Worked Example:

Automatic Generation of New Compound Ideas

This example considers a challenge typically faced when a series of promising compounds has been found, but the chemistry needs to be broadened to further improve one or more properties. Focussing on a single property for optimisation often increases the risk of introducing liabilities for other properties and so in this exercise we're going to see how StarDrop™ can help to explore new chemistry ideas while minimising this risk.
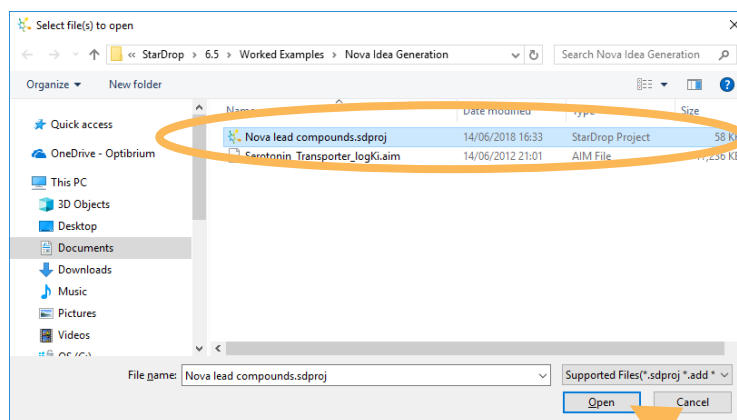
During this example we will consider three compounds from a lead series which we would like to try to evolve into a candidate. The compound has a good profile of ADME properties but insufficient inhibition of the target, the Serotonin transporter. In this example we will use StarDrop's Nova™ module to generate new ideas for compounds to improve the potency while maintaining the balance of other properties.

We will use a model for Serotonin transporter inhibition built with public domain data from the ChEMBL database using StarDrop's Auto-Modeller™ to monitor potency during the exploration.
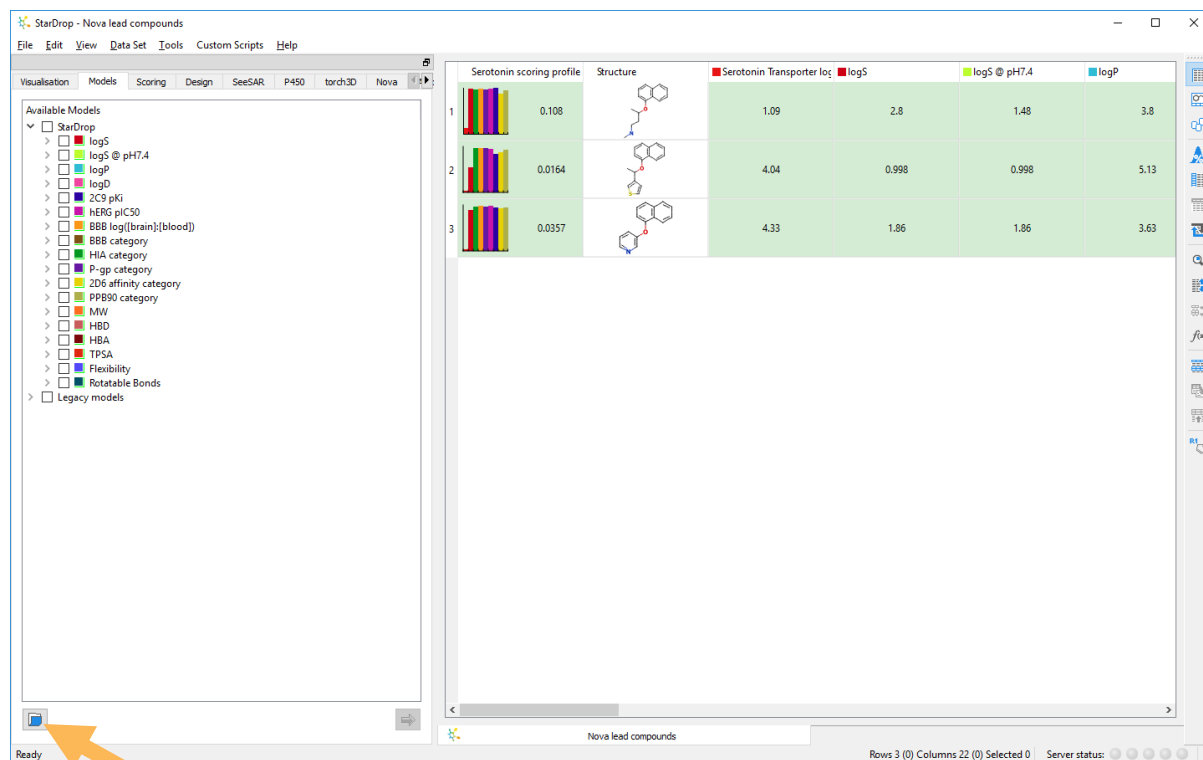
Nova is capable of generating data sets of many hundreds of thousands of compounds if left to run for multiple generations and so in this example we will also explore how this process can be managed to focus on those compounds most likely to achieve our objectives.

