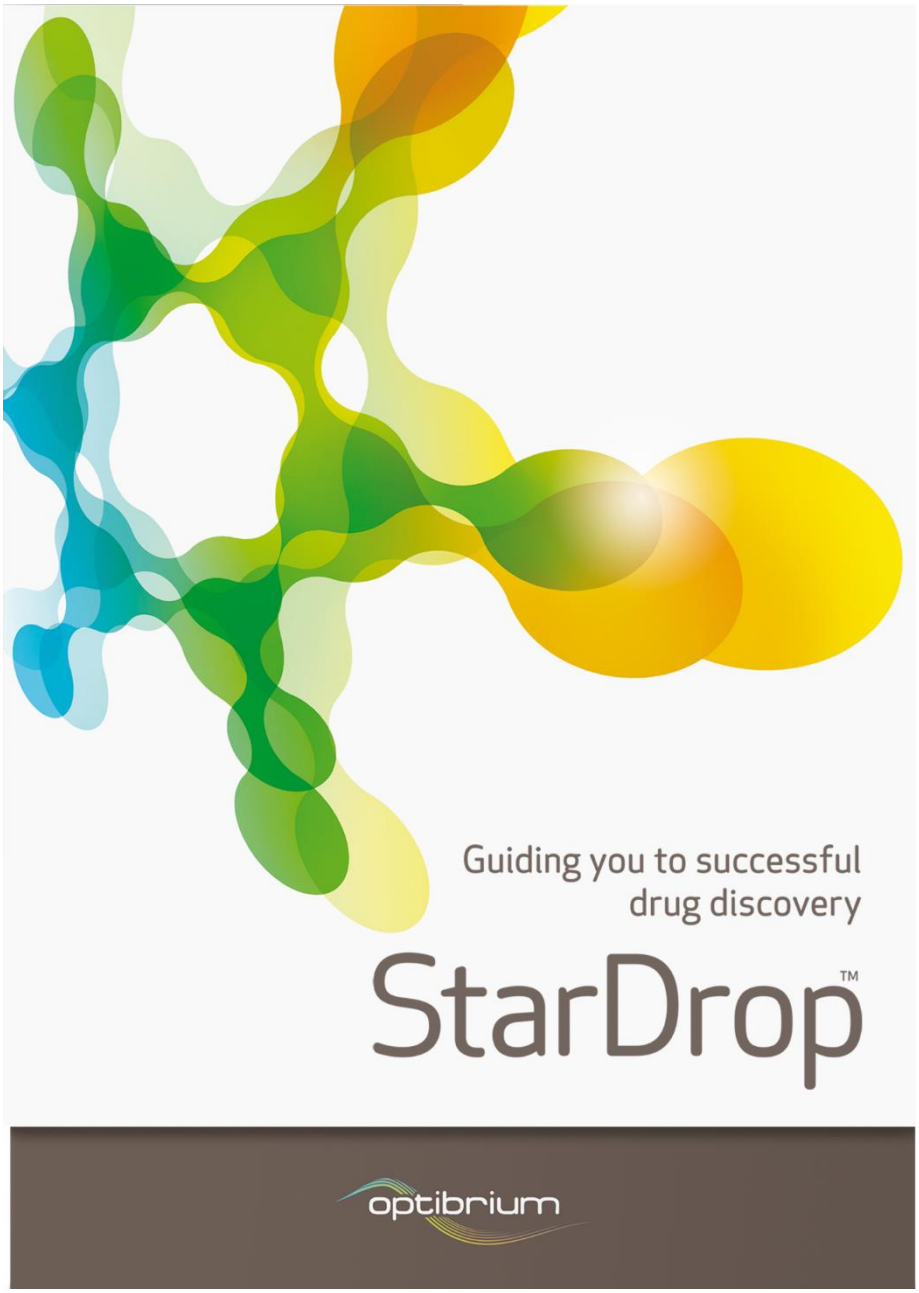
StarDrop™ Training Manual

Version 6.6



Contents

Introduction to StarDrop: Getting Started… 3

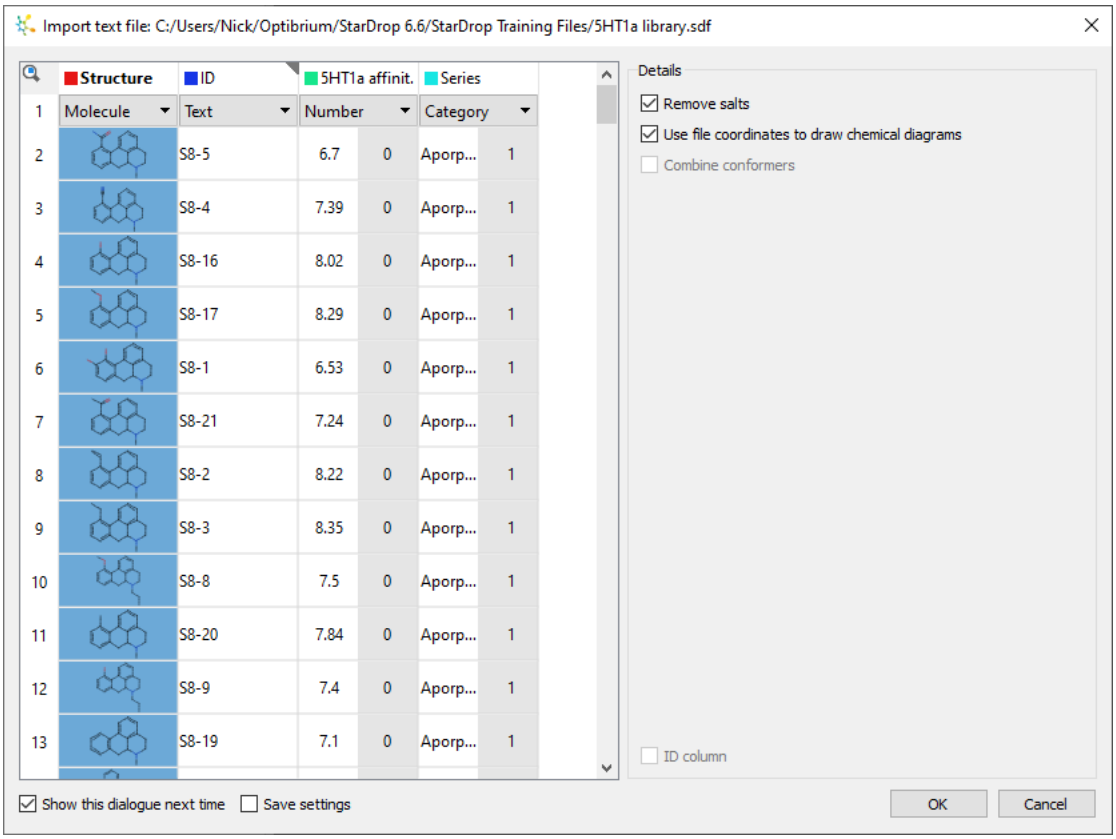
Hit-to-Lead: Targeting High Quality Lead Series..... 14

Lead Optimisation: Guiding the Design of Balanced Compounds 19

Introduction to StarDrop: Getting Started…

In drug discovery, we are faced with many challenges as we look for compounds that have the right balance of properties in order for them to become successful drugs. In the early hit-to-lead stages we often wish to explore a wide range of chemistry in order to find those areas of chemical space which contain lead series with the greatest potential. As we move into lead optimisation we need to focus more closely to learn about the structure-activity relationships (SAR) and design new ideas around specific scaffolds. In this first section, we will become familiar with the StarDrop interface and touch on a number of these themes.

- From StarDrop's **File menu**, choose **Open** and open the file **5HT1a library.sdf**.

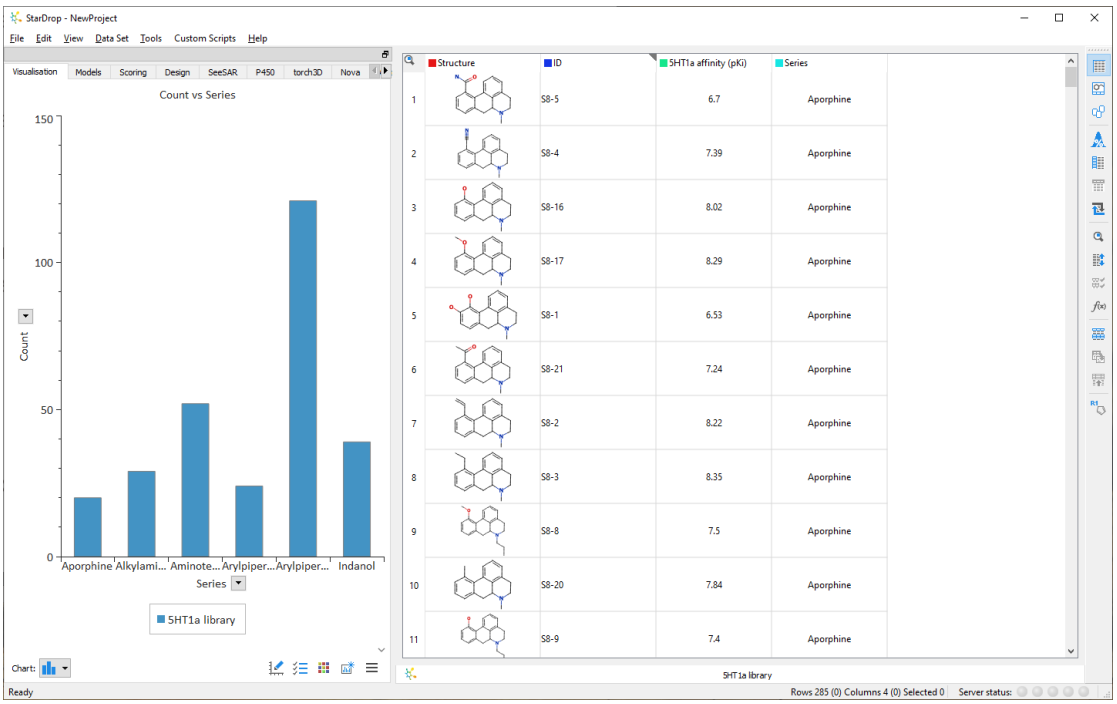


When opening text files (*.txt), Microsoft Excel® spreadsheets saved as comma-separated variable files (*.csv) or SD files, StarDrop will show an **Import text file dialogue** enabling you to confirm the columns of data along with additional details (e.g. units).

Microsoft Excel is a registered trademark of Microsoft Corporation in the United States and/or other countries.

