

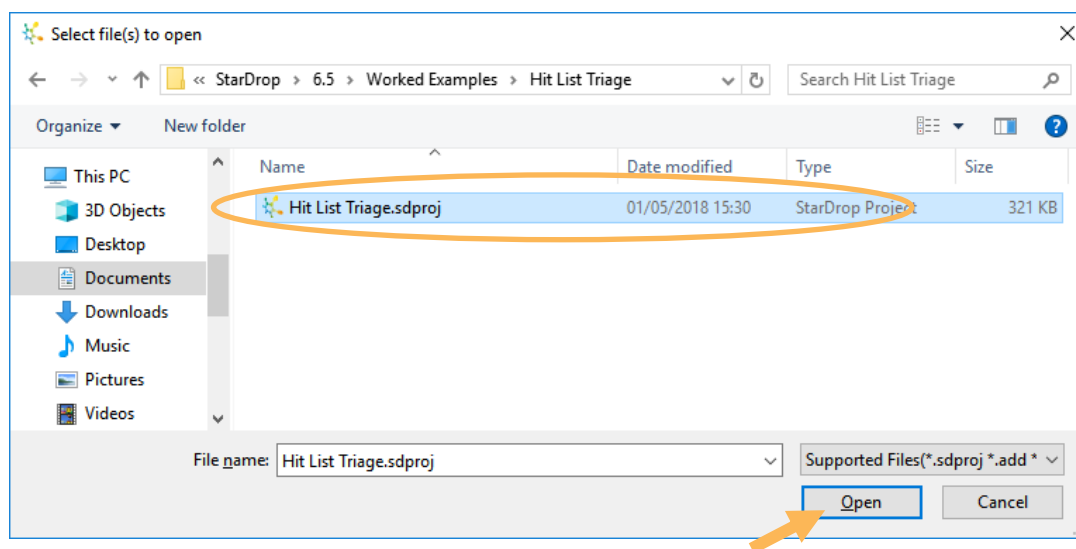
Worked Example:

Hit List Triage 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 pIC₅₀ data measured against the CDK2 kinase and obtained from the ChEMBL database <https://www.ebi.ac.uk/chembl/>

Exercise

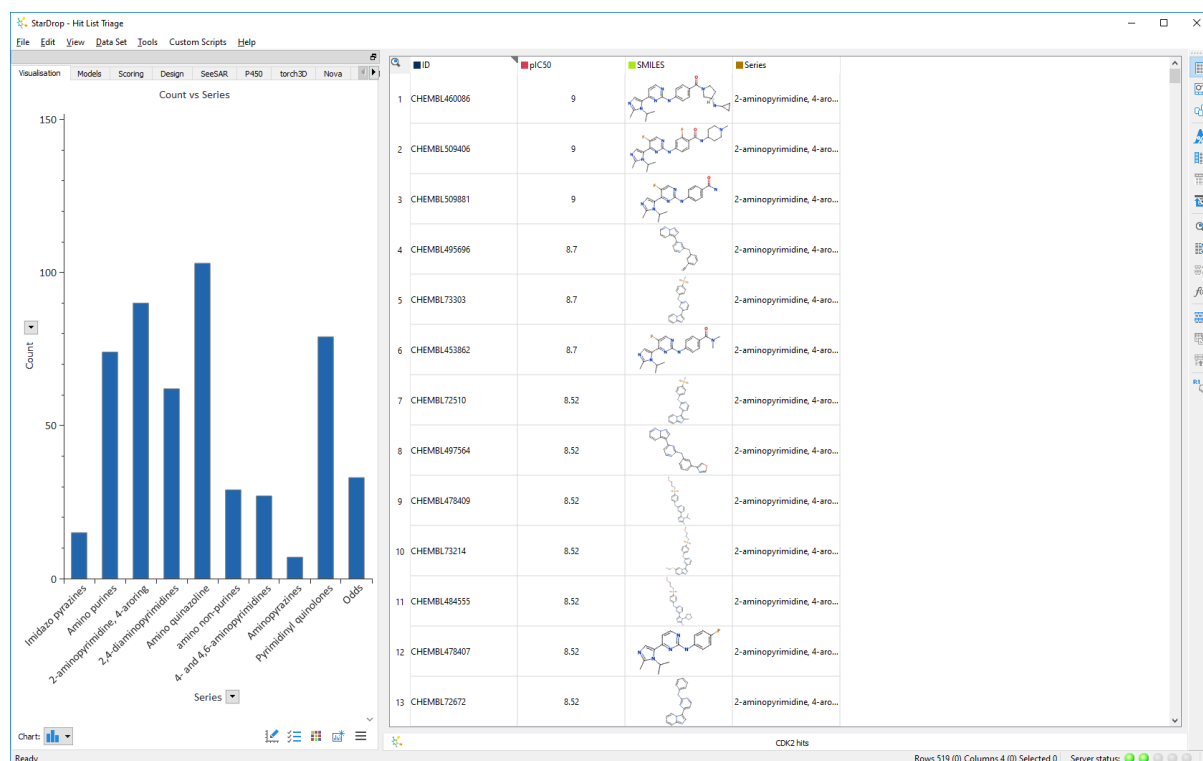
- From StarDrop's **File** menu, select **Open** and open the project file **Hit List Triage.sdproj**.



Optibrium™, StarDrop™, Card View™ and Glowing Molecule™ are trademarks of Optibrium Ltd.


© 2020 Optibrium Ltd.

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 at the bottom.

