



## Worked Example:

### Analysing Cross-series SAR using Matched Series Analysis

This worked example considers a publically available set of PPAR- $\gamma$  pIC<sub>50</sub> data which contains a number of chemical series. Matched Series Analysis is used to identify when a substitution from one series has a high likelihood of improving the binding of another series at PPAR- $\gamma$ . This approach is then further applied to help analyse the project's SAR to generate hypotheses regarding binding modes and identify possible anomalous measurements for further investigation.

A matched series is a series of compounds that are identical except for different substituents at a single point (See section 10.3 of the StarDrop™ Reference Guide for more details). In this worked example suggestions are derived from comparing matched series within the input data, highlighting where SAR does, or does not, track between chemotypes. The suggestions are based on the premise that a matched series with similar activity order in more than one chemotype means that the substituents occupy a similar binding environment. Given a similar binding environment, substituents that have been shown to be better binders within one chemotype have a strong likelihood of being better binders to the target in another. Conversely, the situation where the matched series SAR of one chemotype differs from another also provides useful information because this may indicate that a change in binding mode has occurred or that an assumed overlay of the chemotypes is not valid.



Optibrium™, StarDrop™, Card View™, Nova™, Glowing Molecule™ and Auto-Modeller™ are trademarks of Optibrium Ltd.  
Matsy™ is a trademark of NextMove Software Ltd

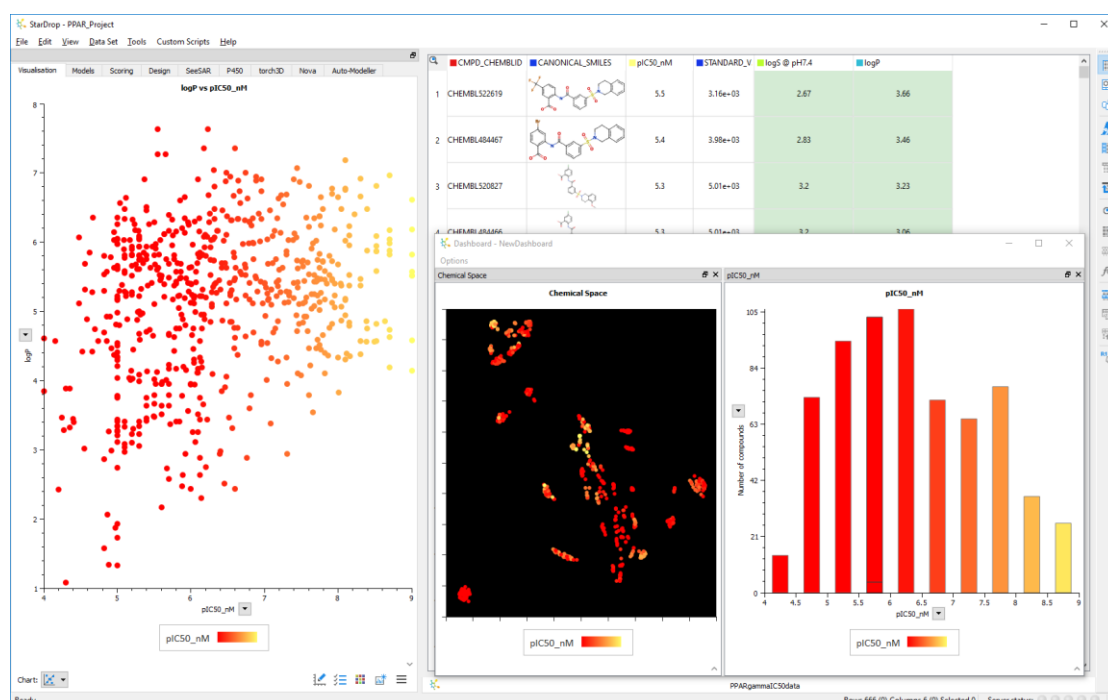
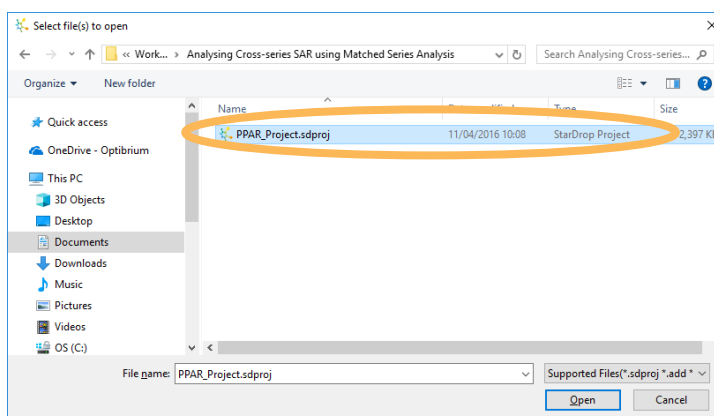
Step-by-step instructions for all the features you will need to use in StarDrop are provided, along with screenshots and examples of the output you are likely to generate. If you have any questions, please feel free to contact [stardrop-support@optibrium.com](mailto:stardrop-support@optibrium.com).

