

Worked Example:

Guiding Selection and Design in Hit-to-Lead

This example explores some of the challenges typically encountered in a hit-to-lead project. The objective in this case is to identify one or more high quality chemistries for progression to detailed *in vitro* and *in vivo* studies, based on initial screening data for potency; ideally the compounds chosen for progression should not only be potent, but also have appropriate ADME properties to result in a high quality lead series. We will also use StarDrop to explore potential modification of one of the existing compounds to improve its properties.

During this exercise we will use a variety of StarDrop's capabilities to explore the data in order to evaluate and select molecules with a good balance of properties. 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 <u>stardrop-support@optibrium.com</u>.

Exercise

- Start StarDrop from the Start menu.
- Open the file **Neurokinin2.add** by using the **File -> Open** menu.

👯 Select file(s) to open			×
G V Vibrarie	s Documents StarDrop Example	▼ 49	Search StarDrop Example
Organize 👻 New fol	der		8= - 🔟 🔞
☆ Favorites ■ Desktop	Documents library StarDrop Example		Arrange by: Folder 🔻
Downloads	Name		Date modified
🕮 Recent Places	roject.apd		08/00/2011-00.45
📤 Google Drive 🕒	👯 Neurokinin2		21/12/2011 16:22
 □ Libraries □ Documents ↓ Music □ Pictures 			
I Videos File	name:	•	Supported Files (*.add *.csv *.tx • Open • Cancel



Optibrium[™], StarDrop[™], Nova[™] Glowing Molecule[™] and Auto-Modeller[™] are trademarks of Optibrium Ltd.

• You will see a spreadsheet containing 309 structures and their measured affinities for Neurokinin 2 (in the column NK2 pKi).

🔩 StarDrop						×
<u>File Edit View Data Set Tools H</u> elp						
Andels Scoring Design Visualisation P450 torch3D		Structure	Compound ID	NK2 pKi		
Available Models	1	in the second	50116745	9.523		⊠ ∕‰
 □ logS □ logS □ logP □ logD □ logD 	2		50105594	9.523		
▷ □ 2.29 pki ▷ □ I-REG pl(250 ▷ ■ BBB log(brain)[blood]) ▷ ■ BBB log(brain)[blood]) ▷ ■ BBB log(brain)[blood]) ▷ ■ HAL category	3	da ja	50116733	9.398		1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
▷ □ P-gp category ▷ □ 206 affinity category ▷ □ PPB90 category ▷ □ PPB90 category	4	مكته	50110673	9.398		l∎‡ ≣⊳ f(x)
 ▷ ■ HBD ▷ ■ HBA ▷ ■ HBA ▷ ■ TPSA ▷ ■ Flexibility 	5	50	50151870	9.301		
 ▷ ■ Rotatable Bonds ▷ □ Legacy models ▷ □ Derek Nexus 	6	to or pr	50116727	9.301		RI
	7	an sh	50110650	9.301		
	8	197 197 197	50116082	9.222		
	9	96° 26°	50071112	9.222		
	*		Neurokinin2			
Ready					Server status: 🗿 🔘 🧿 Rows 309 (0) Columns 3 (0) Se	lected 0

- To explore the chemical diversity of this library and the distribution of potency across it, we are going to first take a look at a chemical space plot of this data. Click on the Visualisation tab on the left.
- Click on the Solution and then click "Create" to create a new chemical space plot.

Models Scoring Design Visualisation P450 torch3D	
Data Set Colour Size 🗘	
Neurokinin2 •	
Plot Customise Filters	
▮ 🕞 🖾 🗰 🖼 🗎 🔪 🖳	
Launched Drug Space Create	
Import	
Delete	
Trelis: <none></none>	
Trelis: <none> 📲 🚺 💆 💽</none>	

 In the dialogue box that appears, type in a name for the chemical space plot you're about to create – in the example below we have called it Neurokinin space.

