BioPharmics™

Version 1.0.1

QuanSA™ Plugin for PyMOL User Guide

optibrium™

Contents

1. Introduction

This guide introduces the graphical interface for the QuanSA plugin for PyMOL[1]. For help with other BioPharmics™ features and operations, please refer to the Surflex™ Manual. For help with PyMOL features and operations, please refer to the PyMOL documentation.

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

[1] The PyMOL Molecular Graphics System, Schrödinger, LLC.

2. How do I... install and run QuanSA in PyMOL

2.1 Pre-requisites

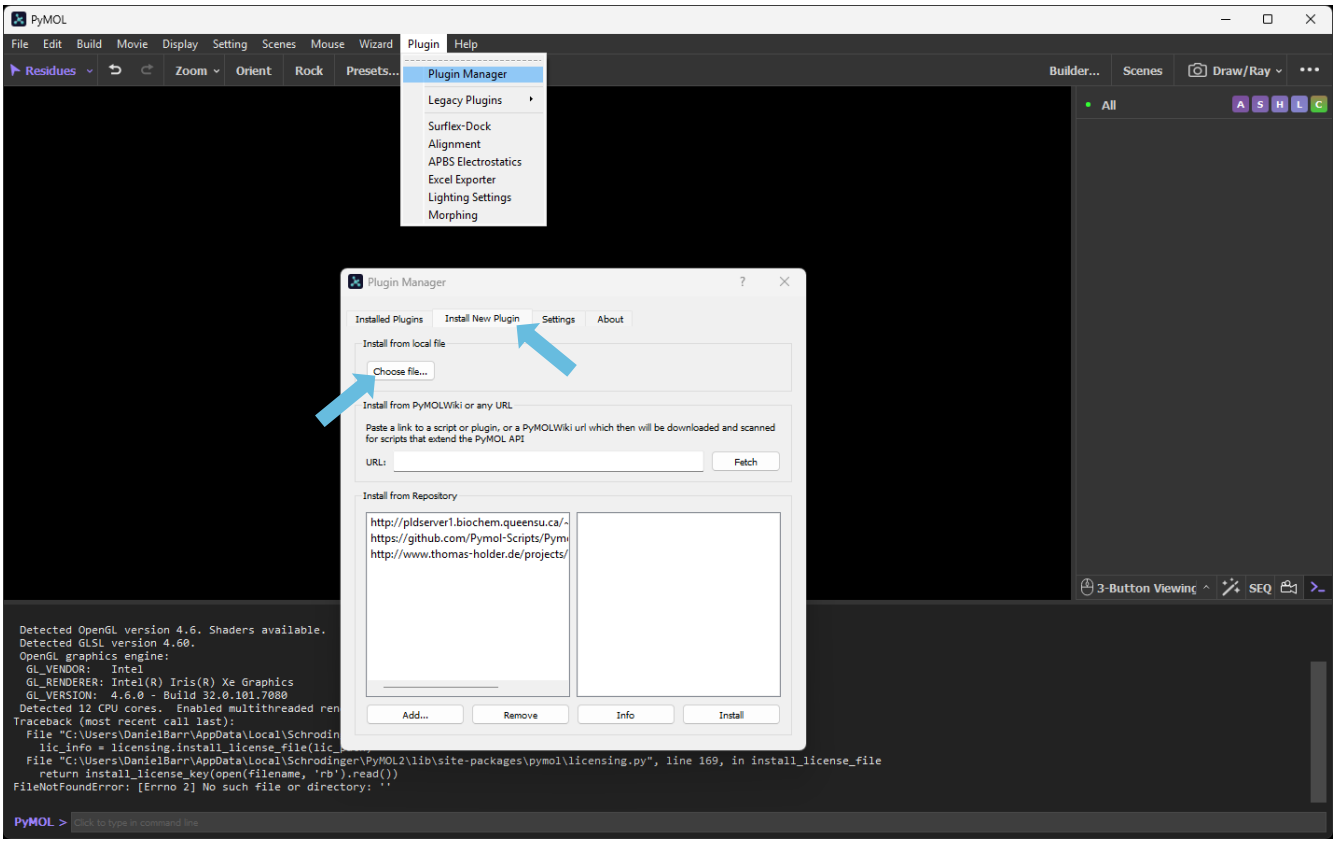
To install the QuanSA plugin for PyMOL, you will need the following:

- A working copy of PyMOL version 3.1 or newer, bundled with Python version 3.10.
- A license for QuanSA.
- The Surflex executables from the Optibrium website from the BioPharmics menu on the left (v5.211 or newer).
- The tar.gz file for the QuanSA plugin for PyMOL from the Optibrium website (version 1.0.1 or newer).

2.2 Installation

To install the QuanSA plugin in PyMOL:

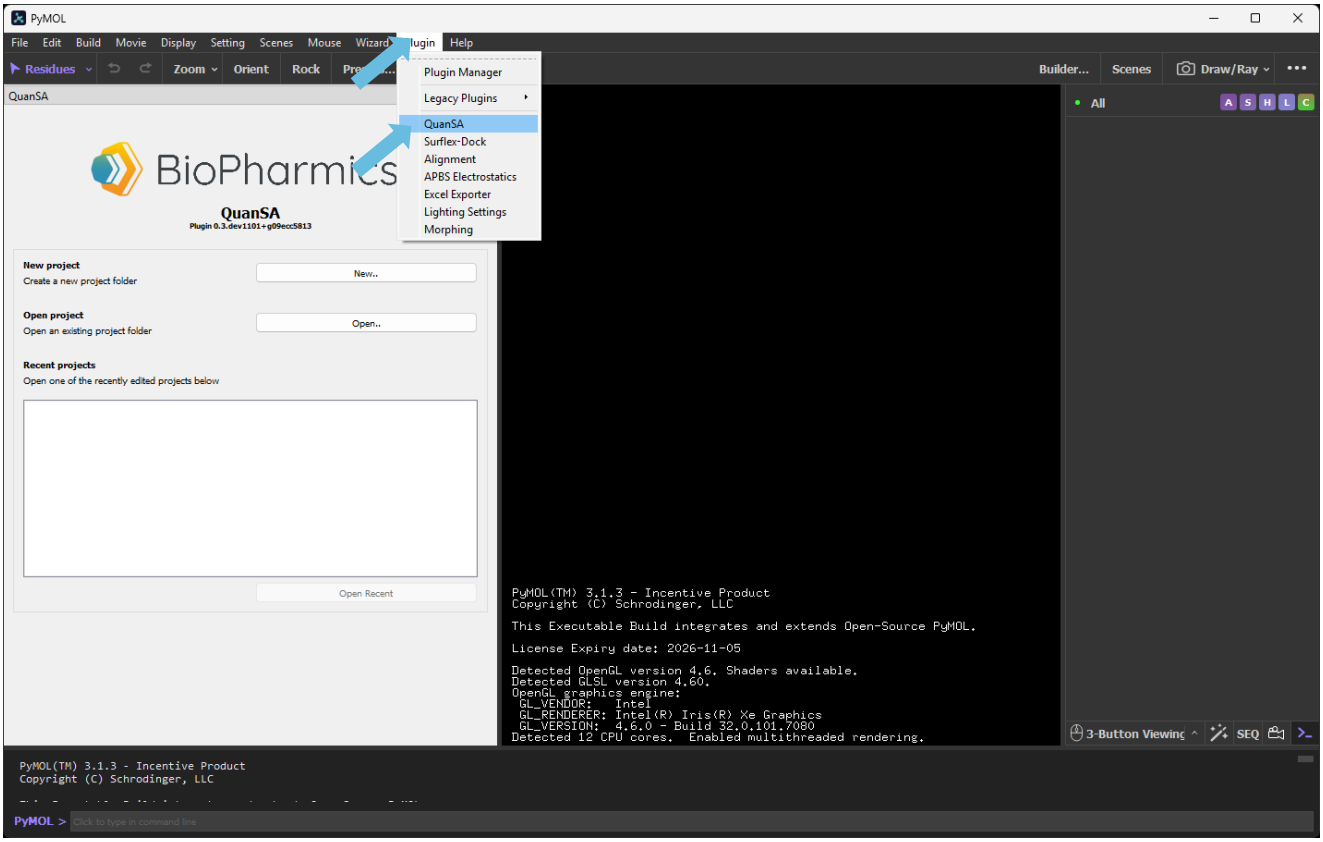
- Select **Plugin Manager** under the **Plugin** menu in PyMOL to launch the Plugin Manager.
- Select the **Install New Plugin** tab at the top of the **Plugin Manager**.
- Select **Choose File**.



- Browse to the location on your computer where you saved the tar.gz file for the QuanSA plugin (you do not need to unzip the archive) and click **Open**.
- A pop-up window will ask you to confirm the directory where the plugin will be installed. Keep the default and click **OK**.
- A pop-up window will confirm that the plugin has been installed. Click **OK** and close the **Plugin Manager**.

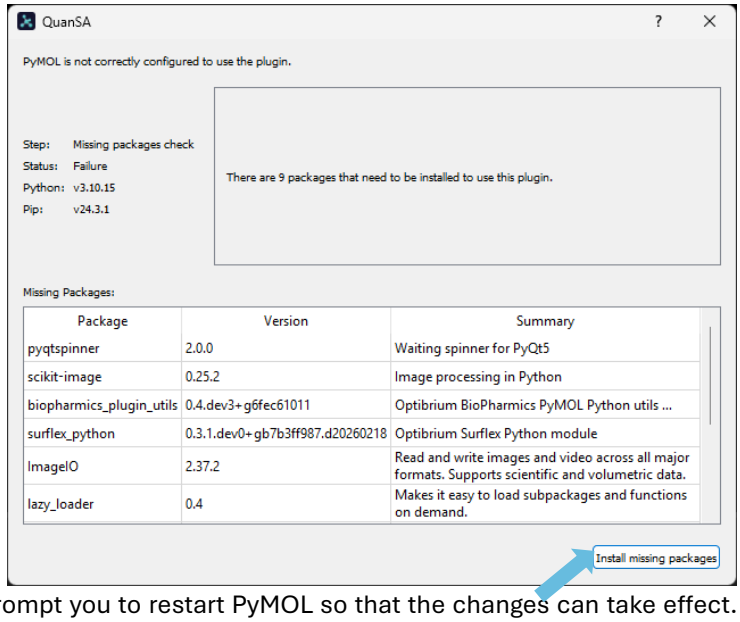
To open the QuanSA plugin:

- Select **QuanSA** under the **Plugin** menu in PyMOL.