Exercise

- In StarDrop, open the project file **Nova lead compounds.sdproj** by selecting **Open** from the **File menu**.
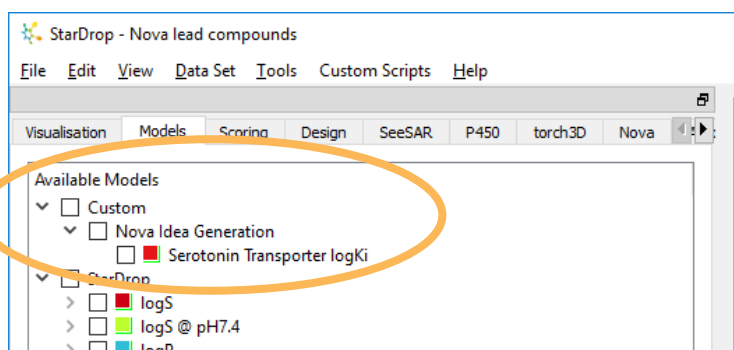
This simple project file contains a data set with three lead compounds for which a range of properties have been predicted. The compounds have been scored using StarDrop's approach to multi-parameter optimisation, called Probabilistic Scoring, against a profile of properties suitable for an orally dosed compound intended for a CNS target.

The third column in the data set is called **Serotonin Transporter logKi (uM)** and the values in this column have been generated by a custom model built using ChEMBL data with the StarDrop Auto-Modeller. These values are used when calculating the overall scores and so in order to calculate this property and scores for any new compounds we need to load this model into StarDrop.

- Click the **Open button** at the bottom of the **Models area** and open the model file **Serotonin_Transporter_logKi.aim**.

You will now see that a new model **Serotonin Transporter logKi** is displayed in the **Custom** branch in the **Models** tab. StarDrop will now be able to use this model in the prioritisation of compounds during the Nova idea generation process.

- Click on the **Nova tab** and select all three rows in the data set by clicking in the top left corner of the data set.

- Click the ➡ button to start the Nova wizard.

The Nova wizard will be displayed. On the first page you can choose whether to apply Chemistry Transformations, run Library Enumeration or generate Matched Series Suggestions.

- For this example, choose **Chemistry Transformations** and click **Next**.

Note: Additional worked examples are available for the other approaches.

The **Specify Input Structure** page of the wizard will show the structures we have selected from the data set.

If you wish to indicate regions of the molecules that must remain unchanged then you can draw round them with the mouse to create a mask.

Selecting the individual molecules on the left enables you to define individual masks for each molecule.

