pKi).



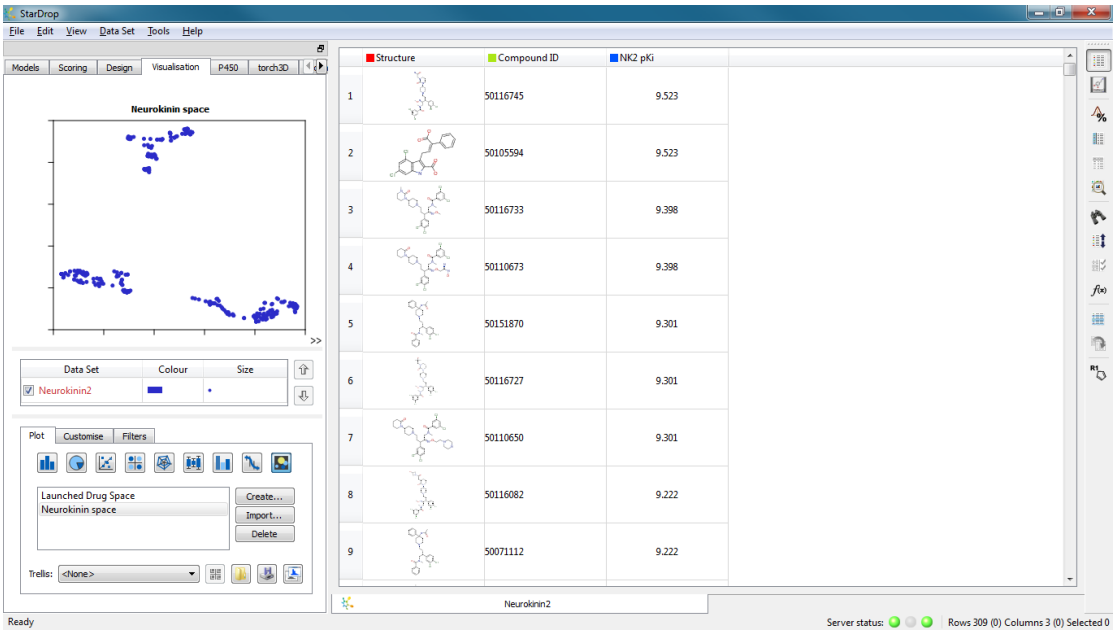
- To explore the chemical diversity of this library and the distribution of potency across it, we are going to first take a look at a chemical space plot of this data. Click on the **Visualisation** tab on the left.
- Click on the  button and then click “Create” to create a new chemical space plot.



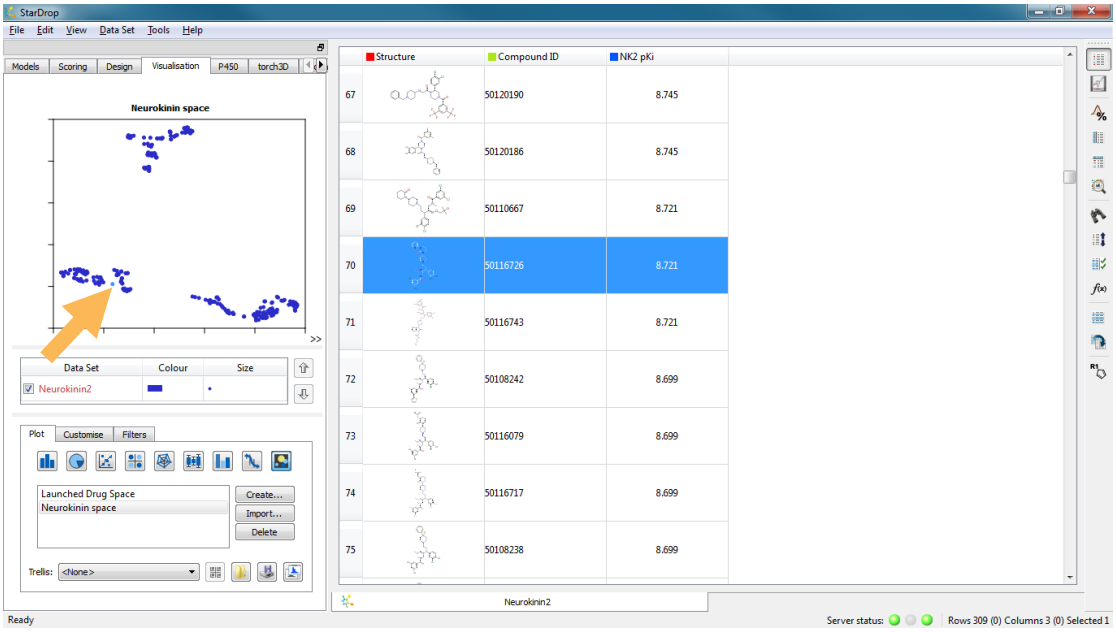
- In the dialogue box that appears, type in a name for the chemical space plot you're about to create – in the example below we have called it **Neurokinin space**.



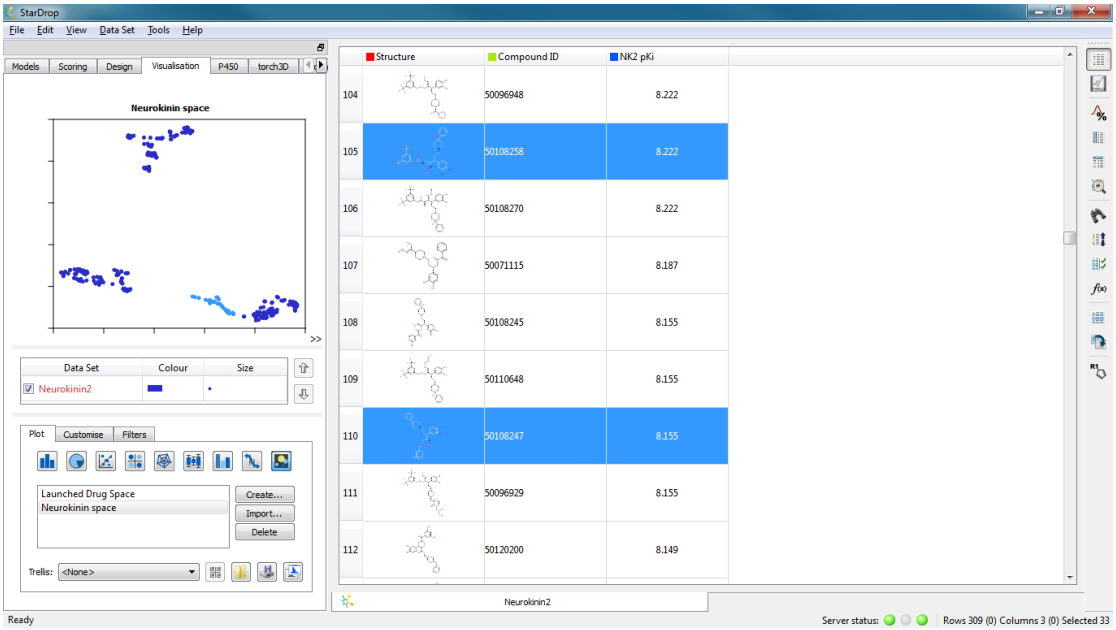
- We can build different types of chemical space plots based chemical structure and/or property data, but in this case we're going to use the defaults, **Chemical Structure Only** and the **Visual Clustering** method, as shown above.
- Click the **OK** button and wait until the chemical space as been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