The first time the QuanSA plugin is opened, a wizard will help you configure the plugin to use the Surflex executables and license. You will need to know the location of the Surflex executable files (sf-tools.exe and sf-quansa.exe) and the location of your Surflex license (surflex_bin.lic). See Section 2.1 for information about how to obtain the executable and license files.

The setup wizard first checks for the Python packages required by the plugin and installs them if necessary. Please note that you will need to be connected to the internet during this process. The Python dependencies for the plugin will be shown in the setup wizard and are listed in **requirements.txt** inside the tar.gz archive. (**Note:** there is no need to extract the archive before installation; everything can be handled through the wizard.)



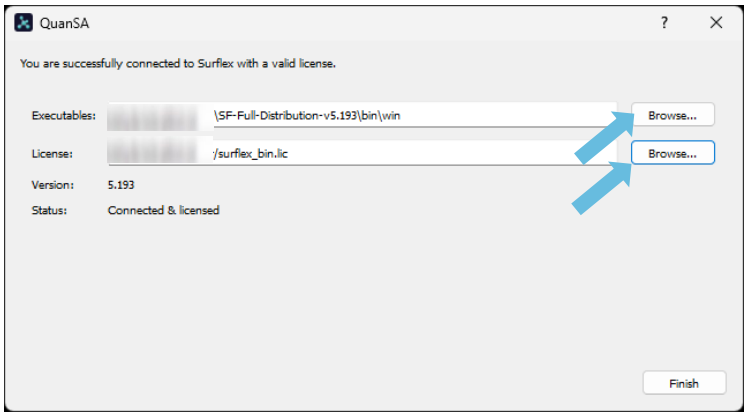
- Click **Install missing packages** to have the wizard download and install the required packages.

After installing the dependencies, the wizard will prompt you to restart PyMOL so that the changes can take effect. Close and re-open PyMOL to continue.

- Select **QuanSA** under the **Plugin** menu in PyMOL again.
- Click **Next** on the Setup Wizard. You will see a message that all required Python packages are installed.
- Click **Next** to proceed to the Surflex configuration.

The wizard will prompt you to browse to the path containing the Surflex executable files and your Surflex license (see Section 2.1 for information about how to obtain these files).

- Click **Browse** next to the **Executables** line of the wizard and navigate to the folder on your computer that contains the Surflex executables (sf-tools.exe and sf-dock.exe) and click **Open**.



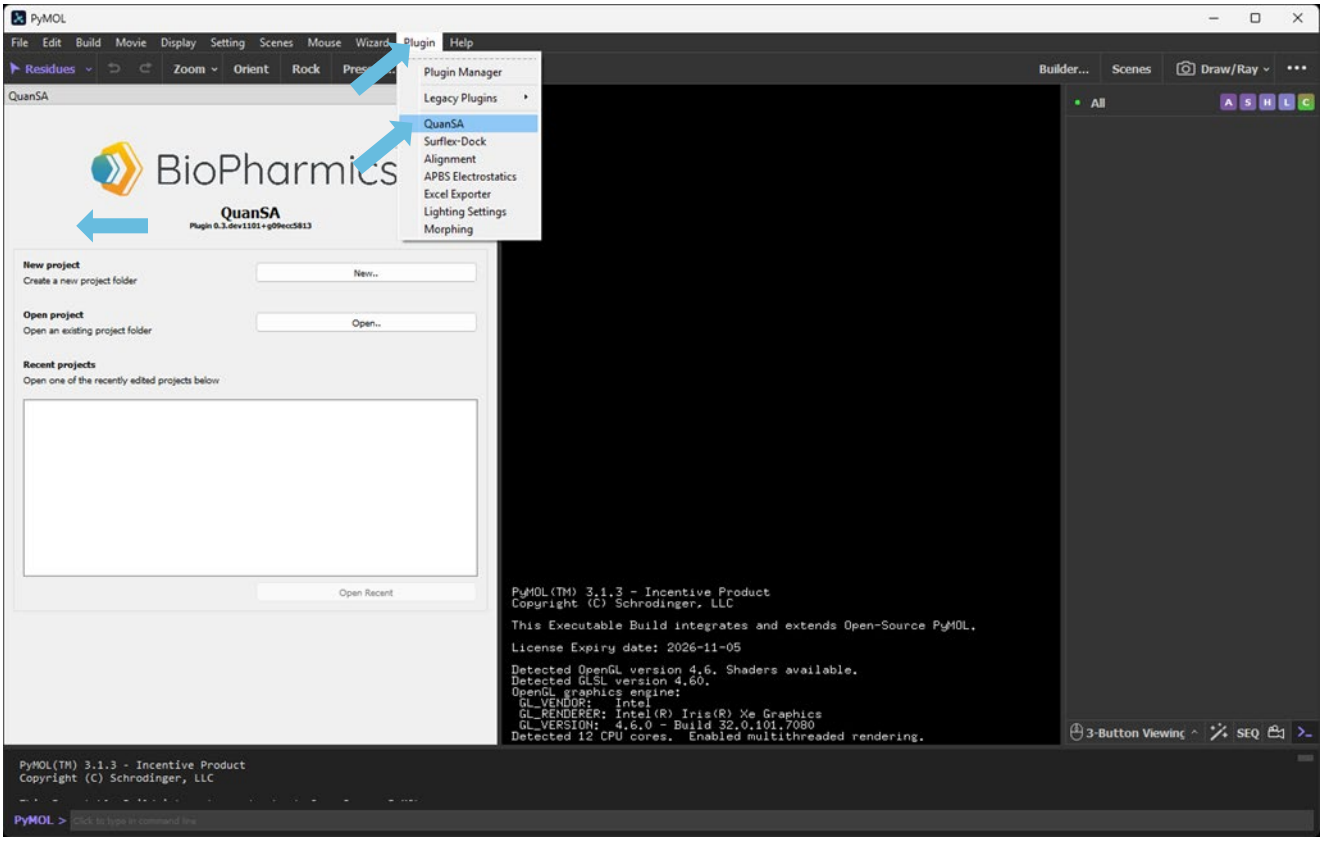
Hint: by default, the Surflex executables are found inside the folder **SF-Full-Distribution-v5.211** inside the bin folder; select the folder in the bin directory that is appropriate for your operating system).

- Click **Browse** next to the **License** line of the wizard and navigate to select the Surflex license file (surflex_bin.lic) on your computer and click **Open**.
- Click **Finish** to close the Setup Wizard.

2.3 Running the QuanSA plugin for PyMOL

To open the QuanSA plugin in PyMOL:

- Select **QuanSA** under the **Plugin** menu in PyMOL.



To dock the QuanSA plugin window inside PyMOL on Windows:

- Double-click on the title bar of the QuanSA plugin to dock the plugin.
- Click the windows icon  at the top right of the QuanSA plugin to return the plugin to its own window.

To dock the QuanSA plugin window inside PyMOL on Mac:

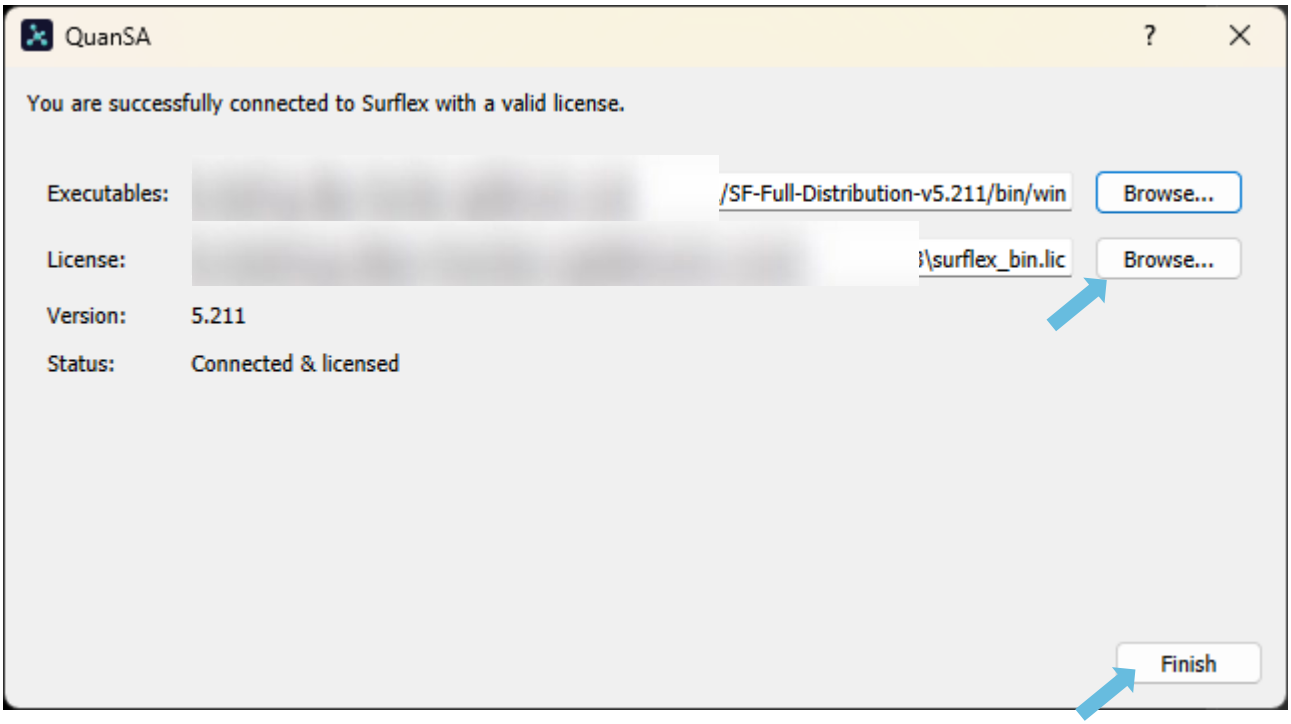
- Click and drag on the title of the QuanSA plugin and drag to the left or right side of the PyMOL display.
- Click and drag on the title of the QuanSA plugin and drag it to the centre of the screen.

On Linux, the plugin will always be docked to the main display and cannot be detached to a separate window.

2.4 QuanSA plugin set up

The first time you launch the plugin, you will be prompted to point to the locations of the Surflex executables and/or license file as necessary. On the **Executables** line, click **Browse** then navigate to the *SF-Full-Distribution-v5.211/bin* folder then choose the folder (win, mac, Linux) that matches your operating system. On the **License** line, click **Browse** then point to your *surflex_bin.lic* file. If the locations or license file are not suitable, the text will appear red.