- In this example all of the molecules contain a common sub-structure. Rather than having to draw this for each compound, click the **Find Common Mask button**.
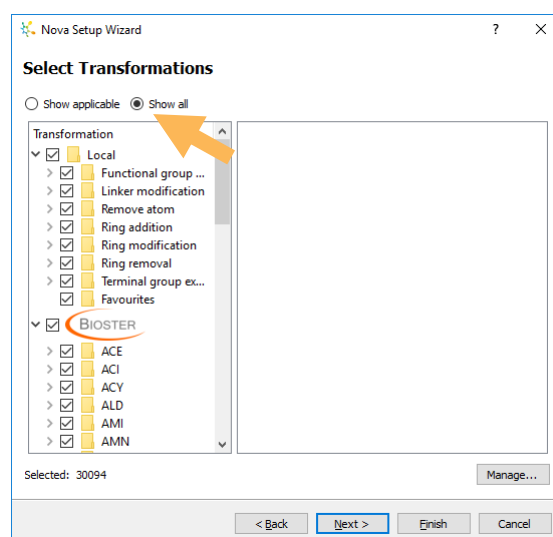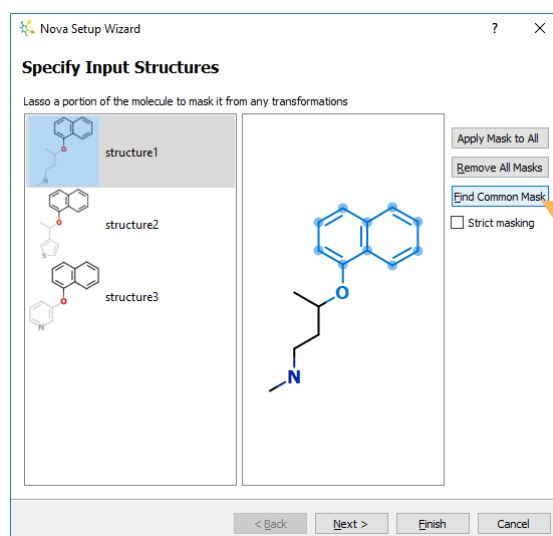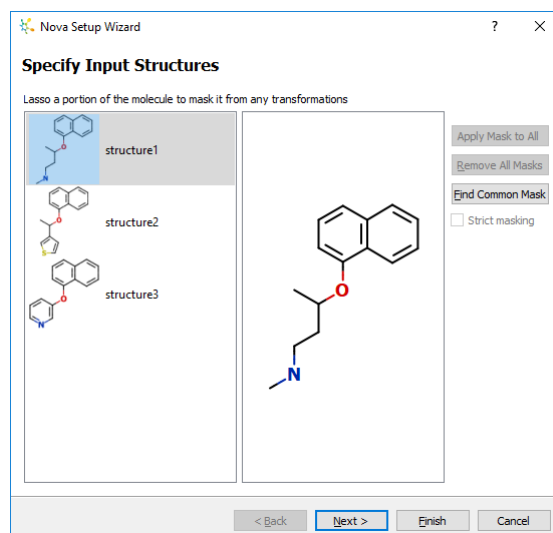
The **napthol** group will now be masked on each of the compounds.

- Click the **Next** button to go to the **Select Transformations** page.

Nova will search the database of available transformations to find those that apply to the input compounds.
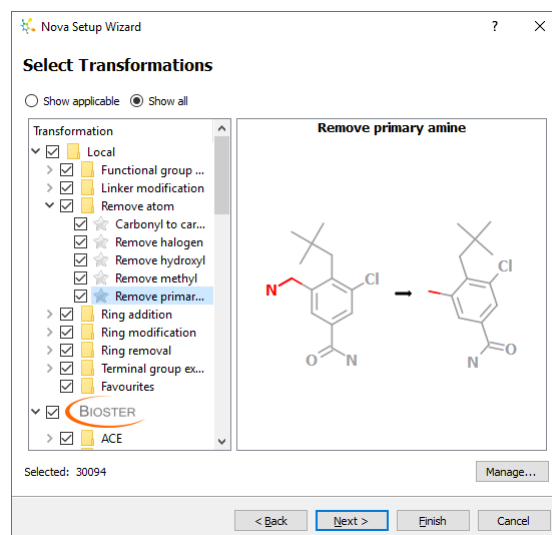
On this page we can select which transformations we want to apply. They are grouped to make it easy to pick certain types of transformation. The **Manage** button enables you to reorganise the transformations or import your own.

