StarDrop™ Worked Example:

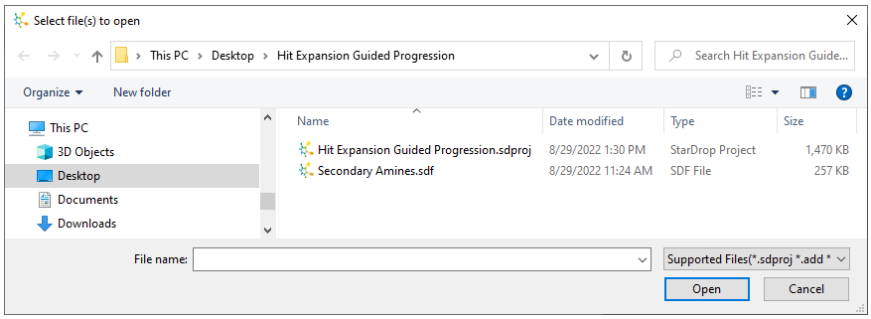
Hit Expansion & Guided Progression using Surflex eSim3D™ and SeeSAR™

This worked example uses StarDrop's Surflex eSim3D and SeeSAR modules to assess a small virtual library of compounds for their similarity to the binding of known Heat Shock Protein 90 (HSP90) inhibitors. The library will be created using StarDrop's R-group clipping tool to quickly transform chemical building blocks into their corresponding substituents through a scaffold-based library enumeration in StarDrop's Nova module. The objective in this example is to use the Surflex eSim3D module first to understand the 3D structure-activity relationships (SAR) and then to create a binding hypothesis. Following the enumeration of the new library, the compounds will be triaged to determine which ones best fit the binding hypothesis. Ultimately, the best scoring compounds will then be docked against HSP90 to provide an estimated ability of the redesigned compounds to bind the active site of the target.

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.

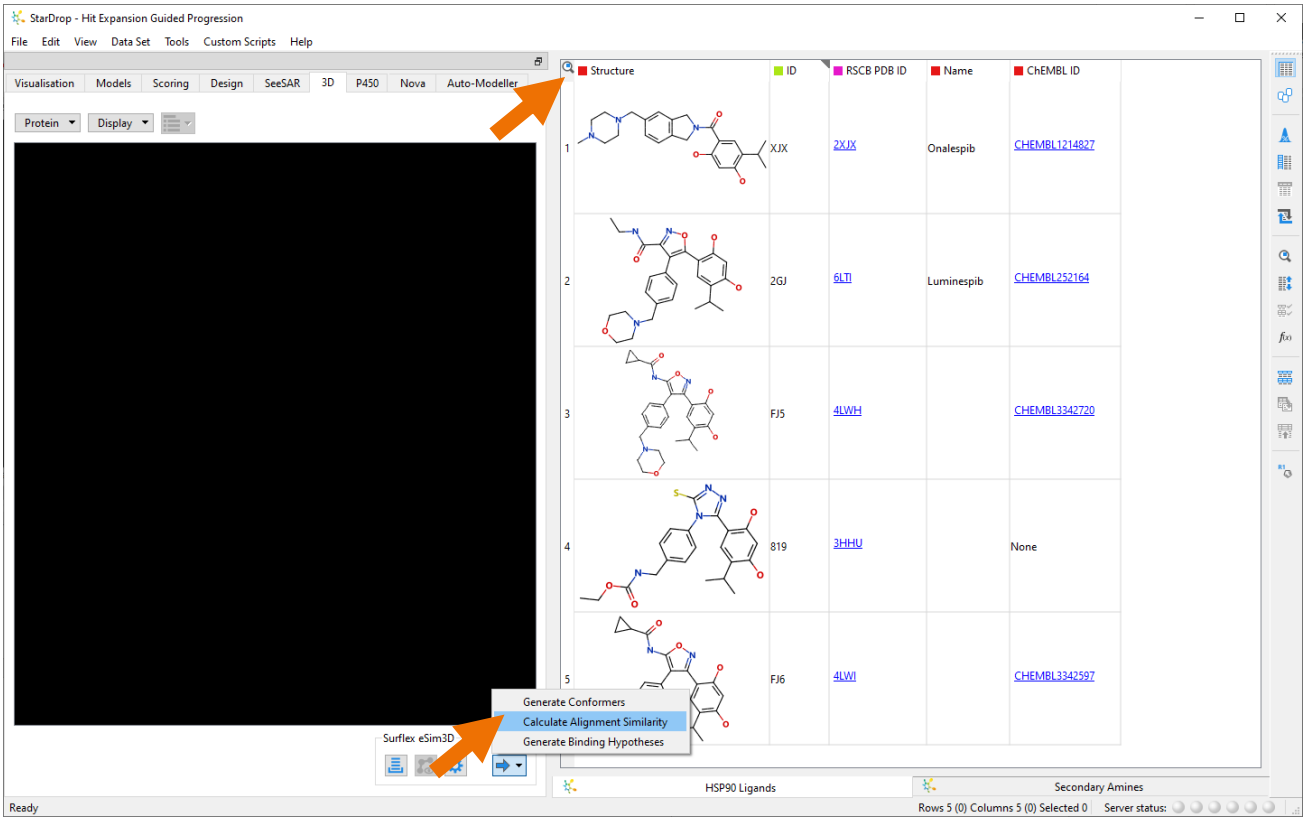
Exercise

- In StarDrop, open the project file **Hit Expansion Guided Progression.sdproj** by selecting **Open** from the **File** menu.



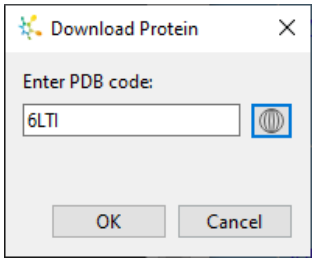
This opens a data set of five known HSP90 inhibitors. The compounds have all been extracted from the RCSB Protein Data Bank (PDB) (https://www.rcsb.org/) along with their ligand and PDB IDs. The compounds share a common resorcylic moiety but differ slightly in their scaffold cores. Three of the molecules (**FJ5**, **FJ6**, and **2GJ**) contain an isoxazole core; however, **2GJ** inverts the ring Nitrogen and Oxygen positions. Analogous to the 5-membered isoxazole, compound **819** possesses a 1,2,4-triazole, whereas the last group member (compound **XJX**) contains an isoindolinamide core. We will first perform an analysis to investigate the similarity of these compounds in terms of their shape, electrostatic potential, and hydrogen bonding potential

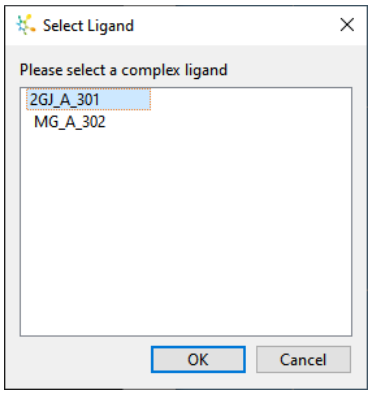
- Select all compounds of the HSP90 ligands data set by clicking in the top left corner of the data set.
- Click the **Go button menu** and select **Calculate Alignment Similarity**.



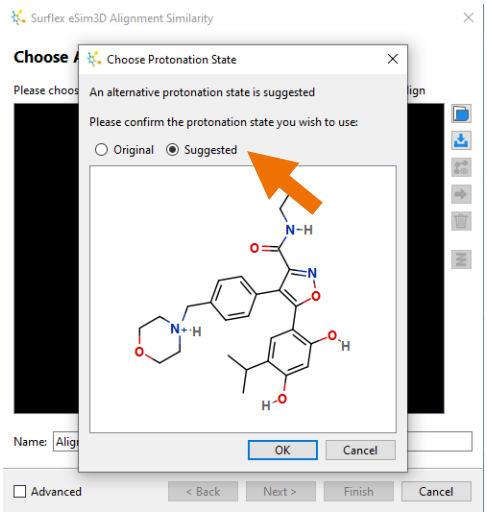
This will prompt you to choose a reference molecule with which to align your compounds. We will use the HSP90 inhibitor **2GJ** (also known as Luminespib), to obtain the bioactive conformation of Luminespib from a co-crystal structure.

- Click the **download** button  to obtain the ligand from a PDB reference.
- Enter the PDB code **6LTI** and click **OK**.
- Select ligand **2GJ_A_301** and click **OK**.





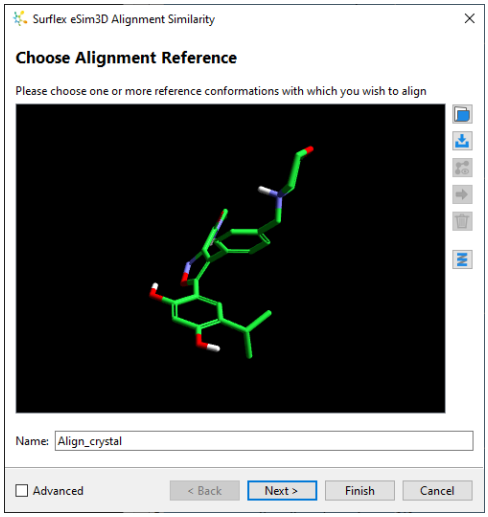
A dialogue box appears, suggesting an alternative protonation state for the nitrogen on the piperazine ring.

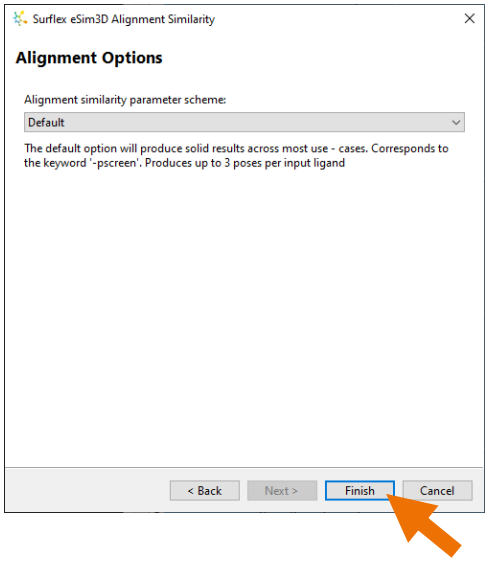


- Click **OK** to accept the protonation.
- In the **Name** textbox, enter **Align_crystal**.

There are further options, such as the ability to add torsional and positional constraints if you select the **Advanced** options. However, we will use the defaults.

- Click **Next**.





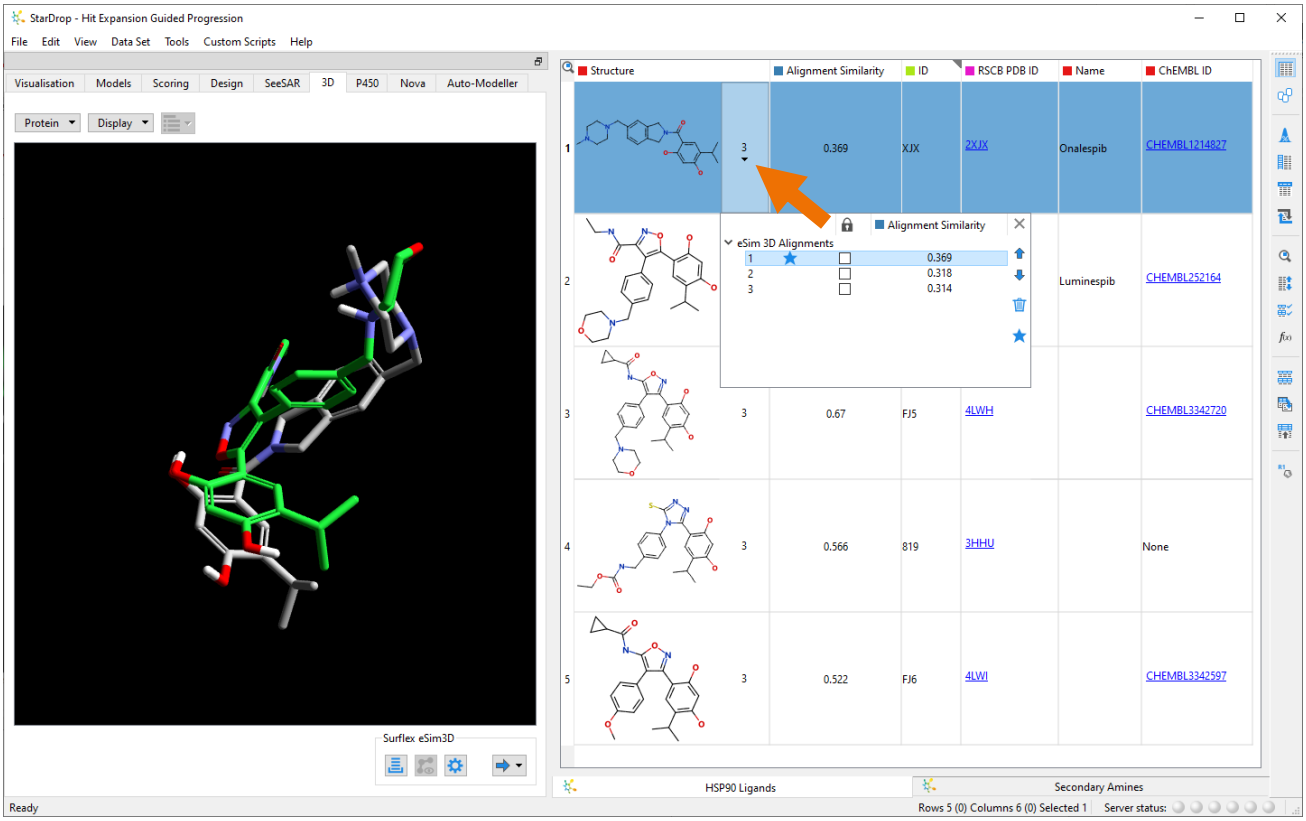
This page gives the option to define the parameter scheme of the alignment similarity analysis. The parameter schemes vary in terms of speed and accuracy.