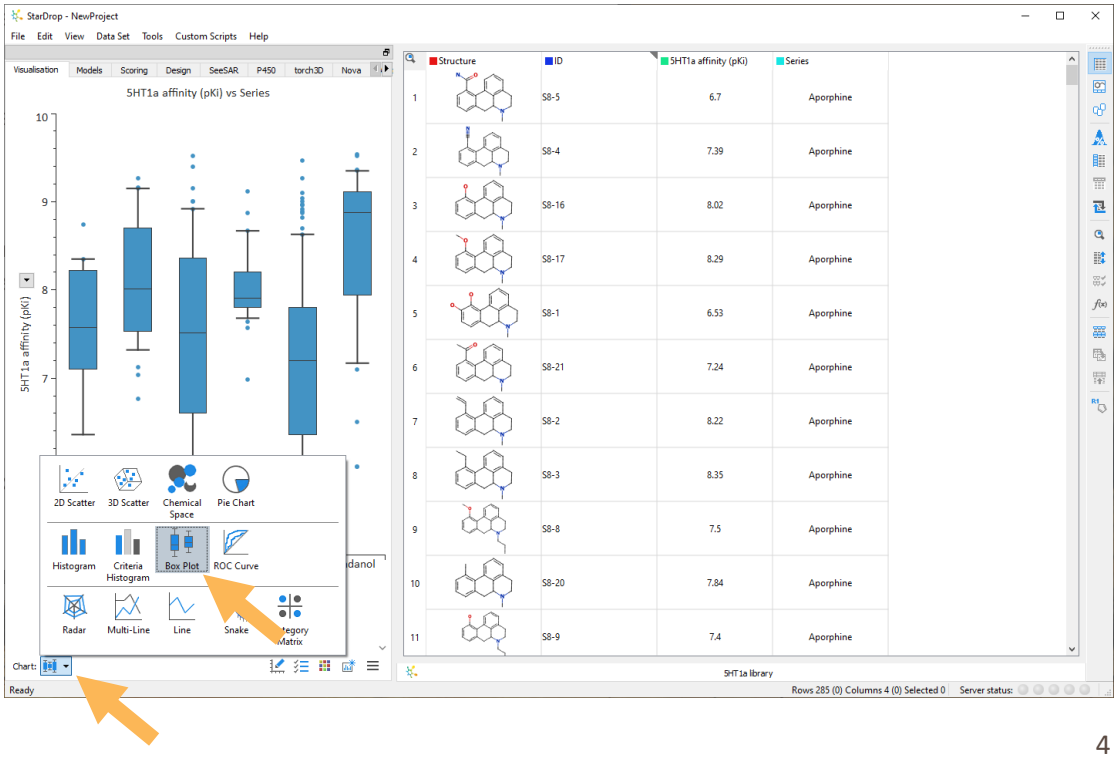
This SD file contains 285 compounds across six chemical series; for each compound there is a structure, ID, measured affinity value (pK_i) against 5HT1a and a category indicating the series to which it belongs.

- Click the **OK** button.



All the structures and data from the SD file appear in a table and a histogram will be displayed showing the number of compounds in each of the series.

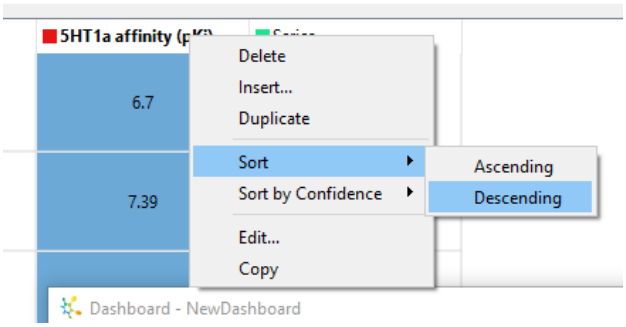
- Click on the **Chart button** and select **Box Plot**, so that we can see the distribution of potency values for each series.



We can see from this that, while all the series have some potent compounds, the Indanol series has the highest average affinity.

- Click the **Detach button**  at the bottom of the Visualisation area to add this chart to a new dashboard.

- Right-click on the **5HT1a affinity (pKi)** column to bring up the menu and choose **Sort** then **Descending**, to bring the compounds with the highest target affinities to the top.



We can see that the most potent compound, S5-34, comes from the Indanol series.

Note: selecting columns is an alternative way to create charts and, as a result, we now have a histogram of the selected 5HT1a affinity (pKi) column in the Visualisation area. Charts in the dashboard will respond to row selections made but will otherwise remain unchanged.

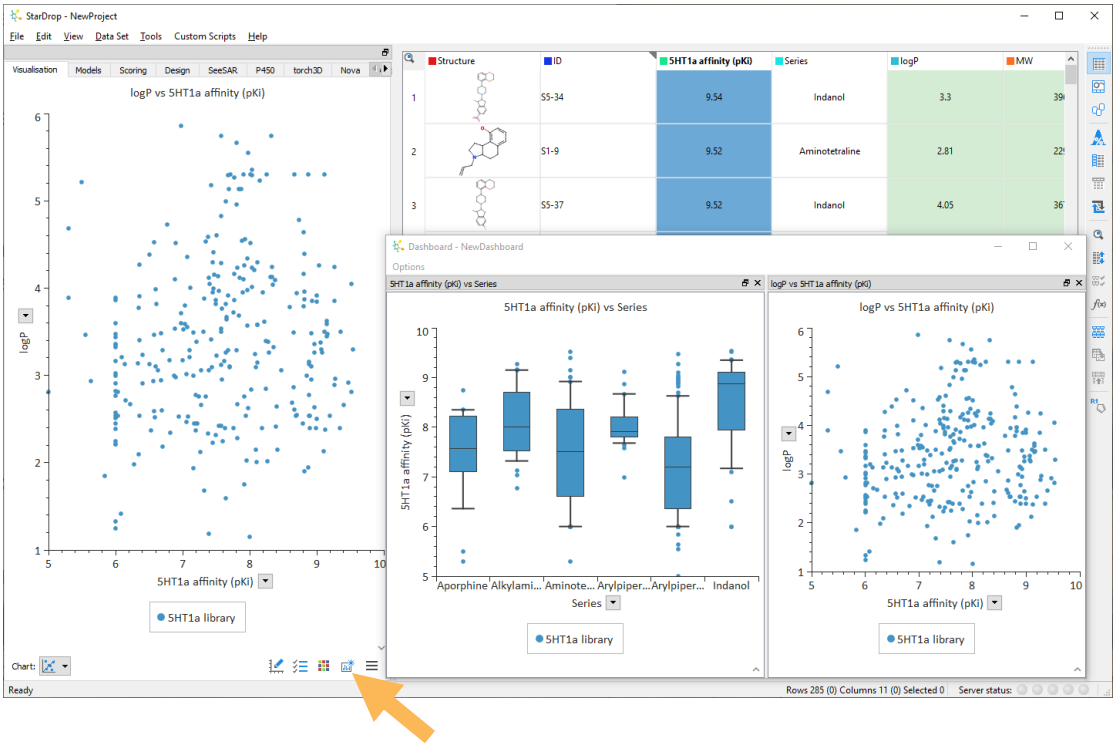
We'd next like to consider the relationship between potency and simple 'drug like' properties of the compounds in this library.

- Click on the **Models tab** to change to the Models area and select **logP**, **MW**, **HBD**, **HBA**, **TPSA, Flexibility** and **Rotatable Bonds** from the list of properties.

- Calculate these properties for all of the compounds in the data set by clicking the  button (this will add a new column to the data set for each property calculated).

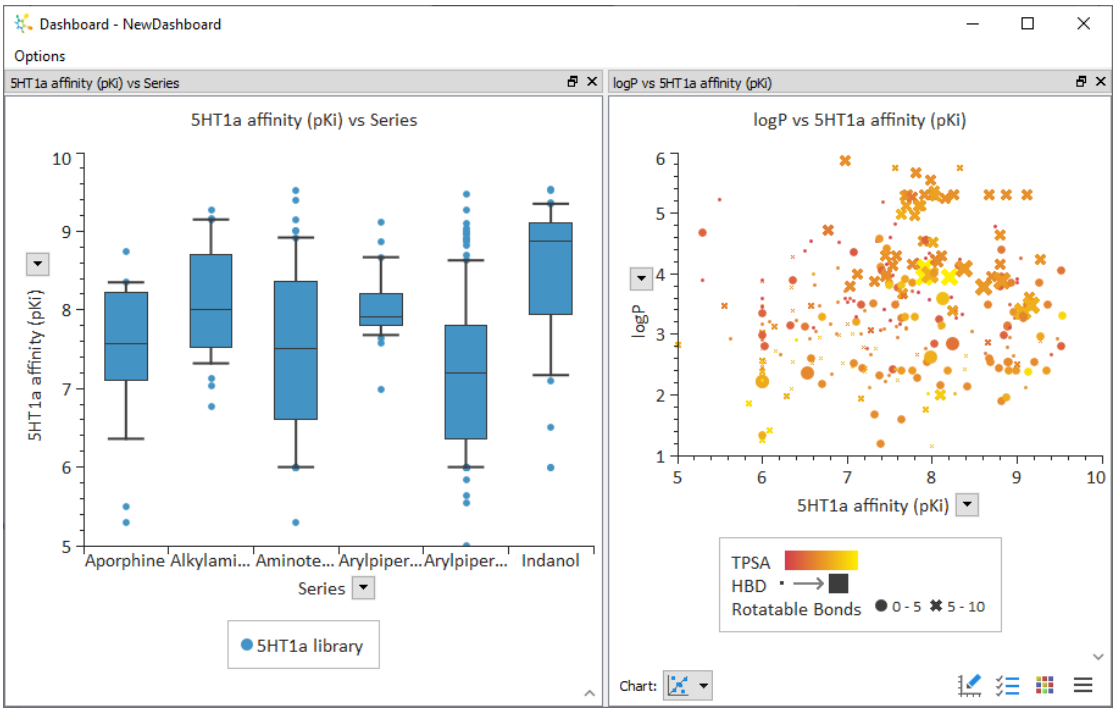
- Change back to the **Visualisation area**, click on the **Chart button** and select **2D Scatter**, with **5HT1a affinity (pKi)** for the x-axis and **logP** for the y-axis.

- Click the **Detach button**  to add the chart to the dashboard (**Note:** you may wish to make the dashboard wider).



- Click on the arrow in the bottom-right corner of the new chart in the dashboard to reveal the controls and click the **Format button**  .

- Colour the points by **TPSA**, size them by **HBD**, set symbols based on **Rotatable Bonds** and then close the **Format Chart Data dialogue**.



Note: the chart and its legend update as you change the formatting options.

- Click on the **Chart Menu button** ☰ and choose **Filter** to display the **Chart Filters dialogue**.
- Click the **Add** button to add a filter based on **MW**, set the upper limit of the range to be **300** and check the **Apply to all** option.

Note: You can invert the range by clicking on it and shift the entire range by dragging it.

From the box plot we can see that only compounds from the Aporphine and Aminotetraline series have not been filtered. Let's now take a closer look at the ranges of properties across the different chemical series in this data set.

