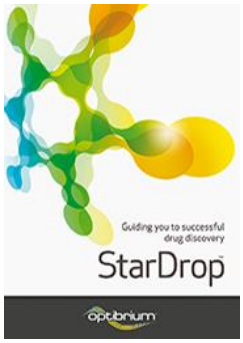


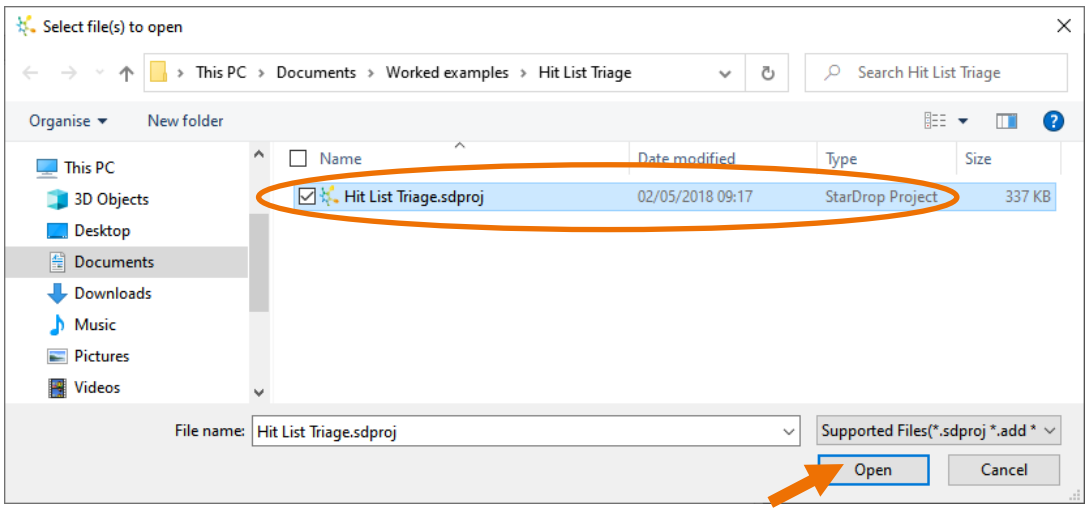
Worked Example:

Hit List Triaging Using Card View™

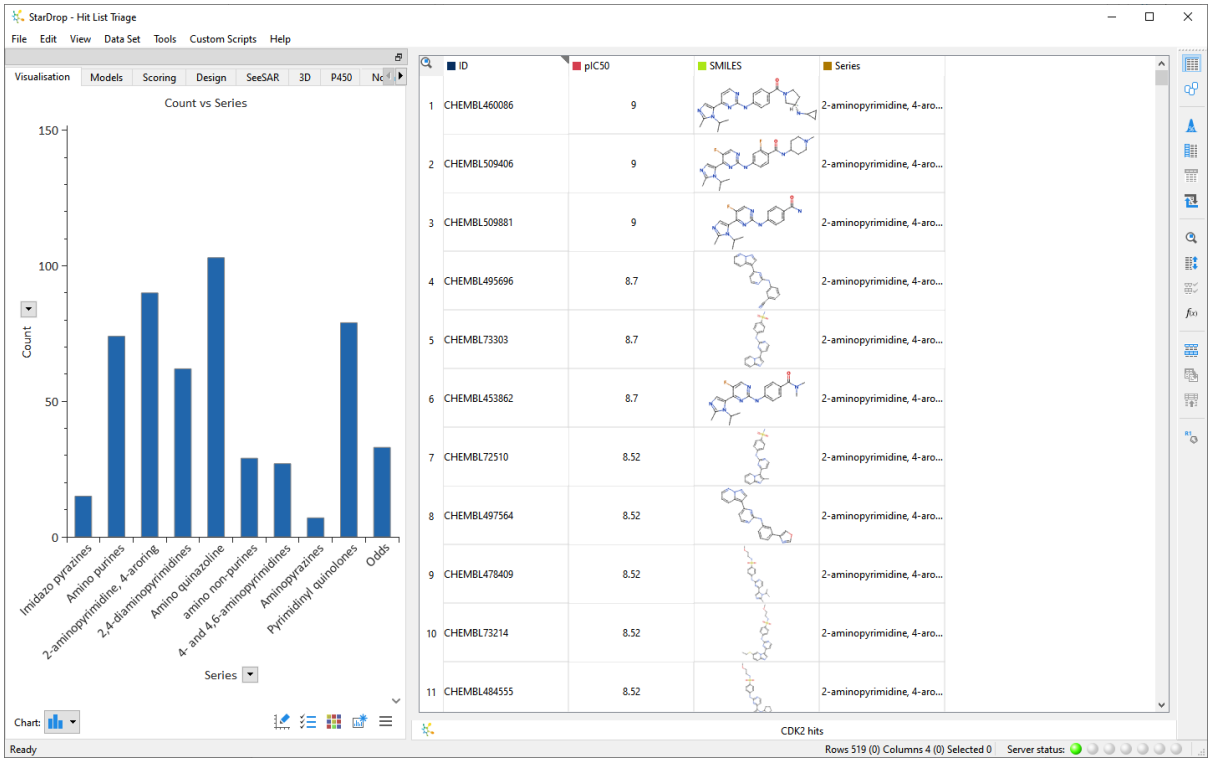
This example is taken from a kinase project in which a large screening campaign has resulted in a hit list containing several chemotypes. The project team wish to evaluate the list and focus their resources on a small number of series that have demonstrable SAR at the target and that are most likely to yield high-quality leads with appropriate physicochemical and ADME properties. In the following steps, we will utilise the features of Card View to quickly and easily accomplish this task. **Note:** The compounds are derived from a set of public domain pIC_{50} data measured against the CDK2 kinase and obtained from the ChEMBL database https://www.ebi.ac.uk/chembl/

Exercise

- From StarDrop's **File** menu, select **Open** and open the project file **Hit List Triage.sdproj**.



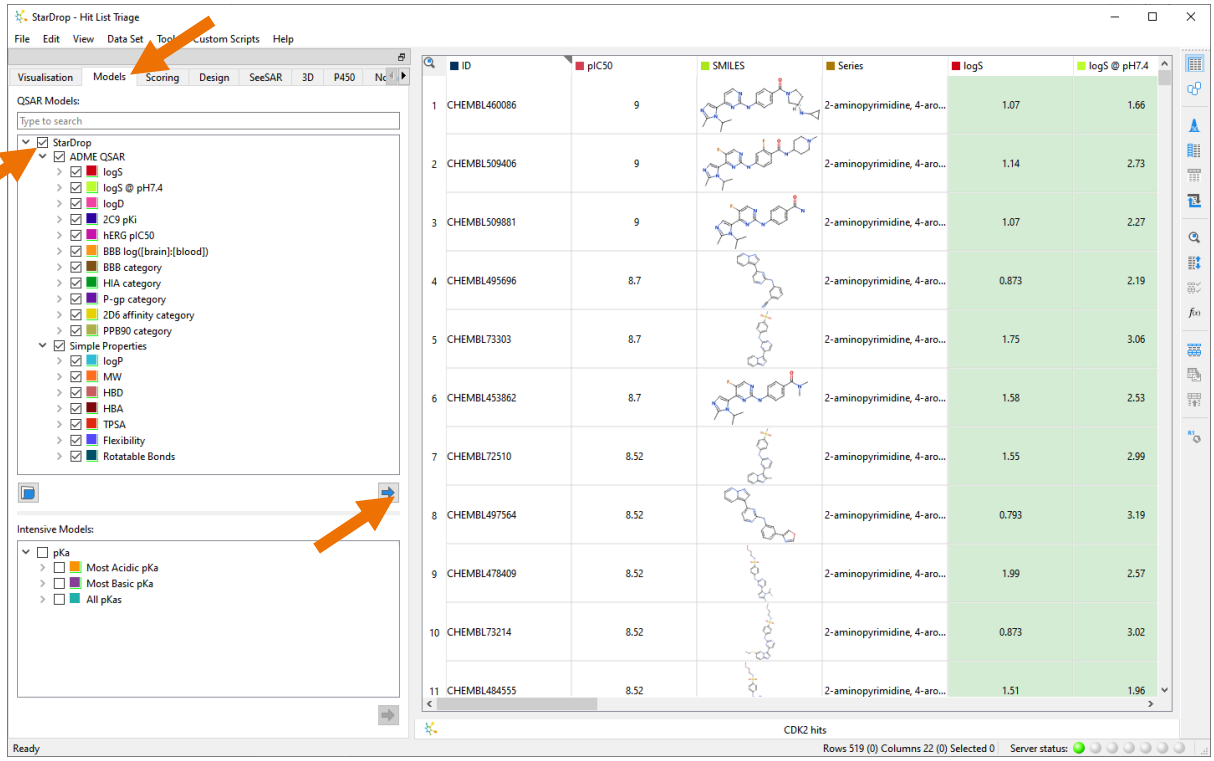
The project contains a single data set containing 519 compounds with experimental potency data (as pIC_{50} values) against Human Cyclin-Dependent Kinase isoform 2 (CDK2), along with a chemical series tag.