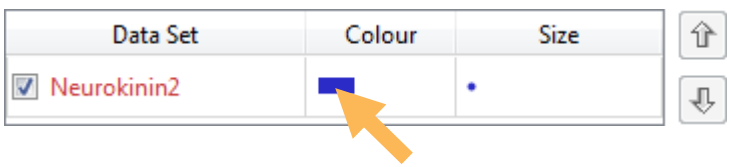
- Each point in the plot corresponds to one compound in the data set. Hover the mouse pointer over a point to see the corresponding structure.
- Selecting a row in the data set will select the point in chemical space and vice versa.



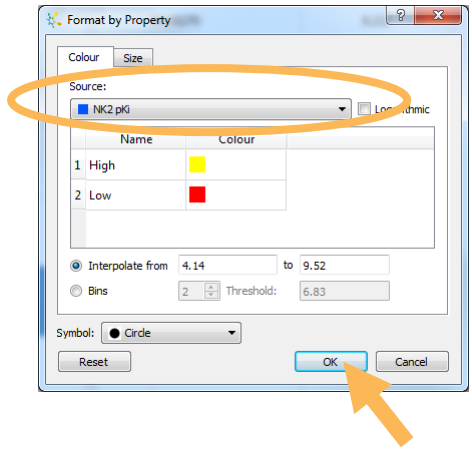
- Use the mouse to draw around a selection of points in the chemical space and they will be selected in the data set.



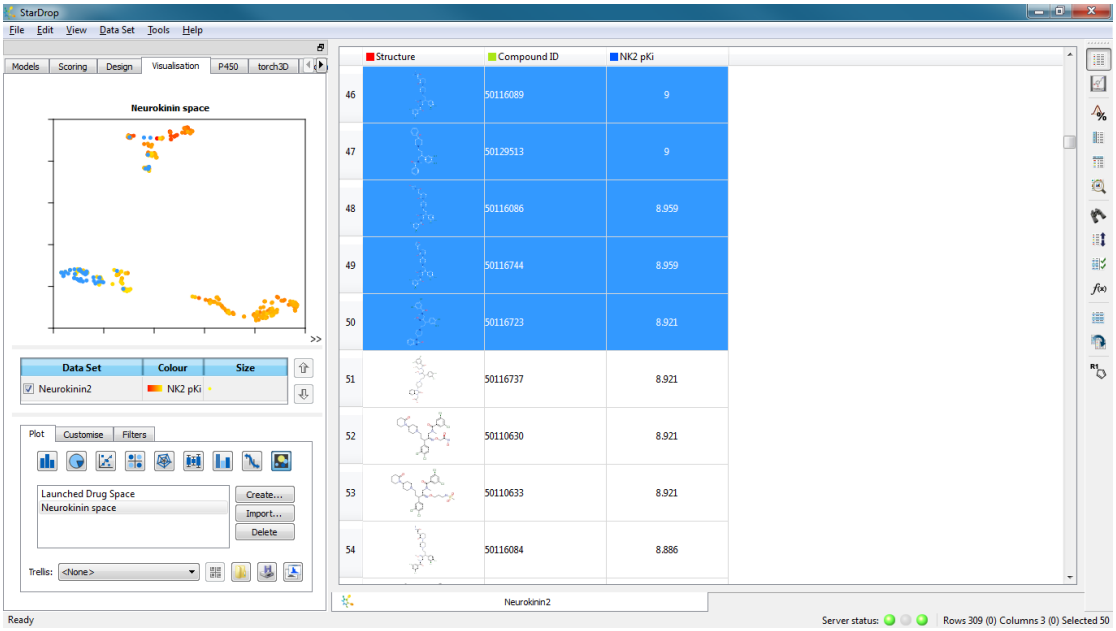
- To see the distribution of pK_i values across the chemical space, we need to change the key. Click on the block of colour next to the 'Neurokinin2 library' in the key, as shown above, to bring up the **Format by Property** dialogue.



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



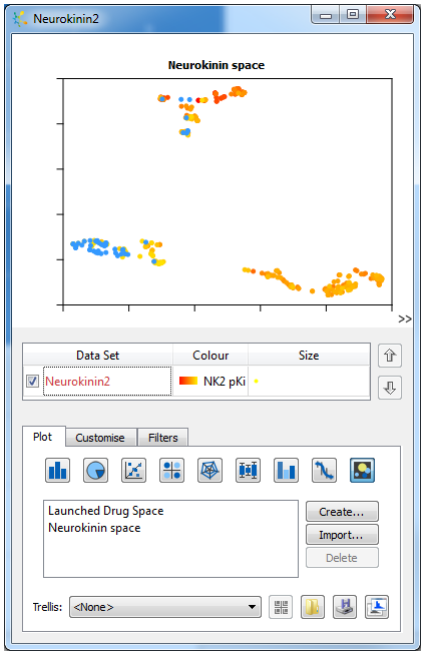
- The compounds with the highest pK_is will be highlighted yellow, the lowest red.
- You will notice that a large proportion of the compounds with the highest pK_i values come from one of the three regions where compounds are clustered. To confirm this, select the first 50 rows (the data is already sorted)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