Click **Finish** to continue to the QuanSA plugin.

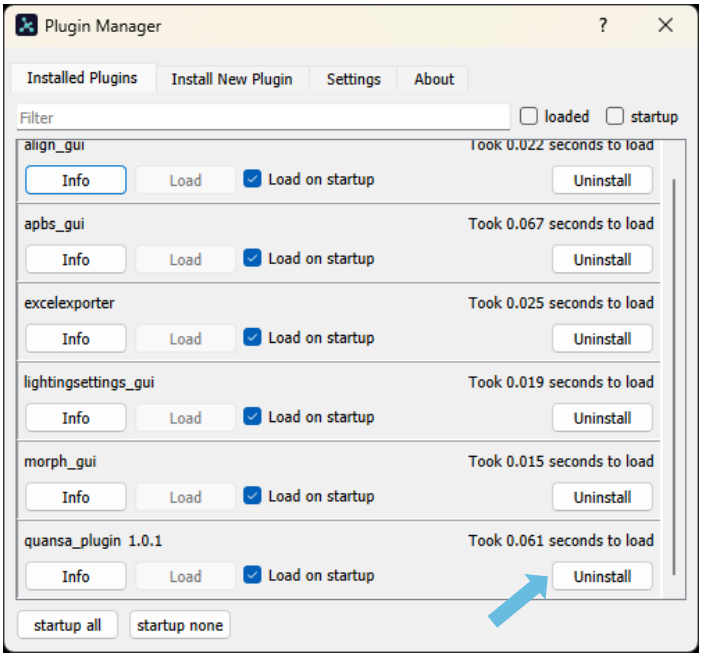


2.5 Uninstalling the QuanSA plugin for PyMOL

To uninstall the QuanSA plugin:

- Open the **Plugin Manager** under the **Plugin** menu in PyMOL.
- Within the **Installed Plugins** tab of the **Plugin Manager**, click **Uninstall** next to the QuanSA plugin.

To perform a complete removal of all plugin files and enable a fresh installation, you will also need to delete the configuration file *pymolpluginsrc.py* in the user's home directory.

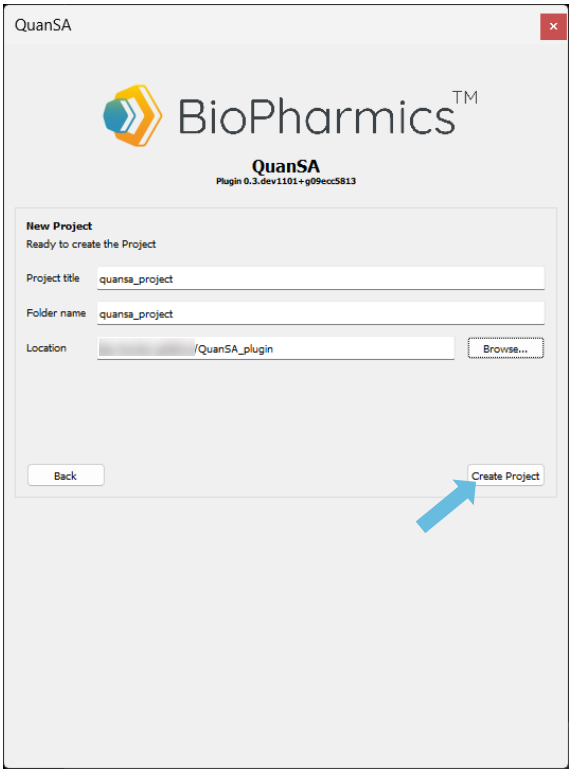


3. How do I... create or open a project

3.1 Creating a QuanSA project

Open the QuanSA plugin (see Section 2.3).

- Click **New** in the QuanSA panel.
- Type a name for your project in the **Project title** box.
- In the **Folder name** box, type the name to use for the new directory where project files will be saved; this will create a new directory with that name that will hold all of the project files.
- **Browse** to select the folder where you would like the project directory to be saved in your filesystem and click **Open**.
- Click **Create project**.



3.2 Saving a QuanSA project

QuanSA automatically saves your work in the directory specified when you created the project (see Section 3.1). There is no manual control for saving a QuanSA project; all files are saved as you work and will be restored automatically when opening the project. Any custom visualisations made in PyMOL will not be saved as part of the QuanSA project. You can save the PyMOL session separately (see Section 8.3).

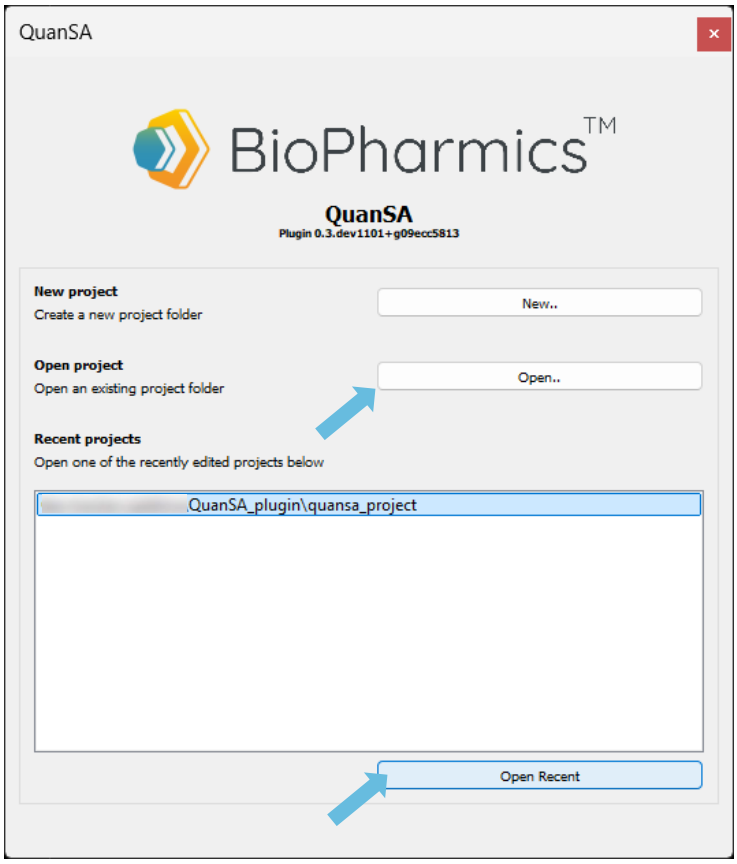
3.3 Opening a QuanSA project

To open a QuanSA project:

- Open the QuanSA plugin (see Section 2.3).

If you have recently worked on a project, the project directory will appear in the **Recent projects** panel of the QuanSA window.

- Click on the recent project you would like to open.
- Click **Open Recent** at the bottom of the QuanSA window.



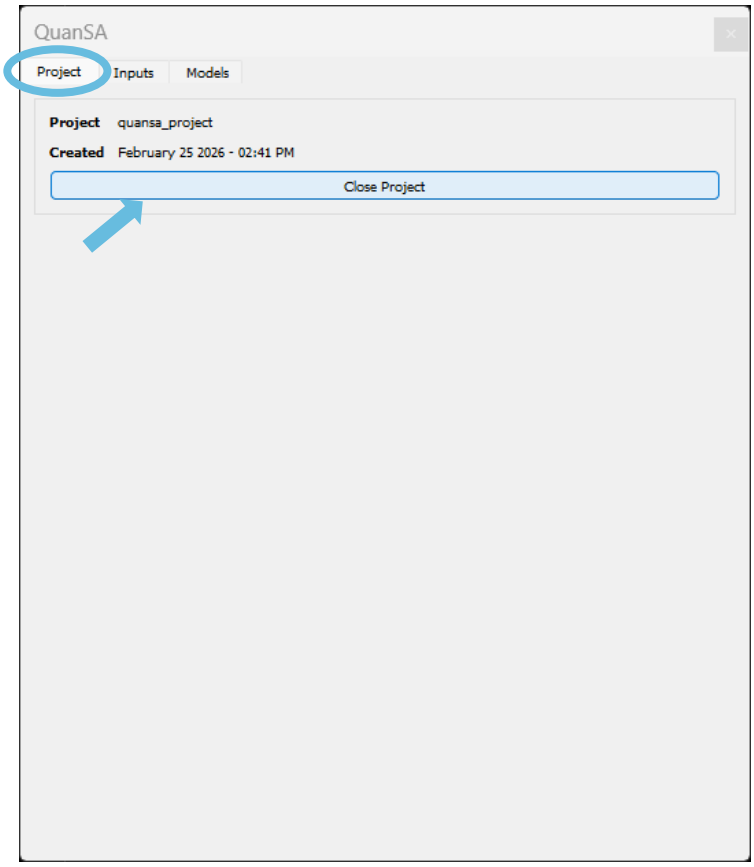
To open a project that does not appear in the **Recent projects** panel of the QuanSA window:

- Click **Open** in the QuanSA window.
- Browse to the project folder that you would like to open.
- Click on the project folder in your file system to select the project.
- Click **Open**.

3.4 Closing a project

To close the project and open another project:

- Click on the **Project** tab of the plugin.
- Click **Close Project**.

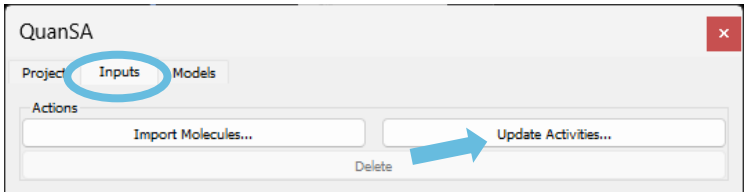


4. How do I... import and export data

4.1 Importing molecules

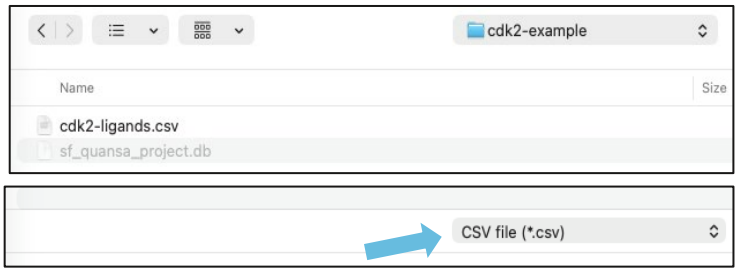
To import molecules for model building, switch to the **Inputs** tab in the plugin.

- Click on the **Inputs** tab in the plugin.
- Click **Import Molecules...** to import the ligand(s) in CSV, MOL2, or SDF format. or as a list of SMILES strings.