- Click on the **Show All** option to see the full library of transformations, including those that do not apply to our specific structures.
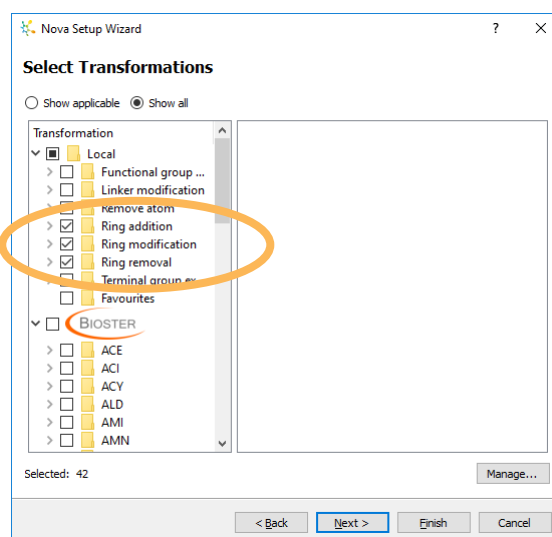
You can open the groups of transformations, by clicking on the arrow next to the group, to browse the individual transformations, as shown to the right.

The **Local** transformations are those provided with the Nova module. If you have a license for the BIOSTER™ module, you will also see a database of >30,000 precedented transformations, manually curated from the literature.

- In this quick example, we're only going to select the three transformation groups relating to rings: **Ring addition**, **Ring modification** and **Ring removal**. Untick the boxes next to the other groups and the BIOSTER library, as shown to the right.
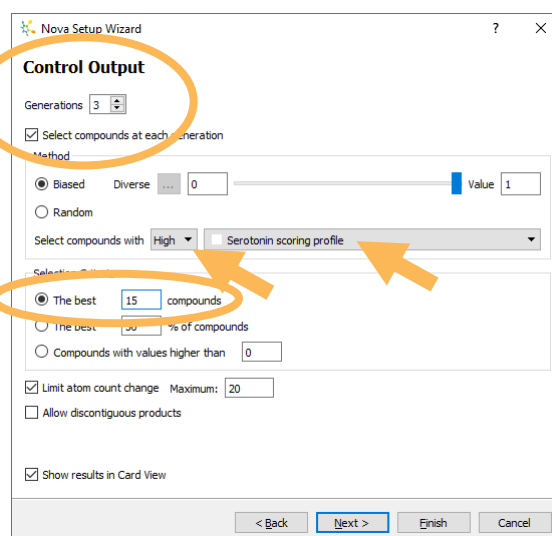
- With **only** these three items selected, click the **Next** button to go to the **Control Output** page

On this page we can choose how many generations of transformations to apply.

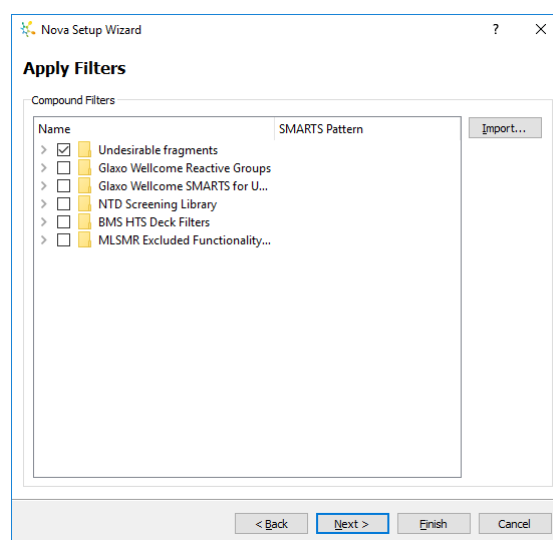- Change the number of **Generations** to **3**.