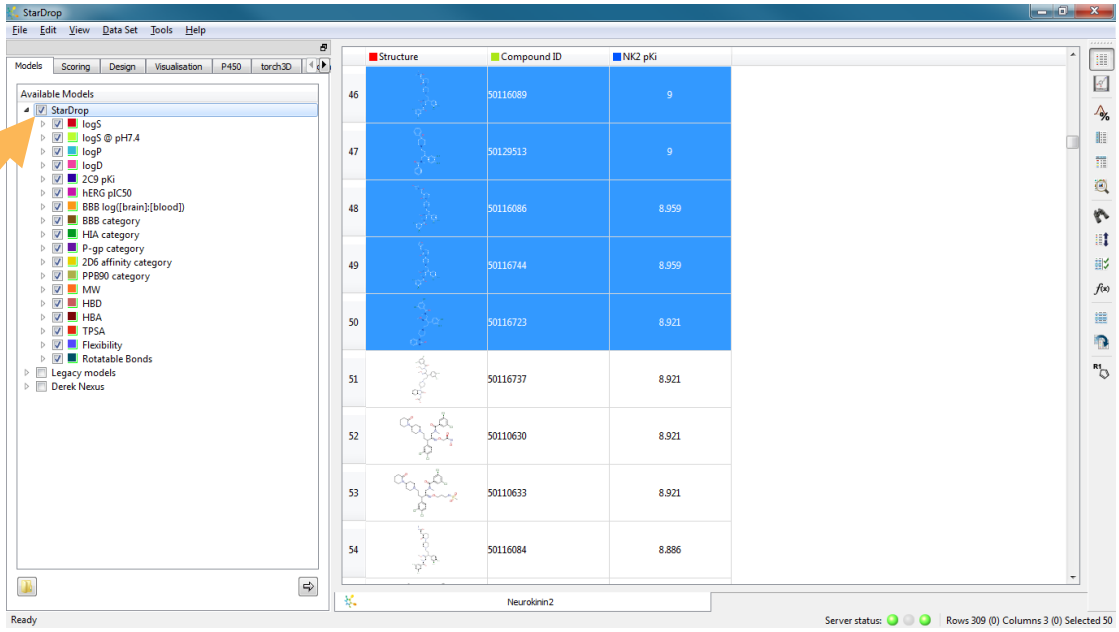
This may be the best region in which to focus for selecting compounds, but first we should consider the other properties that are important in a high quality lead.

- 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.

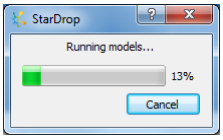
  If you wish, you can minimise this window to keep it out of the way for now.



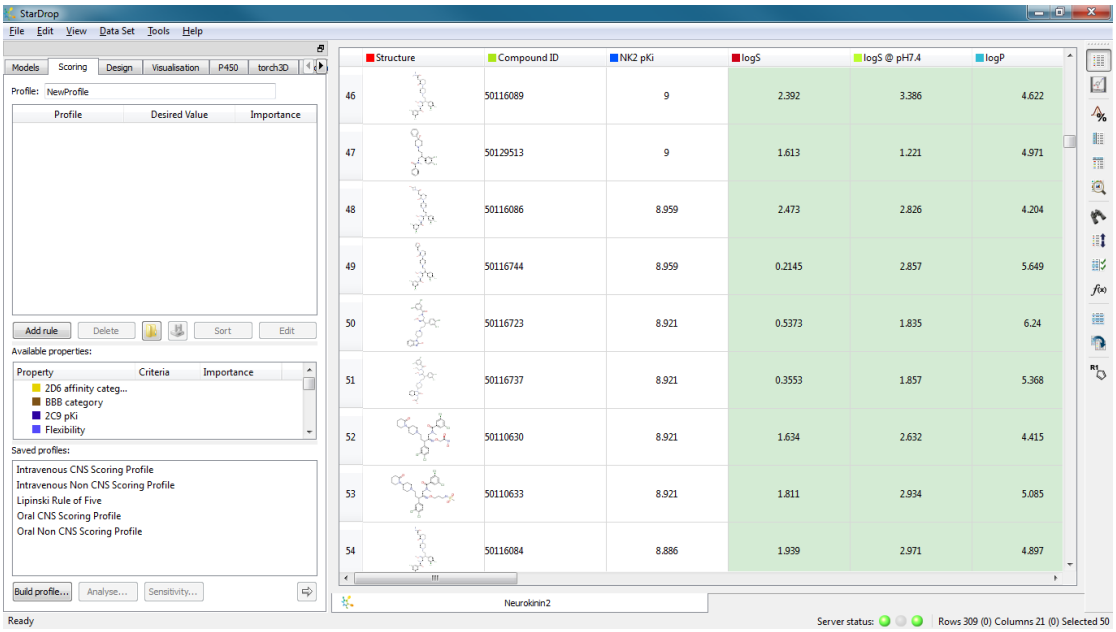
- We have no further measured data, so we're going to generate some predictions of ADME and physicochemical properties. Click on the **Models** tab.
- To run all of the StarDrop models, click the tick box next to the word StarDrop (this selects all of the models under this branch).



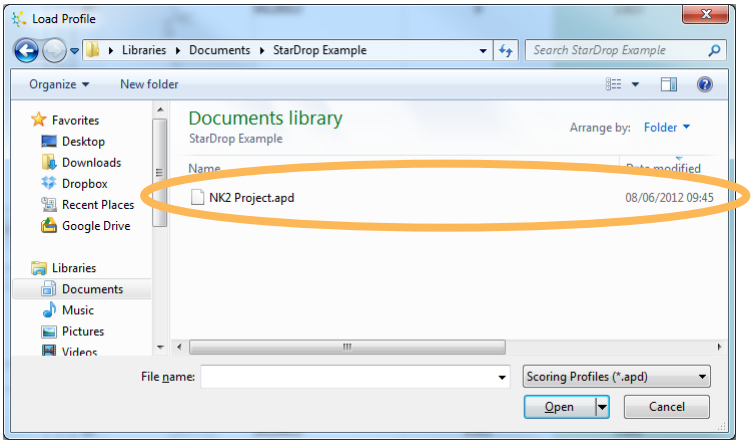
- Click the  button on the **Models** tab and a progress bar will be displayed while the predictions are calculated. When this process is complete you will see that a new column has been added to the data set for each property calculated.



