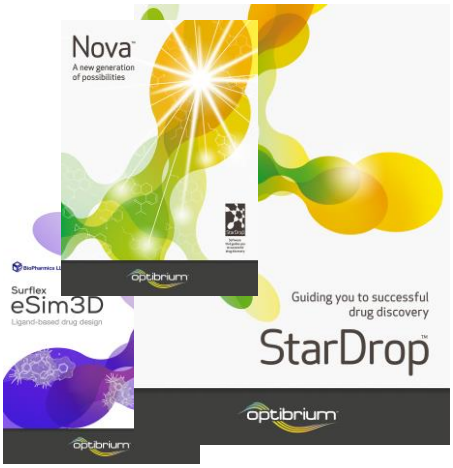


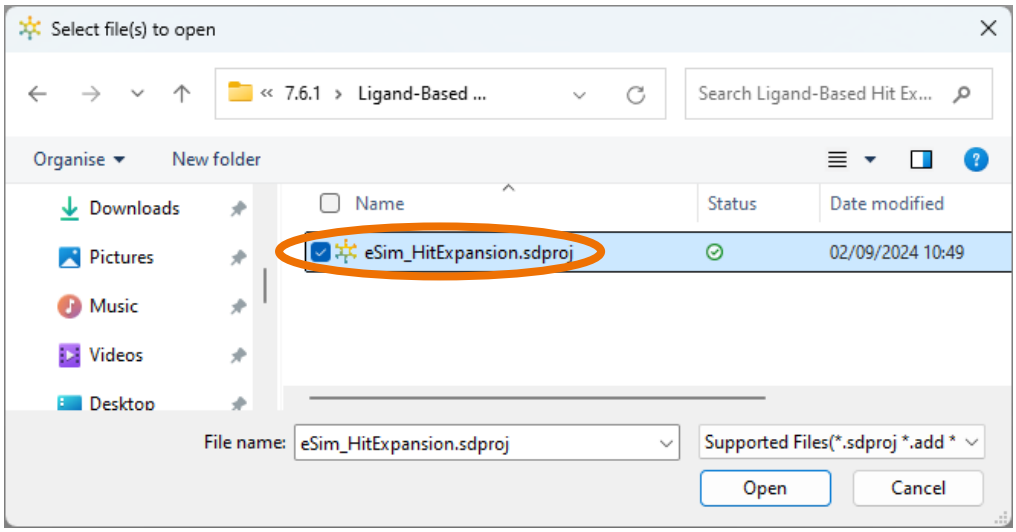

# StarDrop™ Worked Example:

## Ligand-Based Hit Expansion

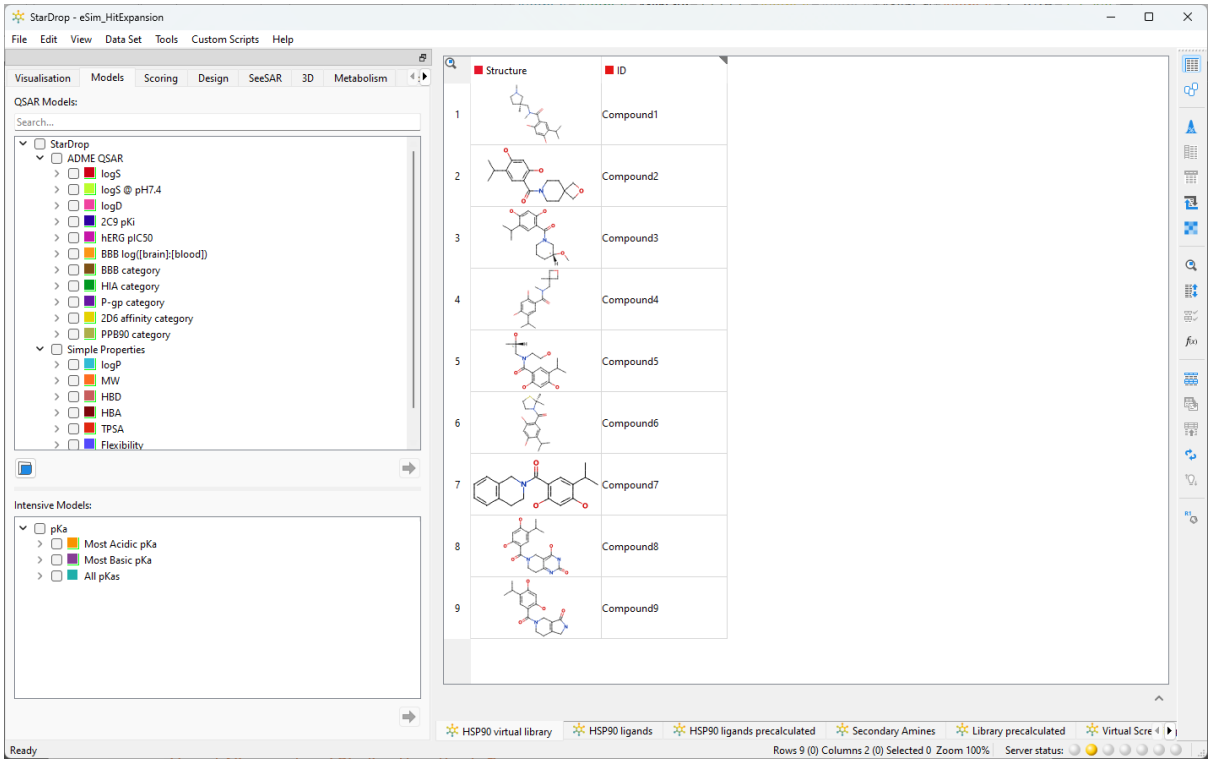
This worked example uses the Surflex eSim3D™ module in StarDrop™ to assess a small library of compounds for their similarity to known Heat Shock Protein 90 (HSP90) ligands. We will design a ligand-based binding hypothesis based on the published crystal structures of five known HSP90 ligands. We will then use StarDrop's Nova™ module to enumerate a library created by a *de novo* design process containing compounds with a beta-resorcylic acid core. We will explore the chemistry space around the lead compound(s) and show that we can move to new regions of the chemical space using virtual screening against a vendor library. 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 support@optibrium.com.

## Exercise

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



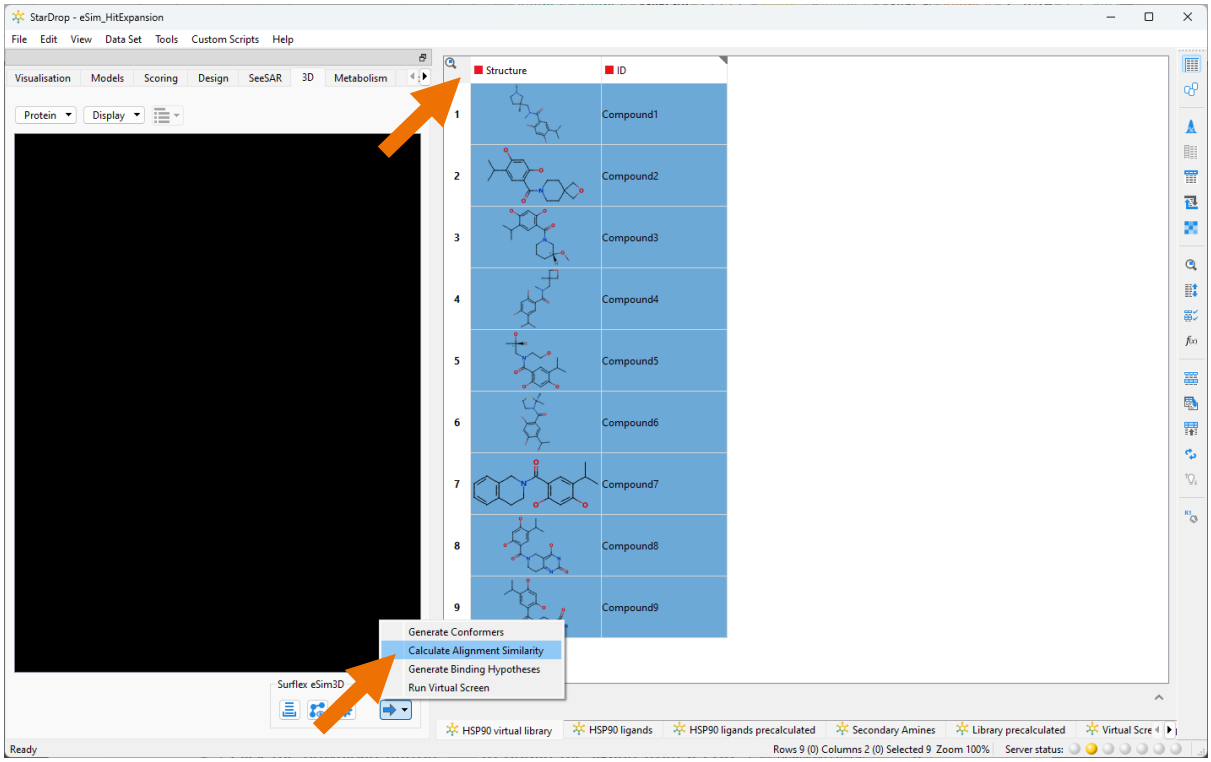
This opens a StarDrop project containing a number of data sets that we are going to explore during this example. The first data set we can see contains a set of nine virtual compounds based around a known HSP90 inhibitor, Onalespib, and share a beta resorcylic acid core.




