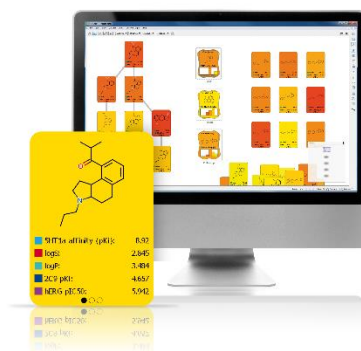


Worked Example:

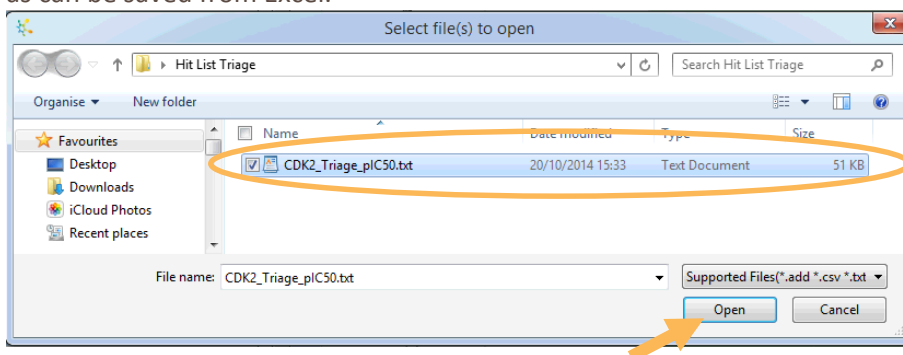
Hit List Triage Using Card View™



This example is taken from a kinase project in which a large screening campaign has resulted in a hit list. The project team wish to evaluate the list to identify the chemotypes within and focus their resources on a small number of series that have demonstrable SAR at the target and 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

- Start the StarDrop application
- From the **File->Open** menu, open the data file **CDK2_Triage_pIC50.txt**. This is a text file such as can be saved from Excel.

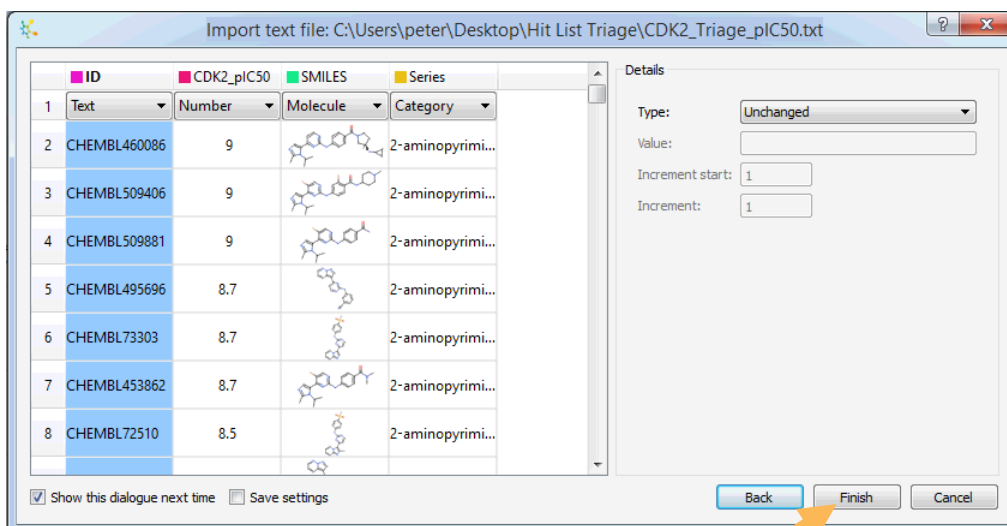


The file import dialogue box shows that the text file contains four columns: the first contains compound identifiers, the second, experimental data for the potency (pIC₅₀) against Human Cyclin-Dependent Kinase isoform 2, the third has compound structures and the last is a chemical series tag.


- Click the **Next** button and then **Finish** to complete the import and the data set will be displayed in StarDrop

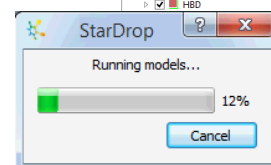
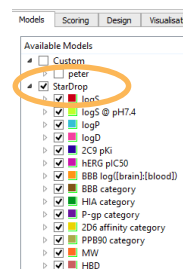


Optibrium™, StarDrop™, Card View™ and Glowing Molecule™ are trademarks of Optibrium Ltd.




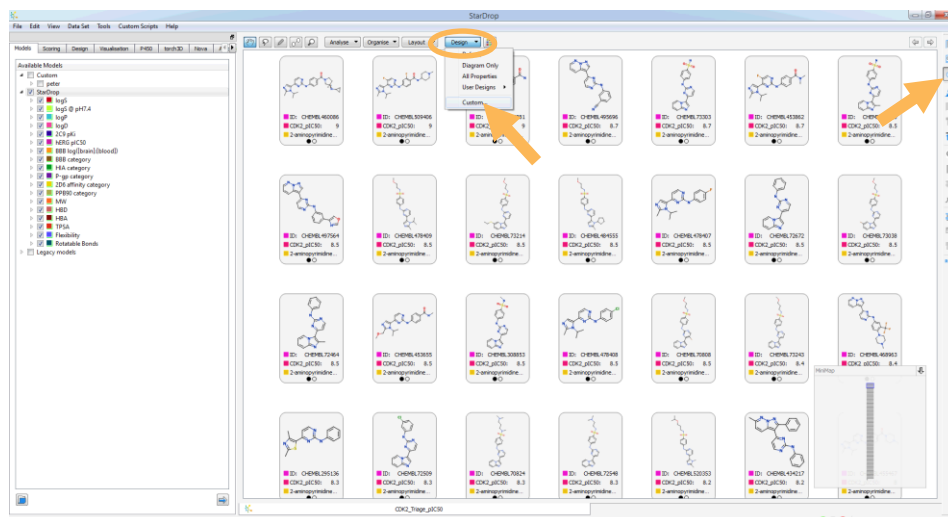
We have no further measured data at this stage of the project, so we're going to generate some predictions of ADME and physicochemical properties.