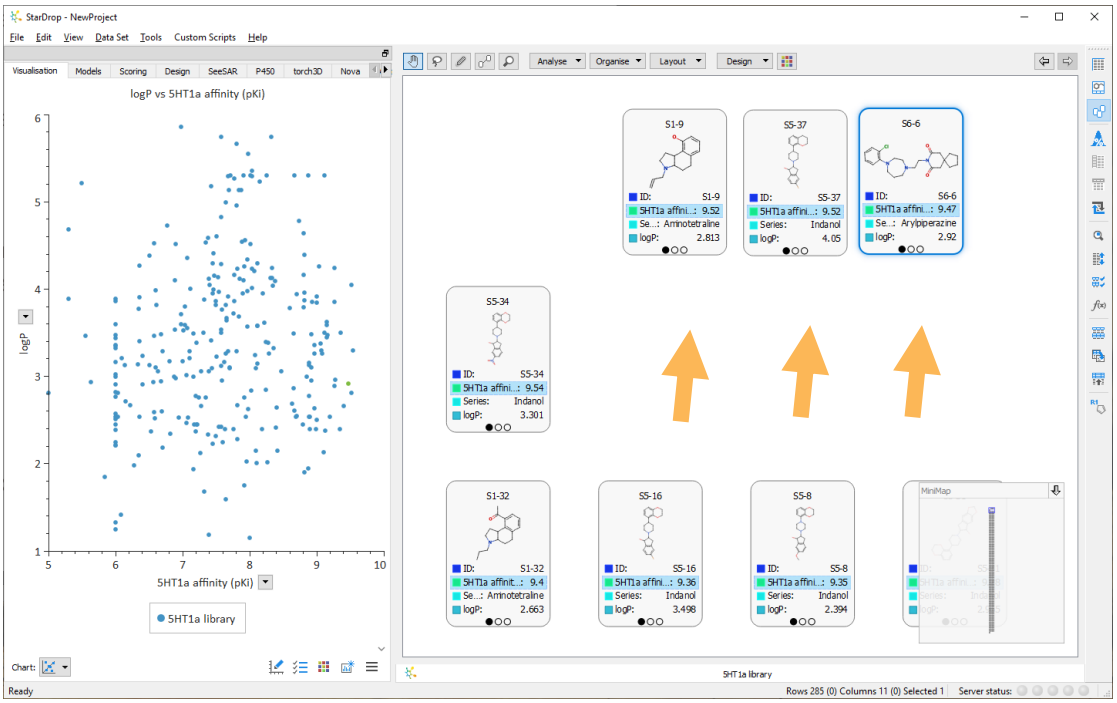
- Clear the filters by clicking the **Clear button**, close the **Chart Filters dialogue** and then minimise the dashboard.
- Click the **Card View button** on the right-hand toolbar to switch into Card View.

The cards already show the original data that we loaded from the SD file, but we can alter the design to add some of the properties we have just calculated.

- Click the **Design menu** at the top of Card View and choose **All Properties**.

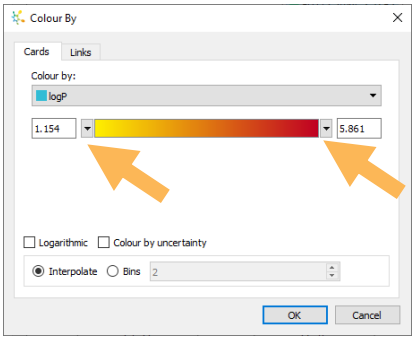
We can easily compare representative, potent compounds from three different series.

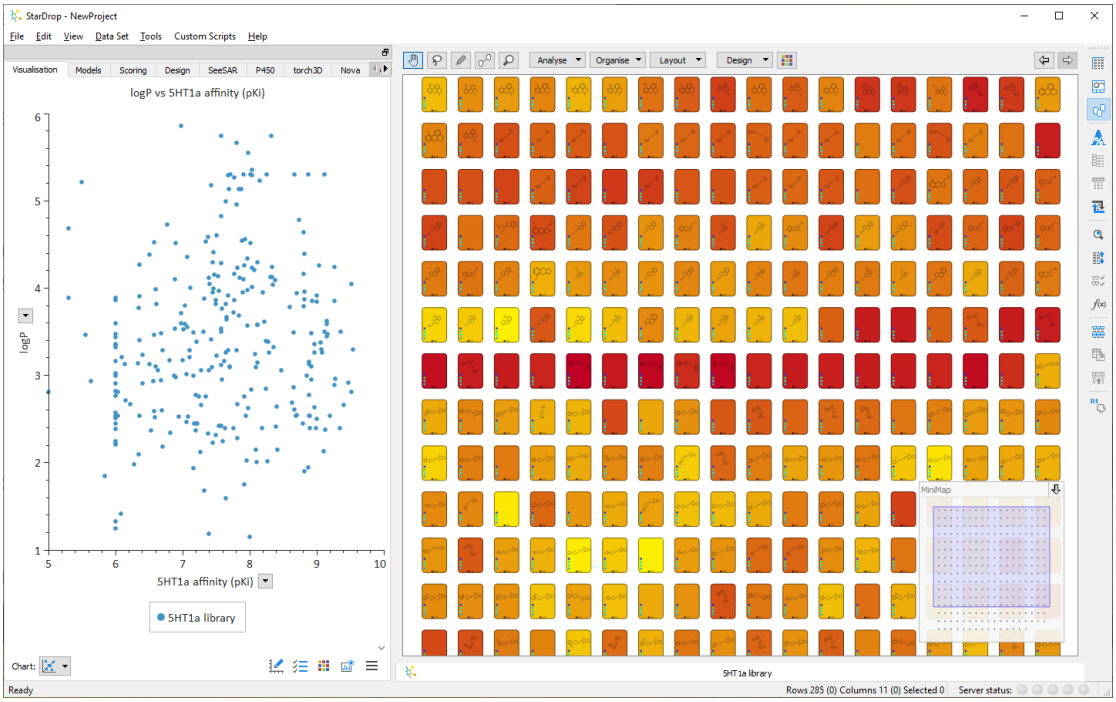
- Move the Card View table top down by clicking in the empty space between cards and dragging down. Move the 2nd, 3rd and 4th cards up into the space above, next to each other, and zoom in using the mouse-wheel or the **Ctrl** and **=** keys.



Let's look at the distribution of logP across the cards representing the compounds in the data set.

- Zoom out using the mouse wheel or **Ctrl** and **–** keys. Then reset the view to a grid, by selecting **Grid** from the **Layout menu** at the top of Card View.
- Change back to the first page on the cards and click the **Format button**  at the top of Card View.
- **Colour By** the property **logP**
- Change the default colour scale to go from yellow (low) to red (high), by clicking on the drop-downs at each end of the colour range, and click **OK**.

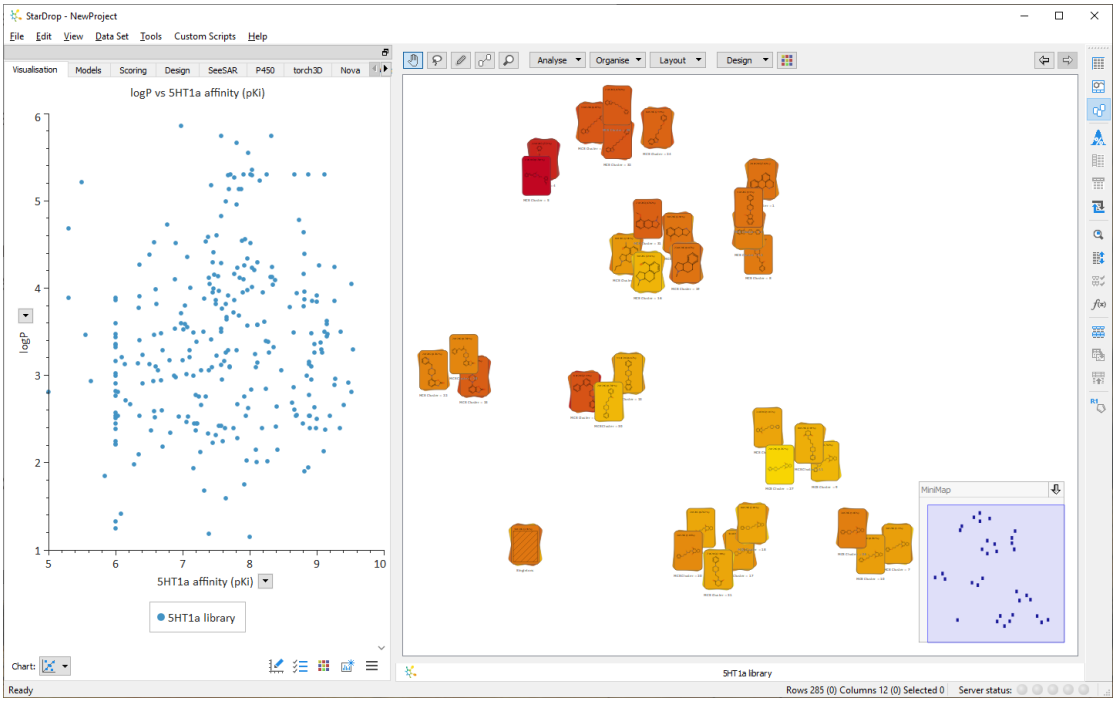




The data set is already sorted with the most potent compounds at the top so yellower cards nearer the top represent compounds with lower logP values and higher 5HT1a pK_i values.

So far we have been considering compounds individually, but at this stage of a project it is common to look for series of compounds with good properties.

- From the **Analyse menu** choose **Clustering**.
- **Common Substructure** is already selected so click the **Cluster button**.

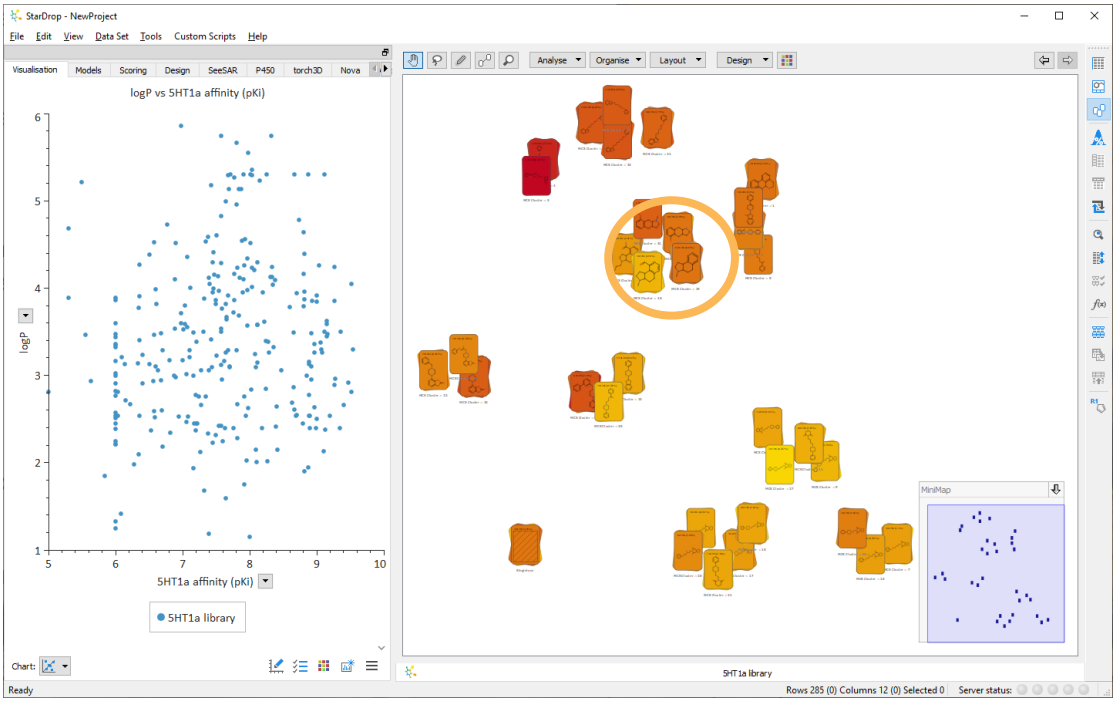


- Right-click on a stack and choose **Inspect** from the menu.