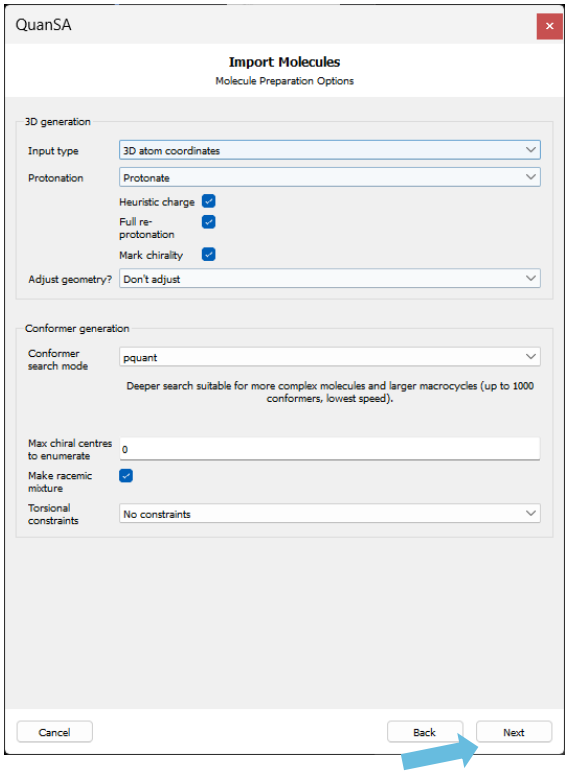
Note: If using a text file in SMILES format for input, put one SMILES string per line, with the compound name or identifier separated by a space after the SMILES string.

Note: After navigating to the folder where you have your data file, you will need to adjust the file type in the bottom of the file input dialog to see your file.



You can include activity values as a column in the same file as the structures you import, or you can upload activities from another text file that contains the compound name and its activity value.

- To update activities using a different file, click **Update Activities** and browse to the file containing the list of compound names and activity values.

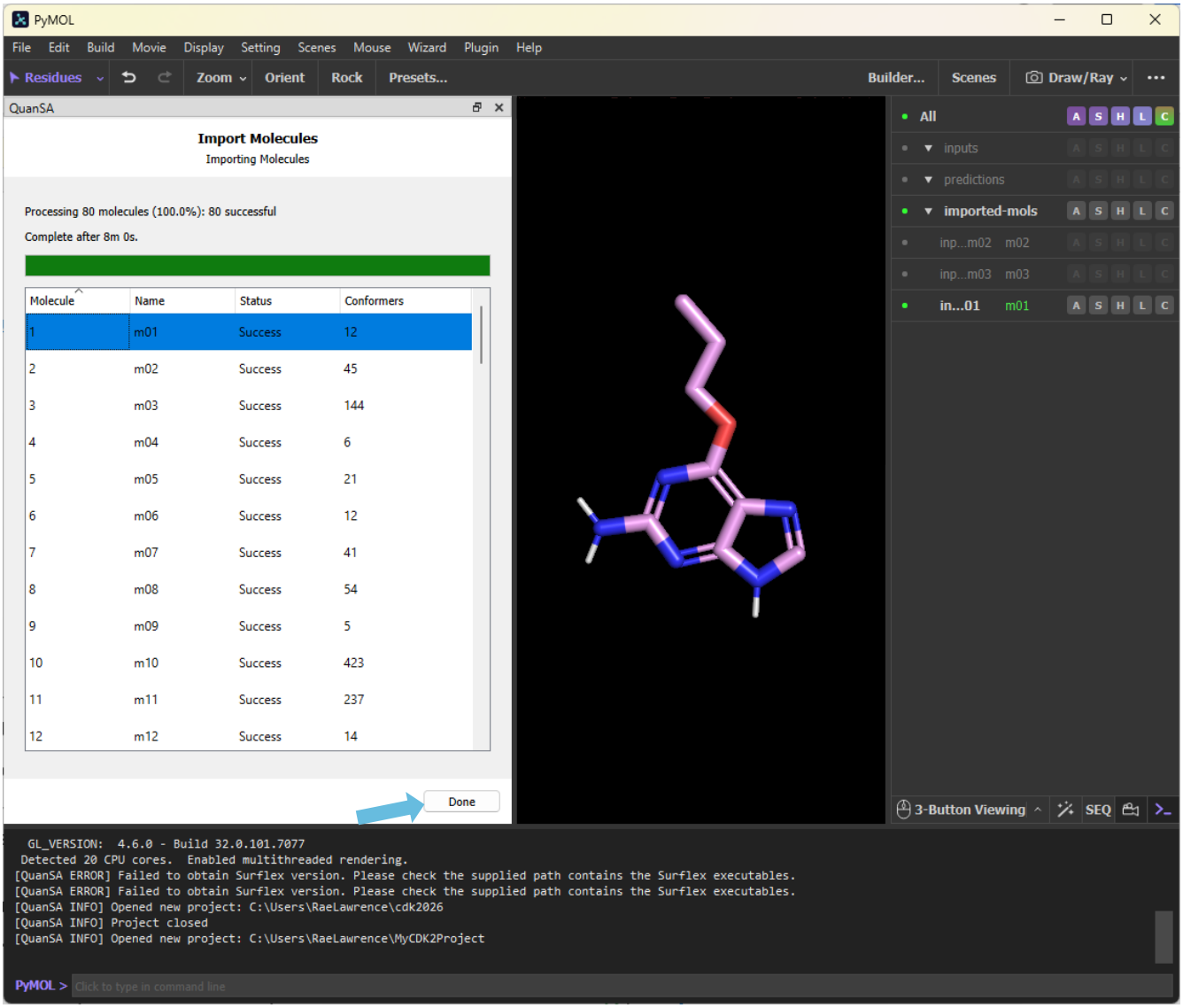


Regardless of the input format, 3D coordinates will be generated for each ligand. You can control the behaviour of protonation, charge manipulation, enumeration of chiral centres, and geometry optimisation. The behaviour of these functions is described in Section 2.3 of the Surflex Manual.

For each ligand, a conformation search is performed using parameter selection schemes, available from the pull-down menu in the **Conformer search** area of the **Import Ligands** wizard. The behaviour of these parameter schemes, with additional options for handling chiral centres and torsional constraints (e.g., from an NMR experiment), is described in Section 2.4 of the Surflex Manual.

When the ligands are imported and the conformer populations calculated, the panel will show the list of molecules with the number of conformations calculated. Clicking on any of the rows will show the 3D structure in the PyMOL.

Click **Done** to continue.



4.2 Exporting results

QuanSA automatically saves all your work in the directory specified when you created the project (see Section 3.1). There is no manual control for saving or exporting results from a QuanSA project.

QuanSA results are saved in the project working directory (see Section 3.1) in MOL2 format with files named:

`model-<i>-<ligand>-poses.mol2` - contains predicted bound poses for a ligand by a given model. These files can be opened in PyMOL or StarDrop, however if the associated interactions need to be re-analysed, the users will have to load them through the plugin.

`init-init-<N>.qml.mol2` - initial alignment results – Can be loaded directly in PyMOL

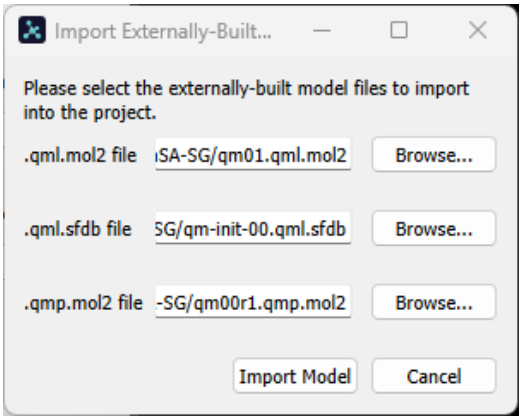
`model-<N>.{qml.mol2,qml.sfdb,qmp.mol2}` - files necessary if the user wants to use their model on the command line

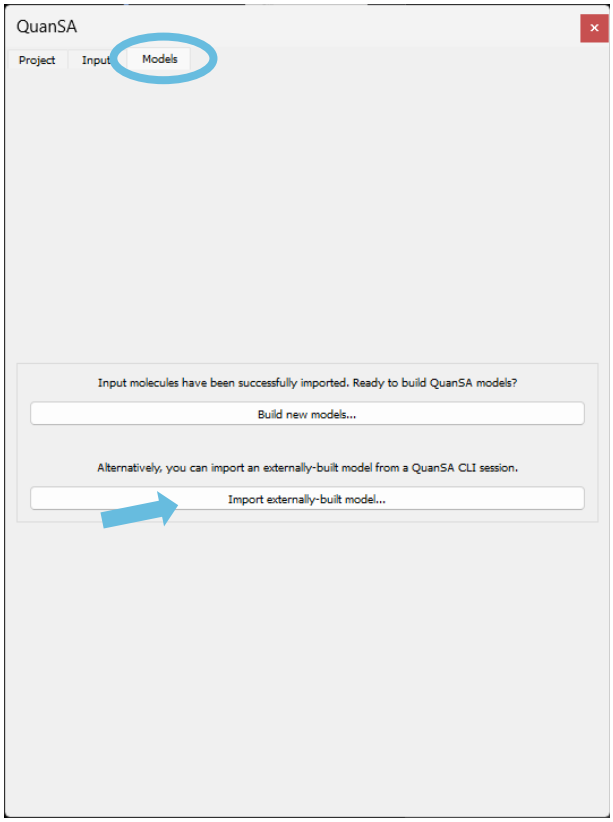
Predictions results can be exported to a text or spreadsheet editor (e.g. MS Excel) using **Ctrl+C** and **Ctrl+V**.

4.3 Importing a QuanSA model created from the Command Line Interface

In the plugin, you can import QuanSA models built previously outside of the plugin. Navigate to the **Models** tab, then click **Import externally-built model**.

You will be prompted to select the externally-built model files as shown in the dialog. Click Browse to navigate to the QuanSA project folder and choose your files. More information about the QuanSA model files can be found in the Surflex Manual.