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

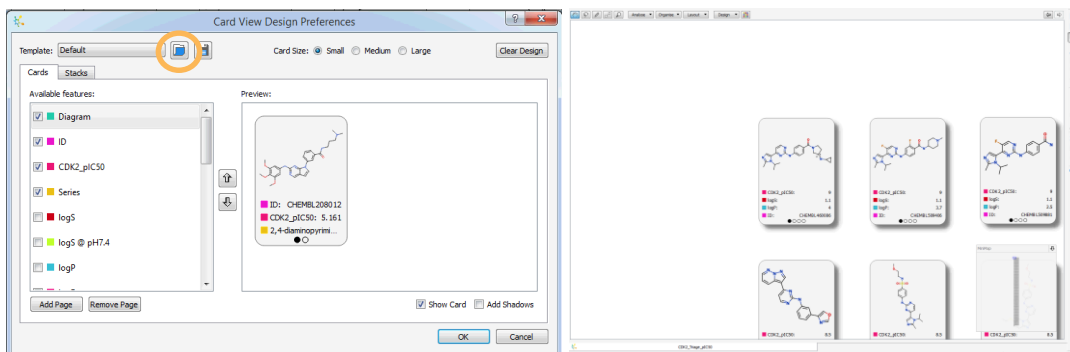


We will now change from the Table View of the data into Card View.

- Click the Card View button  on the toolbar. Click the **Design** button in Card View and select **Custom...** from the menu that appears to open the **Card View Design Preferences** dialogue.



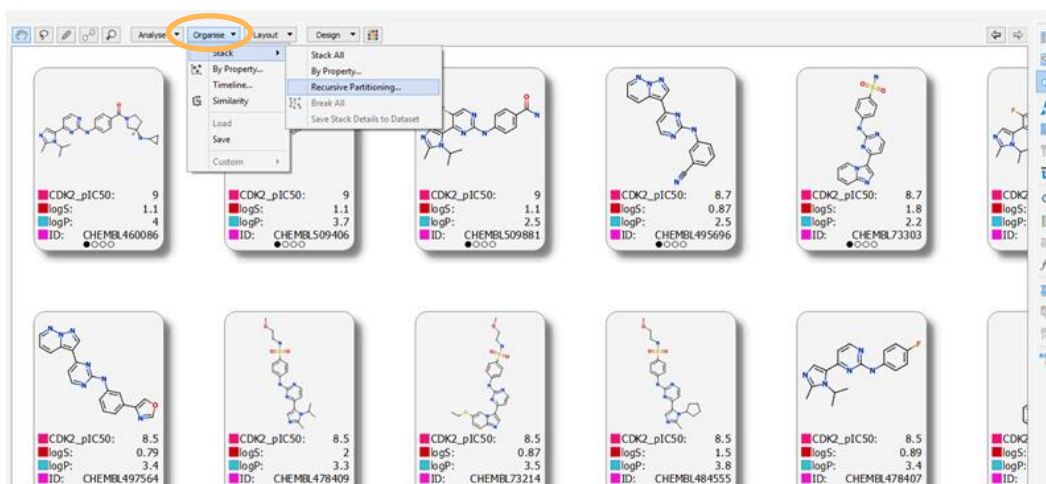
- Click the  button to load in the saved card and stack design template file called **CDK2_Triage_Design.xml** and then click **OK**.



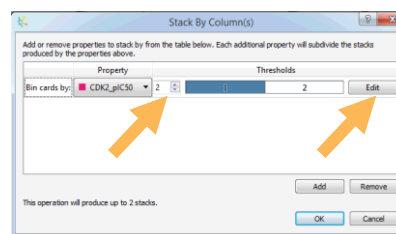
The cards are now displayed using the selected design and the first card is positioned in the centre of the screen; you can move the view of the cards by holding the left mouse button down and dragging the view; you can zoom in or out by using the middle mouse wheel (or holding the **CTRL** key and pressing either the '-' or '=' key). You can move an individual card simply by clicking on it and 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.

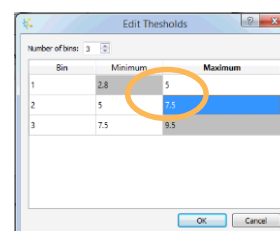
- From **Organise** menu at the top of Card View, select **Stack -> Recursive Partitioning**



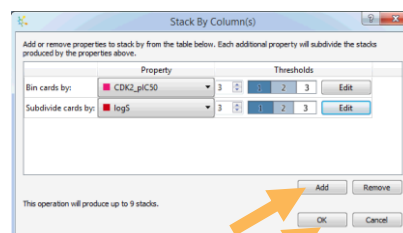
- Choose **CDK2_pIC50** as the property to bin the cards by and increase the number of bins to **3**.