- Click the **Finish** button to begin running the alignment with the default option.

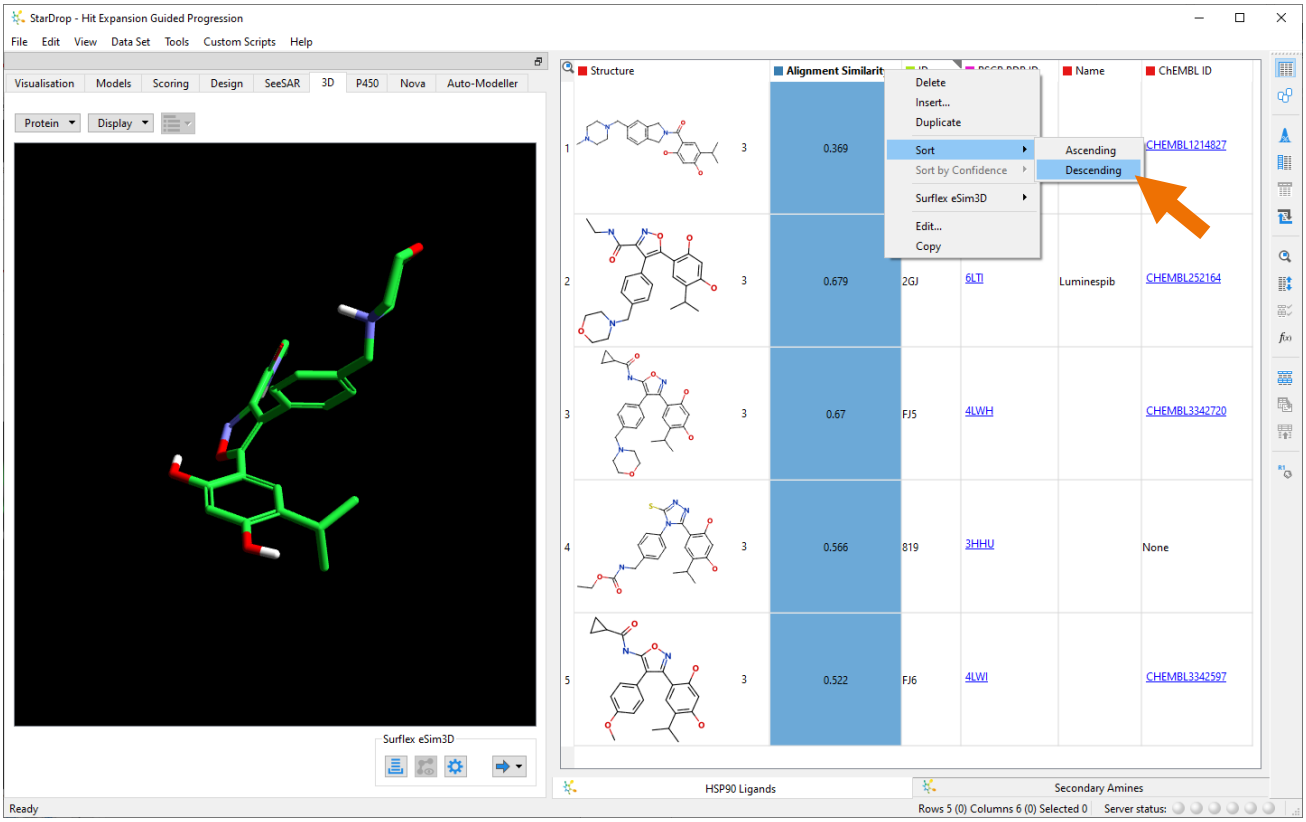
Once the alignment is complete, the calculated similarity value for each molecule is added to our data set in the new **Align_crystal** column. A number is added to the **Structure** column, which is the number of alignments that have been generated. We can view each of the alignments by clicking the number next to the structure and selecting from conformations contained in the pull-down menu. The best scoring conformation is the primary pose denoted by the ★ symbol.

- Select compound **XJX** (Onalespib) to show it in the 3D viewer aligned to **2GJ** (Luminespib).
- Click on the number of alignments to view the aligned conformations.

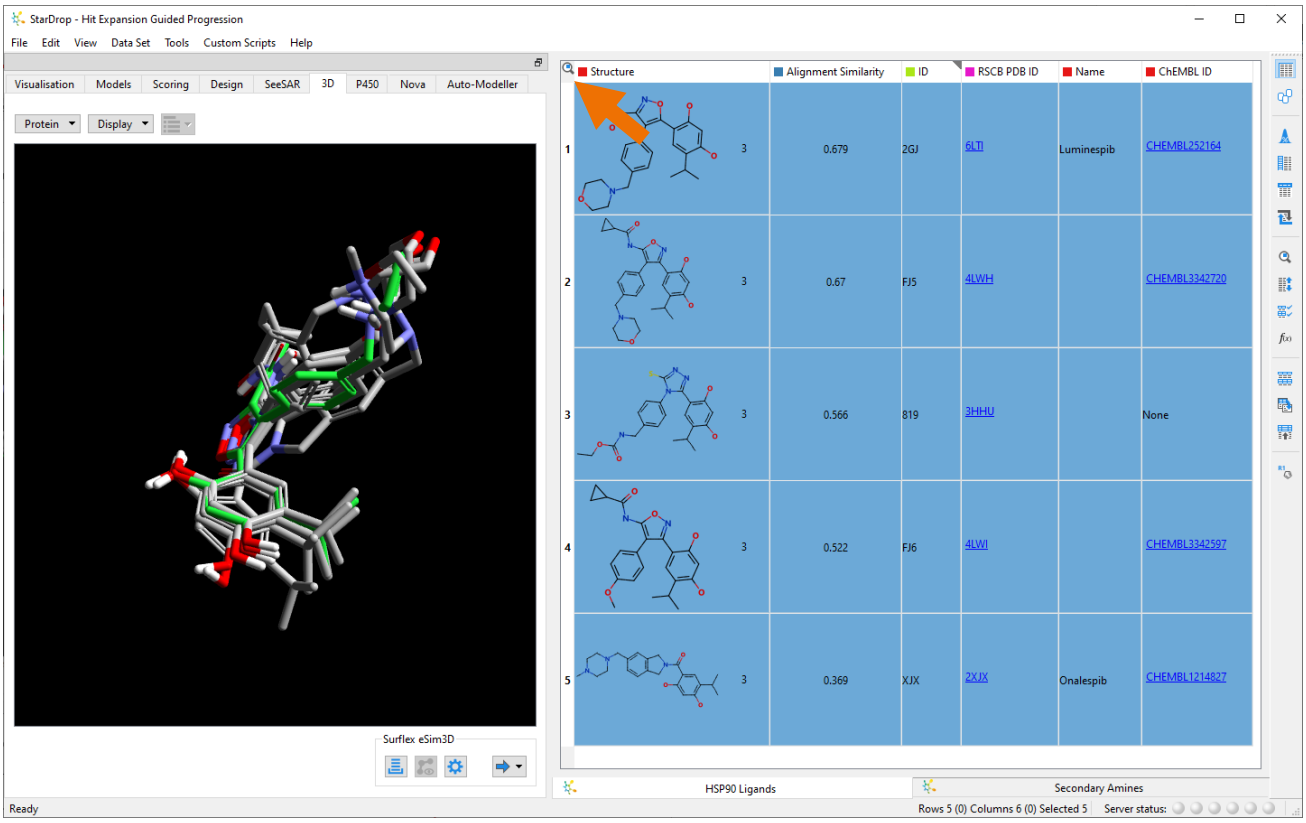
Note: the primary pose can be changed by selecting the conformation and then clicking the star ★ button on the right of the table. The primary pose is the one that is shown in the 3D viewer when the row is selected.



- Right-click on the **Align_crystal** column to bring up the menu and choose **Sort**, then **Descending** to bring the compounds with the highest similarity values to the top.

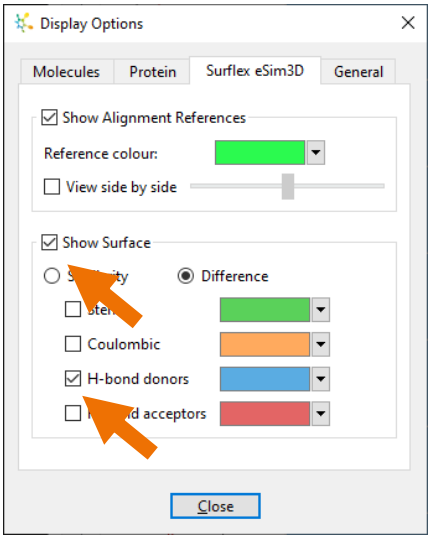


- Select all the ligands by clicking in the top left corner of the data set again.



Following an inspection of the conformations of each ligand, we can see that the 5-membered isoxazole or triazole core-containing compounds are most like the reference Luminespib. However, the isoindolinamide containing Onalespib deviates from that of the reference compound. While the molecules' beta-resorcylic and solvent-exposed tail regions are reasonably aligned, the isoindoline core of Onalespib bridges the two regions differently compared to the other compounds. Beyond the scaffolds' overlap, we can examine the similarities and differences of the Sterics, Coulombics, Hydrogen-bond donor, and Hydrogen-bond acceptor overlap of the molecules under study.

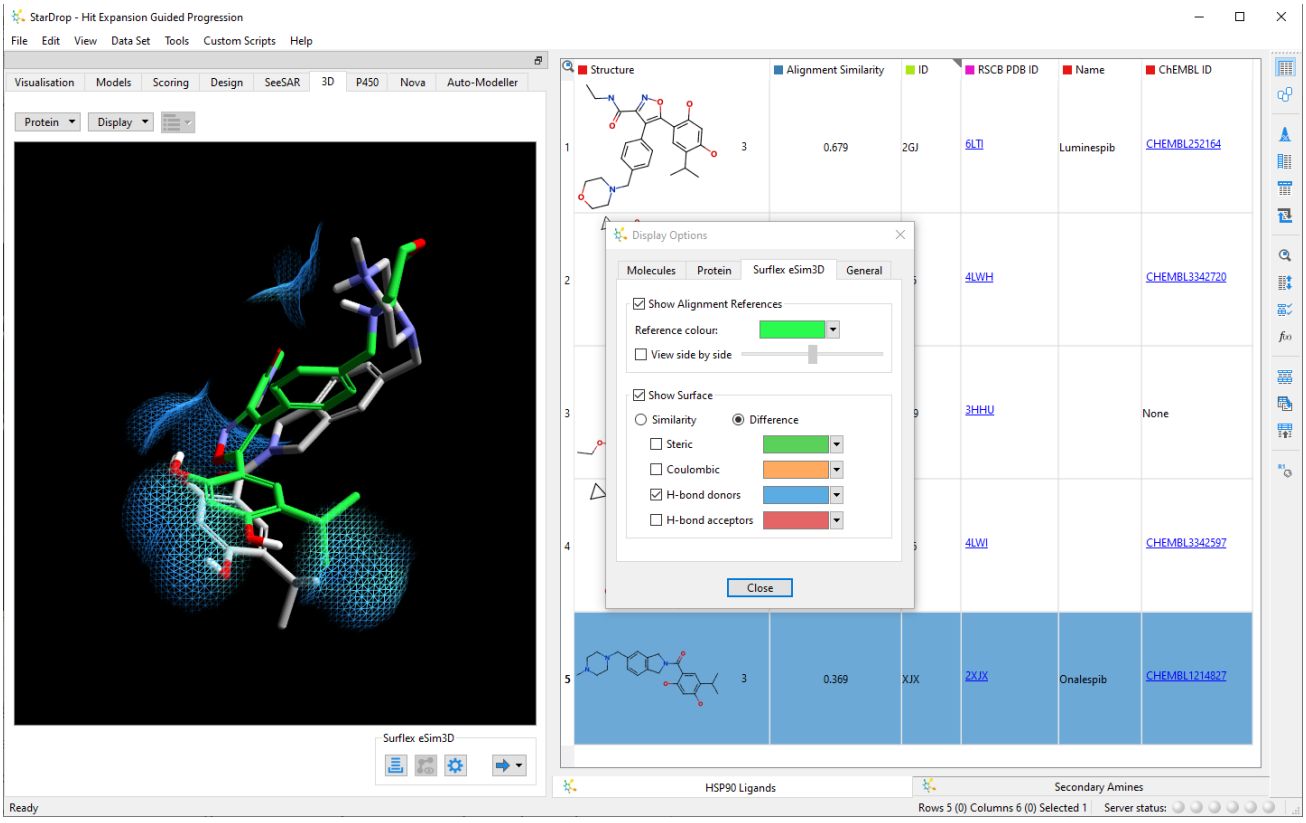
- Select only Onalespib.
- Select the  button to bring up the Surflex eSim3D display options.
- Select the **Show Surface** option and explore the similarity and dissimilarity surface maps to see which regions of Onalespib and Luminespib align well.
- Finally, select the **Difference** option and tick the box next to **H-bond donors**.



These surfaces highlight regions of similarity and dissimilarity in terms of four properties:

1. Steric – the shape of the molecule
2. Coulombic – the electrostatic potentials of the molecule
3. H-bond donors – where these are positioned
4. H-bond acceptors – where these are positioned

We can see a difference in the Hydrogen-bond positioning between the two compounds highlighted around phenolic portions of the resorcylic group. The same can be found with the analysis of the Hydrogen-bond acceptor regions if the differences are also examined.