StarDrop - Hit List Triage

File Edit View Data Set Tools Custom Scripts Help

Visualisation Models Scoring Design SeeSAR P450 torch3D Nova

QSAR Models:

- ☒ StarDrop
 - ☒ logS
 - ☒ logS @ pH7.4
 - ☒ logP
 - ☒ logD
 - ☒ 2C9 pKi
 - ☒ hERG pIC50
 - ☒ BBB log([brain]/[blood])
 - ☒ BBB category
 - ☒ HIA category
 - ☒ P-gp category
 - ☒ ZD6 affinity category
 - ☒ PPB90 category
- ☒ MW
- ☒ HBD
- ☒ HBA
- ☒ TPSA
- ☒ Flexibility
- ☒ Rotatable Bonds
- ☐ Legacy models

Intensive Models:

- ☒ pKas
 - ☒ Most Acidic pKa
 - ☒ Most Basic pKa
 - ☒ All pKas

ID	pIC50	SMILES	Series	logS	logS @ pH7.4	logP
1 CHEMBL460086	9	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	1.07	1.66	3.98
2 CHEMBL509406	9	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	1.14	2.73	3.65
3 CHEMBL509881	9	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	1.07	1.07	2.54
4 CHEMBL495696	8.7	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	0.873	0.873	2.48
5 CHEMBL73303	8.7	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	1.75	1.75	2.2
6 CHEMBL453862	8.7	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	1.58	1.58	3.02
7 CHEMBL72510	8.52	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	1.55	1.55	2.59
8 CHEMBL497564	8.52	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	0.793	0.793	3.39
9 CHEMBL478409	8.52	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	1.99	1.99	3.28
10 CHEMBL73214	8.52	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	0.873	0.873	3.54
11 CHEMBL484555	8.52	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	1.51	1.51	3.8
12 CHEMBL478407	8.52	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	0.891	0.891	3.4
13 CHEMBL72672	8.52	<chem>Nc1nc2c(ncn2C(=O)O1)c3ccccc3</chem>	2-aminopyrimidine, 4-aro...	1.52	1.52	2.57

Ready Rows 519 (0) Columns 22 (0) Selected 0 Server status: ●

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.

StarDrop - Hit List Triage

File Edit View Data Set Tools Custom Scripts Help

Visualisation Models Scoring Design SeeSAR P450 torch3D Nova

QSAR Models:

- ☒ StarDrop
 - ☒ logS
 - ☒ logS @ pH7.4
 - ☒ logP
 - ☒ logD
 - ☒ 2C9 pKi
 - ☒ hERG pIC50
 - ☒ BBB log([brain]/[blood])
 - ☒ BBB category
 - ☒ HIA category
 - ☒ P-gp category
 - ☒ ZD6 affinity category
 - ☒ PPB90 category
- ☒ MW
- ☒ HBD
- ☒ HBA
- ☒ TPSA
- ☒ Flexibility
- ☒ Rotatable Bonds
- ☐ Legacy models

Intensive Models:

- ☒ pKas
 - ☒ Most Acidic pKa
 - ☒ Most Basic pKa
 - ☒ All pKas

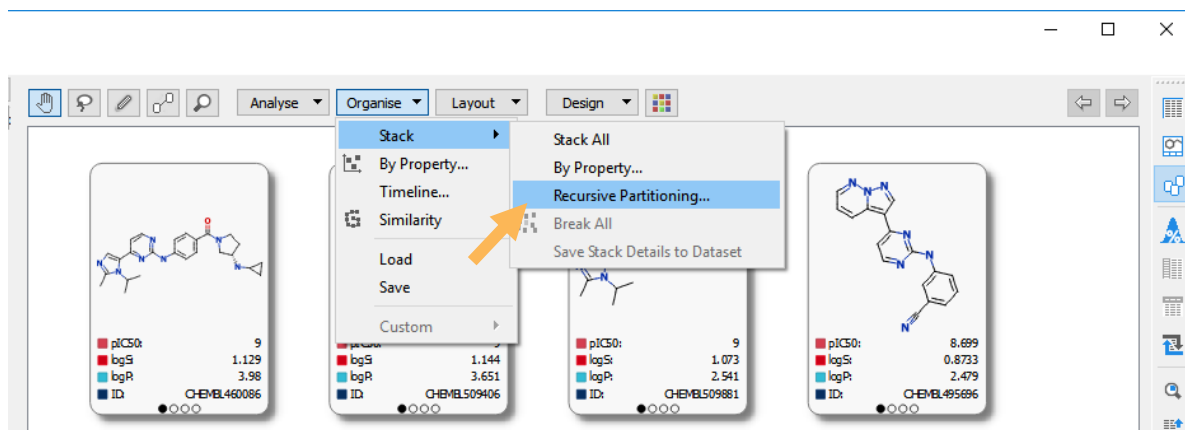
ID	pIC50	logS	logS @ pH7.4	logP
CHEMBL460086	9	1.07	1.66	3.98
CHEMBL509406	9	1.14	2.73	3.65
CHEMBL509881	9	1.07	1.07	2.54
CHEMBL495696	8.7	0.873	0.873	2.48
CHEMBL73303	8.7	1.75	1.75	2.2
CHEMBL453862	8.7	1.58	1.58	3.02
CHEMBL72510	8.52	1.55	1.55	2.59
CHEMBL497564	8.52	0.793	0.793	3.39
CHEMBL478409	8.52	1.99	1.99	3.28
CHEMBL73214	8.52	0.873	0.873	3.54
CHEMBL484555	8.52	1.51	1.51	3.8
CHEMBL478407	8.52	0.891	0.891	3.4
CHEMBL72672	8.52	1.52	1.52	2.57

Ready Rows 519 (0) Columns 22 (0) Selected 0 Server status: ●

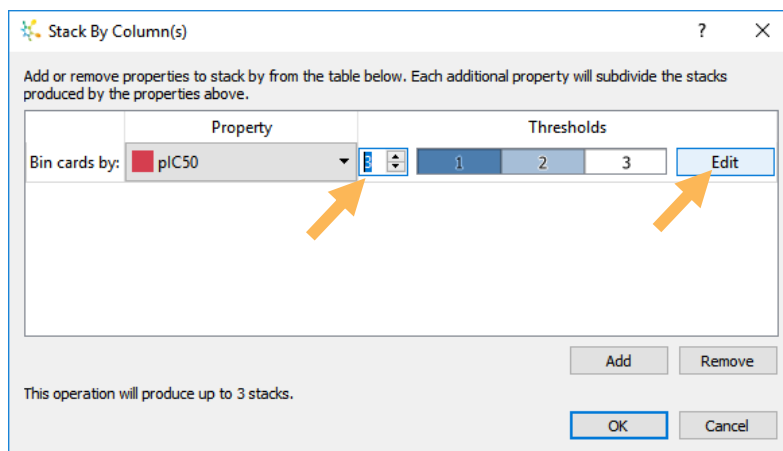
You can move the view of the cards by dragging the view. You can zoom in or out by using a mouse wheel or the **CTRL** and **-/=** keys. You can move an individual card by dragging it to a new location.