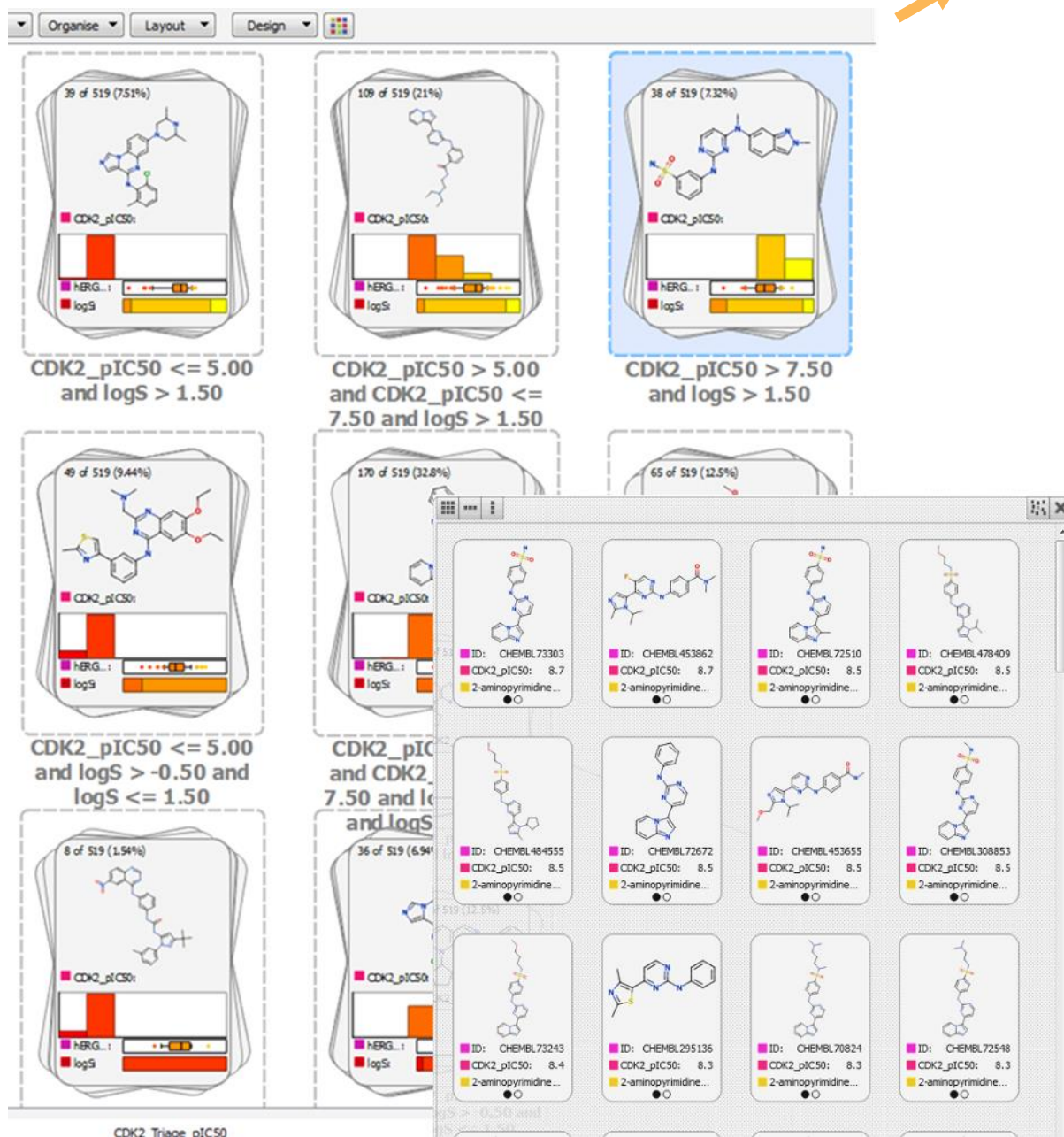
- Click the **Edit** button and change the boundary values to **5** and **7.5**.



- Click the **Add** button to add an additional property of **logS** and then split this also into three bins with -0.5 and 1.5 as boundary values.



- Click **OK** to perform the stacking.

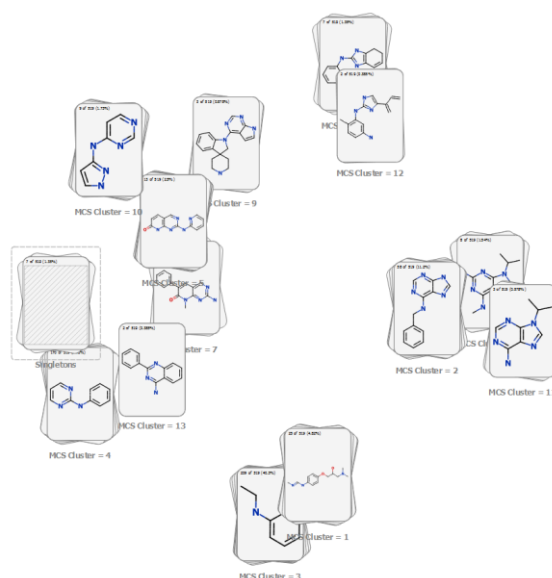
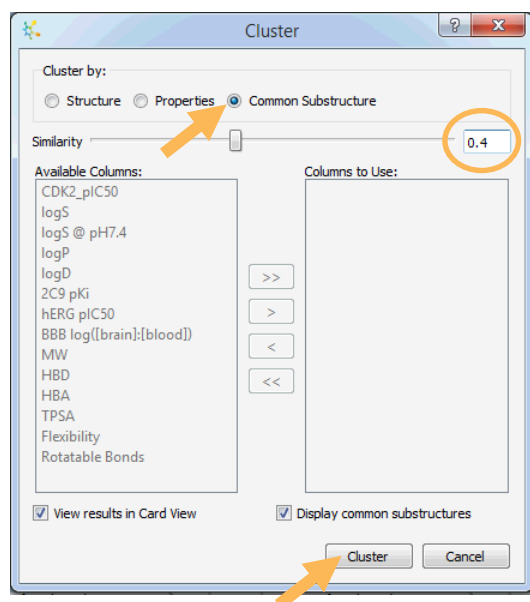