## Ligand Alignment and Binding Hypothesis

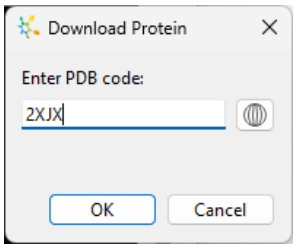
To investigate how similar these compounds are in terms of their shape, electrostatic potential and hydrogen bonding potential to Onalespib, we will perform a similarity analysis.

- Change to the **3D** area by selecting the associated tab.
- Select all compounds of the HSP90 virtual library data set by clicking in the top left corner of the data set.
- Open the Surflex eSim3D menu at the bottom by clicking on the blue arrow  and select **Calculate Alignment Similarity.**




This will prompt you to choose an alignment reference molecule against which to align your compounds. We will use the HSP90 inhibitor Onalespib as a reference, and we will obtain the bioactive conformation of Onalespib from a co-crystal structure.

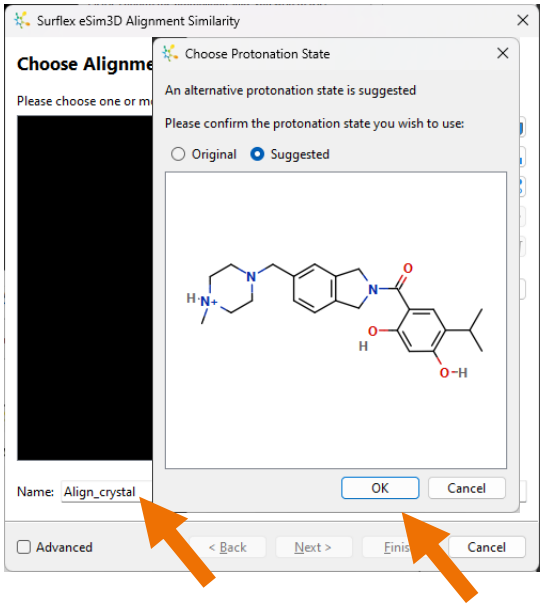
- Click the download button  to obtain the ligand from a PDB reference.
- Enter the PDB code '2XJX' and click **OK**.



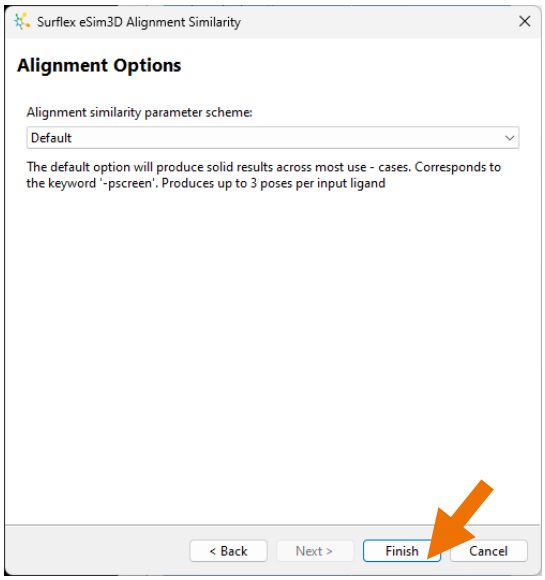

A dialog box appears, suggesting an alternative protonation state for the nitrogen on the piperazine ring.

- 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 (see Section 22.5 of the StarDrop User Guide for more details). However, we will use the defaults.



- Click **Next**.



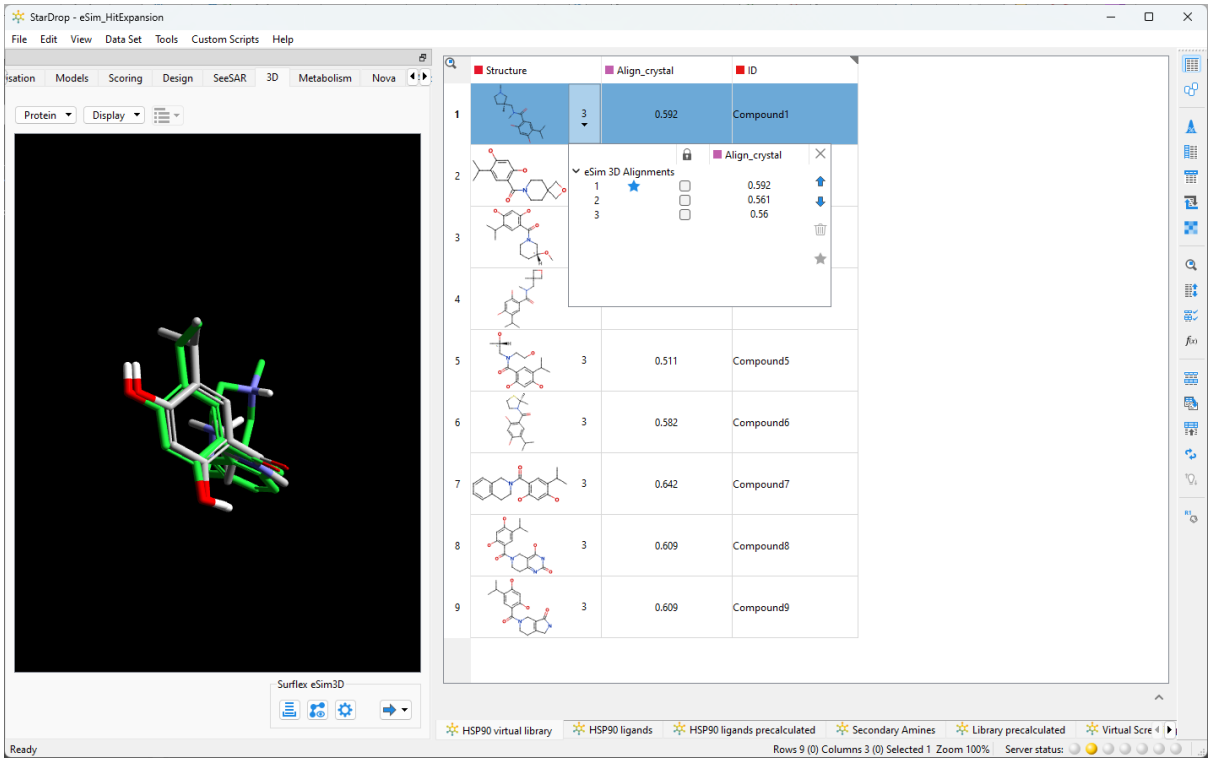

This page gives the option to define the parameter scheme of the alignment similarity analysis. The parameter schemes vary in terms of speed and accuracy.

- Click **Finish** to run the alignment with the default option.

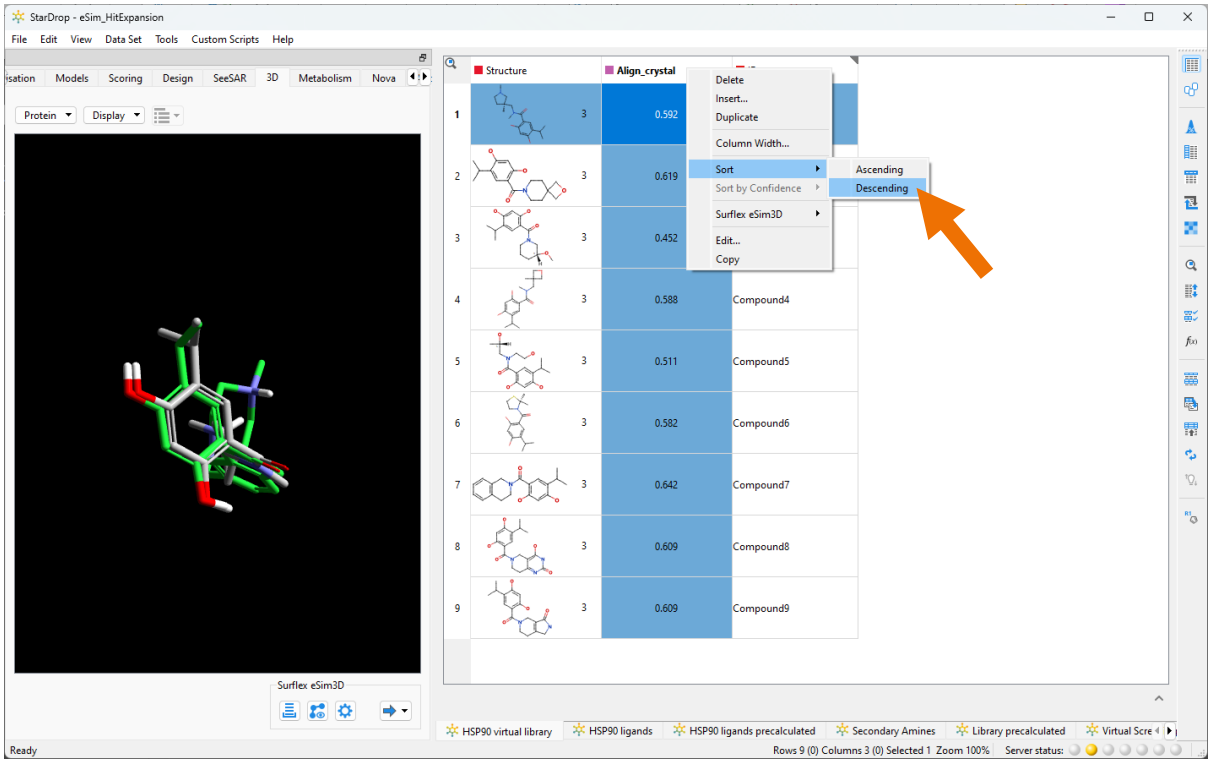
Once the alignment finishes, the calculated similarity value is added to our data set in the new **Align_crystal** column. A number is added to the **Structure** column, which shows the number of alignments it has generated. We can view the alignments by clicking the number next to the structure. The best scoring conformation is the primary pose denoted by the ★ symbol.