The project also contains a histogram showing the number of compounds in each of the series.

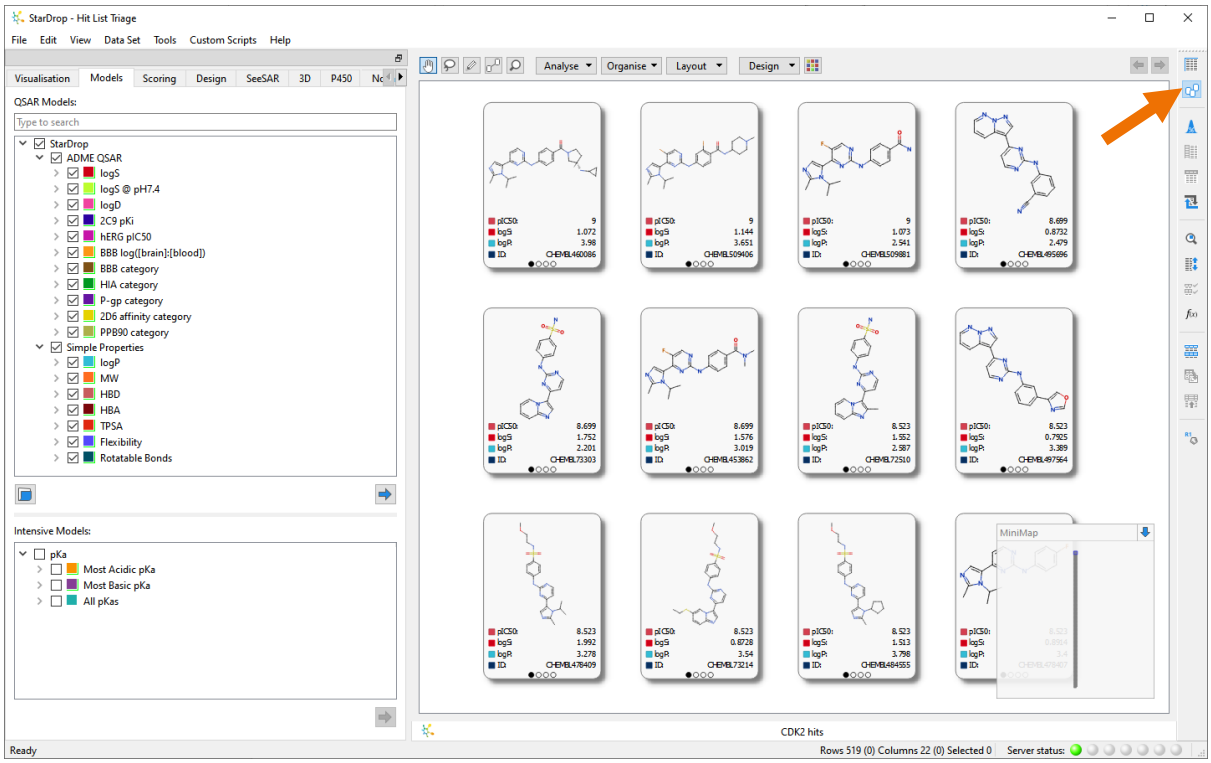
We have no further measured data at this stage of the project, so we're going to generate some predictions of ADME and physicochemical properties.

- Click on the **Models** tab, check the box next to the word StarDrop (this selects all the models under this branch) and click the  button below.



A new column will be added to the data set for each property predicted.

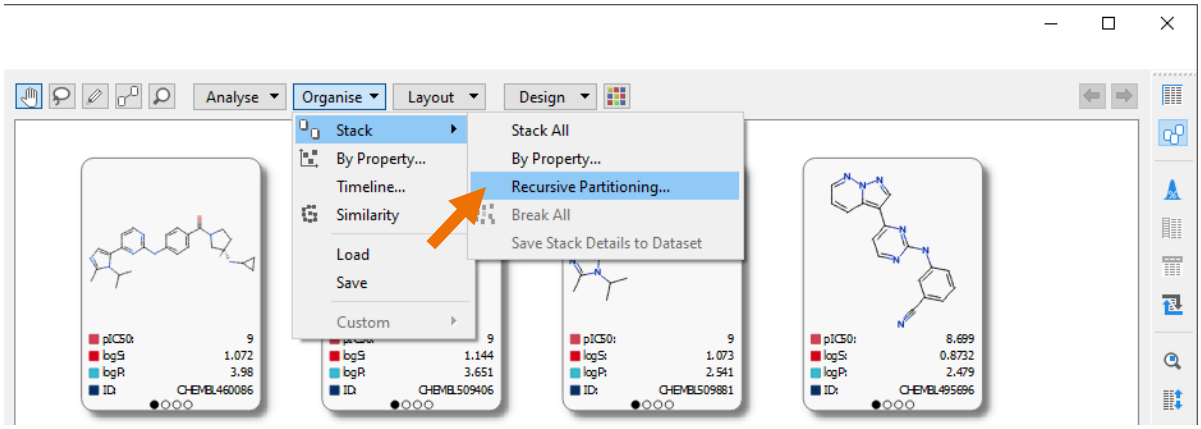
- Click the **Card View** button on the right-hand toolbar to switch from Table View to Card View.



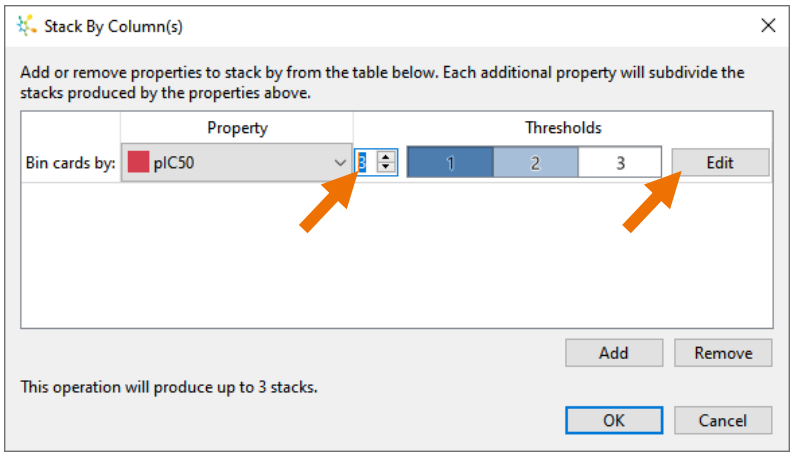
You can move the view of the cards by dragging the view. You can zoom in or out using a mouse wheel or the **CTRL** and **-/=** keys. You can move an individual card by dragging it to a new location.