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

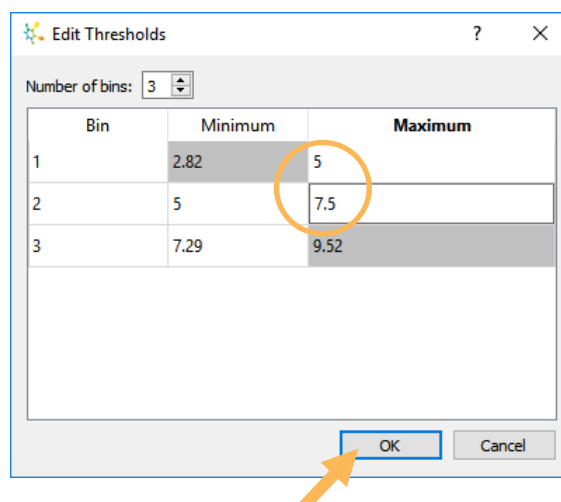
- From **Organise** menu at the top of Card View, select **Stack**, then **Recursive Partitioning**.



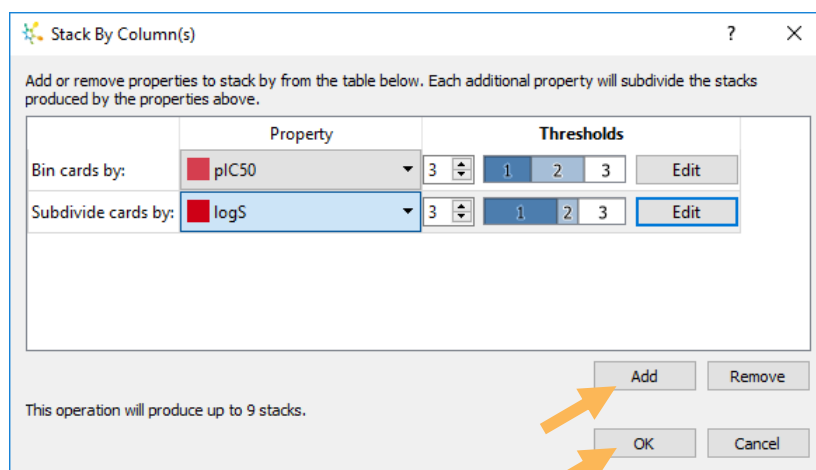
- Choose **pIC50** as the property by which to bin the cards and increase the number of bins to **3**.
- Click the **Edit** button to modify the ranges.



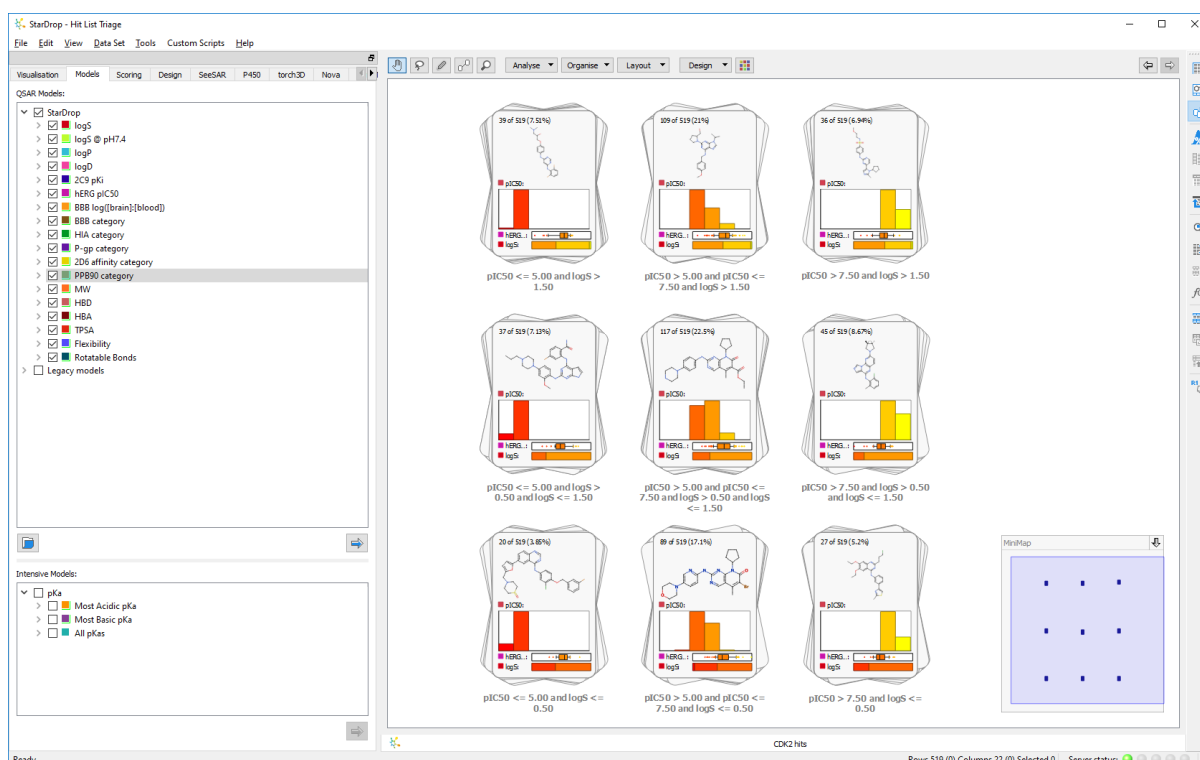
- Change the boundary values to **5** and **7.5** and click **OK**.



- 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.