- The top right hand stack has interesting compounds (38 out of 519) because they are affine and predicted to be soluble. **Double click** (or right click and choose **Inspect**) on the stack to look at the cards within that stack.

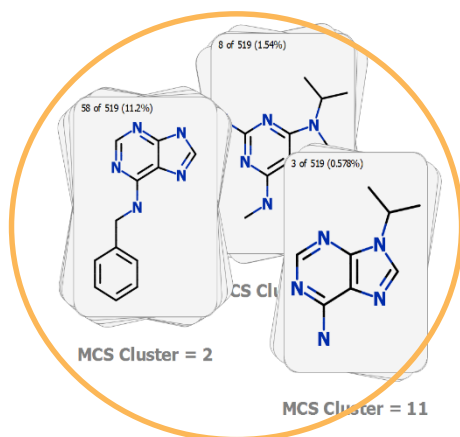
There are a variety of chemical scaffolds in this desirable compounds stack which might increase the project's chance of success, but we would like to know how these series are distributed across the

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 **Cluster by: Common Substructure** option; enter a similarity value of **0.4** and click **Cluster**.




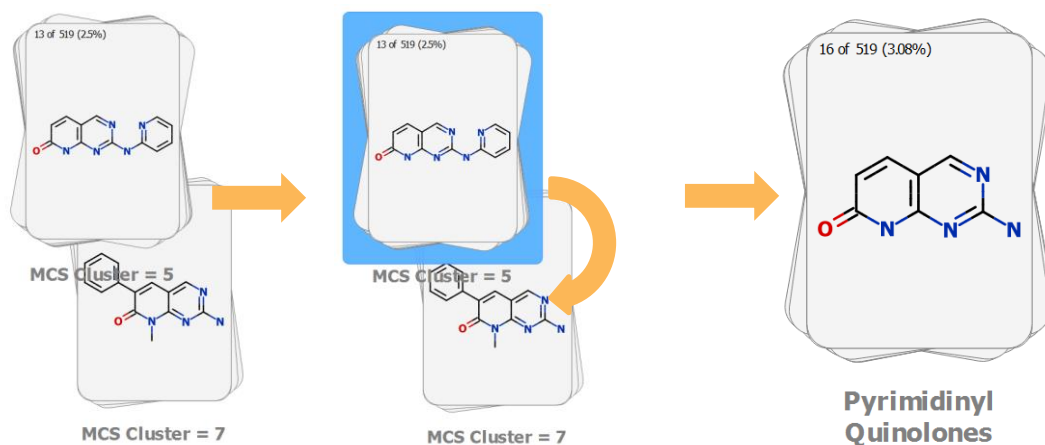
The display will then show the whole data set arranged similarly to that shown above right. Each of the stacks represents a set of compounds which share a common substructure, which is displayed on top of the stack. You can see that there is a group of 3 stacks on the right that all have various decorations 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.

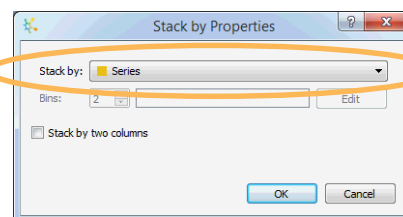
- Choose the **lasso tool**  from the Card View toolbar and draw a ring around the 3 stacks. The selected stacks can now be combined by **right clicking** on the selection and choosing '**Stack Selection**' from the menu.
- The new common substructure of a mono amino purine is now displayed and this stack can be named "amino purines" by selecting and editing the "MCS Cluster =" label below the stack and replacing the text (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.

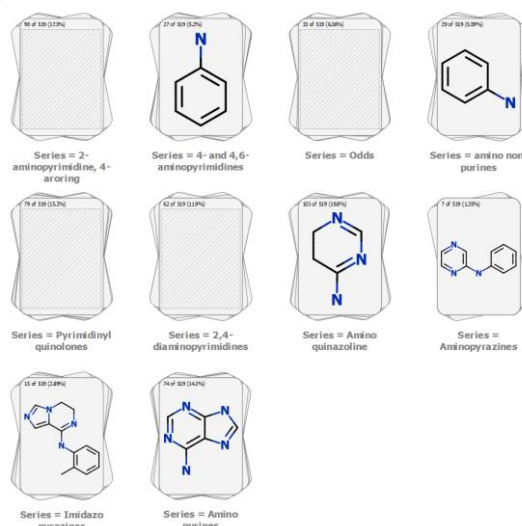


Once all the data set has been allocated to a particular stack then the names of the stacks can be added to the data table as a new category column. (**Hint:** If you wish to do this then click the **Organise** button and select **Stack -> Save stack details to dataset** from the menu.) To save time this process has already been done for you in this data set and the stack names for the chemotypes are entered under the **Series** column.