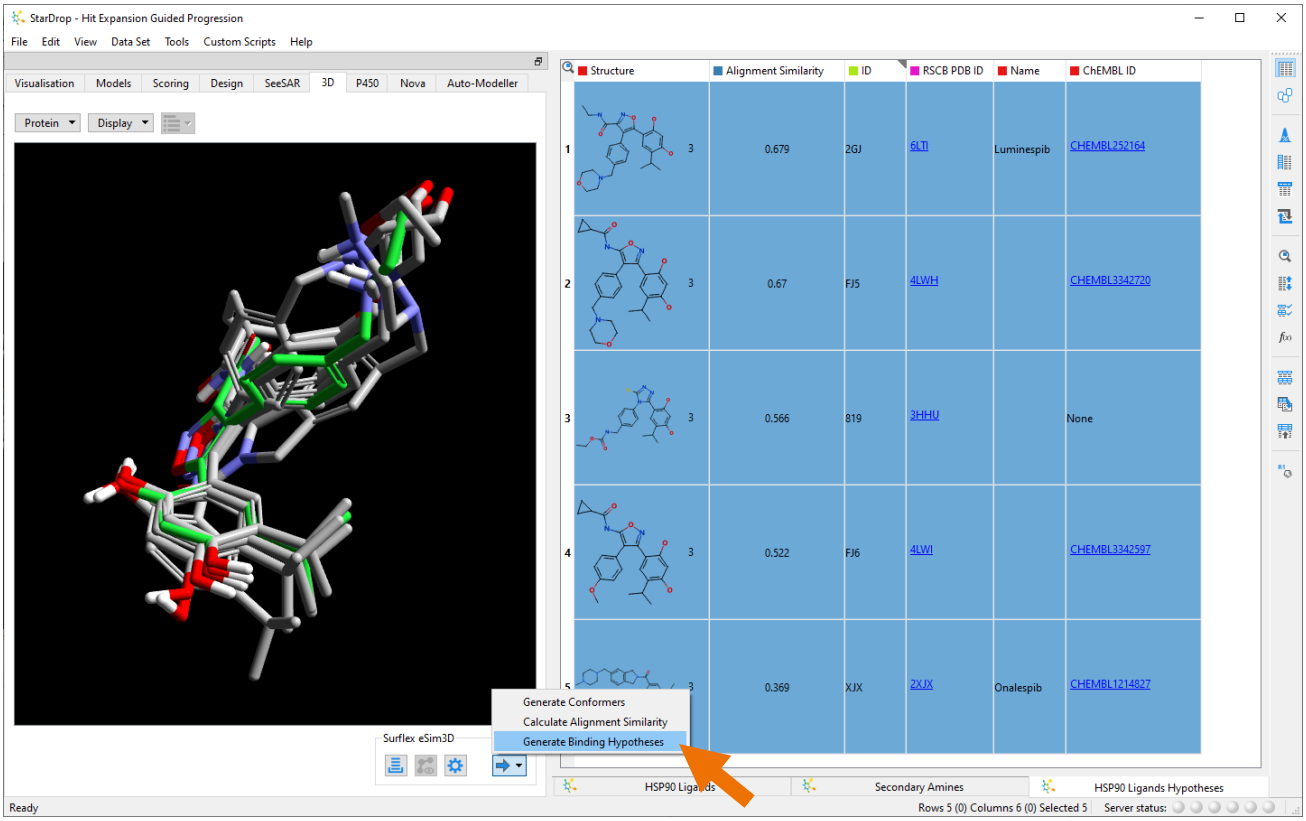
While Onalespib and Luminespib are equally potent inhibitors of HSP90 (18nM and 13nM, respectively), the two compounds bind the target in different conformations. Perhaps there is an opportunity to explore alternative, possibly more flexible, amides around the Onalespib core to find an inhibitor that occupies a bound conformation somewhere in-between the two being studied that may offer improved potency or other beneficial physicochemical properties.

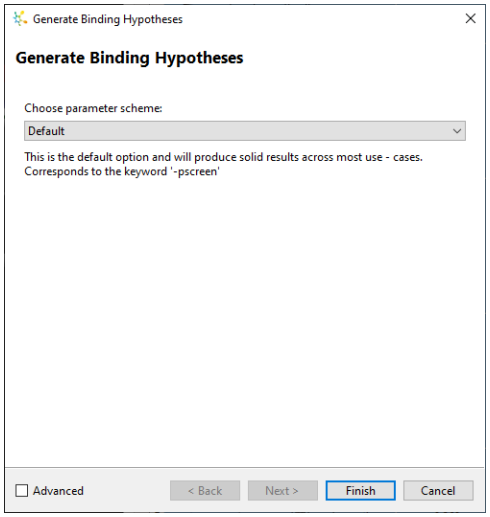
- Untick the **Show Surface** box.
- Click **Close** to close the Display Options.

The Surflex eSim3D technology can simultaneously align multiple ligands to generate a series of consensus alignments that serve as a binding pose hypothesis. This hypothesis can be used as a reference for subsequent compound library alignments. This approach is helpful for projects without structural information about the protein, but known active ligands are available.

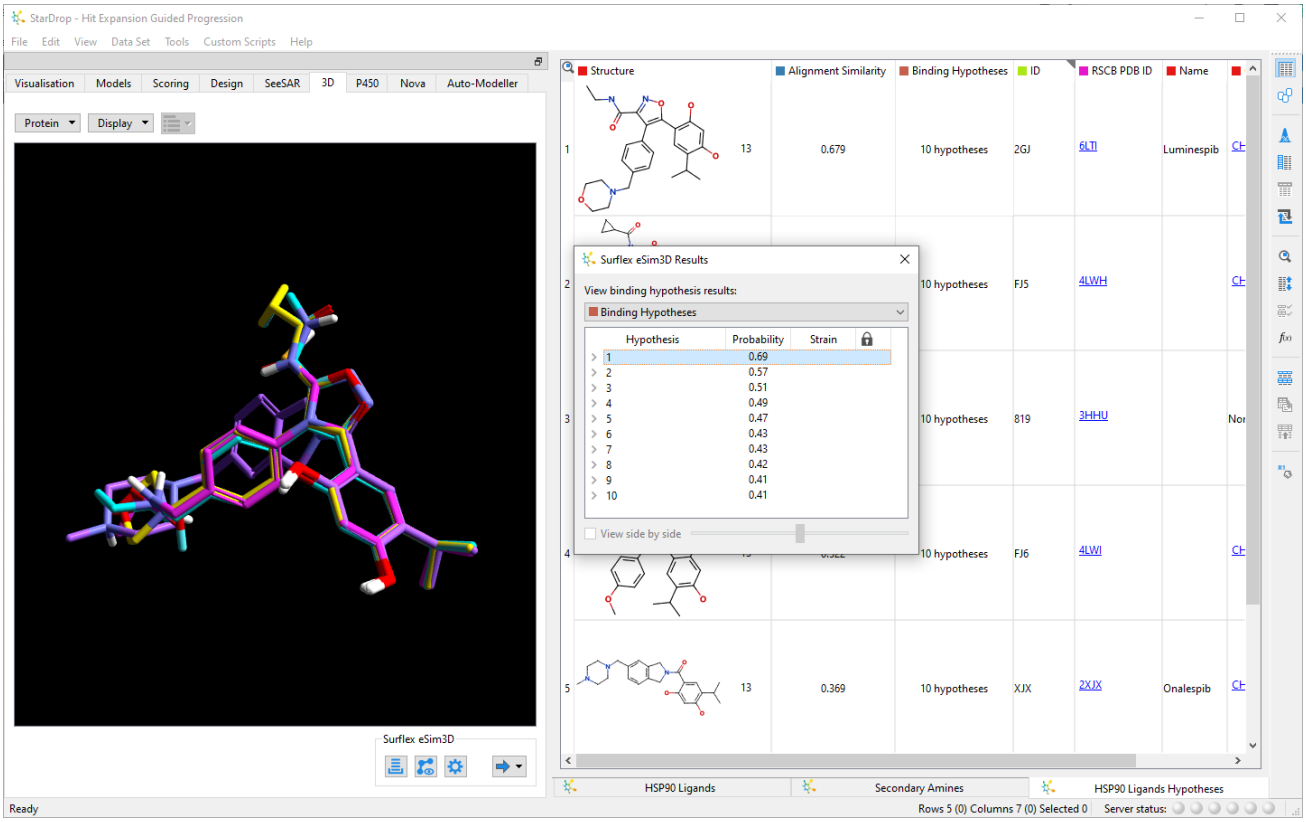
- Again, select all of the compounds in the data set by clicking in the top left corner of the data set.
- Click the **Go button menu**  , but do not make a selection.

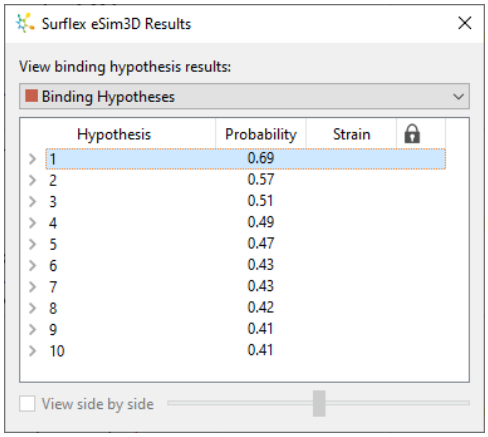
You will see that the bottom choice is listed as **Generate Binding Hypothesis**. Due to the time required to calculate the Binding Hypothesis, the results have already been precalculated and are available in a separate data set (**HSP90 Ligands Hypotheses**). An example image of the selection choices and the Generate Binding Hypothesis dialogue window are shown in the following images.

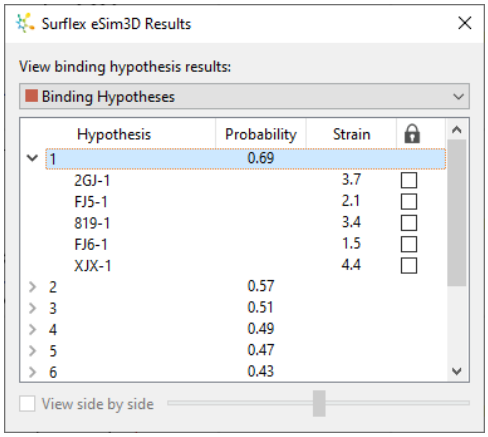




The resulting binding conformations are displayed in the 3D viewer. The hypotheses are shown in the results list with a probability value for each hypothesis. The most likely hypothesis is shown at the top. You can click the arrows to expand the tree and view the conformations that make up the binding hypothesis. The strain on the conformations is shown, the units of which are kcal/mol. The checkbox under the lock symbol can be ticked to keep the conformation(s) shown in the 3D viewer.







Note: If you accidentally close the Binding Hypothesis window and want to access it again, click the **Binding Hypothesis** button under the 3D viewer.

Ten binding hypotheses have been generated, and the probability of each hypothesis being accurate is shown. Let's examine the binding hypothesis with the highest probability and look at the strain associated with Onalespib (XJX). It is noticed that it is calculated to be slightly higher as compared to the other compounds but similar to the reference Luminespib. In other cases, the values for the strain may be something to consider when deciding on a working hypothesis and whether to more closely investigate one of the hypotheses generated with a lower predicted probability. For this exercise, we will proceed from the Binding Hypothesis with the highest probability.

We will now prepare a data set of virtual compounds based on an amide coupling reaction of the beta resorcylic acid core of Onalespib and a library of commercially available secondary amines. Once the new library is generated, we first evaluate their 3D conformations against Luminespib and select analogues that most closely match the binding conformation of Luminespib versus Onalespib. Those molecules will then be docked against the HSP90 target to ascertain if they potentially score higher than either of the two reference ligands.

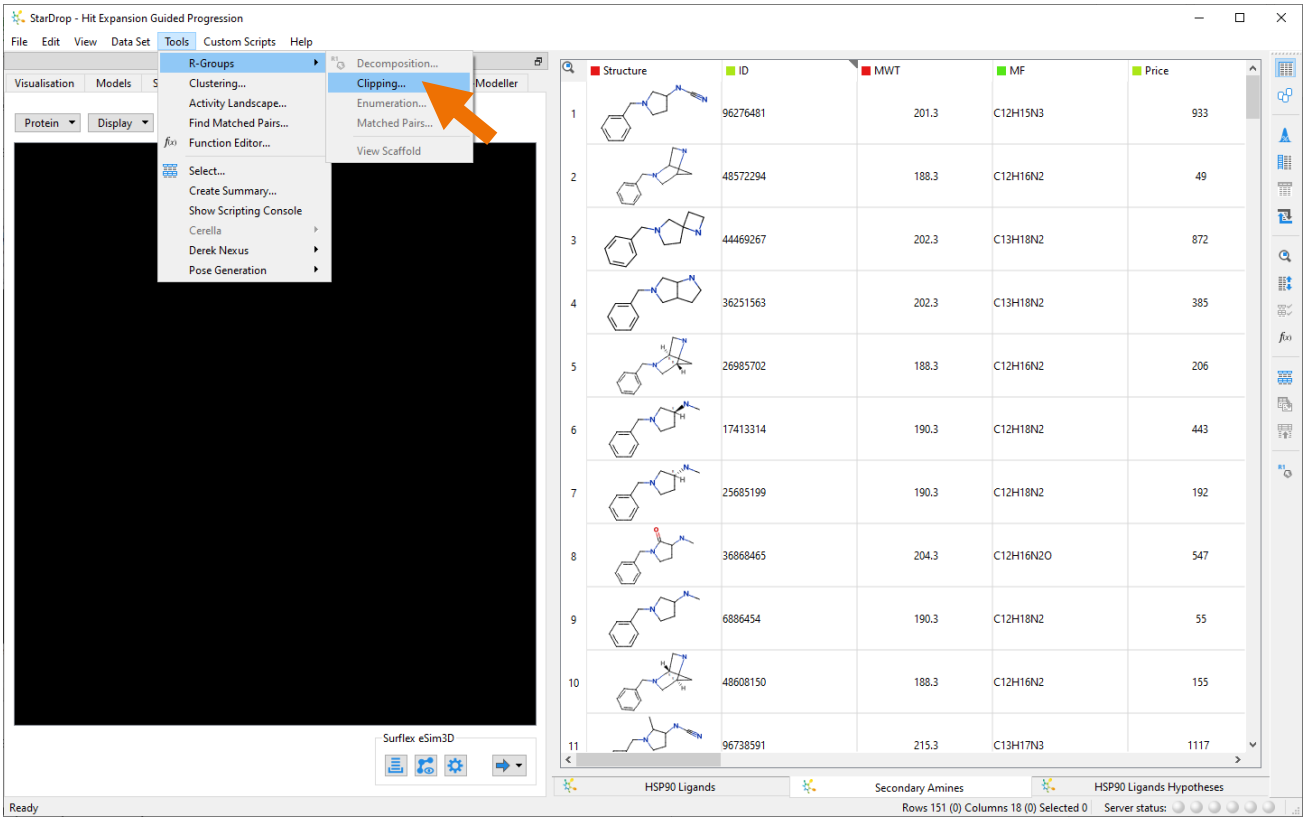
- Select the data set called Secondary Amines using the tab in the bottom left of the StarDrop table view.