## Exercise

- In StarDrop, open the project file **PPAR\_Project.sdproj** by selecting **Open** from the **File** menu.


You will see a spreadsheet containing 666 structures and their measured  $pIC_{50}$  values for

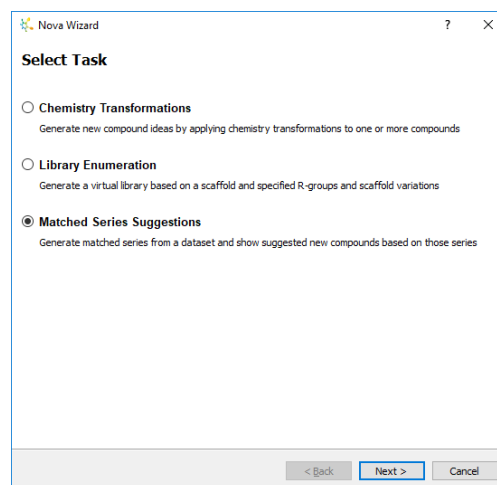
PPAR- $\gamma$  (in the column labelled  $pIC_{50\_nM}$ ) along with two calculated columns for  $\log P$  and  $\log S @ pH7.4$ .



You will also see that there are three visualisations that have been saved in the project file, in all of which the points are coloured from red, for low  $pIC_{50}$  values, to yellow, for high  $pIC_{50}$  values. The histogram shows that there is a wide spread of activities, across 5 orders of magnitude. The scatter plot in the Visualisation area shows that the most active compounds tend to have higher  $\log P$  values, which may not be desirable. The chemical space, which is also

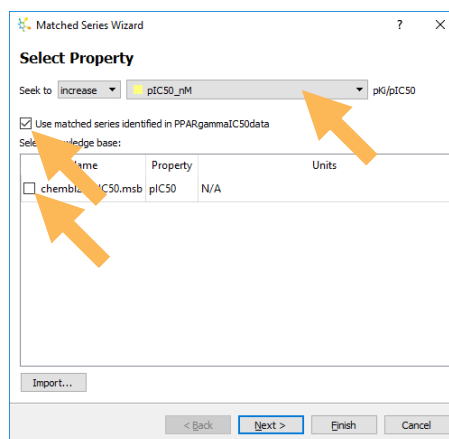
in the dashboard with the histogram, shows the diversity of the set. With good activity spread across the different chemical series, finding the consistent or inconsistent SAR across this set would be a laborious process.

- Minimise the dashboard and change to the **Nova™** area by clicking on the **Nova** tab.
- To start the Matched Series Analysis, click the  at the bottom of the Nova area.
- Select **Matched Series Suggestions** in the Nova Wizard and click **Next**.