Note: This enables us to browse the cards in this stack and, if necessary, remove cards by dragging them out of the window.

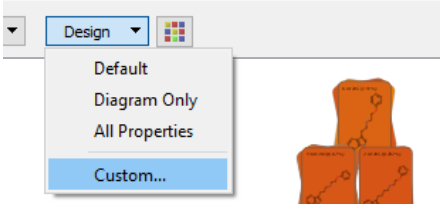
We can apply our expertise to refine the clustering results and better represent the series in the data set by combining the three stacks which contain Aminotetralines (highlighted below).

- Drag each of these stacks on top of one another to combine them.

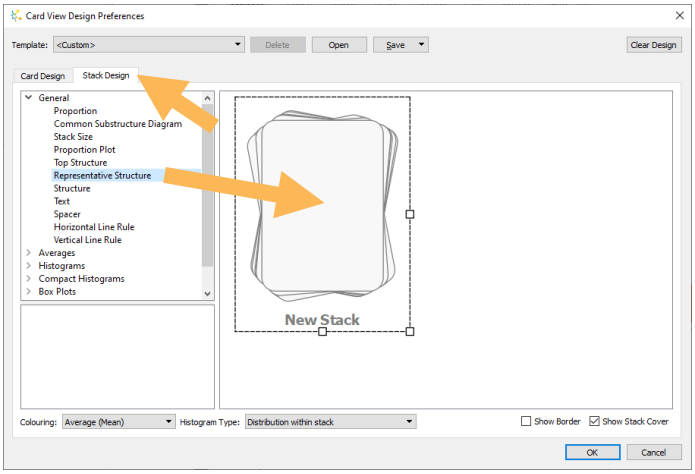


The newly combined stack will show the common sub-structure of all the compounds in the three stacks that we have combined.

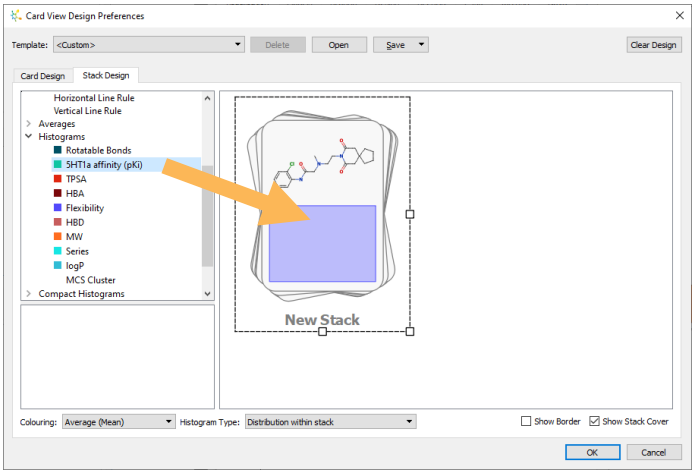
We can change the design of both cards and stacks to show representative information from the data set.



- From the **Design menu** at the top of Card View select **Custom**.

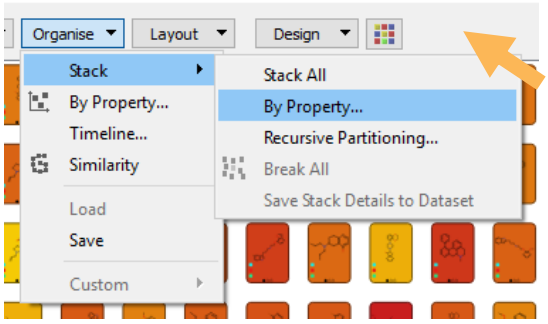


- Click on the **Stack Design tab**, remove the current properties by dragging them from the Stack.
- Expand the **General section** and from it drag **Representative Structure** onto the stack design.

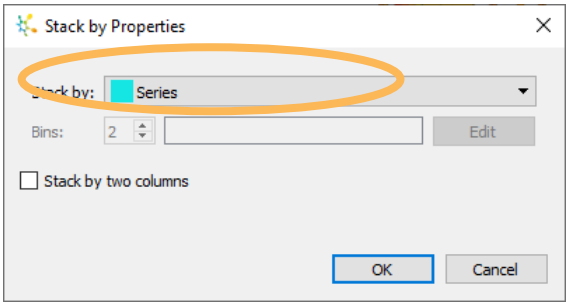


- From the list of **Histograms**, drag **5HT1a affinity (pKi)** below the structure.
- Click **OK**.

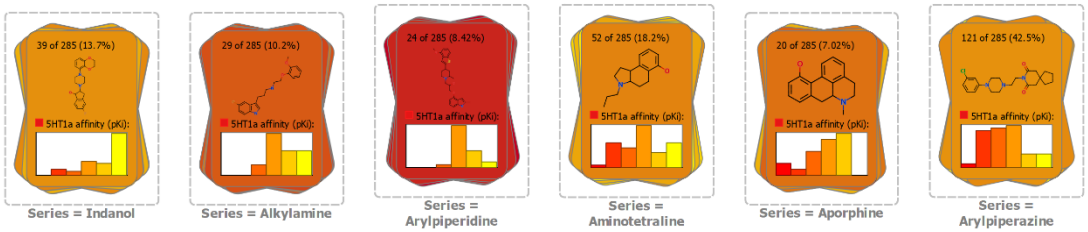
We do not need to refine all the clusters in this data set because the **Series** column already contains this information and we can use this to compare the properties of each series.



- From the **Organise menu** choose **Stack** and then **By Property**.



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



- Compare the stacks with the dashboard, drawing a lasso around the points in the bottom-right corner of the chart where there are potent compounds with low logP values.

From the stacks we can see that it is only the Arylpiperidine and Alkylamine series which have no potent compounds with low logP values.

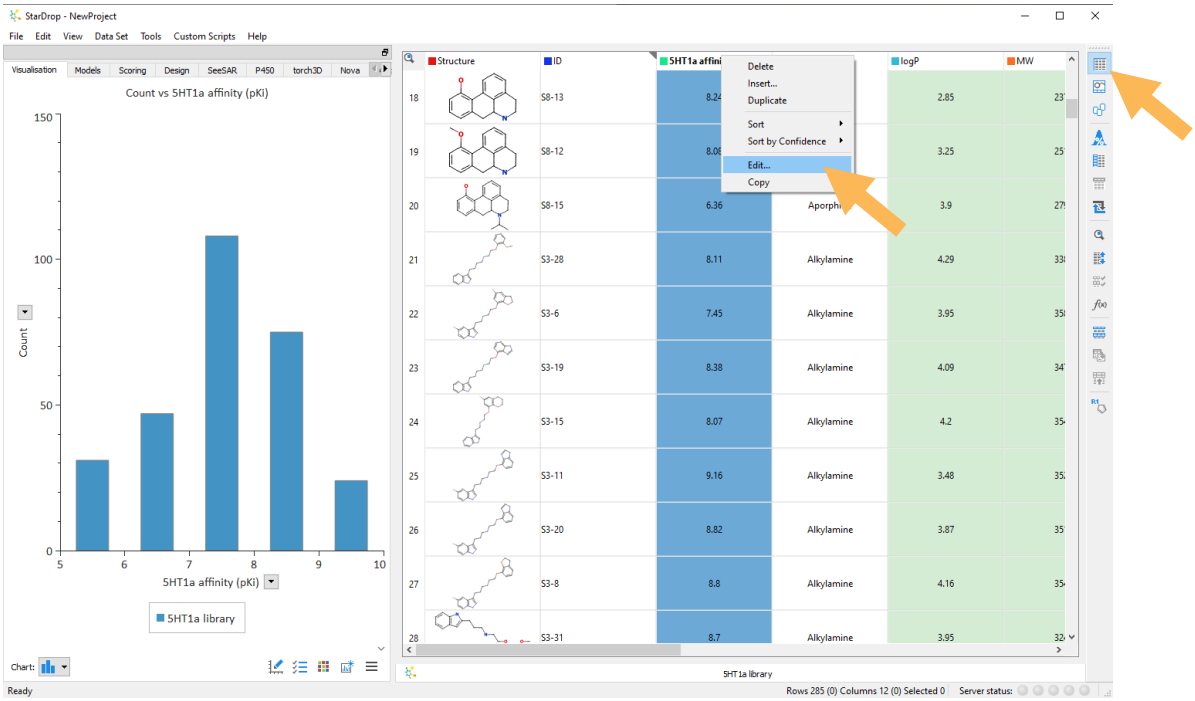
- From the **File menu**, select **Save Project** and choose a name for the project file

Note: If you now close this project and reopen it again all the analyses and visualisations will be restored.

Hit-to-Lead: Targeting High Quality Lead Series

In hit-to-lead, it is important to quickly focus resources on the chemistries that are most likely to yield a high-quality lead series. In this example, we will explore how data from primary screening of a library for potency against the target 5HT1a can be combined with predictions for a range of ADME and physicochemical properties to identify chemistries with a good balance of properties. At the same time, given the uncertainty in the underlying data due to experimental variability and statistical error, it is important that we do not reject compounds inappropriately and risk missing valuable opportunities.

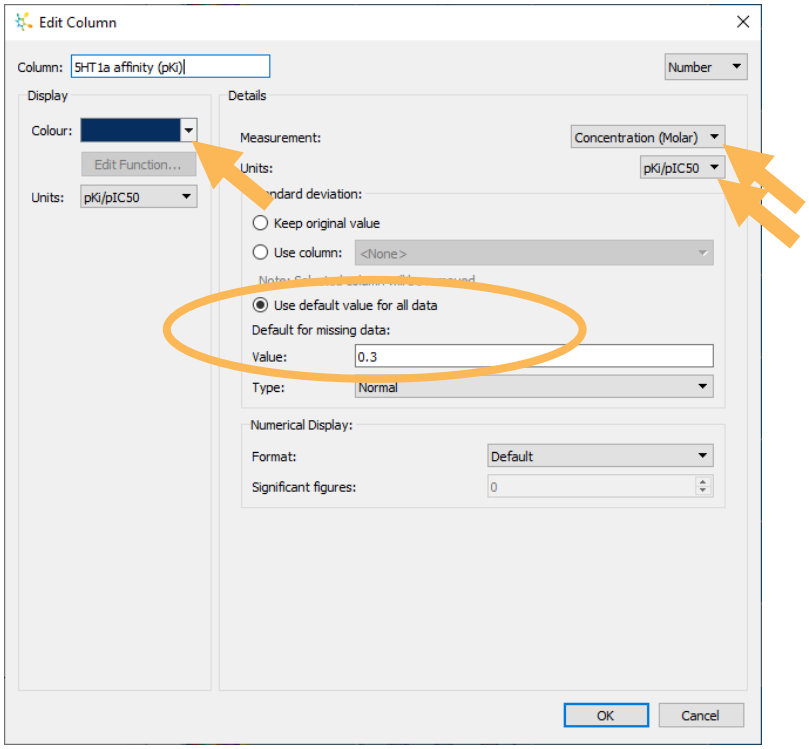
- Clear the selection (click somewhere in a chart where there are no points).
- Minimise the dashboard and switch to Table View by clicking the **Table View button**  on the right-hand toolbar.
- Right-click on the header of the **5HT1a affinity (pKi)** column and from the menu choose **Edit**.