To quickly see which (if any) compounds have an acceptable activity and solubility profile, we can group the compounds into stacks.

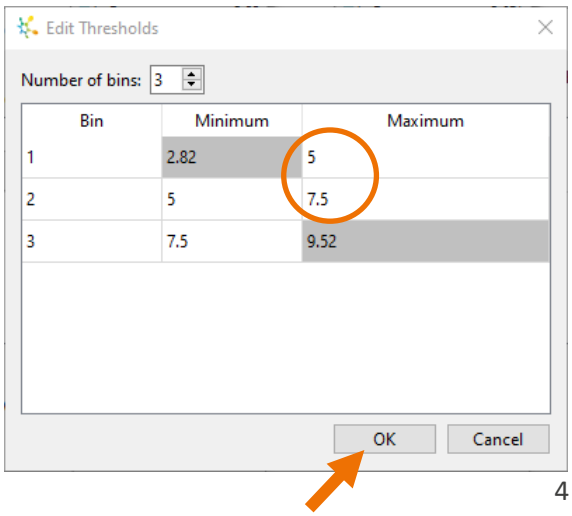
- From **Organise** menu at the top of Card View, select **Stack**, then **Recursive Partitioning**.



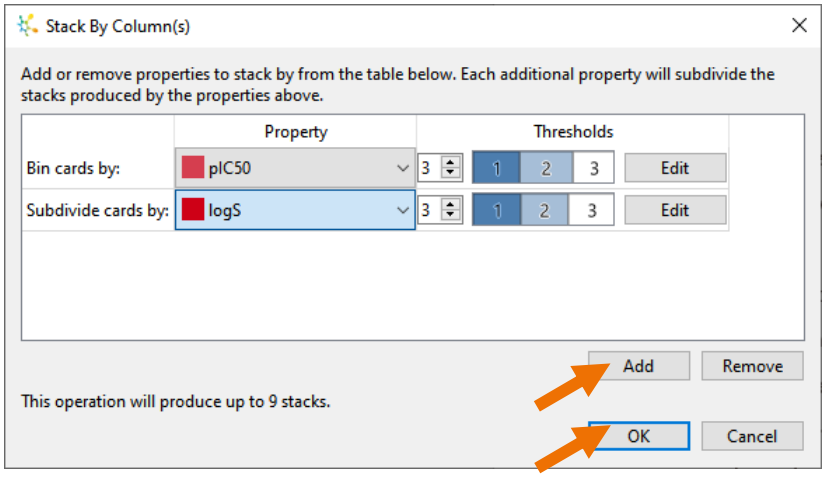
- Choose **pIC50** as the property by which to bin the cards and increase the number of bins to **3**.
- Click the **Edit** button to modify the ranges.



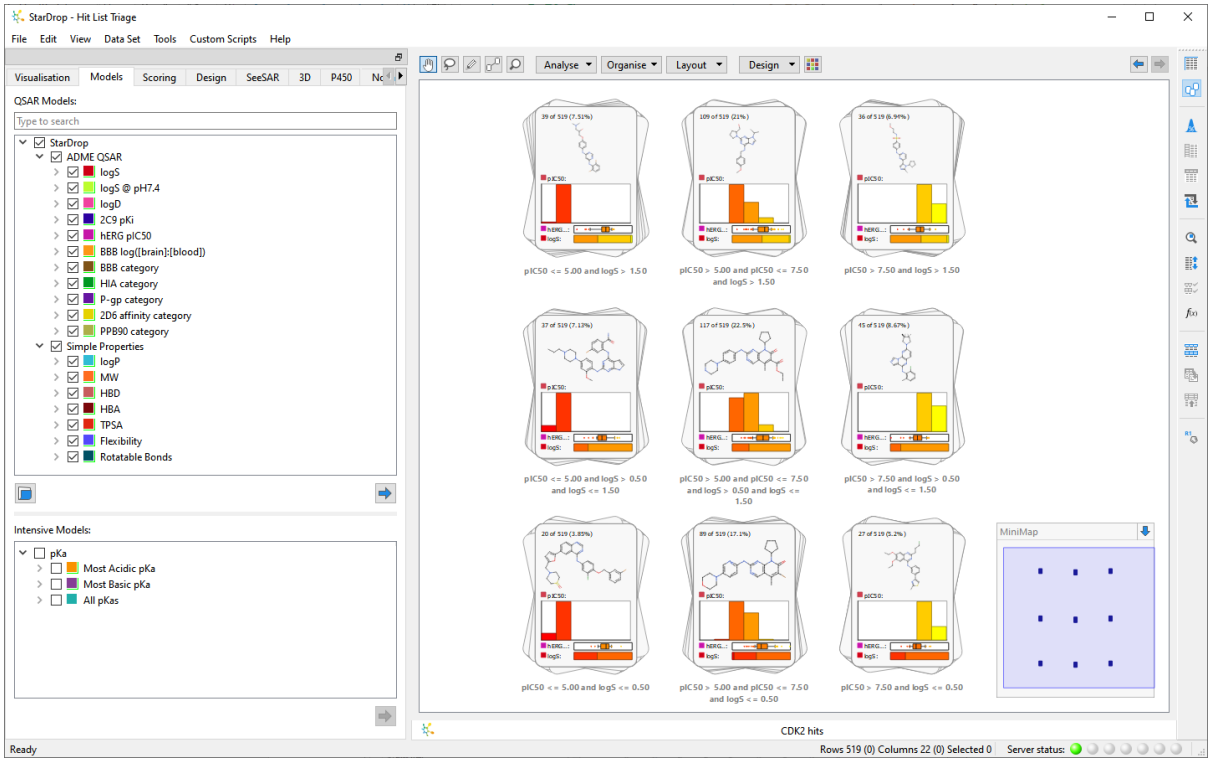
- Change the boundary values to **5** and **7.5** and click the **OK** button.



- Click the **Add** button to subdivide the stacks based on their predicted **logS** values and then split this also into three bins with **0.5** and **1.5** as the boundary values.