In the dialogue box that appears, you can specify which property column you wish to improve. The pIC50\_nM column is the default option and we want to find suggestions that **increase** this value.

- In this exercise we only want to search within the data set, so uncheck the option to use the ChEMBL knowledgebase and check the option above it to **Use matched series identified in PPARgammaIC50data** instead.
- Click **Next** to continue.

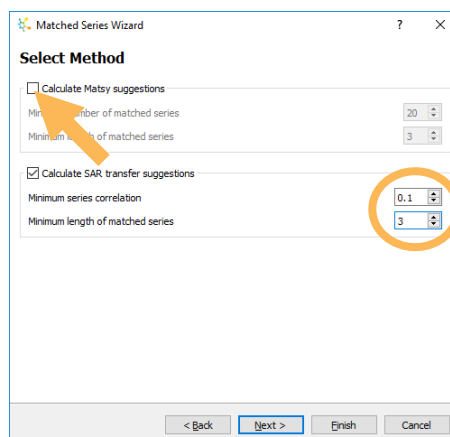


At this point, you can change the parameters for the suggestions returned. In Matsy, the support for a suggestion comes from the number of times it has been observed; the more frequently the input series occurs in the same order, the more likely it is that the suggestion will be an improvement. In a small data set, however, the frequency of occurrence is likely to be low and so in this example we will concentrate on the SAR transfer method.

With SAR transfer, the support for a suggestion comes from a long series of derivatives that show a consistent trend within several chemotypes. The default minimum number of derivatives (8) is too large for this set, so we will reduce this to 3. In this example, we also want

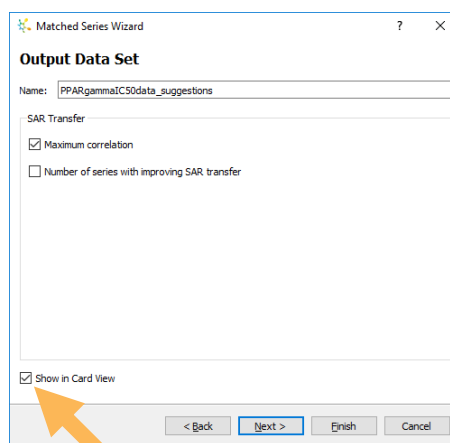
to see if there are cases where the SAR of the different chemotypes does *not* match, so we will also reduce the minimum series correlation for which results will be reported.

- **Uncheck** the box next to **Calculate Matsy suggestions**.
- In the **Calculate SAR transfer suggestions** section, change the **Minimum length of matched series** to **3** and reduce the **Minimum series correlation** to **0.1**.
- Click **Next** to continue.



Here you can give the output data set of suggestions a different name and control the data that are reported in the output.

- Check the **Show in Card View** option, as shown above, and click **Next**.
- In the last step you can choose structural filters to exclude certain chemical groups. In this case we will use the defaults, so click **Finish** to begin the Matched Series Analysis.



The suggestions are returned in Card View as a series of networks and the output should look similar to that shown below.

The 'Matched series' heatmap shows the following data:

Scaffold	Correlation	6.06	5.64	5.15
1	6.43	6.2	5.92	5.45

Each new suggestion in the networks above is coloured by the matched series that inspired it; the matched series from the input data set are the white cards in the networks.

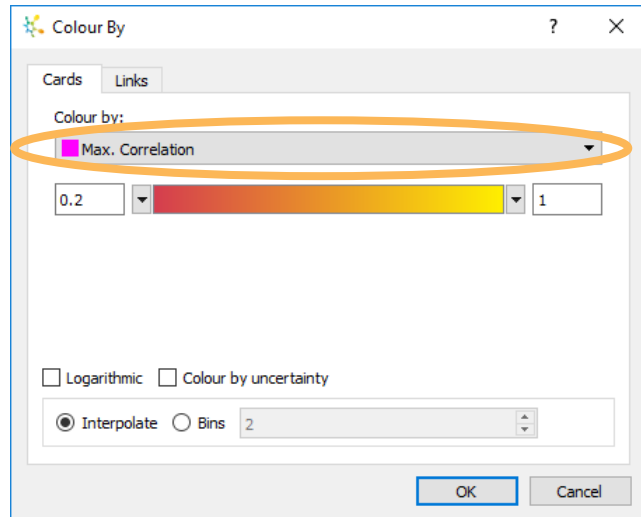
To assess how well these series correlate, and hence how well the SAR transfers between chemotypes, we will colour the cards by the **Max. correlation** property.