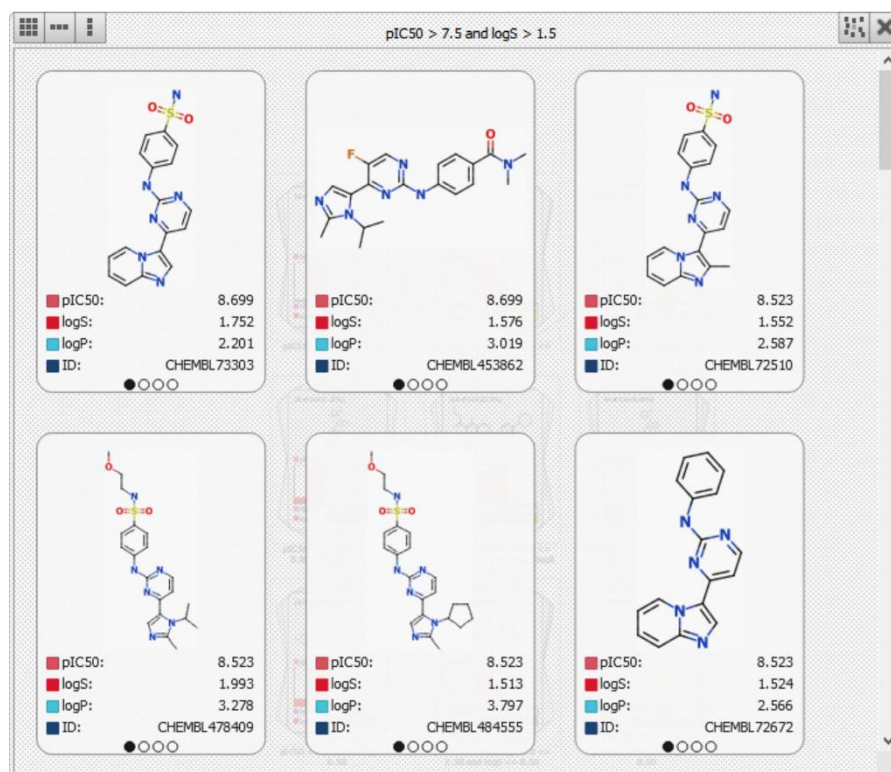
- Click **OK** to stack the cards using these properties.



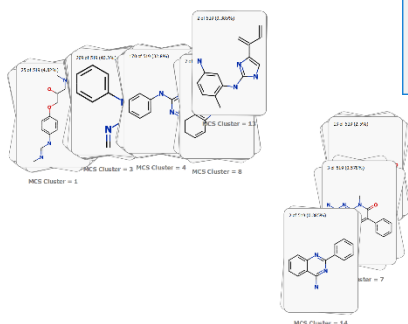
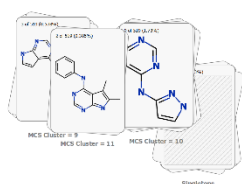
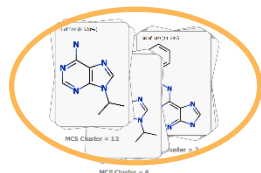
The top right-hand stack has interesting compounds (38 out of 519) because they are potent and predicted to be soluble.

- 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.

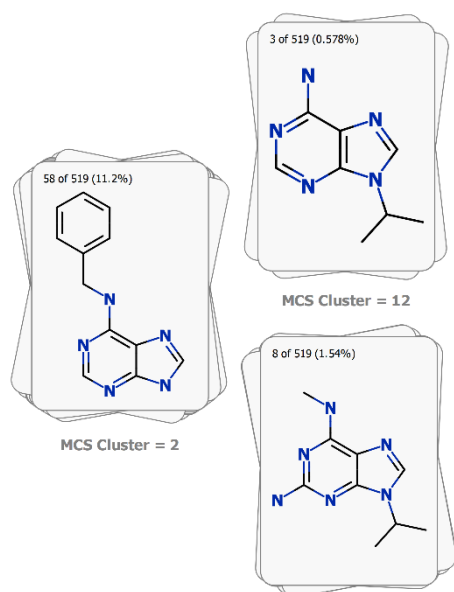


- From the **Analyse** menu at the top of Card View select **Clustering**.
- Choose the **Common Substructure** option, enter a similarity value of **0.4** and click the **Cluster** button.




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 3 stacks on the right that all have various decorations off an amino purine core.




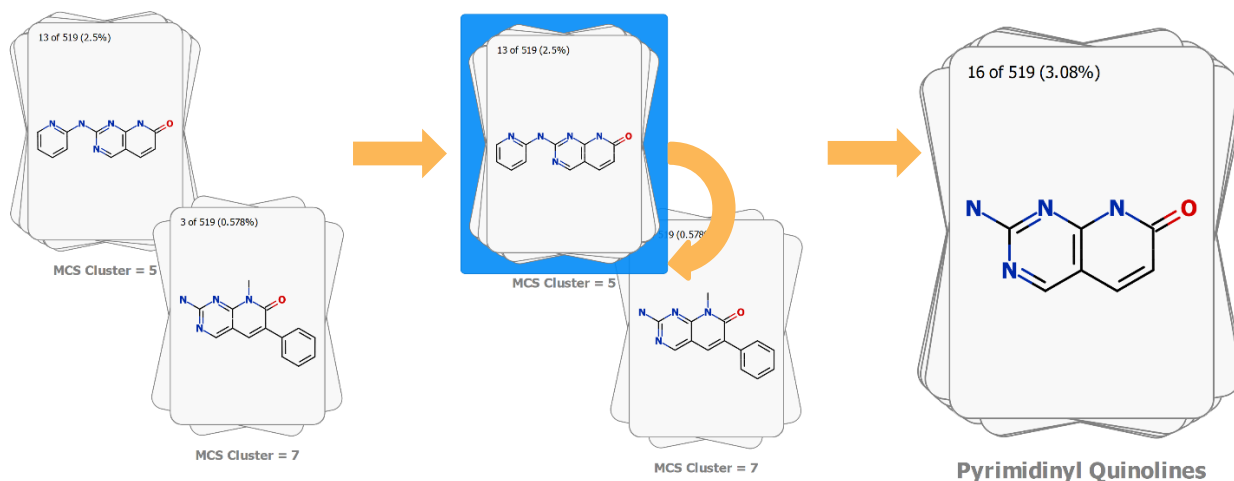
These 3 stacks could be thought of as a single ‘amino purine’ series and can be easily combined manually into one stack.