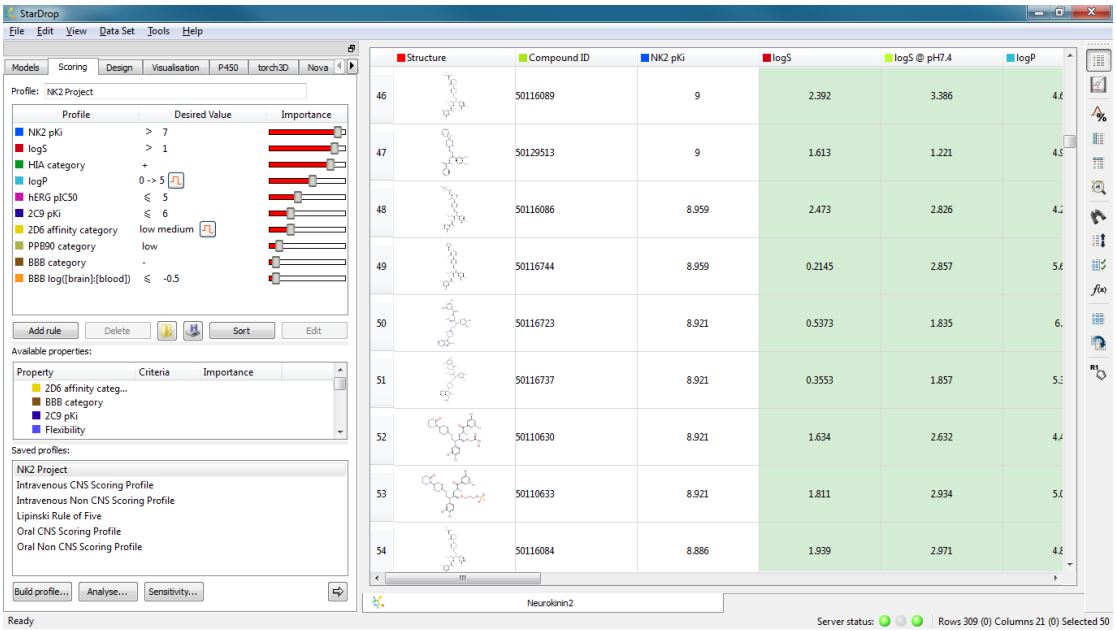
- Scroll across the data set to see the numerical and classification predictions that have been made. Due to the volume and complexity of the data, it is challenging to find which compounds have the best balance of properties. So, we're going to us StarDrop's scoring tool which will make it easy to assess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



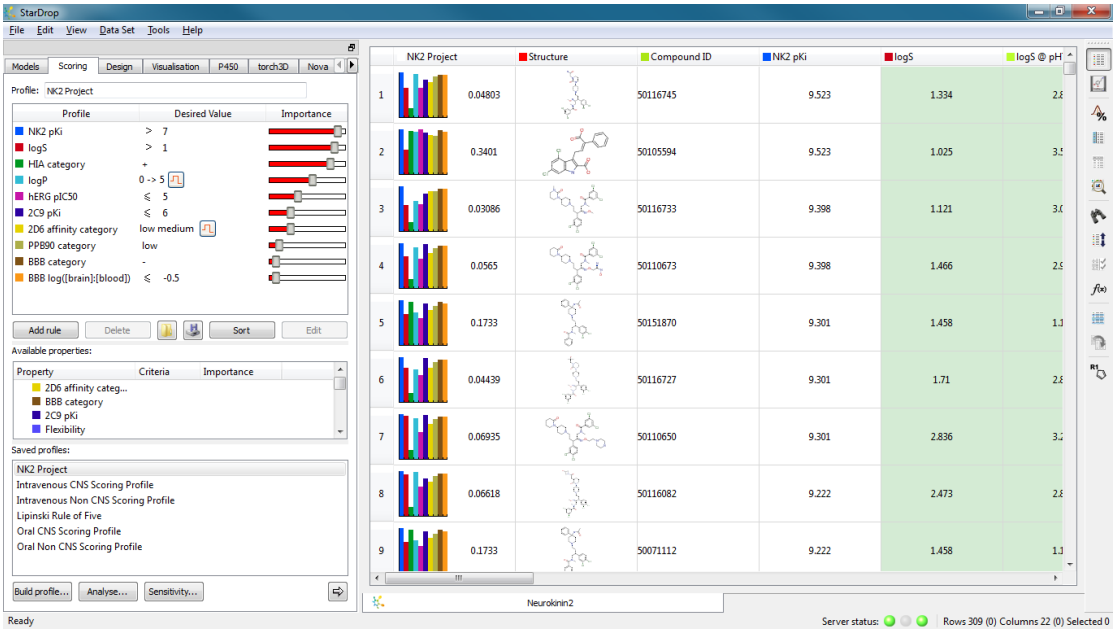
- A scoring profile allows you to define a set of criteria that is important for your project. StarDrop has some example profiles built in, but in this case we're going to load a profile designed for this Neurokinin 2 project.
- Click the  button to open a new scoring profile and load the profile called **NK2 Project.apd**.



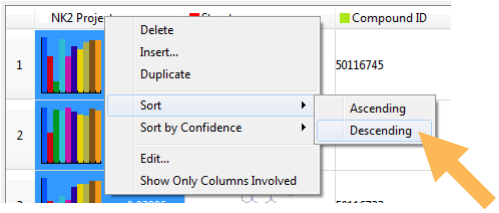
- A scoring profile contains a series of properties along with criteria describing desired values and their relative importance. In this profile we are looking for compounds with good affinity and which are suitable for a peripheral target.



- 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 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.



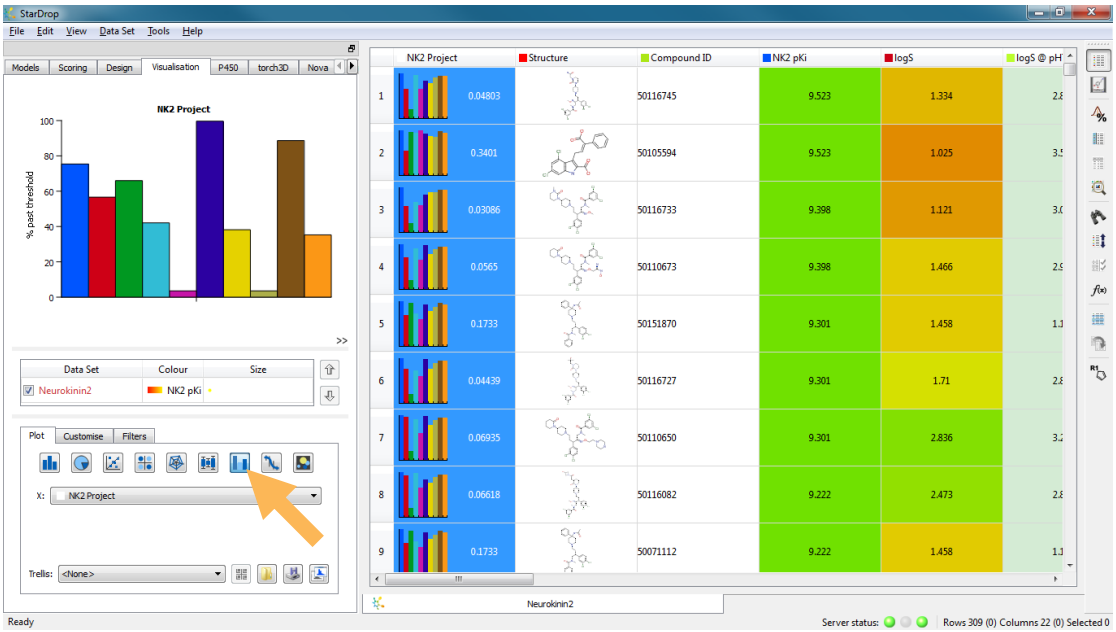
- To find the highest scoring compounds, right-click on the scoring column header and choose **Sort->Descending** to sort the data set from high to low score.