If we apply multiple generations then all of the molecules created in one generation will be used as starting points for the next generation. This can result in very large number of compounds being generated.
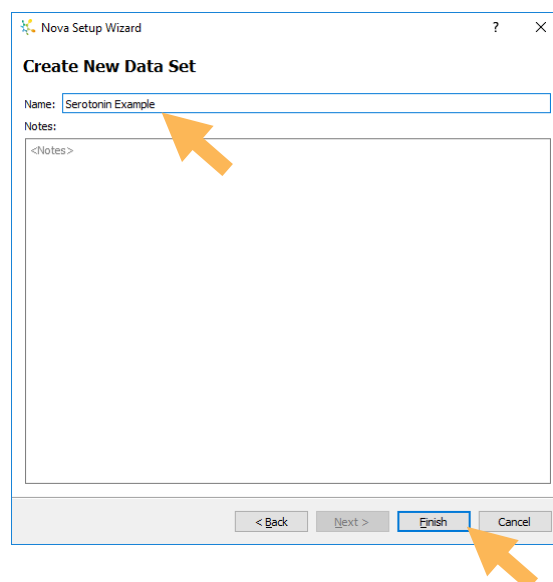
- To restrict the number of compounds used as the basis for subsequent generations, tick the **Select compounds at each generation** checkbox.

- First, we choose a **Method**. We can select compounds based on a property, a score, chemical diversity or a combination of these. If a property or score is involved, then we can indicate whether it is desirable for this to be high or low. In this case we are going to make a selection that is **Biased** entirely towards having a **High** score against the **Serotonin scoring profile**. (The **Biased** slider will have a **Diversity** of 0 and a **Value** of 1 to show that diversity will not be considered in this instance).

- We can then define the **Selection Criteria**. In this case choose **The best 15** compounds.

- Click the **Next** button to go to the **Apply Filters** page.

On this page we can indicate any structural features which we would like to ensure are not present in any of the new compounds. Compounds that contain any of these will be removed at the end of the process. A number of published sets of filters are available and by clicking the **Import...** button you can add your own, but in this example, we will just use the **Undesirable fragments** that are already checked.

- Click the **Next** button to go to the **Create New Data Set** page.

- On the final page we can give the Nova session a name, in this case **Serotonin Example**, and add any other relevant notes. The session name will be used for the new data set that is created.

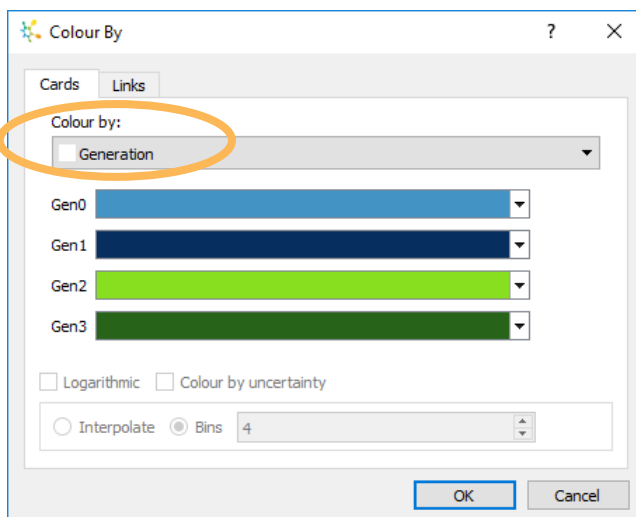- Start the idea generation process by clicking **Finish**.