This data set contains 151 secondary amine structures and their associated meta-data, which were retrieved directly from eMolecules. To learn more about querying and retrieving information on eMolecules compounds directly from StarDrop, please visit:

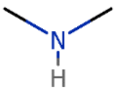
https://www.optibrium.com/community/videos/introduction-to-stardrop-modules-and-features/357-stardropemolecules

The first step in creating the virtual library is to clip the secondary amine reagents into R-group fragments that we can use in the enumeration.

- Open the R-Group Clipper dialogue by selecting **R-Groups** from the **Tools** menu and choosing **Clipping**.

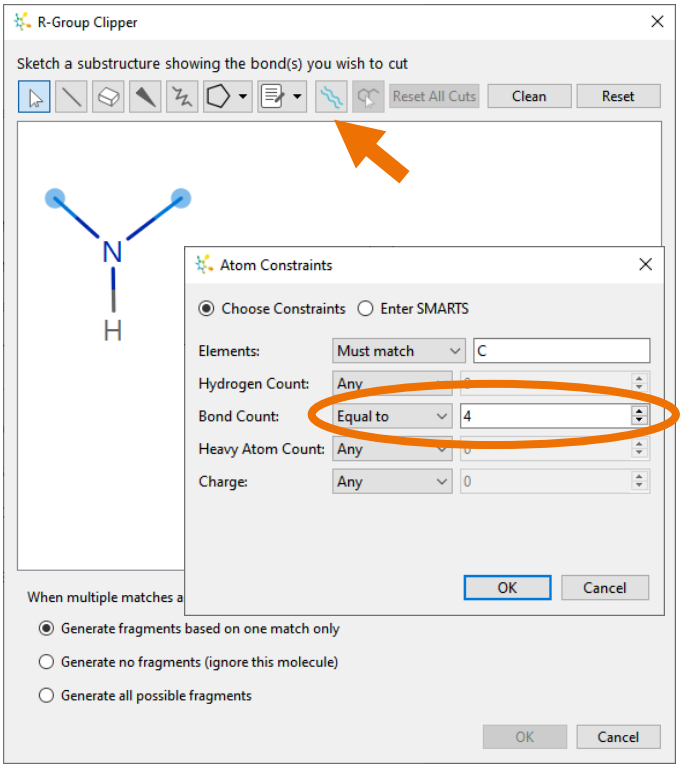


In the R-Group Clipper, we can sketch a substructure that defines how compounds in the data set should be clipped. In this case, we will sketch the secondary amine and impose some bond and atom constraints to limit the fragment to only cyclic, aliphatic, and secondary amines.

- In the sketch area, use the **Bond** tool  to sketch a simple dimethyl amine. 

Hint: To specify an element, hover over an atom and type the element symbol, in this case, “N” and “H”.

- To add atom constraints to the two carbon atoms, select them both by pressing the **CTRL** key while using the **Selection** tool  .



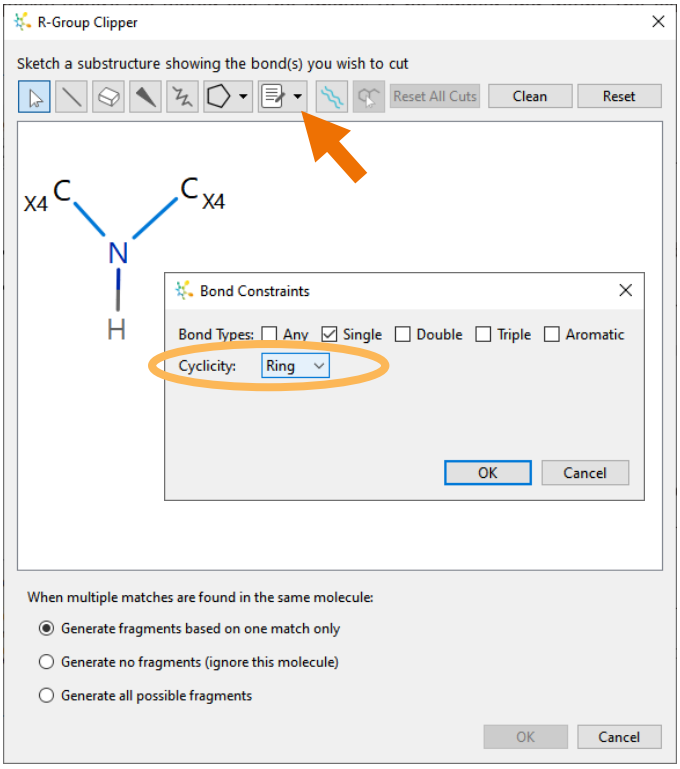
- Click on the **Constraints** menu and choose **Edit Atom Constraints** to display the **Atom Constraints** dialogue.
- Specify that each carbon atom's bond count should be **Equal to 4** and click the **OK** button.

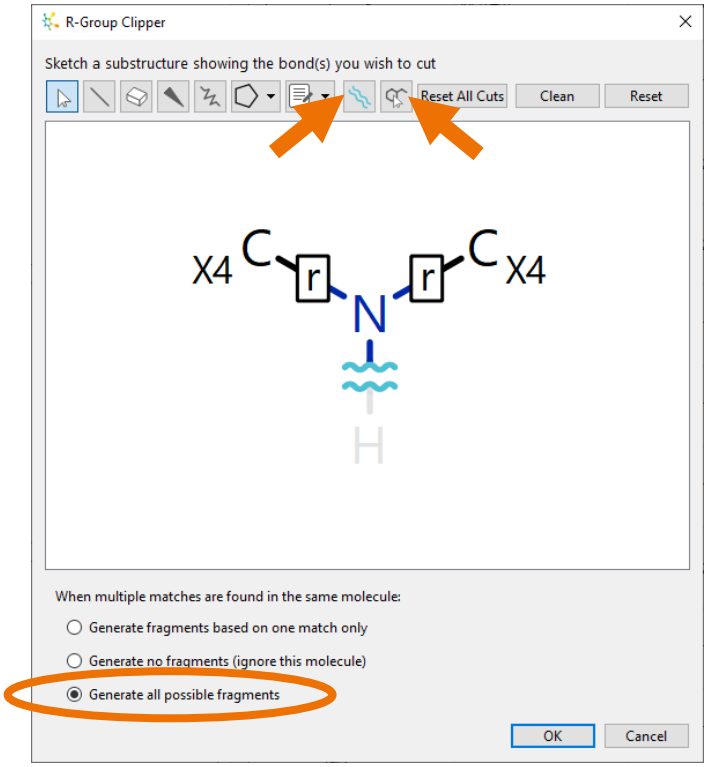
Selecting the Bond count to equal four ensures the carbon is sp^3 hybridised rather than sp^2.

- To add bond constraints, select the two N-C bonds by pressing the **CTRL** key while using the **Selection** tool.

- Choose **Edit Bond Constraints** from the **Constraints** menu to display the **Bond Constraints** dialogue.
- Select **Ring** from the **Cyclicity options** to specify that these bonds must be single bonds that are part of a ring.
- Click the **OK** button.

We next need to specify where we would like the amines to be clipped and define the excluded fragment.





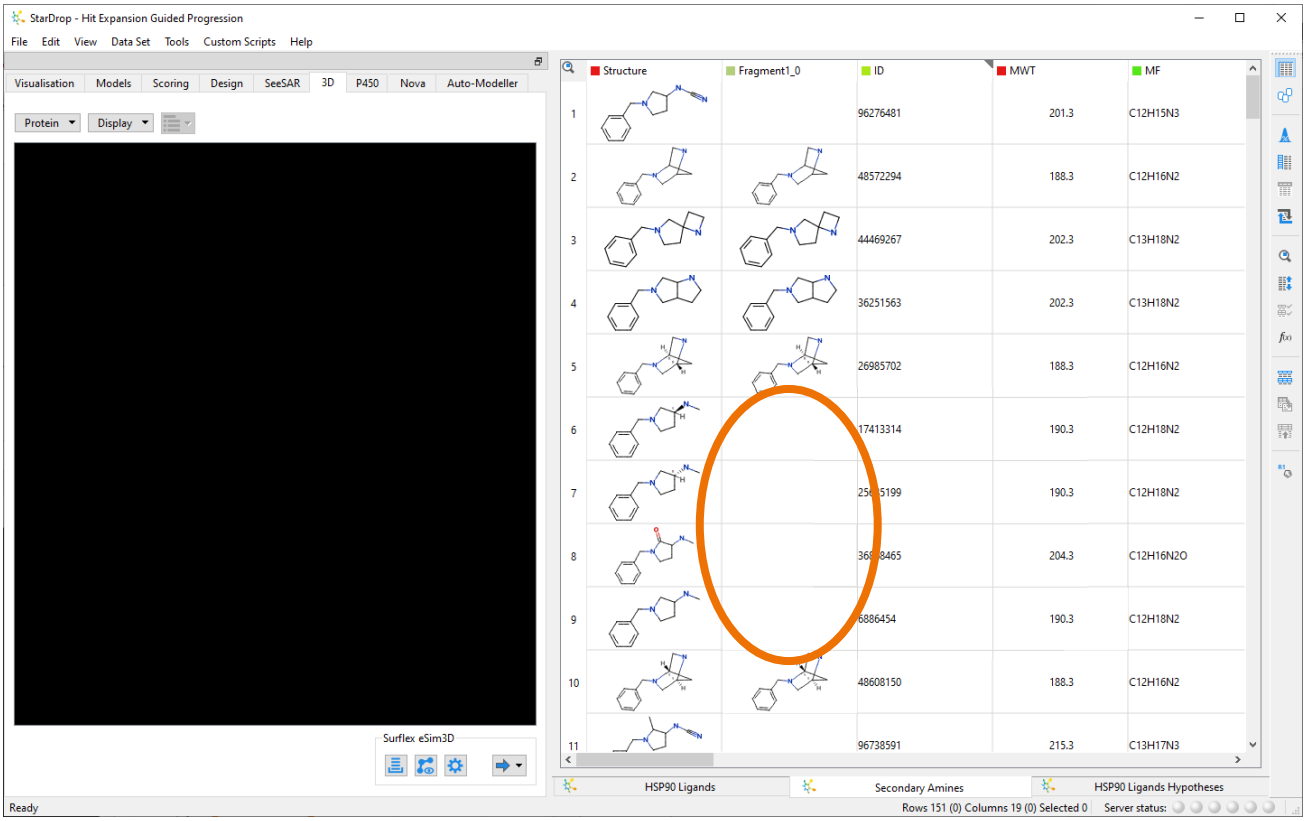
- Select the **Cut** button  and click on the N-H bond to clip this bond.
- Select the **Choose** button  and then click on the Hydrogen to exclude it from the generated fragments.

When comparing molecules in the data set with the specified substructure, multiple matches might be found within the same molecule. At the bottom of the **R-Group Clipper** dialogue, you can specify what should happen when this occurs.

- Select the option to **Generate all possible fragments** and click the **OK** button**.**

As shown in the screenshot below, the fragments will be generated in a new column called Fragment1_0, with an asterisk * indicating the attachment point.

Note: Some rows will not contain a fragment due to the specified exclusion criteria. Examples are highlighted in the screenshot below.

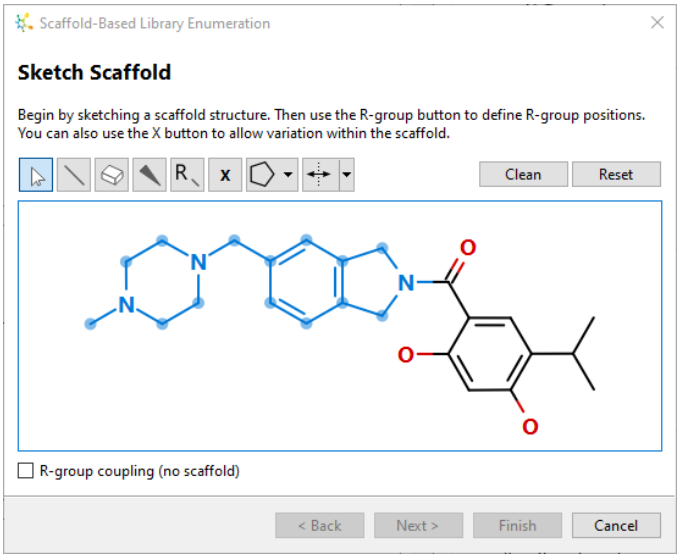


Using this set of fragments, we can now enumerate an amide library using a resorcylic acid scaffold derived from Onalespib. The structure of Onalespib is available in the **HSP90 Ligands** data set.