Data Set:	Neurokinin2		•
Name:	Neurokinin space		
Similarity Model		Metho	d
Ochemical St	ructure Only	Vi	sual Clustering
Properties (Only	© P	CA
Mixed			
Available Columns		Col	umns to Use:
Chemical Struc		>> <	hemical Structure
Progress:			

- We can build different types of chemical space plots based chemical structure and/or property data, but in this case we're going to use the defaults, Chemical Structure Only and the Visual Clustering method, as shown above.
- Click the OK button and wait until the chemical space as been created and the dialogue box disappears. When the process is a chemical space will be plotted in the Visualisation tab.

Bestor Walkation P450 torch30 Image: Compound ID Image: Compound ID Image: Compound ID Image: Compound ID Image: Compound I
Neurokinin space 1 No. S0116745 9.523 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 3 1
0 0 0 0 1 0 0 0 2 0 0 0 3 0 0 0 4 0 0 0 5 0 0 0 5 0 0 0 5 0 0 0 0 0 0 0 0 0 0 0
Image: Colour Size Size Size Image: Deta Set Colour Size Image: Colour Size Image: Colour Size
Joint Set Colour Size Image:
1 2 5 50110673 9.388 5 5 50151870 9.301 Data Set Colour Size 1 5
Data Set Colour Size Image: Colour Size Image:
₩ Ψ Ψ
at Outcomise Filters 7 7 50110650 9:301 at at bit bit <td< td=""></td<>
Launched Drug Space Create Import Import
Delete 9 \$ 50071112 9.222

- Each point in the plot corresponds to one compound in the data set. Hover the mouse pointer over a point to see the corresponding structure.
- Selecting a row in the data set will select the point in chemical space and vice versa.

	8	Structure	Compound ID	NK2 pKi	A
Neurokinin space	67	aordy	50120190	8.745	
	68	and and	50120186	8.745	
-	69	and the	50110667	8.721	
12 a 3	70	an o de de de	50116726	8.721	
	> 71	مېليه ^م ېر	50116743	8.721	
Data Set Colour Size IP Neurokinin2 • J	72	ja Solar	50108242	8.699	
tot Customise Filters	73	A. S.	50116079	8.699	
Launched Drug Space Create Neurokinin space Import	74	197 197 197	50116717	8.699	
Delete	75	14. 20 20	50108238	8.699	

 Use the mouse to draw around a selection of points in the chemical space and they will be selected in the data set.

	8	Structure	Compound ID	NK2 pKi	
Scoring Design Visualisation P450 torch3D		de la	Compound to	- the pit	
	104	saira	50096948	8.222	
Neurokinin space		20			
*		. 8			
	105	- Alexandra			
		-+			
-	106	ztyte	50108270	8.222	
		8			
-	107	~al	50071115	8.187	
120.2	107	, Ă	500/1115	8.18/	
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		8			
	108	900.00 24	50108245	8.155	
	>				
Data Set Colour Size 🕆	109	za Laz	50110648	8.155	
leurokinin2 •		8			
Customise Filters		90. or			
	110	a de la compañía de l			
L 🕞 🗵 👬 🖗 🗎 📘 🔪 📓		, tite			1
aunched Drug Space Create	111		50096929	8.155	
Neurokinin space		~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~			
Delete	112	and. ma	50120200	8.149	
is: <none> 💌 🔡 👪 💽</none>		6			
	*				1

 To see the distribution of pK_i values across the chemical space, we need to change the key. Click on the block of colour next to the 'Neurokinin2 library' in the key, as shown above, to bring up the Format by Property dialogue.



• In the resulting dialogue, change the **Source** for the colour to **NK2 pKi** and click **OK**.

	urce: NK2 pKi		▼ 🗖 Los , dmic
	Name	Colour	
1	High		
2	Low		
	Interpolate from	4.14 ta 2 - Threshold:	o 9.52

- The compounds with the highest pK_is will be highlighted yellow, the lowest red.
- You will notice that a large proportion of the compounds with the highest pK_i values come from one of the three regions where compounds are clustered. To confirm this, select the first 50 rows (the data is already sorted) by clicking on row 1, holding down the shift-key and then scrolling to and selecting row 50.