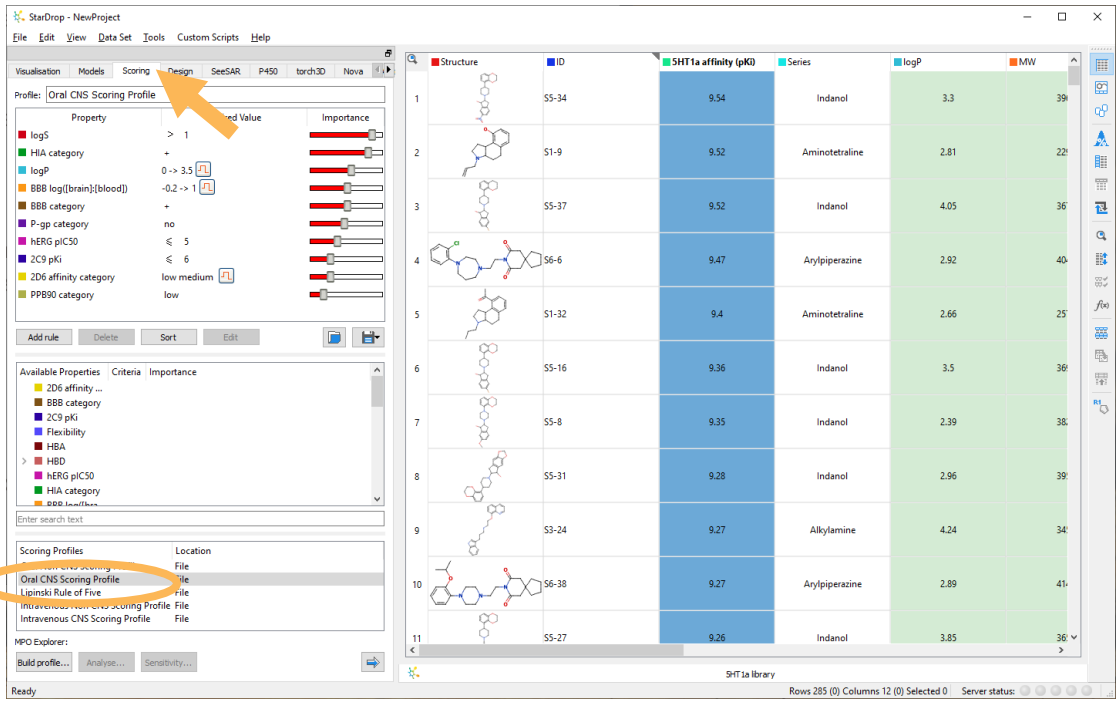
The data we imported contained no information about the errors associated with the measured potency values and so we will estimate that the standard deviation is 0.3 log units (equivalent to approximately a 2-fold error in the K_i values).

- In the Edit column dialogue, change the **Measurement** and **Units** to **Concentration (Molar)** and **pKi/pIC50** respectively.

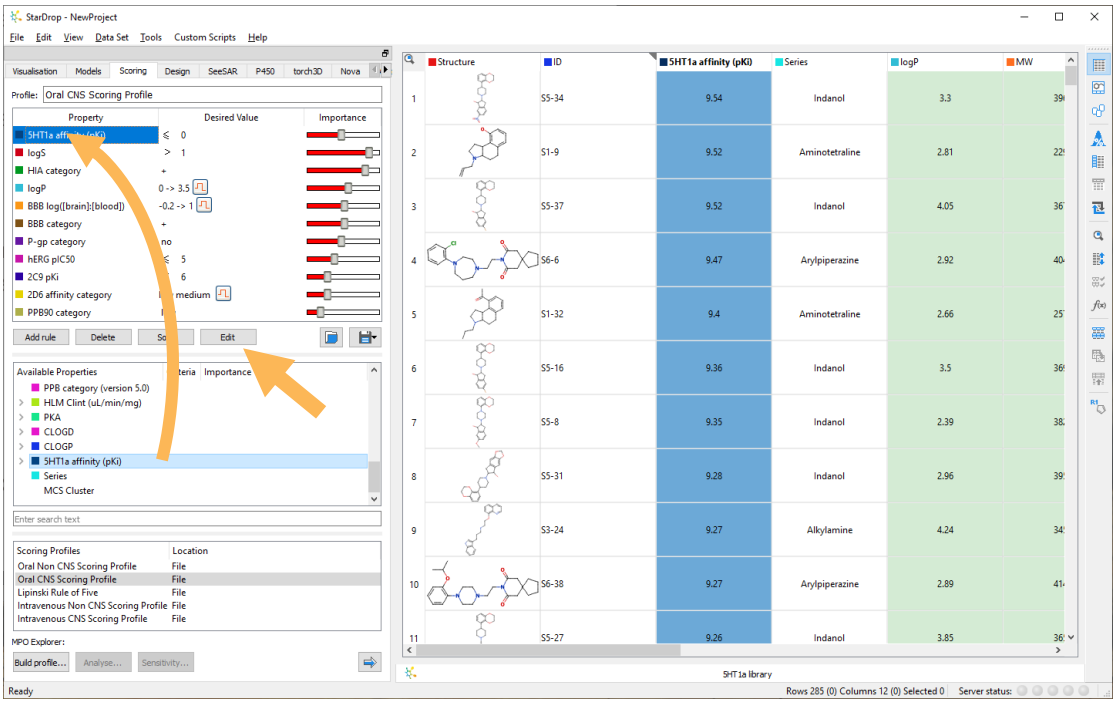
- Set the default standard deviation to be **0.3**.



- Finally, change the **Colour** to dark blue and click **OK**.
- In the **Models area** select all the StarDrop QSAR Models and click the  button to run them.
- Change to the **Scoring area** and select the **Oral CNS Scoring Profile** from the list.

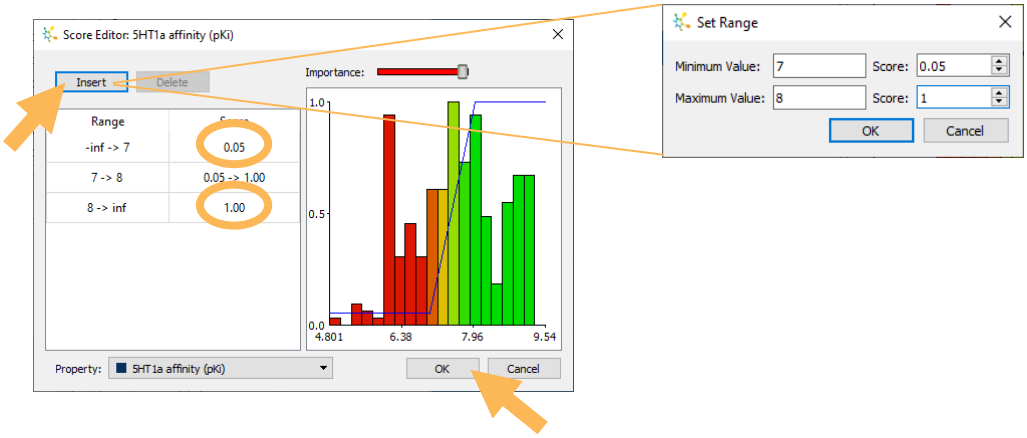


- Add **5HT1a affinity (pK$_i$)** from the list of **Available properties** to the scoring profile by dragging the property into the profile editor.



A simple cut-off criterion can be easily set by editing the **Desired Value** and **Importance** in the profile itself, but in this case, we will set a more sophisticated trend scoring function.

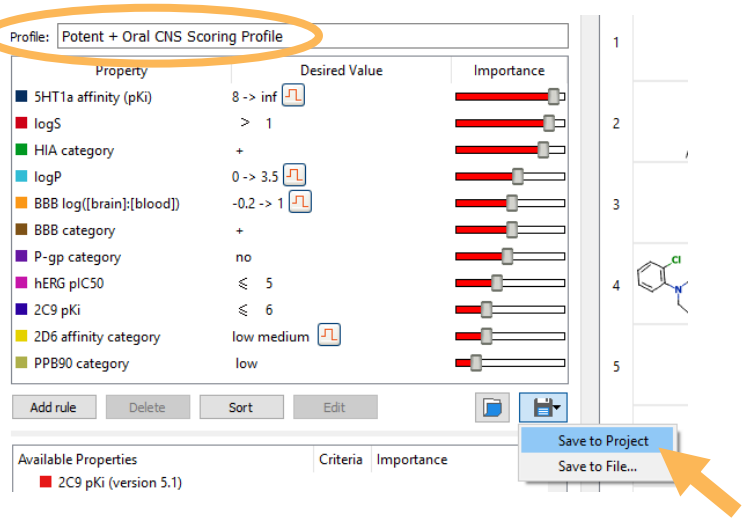
- Select **5HT1a affinity (pKi)** in the profile editor and click the **Edit button**.
- In the **Score Editor**, click the **Insert button** to insert a range from 7 to 8, with a score of 0.05 for pK$_i$ = 7 and a score of 1 for pK$_i$ = 8, as shown below.
- Edit the score for pK$_i$ below 7 to be **0.05** and for pK$_i$ above 8 to be **1**, by double-clicking on them, as shown below.



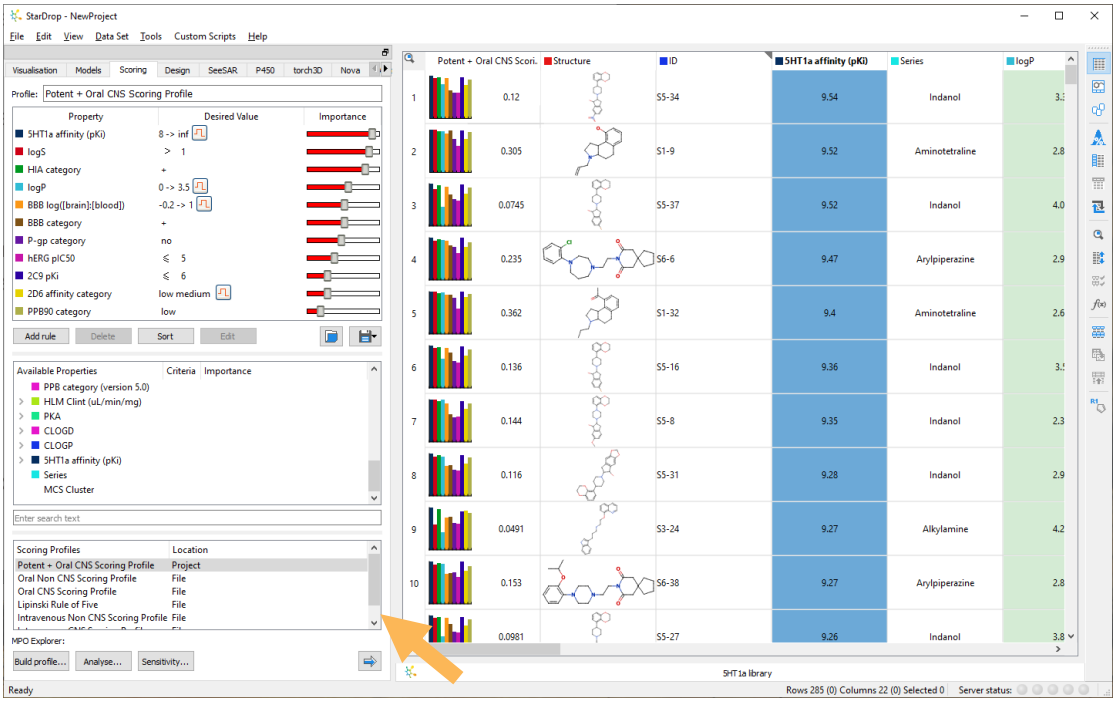
- Click **OK** when you have finished.