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 3 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.

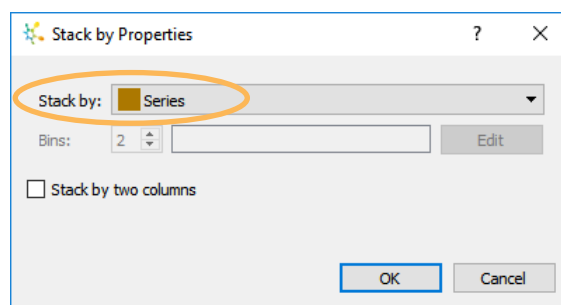
Here with only 2 stacks we can use the **hand tool**  to simply drag one on top of the other as shown.



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 -> Save stack details to dataset** from the menu.

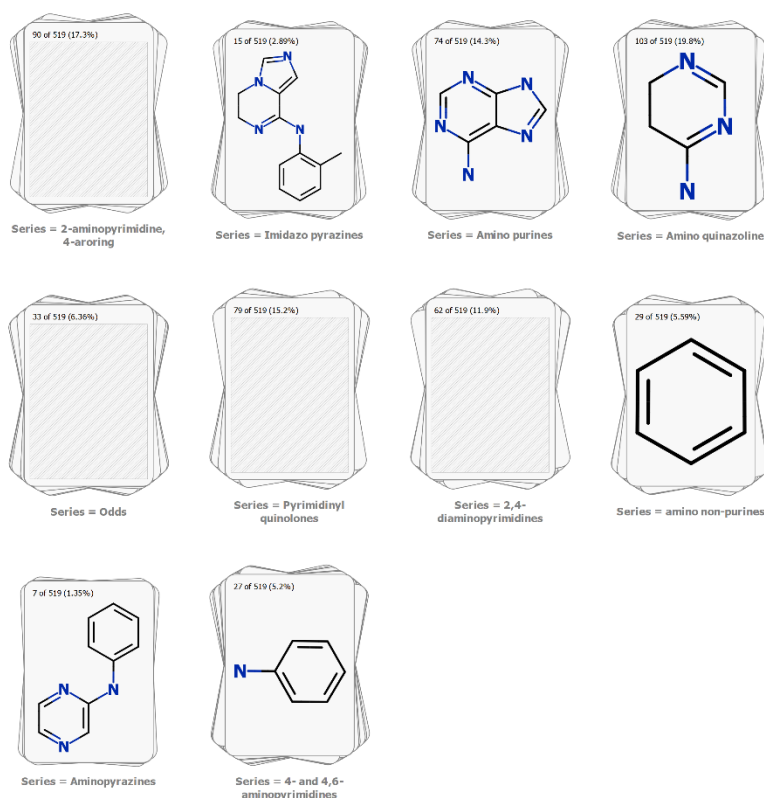
To save time this process has already been done for you in this data set and the stack names for the chemotypes are provided in the **Series** column.

- 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 **OK**.



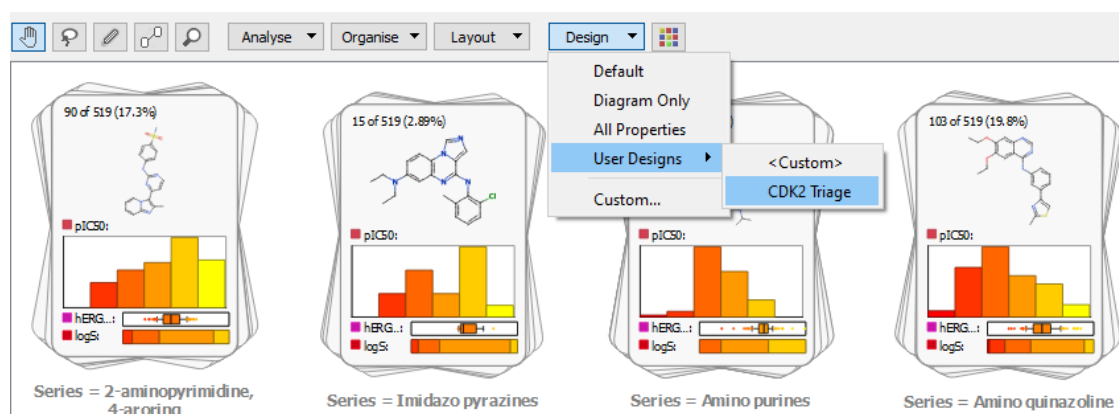
This stacks the cards into 10 stacks which have the common substructure detected within the stack. As the series definitions here are human derived there are some stacks that do not have a large enough common substructure for display.

- 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**.



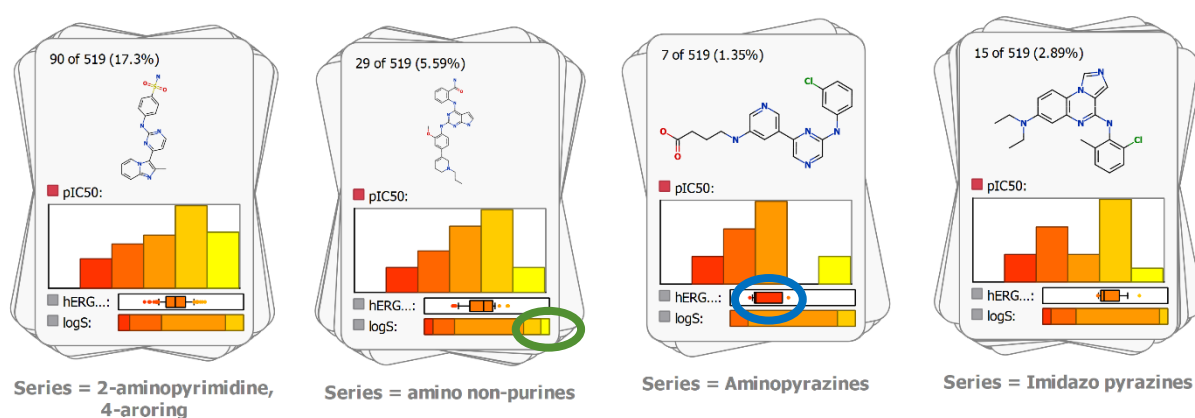
More relevant information about the property distributions for the cards in the stacks and a representative compound from the stack can be seen by changing the card template.