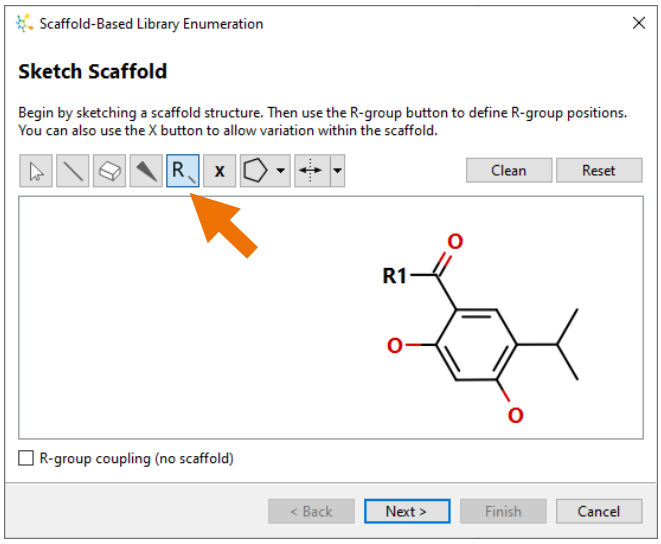
- Click on the **HSP90 Ligands** data set tab.
- Click on the row with Onalespib.

- Click on the **Nova** tab.
- Click the arrow button  at the bottom of the Nova area to start the enumeration.
- In the wizard that appears, select **Scaffold-Based Library Enumeration** and click the **Next** button.

The **Sketch Scaffold** page will be shown containing Onalespib. If desired, we could sketch a new scaffold by clicking the **Reset** button, but in this case, we'll edit the displayed compound to create the scaffold for our new library.



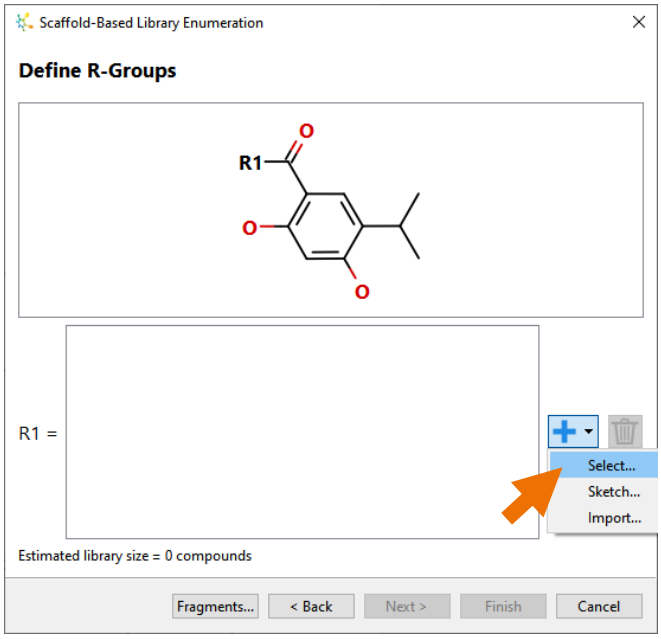
- Use the **Select** tool  to lasso the amine portion of the molecule.
- Click the **Delete** key to delete the highlighted atoms.
- Use the **R-group** tool  to add an R-group by clicking on the carbonyl atom to which it should be connected.



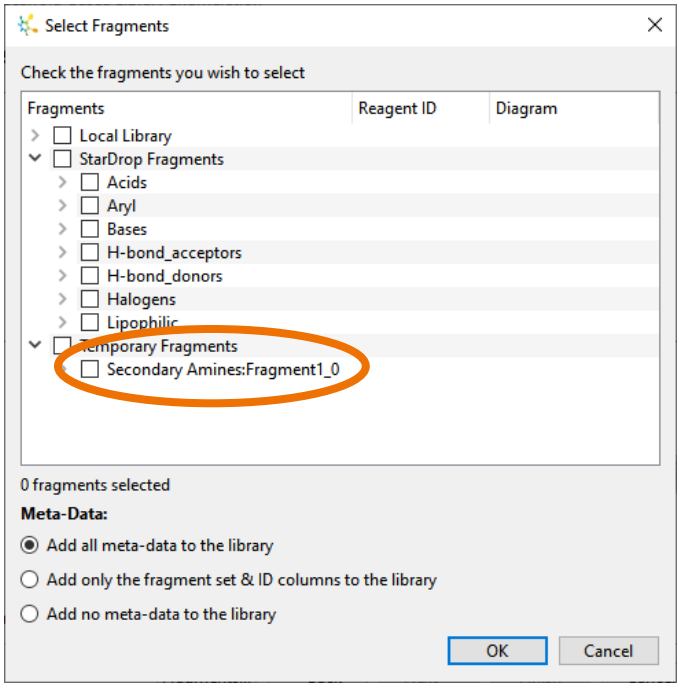
- Click the **Next** button.

The **Define R-groups** page is displayed. Here we will define the list of secondary amine fragments to use in the enumeration.

- Click the **Add** button  next to **R1** and choose **Select** to open the library of predefined substituent groups.

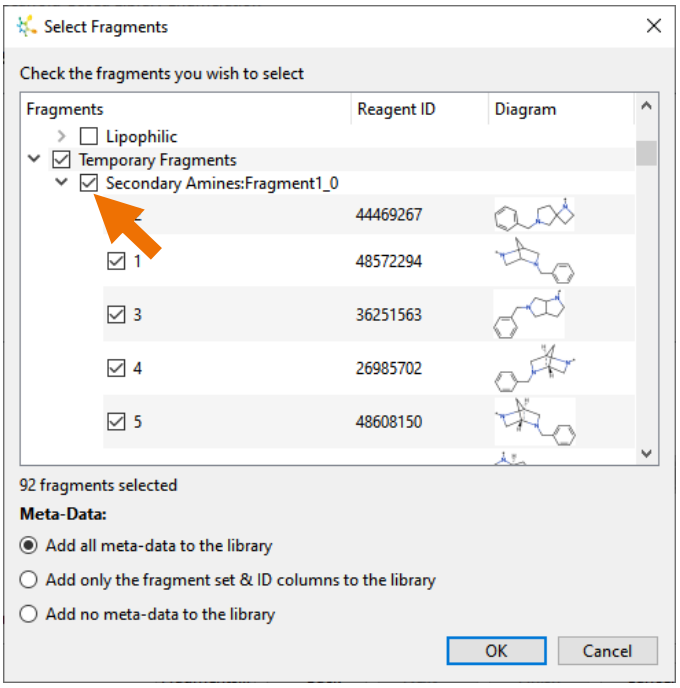


In the fragment library, you will see all the fragments that have been previously saved. The fragments derived from the R-group clipping of the amine library are available at the bottom of the list. They are listed as “Temporary Fragments” because they are from one of the project data sets and have not explicitly been added to the library for future use in other StarDrop projects.



Hint: To permanently add a set of fragments to the library, right-click on the fragment column header in the data set and choose **Add Data Set to Fragment Library** from the menu.

- Tick the box next to **Secondary Amines:Fragment1_0** to select these fragments.



The **Meta-Data** options enable you to specify what data from the fragment library are added to the new series data set.

- Select the **Add all meta-data to the library** option and click the **OK** button.

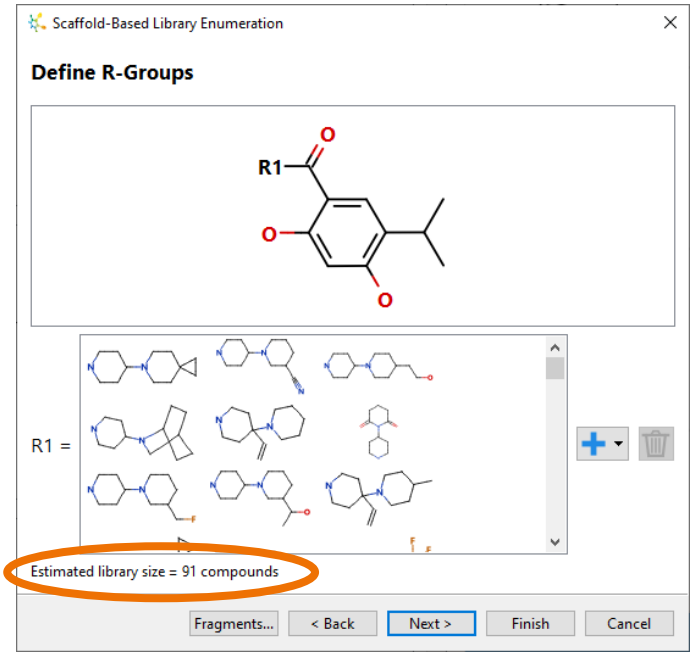
Note that with this selection, the columns of information imported from eMolecules will be added to the enumerated library, making it easy to see which reagents are required for each virtual compound.

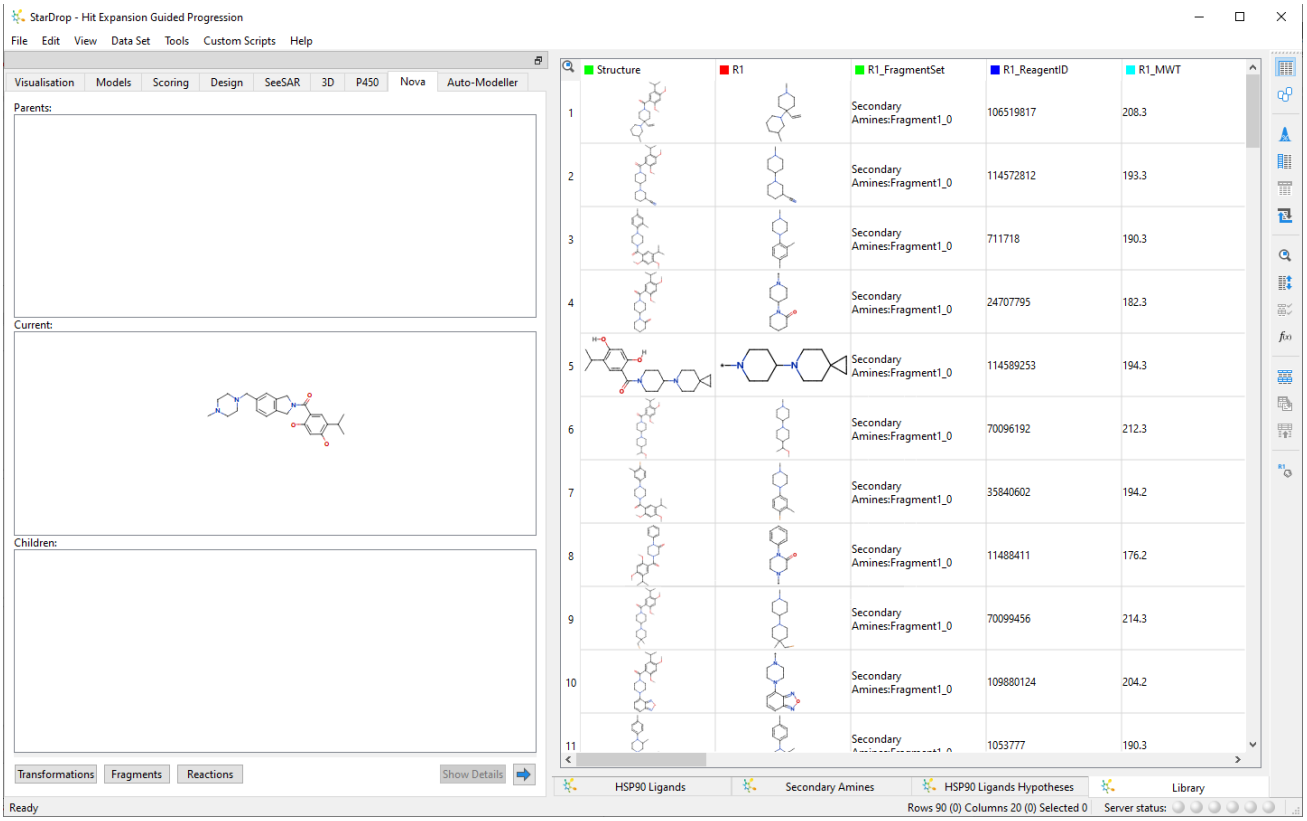
The fragments selected will be shown next to R1. If we wish to add more fragments, we can do so by clicking the **Add** button  again, but in this case, we will only use the fragments we already have.

The estimated library size is 91 compounds.

- Click the **Finish** button.

A new data set will be added to the project called **Library**. It contains 90 structures along with all the reagent meta-data from eMolecules.