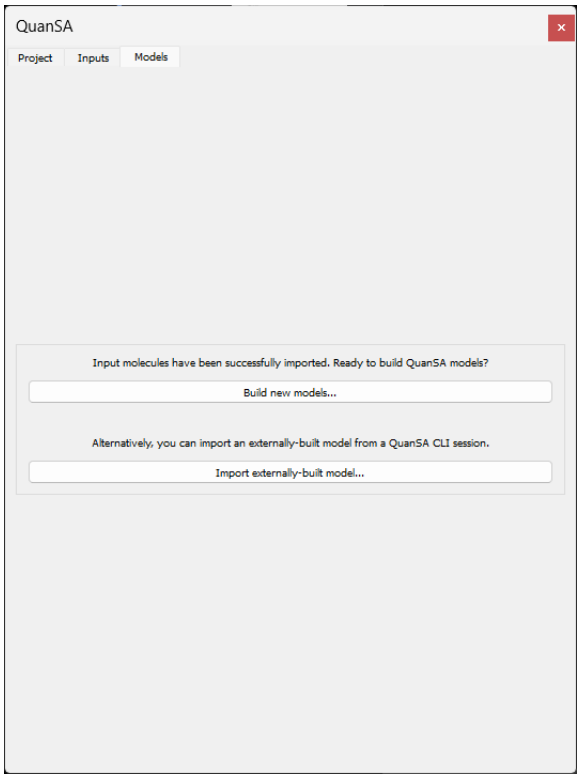


5. How do I... build a QuanSA model

Protonation and 3D conformations are calculated when molecules are imported (see Section 4.1). The basic workflow for calculating a QuanSA model is:

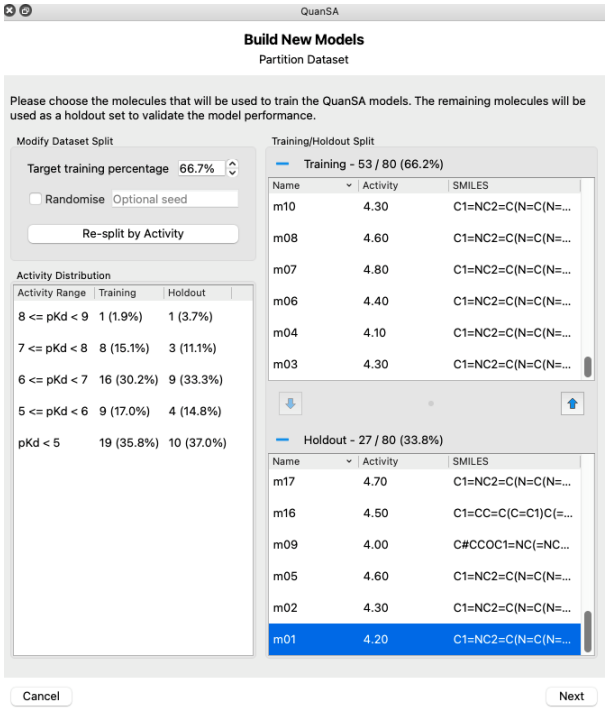
1. Partition the data set into training and holdout sets (Section 5.1)
2. Generate initial alignment hypotheses, which may be guided by 3D alignment references, if available. (Section 5.2)
3. QuanSA model induction using the alignment hypotheses. (Section 5.3)
4. Predict affinity for new compounds. (Section 6)

Once molecules are prepared, navigate to the **Models** tab in the QuanSA dialog and click **Build new models**. This will open a dialog in which you will partition your data into a training set and a holdout set which will be used to assess model performance.



5.1 Creating training and holdout sets

In the Partition Dataset dialog, there are several panels where you can define your partitions.



- **Modify Dataset Split**: By default, one third of the molecules are used for training, split by activity to attempt to make the activity distribution in the training and holdout sets roughly equal. Alternatively, you can choose to randomise the split.

  When **Randomise** is chosen, the data set will be split randomly. An optional seed can be supplied to ensure the random split is reproducible.

- **Activity Distribution**: This panel shows the distribution of activities in your training and holdout sets. You can confirm that the activities in either set are not skewed.

- **Training/Holdout Split**: In this panel, you can explore the collections of compounds that have been assigned to the sets. You can manually move molecules between sets by selecting the row and using the blue up/down arrow buttons.

Click **Next** to proceed to the **Initial Alignment Hypotheses** panel.

5.2 Initial alignment hypotheses generation

The next step is to produce an initial set of alignment hypotheses which are used to induce the QuanSA model.

In most cases, the default settings are appropriate, however, you can change various options in the **Initial Alignment Hypotheses** panel:

- **Alignment hypotheses:** The maximum number of alignment hypotheses to produce. Each hypothesis represents a different set of potential poses of the training molecules in the binding site. A QuanSA model can be built from each alignment hypothesis.

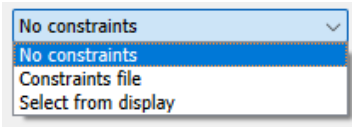
- **RMS:** Determines how close in distance (Angstroms) alternative pose groups must be to be considered separate. Increasing RMS will produce a smaller number of alignment hypotheses with coarser structural variation.

- **Seed molecules:** The maximum number of molecules to select from the training set to seed the alignment hypotheses.

- **Seed selection window:** The molecules with an activity within this value of the most active molecule in the training set is used to seed the alignment.

- **Assay delta:** The difference between the measured and predicted activity of a molecule below which a model is not penalised during training.

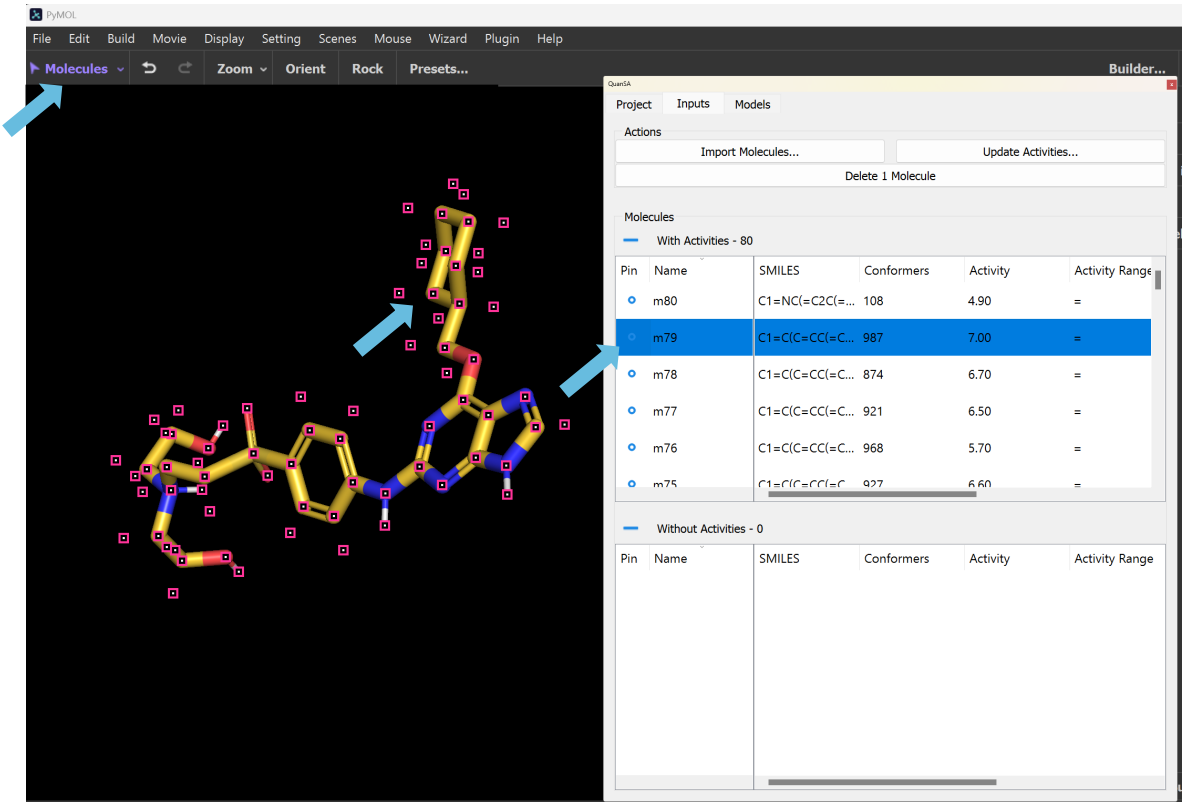
- **Positional constraints**: Molecular fragments to constrain position in the alignment. Constraints can be introduced from a Constraints file or by Selecting from display.



Constraints file: You can load a constraints file (mol2).

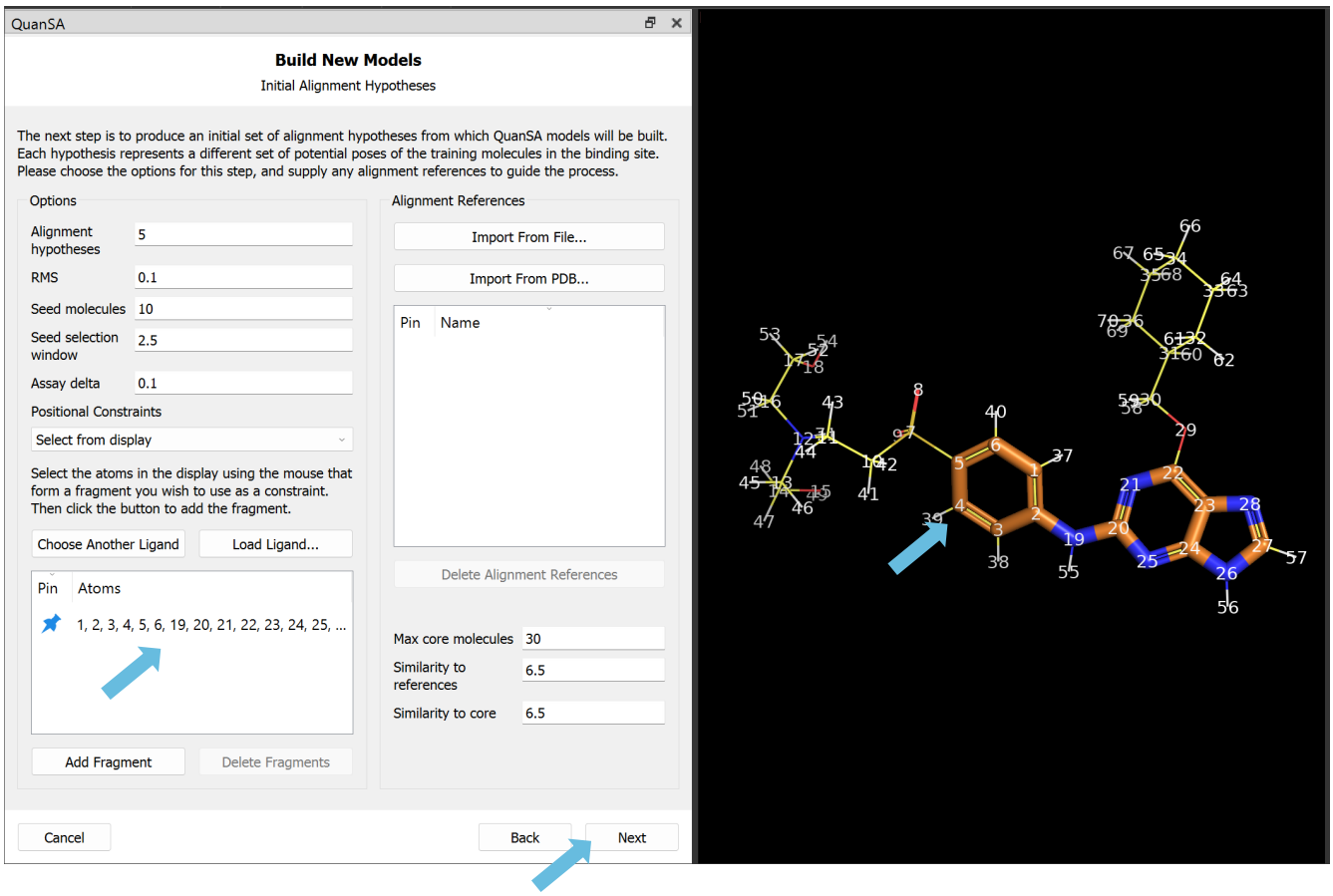
Select from display: When this option is chosen, the panel updates to give you the option to choose a ligand and fragments.