- 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.

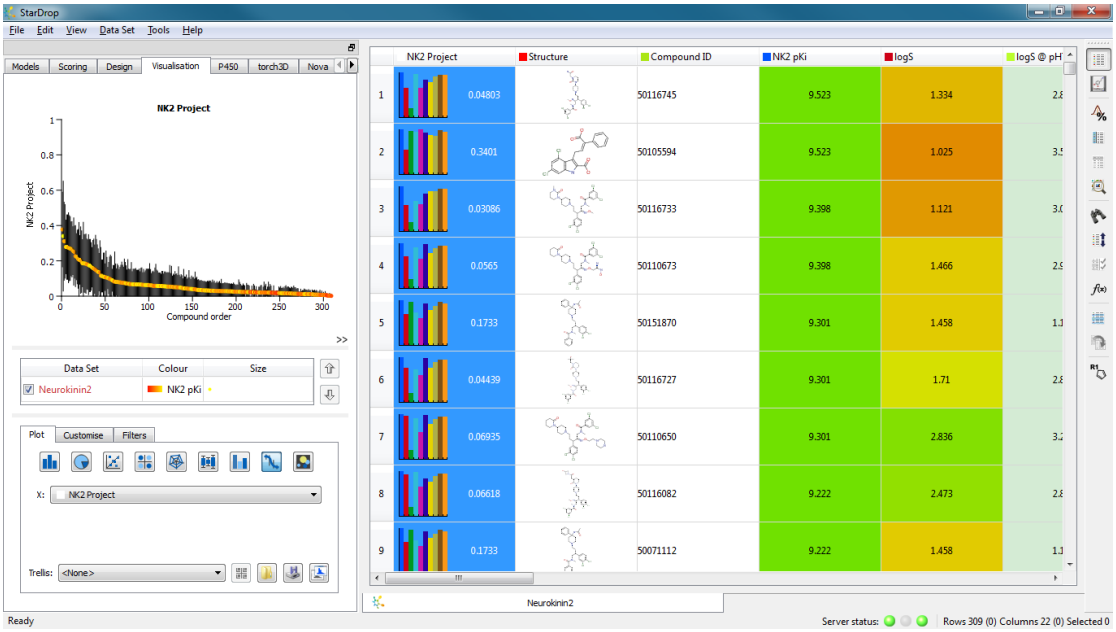
Having scored the compounds we're now going to take a look at the overall trends across the data set and consider which compounds we should select.

- Click on the **Visualisation** tab again.
- Select the data set column containing the scores (NK2 Project) and choose the  button as shown below. The resulting scoring histogram shows the profile of properties for the overall set; each bar represents one property in the profile, identified by the colour. The y-axis shows the percentage of compounds in the set that meet the ideal criterion for each property, as defined in the scoring profile.

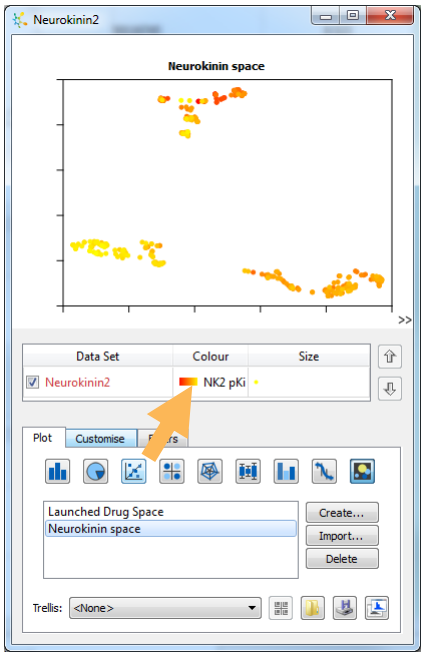


We can see that for some of the property criteria we have specified in our scoring profile, in particular hERG inhibition and Plasma Protein Binding, there are very few compounds which meet the criteria and hence these may be consistent issues for the chemistry in this data set.

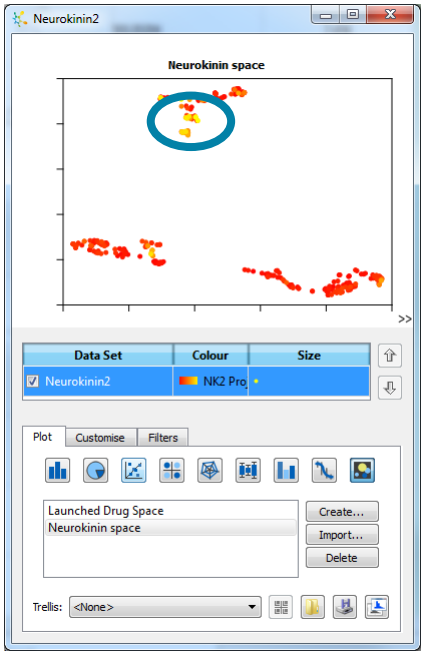
- Now select the  button on the **Visualisation** tab to change the plot type. A 'snake plot' shows the scores (on the y-axis) for all compounds in order from highest score to lowest score (along the x-axis). The overall uncertainty in each score, due to the uncertainty in the underlying data, is also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



