Identifying Interaction Hot Spots with SuperStar

(SUP-001)

2025.1 CSD Release
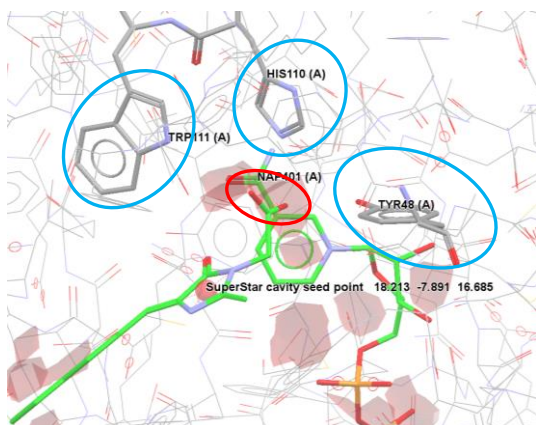

CCDC
advancing structural science

Table of Contents

Introduction

SuperStar, a component of the CSD-Discovery Suite, is a program for identifying regions within a protein binding site or around a small molecule where particular functional groups (probes) are likely to interact favourably. SuperStar uses real experimental information about intermolecular interactions, derived from either the Cambridge Structural Database (CSD) or the Protein Data Bank (PDB). It applies a fully knowledge-based approach to identify protein-ligand interaction hotspots.

Before beginning this workshop, ensure that you have a registered copy of CSD-Core or above installed on your computer

Learning Outcomes

In this workshop you will learn how to use SuperStar to visually identify the most likely interaction hotspots within a protein's binding pocket. At the end of this workshop, you should be able to:

- Generate SuperStar maps
- Create and analyse propensity maps

This workshop will take approximately **30** minutes to be completed. The words in *Blue Italic* in the text are reported in the Glossary at the end of this handout.

Pre-required skills

There are no pre-required skills, however basic familiarity with Hermes would be advantageous. A summary of the Hermes interface is given at the end of this document.

Materials

For this workshop we will use the files that have been shared by the tutors or that you can download from the `workshop` folder.
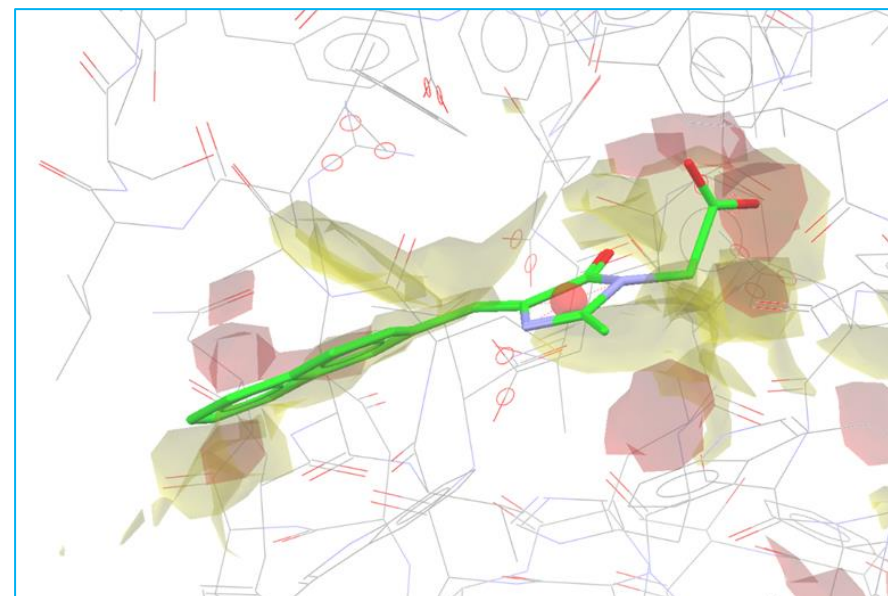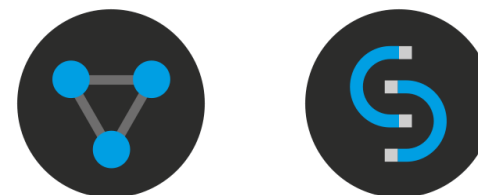
Exploring interactions of aldose reductase inhibitors using SuperStar

Introduction

Aldose reductase (ALR2) is an NADPH-dependent enzyme that catalyses the reduction of D-glucose to D-sorbitol using NADPH as reductant in the polyol pathway. Under normal glycemic conditions, only a small fraction of glucose is metabolised through the polyol pathway, as the majority is metabolised via the tricarboxylic acid cycle in the glycolytic system. However, in response to the chronic hyperglycaemia found in diabetes, glucose flows through the polyol pathway where ALR2 is activated to consume the flooded glucose. Therefore, high levels of D-sorbitol are formed intracellularly, and a large amount of NADPH is consumed. This results in an osmotic pressure imbalance that causes some diabetic complications such as cataract and nephropathy. Hence the inhibition of ALR2 activity may offer a promising option for the alleviation or prevention of complications and symptoms associated with chronic hyperglycaemia. Many ALR2 inhibitors have been developed as drug candidates; virtually all have failed, although some such as epalrestat are commercially available in several countries.

A recent study published by R. Saito *et al.* (*Eur. J. Med. Chem.* 2017. **125**, 965-974, DOI: 10.1016/j.ejmech.2016.10.016) proposes a new class of highly potent ALR2 inhibitors, which are related to the fluorescent chromophore of the *Aequorea* green fluorescent protein (GFP). By using docking experiments and SuperStar maps, Saito *et al.* determined the important structural features needed for the GFP chromophore model series to exhibit high ALR2 inhibitory activity (Figure 1).

In this workshop, we will replicate this work, demonstrating how Saito *et al.* used SuperStar maps to efficiently identify protein-ligand interaction hotspots that can be linked to the inhibitor activity of the compounds investigated.

Provided input files in `workshop`:

- *4IGS_protein.mol2,* the coordinates of the ALR2 protein derived from the co-crystal structure of ALR2 in complex with JF0064 inhibitor (2,2',3,3',5,5',6,6'-
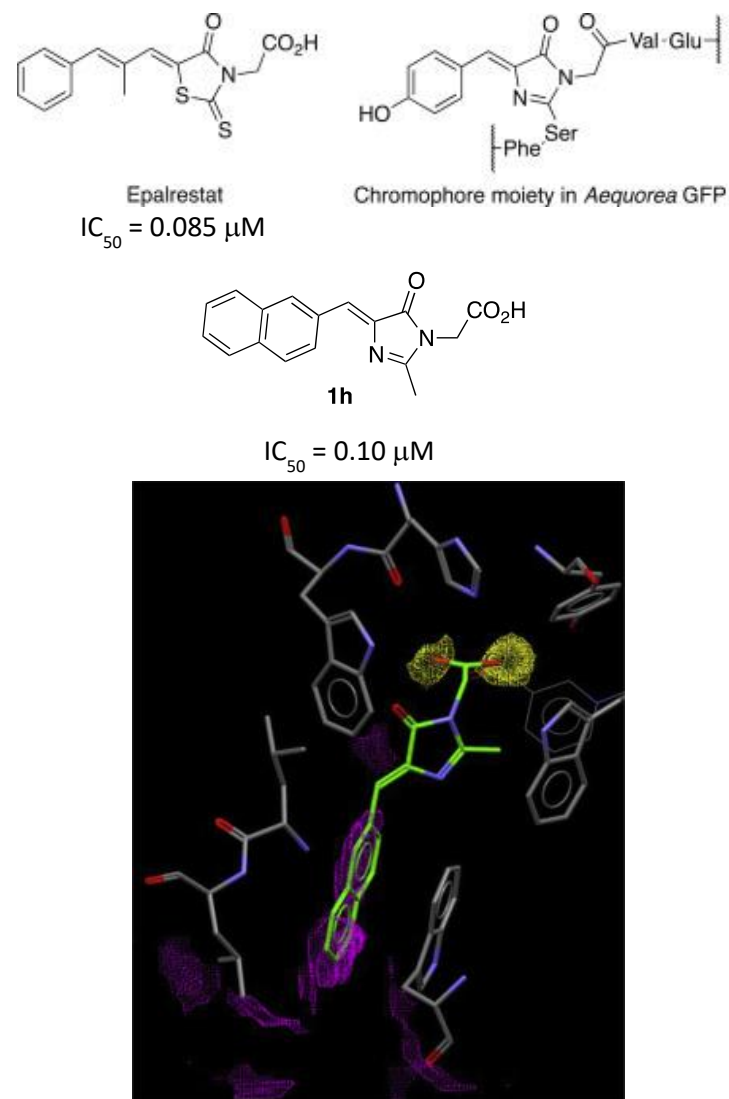