Note: To use **Select from display**, the ligand selection must be made prior to clicking **Build New Models** in the **Models** tab. You must return to the **Inputs** tab and **select** the ligand that you wish to use for constraints in the panel. The structure will display in PyMOL. Select the atoms in PyMOL by clicking on one of the atoms. You may have to adjust the mouse-action from the drop-down list in the upper left-hand corner of the PyMOL interface. When atoms are selected, they are highlight with a red box in PyMOL.



With the ligand selected in PyMOL, return to the **Models** tab and click **Build New Models**. When you return to the **Initial Alignment Hypothesis** dialog, choose **Select from Display** from the **Positional Constraints drop-down list**.

Next, click **Choose Ligand**. Your selected ligand will appear in PyMOL. Click on the atoms that you wish to constrain, then click **Add Fragment** in the **Build New Models** panel. The atom numbers of the fragment will be added to the list in the panel. Multiple fragments and multiple ligands can be used to add constraints.



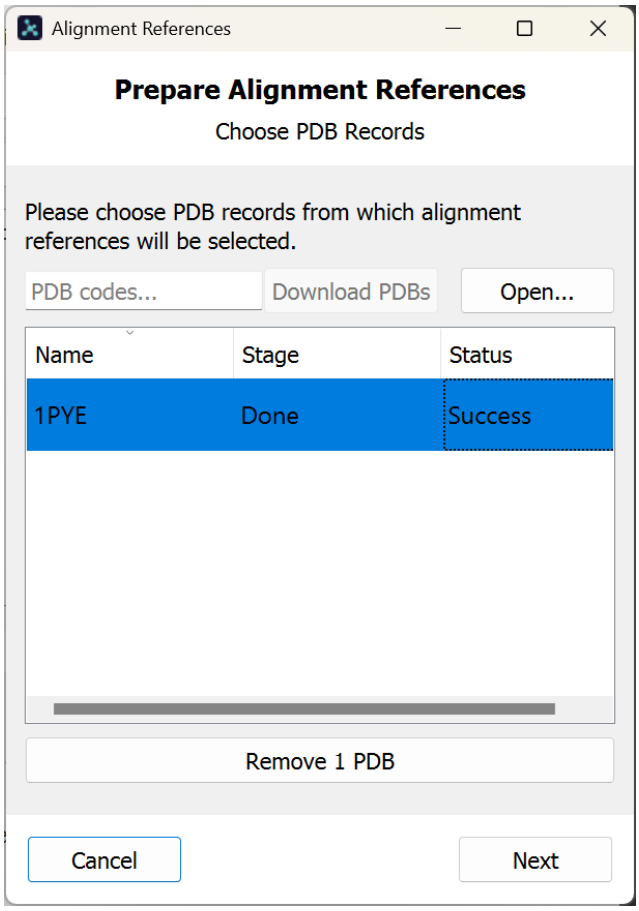
When suitable options are selected, click **Next** to initiate the calculation.

5.2.1 Using an alignment reference

If structural information for the binding mode is available, it can be included as an alignment reference. The structures can be imported from a file or can be imported from the Protein Data Bank by clicking the appropriate button in the **Alignment References** section of the panel.

- **Import from file:** Use this option if you have 3D ligand structures available in a mol2 file. Navigate to the mol2 file, then press **Open**.
- **Import from PDB:** Use this option if you have a PDB file or will download the structure from the PDB. When you click the **Import from PDB button**, the **Prepare Alignment References** dialog will open. Type in a PDB code,

then choose to download or open the PDB file. If you choose Open, you will be able to navigate to your file and click **Open**. **Download from PDB** retrieves the structure from the PDB. After either action, the protein structure is prepared and the ligand(s) extracted to be used as references to guide the alignment. You can delete prepared structures by selecting them in the list and clicking the **Remove X PDB** button.



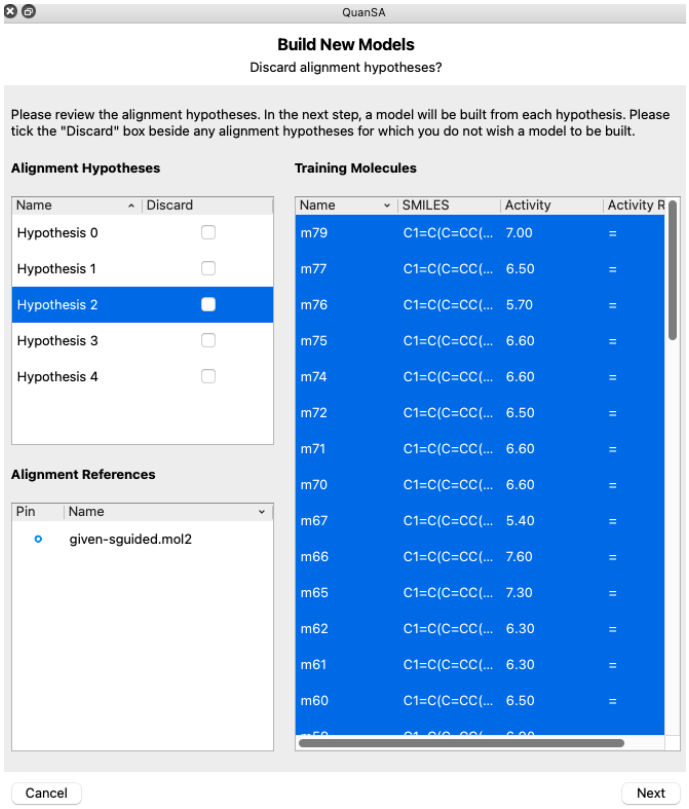
Other options in the **Alignment References** section include:

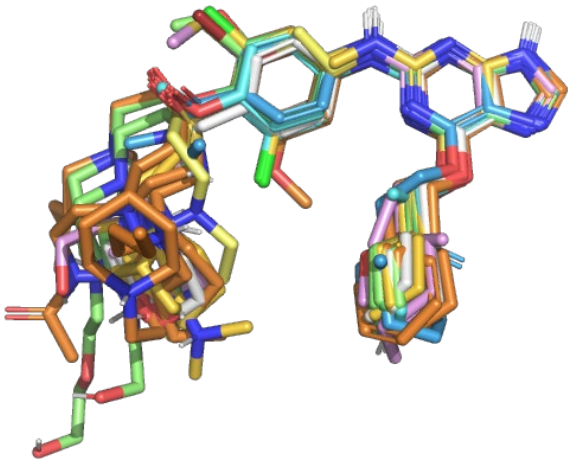
- **Max core molecules:** The maximum number of molecules from the training set to align directly to the supplied alignment reference to form the core alignment set. The remaining molecules will be aligned to the core alignment set.
- **Similarity to references:** The minimum similarity between a training molecule and the alignment references to be selected as part of the core alignment set.
- **Similarity to core:** The similarity threshold for matching training molecules to those selected for the core alignment.

5.3 Reviewing alignment hypotheses

Once the alignment hypotheses are calculated, you will be notified in the panel, and the calculation time will be reported. Click **Next** to view the alignment hypotheses as shown in the image below. You can click through the hypotheses in the Alignment Hypotheses section of the panel. By default, all of the molecules in the training set are selected and displayed in PyMOL. You can select individual molecules, use **Shift-click** /**Ctrl-click** to select multiple molecules, or use **Ctrl-A** to select all.

If there are any spurious alignments, you can mark them for deletion in the **Alignment References** section by clicking the checkbox in the **Discard** column. When you click **Next**, the QuanSA models will be built, with the discarded hypotheses not taken forward to model building.

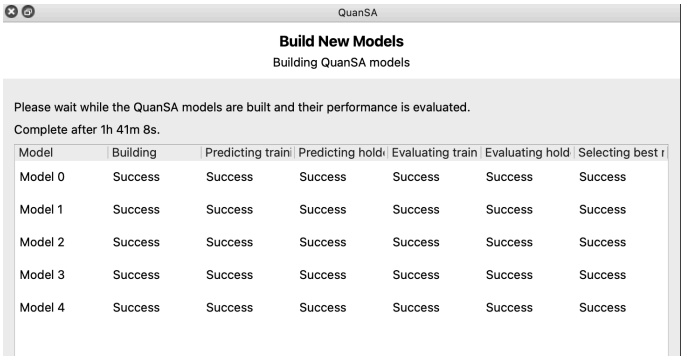




5.4 Reviewing QuanSA models

Depending on the number of hypotheses and the number of molecules in your project, the QuanSA model induction can take a few hours on an average consumer-grade laptop. When the models are being calculated, each column will be updated with the status of the calculation: queued, running, success.