- 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 pIC₅₀ 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 pIC₅₀?”

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).



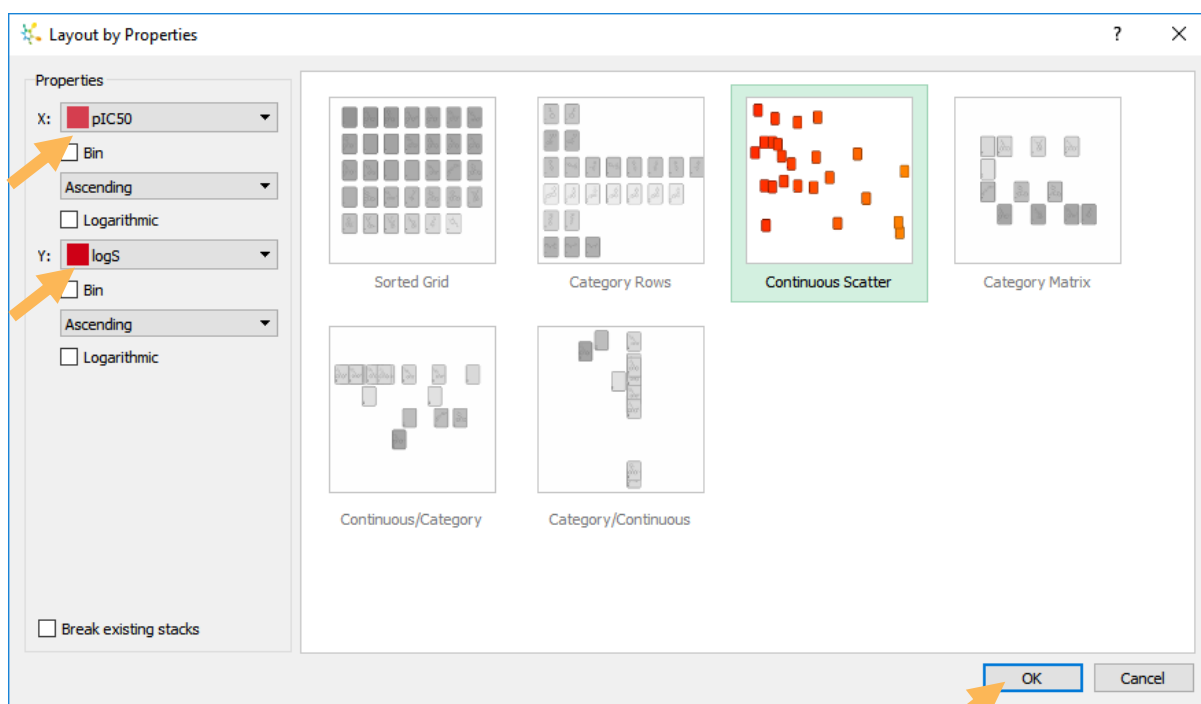
Note: You can, at any time, **right-click** and **Copy Image** of the stacks or cards displayed in order to paste the view into a slide for presentation purposes.


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.

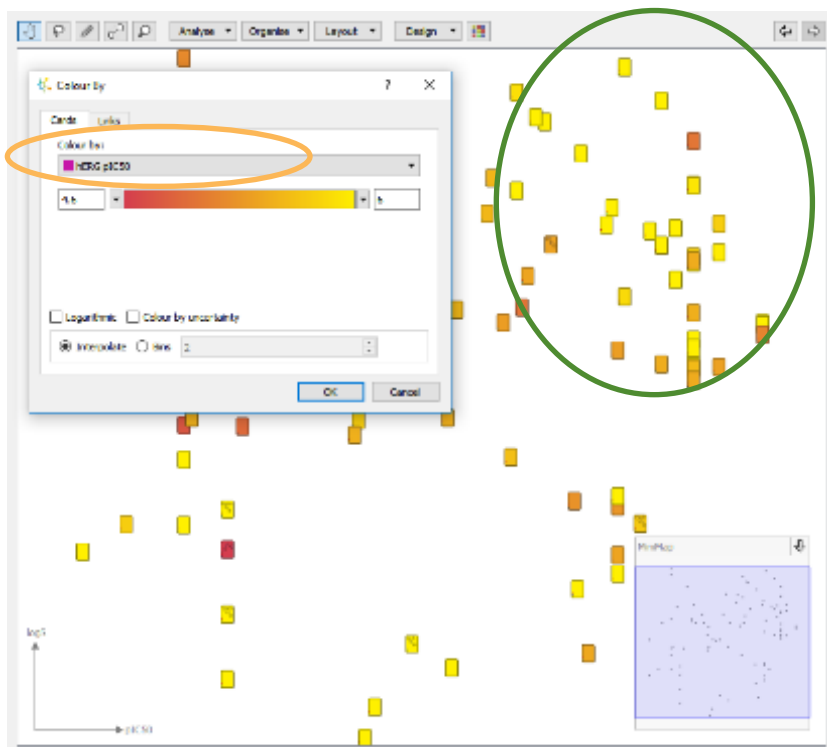
- Select the ‘2-aminopyrimidine, 4-aroring’ stack and click the **Create new data set from selection** button  on the right-hand toolbar (or from the **Data Set menu** choose **Create from selection**).

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 **OK**.