Epalrestat
IC$_{50}$ = 0.085 μM

Chromophore moiety in *Aequorea* GFP

1h

IC$_{50}$ = 0.10 μM

Figure 1. *Structures of epalrestat, the chromophore moiety in the Aequorea green fluorescent protein (GFP) and the GFP chromophore model 1h. PDB-based SuperStar propensity map of the 4IGS binding site using the aliphatic CH (magenta) and carbonyl oxygen (yellow) probes, accompanied with the predicted docking pose 1h.*

octafluoro-4,4'-biphenyldiol) (PDB code: 4igs). The *4IGS_protein.mol2* includes only one (*A*) of the alternative conformations listed in the original *4igs.pdb* and has been set up in accordance with the guidelines for the preparation of input files (see the GOLD user manual, for Setting Up the Protein(s) and Setting Up Ligands). Note that HIS100 was set up as the tautomer where the *NE2* atom of the imidazole ring is protonated. Note that this structure includes the NADP+ cofactor (with ligand ID NAP).

- *compound_1h.mol2,* the predicted docking pose of **1h** compound, a new highly potent ALR2 inhibitor related to the fluorescent chromophore of the *Aequorea* green fluorescent protein (GFP). The predicted docking pose of **1h** compound within the ALR2 protein binding site derived from *4igs.pdb* has been downloaded from the Supplementary data of Saito *et al.* (*Eur. J. Med. Chem.* 2017. **125**, 965-974, DOI: 10.1016/j.ejmech.2016.10.016). The compound **1h** has been extracted from the complex and has been set up in accordance with the guidelines for the preparation of input files (see the GOLD user manual, for Setting Up the Protein(s) and Setting Up Ligands).

The example used here mimics the situation where a researcher has a crystal structure of a protein and wants to explore the binding site of that protein to determine the key features required for a ligand to strongly and selectively bind that protein.

Generate SuperStar Maps with the Hermes GUI

Generating Propensity Maps using the Carbonyl Oxygen Probe

1. Open Hermes and click on **File** and select **Open…** from the pull-down menu to load both *4IGS_protein.mol2* and *compound_1h.mol2.* The loaded molecules will be displayed in the Hermes 3D view and listed in the *Molecule Explorer* window in the left-hand panel.

2. Click on **Calculate** from the Hermes top-level menu and select **SuperStar** from the drop-down menu

3. Choose *4IGS* in the *Use Protein* pop-up window and click **OK.** This will open the *SuperStar* set-up window.

4. In the *SuperStar* window, click on **Settings**. In the *Ligsite Cavity Detection* dialogue, choose *Centroid* from the *Grow cavity from* drop-down menu.
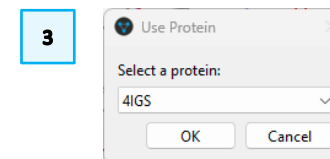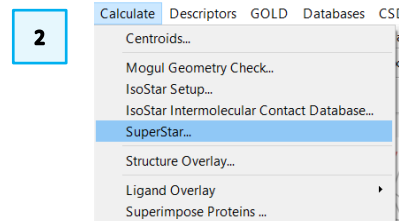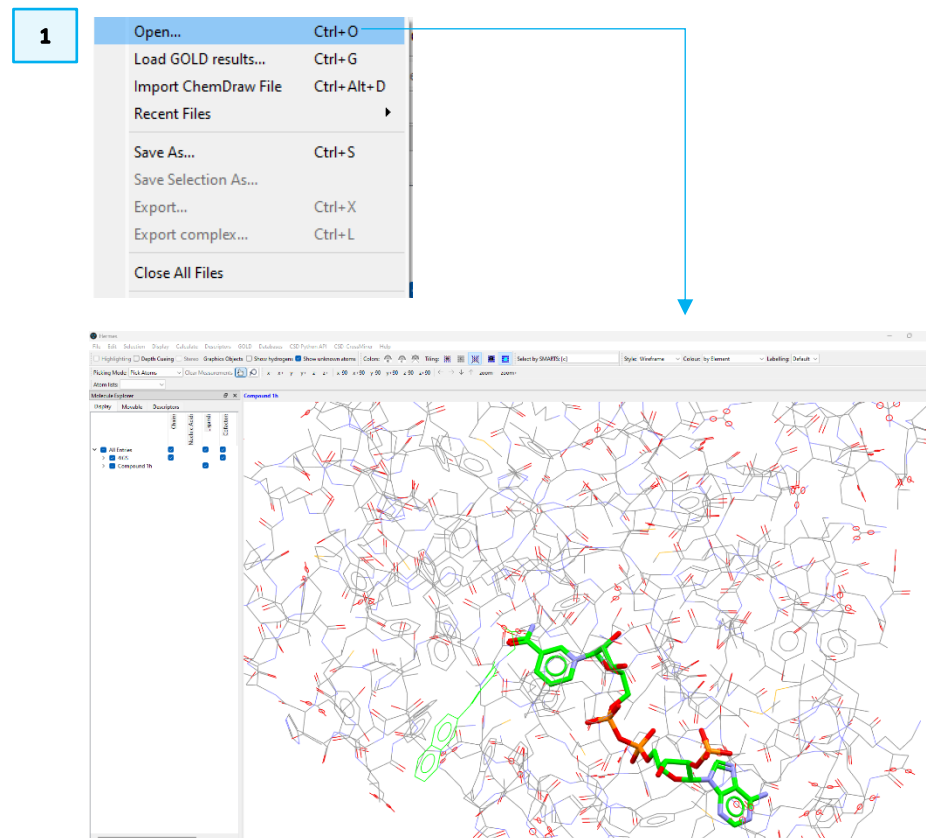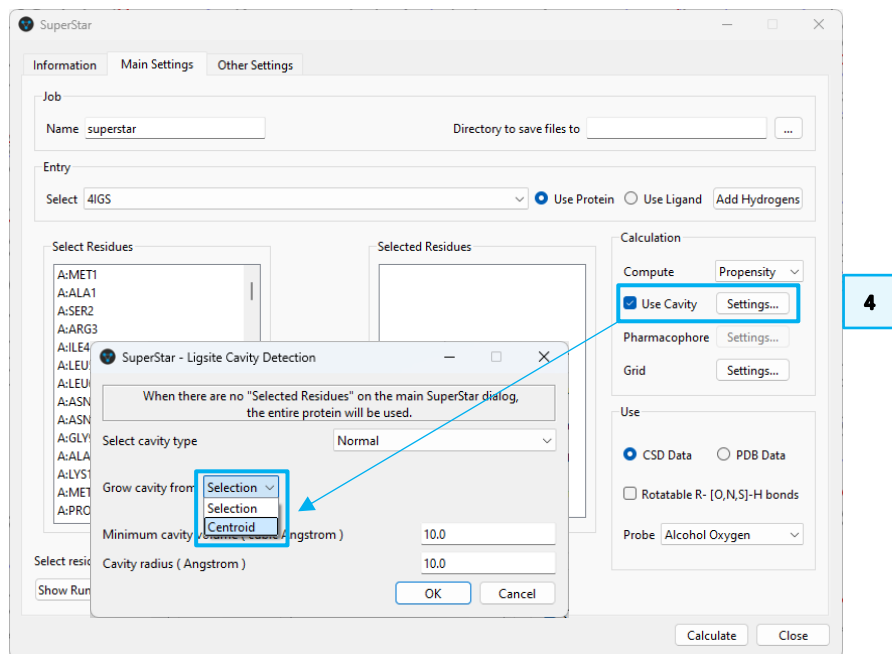
When working with protein structures, cavity detection can be used to determine the extent of the _binding site_. All residues adjacent to the cavity found will then be used for the subsequent map calculation. In this case we will define the binding site starting from a point.