- Select Compound1 to show it in the 3D viewer aligned to Onalespib.
- Click 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 the header of 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.




We can see that the most similar compound to the co-crystallised ligand is Compound7.

- Select Compound7 in row 1 to see the alignment in the 3D viewer.

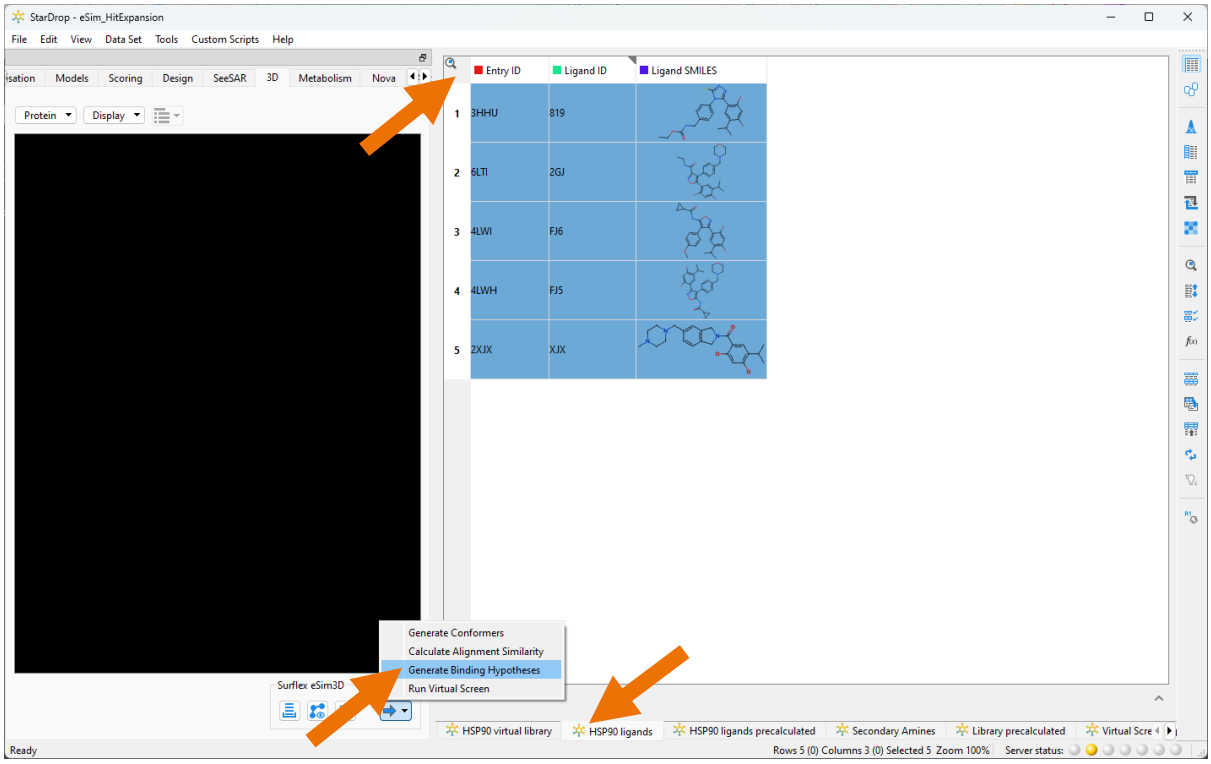
As you rotate the 3D display with your mouse you can see that there is good alignment with the beta resorcylic acid and isoindoline regions of our reference compound, Onalespib. However, Compound7 is much smaller and doesn't have a portion that aligns with the piperazine ring. Therefore, despite its alignment, the potency might be much lower.

Just as designing against one property could lead to bias and missed opportunities, comparing the alignment similarity to only one ligand might introduce bias. There are several ligands that bind to HSP90, and five are included in the 'HSP90 ligands' data set. We shall use all of these ligands to generate a binding hypothesis to gain an understanding of the 3D shape and interactions that can be tolerated by the binding pocket, even in the absence of a crystal structure.

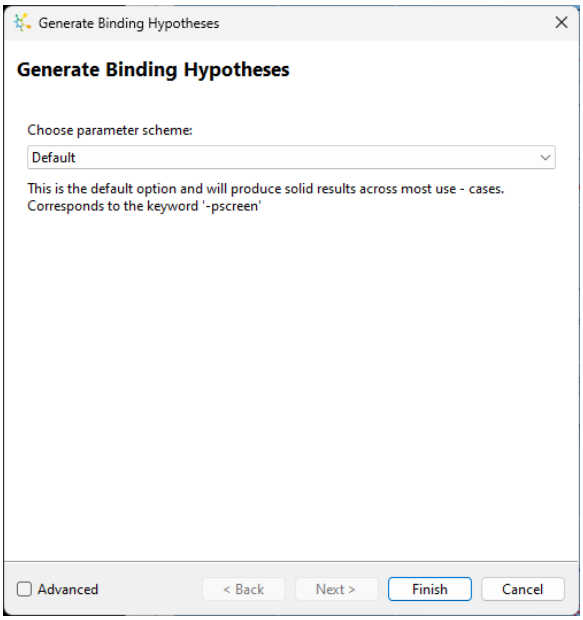
- Select the second data set of the project called **HSP90 ligands** by clicking the tab at the bottom.

This data set shows Onalespib and four other compounds with their HSP90-bound conformation determined. These compounds all contain the beta-resorcylic acid core.

- Select all compounds in this data set by clicking in the top left corner of the data set.
- Open the Surflex eSim3D menu at the bottom by clicking on the blue arrow  and select **Generate Binding Hypotheses**.



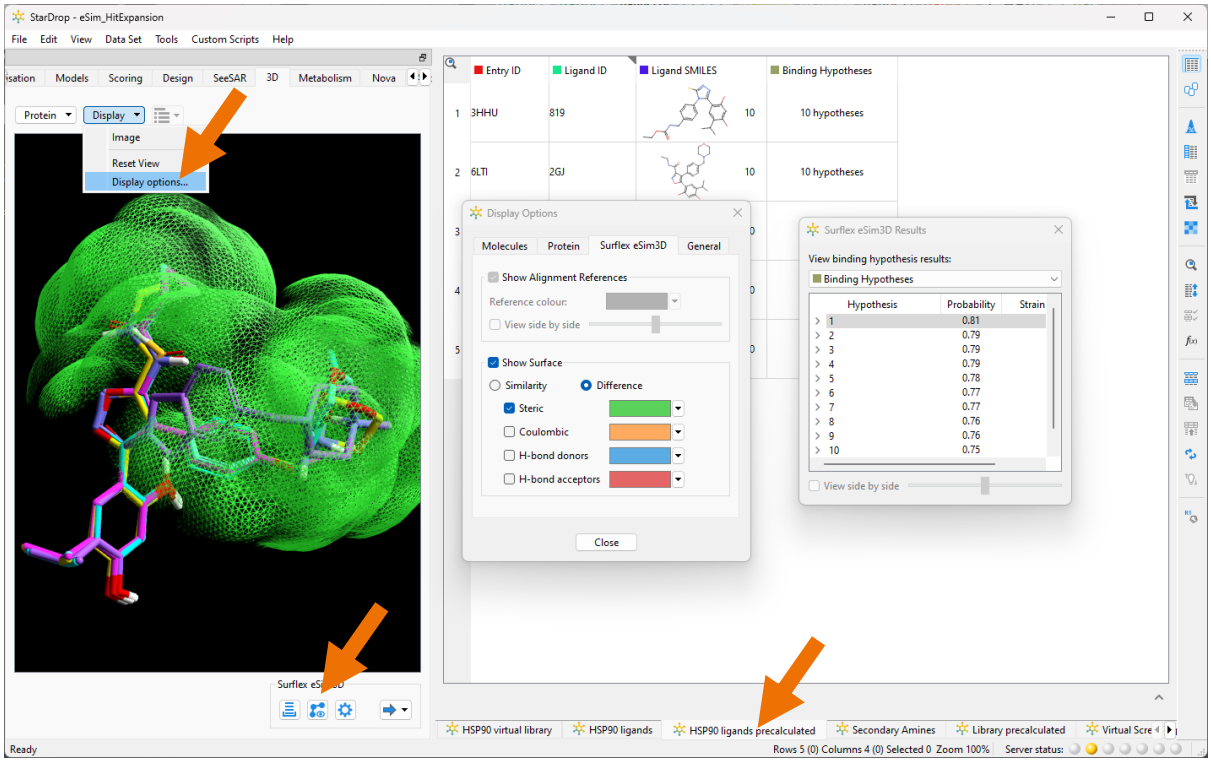
The parameter scheme specifies how intensive a search to conduct for binding hypotheses. Choosing an intensive calculation will lead to more accurate results but will take longer to compute. You can set additional configuration options by selecting **Advanced**, but in this case, we will use the default setting.