- Click on the **Format button**



at the top of Card View.

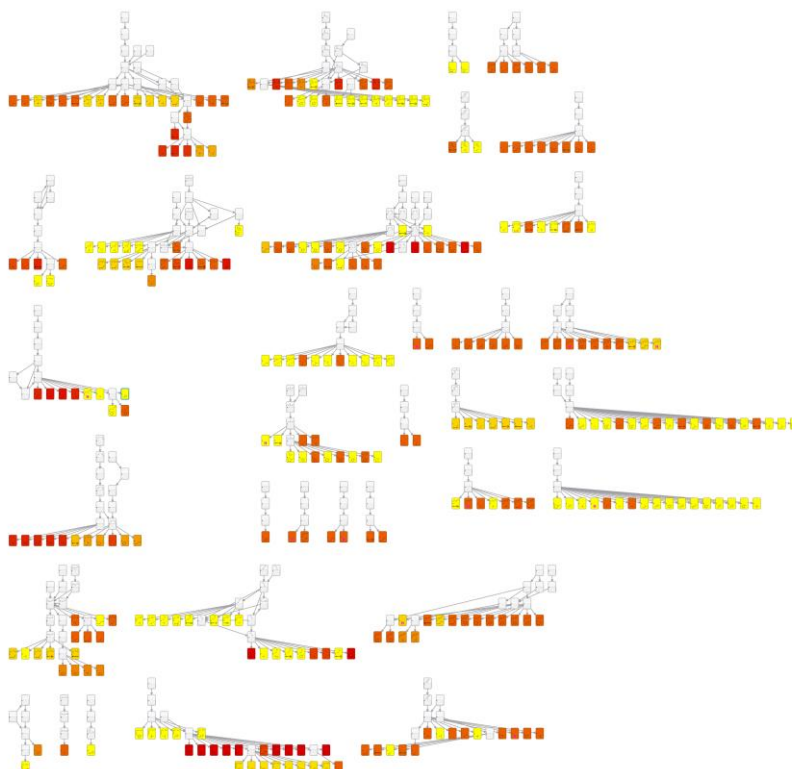
- In the **Colour By** dialogue change the **Colour by** drop-down to **Max. Correlation**.
- Click **OK**.




Card View display will look similar to that shown below and we can see where the red and yellow cards

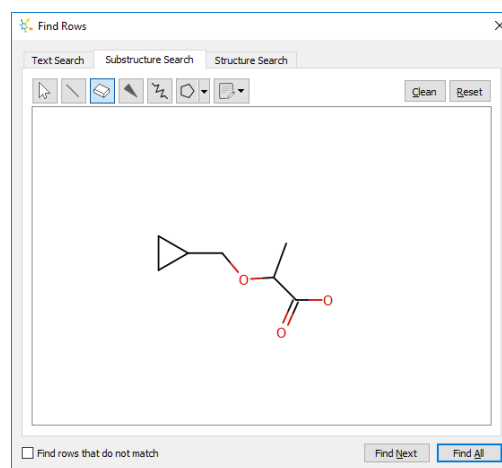
indicate poor and good SAR transfer respectively. In the next section we will examine the networks highlighted by the coloured circles.

**Note:** depending on the dimensions of your Card View area, the networks may be laid out in a different configuration.



To help us to explore the underlying data in more detail, when the card representing a suggested compound is selected, the matched series of compounds giving rise to the suggestion is highlighted by the blue links between the cards. The display in the Nova tab also changes to show the correlations in SAR that prompted the matched series suggestion. To see this, we will search for a specific example to study in more detail.