5. Select the atoms of the imidazole ring of compound **1h** in the visualiser. The centroid will be displayed in the visualiser and its XYZ coordinates will be automatically entered to the right of the _Grow cavity from_ option (18.2126, -78912, 16.6852).

Note that you can select different cavity types. _Normal_ is the default. Options are: _Shallow, Shallow/Normal, Normal, Normal/Buried and Buried_ (See the SuperStar User Guide for more details).

Leave the cavity type as _Normal_ for the purposes of this workshop. Also, the cavity size can be adjusted by setting the **Cavity radius** (Å) and the **Minimum cavity volume** (Å³) parameters. Leave the default parameters here. Click OK to close the _Ligsite Cavity Detection_ dialogue.

6. Be sure that **Propensity** is selected next to _Compute_ in the _Calculation_ section of the _SuperStar_ window.
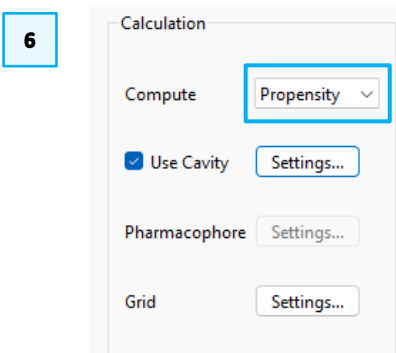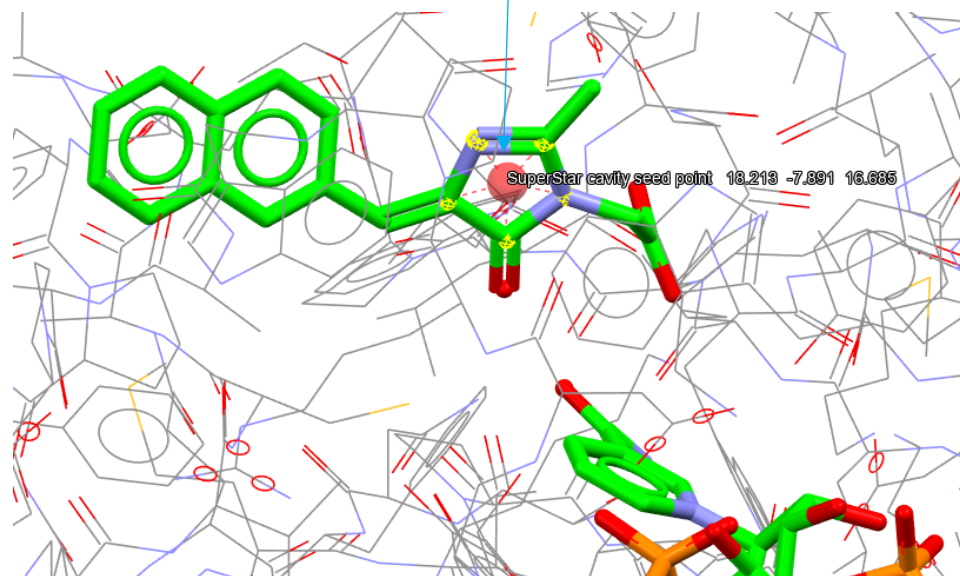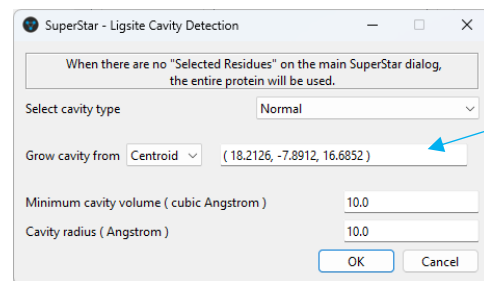
Propensities are measured relative to the expected (random) chance of finding a group at a certain position, e.g. a propensity of 4 indicates that the chance of finding the probe group at that point is 4 times as high as random. As a general rule, propensities of 2 and above indicate favourable interaction sites (although for *hydrophobic probes*, values between 1 and 2 can also be meaningful).

7. Select the output directory, e.g. `SuperStar_workshop`.

8. SuperStar allows data from either the Cambridge Structural Database (CSD) or the Protein Data Bank (PDB) to be used for propensity map calculation. The *Data source* to be used for the map calculation can be specified within the *Use* section of the *SuperStar* window. Enable the **PDB Data** button to retrieve PDB-based scatterplot data for the compilation of the SuperStar map.

9. For protein residues, SuperStar can take into account the fact that OH, NH and SH groups may rotate. To enable this feature, tick the **Rotatable R-[O,N,S]-H bonds** tick box in the *Use* section of the *SuperStar* window. This feature applies to serine, threonine, tyrosine, lysine and cysteine residues. If this tick box is not enabled, the -XH bond for the above residues will be treated as being fixed in space.

Before a SuperStar map can be calculated, a probe needs to be selected. Contouring will subsequently take place for one of the atoms of the chosen probe. The list of available probes for CSD and PDB data are listed in the **Probe** drop-down menu within the *Use* section of *SuperStar* window.

10. Select **Carbonyl Oxygen** as the probe atom from the **Probe** drop-down menu. The carbonyl oxygen can accept *hydrogen bonds*, so favourable spots for *acceptor groups* in the ligand will be detected.