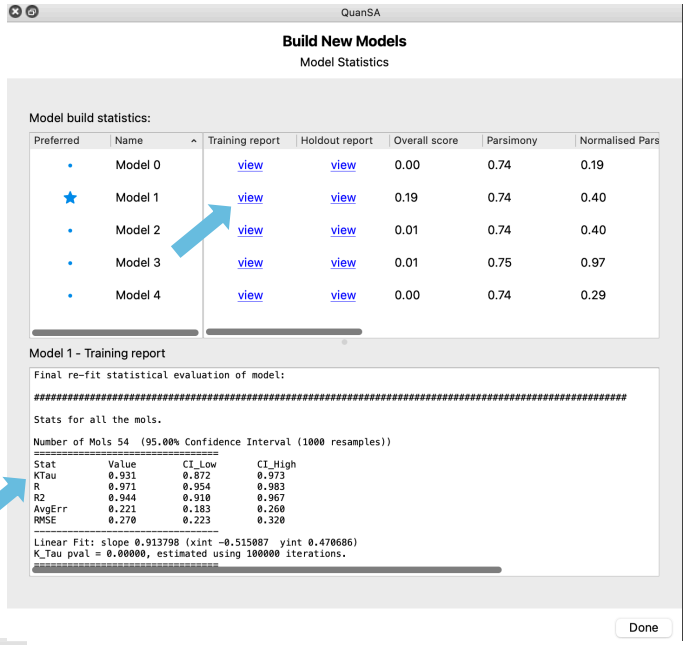
When the process completes, the dialog will reflect the success. Click **Next** to explore the **Model Statistics** dialog.

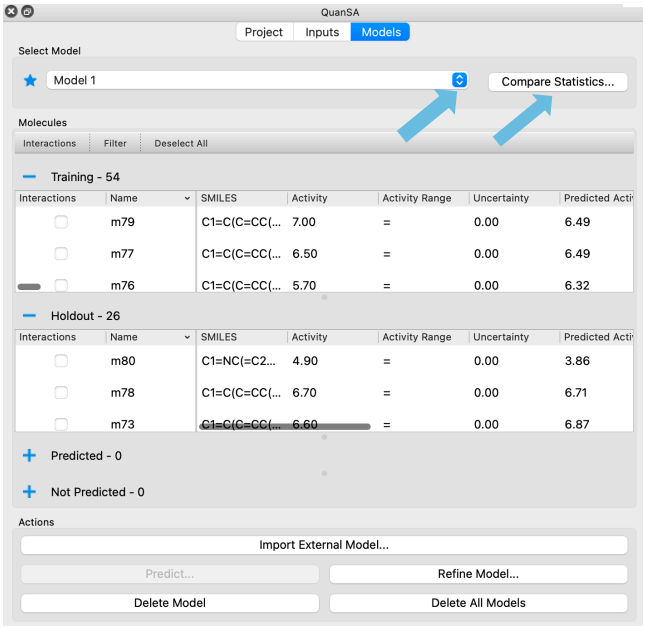


After model building is complete, the model statistics will be loaded into the dialog. The preferred model, based on the overall score, is highlighted with a ★.

The **Training report** and **Holdout report** can be viewed in the bottom panel to assess model performance. For a detailed description of various metrics (parsimony, exclusion, novelty and confidence), please refer to the Surflex Manual.

Once you have analysed the reports, click **Done**.





The **Models** tab is now populated with the QuanSA models where you can explore the models' performance on the training and holdout sets and visually assess which properties contribute most to the activity and leverage this information to guide further molecular design and optimisation (**Interactions**) as shown in section 6.

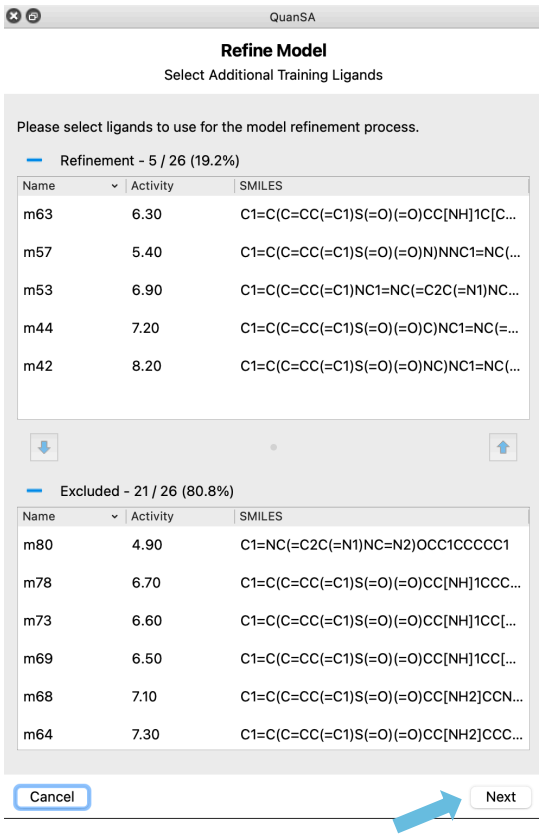
You can choose from the models from the **Select Model drop-down list**. Clicking **Compare Statistics** will return your view to the **Model Statistics** panel as shown above.

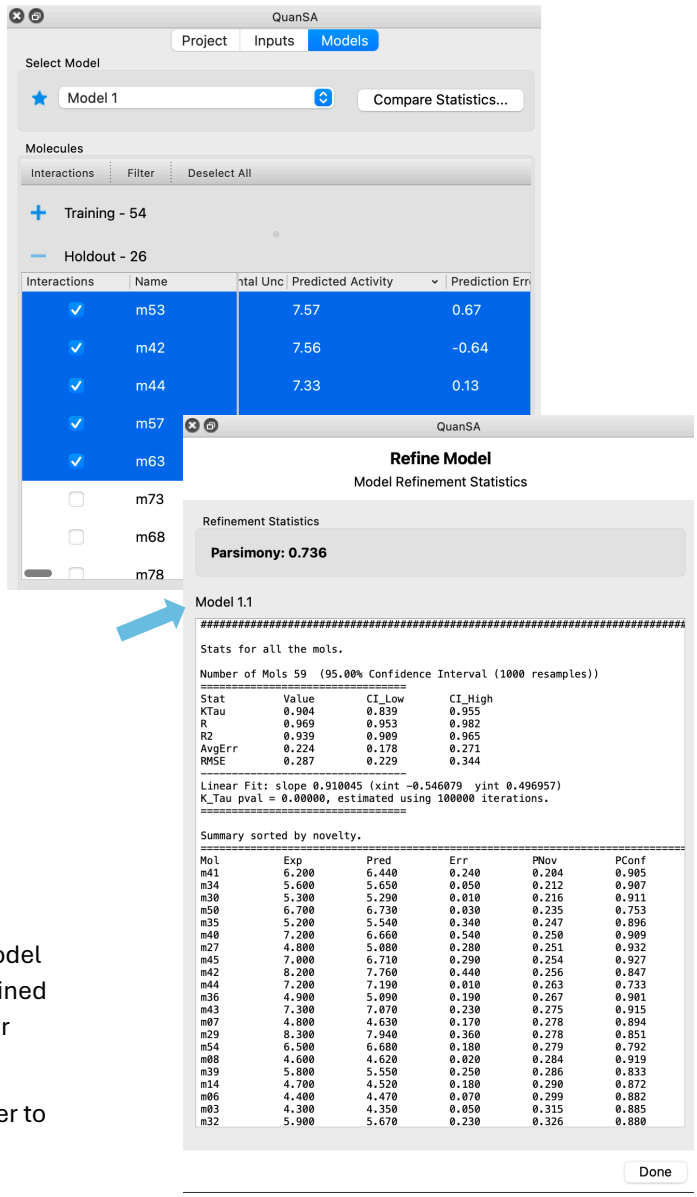
5.5 Refining a model

QuanSA models can be refined within the plugin. To refine a model, first select the **model** to refine from the drop-down list and then click on **Refine Model**.

In the next dialog, you can choose the molecules with which to refine the model by selecting them from other imported molecules. In the image, the selected molecules are in the holdout set.

Clicking **Next** starts the model refinement.





When the refinement completes, the model name is appended with a digit (e.g., Model 1.1) to reflect the model that was refined and the subsequence refinement. Refined model statistics can be viewed and compared to earlier models.

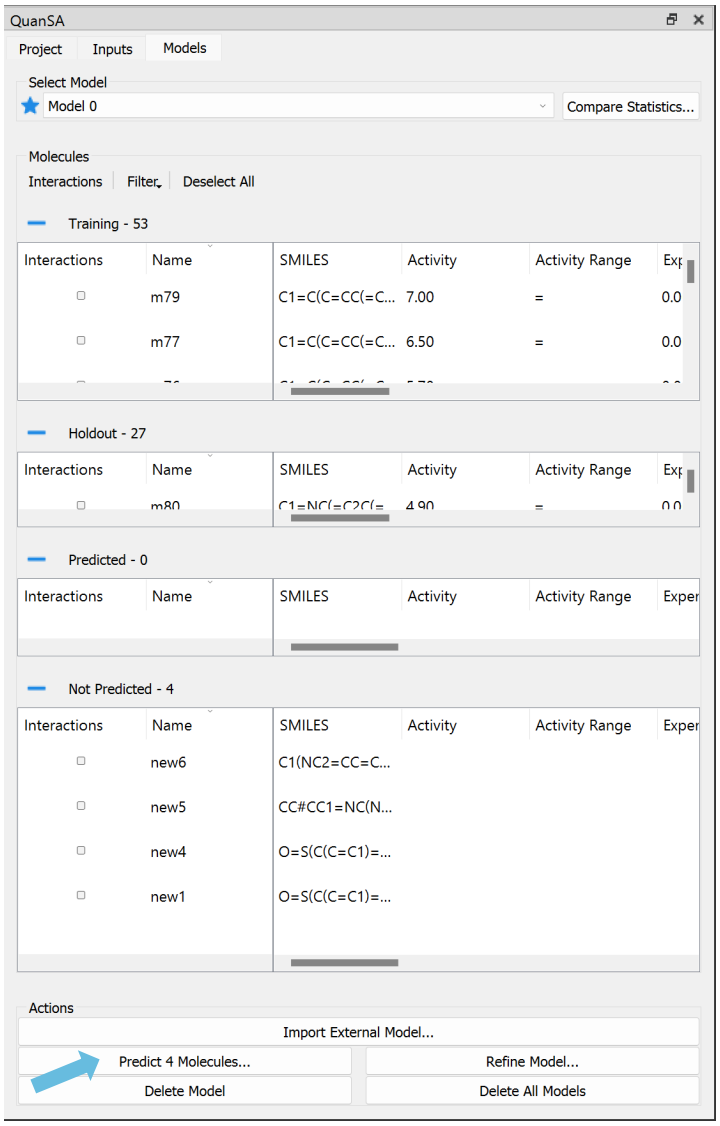
For details on the importance of parsimony, please refer to the Surflex Manual.