Is Scoring Design Visualisation P450 torch3D		Structure	Compound ID	NK2 pKi	
Neurokinin space	46				
	47				
-	48				
10 a X	49				
· · · · · · · · · · · · · · · · · · ·	50				
Data Set Colour Size Image: Colour Image: Colour <t< td=""><td>51</td><td>Angel a</td><td>50116737</td><td>8.921</td><td></td></t<>	51	Angel a	50116737	8.921	
ot Customise Filters	52	and the second	50110630	8.921	
Launched Drug Space Create Neurokinin space Import	53	and the	50110633	8.921	
rells:	54	in the second se	50116084	8.886	

This may be the best region in which to focus for selecting compounds, but first we should consider the other properties that are important in a high quality lead.

- 0 - X

We're going to keep this plot to use later, so
 click on the Sutton in the Visualisation tab
 to create a separate window containing this
 plot.

If you wish, you can minimise this window to keep it out of the way for now.

👯 Neurokinin2		
N	eurokinin spa	ace
-		
	-	
Data Set	Colour	Size
Neurokinin2	NK2 pKi	•
Plot Customise Filters		i 🖬 🔪 🗳
Launched Drug Space Neurokinin space		Create Import Delete
Trellis: <none></none>		- # 🚺 🛃 💽

- We have no further measured data, 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).

todels Scoring Design Visualisation P450 torch3D 4		Structure	Compound ID	NK2 pKi	
Available Models	46			9	
▷ ♥ IogS IogS © pH7.4 ▷ ♥ ■ IogP ▷ ♥ ■ IogP ▷ ♥ ■ IogP ▷ ♥ ■ IogP	47			9	
 ✓ ■ LSO pN ✓ ■ LERG p1CS0 ✓ ■ BBB log[[brain];[blood]] ✓ ■ BBB category ✓ ■ LA category 	48			8.959	
	49			8,959	
 ▷ ♥ ■ HBD ▷ ♥ ■ HBA ▷ ♥ ■ HBA ▷ ♥ ■ FISA ▷ ♥ ■ Flexibility 	50	्व स्व ब्रह्म	50116723	8.921	
▷	51	مايي م ^{ول.}	50116737	8.921	
	52	مكمته	50110630	8.921	
	53	a to	50110633	8.921	
	54	16.08°	50116084	8.886	

 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.



- Scroll across the data set to see the numerical and classification predictions that have been made. Due to the volume and complexity of the data, it is challenging to find which compounds have the best balance of properties. So, we're going to us StarDrop's scoring tool which will make it easy to assess all of this information together.
- Click on the **Scoring** tab.

e <u>E</u> dit <u>V</u> iew <u>D</u> ata Set <u>T</u> ools <u>H</u> elp							
dels Scoring Design Visualisation P450 torch3D (*)		Structure	Compound ID	NK2 pKi	logS	logS @ pH7.4	■logP ▲
offie: NewProfile Profile Profile Desired Value Importance	46	مير مري	50116089	9	2.392	3.386	4.622
	47	2007 S	50129513	9	1.613	1.221	4.971
	48	2. 19 19	50116086	8.959	2.473	2.826	4.204
	49	and the second	50116744	8.959	0.2145	2.857	5.649
kald rule Delete	50	444	50116723	8.921	0.5373	1.835	6.24
lable properties: operty Criteria Importance 206 affinity coteg BB8 category	51	Angeler and	50116737	8.921	0.3553	1.857	5.368
2C9 pKi Flexibility ed profiles:	52	هريط	50110630	8.921	1.634	2.632	4.415
ravenous CNS Scoring Profile ravenous Non CNS Scoring Profile inski Rule of Five I CNS Scoring Profile	53	and the	50110633	8.921	1.811	2.934	5.085
al Non CNS Scoring Profile	54	d d	50116084	8.886	1.939	2.971	4.897
Id profile Sensitivity	×	III	Neurokinin2				F

- A scoring profile allows you to define a set of criteria that is important for your project. StarDrop has some example profiles built in, but in this case we're going to load a profile designed for this Neurokinin 2 project.
- Click the button to open a new scoring profile and load the profile called NK2
 Project.apd.