11. Click **Calculate.** The resulting PDB-based map will be displayed as a 3D contoured map in the Hermes 3D view.
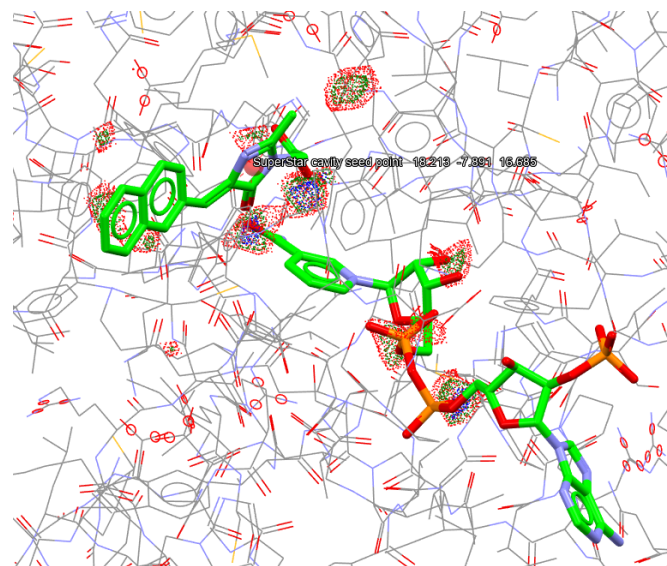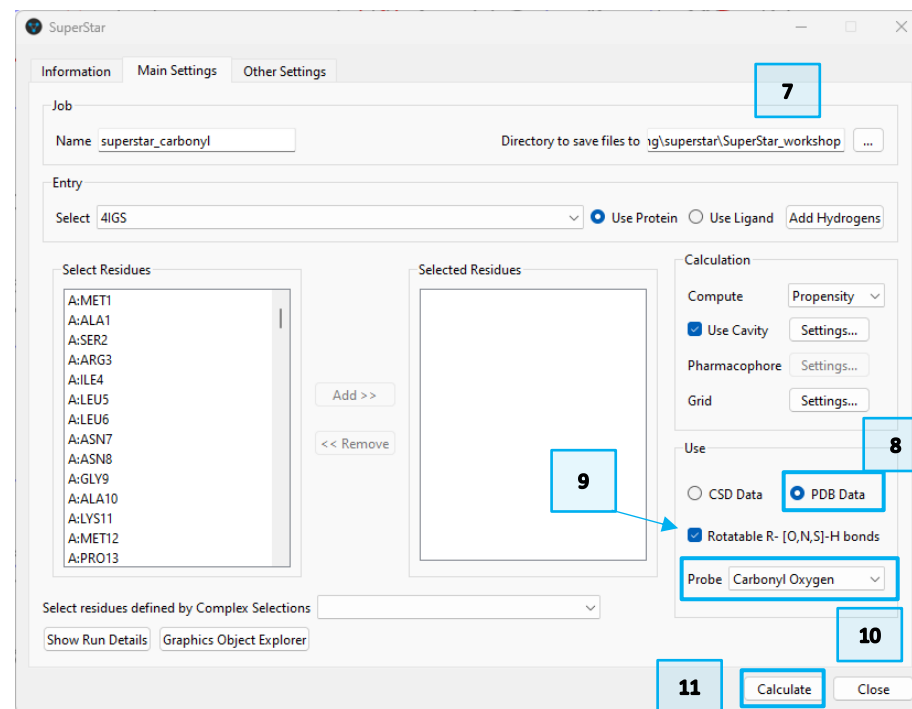
Once the SuperStar maps have been created, their display can be controlled in the Hermes *Graphics Objects Explorer*. This will be opened the first time a map is calculated. It can also be opened by clicking on the **Graphics Object Explorer** button at the base of the *SuperStar* window. In addition, it can be opened from the **Display** drop-down menu in the top-level Hermes menu bar.
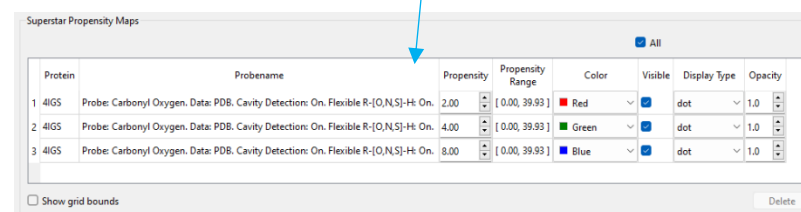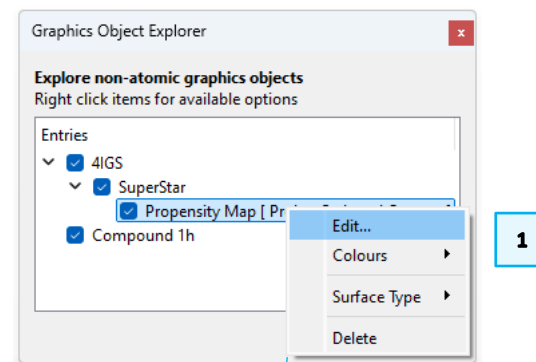
Customising Carbonyl Oxygen Propensity Maps

1. Right-click on **Propensity Map [Probe: Carbonyl Oxygen]**, in the *Graphics Objects Explorer* window and then choose **Edit**. This will open the *SuperStar Propensity Maps* window. From here, for each contour surface, it is possible to edit the propensity level, colour, display type, opacity and whether each contour surface is visible in Hermes or not.
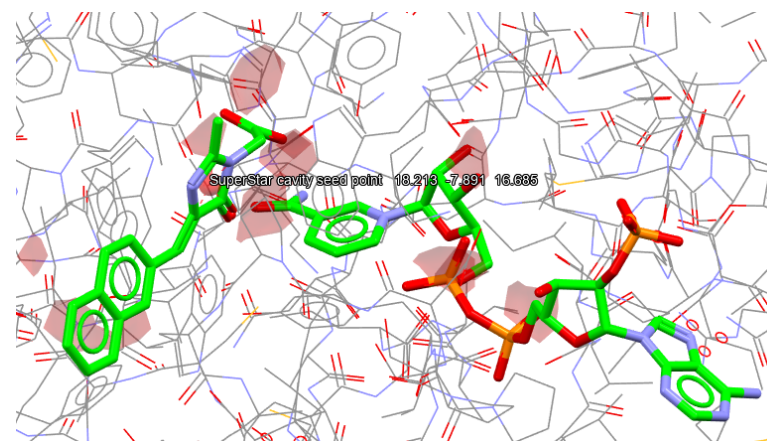
Three propensity maps are listed, corresponding to the contoured maps at levels 2.0 (red) and 4.0 (green) and 8.0 (blue). These SuperStar contour maps depict the spatial distribution of propensities.

2. From the *SuperStar Propensity Maps* window, turn-off the second and third propensity maps (corresponding to **Propensity** of *4.0* and *8.0*) by disabling their tickboxes in the **Visible** column.

3. Customise the first propensity map, corresponding to the map with **Propensity** of *2.0*.

   - Change its **Display Type**. By clicking on *dot* and select *triangle* from the drop-down menu.

   - Change the opacity level from *1.0* to *0.2* using the **Opacity** spin-box.

This will result in the Hermes 3D view only showing the map of propensity 2 indicating that the chance of finding the carbonyl oxygen probe group at that point is twice as high as random.

|	Protein	Probename	Propensity	Propensity Range	Color		Visible	Display Type	Opacity
1	4IGS	Probe: Carbonyl Oxygen. Data: ...	2.00	[0.00, 39.93]	■	Red	☑	triangle	0.2
2	4IGS	Probe: Carbonyl Oxygen. Data: ...	4.00	[0.00, 39.93]	■	Green	☐	dot	1.0
3	4IGS	Probe: Carbonyl Oxygen. Data: ...	8.00	[0.00, 39.93]	■	Blue	☐	dot	1.0

Analysing Carbonyl Oxygen Propensity Maps

The propensity map derived by using the carbonyl oxygen probe reveals that the carboxylate group of compound **1h** matches very well in the H-bond acceptor areas predicted by SuperStar (see blue circle in the figure on the right).

1. In the Hermes 3D view, in the *Molecule Explorer* click on the ">" adjacent to 4IGS and underneath *All Entries*. Then click on ">" adjacent to *Chains* and then on ">" adjacent to *A.* This will show the list of all the residues of the protein.

2. Scroll through the list in the *Molecule Explorer* and select TYR48, HIS110 and TRP111. Multi-selection is possible by using **CTRL + left-click.**

3. Right-click in the *Molecule Explorer* area and select **Style** > **Capped Sticks** from the pull-down menu. This will change the style of the three selected residues to capped sticks in the Hermes 3D view.