- Click **Finish** to start the generation of binding hypotheses.

This will take a few minutes to complete, but if you prefer, you can view the saved results stored in the third data set in this project called **HSP90 ligands precalculated**.

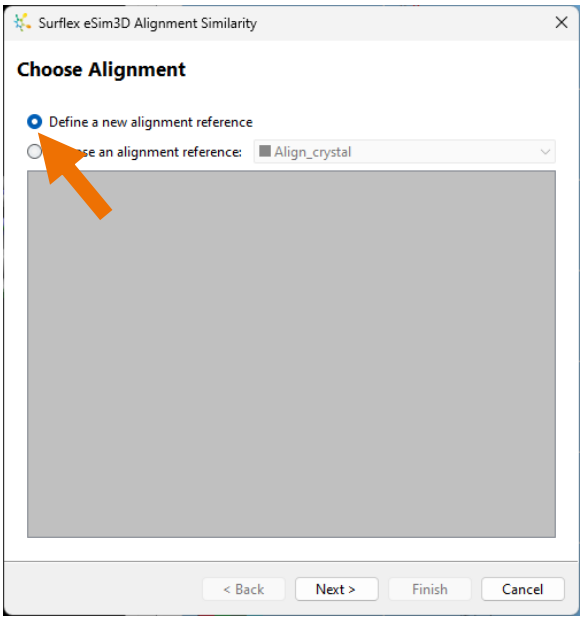
- Click the **Binding hypothesis** button  under the 3D viewer to explore the binding hypotheses. Ten binding hypotheses have been generated, and the probability of each of the hypotheses being accurate is shown.
- With hypothesis 1 selected, choose **Display options** from the **Display** drop-down menu at the top of the 3D viewer (or click the settings button  below the 3D area).
- Select the **Surflex eSim3D** tab in the display options and tick the box to **Show Surface**.
- Select the **Difference** radio button and tick the **Steric** box to show differences in the steric surface of the ligands in the binding hypothesis.
- Turn off the surface representation by un-ticking the **Show Surface** box.
- Close the **Display options** and **Surflex eSim3D Results** windows.

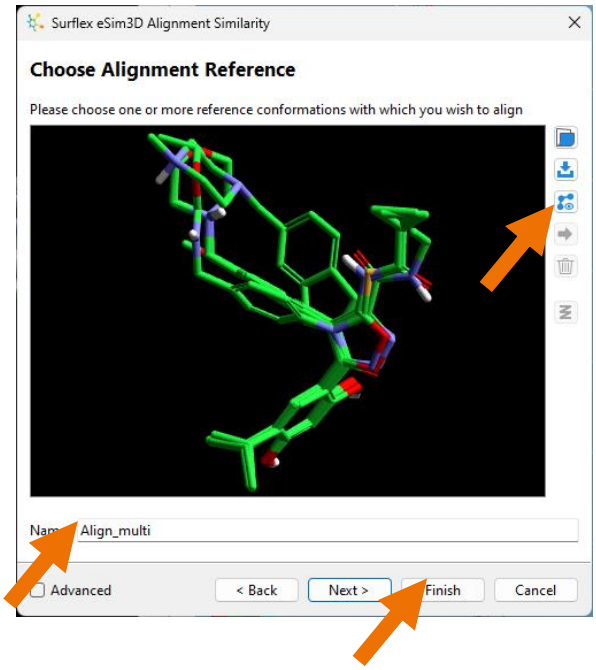



We can see good alignment around the beta-resorcylic acid region, where similarities between the shapes of the ligands indicate likely steric constraints or conserved interactions. Areas of steric dissimilarity, as shown above, indicate regions where there is more freedom within the binding site. We now have a better reference for the virtual compound alignment, as it is made up of several examples of HSP90 ligands. We will now compare our virtual compounds to the highest probability binding hypothesis.

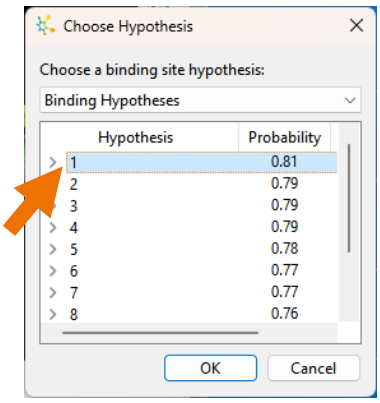
**Note**: The results can be accessed at any time by clicking the **Binding hypothesis** button  under the 3D viewer.

- Re-open the first data set, **HSP90 virtual library**.
- Select all compounds by clicking in the top left of the data set.
- Open the Surflex eSim3D menu at the bottom by clicking on the blue arrow  and select **Calculate Alignment Similarity**.
- Select **Define a new alignment reference** and click **Next.**






- Click the **Binding hypothesis** button to select a binding hypothesis to use as a reference.
- Choose the first binding hypothesis and click **OK**.




- Give the alignment the name 'Align_multi' and then click **Finish**.

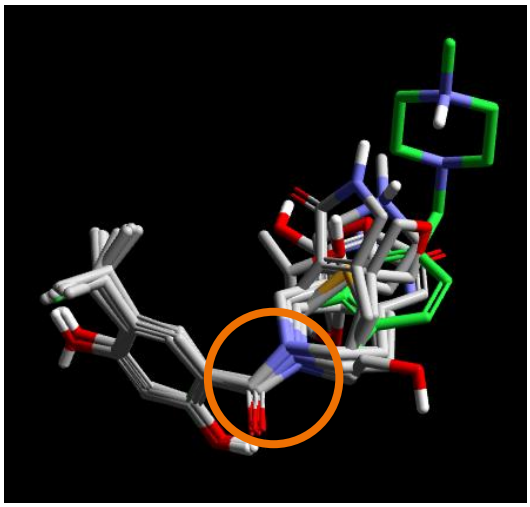
The calculated similarity is again shown in a new column, this time called 'Align_multi'.

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

Notice that the compounds' ranking has changed, and, in many cases, a higher scoring alignment is obtained when aligning to the multi-ligand hypothesis. By using multiple inhibitors to generate our binding hypothesis, we are able to consider a range of conformations and interactions that could be accommodated by the target binding site.

## Library Enumeration

The virtual library we have considered above shows excellent alignment between compounds with the beta resorcylic acid core. This part of the molecule binds deep in the pocket of the HSP90 protein. This core is connected via a secondary amide (circled in the image to the right) to a variety of different chemical groups at the far end of the molecule. We will explore the chemical space around this region more thoroughly by preparing a focused library 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 will evaluate the 3D alignment of every compound against the binding hypothesis we generated above.

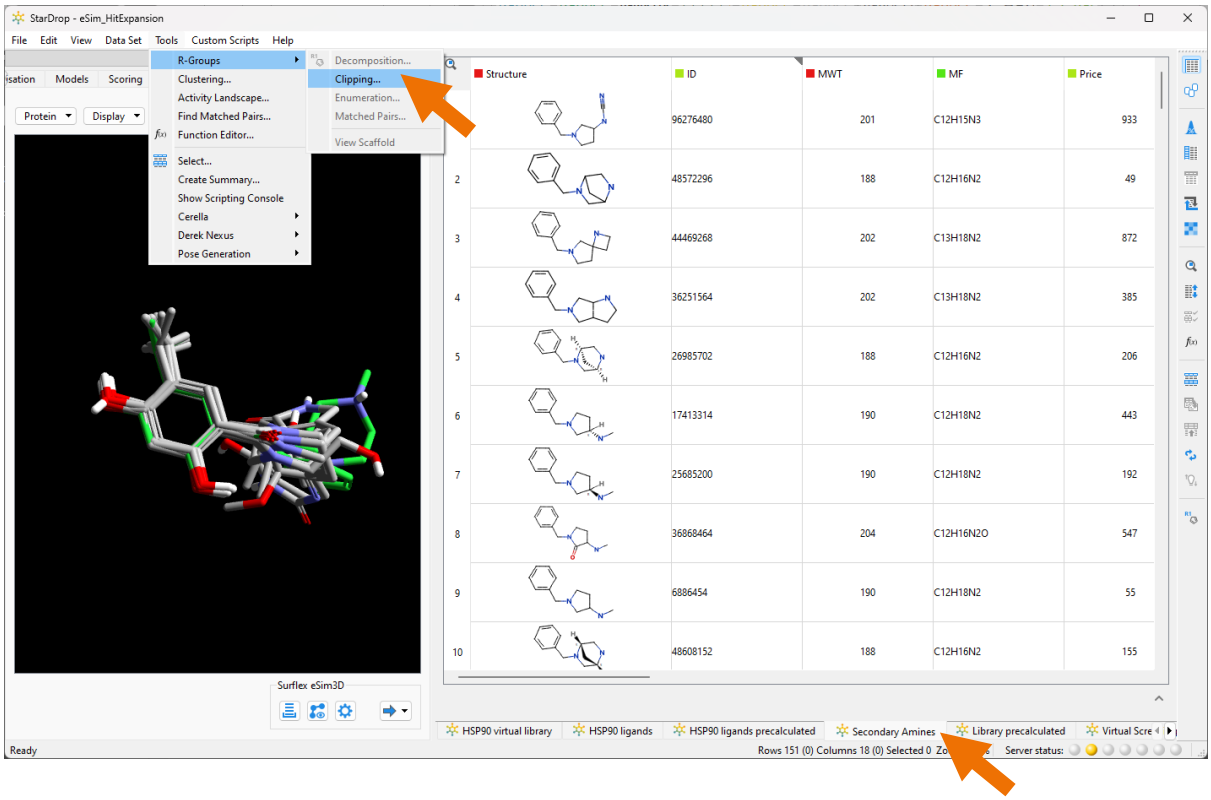
- Select the data set called **Secondary Amines** using the tabs at the bottom.