- Give the resulting profile a name by editing it at the top.
- Save it into the project by clicking on the **Save button** below and choosing **Save to Project**.

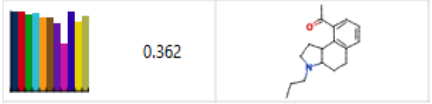
It will appear in the list of scoring profiles at the bottom, so that you can retrieve it easily.



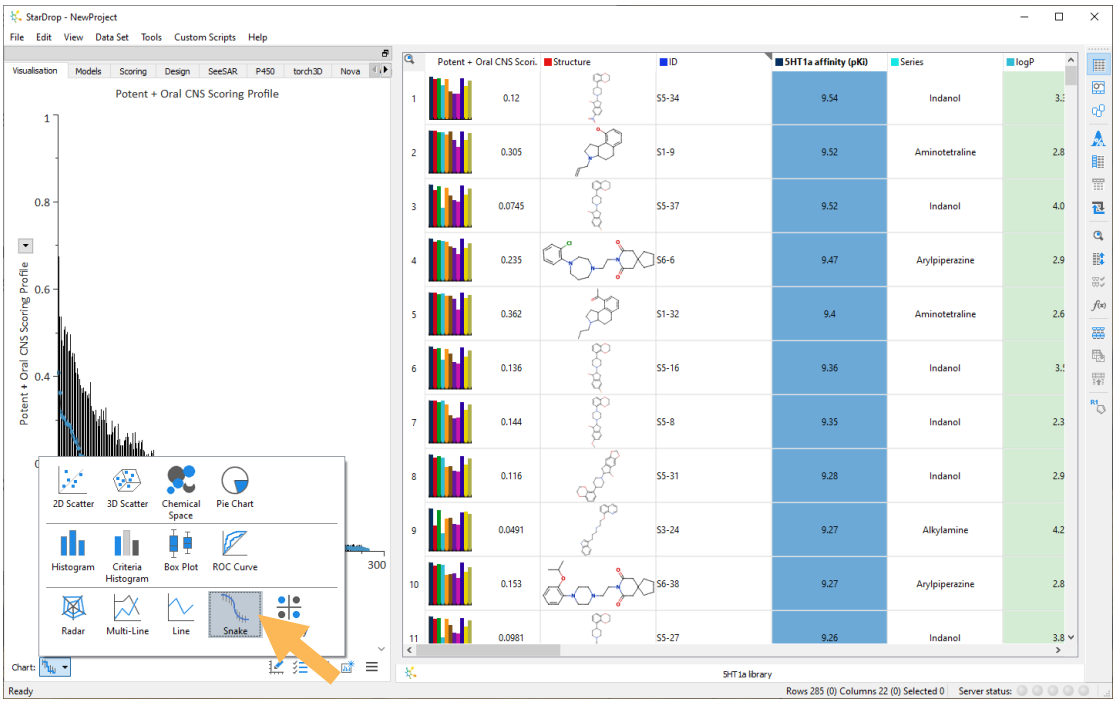
- Run this scoring profile by clicking the button at the bottom of the **Scoring area**.



We can see that in the first few rows where we have the most potent compounds there are some good scores (> 0.3). Looking at row 5 we can see that it is only the predicted hERG pIC_{50} (represented by the pink bar in the score histogram) that appears to be a problem for this compound.

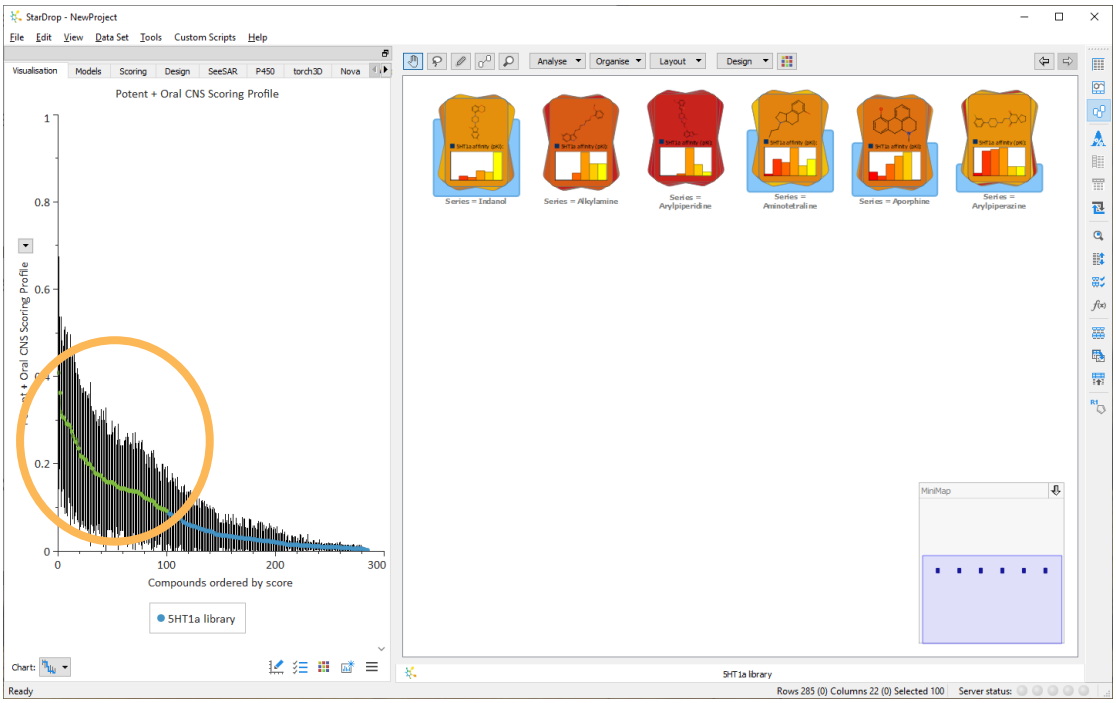


- Change to the **Visualisation area** and from the **Chart button** select **Snake** to create a **Snake Plot** showing all the compounds, ordered by score along the x-axis.



The y-axis shows the score for each compound and the error bars indicate the level of confidence we can have in the scores, given the uncertainty in the underlying data. Points in the snake plot for which the error bar overlaps with that of the first compound cannot be confidently distinguished from the highest scoring compound.

- Change back to Card View by clicking the **Card View button** again.
- Select the top 100 compounds by lassoing them in the snake plot.



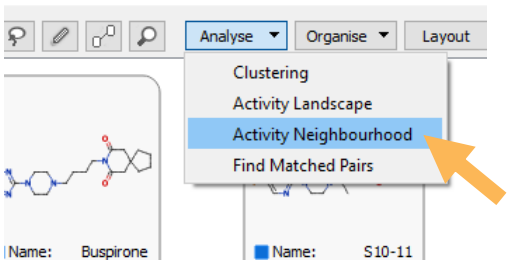
These compounds' scores all overlap with the top scoring compound's error bars showing that there are compounds from four series which have a good overall balance of properties.

Lead Optimisation: Guiding the Design of Balanced Compounds

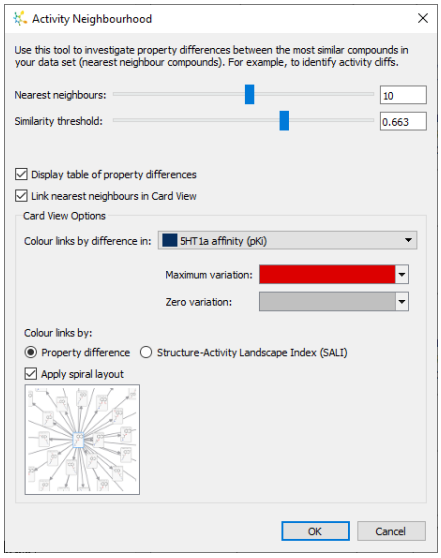
One of the chemical series chosen for progression from the hit-to-lead project, explored in the previous section, was a series of Arylpiperazines (series S10). Due to their similarity with the drug Buspirone, which was known to have issues due to rapid metabolism by Cytochrome P450 CYP3A4, the compounds were initially tested for stability against this P450 isoform.

In the following example, we will explore approaches to investigate the structure-activity relationships (SAR) for these properties along with additional predicted properties. Furthermore, we will see how the Glowing Molecule™ visualisation can help to guide the design of compounds to overcome potential liabilities, while monitoring other properties to ensure that improvements to one property do not have a negative impact on other important factors.

- Open the file **Arylpiperazine series S10.add** containing 20 compounds from series S10, for which both potency (pK_i) against the 5HT1a target and half-life for metabolism by CYP3A4 have been experimentally measured.
- Switch into Card View by clicking the **Card View button** .
- Select compound **S10-14** (using the **Find** tool on the toolbar if needed).
- We are going to investigate the SAR around this compound, so from the Card View **Analyse menu** select **Activity Neighbourhood**.