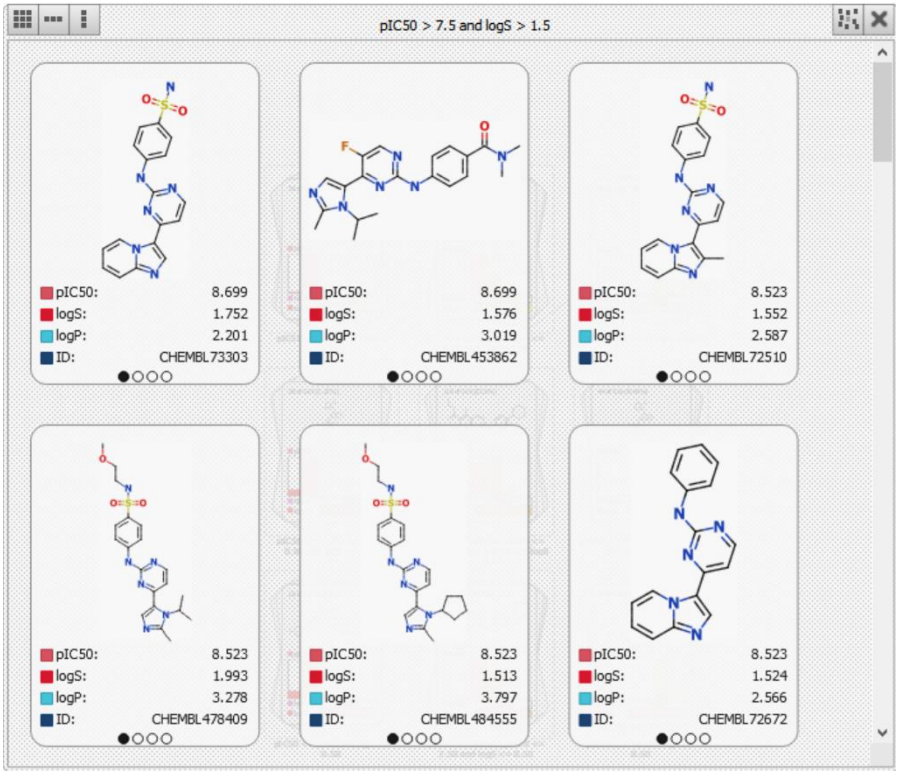
- Click the **OK** button to stack the cards using these properties.



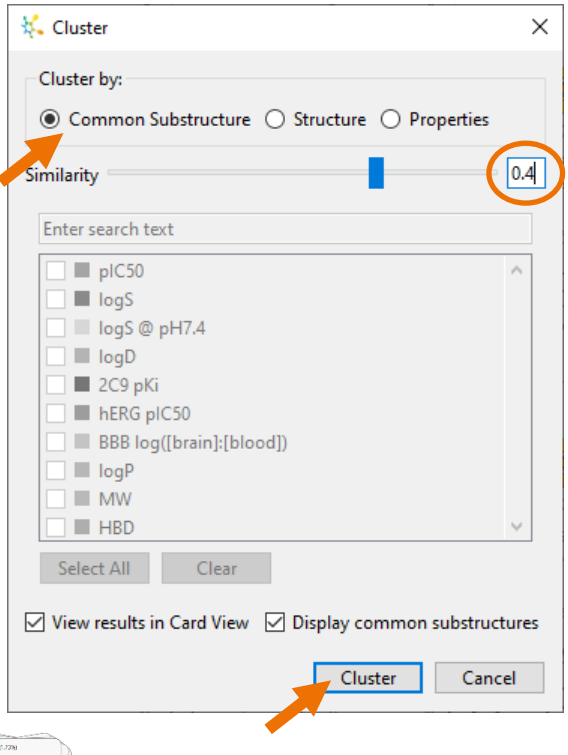
The top right-hand stack has interesting compounds (36 out of 519) because they are potent and predicted to be soluble.

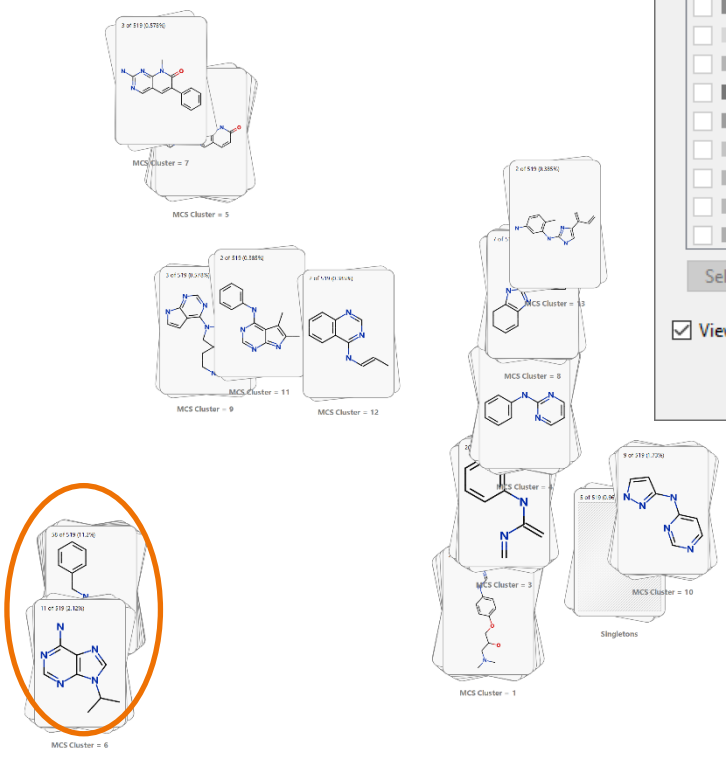
- **Double-click** (or right-click and choose **Inspect**) on the stack to look at the cards within that stack.

The stack contains a variety of chemical scaffolds, which might increase the project's chance of success, but we would like to know how these series are distributed across the whole data set. For this, we need to define the chemotypes in the set, which we can do by clustering the data set based upon the common substructures.



- From the **Analyse** menu at the top of Card View, select **Clustering**.
- Choose the **Common Substructure** option, enter a similarity value of **0.4** and click the **Cluster** button.





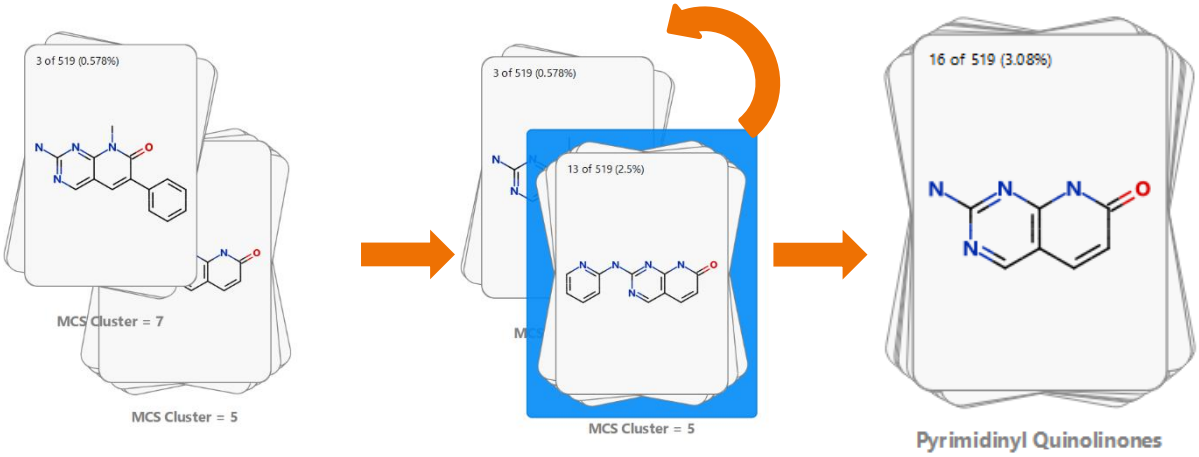
The cards will be stacked, representing sets of cards that share a common substructure, shown on the stack cover. The stacks are laid out, as shown above, based upon the relative similarities of the common substructures. You can see that there is a group of 2 stacks on the left that all have various decorations off an amino purine core.

These 2 stacks could be thought of as a single 'amino purine' series and can be easily combined manually into one stack. **Note**: you can move them to see the substructures better.

- Choose the **lasso** tool from the Card View toolbar and draw a ring around the 2 stacks. The selected stacks can now be combined by **right-clicking** on the selection and choosing **Stack Selection** from the menu.