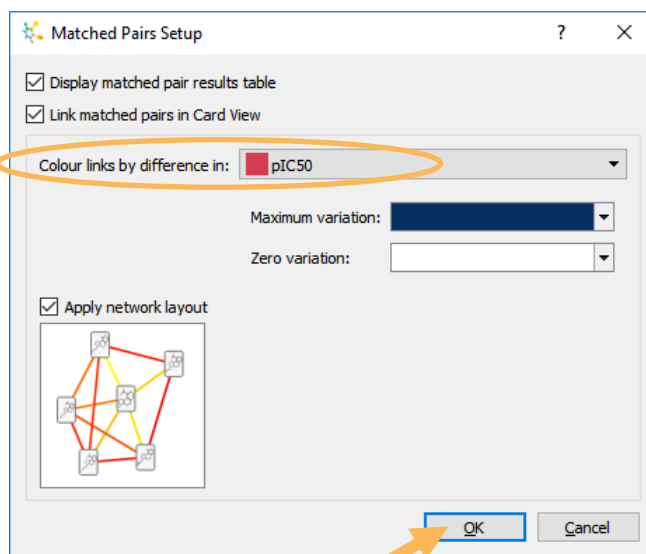
- These cards can be coloured by their predicted **hERG pIC₅₀** using the **format button** . In the **Colour By** dialogue select **hERG pIC₅₀** and change the interpolation scale to go from **4.6** to **6** as shown right.



Note that the top right-hand corner of the display (which corresponds to good pIC₅₀ 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.

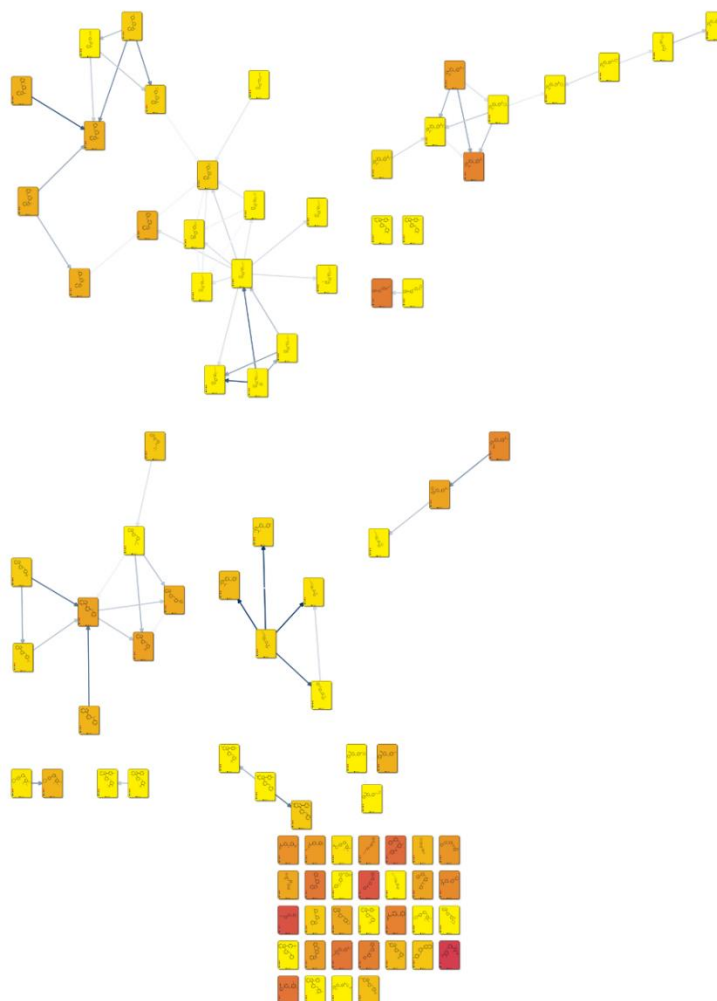
The more detailed SAR in this chemotype can be investigated by finding the matched pairs within the series.

- 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 **OK** 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 show 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



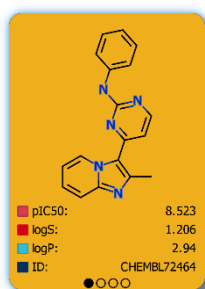
and CDK2 pIC₅₀ SAR can be separated. This provides a further indication of the potential developability of this chemotype.

- Sort the **Transformations** table by clicking 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 Δ pIC₅₀ column shows that this change is associated with a major change in activity (an average of 1.8 log units).

Transformations:		Count	Δ pIC ₅₀	Δ Series	Δ logS	Δ logP
Cl → *	H	5	1.79	0	0.571	0.571
* → *	H	5	0.475	0	0.262	0.262
* → F	H	5	0.117	0	0.626	0.207
H	* → F	5	-0.117	0	-0.626	-0.207
H	* → *	5	-0.475	0	-0.262	-0.262

Matched Pairs:		Transfo	Δ pIC ₅₀	Δ Series	Δ logS	Δ logP
compound1	compound2	16	3.47	2-amino...	0.644	0.644
		16	2.31	2-amino...	0.659	0.659

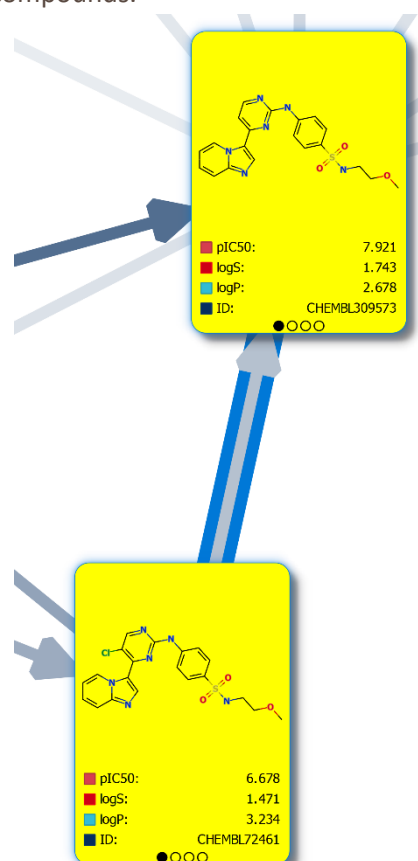
- As you click on the different examples of this transformation in the **Matched Pairs** table below, the network display in Card View zooms in to that pair of compounds.