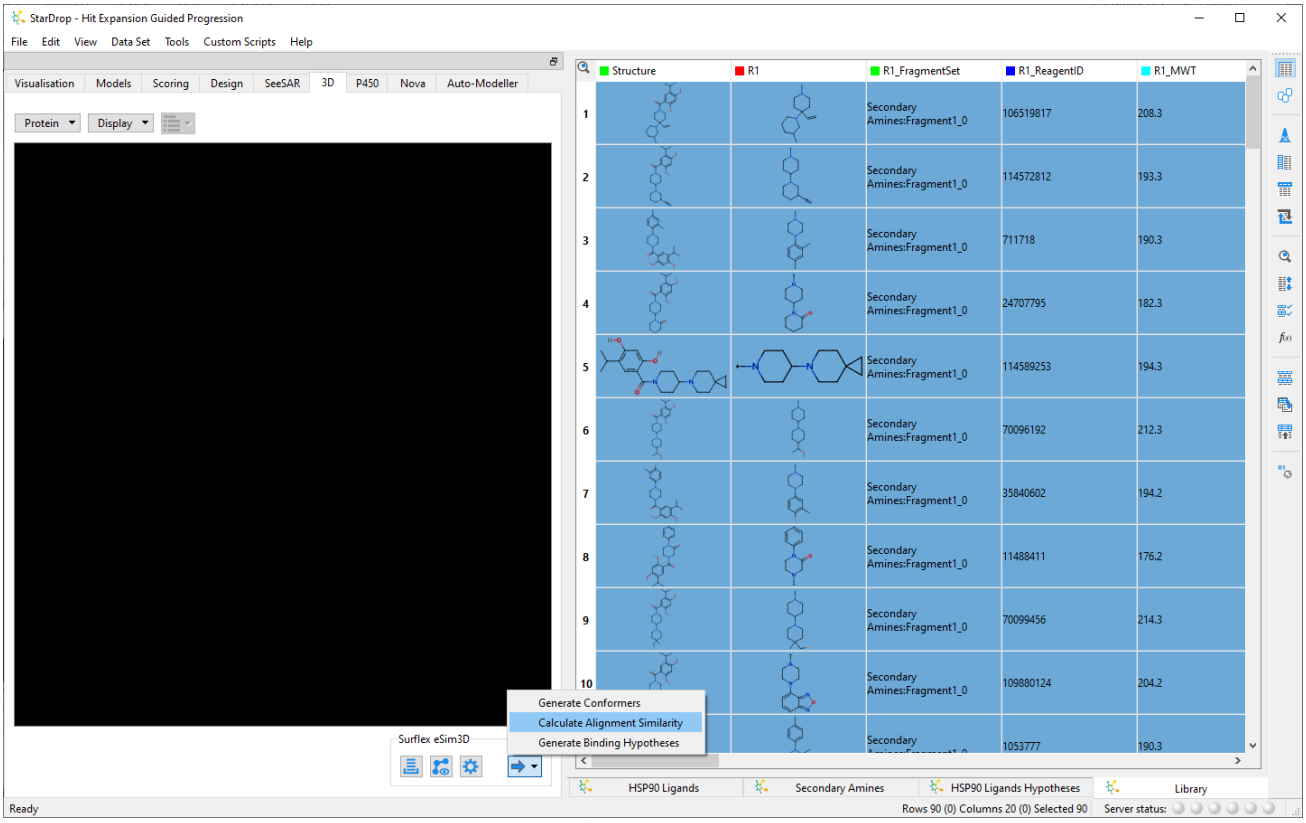




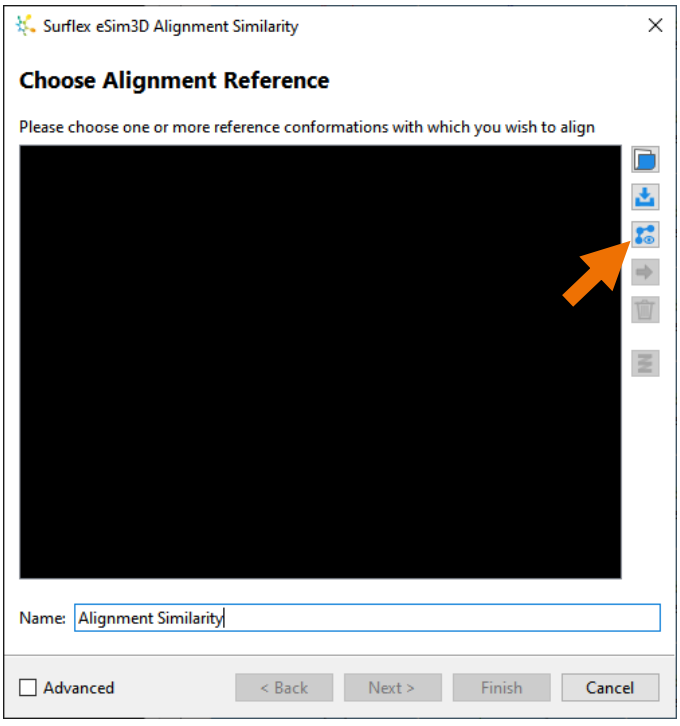
Suppose we want to evaluate the new analogues of Onalespib. In that case, we can use the eSim3D module to compare the alignment similarity of each compound against the Binding Hypothesis we previously generated.

- With the new Library data set open, select the **3D** tab in the left-hand window.
- Select all the compounds in the library set by clicking in the top left corner of the data set.

- Click the **Go button menu** and select **Calculate Alignment Similarity**.

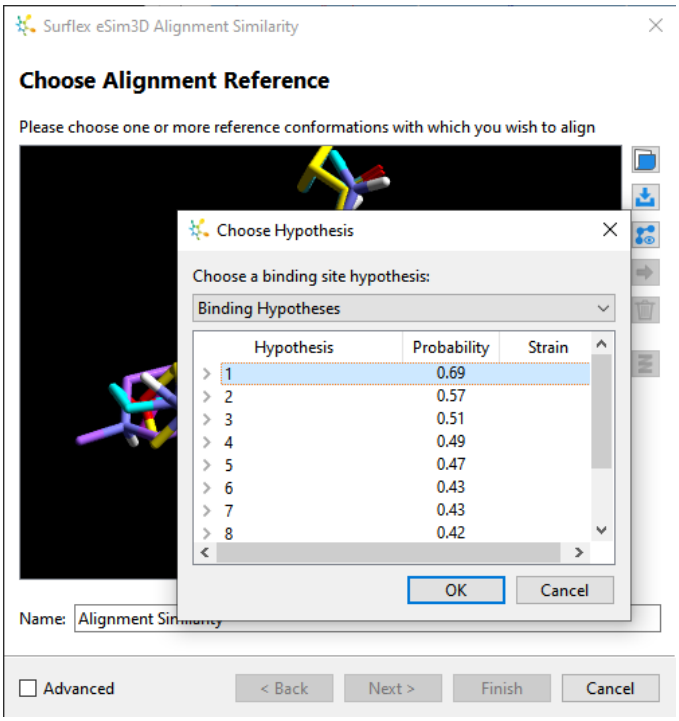


- Click the **Binding Hypothesis** button to select an alignment reference from the binding hypotheses.



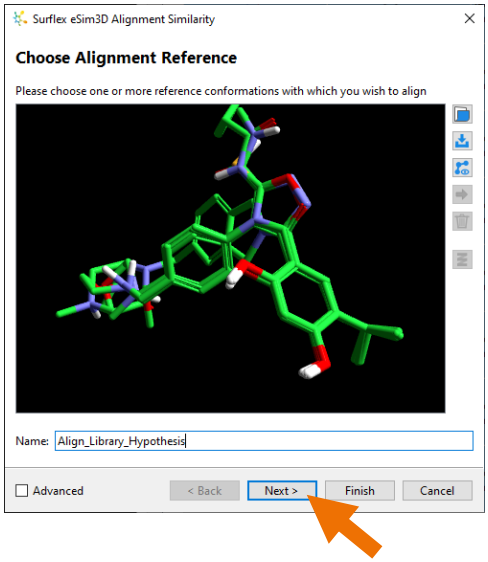
This will prompt you to choose a binding hypothesis alignment reference against which to align your compounds. We will use the hypothesis with the highest probability for this exercise.

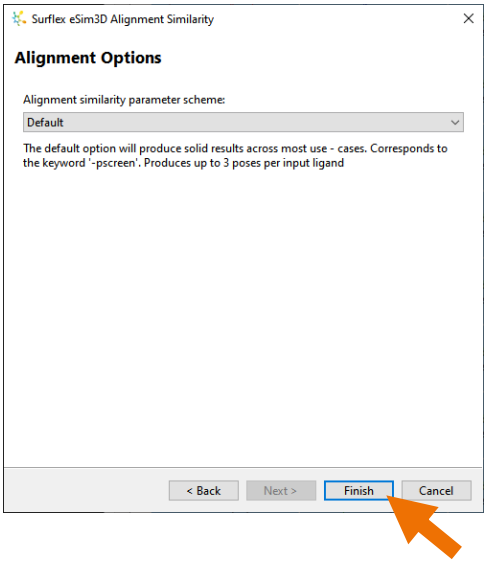
- Select Hypothesis 1.
- Click **OK**.



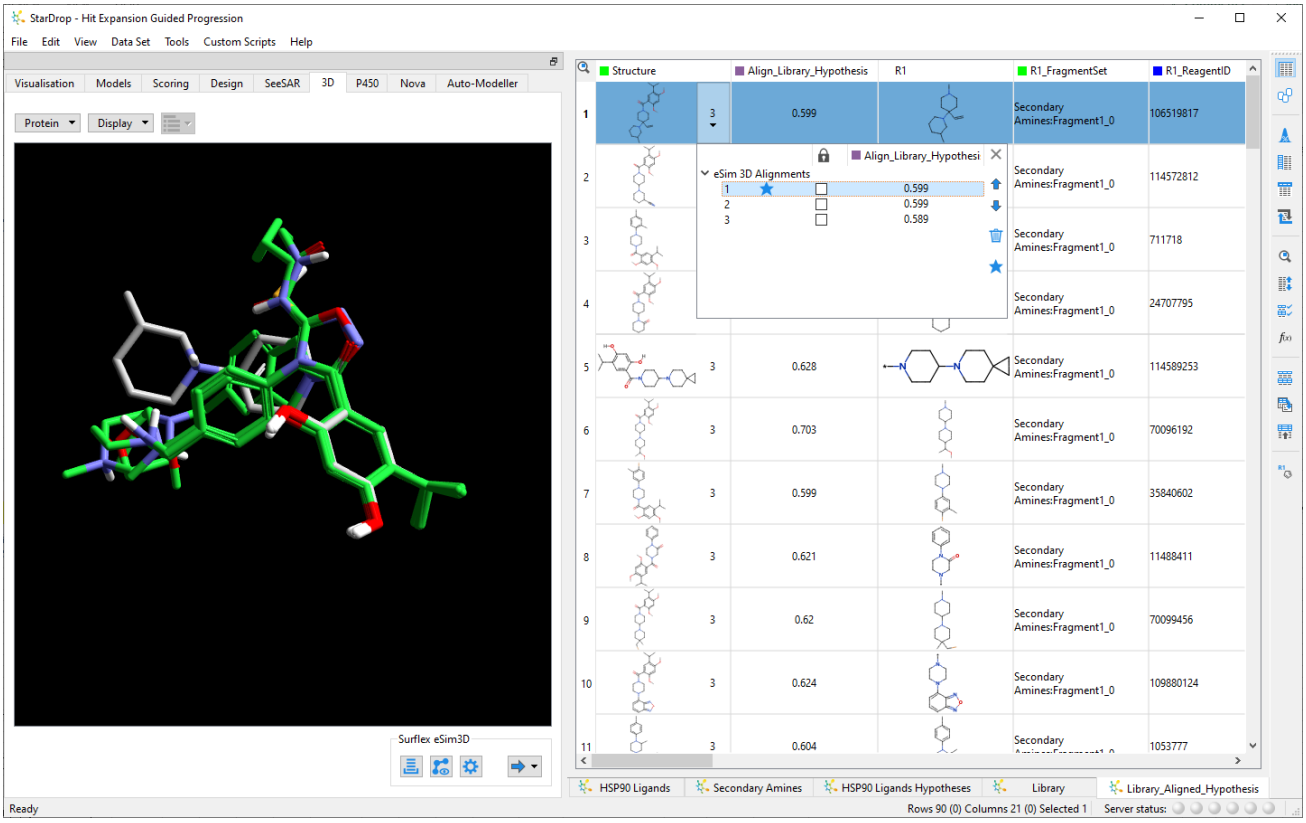
In the Choose Alignment Reference window:

- Rename the analysis **Align_Library_Hypothesis**.
- Click **Next**.
- If working through this exercise independently, click the **Finish** button to begin running the alignment with the default option. If you do not wish to wait for the results, you can click **Cancel** and switch to a data set with pre-calculated results in the **Align_Library_Hypothesis** tab.

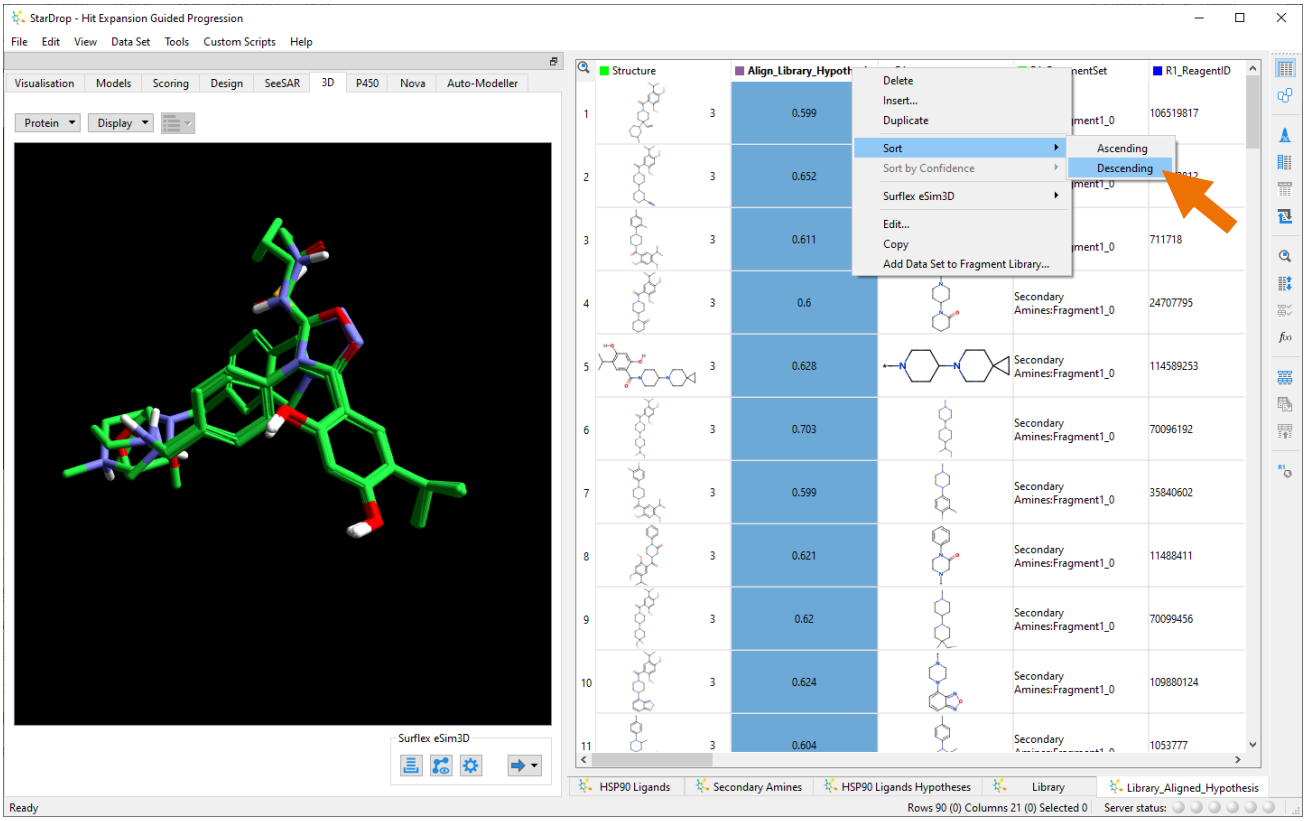




When the alignment is complete, the calculated similarity value for each molecule will be added to our data set in the new **Align_Library_Hypothesis** column. A number will also be added to the **Structure** column, which is the number of alignments that have been generated. We can view alignments by clicking the number next to the structure and selecting from the drop-down menu. The best scoring conformation is the primary pose denoted by the ★ symbol.