As highlighted with blue circles in the figure on the right, the predicted (docked) pose of compound **1h** in the protein binding site is able to form hydrogen bond interactions between the carboxylate group of the **1h** compound (as H-bond acceptors) and the Tyr48/His110/Trp111 sidechains (as H-bond donors).
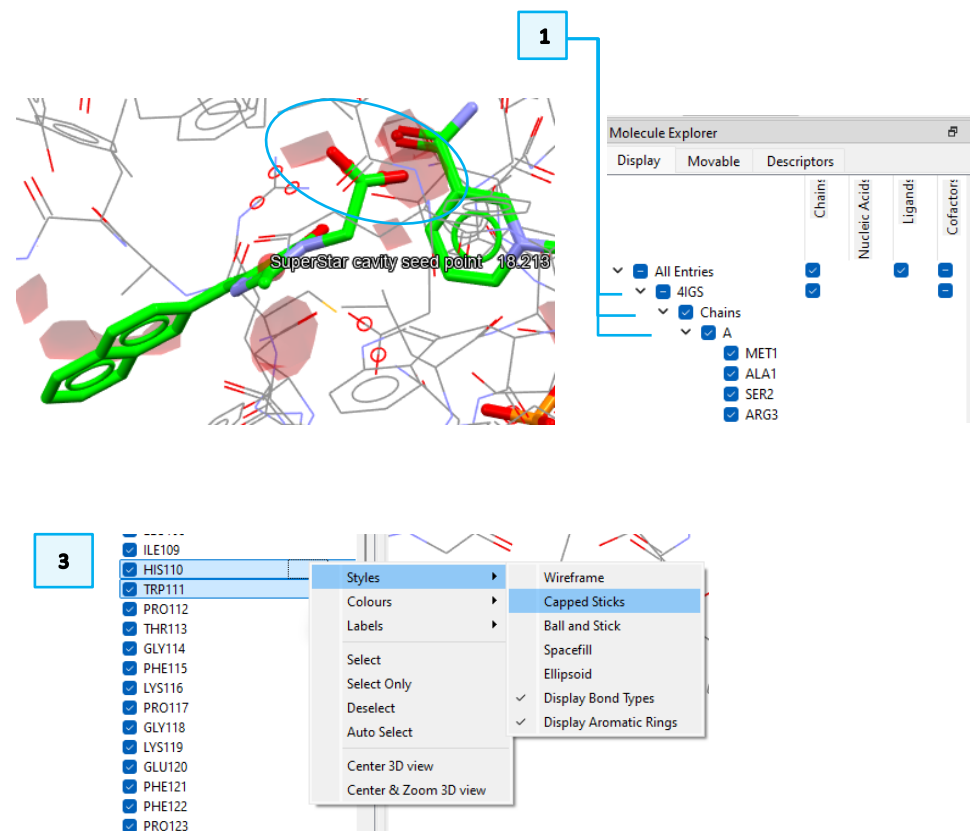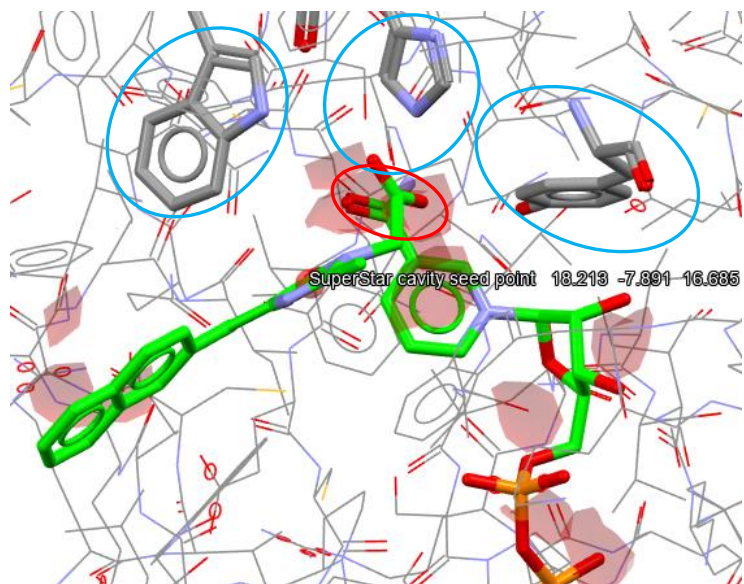
Generating & Customising Propensity Maps using the Aliphatic CH Carbon Probe

We are now going to explore the hydrophobic interactions in the protein binding site. To do so, please return to the *SuperStar* window. If you have closed it, please click on **Calculate** > **SuperStar** from the Hermes top-level menu.

1. Select **Aliphatic CH Carbon** as the probe atom from the **Probe** drop-down menu. The aliphatic CH carbon will detect favourable spots for hydrophobic groups in the ligand.

2. In the *Job* section of the *SuperStar* window, change the name of the job e.g. *superstar_aliphatic*. This will ensure that the previous maps are not overwritten.

3. Click **Calculate.** The resulting PDB-based maps will be displayed as 3D contoured maps in the Hermes 3D view.

4. Hide the Propensity Maps previously obtained using the carbonyl oxygen probe by disabling the **Propensity Map [Probe: Carbonyl Oxygen]** tick box in the *Graphics Object Explorer* window.

5. Right-click on **Propensity Map [Probe: Aliphatic CH Carbon]** and choose **Edit.** This will open the *SuperStar Propensity Maps* window. This time, the six propensity maps calculated up to now are listed, but the maps with the carbonyl oxygen probe have the **Visible** tick box disabled (and are thus not displayed in the Hermes 3D view).
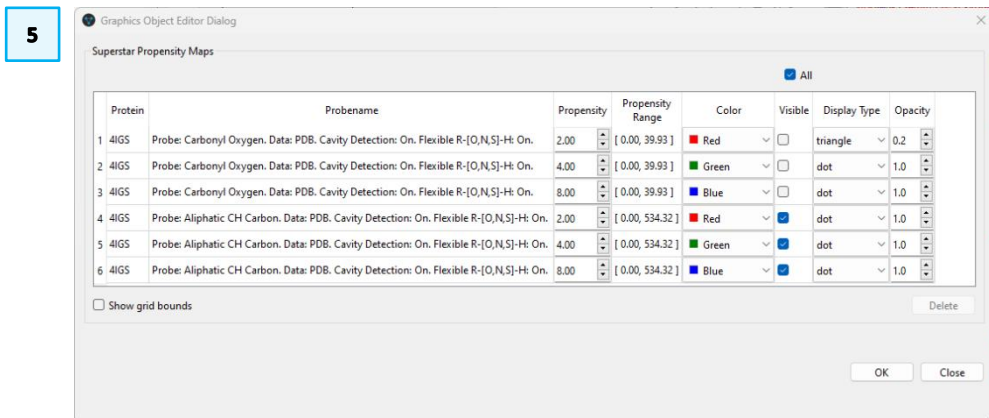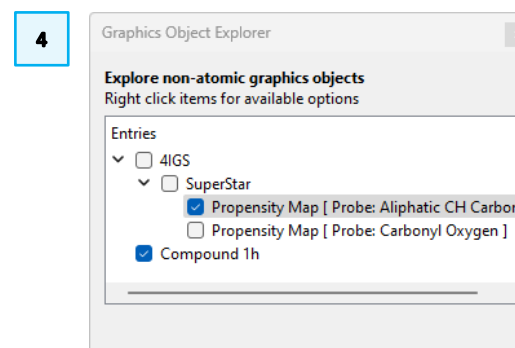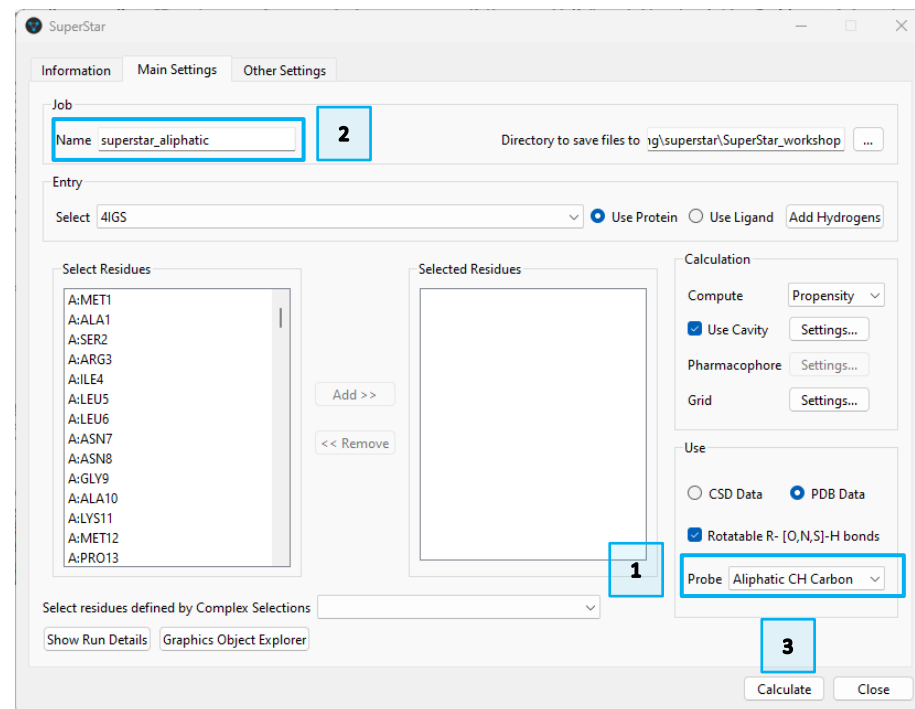
6. In the *SuperStar Propensity Maps* window, turn off the propensity maps with the aliphatic CH carbon with a **Propensity** level of *4.0* and *8.0,* by disabling their tick boxes in the **Visible** column (row 5 and 6 in the figure on the right).