👯 Load Profile	10 m		×
🕞 🗢 🕨 🕨 Librar	ries + Documents + StarDrop Example	• • j	Search StarDrop Example
Organize 👻 New f	older		8≡ ▼ 🚺 🔞
☆ Favorites ■ Desktop	Documents library StarDrop Example		Arrange by: Folder -
🗼 Downloads	E Name		Data modified
 Dropbox Recent Places Google Drive 	NK2 Project.apd		08/06/2012 09:45
Libraries Documents Music Fictures	- 4 [III		,
🖼 Videos Fi	le <u>n</u> ame:	• [coring Profiles (*.apd)

 A scoring profile contains a series of properties along with criteria describing desired values and their relative importance. In this profile we are looking for compounds with good affinity and which are suitable for a peripheral target.

Models Scoring Design Visualisation P450 torch3D Nova		Structure	Compound ID	NK2 pKi	logS	logS @ pH7.4	■logP ^
Visualisation P450 torch30 Nova Visualisation rofile: N42 Project Profile Designed Value Importance	46	46.081.	50116089	9	2.392	3.386	4.£
NK2 pKi > 7 log5 > 1 HIA category + ooP 0 -> 5 / 1	47	60.75 76	50129513	9	1.613	1.221	4.5
ingp 0→3 5 1 hERG pIC50 ≤ 5 220 pKi ≤ 6 200 affinity category low medium 1	48	and the second	50116086	8.959	2.473	2.826	4.2
PPB90 category low ■ BBB category - - ■ BBB log([brain]:[blood]) ≤ -0.5 ■	49	and	50116744	8.959	0.2145	2.857	5.6
Add rule Delete 🔐 😹 Sort Edit	50	to the second	50116723	8.921	0.5373	1.835	6.
Property Criteria Importance 206 affinity categ BBB category	51	ىلى مىرىيە	50116737	8.921	0.3553	1.857	5.3
2C9 pKi Flexibility ved profiles:	52	مركبها	50110630	8.921	1.634	2.632	4.4
IK2 Project Irtavenous Non CNS Scoring Profile Irtavenous Non CNS Scoring Profile ipinski Rule of Five	53	and the second	50110633	8.921	1.811	2.934	5.0
Oral CNS Scoring Profile Oral Non CNS Scoring Profile	54	A. O.	50116084	8.886	1.939	2.971	4.5
Build profile Analyse Sensitivity	×	m	Neurokinin2				۶.

- Run the scoring by clicking the button in the Scoring tab.
- A new column will be added to the data set containing a score for each compound, taking into account the property criteria and their importance in the profile, as well as the uncertainty in the underlying experimental and predicted data. The score is a value between 0 and 1, representing the likelihood of success of the compound against the profile we have defined.

odels Scoring Design Visualisation P450 torch3D Nova 4	NK2 Project	Structure	Compound ID	NK2 pKi	logS	logS @ pH 🕯
Otes Otes <th< th=""><th>1 0.04</th><th>303 J</th><th>50116745</th><th>9.523</th><th>1.334</th><th>2.8</th></th<>	1 0.04	303 J	50116745	9.523	1.334	2.8
INK2 pki > 7 I log5 > 1 H4 category + log9 0 -> 5 1	2 0.34		50105594	9.523	1.025	3.5
hERG plC50 ≤ 5 ■ 2C9 pKi ≤ 6 ■ ■ 2D6 affinity category low medium 1 ■ ■	3 0.03	386 ×	50116733	9.398	1.121	3.0
PPB90 category low ∎ BBB category - - - BBB log([brain]:[blood]) ≤ -0.5 -	4	-Q	50110673	9.398	1.466	2.5
Add rule Delete	5 0.17	33	50151870	9.301	1.458	1.3
roperty Criteria Importance 206 affinity categ BBB category	6 0.04	139 5 p.	50116727	9.301	1.71	2.8
2C9 pKi Flexibility ved profiles:	7 0.06	335	50110650	9.301	2.836	3.2
IK2 Project Itravenous CNS Scoring Profile Itravenous Non CNS Scoring Profile Ipinski Rule of Five	8 0.06	518	50116082	9.222	2.473	2.8
Dral CNS Scoring Profile Dral Non CNS Scoring Profile	9 0.17	33	50071112	9.222	1.458	1.1