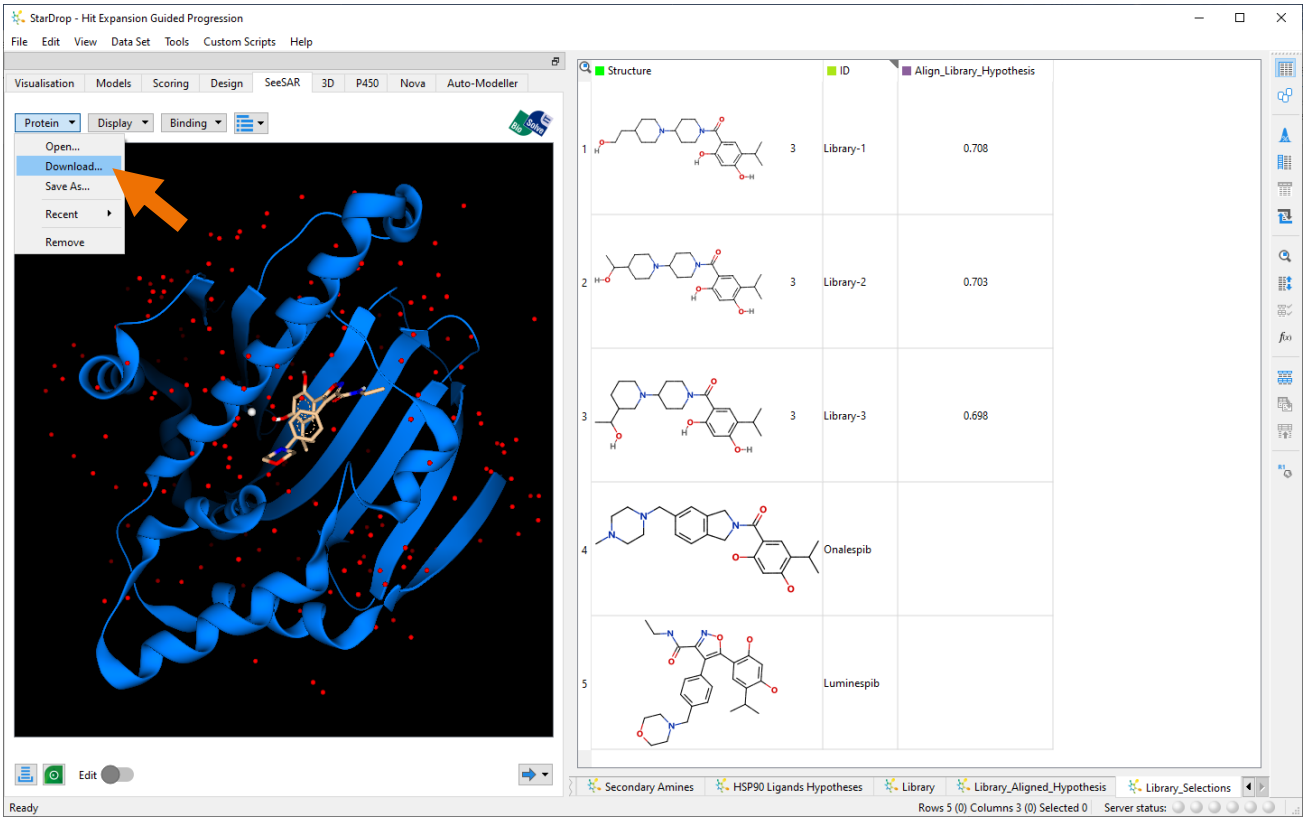
- Once the alignment is complete, right-click on the **Align_Library_Hypothesis** column and select **Sort** and then **Descending**.



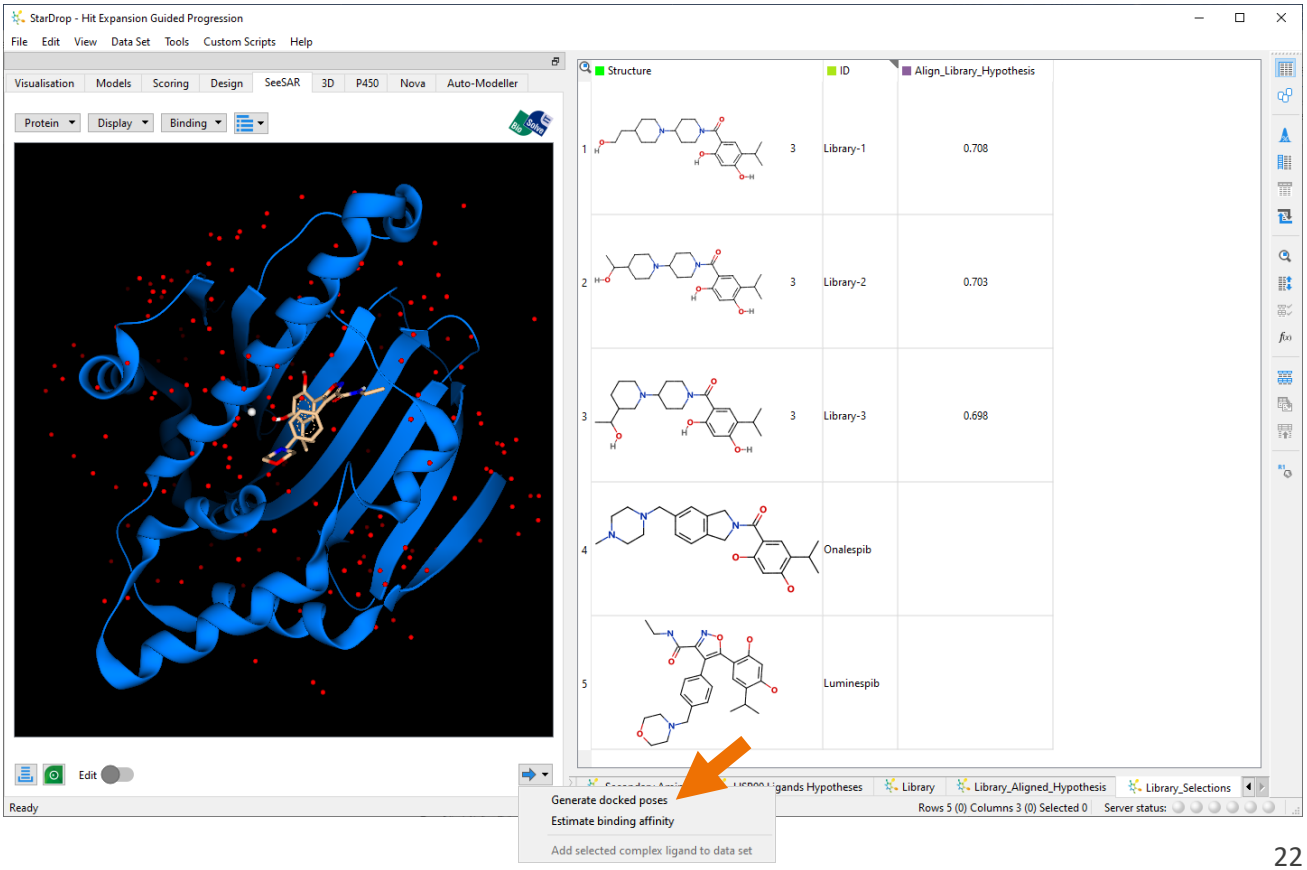
We will now look at the three top-scoring compounds with an alignment score close to or equal to 0.70. For comparison, each of these compounds, as well as Luminespib and Onalespib, will be docked against the target HSP90. For convenience, the three library compounds and the two references are already provided for you in the **Library_Selections** data set. The StarDrop SeeSAR module allows viewing of the 3D structure-based design docking results, generation of compound poses for virtual screening, interactive 3D design, and identification of critical interactions driving binding affinity. We will use the SeeSAR module to perform docking experiments of the new ideas against the target. Our initial 3D alignment of the library-designed compounds to the hypothesis was a ligand-based approach. While this was a reasonable first approximation and relatively fast, the alignment to the hypothesis does not contain information about any “excluded volumes” within the binding site, which can be identified by performing docking experiments.

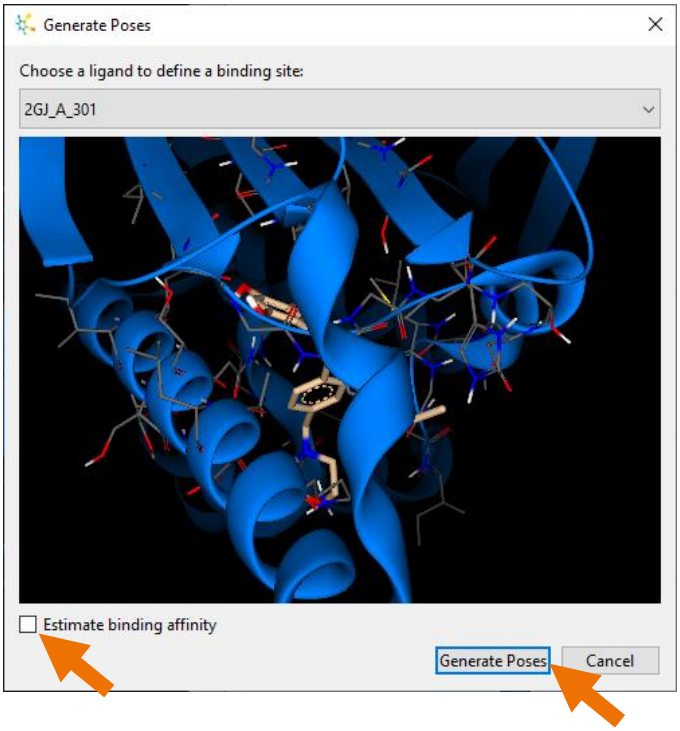
We will now evaluate how well our top-scoring compounds from the library generation bind the active site.

- Change to the **Library _Selections** data set.
- Click on the **SeeSAR** tab.
- In the **SeeSAR** area, select the **Protein** menu button and select **Download**.
- Enter the PDB code **6LTI**.
- Click **OK**.



- From the  menu at the bottom of the SeeSAR area, choose **Generate docked poses**. **Note:** By default, you will see that the co-crystallised ligand, Luminespib, is already placed in the binding site.