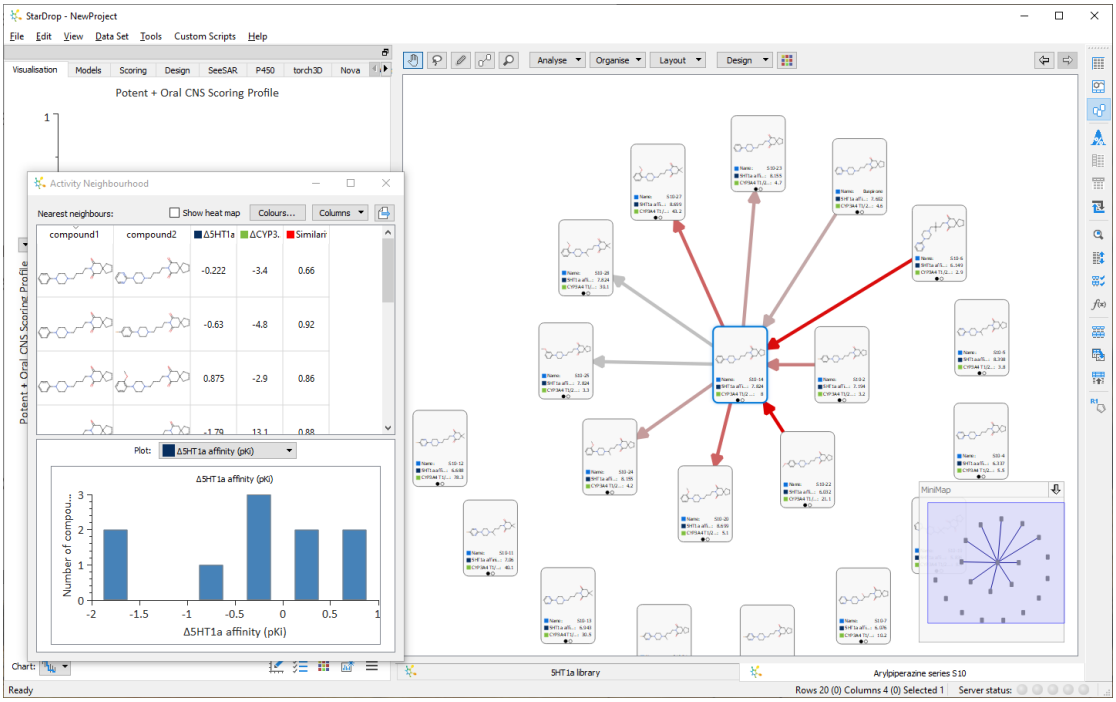


We will use the default values for all the settings, linking the 10 most similar compounds to the selected reference compound. This will help us to find any 'activity cliffs', where a small variation in the compound structure has a large impact on the potency.

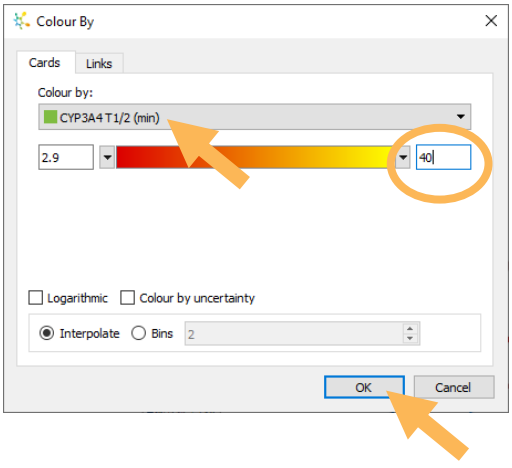


- Click **OK** to run the analysis.

The compounds are laid out in a spiral with the reference at the centre and the most similar compounds close to it. The arrows show the direction in which potency increases with red links indicating the largest differences.

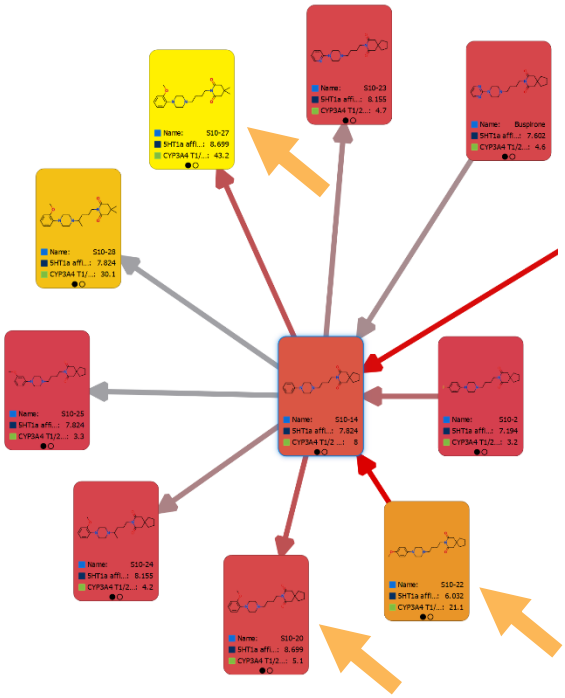


- Close the **Activity Neighbourhood dialogue** and click the **Format button**  at the top of Card View.



- **Colour By** the property **CYP3A4 T1/2 (min)** setting the from red (low) to yellow (high) setting the upper threshold to **40** minutes (any values greater than this will also be yellow) and click **OK**.

Of the most similar compounds, the second one (S10-22) is coloured more orange than the reference, indicating greater stability. However, it has a red arrow pointing towards the reference indicating that it is significantly less potent. The only difference in structure is the presence of the methoxy group at the para- position on the benzyl.

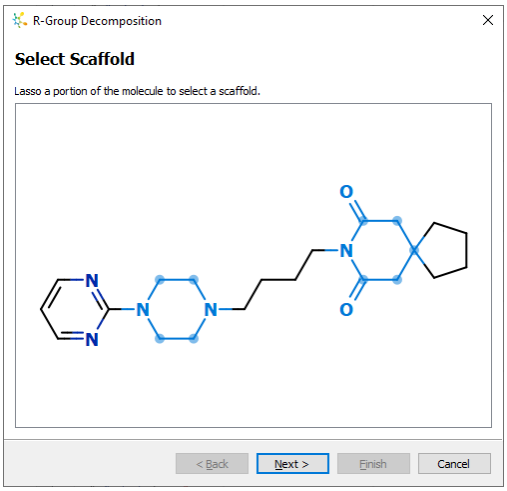


The next compound (S10-20) has a red arrow pointing the other way, indicating that it is more potent than the reference. It also has a methoxy on the benzyl but this time at the ortho- position. The change of position of the methoxy looks interesting in terms of potency, but we can see that this compound lacks stbility, like the reference.

Further round is a bright yellow card, compound S10-27, indicating good stability. It has a red arrow pointing towards it indicating that it is also significantly more potent than the reference. It contains the ortho- methoxy substitution which perhaps explains its increased potency and in combination with the gem-dimethyl replacement for the spiro- group at the other end of the structure appears to satisfy both properties simultaneously.

An alternative way we could investigate the same problem is by carrying out an R-group decomposition.

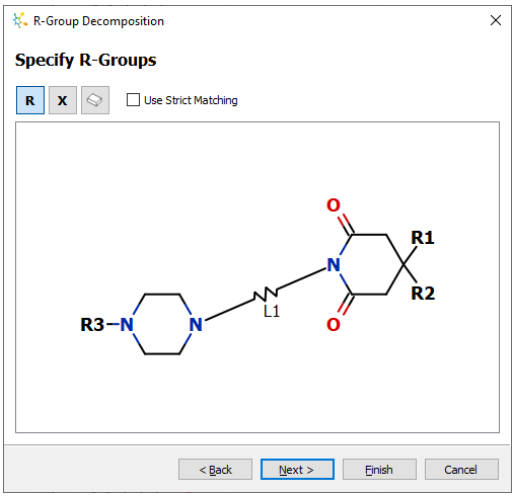
- Change back to the **Table View** by clicking the **Table View button** on the right-hand toolbar.
- Select Buspirone and click the **Run R-group Decomposition button** on the toolbar.



First, we will define the regions representing the common scaffold within the chemical series.

- Select the piperidinedione and piperazine rings by drawing around the, holding down the **Ctrl key** to select a second region.

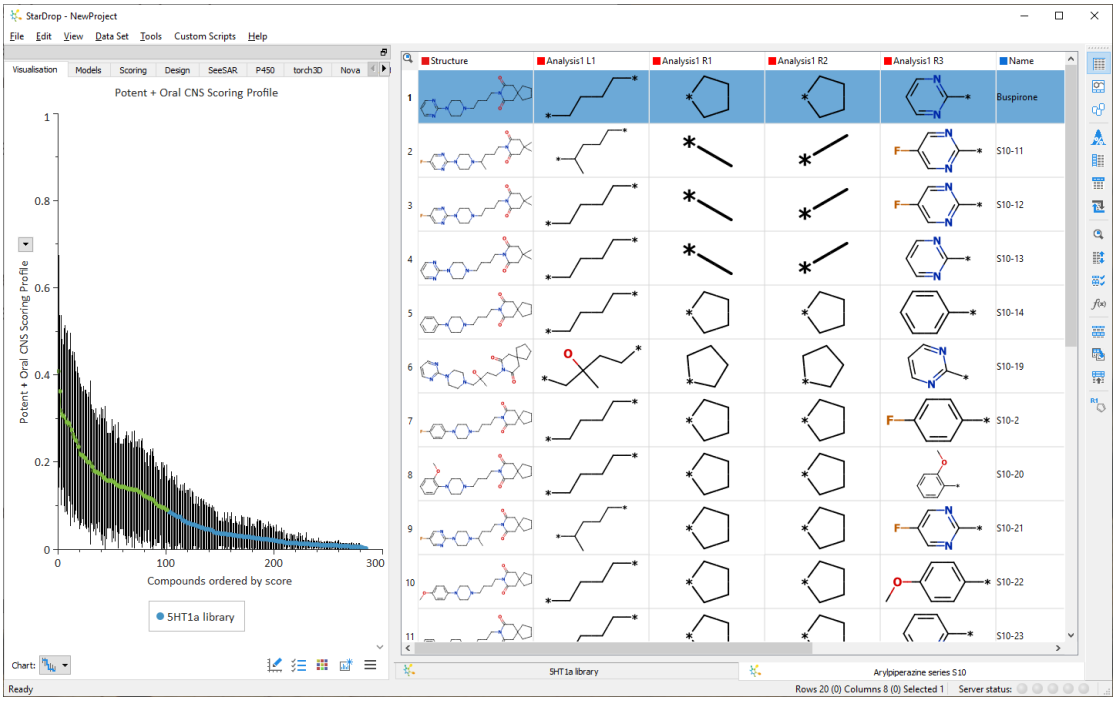
- Click **Next** to confirm the R-Groups that will be analysed.



In this example, the scaffold represents two separate regions, so the connecting group will be considered to be a variable linker.