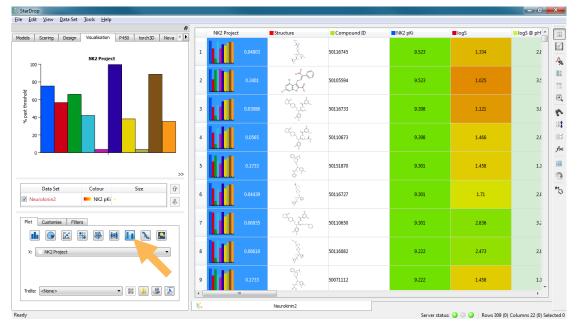
 To find the highest scoring compounds, right-click on the scoring column header and choose Sort->Descending to sort the data set from high to low score.



The histogram in each cell gives a quick overview of the impact of each property on the score for a compound. As you scroll down the data set, and the scores become lower, the histograms with low bars indicate the properties that have not met the requirements defined in the project profile, taking into account the confidence in the data and importance of each property.

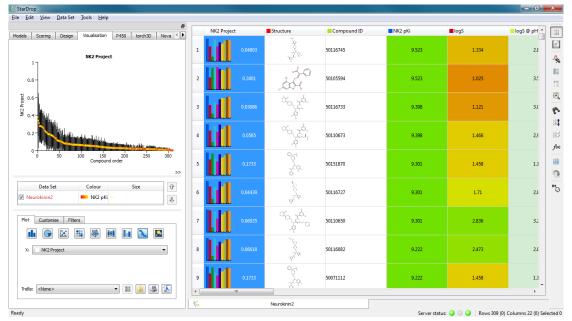
Having scored the compounds we're now going to take a look at the overall trends across the data set and consider which compounds we should select.

- Click on the **Visualisation** tab again.
- Select the data set column containing the scores (NK2 Project) and choose the button as shown below. The resulting scoring histogram shows the profile of properties for the overall set; each bar represents one property in the profile, identified by the colour. The y-axis shows the percentage of compounds in the set that meet the ideal criterion for each property, as defined in the scoring profile.

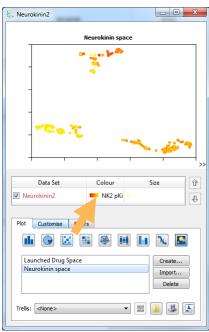


We can see that for some of the property criteria we have specified in our scoring profile, in particular hERG inhibition and Plasma Protein Binding, there are very few compounds which meet the criteria and hence these may be consistent issues for the chemistry in this data set.

button on the **Visualisation** tab to change the plot type. A Now select the 'snake plot' shows the scores (on the y-axis) for all compounds in order from highest score to lowest score (along the x-axis). The overall uncertainty in each score, due to the uncertainty in the underlying data, is also displayed as an error bar around each point.



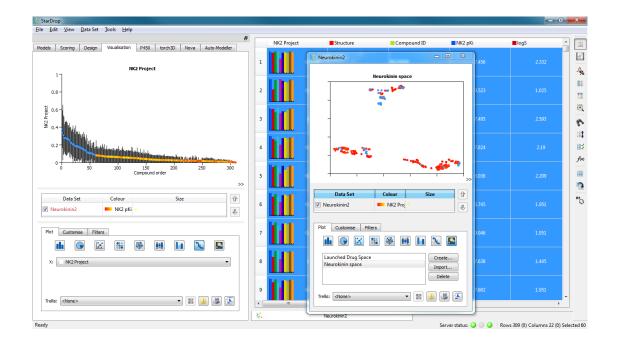
- Now open the detached chemical space plot we created previously.
- Click on the block of colour next to the 'Neurokinin2 library' in the key in this window to bring up the Format by Property dialogue again and, as we did before, change the **Source** property to be the NK2 Project values. This way the data will be coloured based upon the overall scores of the compounds, not only the pK_i values.