7. Customise the aliphatic CH carbon map with a **Propensity** level of *2.0.*

   - Change its **Display Type** by clicking on *dot* and selecting *triangle* from the drop-down menu.

   - Change the opacity from *1.0* to *0.2* using the **Opacity** spin-box.

   - Change the colour of the contour surface by selecting **Yellow** from the **Color** drop-down menu.

This will result in the Hermes 3D view only showing the map of propensity 2 indicating that the chance of finding the aliphatic CH carbon probe group at that point is twice as high as random.

	Protein	Probename	Propensity	Propensity Range	Color	Visible	Display Type	Opacity
1	4IGS	Probe: Carbonyl Oxygen. Data: PDB. Cavity Detection: On. Flexible R-[O,N,S]-H: On.	2.00	[0.00, 39.93]	Red	☐	triangle	0.2
2	4IGS	Probe: Carbonyl Oxygen. Data: PDB. Cavity Detection: On. Flexible R-[O,N,S]-H: On.	4.00	[0.00, 39.93]	Green	☐	dot	1.0
3	4IGS	Probe: Carbonyl Oxygen. Data: PDB. Cavity Detection: On. Flexible R-[O,N,S]-H: On.	8.00	[0.00, 39.93]	Blue	☐	dot	1.0
4	4IGS	Probe: Aliphatic CH Carbon. Data: PDB. Cavity Detection: On. Flexible R-[O,N,S]-H: On.	2.00	[0.00, 534.32]	Yellow	☑	triangle	0.2
5	4IGS	Probe: Aliphatic CH Carbon. Data: PDB. Cavity Detection: On. Flexible R-[O,N,S]-H: On.	4.00	[0.00, 534.32]	Green	☐	dot	1.0
6	4IGS	Probe: Aliphatic CH Carbon. Data: PDB. Cavity Detection: On. Flexible R-[O,N,S]-H: On.	8.00	[0.00, 534.32]	Blue	☐	dot	1.0
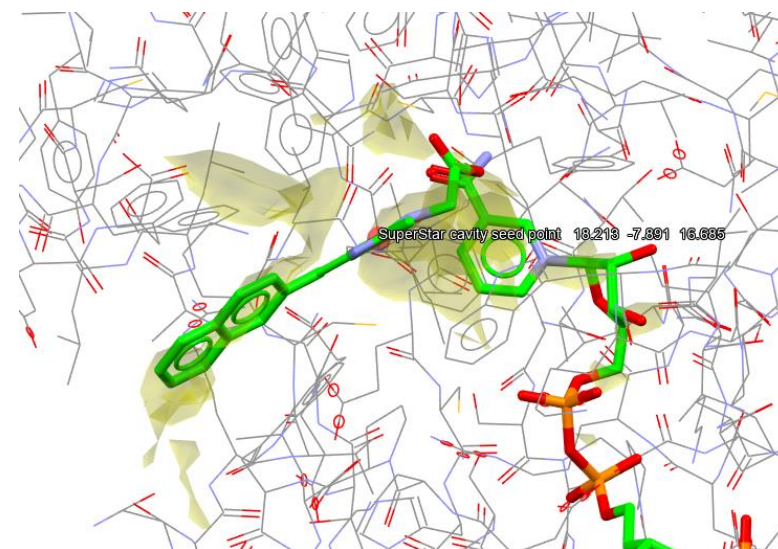
Tip: You can reset the style of all residues by right-clicking on the chain in the molecule explorer.
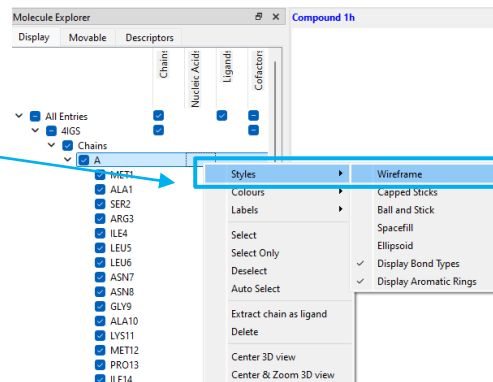
Analysing Aliphatic CH Carbon Propensity Maps

The propensity map derived by using Aliphatic CH probe reveals that the docked pose of compound **1h** has a naphthyl group that matches very well in the hydrophobic areas predicted by SuperStar (see blue circle in the figure on the right).

1. In the Hermes 3D view, in the *Molecule Explorer* click on the ">" adjacent to 4IGS and underneath *All Entries*. Then click on ">" adjacent to *Chains* and then on ">" adjacent to *A*. This will show the list of all the residues of the protein.

2. Scroll through the list in the *Molecule Explorer* and select TRP219, LEU300 and LEU301. Multi-selection is possible by using **CTRL + left-click**.

3. Right-click in the *Molecule Explorer* area and select **Style** > **Capped Sticks** from the pull-down menu. This will change the style of the three selected residues to capped sticks in the Hermes 3D view.

The predicted docked pose of compound **1h** in the protein binding site forms a nice *aromatic face-to-face (π-π) interaction* between the naphthyl group of the **1h** compound and the indole ring of Trp219. In addition, lipophilic contacts can be formed between the same naphthyl group and the Leu300 and Leu301 residues (see blue circles in the figure on the right).