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:

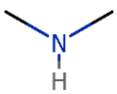
https://optibrium.com/resources/emolecules-plug-in-for-stardrop/

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 dialog by selecting **R-Groups** from the **Tools** menu and choosing **Clipping**.

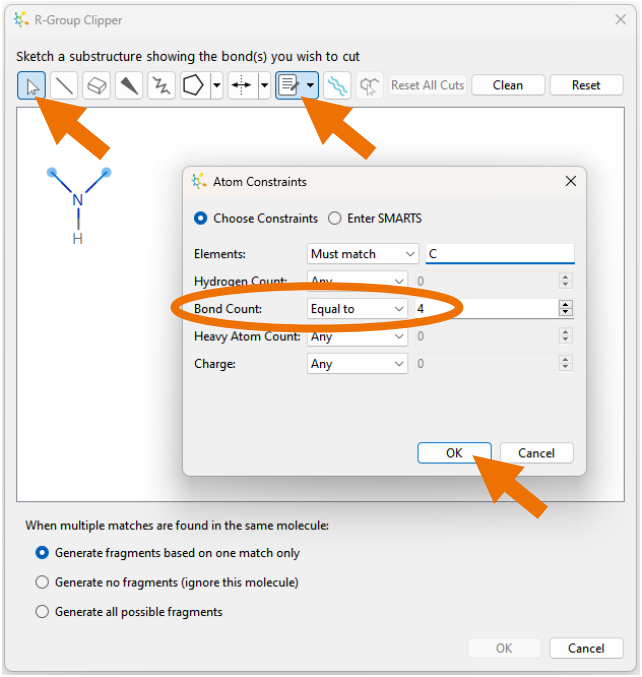



In the R-Group Clipper, we can sketch a substructure that defines how compounds in the data set should be clipped. In this case, we will sketch the secondary amine and impose some bond and atom constraints to limit the fragment to cyclic, aliphatic, and secondary amines.

- 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”.

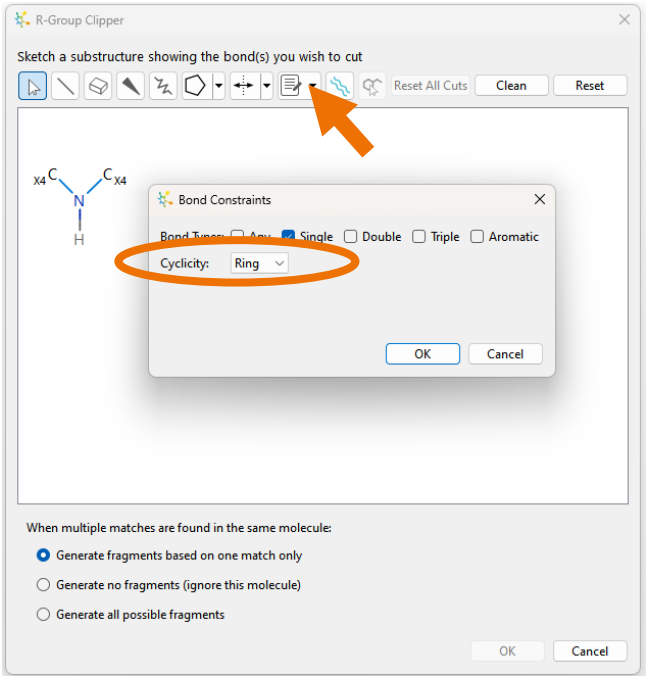
- To add atom constraints to the two carbon atoms, select them both by pressing the **Ctrl** key while using the **Selection** tool  .




- Open the **Constraints**  menu and choose **Edit Atom Constraints** to display the **Atom Constraints** dialog.

- Specify that each carbon atom’s bond count should be **Equal to 4** and click the **OK** button.

Selecting a bond count of four ensures the carbon is $sp^3$ hybridised rather than $sp^2$.

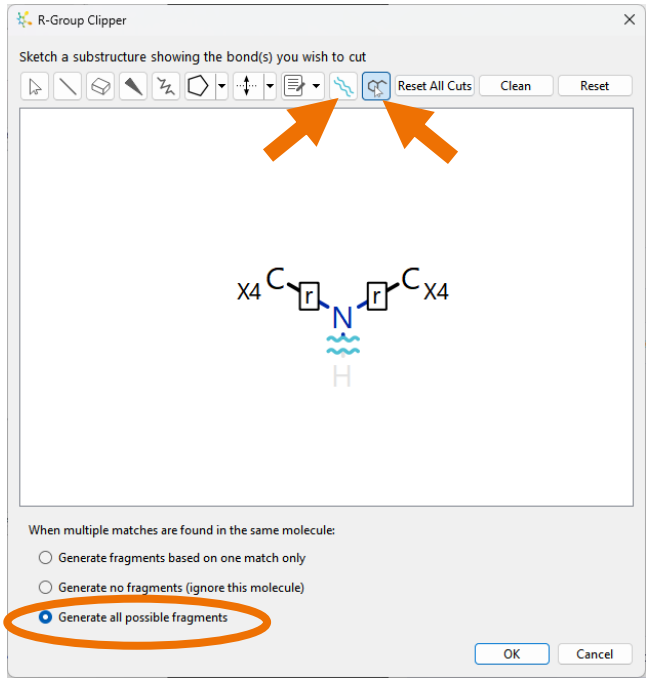



- To add bond constraints, select the two N-C bonds by pressing the **CTRL** key while using the **Selection** tool .
- Choose **Edit Bond Constraints** from the **Constraints** menu to display the **Bond Constraints** dialog.
- Select **Ring** from the **Cyclicity** options to specify that these bonds must be single bonds that are part of a ring, and click the **OK** button.

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

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

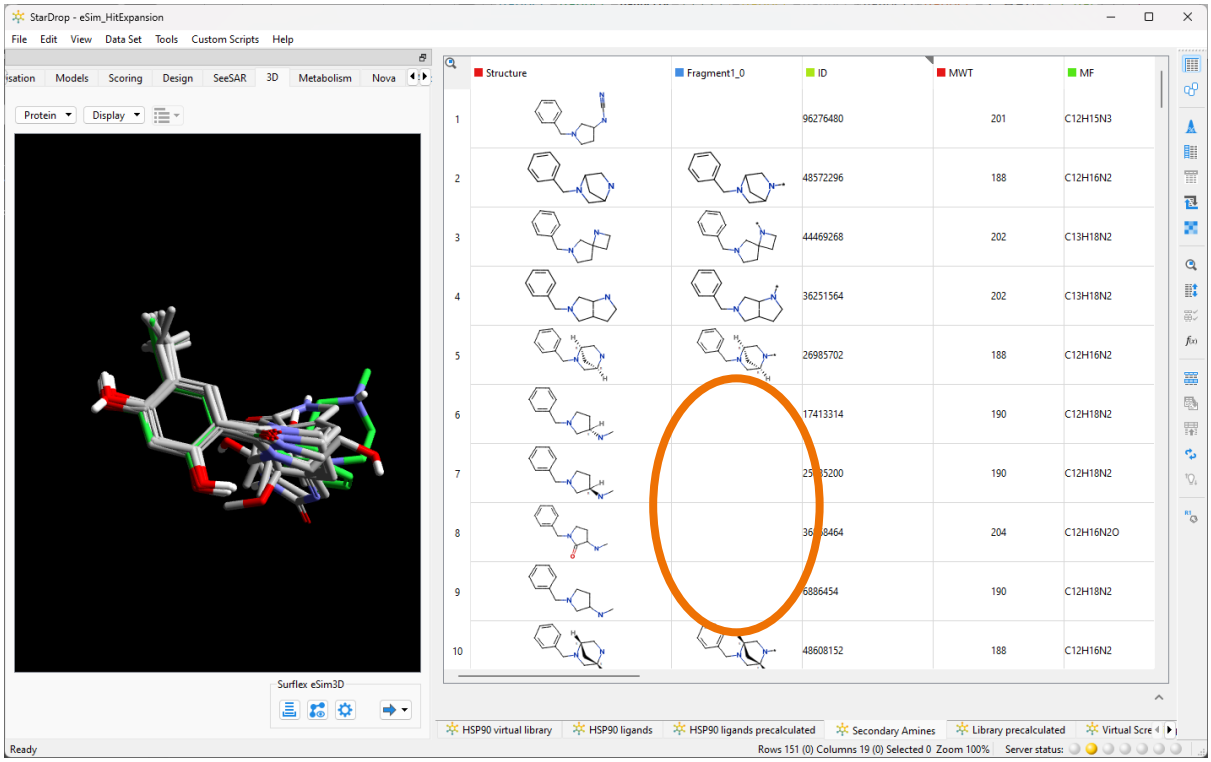
When comparing molecules in the data set with the specified substructure, multiple matches might be found within the same molecule. At the bottom of the **R-Group Clipper** dialog, you can specify what should happen when this occurs.




