

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.



© 2018 Optibrium Ltd. Optibrium™, StarDrop™, Glowing Molecule™, Nova™, Auto-Modeller™ and Card View™ are trademarks of Optibrium Ltd. BIOSTER™ is a trademark of iKem Szolgáltató és Kereskedelmi BT

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.

K StarDrop - Nova lead compounds						- 0	×
Die Zur Ten Zuster Tens enternanden Teib	Serotonin scoring profile	Structure	Serotonin Transporter loc	■ logS	logS @ pH7.4	logP	
Visualisation Models Scoring Design SeeSAR P450 torch3D Nova 😯 Parents:	1 0.108	, 20	1.09	2.8	1.48	3.8	8
	2 0.0164	<u>s</u>	4.04	0.998	0.998	5.13	
	3 0.0357	¢ ^{\$\$\$}	4.33	1.86	1.86	3.63	2
Current:							
							f(x)
							R10
Children:							
Transformations Fragments	<	Nova laad compounde				2	·
Ready		Nova icoa compositos		Rows 3 (0) Columns	22 (0) Selected 3 Serv	ver status: 🔵 🔵 🔘	

• 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.

🔆 Nova Wizard	?	×
Select Task		
Chemistry Transformations		
Generate new compound ideas by applying chemistry transformations to one or more co	ompounds	
○ Library Enumeration		
Generate a virtual library based on a scaffold and specified R-groups and scaffold varia	tions	
O Matched Series Suggestions		
Generate matched series from a dataset and show suggested new compounds based or	n those seri	es
< <u>B</u> ack Next > ,	Can	cel
		-

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

👯 Nova Setup Wizard	?	х
Apply Filters		
Compound Filters		
Name SMARTS Pattern >	Import	
< Back Next > Finish	Cance	el

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.

🕻 Nova Setup Wiz	ard				?	×
Create New	Data Set					
lame: Serotonin Ex	ample					
otes:		•				
<notes></notes>						
		•				
		< <u>B</u> ack	<u>N</u> ext >	Einish	Cano	el:

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 III** 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

🐛 Colour By	?	×
Cards Links Colour by: Generation	•	
Gen0 🗸		
Gen1 💌		
Gen2		
Gen3		
Logarithmic Colour by uncertainty Interpolate Bins 4		
ОК	Cance	1

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).

🐛 Colour By	?	×
Cards Links		
Colour by: Serotonin scoring profile	•	
0.016	▼ 0.526]
Logarithmic Colour by uncertainty		
Interpolate O Bins 2	A V	
Ot	(Cance	ł



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.

StarDrop - Nova lead compounds						- 0	×
e Zur Trea Zaroar Toola caroniaribra Treb	Structure	ID	Transform	Generation	Serotonin scoring profile	■2C9 pKi ^	
sualisation Models Scoring Design SeeSAR P450 torch30 Nova 120		Serotonin Example: Gen 2: 66	Aromatic to thiophene	Gen2	0.526	4.61	65 65
ng ng ng	2	Serotonin Example: Gen 2: 89	Aromatic to thiophene	Gen2	0.488	4.61	<u>▲</u> ■
0.292 0.289 0.3	3	Serotonin Example: Gen 2: 97	Aromatic to thiophene	Gen2	0.488	4.61	1
ent:	4	Serotonin Example: Gen 2: 114	Aromatic to thiophene	Gen2	0.474	4.61	e 1111 1111
Serotonin Example: Gen 2: 66 Serotonin scoring profile: 0.526	5	Serotonin Example: Gen 2: 86	Six membered aromatic to	Gen2	0.452	4.54	f(x)
\checkmark°	6	Serotonin Example: Gen 2: 96	Six membered aromatic to	Gen2	0.452	4.54	₽
N CS	7	Serotonin Example: Gen 2: 87	Six membered aromatic to	Gen2	0.45	4.54	RI O
en:	8	Serotonin Example: Gen 2: 101	Aromatic to pyrrole	Gen2	0.409	4.57	
0.35	, ,	Serotonin Example: Gen 2: 108	Six membered aromatic to	Gen2	0.402	4.61	
	10	Serotonin Example: Gen 3: 102	Pyridine to thiophene	Gen3	0.378	4.47	
		Serotonin Example: Gen 3:	Methyl closure of six sided.	Gen3	0.372	4.59 × >	
y	K Nova lead co	mpounds 🠇	Serotonin Exa	mple Rows 46 (0) Columns	14 (0) Selected 1 Server stat	:us: 0 0 0 0 0	

• 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.

🤽 StarDrop - Nova lead compounds							- 0	×
Eile Edit View Data Set Lools Custom Scripts Help								
Mousiestine Meddle Coorine Decise CoorCAD D4E0 teach2D Nova	Stru	icture	ID	Transform	Generation	Serotonin scoring profile	2C9 pKi	^
Parents:	13	\sim	Serotonin Example: Gen 2: 99	Six membered aromatic to	Gen2	0.367	4.54	69 69
	14	afor	Serotonin Example: Gen 3: 92	Ethyl closure of six-sided r	Gen3	0.362	4.69	<u>∧</u> ≣≣
0.148	15		Serotonin Example: Gen 2: 92	Aromatic to pyrrole	Gen2	0.358	4.57	1
Qurrent:	16	-45 	Serotonin Example: Gen 3: 7	Methyl closure of six sided	Gen3	0.35	4.64	
Serotonin Example: Gen 2: 116 Serotonin scoring profile: 0.344	17	ц Д	Serotonin Example: Gen 3: 73	Methyl closure of six sided	Gen3	0.349	4.57	f(x)
_N0	18		Serotonin Example: Gen 3: 30	Methyl closure of six sided	Gen3	0.347	4.57	
S	19	\sim	Serotonin Example: Gen 2: 116	Aromatic to thiophene	Gen2	0.344	4.61	RI
Children:	20		Serotonin Example: Gen 3: 33	Methyl closure of six sided	Gen3	0.338	4.57	
	21	40 •	Serotonin Example: Gen 3: 78	Methyl closure of six sided	Gen3	0.326	4.64	
	22		Serotonin Example: Gen 3: 43	Methyl closure of six sided	Gen3	0.325	4.64	
	23 «	<u>à</u>	Serotonin Example: Gen 3:	Methyl closure of six sided	Gen3	0.313	4.78	~
Transformations Fragments Show Details 🖨	*	Nova lead cor	npounds 🠇	Serotonin Exam	ple			

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 2967-2976, available (11), рр also as preprint from а http://www.optibrium.com/community/publications/drug-discovery-applications/123-preprintapplying-med-chem-transformations-and-mpo-to-guide-the-search-for-high-quality-leads-andcandidates.

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