- The new common substructure of a mono amino purine is now displayed, and this stack can be named "amino purines" by selecting and editing the "MCS Cluster = " label below the stack (click on the label to edit it).

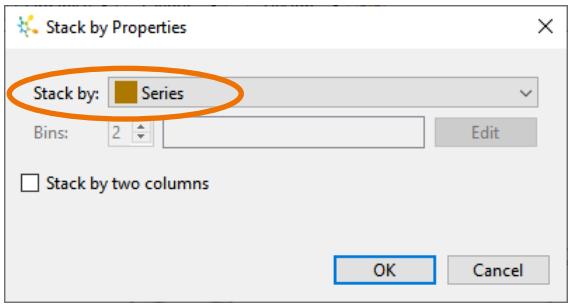
We can repeat the grouping process for a set of pyrimidinyl-quinolinones in the main group of stacks. Here, with only 2 stacks, we can use the **hand** tool to simply drag one on top of the other as shown.



Once you have refined the stacks, the names you have given them can be added to the data set in a new category column by clicking the **Organise** button and selecting **Stack**, then **Save stack details to dataset** from the menu.

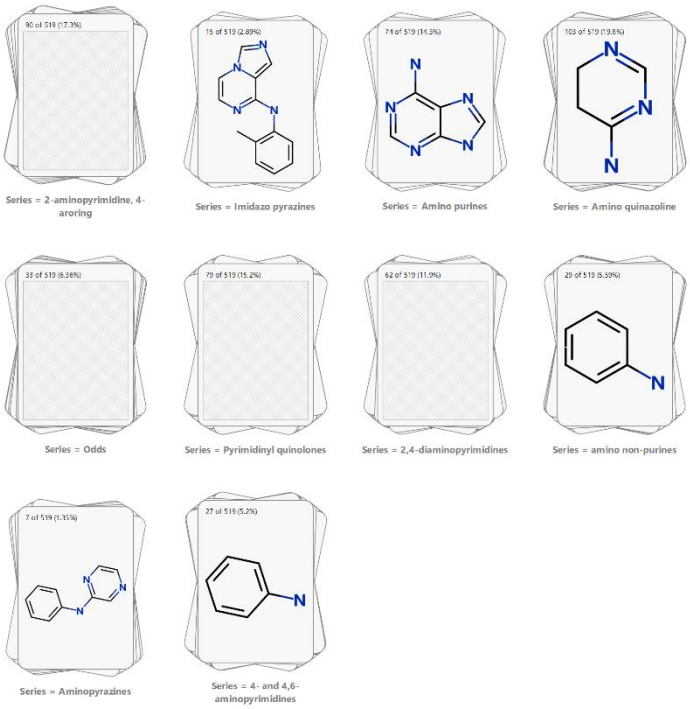
To save time, this process has already been done for you in this data set, and the stack names for the chemotypes are provided in the **Series** column.

- To view the data set with these defined series, click the **Organise** menu at the top of Card View and from the **Stack** menu, choose **By property**. In the **Stack by Properties** dialogue that appears, choose the **Series** column from the drop-down list and click the **OK** button.



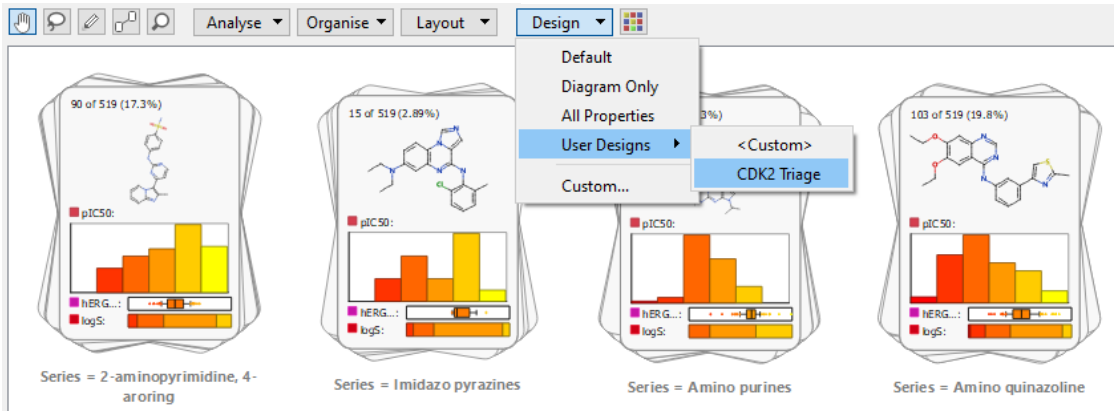
This stacks the cards into 10 stacks which have the common substructure detected within the stack. As the series definitions here are human-derived, there are some stacks that do not have a large enough common substructure for display.

- The stacks can be laid out in a more compact grid by clicking the **Layout** menu at the top of Card view and choosing **Grid**.



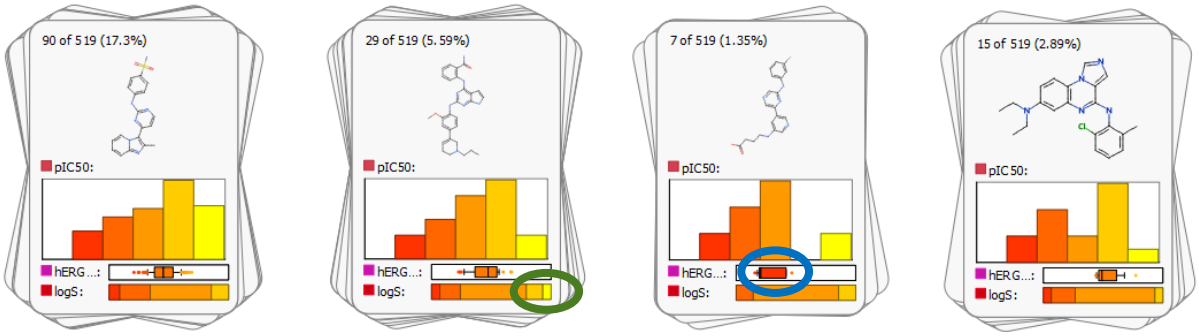
More relevant information about the property distributions for the cards in the stacks and a representative compound from the stack can be seen by changing the card template.

From the **Design** menu select **User Designs** then **CDK2 Triage**.



The histograms on the stack covers show the distribution of key properties of the compounds within each stack; potency as a histogram, hERG pIC_{50} as a box plot and logS as a compact histogram. These displays can help to answer the question of "which series has a good range of potencies as well as acceptable logS and/or hERG pIC_{50}?"

We can see from the list of stacks that the 'amino non-purines', '2-aminopyrimidine, 4-aroring', and the two 'pyrazines' series all have compounds with good activity (the yellow bars in the potency histogram); however, the solubility distribution looks more promising for the 'amino non-purines' (as indicated by the green ring) whilst the 'aminopyrazines' appear to have the better hERG profile (as indicated by the blue ring).