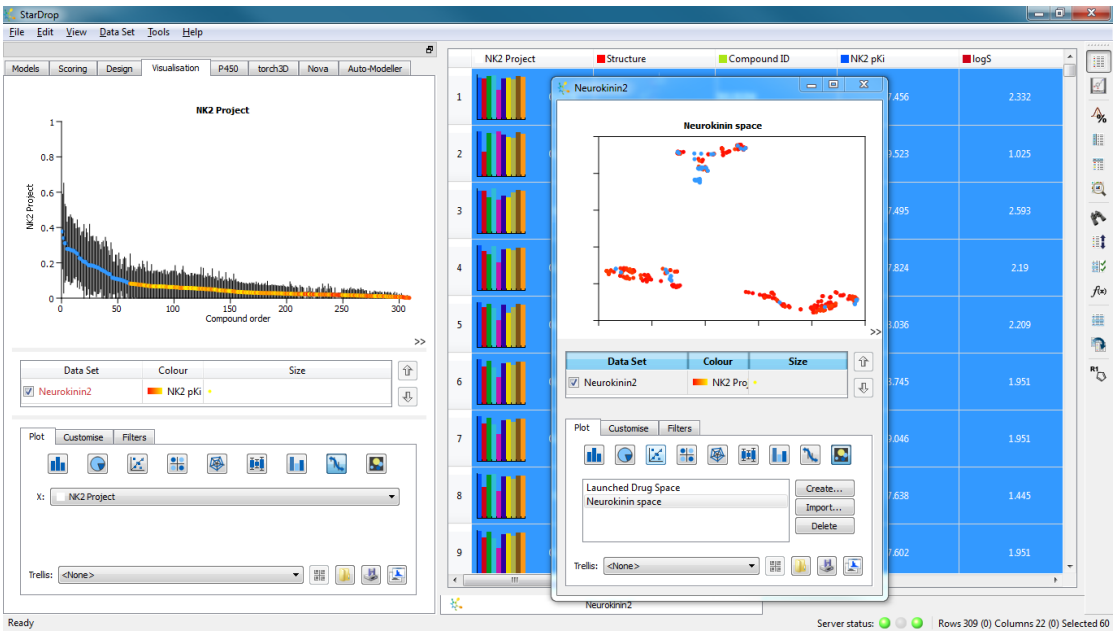
- Now open the detached chemical space plot we created previously.
- Click on the block of colour next to the 'Neurokinin2 library' in the key in this window to bring up the **Format by Property** dialogue again and, as we did before, change the **Source** property to be the **NK2 Project** values. This way the data 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. The cluster with the highest potency is now quite red, because it is unlikely to have acceptable ADME properties. However, we can identify 'hot spots' in yellow which are likely to achieve both good potency and ADME properties.



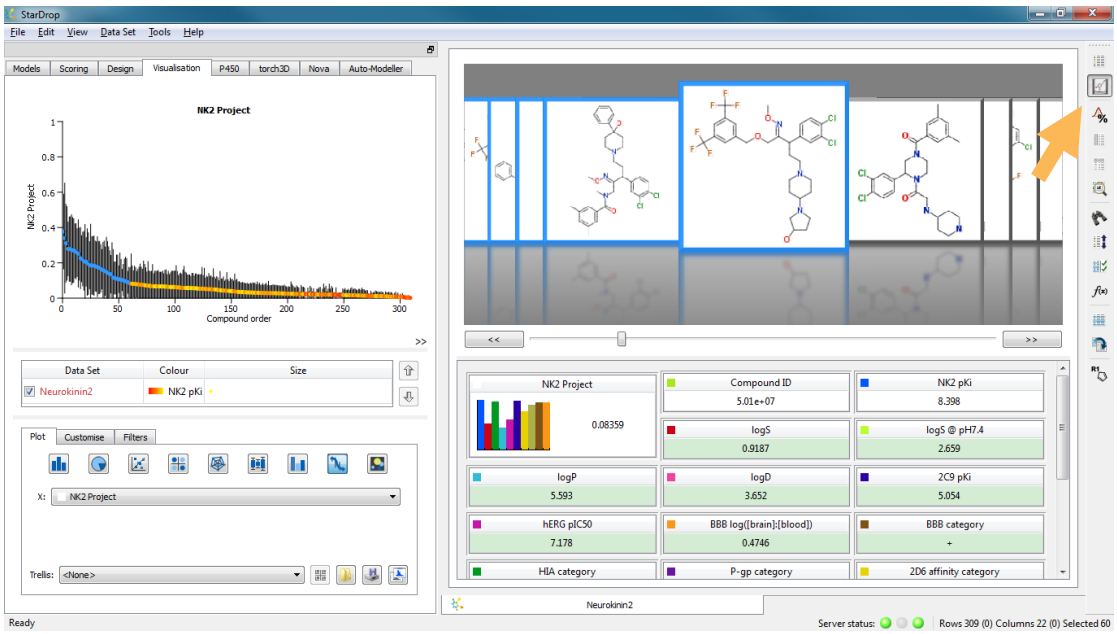
- Looking at the 'snake plot' we can see that the top 50 compounds cannot be confidently distinguished from the top scoring compound (notice that the error bar for the top-scoring compound overlaps with the error bars of approximately the top 50 compounds). Therefore, we should consider selecting compounds from this range to further explore their properties and make a confident selection of potential lead series. Selecting the top 50 compounds in the snake plot (by drawing a ring round them with the mouse) highlights where these lie within the chemical space.

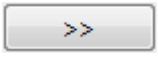
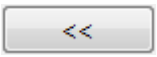


Notice that this selection includes compounds from different regions of the chemical space. This suggests that some of these chemistries cannot be rejected with confidence. Therefore, it may be more appropriate to sample some of these alternative chemistries to generate some experimental ADME data. These data will have greater confidence that predicted values and hence we will be able to identify with confidence the chemistries that will yield a high quality lead series.

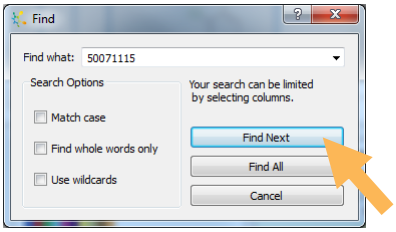
Having looked at the data on a project scale, the next question is whether there are any compounds that, with minor modification, could also achieve a good balance of properties.

- Close the window with the chemical space and click the  button on the right-hand tool-bar to switch into the molecule view.