The Nova process will take a couple of minutes to run. While the process is running an indicator will show which generation is currently underway along with the best property or score value achieved so far.
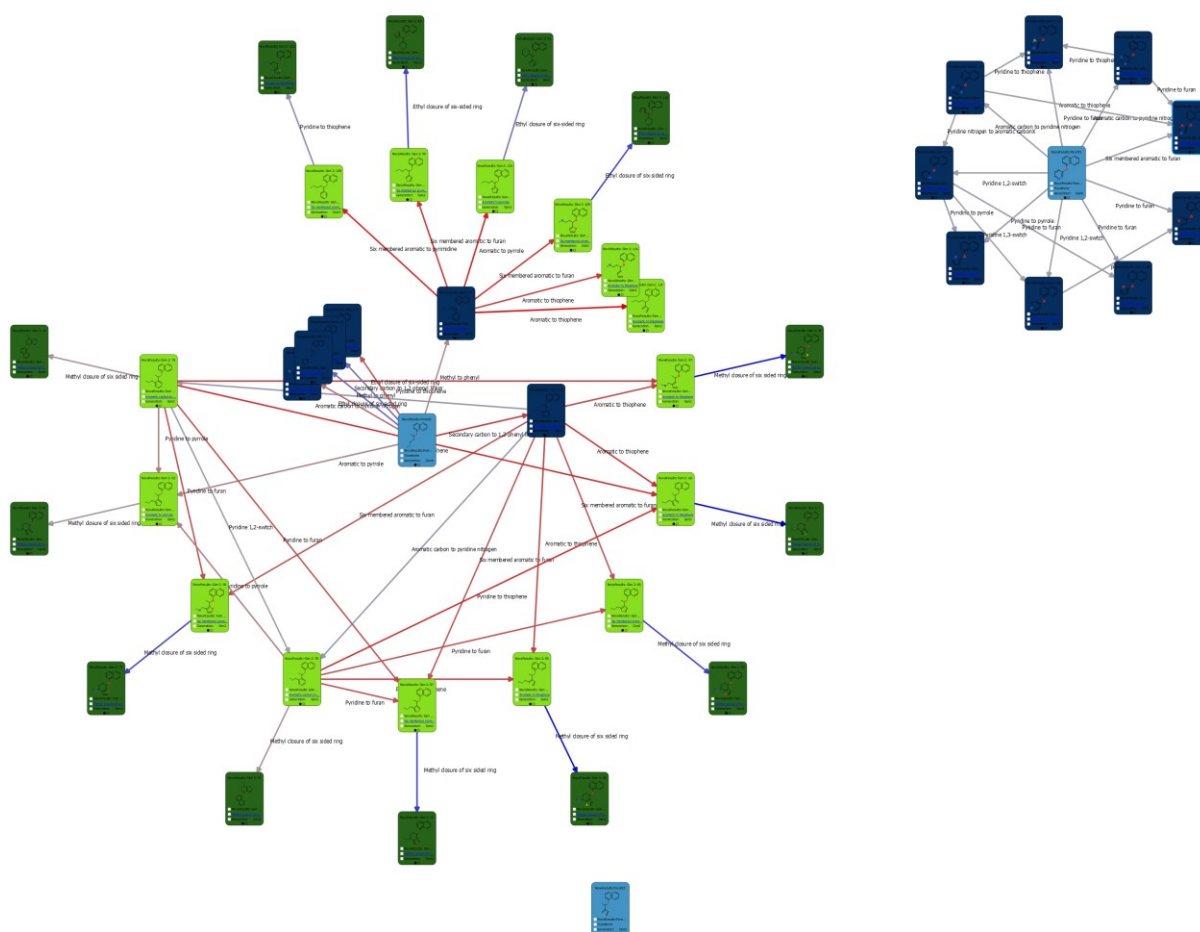

Once complete, a new data set called **Serotonin Example** will be displayed in Card View.
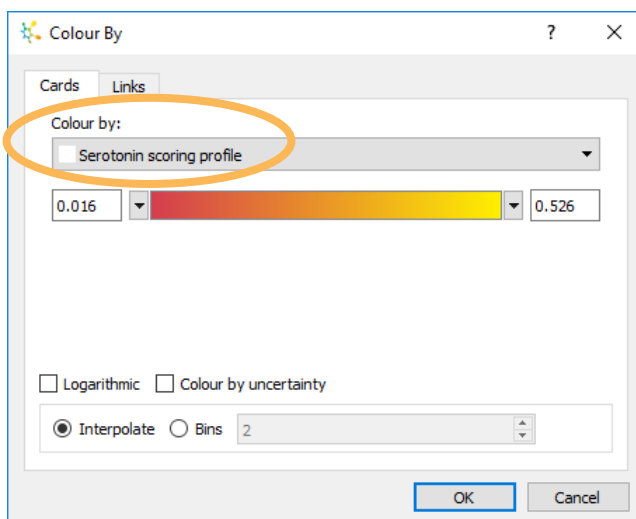
- We can colour the cards to show the different generations of compounds by clicking the **Format button** [icon] at the top of Card View.