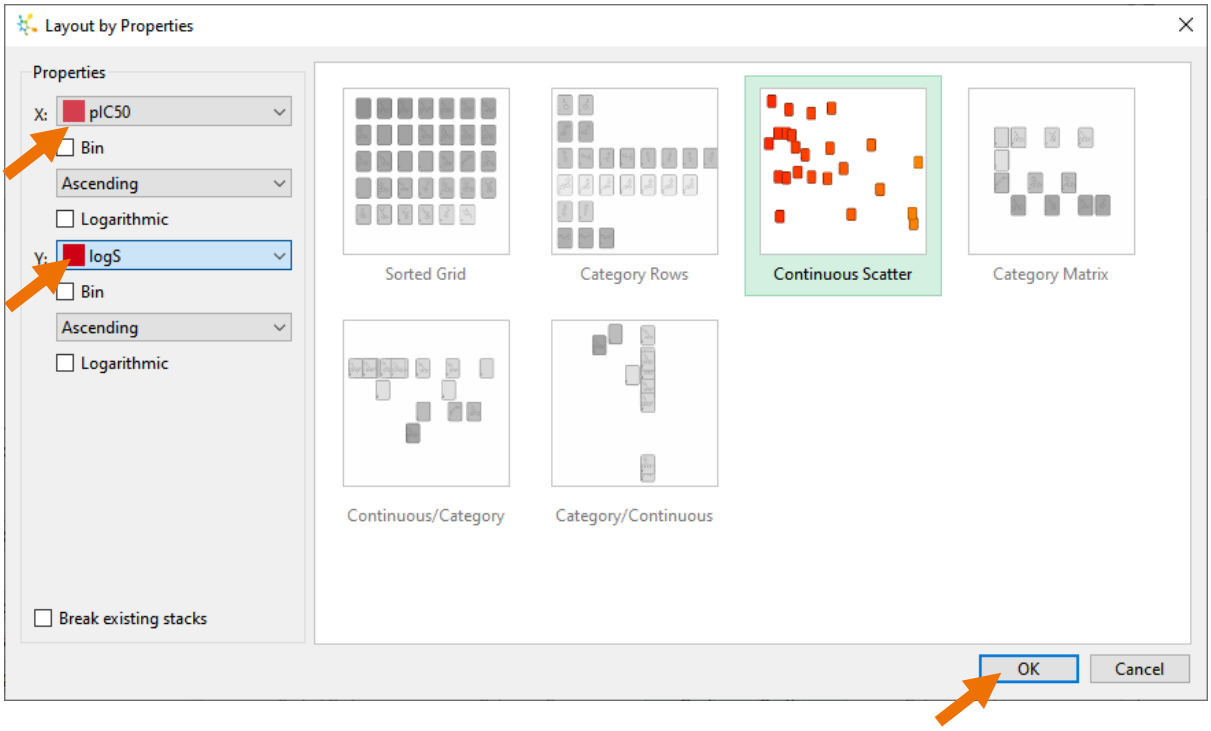
Note: You can, at any time, **right-click** and **Copy Image** of the stacks or cards displayed in order to paste the view into a slide for presentation purposes.

Next, we shall examine one of these clusters to see what SAR points can be learned from a particular series. A co-crystal structure of one of the compounds in the '2-aminopyrimidine, 4-aroring' stack has been published (PDB code 1OIT) and so with the possibility of using the available 3D information, this cluster is of interest to follow-up.

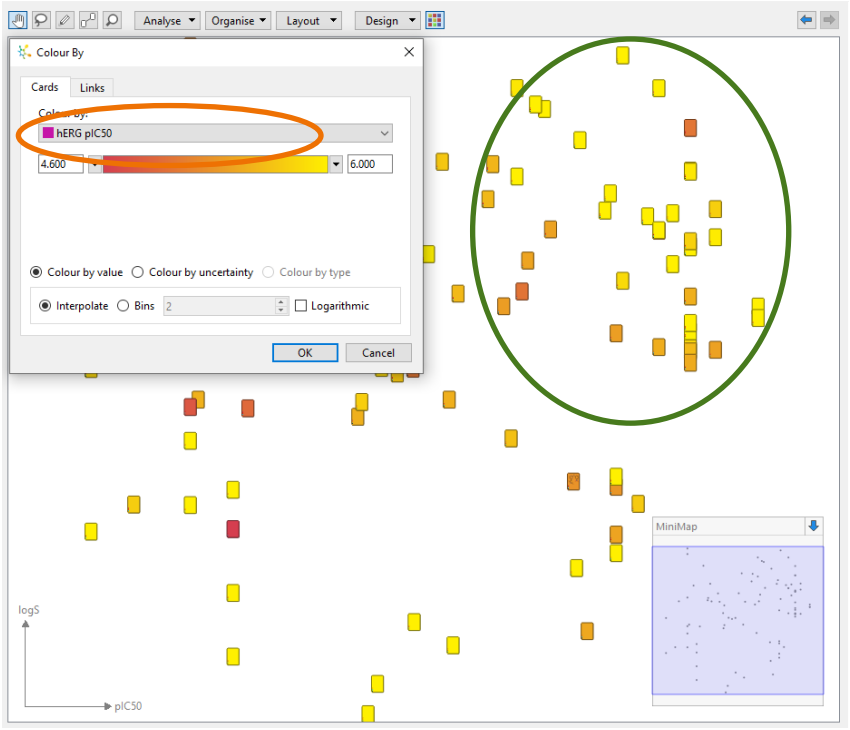
- Select the **'2-aminopyrimidine, 4-aroring'** stack and click the **Create new data set from selection** button  on the right-hand toolbar (or from the **Data Set** menu choose **Create from selection**).

The selection is copied from the original data set into a new data set.

- You can see how properties are distributed across this chemotype by clicking the **Organise** menu at the top of Card View and choosing **By Property**.
- Set **pIC50** as the X direction property and **logS** as the Y direction property and click the **OK** button.



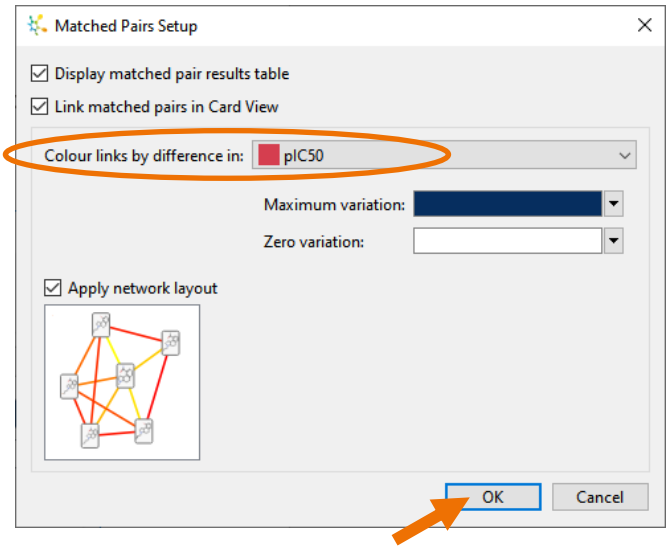
- These cards can be coloured by their predicted **hERG pIC$_{50}$** using the **format** button  . In the **Colour By** dialogue, select **hERG pIC$_{50}$** and change the interpolation scale to go from **4.6** to **6** as shown right.



Note that the top right-hand corner of the display (which corresponds to good pIC$_{50}$ and solubility values) has very few red cards (indicating inactivity at hERG). This means that there is a conflict in this series between having soluble, potent compounds and low hERG activity. This could be addressed using StarDrop's Probabilistic Scoring approach to multi-parameter optimisation to identify alternative structures within this series with a better balance of properties. These might act as more useful starting points for further optimisation than simply the most active or the most soluble. Take a closer look at the use of Probabilistic Scoring in some of the other tutorials available in the Optibrium Community.