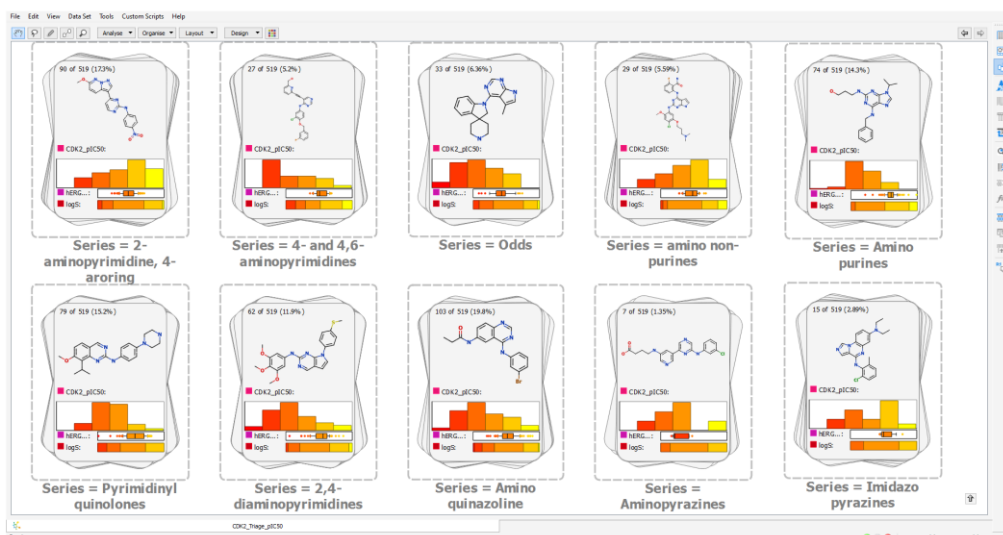
- To view the data set with these defined series, click the **Organise** menu at the top of Card View and choose **Stack -> By property**. In the **Stack by Properties** dialogue that appears choose the **Series** column from the drop-down list. This stacks the cards into 10 stacks which have the common 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**.

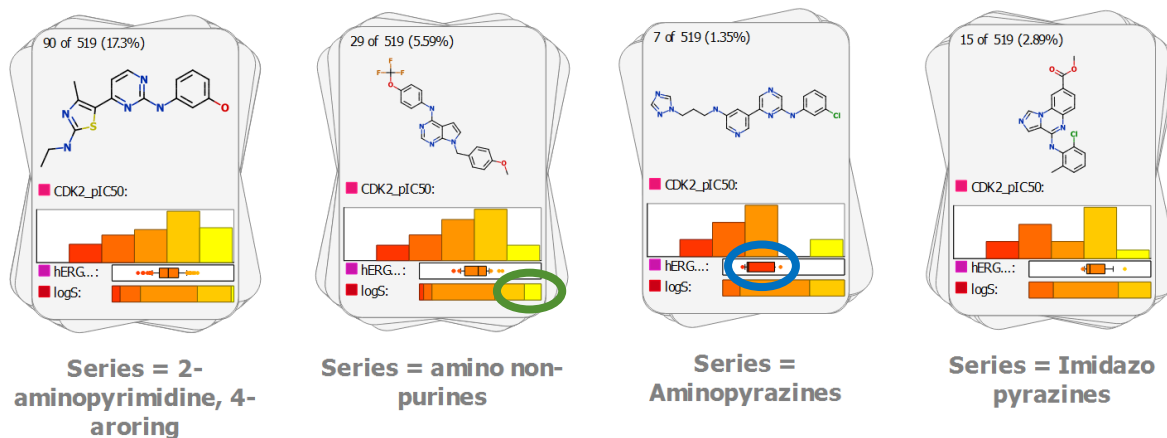


- More relevant information about the property distributions for the cards in the stacks and a representative compound from the stack can be seen by applying the **CDK2_Triage_design** template again (click the **Design** menu at the top of Card View and choose **User designs -> CK2_Triage_design** from the menu).




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 ringed area) whilst the ‘aminopyrazines’ appear to have the better hERG profile (as indicated by the blue ringed area).



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.

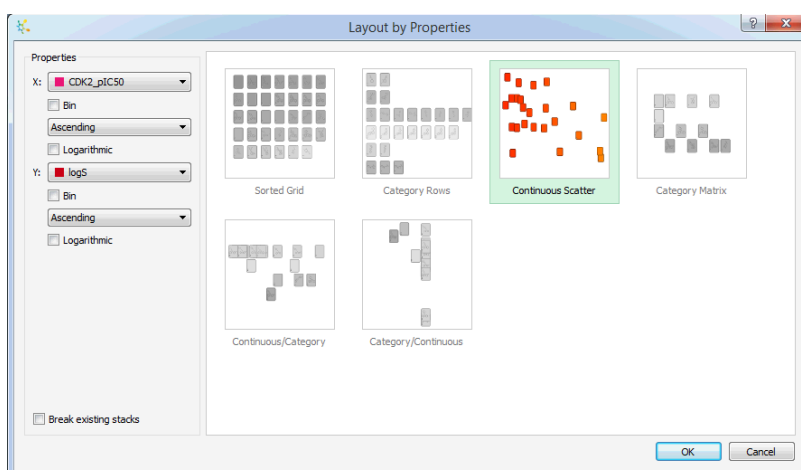
Next we shall examine one of these clusters to see what SAR points can be learned from a particular series. A co-crystal structure of one of the compounds in the ‘2-aminopyrimidine, 4-aroring’ stack has been published (PDB code 1OIT) and so with the possibility of using the 3D information available,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


- Left click on the '2-aminopyrimidine, 4-aroring' stack to select it and then create a new set from the selection using the  button on the toolbar or using the **Data Set -> Create from selection..** menu.