Tip: You can label the alpha carbon of each amino acid according to the protein residue by right-clicking and choosing Label Alpha Carbons by Protein Residue.
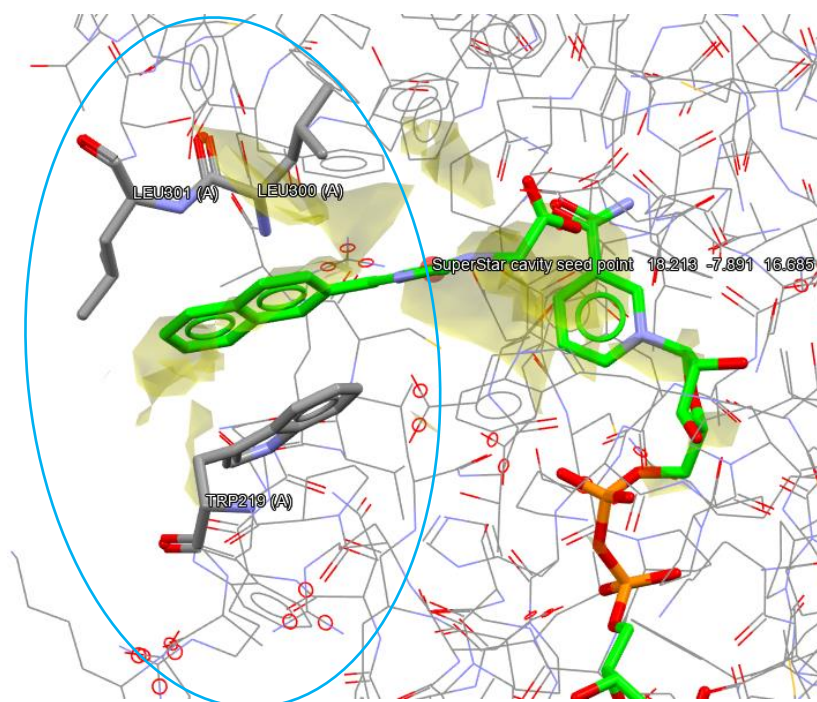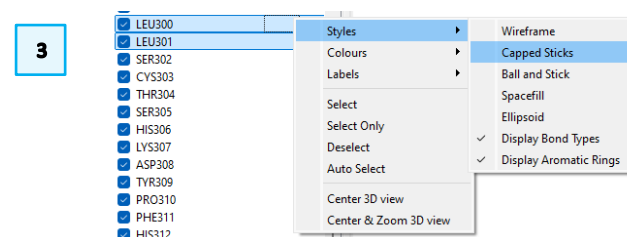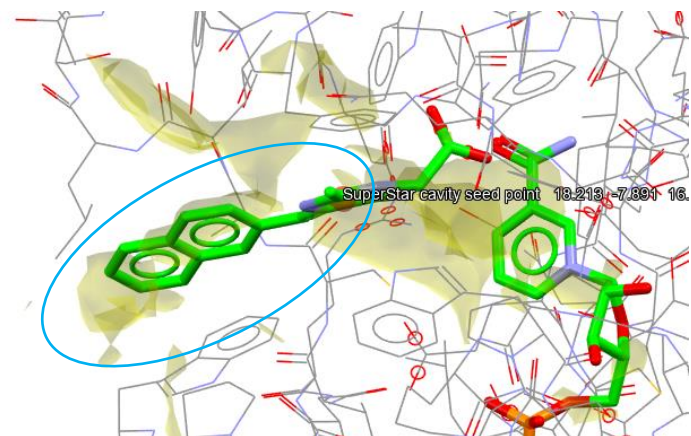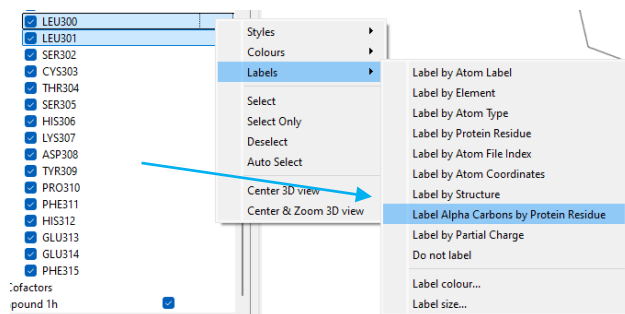
Conclusions

Using SuperStar, we were able to map the regions of preferred H-bond acceptors or preferred hydrophobic groups in the ligand. This analysis validated the binding mode of compound **1h** to the ALR2 receptor as obtained from docking studies and helped explain the variation of activity for a series of potential AR2 inhibitors as an alternative to epalrestat. In particular, SuperStar showed that the activity can be linked to the degree of lipophilic contact between the aryl moiety and the enzyme lipophilic pocket (Trp219/Leu300/Leu301) and confirmed what structural features are important for the GFP chromophore model to be a good ARI candidate and particularly for compound **1h** to be further evaluated *in vivo.*
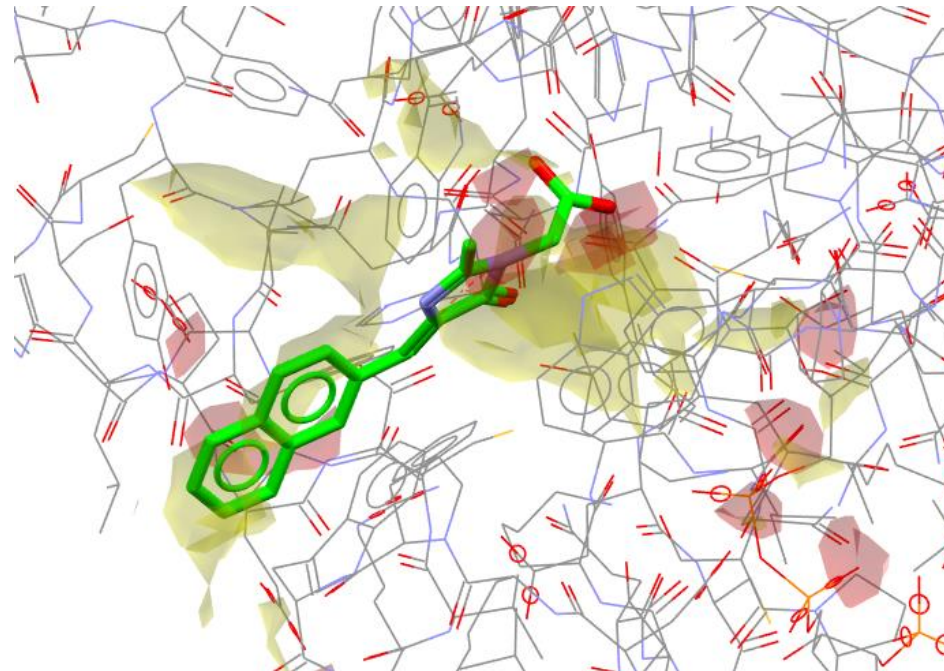
Summary

This workshop introduced how to identify interaction hotspots using SuperStar. You should now be familiar with:

- Generating SuperStar maps
- Creating and analysing propensity maps

For your reference, more details can be found in the SuperStar user guide.

Next Steps

If you have enjoyed this workshop, you might like to explore other workshops from the such as "Understanding Protein-Ligand Interaction Patterns Using Full Interaction Maps" (MER-004), which can be found on the CSD-Discovery self-guided workshop webpages.

https://www.ccdc.cam.ac.uk/community/training-and-learning/workshop-materials/csd-discovery-workshops/

Feedback

We hope this workshop improved your understanding of SuperStar for identifying interaction hotspots and you found it useful for your work. As we aim to continuously improve our training materials, we would love to hear your feedback. Follow the link on the workshop homepage and insert the workshop code, which for this self-guided workshop is *SUP-001*. It will only take 5 minutes and your feedback is anonymous. Thank you!
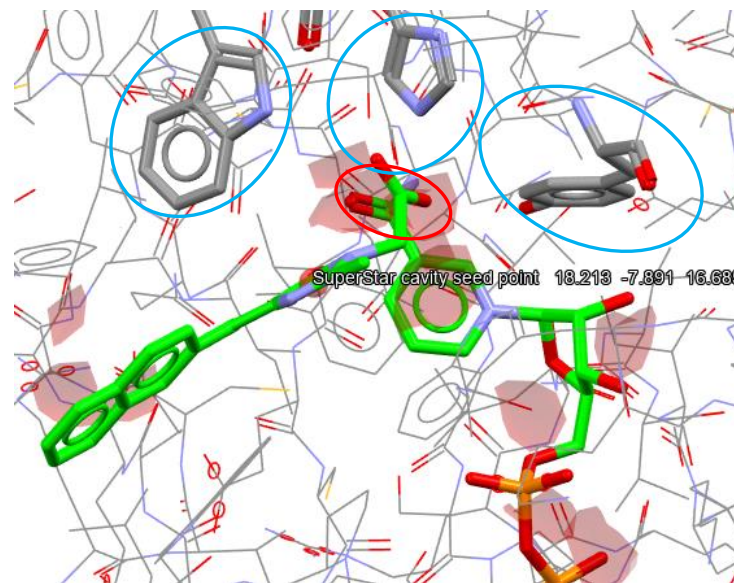
Glossary

Aromatic Interactions are noncovalent interactions formed between aromatic rings. These interactions are important in material science since they will contribute to the overall crystal structure stability. The orientation of the aromatic ring can vary from parallel to T-shape, and we found during our DFT calculations that the T-shape interactions are very close in strength to the parallel displaced ones. Their strength is found between 0 and 16 kJ/mol based on DFT calculations.