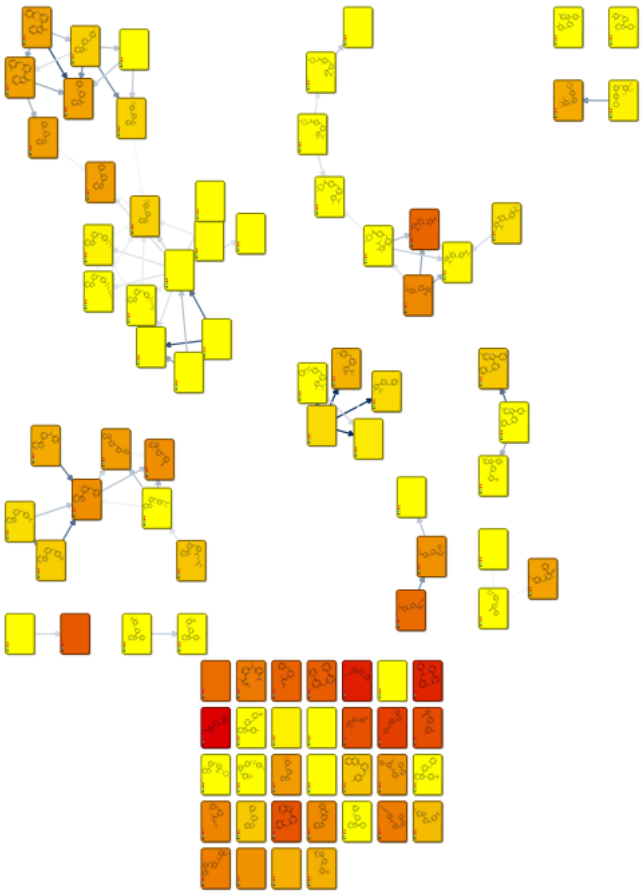
The more detailed SAR in this chemotype can be investigated by finding the matched pairs within the series.

- Click the **Analyse** menu at the top of Card View and choose **Find Matched Pairs**.
- Colour the links by *differences* in the **pIC50**, with the maximum variation coloured dark blue and the zero-variation white (you can change the colours by clicking on them).
- Click the **OK** button to run the matched pairs analysis.

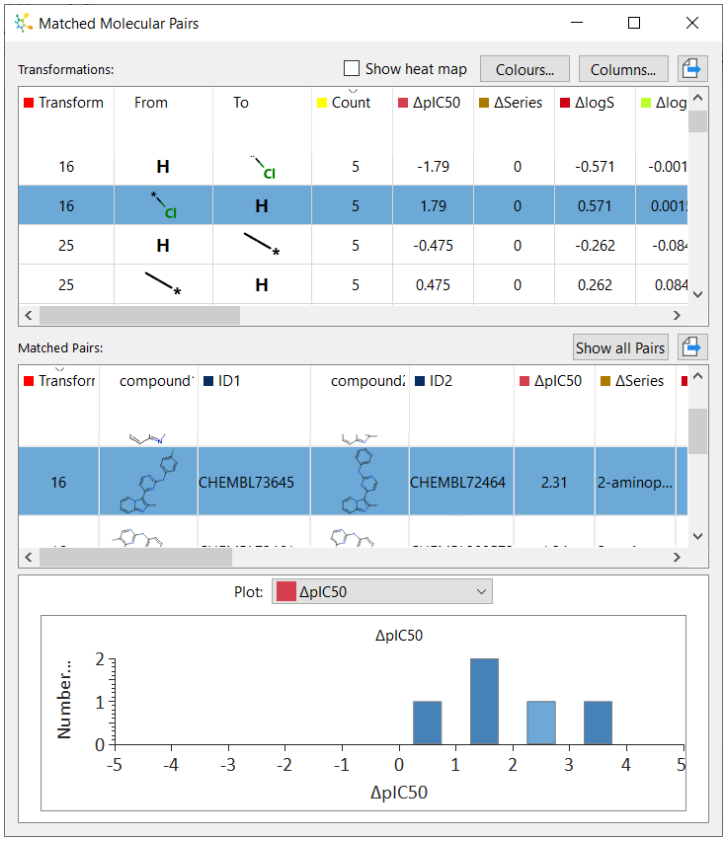


The matched pairs results will be shown in Card View as a network in which each matched pair is indicated with a link between the corresponding cards. A table will also be created showing a summary of the matched pairs transformations, the individual matched pairs and a histogram enabling you to see the distribution of changes in a property for the matched pairs.

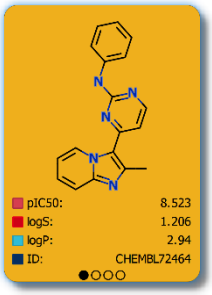
As the cards are still coloured by predicted hERG, we can see that the non-paired molecules in the grid at the bottom of the display have a lot of the lowest hERG activity values, as shown by the red cards (you may need to zoom out to see these). Some of these compounds still have activity at CDK2, so they have the potential to be exploited as new leads in this chemotype. The different coloured cards in the main networks show that variations in hERG activity can be achieved; also, similarly coloured cards with deep blue links connecting them indicate that the hERG activity and CDK2 pIC_{50} SAR can be separated. This provides a further indication of the potential developability of this chemotype.



- Sort the **Transformations** table by left-clicking twice on the header of the **Count** column and then select the **chlorine to hydrogen** change (which has 5 occurrences) so that we can focus on those pairs of molecules. The **ΔpIC50** column shows that this change is associated with a major change in activity (an average of 1.8 log units).

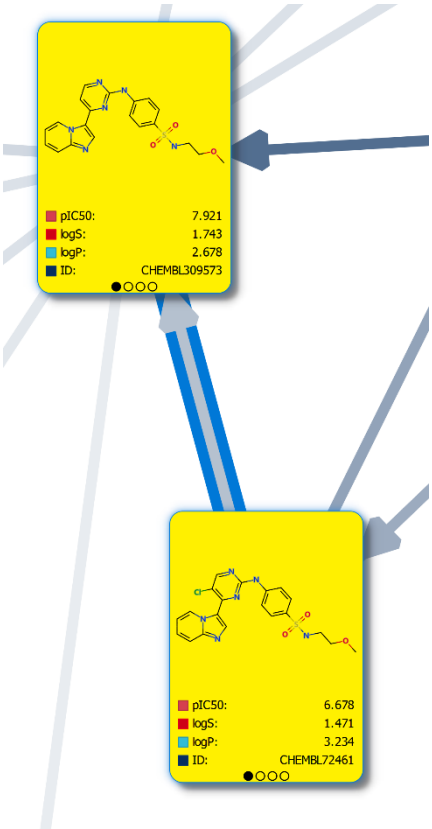


- As you click on the different examples of this transformation in the **Matched Pairs table** below, the network display in Card View zooms in to that pair of compounds.



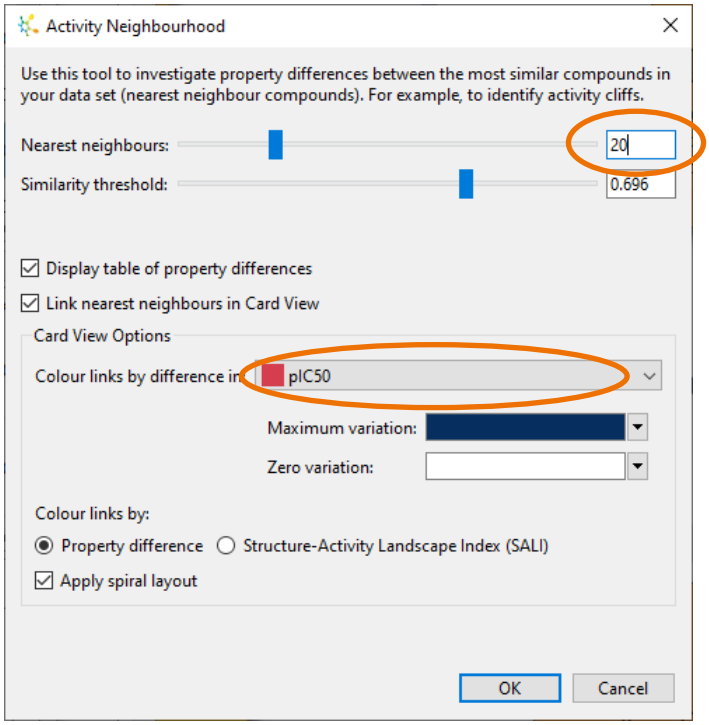
We can see that it is the unsubstituted phenyl ring derivative (CHEMBL72464) which is the best compound and the various chlorine substitutions are much less active.