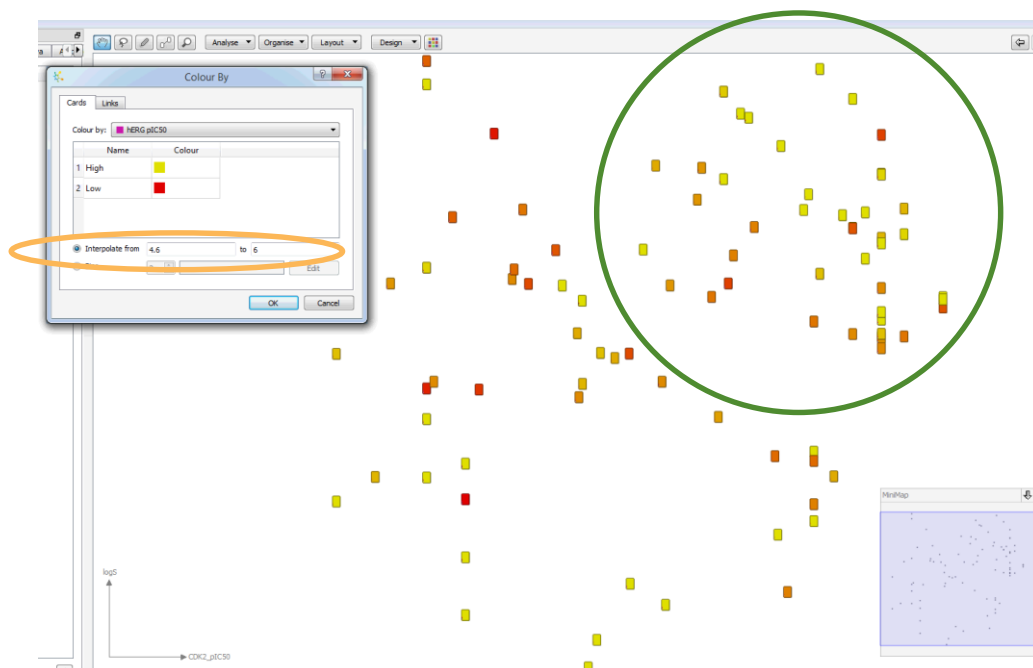
- Rename the set as you wish and then click the  button on the toolbar to view this data set in Card View.

- You can show how properties are distributed across this chemotype by clicking the **Organise** menu at the top of Card View and choosing **By Property**.

- Set **CDK2_pIC50** as the X direction property and **logS** as the Y direction property and click **OK**.



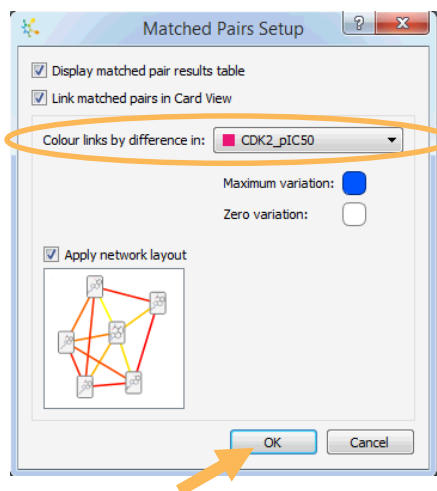
- These cards can be coloured by their predicted **hERG pIC50** using the colour button . In the **Colour By** dialogue select **hERG pIC50** and change the interpolation scale to go from 4.6 to 6 as shown below.



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 would mean 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 identify alternative structures within this series, with a better balance of properties, to act as more useful starting points for 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.

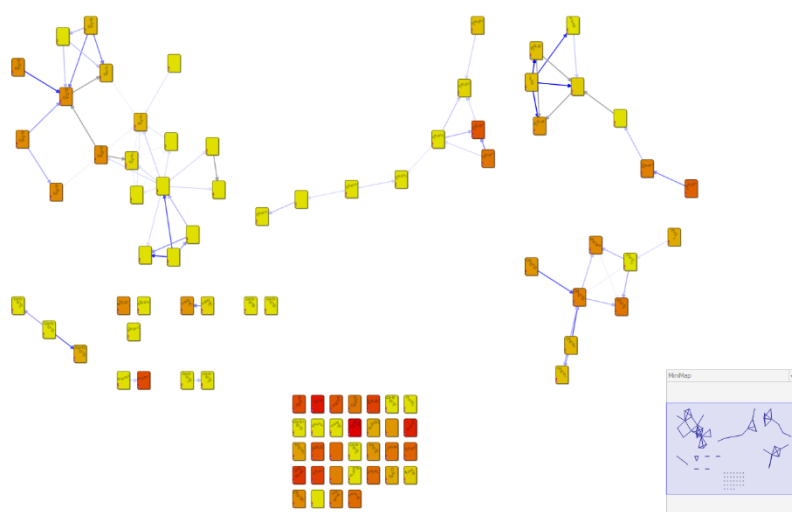
The more detailed SAR in this chemotype can be investigated by finding the matched pairs within the set.