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

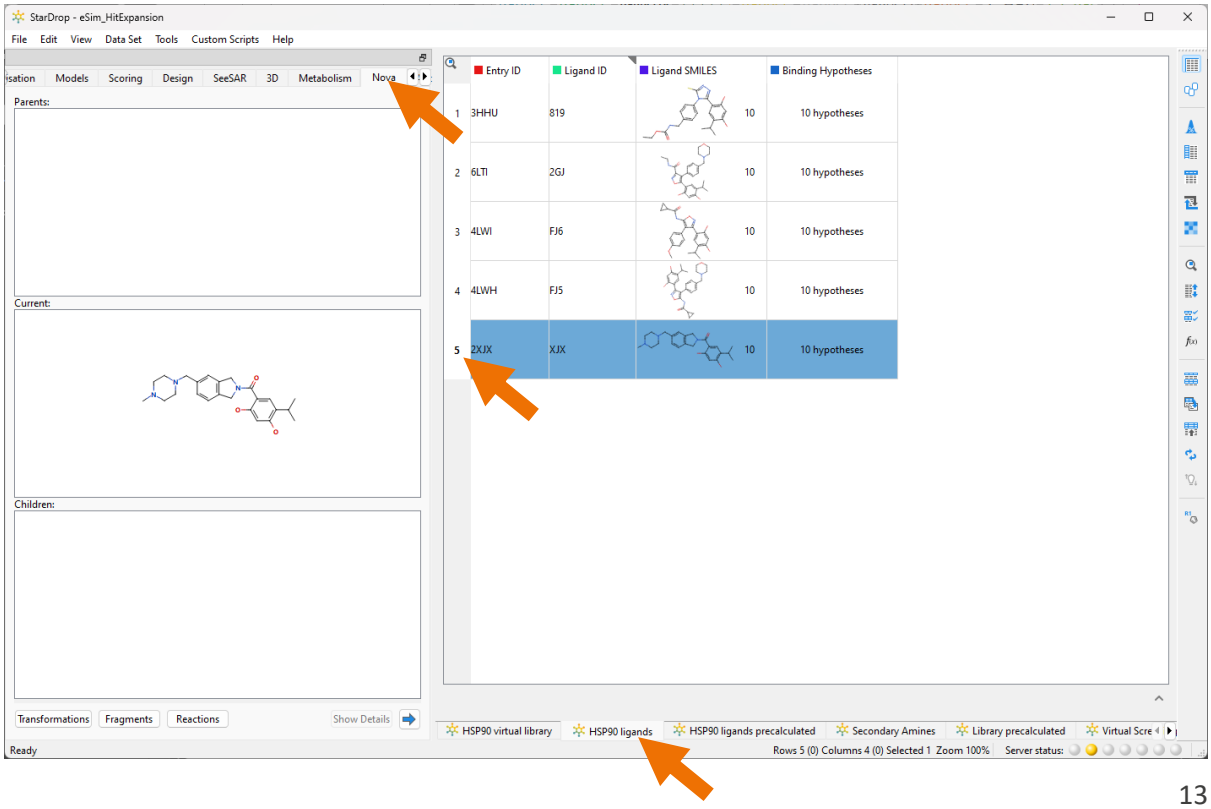
As shown in the screenshot below, the fragments will be generated in a new column called Fragment1_0, with an asterisk * indicating the attachment point.

**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 (XJX) is available in the **HSP90 Ligands** data set.

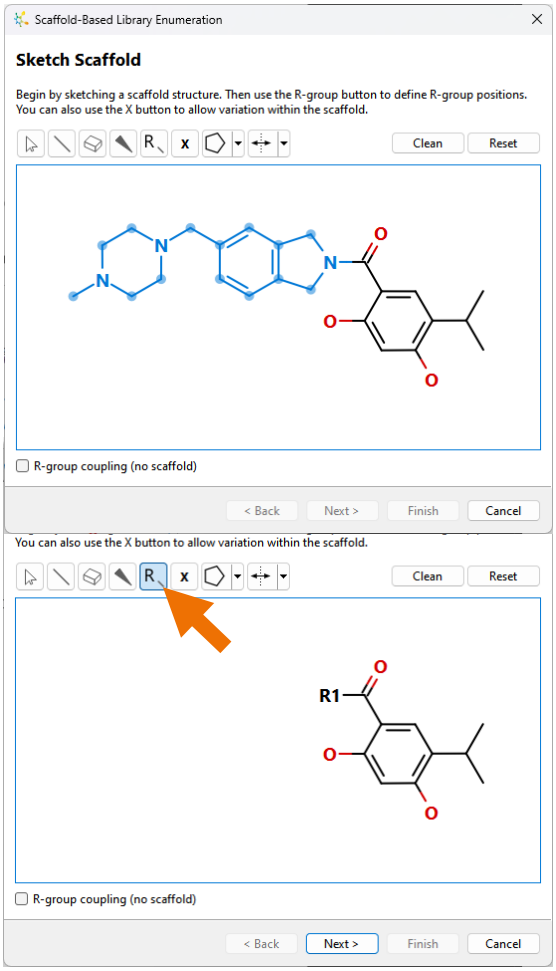
- Click the **HSP90 Ligands** data set at the bottom.
- Select the row with Onalespib (XJX).
- Click the **Nova** tab.

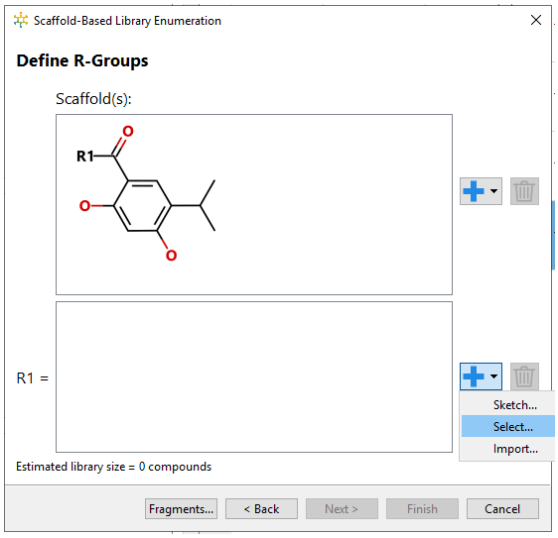


- Open the Surflex eSim3D menu at the bottom by clicking on the blue arrow  at the bottom of the **Nova** area to start the enumeration.
- In the wizard that appears, select **Scaffold-Based Library Enumeration** and click the **Next** button.

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.



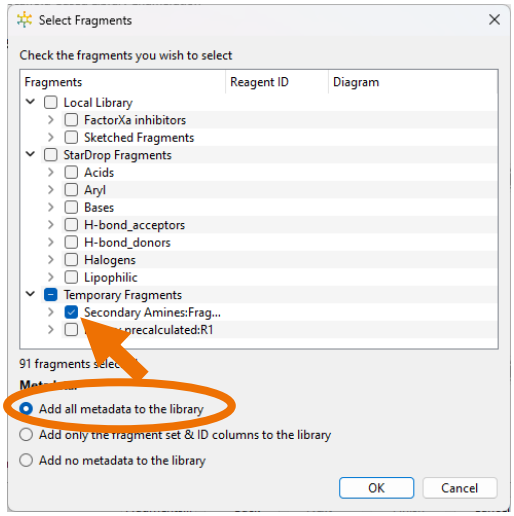

The **Define R-groups** page is displayed. Here we will define the list of secondary amine fragments to use in the enumeration.

- 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 named **SecondaryAmines:Fragment1_0** in the "Temporary Fragments" list 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 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, along with their price, quantity, supplier and other useful information for proceeding to synthesis.

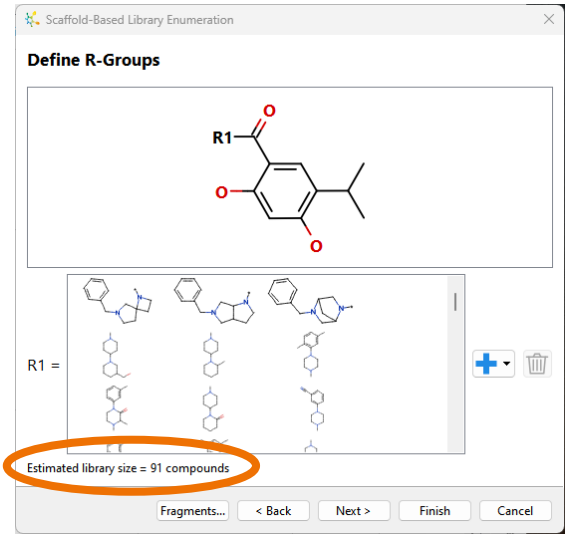
The fragments selected will be shown next to R1. If we wish to add more fragments, we can do so by clicking the **Add** button  again, but in this case, we will only use the fragments we already have.

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 the newly enumerated structures along with all the reagent meta-data from eMolecules.