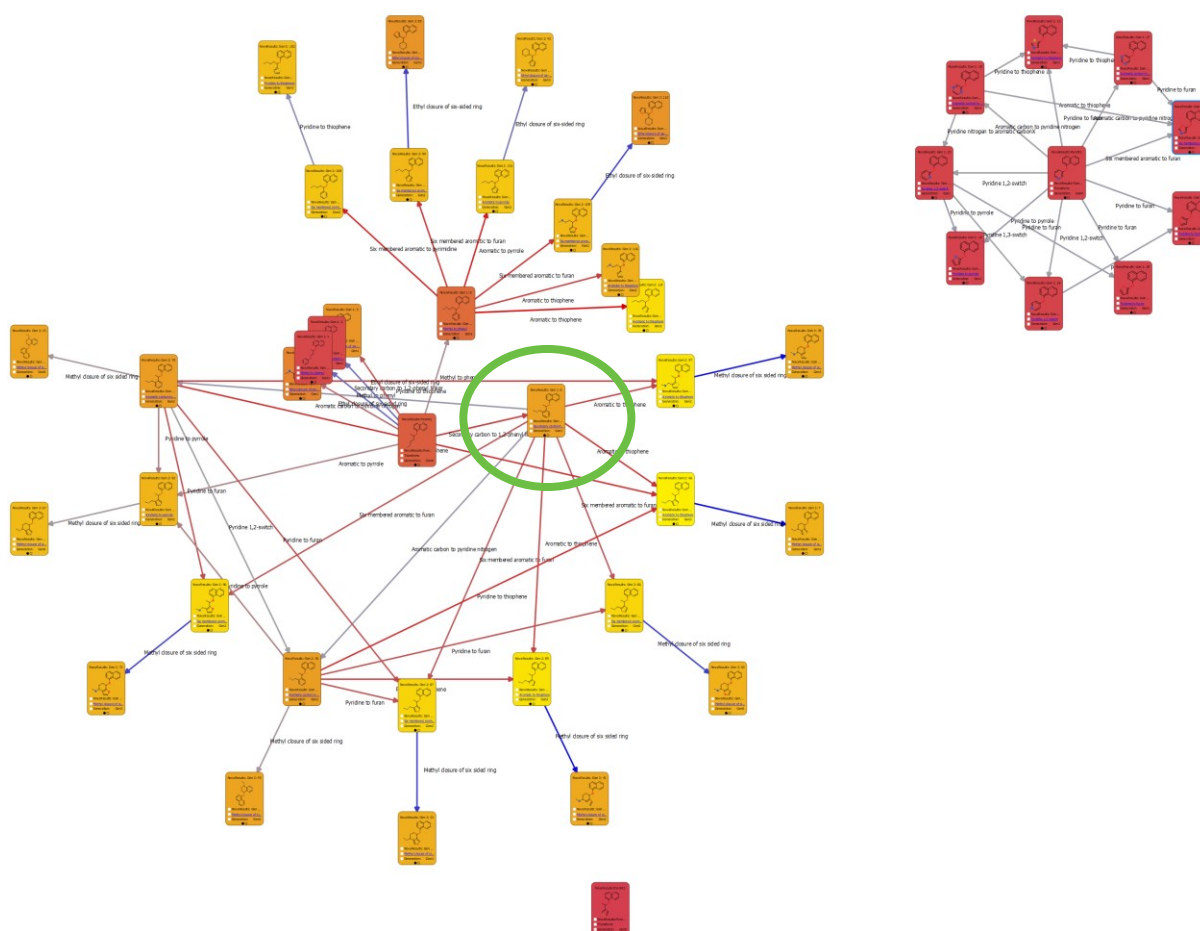- In the **Colour By dialogue**, choose the property **Generation**.