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 a 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 which 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.

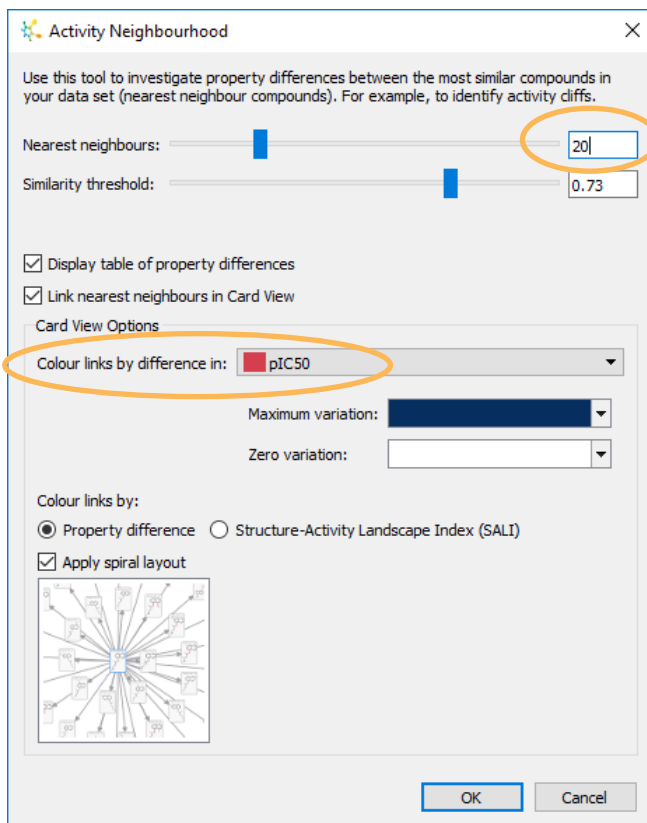


- Select the unsubstituted derivative (CHEMBL72464) in Card View, and from the **Analyse** menu at the top of Card view select **Activity Neighbourhood**.

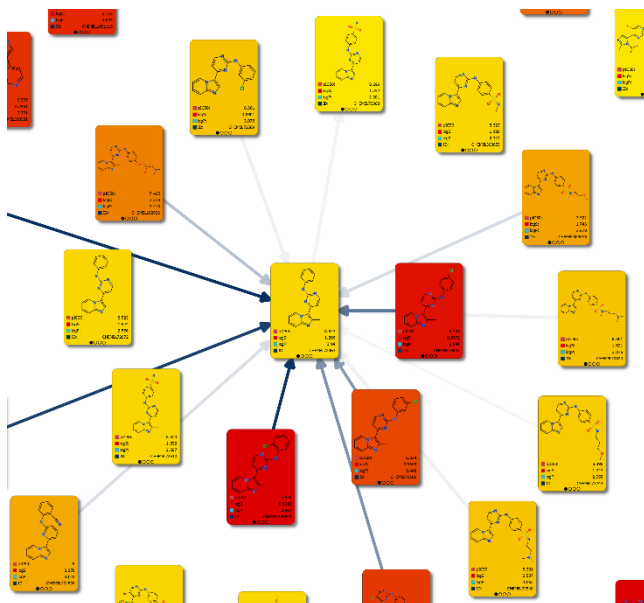
- In the **Activity Neighbourhood** dialogue choose to colour the links by **pIC50**, choose **20** as the number of **Nearest neighbours** and change the colours of the **Maximum** and **Zero** variations to the blue and white as before (shown right).

The resulting display will have CHEMBL72464 at the center 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.



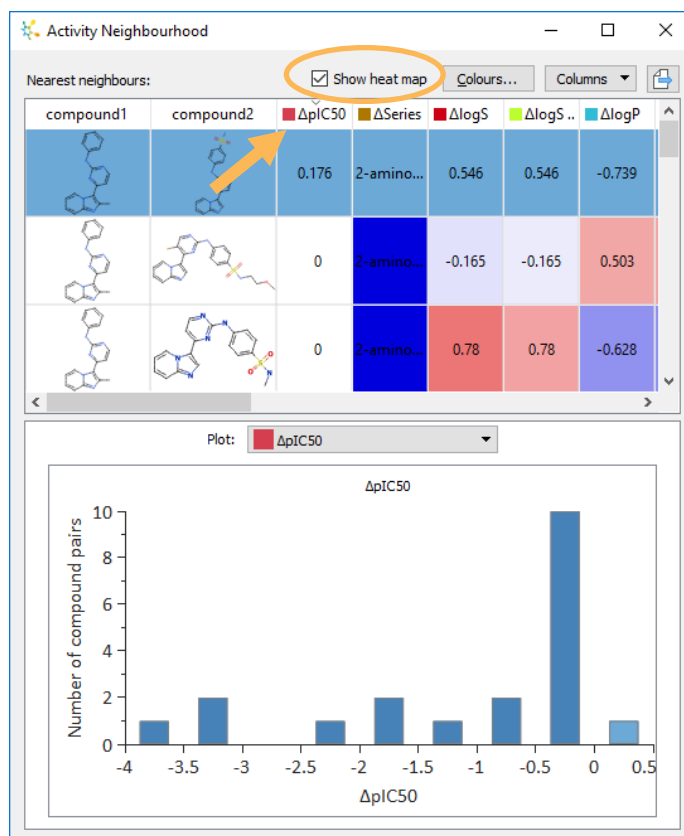
Note: Hover the mouse over any link to compare that pair of cards side-by-side.

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 **Δ pIC50** 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 which keep activity (the pale or white colours in the ΔpIC_{50} column), decrease $\log\text{P}$ (the blue colour in the $\Delta\log\text{P}$ column) and also increase $\log\text{S}$ (the red colour in the $\Delta\log\text{S}$ 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.