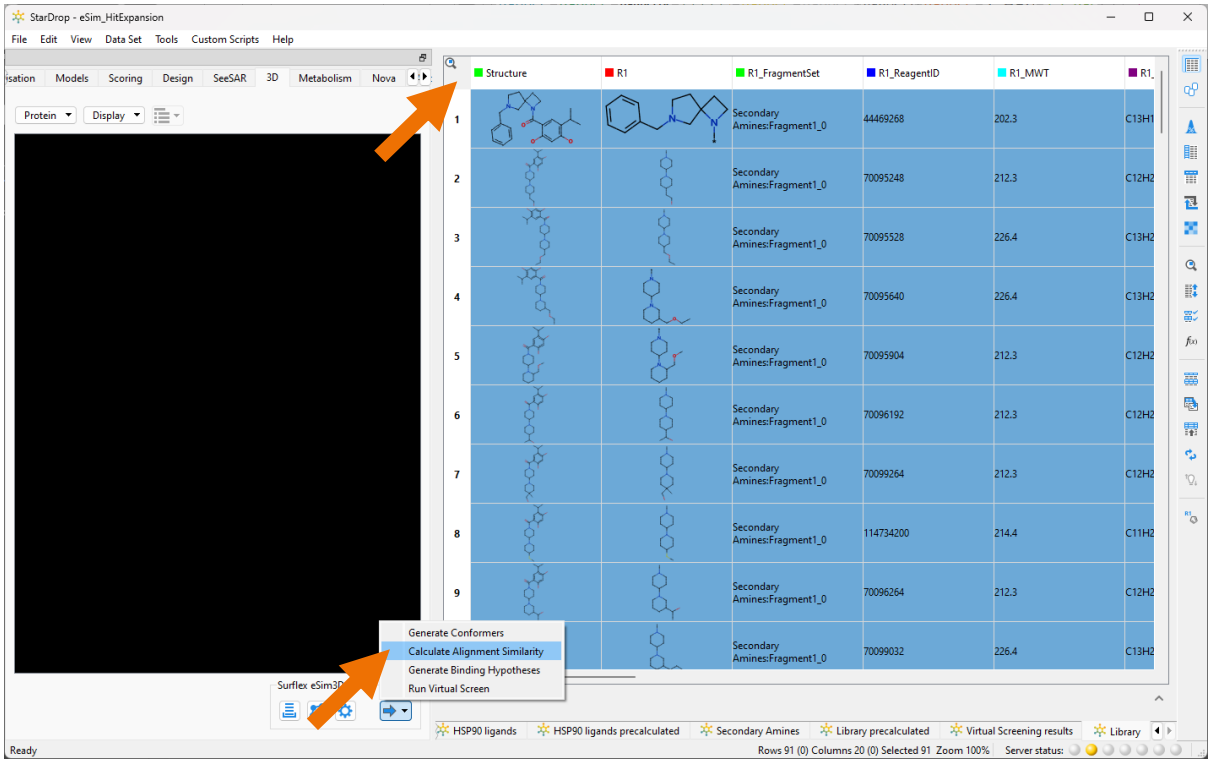
- Switch to the new **Library** data set at the bottom.

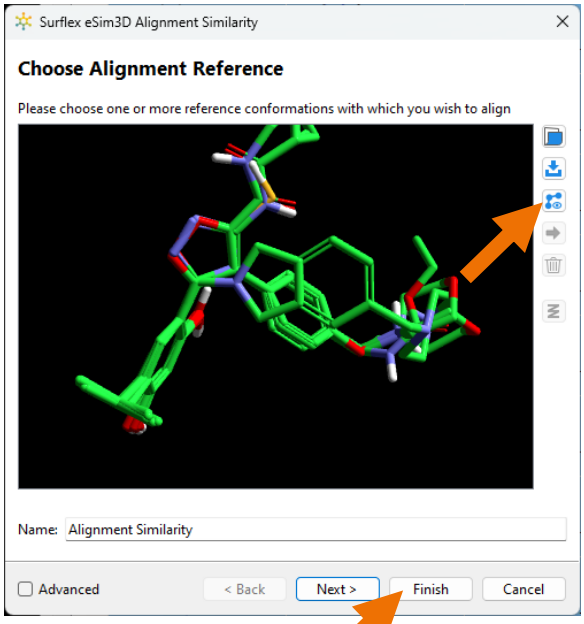


## Prioritising the Best Compounds

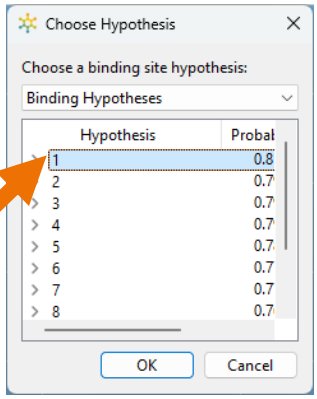
One way to evaluate the new analogues of Onalespib is to use the Surflex eSim3D module to compare the alignment similarity of each compound against the binding hypothesis we previously generated.

- Select the **3D** tab on the left.
- Select all the compounds in the **Library** data set by clicking in the top left corner of the data set.
- Open the Surflex eSim3D menu at the bottom by clicking on the blue arrow  and select **Calculate Alignment Similarity** to run the alignment.






- 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.

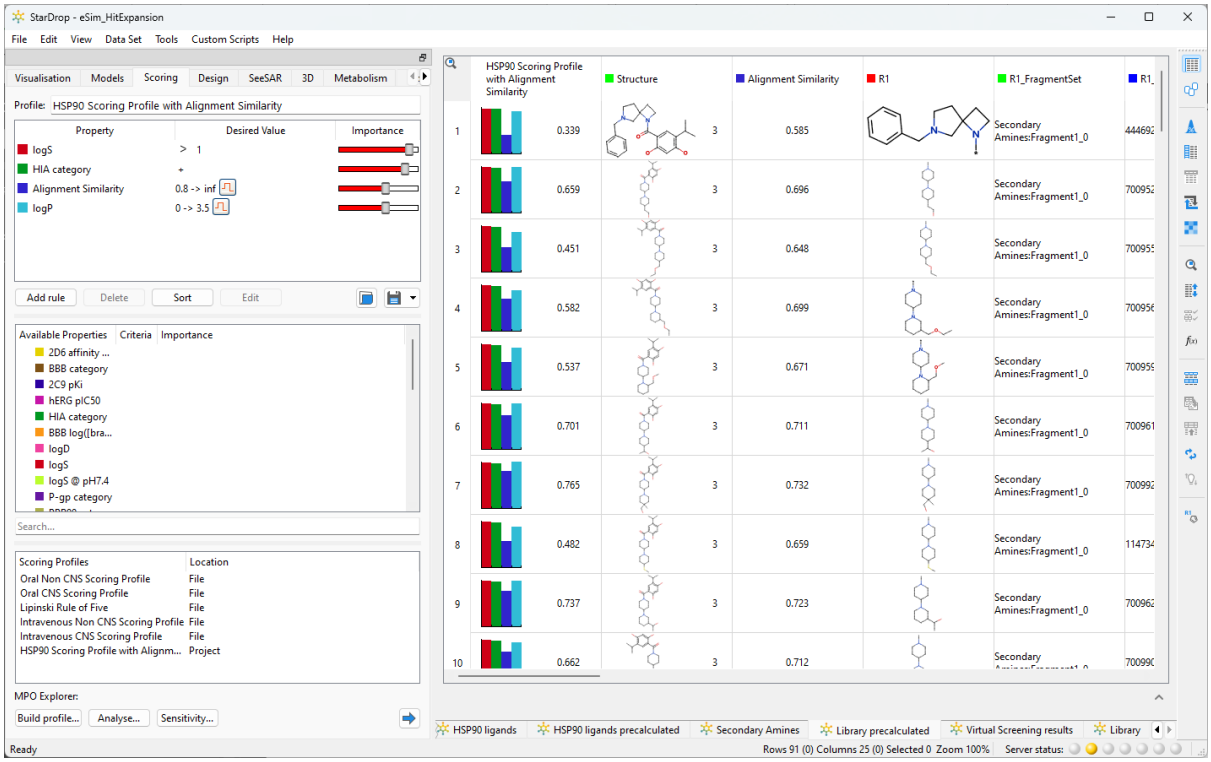
- Select Hypothesis 1.
- Click **OK**.
- In the **Choose Alignment Reference** window, click **Finish** to run the alignment with the default options.

When the alignment is complete, the calculated similarity value for each molecule will be added to our data set in the new **Alignment Similarity** column. If you prefer, the results of the alignment have been saved with the project in the **Library precalculated** data set.

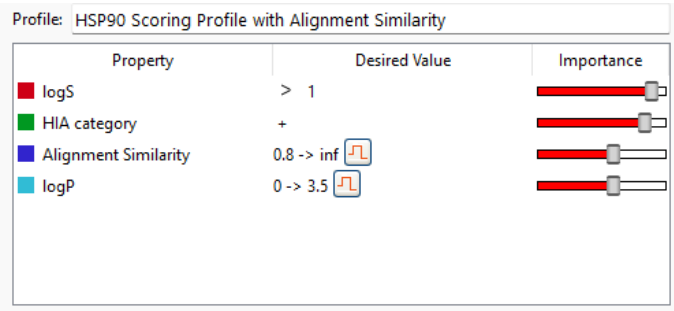
- Right-click the header of the **Alignment Similarity** column, select **Sort** and then **Descending**.

While the alignment score gives a measure of affinity to our protein target, we know that there are other important properties required for a compound to be a good drug candidate. We can examine how our compounds perform against multiple property criteria simultaneously by using StarDrop's approach to multi-parameter optimisation called Probabilistic Scoring.