- Click on the **Find button**  on the right-hand toolbar to bring up the **Find dialogue**.
- Click on the **Substructure Search** tab and draw the substructure shown on the right.
- Click **Find All**.



**Note:** you may need to move the Find dialogue out of the way.

This will highlight a series within one of the sub-networks and display the first compound found in the Nova area.

**Note:** the suggestions at the bottom of the network may be ordered differently.

- Zoom into the selected region using the mouse wheel or **Ctrl** and = keys.

**Matched series:** Increasing pIC50<sub>μM</sub>

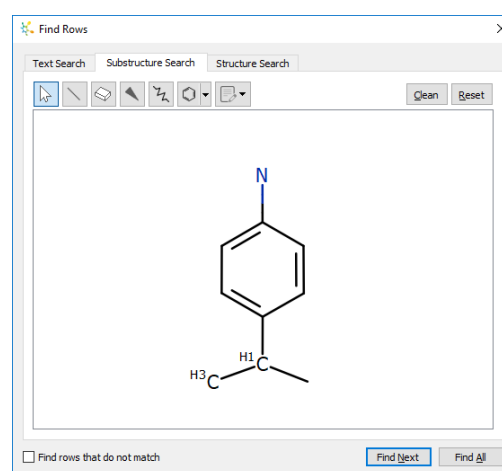
Scaffold	Correlation	8.15	7.52	7.44	6.95	5.8	
1		8.7	8.4	8.15	7.68	7.34	5.8

Where there is a high correlation between two matched series, we can have confidence that the SAR may be transferred. Here we can see that the ether substituent in the benzothiophene series increases the activity as the size increases, with the cyclopropylmethyl group being the most active in the series of 6. A pop-up showing the scaffold for each series can be seen by hovering the mouse over structure, as shown above. As one might have expected, the naphthyl series shows the same SAR for the first five derivatives and so the suggestion given is to make the cyclopropylmethyl derivative in the naphthyl series, because it is expected to be more active.

We will now consider another series and the compound shown below.

- To find it, again use the **Find tool** and search for the substructure shown right. You can specify Hydrogen counts by right-clicking on the atoms to bring up the menu.

**Note:** there are two structures where this substructure is found, and you may need to click **Find Next** to find the one shown below.



The cards are coloured by the correlation of the SAR in the suggestion series to the SAR in the rest of the data set. This group contains different suggestions for substituted phenyls adjacent