A protein displayed as a ribbon with its binding site represented in capped sticks style in green colour.

Binding Site

A specific region in a molecular entity that is capable of entering into a stabilizing interaction with another molecular entity. An example of such an interaction is that of an active site in an enzyme with its substrate. Typical forms of interaction are by hydrogen bonding, coordination and ion pair formation. Two binding sites in different molecular entities are said to be complementary if their interaction is stabilizing. *Source*: PAC, 1994, 66, 1077. (Glossary of terms used in physical organic chemistry (IUPAC Recommendations 1994)) on page 1089.

Centroid

In Hermes, the centroid is the geometric centre of a group of selected atoms.

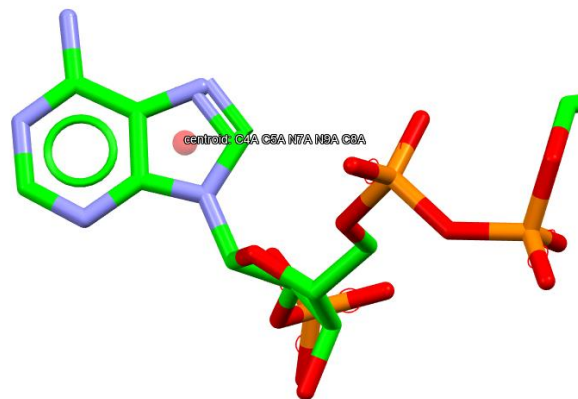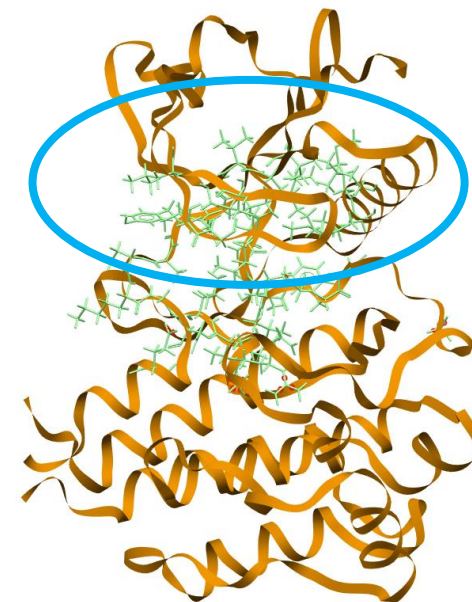*The centroid of the imidazole ring is represented as a translucent red sphere.*

Contour probe

An atom of the contact group from which it is possible to compute a contoured density surface in SuperStar e.g. for carbonyl, the O or C atom are potential contour probes.
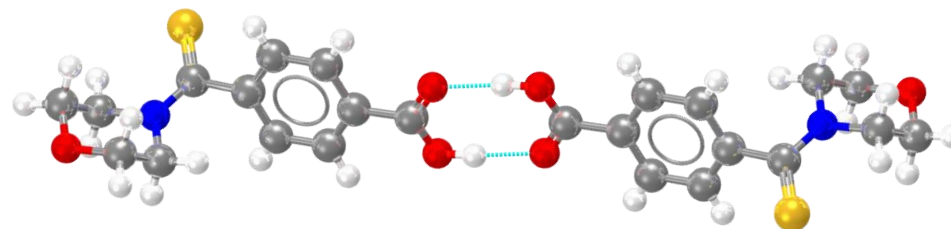
Hydrogen Bonds

Hydrogen bonding occurs between donor-acceptor interactions precisely involving hydrogen atoms. The H-bonds interactions are classified as: strong (mostly covalent), moderate (mostly electrostatic) and weak (electrostatic). Their strength is observed to be between 12 and 30 kJ/mol.

In light blue, example of hydrogen bonds for refcode MULWIC.

Hydrogen Bond Donor/Acceptor

If a typical hydrogen bond is depicted as X—H⋯Y—Z, where the dots denote the bond, X—H represents the hydrogen bond *donor*. The *acceptor* may be an atom or anion Y, or a fragment of a molecule, Y—Z, where Y is bonded to Z. The acceptor is an electron-rich region such as, but not limited to, a lone pair on Y or a π-bonded pair of Y—Z. [Source: E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci and D. J. Nesbitt, *Pure Appl. Chem*., 2011, **83**, 1637 – 1641.]

Illustration of a hydrogen bond interaction with between hydrogen bond donor X–H and hydrogen bond acceptor Y–Z.

Hydrophobic

Hydrophobic molecules effectively "repel" water and thus have a tendency to self-aggregate in aqueous media, excluding water in so doing. On a structural level, these are non-polar groups such as alkyl or aryl moieties.
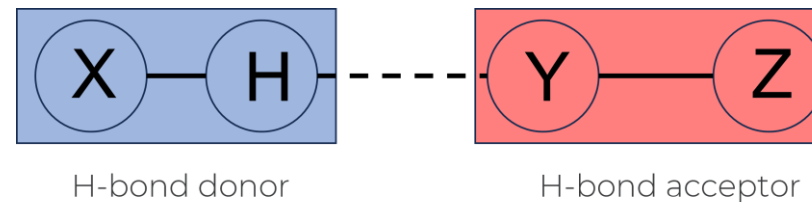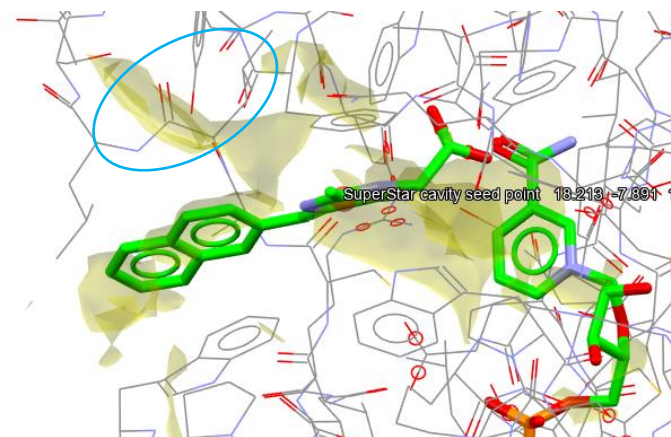

Naphthalene is a non-polar, hydrophobic group. The SuperStar map shows high probability of finding
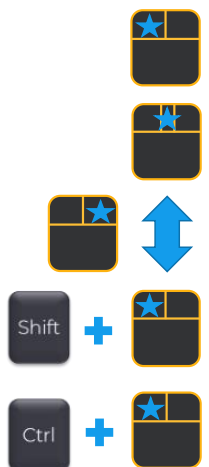
Basics of Hermes Visualisation

Hermes is the CCDC's visualization software to study protein-ligands interaction, and it is the interface for protein docking with GOLD.

In the following we will see some of the basics of navigation and visualization in Hermes that you will find helpful to support your analysis.

In the **Hermes interface** we find:

- **At the top**: list of menus from which we can access visualization and analysis options, and other CSD-Discovery components.
- **On the left-hand side**: the **Molecule Explorer**, where the proteins, ligands, water molecules, and other molecules present in the structures are listed in the *Display* tab. You can expand each entry to see the components, and tick or untick the corresponding boxes to display / not display them. The solutions from a GOLD run will also be displayed in the Molecule Explorer.
- **On the left-hand side**, below the Molecule Explorer: the **Contact Management** box, which allows to display H-bonds and short contacts for selected molecules by ticking the corresponding box.

Using the **mouse to enhance visualization**:

- Left mouse button and move – rotate molecules

- Middle Mouse wheel – move molecules up and down

- Right mouse button and move up and down – zoom in and out of molecules

- Shift + Left mouse button and move - rotate in the plane molecules

- Ctrl + Left mouse button and move - translate molecules