- Click the **Analyse** menu at the top of Card View and choose **Find Matched Pairs**.
- Colour the links by *differences* in the **CDK2_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 pair analysis.

The matched pair analysis will run and the results will be shown in Card View as a network in which each matched pair is shown as a link between the corresponding cards. A table will also be created showing a summary of the matched pair 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 (move the visible area, if necessary, using the MiniMap in the bottom-right corner or zoom with the mouse wheel). Some of these compounds still have activity at CDK2 so they have the potential to be exploited



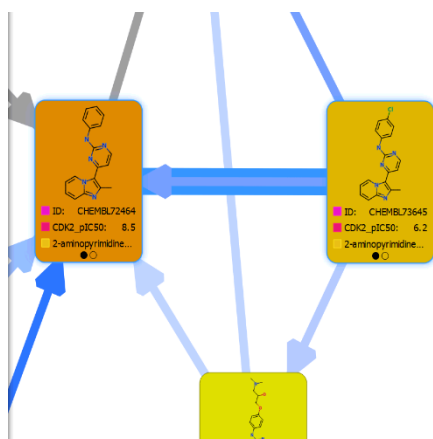
as new leads in this chemotype. The coloured cards in the main networks show that variations in hERG activity can be achieved; also similarly coloured cards with deep blue links connecting them indicate that the hERG activity and CDK2 pIC₅₀ SAR can be separated. This provides a further indication of 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 $\Delta\text{CDK2_pic50}$ column shows that this change is associated with a major change in activity (an average of 1.8 log units).

Matched Molecular Pairs									
Transformations:		Count	ΔCDK2	ΔSeries	ΔlogS	ΔlogP	ΔlogD	$\Delta\text{C9 pi}$	
H		6	-0.14	0	-0.27	-0.27	0.56	0.56	0.24
H		6	0.14	0	0.27	0.27	-0.56	-0.56	-0.24
H		5	-0.12	0	-0.63	-0.21	0.13	0.13	0.088
*-F		5	0.12	0	0.63	0.21	-0.13	-0.13	-0.088
H		5	-1.8	0	-0.57	-0.57	0.55	0.55	0.27
Cl		5	1.8	0	0.57	0.57	-0.55	-0.55	-0.27
H		2	0.66	0	-0.022	-0.022	-0.42	-0.42	-0.2

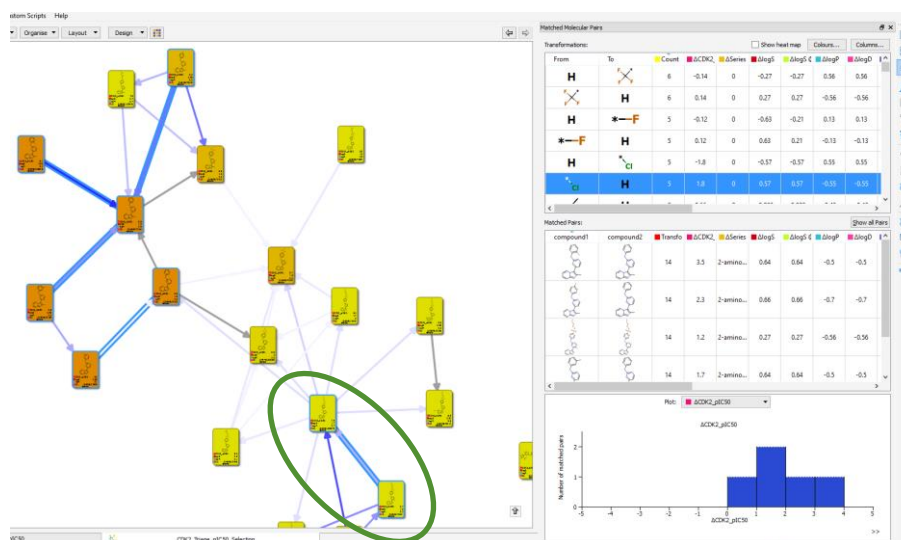
- As you click on the different examples of this transformation in the **Matched Pairs** table below, the network display in Card View zooms in to that pair of compounds.

Matched Pairs									
compound1	compound2	Transfo	ΔCDK2	ΔSeries	ΔlogS	ΔlogP	ΔlogD	$\Delta\text{C9 pi}$	
		14	3.5	2-amino...	0.64	0.64	-0.5	-0.5	-0.23
		14	2.3	2-amino...	0.56	0.56	-0.7	-0.7	-0.24
		14	1.2	2-amino...	0.27	0.27	-0.56	-0.56	-0.096
		14	1.7	2-amino...	0.64	0.64	-0.5	-0.5	-0.25



Here 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.

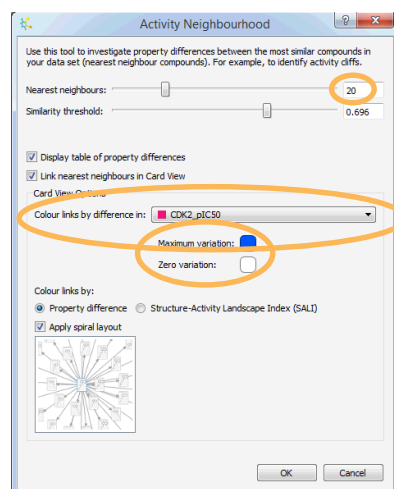
One example of this transformation shows the chlorine changing to a hydrogen on the pyrimidine rather than the phenyl ring and this is obvious too by seeing the slightly separate pair of selected cards in the network plot (circled below).