From the resulting display, shown below, we can see that, of the three initial compounds, one (top-right) resulted in the selection of compounds only from the first generation and none of the subsequent generations had a sufficiently high score to be selected. Another compound (bottom) did not result in any 'children' that were selected, while all the second and third generation compounds were generated from only one of the initial compounds (centre).

- We can now choose to colour the cards by the overall score by clicking again on the **Format button** at the top of Card View

- This time, in the **Colour By dialogue** choose the property **Serotonin scoring profile**.

From this display (below), we can see that most of the high scoring compounds are in the second generation (the yellowest cards).
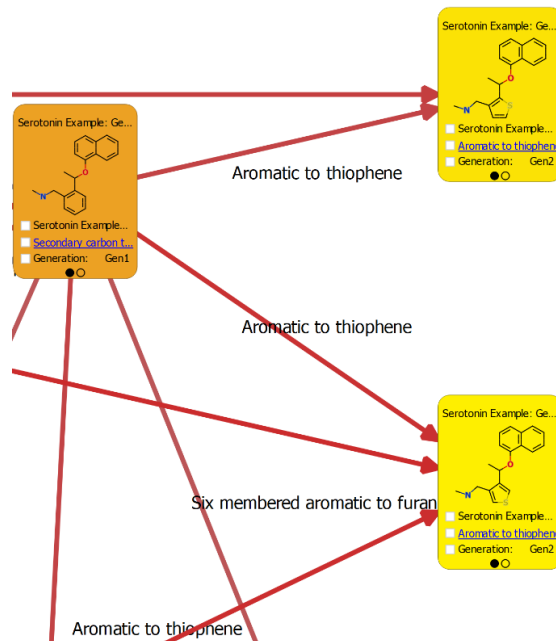

You will also note that the links between cards are coloured by the change in score between the corresponding compounds; a large increase is shown by a red link, a large decrease by a blue link and a small change in grey.

- Zoom in to the compound circled in green above using the mouse wheel (or by simultaneously pressing the **CTRL** and **=** keys)

From this, you will note that the biggest improvements in score from this compound (the red links) result from the "Aromatic to thiophene" and "Six membered aromatic to furan" transformations.

You can see this transformation in more detail by clicking on the hyperlink on the cards. You can also choose the see more or different properties on the cards by clicking on the **Design menu** at the top of Card View.

- Switch to Table View by clicking on the **Table View button** .

The compounds in the data set are sorted to show the compounds with the best property or score at the top. The data set contains a column showing which transformation was applied to generate each compound from its immediate 'parent' and a column indicating in which generation this occurred.

- Select the first row.

The first row contains the top scoring compound and we can see this compound's parent molecules and any subsequent children in the viewer on the left. This compound was created in generation 2.

Sometimes the same molecule can be generated from different parent molecules. In this case, all three parent compounds could be transformed with the Aromatic to thiophene transformation to produce this molecule. Notably, all three parent compounds had scores of under 0.3 but the introduction of the thiophene ring significantly improves this.

- Finally, notice that row 19 is actually the drug Duloxetine, a serotonin-norepinephrine reuptake inhibitor.

Further details about using Nova in conjunction with StarDrop's Probabilistic Scoring can be found in the following publication "Applying Medicinal Chemistry Transformations and Multiparameter Optimization to Guide the Search for High-Quality Leads and Candidates", Segall *et al.*, J. Chem. Inf. Model., 2011, 51 (11), pp 2967–2976, also available as a preprint from http://www.optibrium.com/community/publications/drug-discovery-applications/123-preprint-applying-med-chem-transformations-and-mpo-to-guide-the-search-for-high-quality-leads-and-candidates.

This example does not explore the potential of applying over 30,000 precedented transformations using the BIOSTER module (http://www.optibrium.com/stardrop/stardrop-bioster.php). If you have any questions, please feel free to contact stardrop-support@optibrium.com.