Here we can see that the pattern has changed dramatically from that we could see when looking only at potency. The cluster with the highest potency is now quite red, because it is unlikely to have acceptable ADME properties. However, we can identify 'hot spots' in yellow which are likely to achieve both good potency and ADME properties.

K Neurokinin2	
Neurokinin space	
	~
Data Set Colour Size 🗘]
V Neurokinin2 NK2 Pro	
Plot Customise Filters]
Launched Drug Space Create Neurokinin space Import Delete	
Trelis: <none></none>	

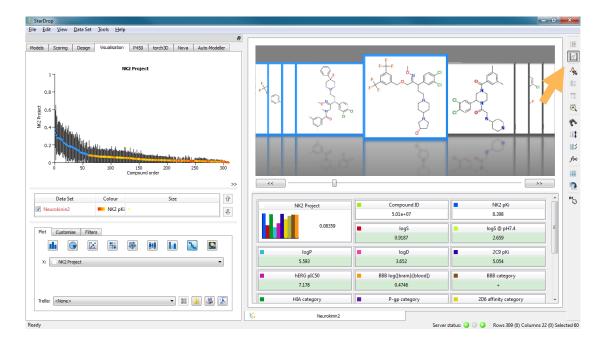
Looking at the 'snake plot' we can see that the top 50 compounds cannot be confidently distinguished from the top scoring compound (notice that the error bar for the top-scoring compound overlaps with the error bars of approximately the top 50 compounds). Therefore, we should consider selecting compounds from this range to further explore their properties and make a confident selection of potential lead series. Selecting the top 50 compounds in the snake plot (by drawing a ring round them with the mouse) highlights where these lie within the chemical space.



Notice that this selection includes compounds from different regions of the chemical space. This suggests that some of these chemistries cannot be rejected with confidence. Therefore, it may be more appropriate to sample some of these alternative chemistries to generate some experimental ADME data. These data will have greater confidence that predicted values and hence we will be able to identify with confidence the chemistries that will yield a high quality lead series.

Having looked at the data on a project scale, the next question is whether there are any compounds that, with minor modification, could also achieve a good balance of properties.

 Close the window with the chemical space and click the button on the righthand tool-bar to switch into the molecule view.



When the cell containing the score is selected a heat map is displayed highlighting the properties included in the score. Those meeting the criteria with confidence are highlighted green and those that may pose a risk are coloured yellow through to red, taking into account the importance of the property and the uncertainty in the data.

Scroll through the set by clicking on the molecules or using the \square

and

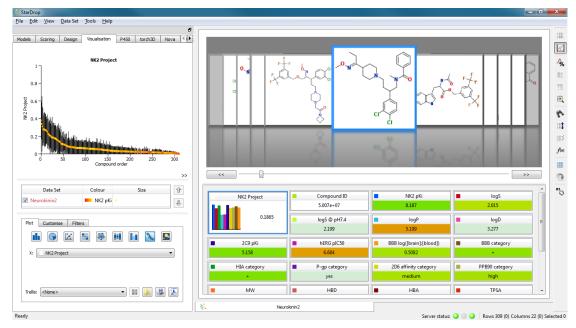
>>

_____ buttons.

Select the cell titled Compound ID and click the button on the toolbar to the right to bring up the Find dialogue.

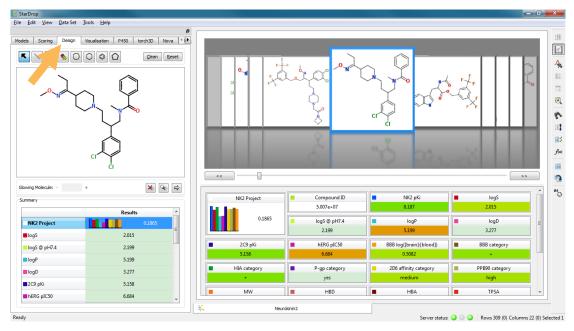
Find	? ×
Find what: 50071115	
Search Options	Your search can be limited by selecting columns.
Match case	
Find whole words only	Find Next
	Find All
Use wildcards	Cancel

Type in 50071115 and click Find Next (as shown above) to search for this compound.
 Click the Cancel button to close the Find dialogue and select the score cell again to see the heat map.