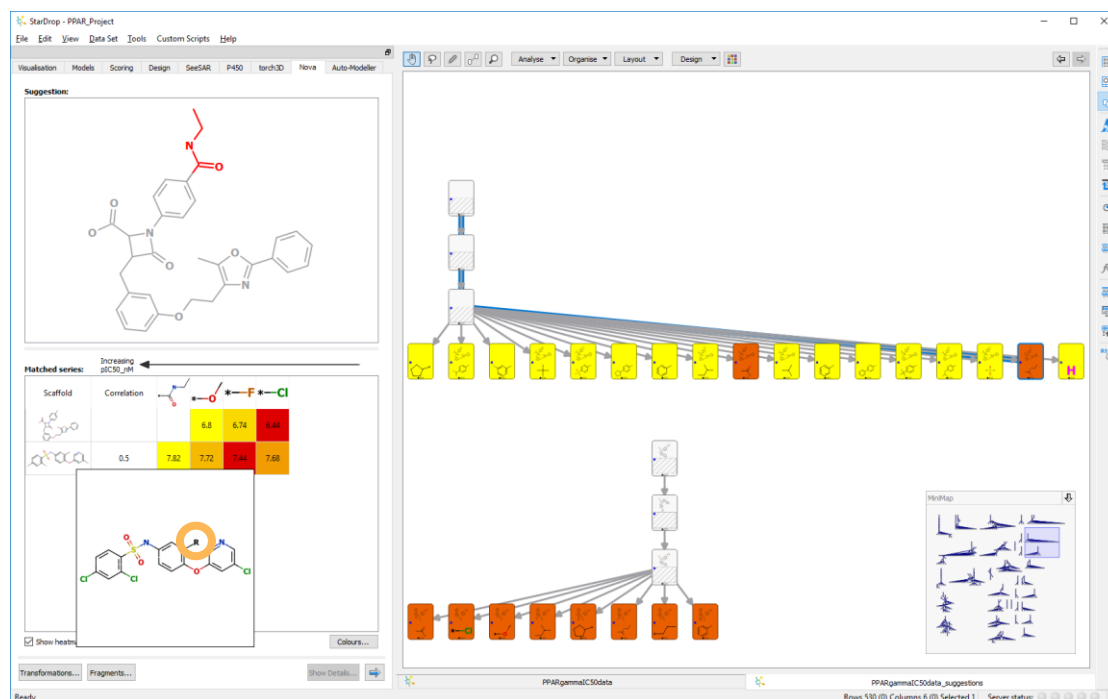
to a beta-lactam group. The SAR for the supporting series is consistent for an overlay of those series where the hydrophobic groups are binding in the same pocket and then leading, via a flexible linker, to the acidic group at the other end of the molecules.

The selected card has one scaffold series supporting it, while other suggestions in the network have two additional supporting series.

Interestingly, the suggestion for derivatisation is adjacent to the acidic group in the beta-lactam series, i.e. at the other end of the molecule. This SAR transfer indicates three possibilities:

- An overlay of the series where the paraquinone ether is bent back towards the acid (the linker is sufficiently flexible)
- There is another, similar aryl binding pocket not yet explored in the thiazolidinedione or the benzopyran series
- The SAR transfer is simply a coincidence

We typically we look for matched series with a high correlation, to give confidence that the suggestion will have improved activity. Therefore, suggestions from matched series with a poor correlation are not likely to be relevant. Examples of these are shown by the orange coloured cards in this network and, as shown below, they are derived from a diarylsulfonamide series where one might not expect the rings to be binding in the same area as this series does not have an obvious acidic group.