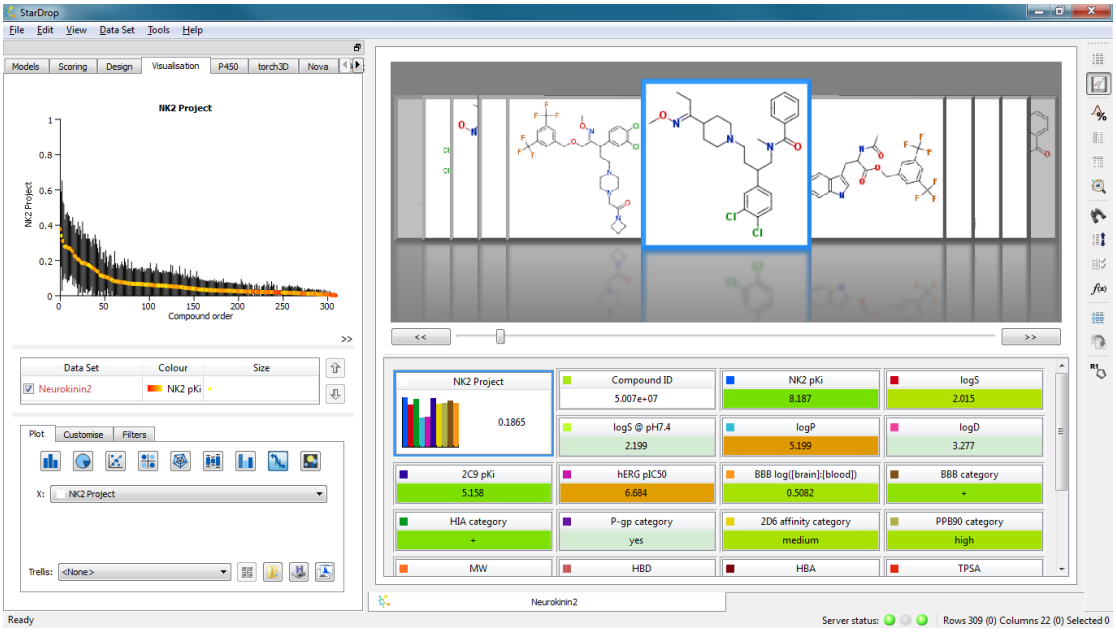


- When the cell containing the score is selected a heat map is displayed highlighting the properties included in the score. Those meeting the criteria with confidence are highlighted green and those that may pose a risk are coloured yellow through to red, taking into account the importance of the property and the uncertainty in the data. Scroll through the set by clicking on the molecules or using the  and  buttons.

- Select the cell titled **Compound ID** and click the  button on the toolbar to the right to bring up the **Find** dialogue.

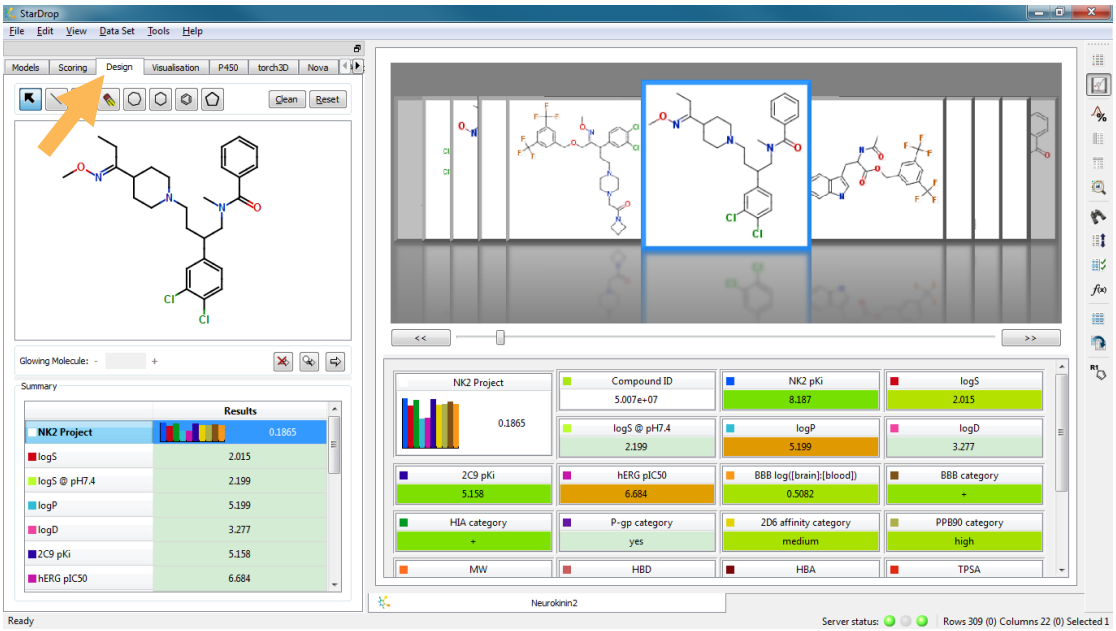


- Type in **50071115** and click **Find Next** (as shown above) to search for this compound. Click the **Cancel** button to close the **Find** dialogue and select the score cell again to see the heat map.