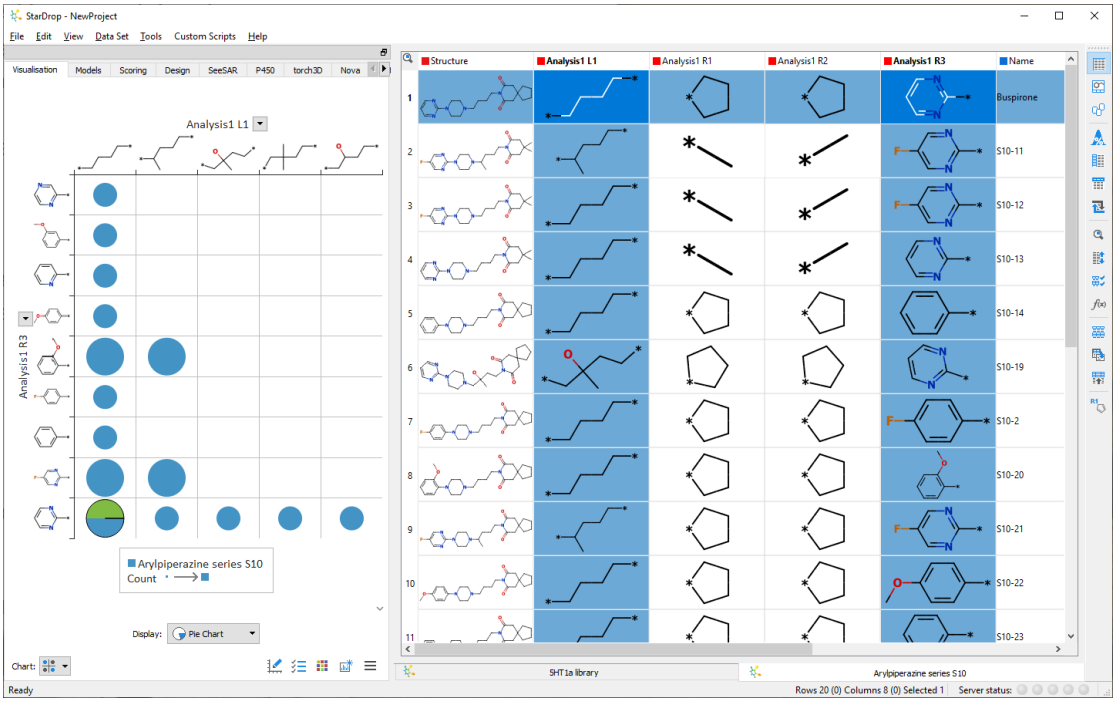
Note: You can indicate additional R-Groups by selecting the R button and then clicking at the point on the scaffold where they should be included. To specify variable atoms or fragments, click the X button and then select atoms on the scaffold which may vary across the series. If you have multiple scaffolds decorated with the same R-groups, you can decompose them at the same time by selecting one example of each scaffold at the start.

- Click the **Finish button**.



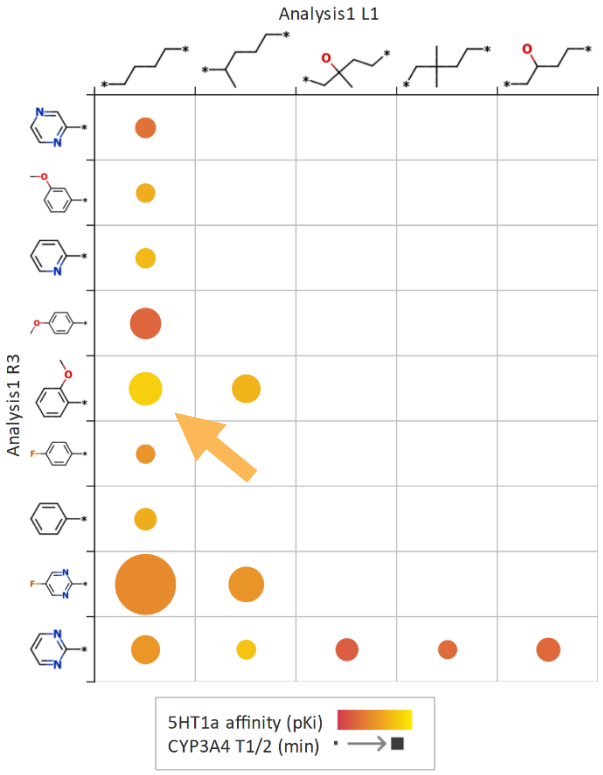
New columns are added to the data set indicating the R-Groups and linkers present.

- Select columns **Analysis1 L1** and **Analysis1 R3** at the same time (hold down the **Ctrl key** while selecting) to display an SAR table in the Visualisation area.



Note: You can also create an SAR table by selecting **Category Matrix** from the **Chart menu**.

- Format the SAR table, colouring by the potency, **5HT1a affinity (pKi)**, and sizing by the stability, **CYP3A4 T1/2 (min)**.



In the SAR table, the most interesting cell is the medium sized yellow circle for the ortho- methoxy combination with the tetramethylene linker.

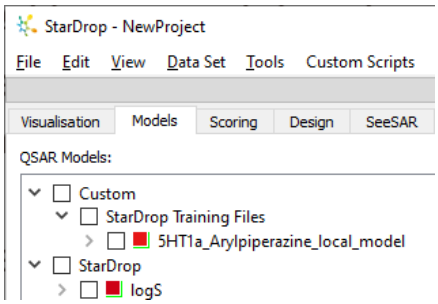
- Click on this cell to select the associated rows in the data set.

This highlights the two compounds in the data set containing this combination of R-groups, one of which is the very promising compound S10-27, highlighted earlier.

One of the potential issues identified for this chemical series is inhibition of the hERG ion channel, indicating a risk of QT prolongation and cardiotoxicity. Therefore, we will use the interactive designer to explore potential strategies to reduce the predicted hERG pIC_{50}, guided by the Glowing Molecule.

To enable us to consider the effect of any changes we make to the potency, we will use a QSAR model developed for this series to predict this property.

- Change to the **Models area** and click the button to load the model file named **5HT1a_pKi_Arylpiperazine_local_model.aim**.



This will appear in the list of models under the branch alled **StarDrop Training Files**.

To enable us to consider the overall scores for our compounds, taking into consideration the different property criteria and their relative importance, we will also load a new scoring profile that uses this model.

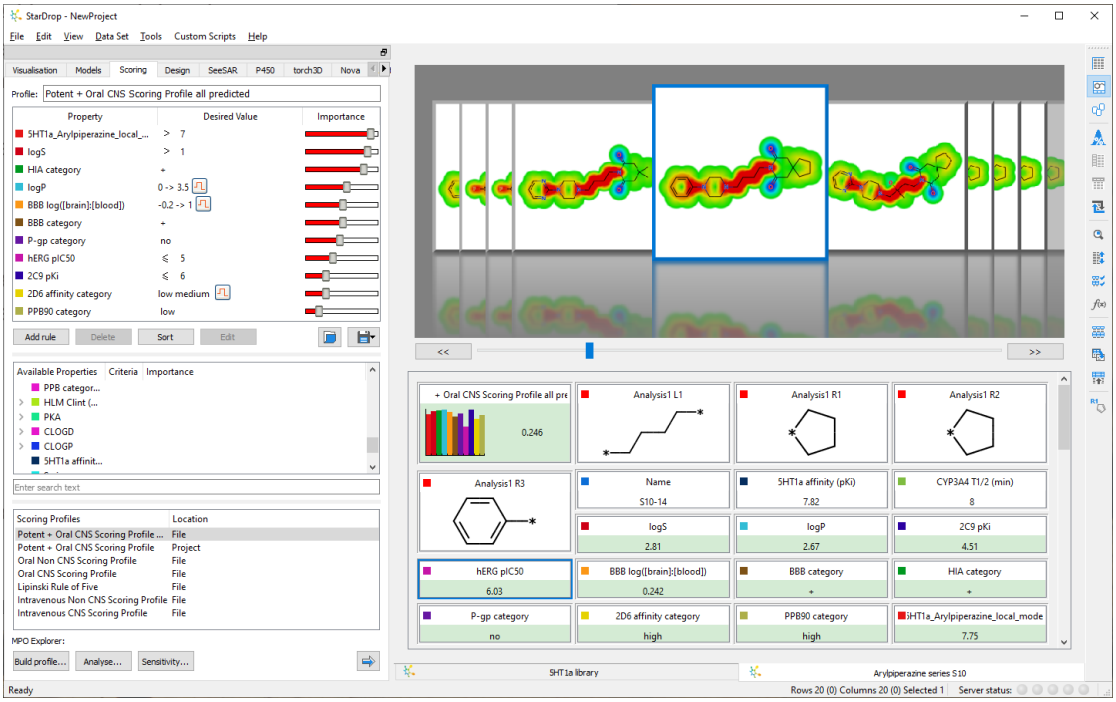
- Change to the **Scoring area**, click the button and load the scoring profile file named **Potent + Oral CNS Scoring Profile all predicted.apd**.

This is similar to the profile we created earlier, using predicted values of the 5HT1a pK_i in place of experimental pK_i.

- Score the compounds using the new profile by clicking the button.

Note: the models used in the scoring profile will run automatically.

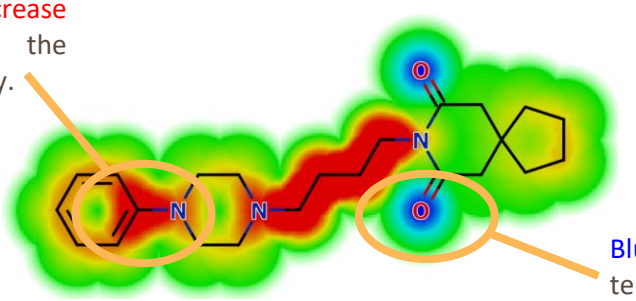
- To help us to get an overview of each compound's properties change to the **Molecule View** by clicking the **Molecule View button** on the toolbar.
- As before, select compound **S10-14**, which is the highest-scoring compound.
- Click on the **hERG pIC50** model result to display the Glowing Molecule for this property, as shown below.



Change to the **Design area** to explore strategies to reduce the potential for hERG inhibition.

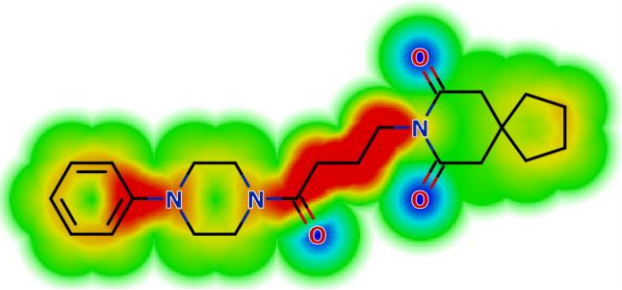
- The selected structure will be displayed in the structure editor. When the **hERG pIC50** prediction is selected, the Glowing Molecule will also be displayed.

Red regions have a tendency to increase the value of the predicted property.



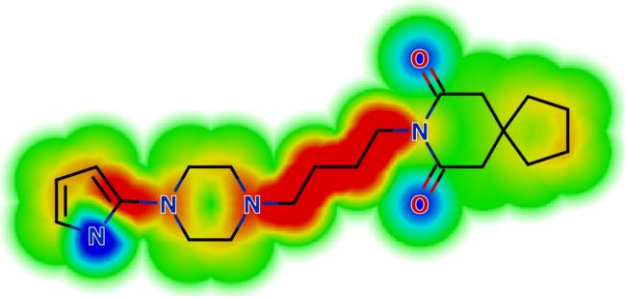
Blue regions have a tendency to decrease the value of the predicted property.

- Edit the structure, as shown below.



The predicted hERG pIC50 decreases to 5.38, but despite improving the hERG pIC50 the overall score has decreased to 0.152 because the predicted blood-brain barrier penetration is now lower and this property is more important in terms of the overall score than the hERG pIC50.

- Try drawing the structure below instead, replacing the phenyl of compound S10-14 with a pyrole.



Hint: To quickly delete the phenyl ring in the previous compound, draw around the phenyl ring with the **select tool** and type **Ctrl-X**.

The predicted hERG pIC50 decreases to 5.37 and now the overall score increases slightly to 0.252, suggesting that this may be an interesting idea to explore further.

- Add this compound to the data set by clicking the button below the editor. If you wish, you can give the compound a name by double-clicking in the **Name** cell.

Feel free to explore some additional ideas for how to reduce the predicted hERG pIC50 without having a detrimental effect on the overall balance of properties.