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

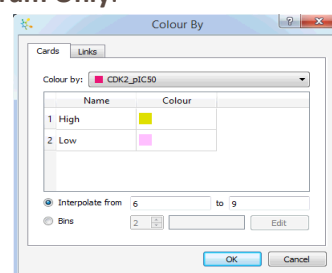
first need to select a specific compound which in this case will be the unsubstituted phenyl compound, CHEMBL72464.

- Select the unsubstituted derivative (CHEMBL72464) in Card View, click the **Analyse** menu at the top of Card view and select **Activity Neighbourhood**.
- In the **Activity Neighbourhood** dialogue choose to colour the links by CDK2_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 used before (shown right).

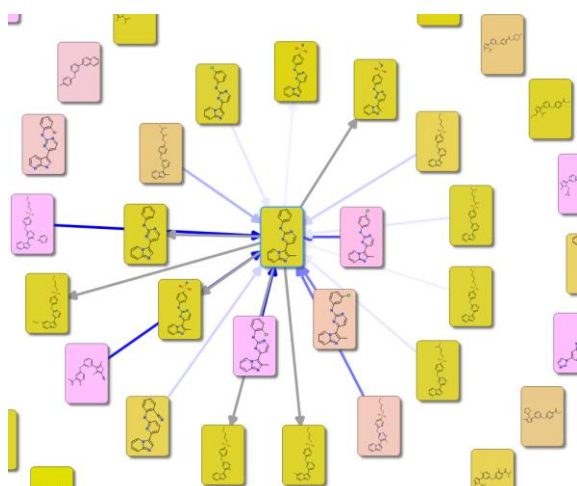


To make the structures more prominent we can change the card design to show the structure only

- Click the **Design** menu at the top of Card View and choose **Diagram Only**.
- Click the  button to bring up the **Colour By** dialogue again and colour the cards by the **CDK2_pIC50** value, interpolating from 6 (pink) to 9 (yellow)



The image below shows the resulting activity neighbourhood around the selected unsubstituted phenyl compound with the cards



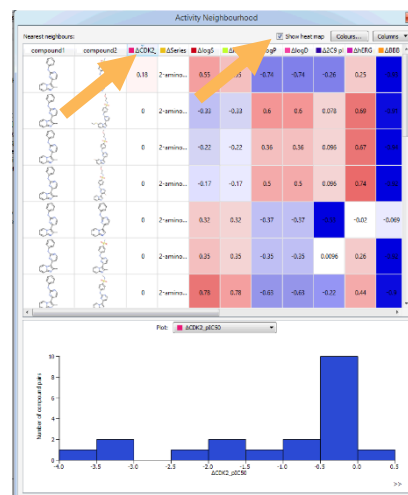
From this analysis we can see that not all substitutions on that phenyl ring are disfavoured as those compounds having a para-sulfonamide substituent are equally affine.

- Hover the mouse over a link to compare that pair of cards side-by-side

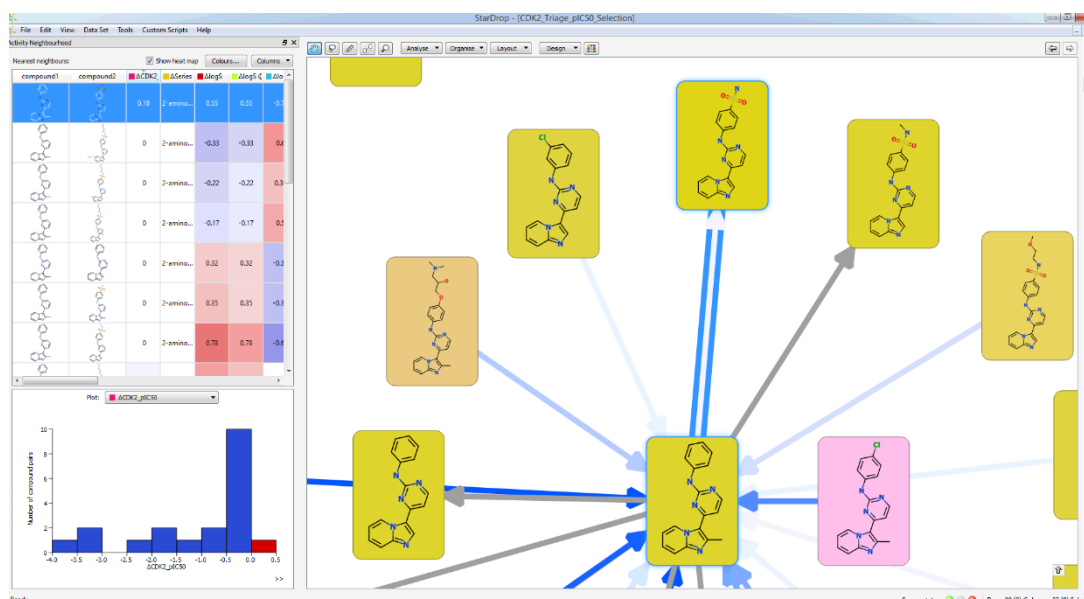
In the Activity **Neighbourhood** table, The **Nearest neighbours** table can be used to view the effects of these changes across the range of properties in the data table.

- **Left click** twice on the header of the $\Delta\text{CDK2_pic50}$ 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 $\Delta\text{CDK2_pic50}$ 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 also separation of 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.