One example of this transformation shows the chlorine changing to hydrogen on the pyrimidine rather than the phenyl ring (shown right).



The Matched Pairs analysis has highlighted single point changes that occur within the data set, but we are also interested to see if there are other changes that have a significant impact. For this, we first need to select a specific compound which in this case will be the unsubstituted phenyl compound, CHEMBL72464.

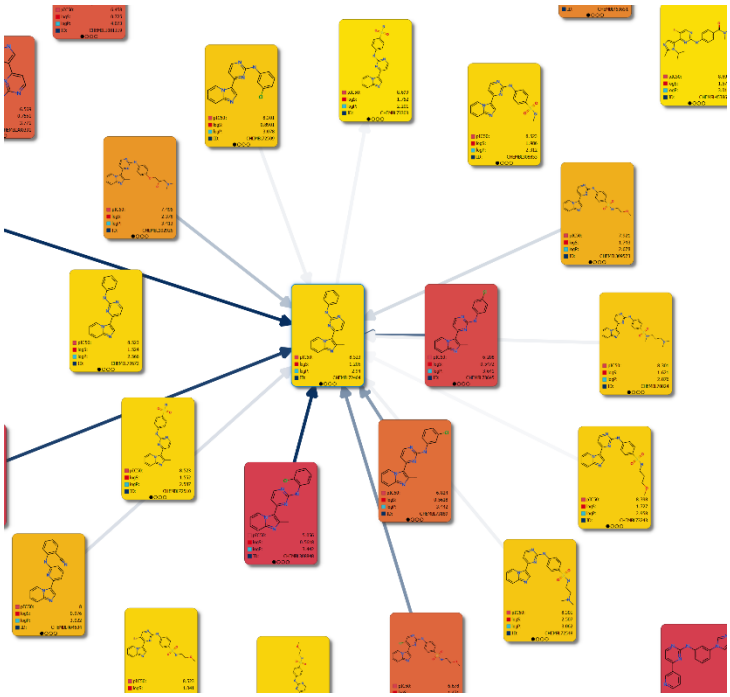
- Select the unsubstituted derivative (CHEMBL72464) in Card View, and from the **Analyse** menu at the top of Card view, select **Activity Neighbourhood**.
- In the **Activity Neighbourhood** dialogue, choose to colour the links by **pIC50**, choose **20** as the number of **Nearest neighbours** and change the colours of the **Maximum** and **Zero variations** to the blue and white as before (shown right).



The resulting display will have CHEMBL72464 at the centre and the other cards spiraling outwards in descending similarity.

- Click the **format** button  to bring up the **Colour By** dialogue again and colour the cards by the **pIC50** value, interpolating from **6** (red) to **9** (yellow)

The image below shows the resulting activity neighbourhood. From this analysis, we can see that not all substitutions on that phenyl ring are disfavoured as those compounds having a para-sulfonamide substituent are equally potent.

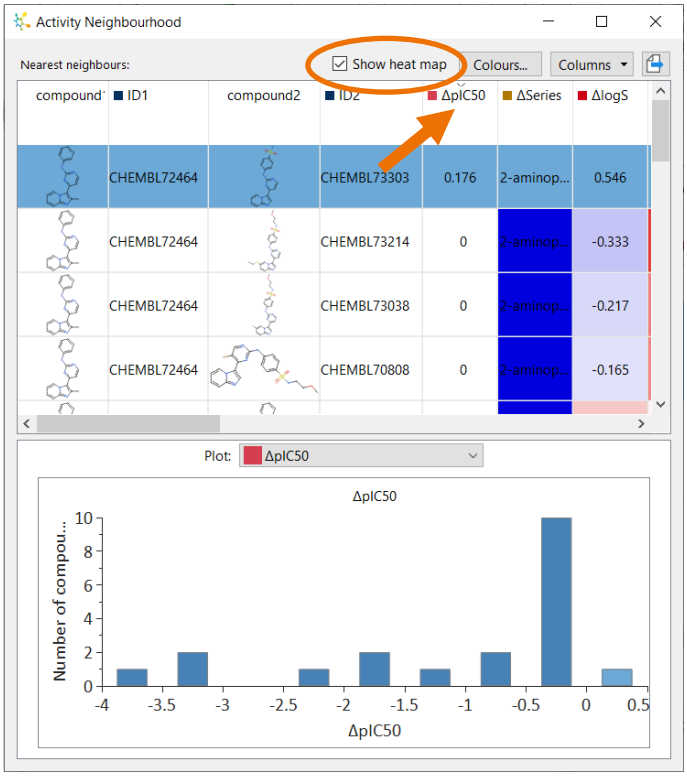


Note: Hover the mouse over any link to compare that pair of cards side-by-side.

In the **Activity Neighbourhood** dialogue, the **Nearest neighbours** table can be used to view the effects of these changes across the range of properties in the data table.

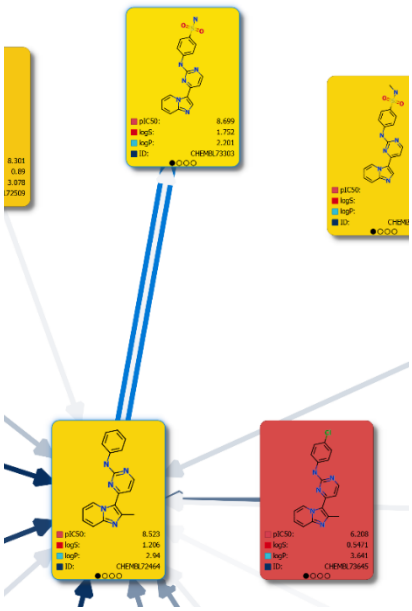
- **Left-click** twice on the header of the **ΔpIC50** column to sort the column such that the positive changes are at the top and then **check** the **Show heat map** box.

- Here we can see that there are changes that keep activity (the pale or white colours in the **ΔpIC50** column), decrease **logP** (the blue colour in the **ΔlogP** column) and also increase **logS** (the red colour in the **ΔlogS** column), indicating that these features can be optimised independently of each other in this series. Again, this is a positive indicator for the potential successful optimisation of this chemotype.



- **Click** on the first row in the **Nearest Neighbours** table, and the display will focus on that pair of molecules (see below).

This view can summarise the SAR in the series and can be captured by **right-clicking** on the view and selecting **Copy Image** from the menu, enabling you to paste this into a slide or document. Similar analyses can be done for the other chemotypes to produce a full report for an HTS triage.



Conclusions

With this worked example, we have demonstrated a number of approaches for using Card View along with some of the analyses within StarDrop to do HTS triaging. In this example, we have explored the subdivision of the hit set into chemotypes, assessment of those chemotypes by desirable and undesirable property distributions, identified SAR at the target and separated that SAR from the SAR of counter targets. As described, the results of these analyses can be easily captured for presentation within slides or documents.