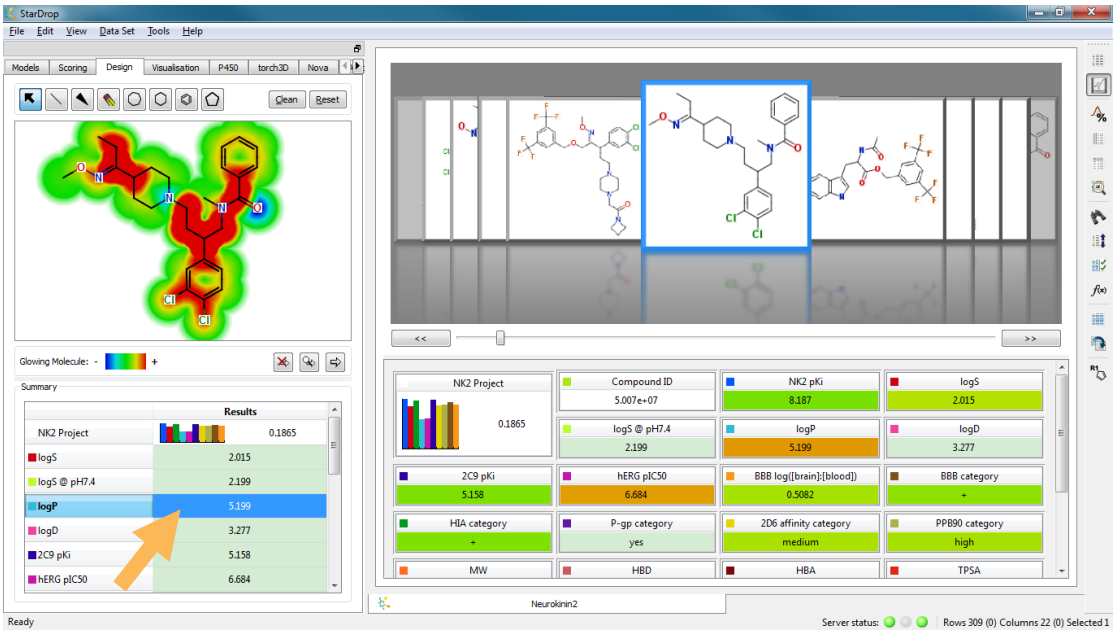


This compound has good affinity for the target and most of the predicted properties are acceptable. However, from the heatmap and score histogram we can see that the predictions suggest potential issues with logP (light blue bar) and hERG inhibition (pink bar). Therefore, this may be an opportunity to attempt to improve on the compound by making minor modifications.

- Switch to the **Design** tab and click on the compound structure in the Molecule View. (If necessary drag the bar in 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 table to make it wider).

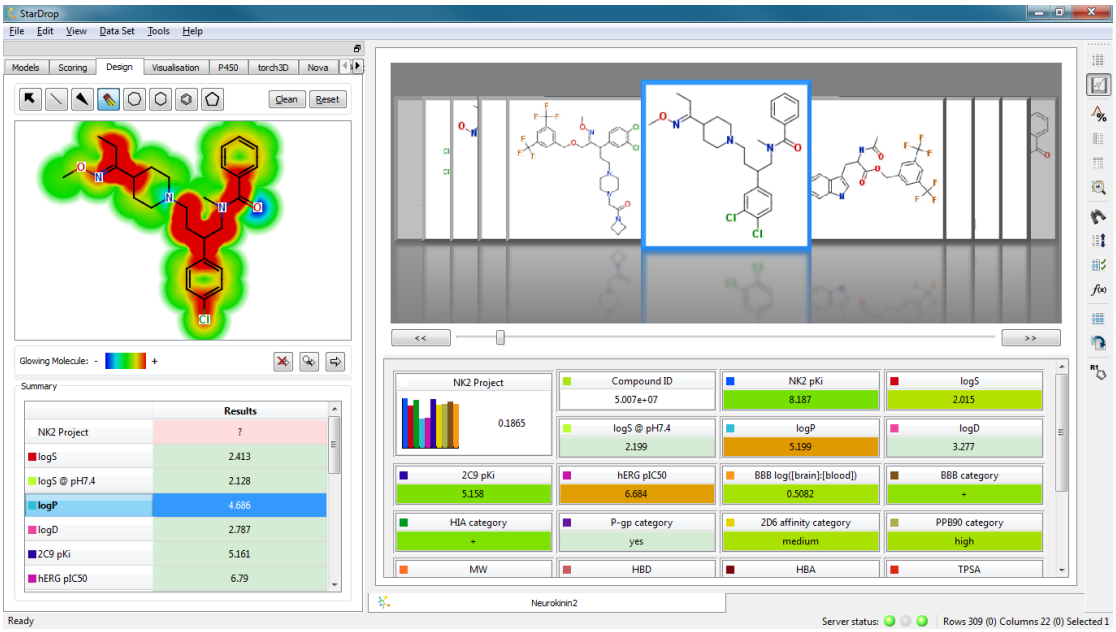


- Click on the logP value in the summary table below the Design window.



This highlights the molecule, using StarDrop's Glowing Molecule, to indicate which regions are having the greatest influence on the predicted value. Areas highlighted red tend to increase the property value and areas highlighted blue tend to decrease the property value. For logP we would like the value to be less than 5 so we should concentrate on making changes to the regions which are red to reduce the logP.

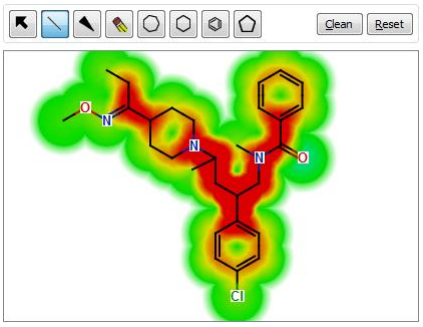
- To remove one of the chloro substitutions on the phenyl ring, click the eraser button  at the top of the design window and then click on the chlorine to remove it.



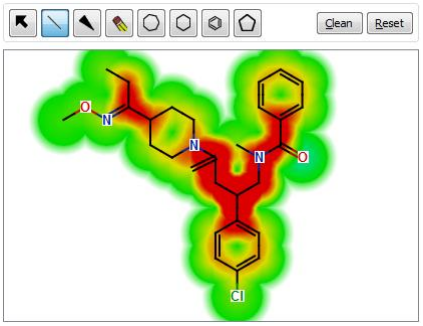
- In the summary table below, notice that the predictions for this new molecule are updated instantly. Importantly the logP is now lower.
- Scroll down the summary table and select the hERG prediction to see the Glowing Molecule for this property. We would ideally like the hERG pIC_{50} to be under 6.

Here we can see that the basic Nitrogen on the piperidine is having a significant effect to increase the hERG affinity. We will change this amine to an amide and explore the effect this has.

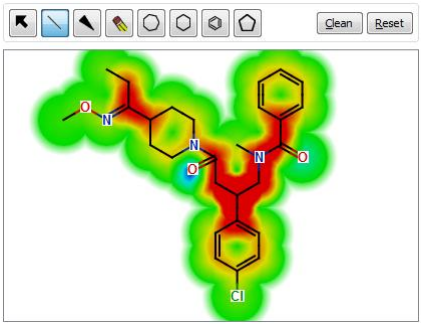
- Select the bond tool  and click on the carbon chain atom next to the amine and while the mouse is clicked, drag away and then release to draw a methyl group. Make sure you don't connect this to any other atom! (If you do, hover the mouse pointer over the drawing and press ctrl-Z to undo it).



- 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 pIC50 and that these changes have not-only reduced the predicted hERG affinity, but have also reduced the logP value further.

If you scroll up the summary table to the top again you will see a question mark against the score. This is because we do not yet have an experimental pK_i 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.

This example illustrated how we are able to explore the chemical space of our project to look for compounds which have a good overall balance of properties. 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.