6. How do I... view model results

6.1 Predicting affinity for new compounds

To use a model to predict affinities for new compounds, return to the Inputs tab and repeat the import process described in Section 4.1. This time, as there is no activity data available, the molecules will have conformations generated. Return to the **Models** tab.

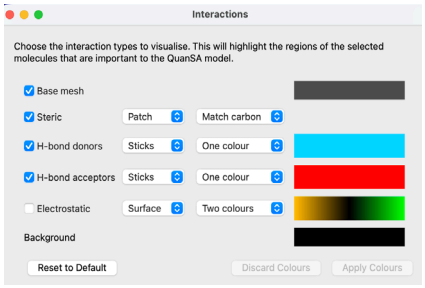
To predict activities, click the **Predict X Molecules** button. Alternatively, **select** one more of the unpredicted compounds then click **Predict**. You will be asked to confirm the calculation in a follow up dialog: click **Next**. As predictions are made, the molecules will move from the **Not Predicted** panel to the **Predicted panel**.



6.2 Displaying 3D markers

A unique and powerful feature of QuanSA is that you can visually assess which properties contribute most to the activity and leverage this information to guide further molecular design and optimisation.

Select a molecule in the **Models** tab to show it in PyMOL. Click the **Interactions** button to launch the Interactions dialog, in which you can choose how to display mode information.



In the **Interactions** dialog, you can choose to show a surface mesh on the structure, as well as a variety of options to display and colour steric, H-bond donor, H-bond acceptor, and electrostatic contributions to the QuanSA activity prediction.

Use the checkboxes to control the visibility of the mesh and interactions.

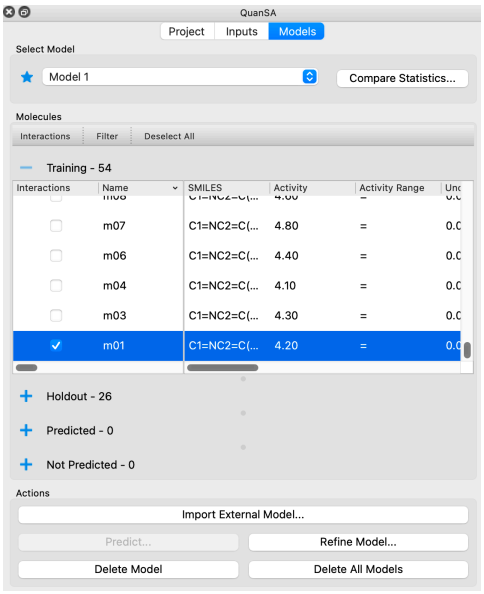
Three display options are available: surface, sticks, and patches.

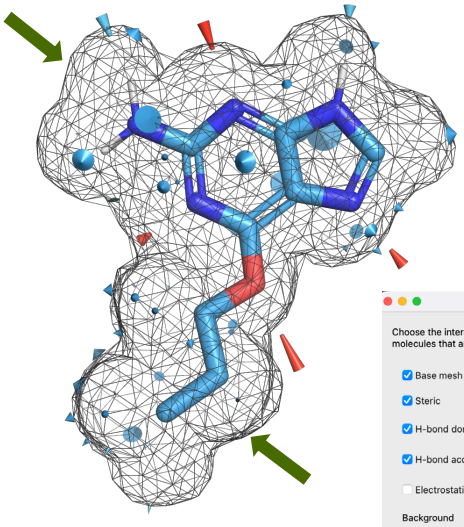
When Surface is chosen, the contribution and its extension is mapped onto the mesh surface. When contributions are shown as sticks, the length of the stick is proportionate to the importance of the contribution. When contributions are shown as patches, a patch is displayed in which the diameter of the patch corresponds to the importance of the interaction.

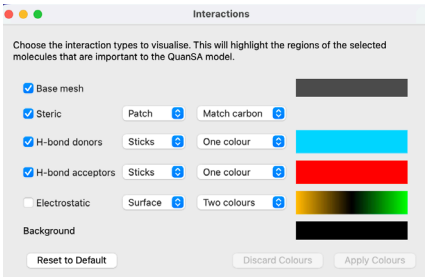
You can also control the colouring of the interactions, by choosing to match the interaction to carbon colour, use a single colour (e.g., H-bonds and H-acceptors) or two colours (e.g., electrostatic). When two colours are used to display electrostatic contribution, the yellow refers to favourable and the green refers to unfavourable electrostatics contributions.

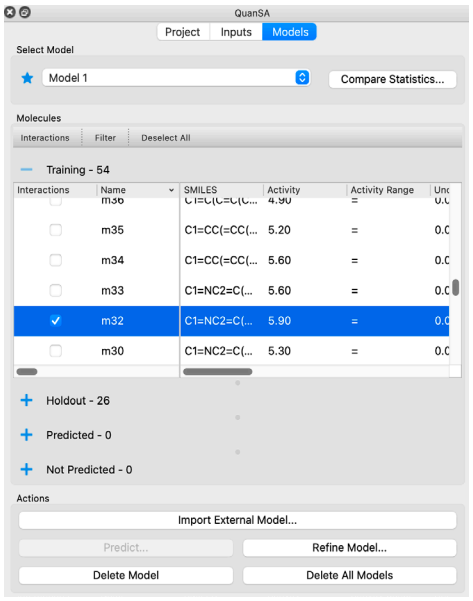
The colours for the interactions and the background in PyMOL can be changed by clicking on the coloured bars and choosing a new shade.

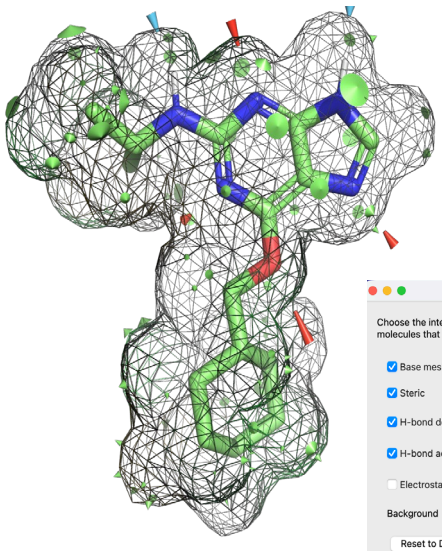
Two examples of visualised contributions are shown below.

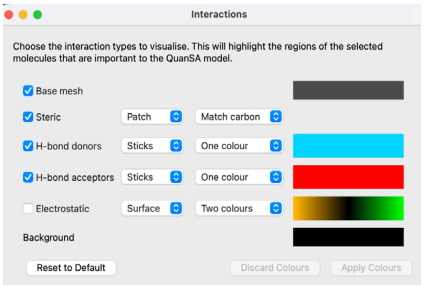

QuanSA
Project
Inputs
Models
Select Model
Model 1
Compare Statistics...
Molecules
Interactions
Filter
Deselect All
Training - 54
Interactions
Name
SMILES
Activity
Activity Range
m07
C1=NC2=C(...
4.80
=
m06
C1=NC2=C(...
4.40
=
m04
C1=NC2=C(...
4.10
=
m03
C1=NC2=C(...
4.30
=
m01
C1=NC2=C(...
4.20
=
Holdout - 26
Predicted - 0
Not Predicted - 0
Actions
Import External Model...
Predict...
Refine Model...
Delete Model
Delete All Models




Interactions
Choose the interaction types to visualise. This will highlight the regions of the selected molecules that are important to the QuanSA model.
Base mesh
Steric
Patch
Match carbon
H-bond donors
Sticks
One colour
H-bond acceptors
Sticks
One colour
Electrostatic
Surface
Two colours
Background
Reset to Default
Discard Colours
Apply Colours


QuanSA
Project
Inputs
Models
Select Model
Model 1
Compare Statistics...
Molecules
Interactions
Filter
Deselect All
Training - 54
Interactions
Name
SMILES
Activity
Activity Range
m35
C1=CC(=CC(...
5.20
=
m34
C1=CC(=CC(...
5.60
=
m33
C1=NC2=C(...
5.60
=
m32
C1=NC2=C(...
5.90
=
m30
C1=NC2=C(...
5.30
=
Holdout - 26
Predicted - 0
Not Predicted - 0
Actions
Import External Model...
Predict...
Refine Model...
Delete Model
Delete All Models




Interactions
Choose the interaction types to visualise. This will highlight the regions of the selected molecules that are important to the QuanSA model.
Base mesh
Steric
Patch
Match carbon
H-bond donors
Sticks
One colour
H-bond acceptors
Sticks
One colour
Electrostatic
Surface
Two colours
Background
Reset to Default
Discard Colours
Apply Colours

7. How do I... troubleshoot problems

7.1 Installation problems

7.1.1 The QuanSA plugin is unable to connect to my Surflex executable

Your Surflex executable should be saved locally on the machine and not on a cloud-based drive (e.g. OneDrive). For installation instructions, see section 2. If you have any questions, please contact support@optibrium.com.

7.1.2 The QuanSA plugin doesn't start, showing the error message "An error occurred when trying to initialise the plugin"

If the plugin fails to start, the primary reason is that an unsupported version of Python is used. This can happen if you update PyMOL using the option **Help->Check for Updates**, but PyMOL starts using an unsupported Python version. Use the following commands to check the Python version PyMOL is using:

```
import sys
print(sys.version)
```

7.2 PyMOL crashes

PyMOL crashes are rarely observed when using the QuanSA plugin. We have done our best to investigate and mitigate these events; because the crash arises from the underlying PyMOL code and not the plugin itself, it is not always straightforward to diagnose or resolve. If you experience crashes, please send the crash report and log files for the PyMOL session to support@optibrium.com and we will be happy to investigate.

7.3 Saving and restoring a PyMOL session

It is not recommended to alter the visualisations in PyMOL while using the plugin; the plugin will override any visualisations made in PyMOL. After running all calculations, you can select or pin the structures of interest in the plugin to load them in the PyMOL visualisation state. Note that only the structures that are selected or pinned in the plugin will be available to view in the PyMOL interface. After selecting all structures of interest, you can close the QuanSA plugin and customise the visualisation using the selection and display features in PyMOL.

Any custom visualisations made in PyMOL will not be saved as part of the QuanSA project. It is recommended to only use the QuanSA visualisation tools while performing calculations. When you have completed the calculations, you can close the plugin and use PyMOL to make other visualisations. You can save and open PyMOL sessions under the **File** menu in PyMOL.

You can also create visualisations in PyMOL (or another favourite visualisation program) by loading the output files from the QuanSA plugin calculations.