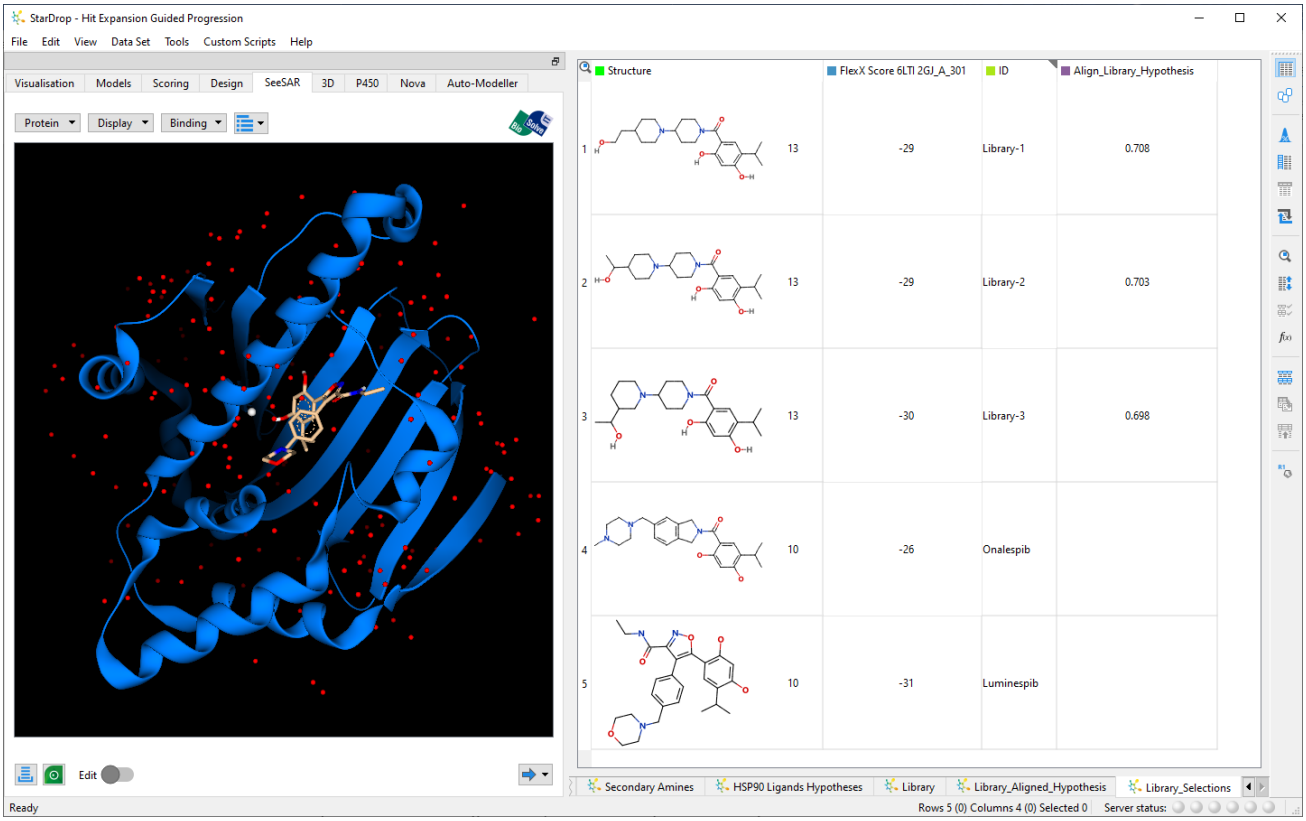




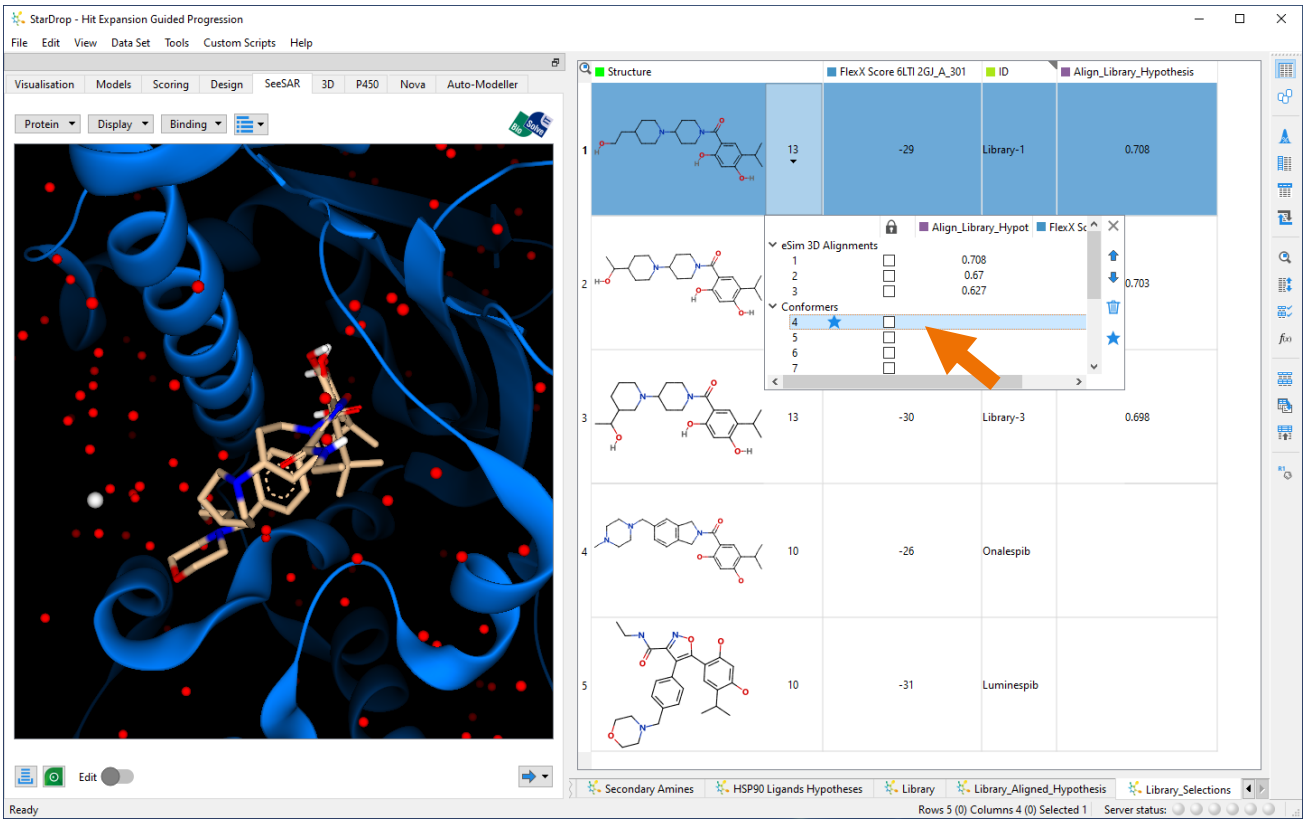
When generating poses, we need to define a binding site. By default, you will see that the co-crystallised ligand, Luminespib (2GJ_A_301), is already selected, and we will use this (**Note**: if we had already docked other ligands, then we could use one of these to define the binding site instead).

- First, uncheck the option to **Estimate binding affinity**. **Note**: If you have the SeeSAR Affinity module, then the Estimate binding affinity option can remain checked, but we will not use it in this example.
- Then click the **Generate Poses** button to start the process.

You will see that a new column is added to the data set in which the FlexX Score for each compound will be displayed once the poses have been generated. While the calculations take place, the compounds will be listed as either Running (the calculation is taking place) or Queued (the calculation will start when this compound reaches the front of the queue).

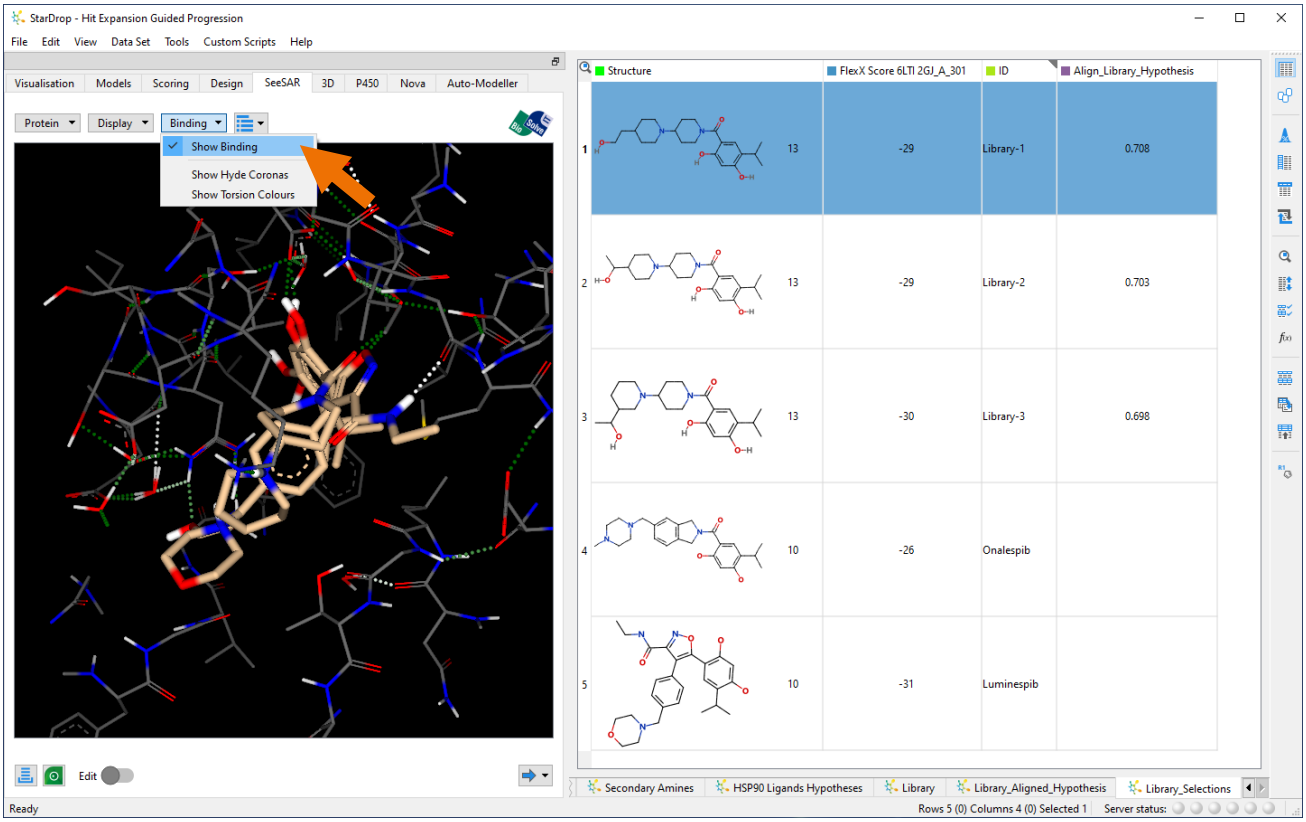


- To view some of the poses that have been generated, select the compound in **row 1,** which docks with a score higher than Onalespib.

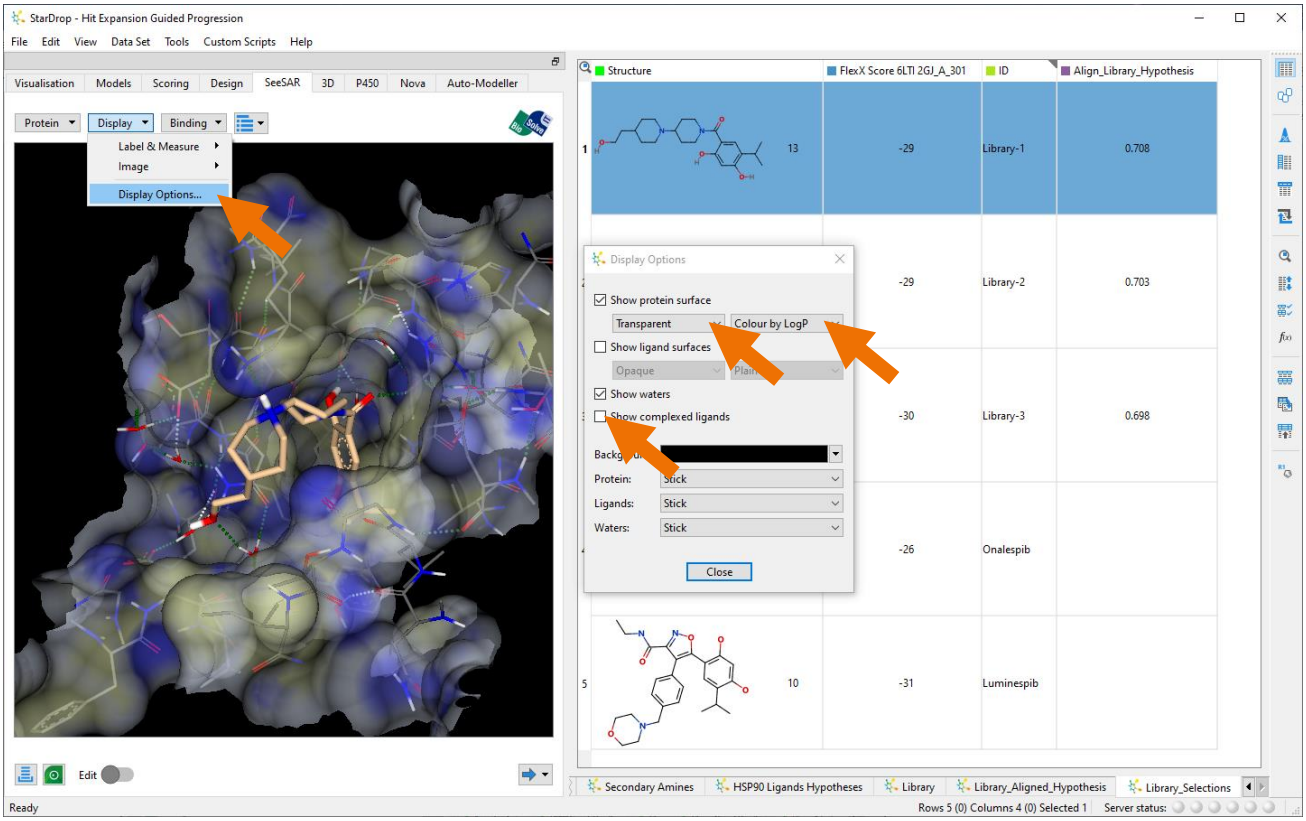


The binding of this new analogue is shown to overlap with the co-crystallised ligand. The view can be manipulated using your mouse to change the image's perspective.

- The left mouse button enables you to rotate the image.
- The right mouse button enables you to pan in any direction.
- The scroll wheel allows you to zoom in and out of the image.
- To view the binding site in more detail, in the SeeSAR area, select the **Binding** pull-down menu and then select **Show Binding**.



- From the **Display** menu, choose **Display Options**.
- In the Display Options dialogue, tick the Show protein surface option and choose **Transparent** and **Colour by LogP** from the menus.
- In the **Display Options** dialogue, untick the **Show complexed ligands** option to hide the ligand.
- Click **Close** to remove the **Display Options** dialogue.



From the docking experiment, the new compound from the library enumeration does seem to be able to bind the active site in a conformation that partially overlaps the co-crystallised ligand. While not definitive proof of improved binding, the molecule can be considered for synthesis and follow-on testing.

Conclusion and Additional Resources

With this worked example, we have demonstrated several options for using StarDrop's Surflex eSim3D and SeeSAR modules to assess a small virtual library of compounds for their similarity to the binding of known inhibitors to a target of interest. We first used the Surflex eSim3D module to understand the 3D structure-activity relationships (SAR) and create a binding hypothesis. Following the enumeration of a new library based on one of the known inhibitors, we then evaluated which ones best matched the generated binding hypothesis. Ultimately, the best scoring compounds were docked against the target to estimate the redesigned compounds' ability to bind the active site.

If you have any questions, please feel free to contact stardrop-support@optibrium.com.