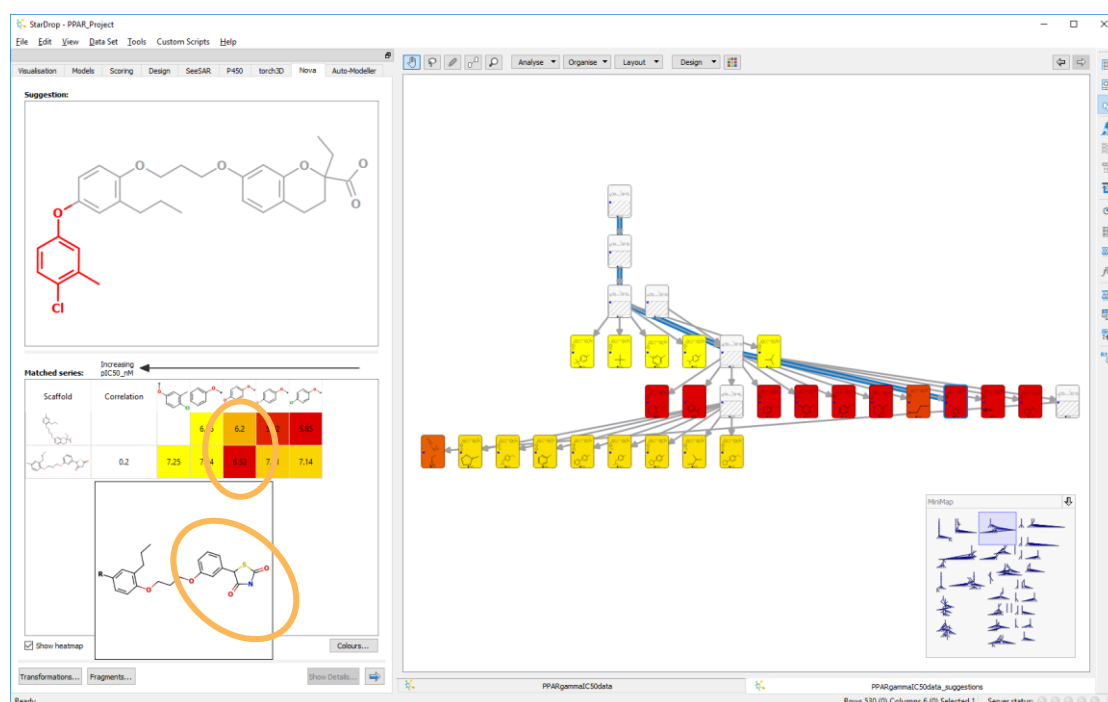
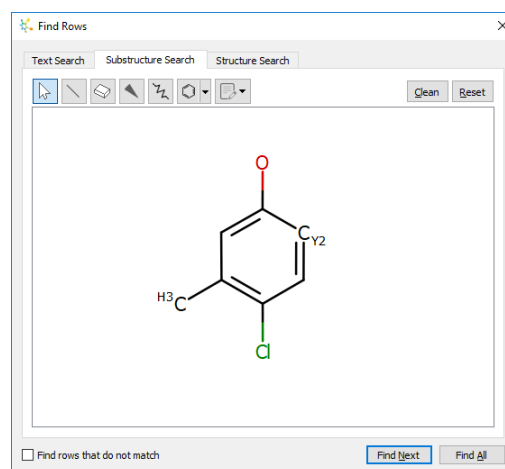




However, looking for series with low correlation can highlight discontinuous SAR that might be derived from rogue assay values.

- To find the final series we will examine, again use the **Find tool** and search for the substructure shown right. You can specify exactly two heavy atoms connected to the aromatic carbon by right-clicking on it to bring up the menu.

**Note:** as before there are multiple structures where this substructure is found, and you may need to click **Find Next** to find the one shown below.



The network above shows that the SAR in the hydrophobic tail region of the benzopyran series matches that of the glycine derivative series (the top row of yellow cards) when the series is only 3 derivatives. However, when the fourth derivative is included in the matched series the SAR suddenly does not match derivatives in the thiazolidinedione series (the second row containing red cards).

From the inspection of one of these thiazolidinedione series we can see that the central anisole derivative has lower activity than one would expect. The differences between

compounds in this SAR series as a whole are very small and within the error of an assay, so lower correlations will occur just by chance, but the reduction in activity of the methoxy derivative significantly reduces the correlation and appears anomalous. This is emphasised by examination of the longer series of 5 compounds where the suggestion cards are all coloured orange (bottom row). These compounds come from a very similar thiazolidinedione series where the thiazolidinedione is para to the linker ether rather than meta as in the case above, and again there is a good correlation with the benzopyran series. This would suggest that it would be useful to retest the activity of the methoxy derivative of the meta thiazolidinedione series to confirm this result.

This worked example has shown that the analysis of the correlating and non-correlating SAR series enables hypotheses regarding compound alignments or their binding modes to be formulated and highlights anomalous data points in a way that is complementary to simple Matched Molecular Pairs analysis. The Matched Series Analysis tool within Nova also provides suggestions for further derivatives that are likely to improve a desired endpoint and can help support or disprove these binding hypotheses.

The applicability and suitability of suggestions also relies on many other compound properties so the output data set can be further prioritised in StarDrop, for example using predictive models of physicochemical, ADME or target activity properties and Probabilistic Scoring for multi-parameter optimisation.

Examples of the use of these methods can be found in some of the other worked examples, but if you have any questions or would like to see a demonstration then please contact [stardrop-support@optibrium.com](mailto:stardrop-support@optibrium.com).