This compound has good affinity for the target and most of the predicted properties are acceptable. However, from the heatmap and score histogram we can see that the predictions suggest potential issues with logP (light blue bar) and hERG inhibition (pink bar). Therefore, this may be an opportunity to attempt to improve on the compound by making minor modifications.

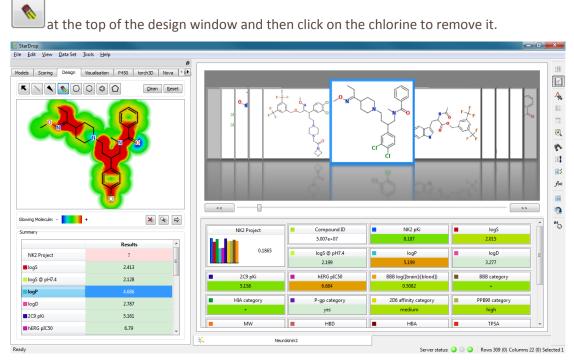
 Switch to the **Design** tab and click on the compound structure in the Molecule View. (If necessary drag the bar in between the drawing area and the summary table below the editor to make the drawing area taller and drag the bar between the Design tab and the data table to make it wider).



• Click on the logP value in the summary table below the Design window.



This highlights the molecule, using StarDrop's Glowing Molecule, to indicate which regions are having the greatest influence on the predicted value. Areas highlighted red tend to increase the property value and areas highlighted blue tend to decrease the property value. For logP we would like the value to be less than 5 so we should concentrate on making changes to the regions which are red to reduce the logP. • To remove one of the chloro substitutions on the phenyl ring, click the eraser button



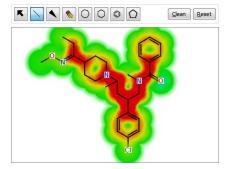
- In the summary table below, notice that the predictions for this new molecule are updated instantly. Importantly the logP is now lower.
- Scroll down the summary table and select the hERG prediction to see the Glowing Molecule for this property. We would ideally like the hERG pIC₅₀ to be under 6.

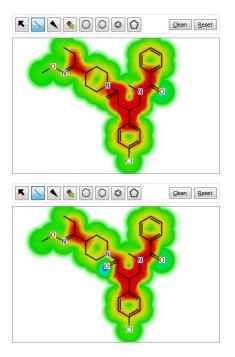


Here we can see that the basic Nitrogen on the piperidine is having a significant effect to increase the hERG affinity. We will change this amine to an amide and explore the effect this has.

- Select the bond tool and click on the carbon chain atom next to the amine and while the mouse is clicked, drag away and then release to draw a methyl group. Make sure you don't connect this to any other atom! (If you do, hover the mouse pointer over the drawing and press ctrl-Z to undo it).
- Now click on the new bond you have drawn to make this a double bond.

 Finally, point the mouse at the carbon at the end of the double bond and press the 'O' key to replace the Carbon with Oxygen and complete the amide.





 Notice that the amide O has a blue glow, indicating that it is having an effect to decrease the predicted hERG pIC50 and that these changes have not-only reduced the predicted hERG affinity, but have also reduced the logP value further.

If you scroll up the summary table to the top again you will see a question mark against the score. This is because we do not yet have an experimental pK_i value for the NK2 affinity. To generate predictions of NK2 affinity we can use the StarDrop Auto-modeller to build a new model of the data we have. We will not cover the Auto-Modeller in this exercise, but if you would like to see a demonstration please contact <u>stardrop-support@optibrium.com</u>.

This example illustrated how we are able to explore the chemical space of our project to look for compounds which have a good overall balance of properties. We have also seen how we are able to focus in on individual compounds and use the Glowing Molecule to help design compounds which better meet our project needs.