- Change to the **Library precalculated** data set.
- Select the **Scoring** tab.




A scoring profile enables you to define a set of criteria that are important for your project. StarDrop provides some example profiles, but here we would like to find compounds with good solubility and absorption, an alignment score greater than 0.8 and logP between 0 and 3.5.




For this example we have already created the profile and run it against the library, saving the results in the **Library precalculated** data set.

The first column of the data set contains the pre-calculated score. The score is a value between 0 and 1, representing the likelihood of the compound meeting all the criteria in the profile allowing for the uncertainty in any measurements or predictions used. The histogram in each cell gives a quick overview of each property's impact on the compound's score.

- Right-click the **HSP90 Scoring Profile with Alignment Similarity** column, select **Sort** and then **Descending**.

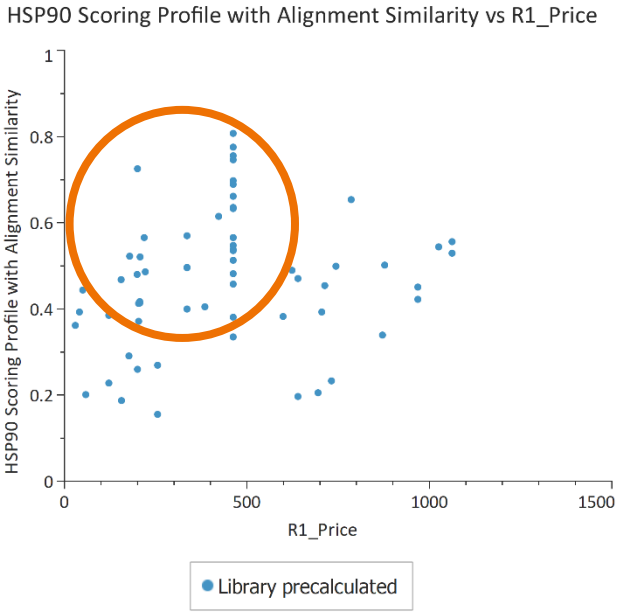
Not surprisingly, many of the best-scoring compounds show high alignment similarity, but there is some variation due to differences in the predicted ADME properties. We can use the metadata we imported from eMolecules to consider the cost of the reagents associated with the library compounds.

We have created a dashboard with two plots to explore this library further. A dashboard allows you to copy charts in the Visualisation area into a separate window.

- Go to the **Visualisation** area, click the ☰ button and choose **Load Dashboard**.
- In the **Load Dashboard** dialog, choose **NewDashboard.**

We will look at the first plot titled "HSP90 Scoring Profile with Alignment Similarity vs R1_Price".

In the top left corner, we can see that a few of the virtual compounds achieve a high score and use inexpensive reagents. These compounds might be worth progressing to further investigation, perhaps by docking into the crystal structure and/or obtaining experimental measurements.

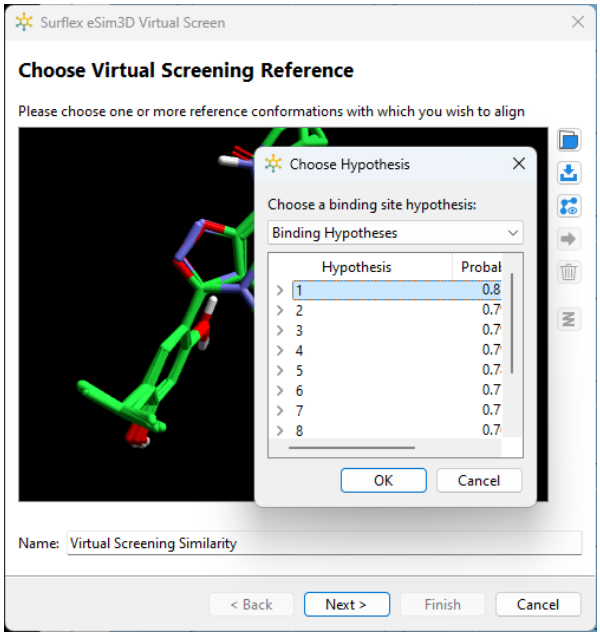



## Virtual Screening

Another way that we can explore new chemical space around our lead compound(s) is to perform a virtual screen against a large library. Virtual screening allows us to identify scaffold hops that have similar 3D shape and electrostatic profiles and are, therefore, likely to bind to our target but have different 2D structural motifs. In this example, we will use our binding hypothesis for virtual screening against commercially-available compounds from the Enamine stock collection. You can download prepared virtual screening collections with conformers pre-generated for the Enamine, MolPort, and eMolecules libraries from the Customer Hub area of Optibrium's website:

https://optibrium.com/resource-type/virtual-screening-collections/

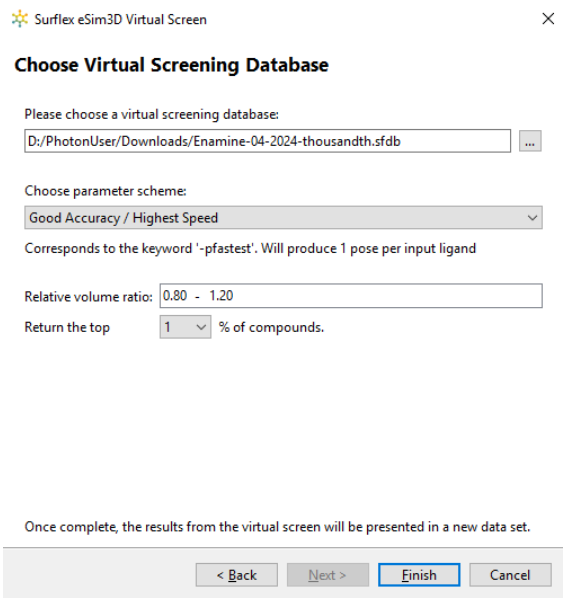
- Minimise the dashboard and return to the **3D** area.




- Open the Surflex eSim3D menu at the bottom by clicking on the blue arrow  at the bottom of the 3D area and select **Run Virtual Screen**.
- Click the **Binding hypothesis**  button.
- Select the first binding hypothesis and click **OK** to continue.
- Click **Next**.

The following window asks you to select the Virtual Screening Database and set parameters for the calculation.

We will use the smallest database (Enamine 4-2024 Thousandth) and the fastest settings for this example to provide a preview of the type of results we would obtain if screening the full library. When screening an extensive database, you will probably want to invest the time for a more accurate calculation on the full database and return only the top-scoring results.



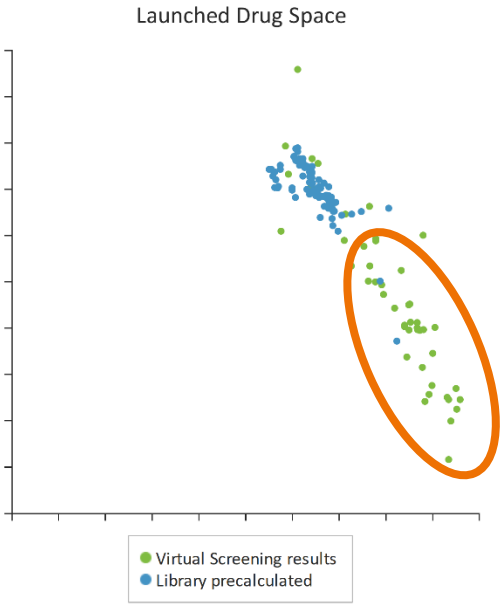

- Click the three dots next to the file selection and choose **Enamine-04-2024-thousandth.sfdb**. You can download the file from this link if necessary.
- Choose **Good Accuracy/Highest Speed** from the menu.
- Keep the default relative volume ratio. This pre-selects which compounds to screen based on a comparison of the molecular volumes of each compound with your alignment reference. A range around 1.0 will return molecules approximately the same size as your reference. You can search for larger or smaller molecules by changing the relative volume ratio.
- Because we are using a very small screening library, choose to return the top **1%** of results.
- Click **Finish**.

The results of the virtual screen will be returned in a new dataset once the job has been completed. If you prefer not to wait for the calculations to run, the results have been saved with the project in the **VS precalculated** data set.

### Comparison of the Libraries

We will use a chemical space visualisation to compare the virtual screening and the library enumeration results.

- Re-open the dashboard by clicking the ☰ button at the bottom of the **Visualisation** area and choosing **Load Dashboard**
- Select the second tab in the dashboard containing the chemical space plot titled "Launched Drug Space".




The chemical space plot shows that our virtual screening results (shown as green points in the figure) explore new regions of chemistry space and may be useful for scaffold hopping. We could proceed to evaluate these compounds using StarDrop's predictive models and MPO scoring to select some of the best results to advance for further study.

## Conclusions and Additional Resources

In this worked example, we have demonstrated several ways that we can use StarDrop's Surflex eSim3D module 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. We enumerated a focused library around one of the known inhibitors using the Nova module, and we evaluated which compounds best matched the generated binding hypothesis. We finished with scaffold hopping using a virtual screening approach to explore new areas of chemistry space.

StarDrop contains many more features to continue with further SAR analysis, property prediction, and MPO scoring for prioritisation of these leads. For additional StarDrop tutorials, please visit www.optibrium.com/tutorials. If you have any questions or feedback, please feel free to contact support@optibrium.com.