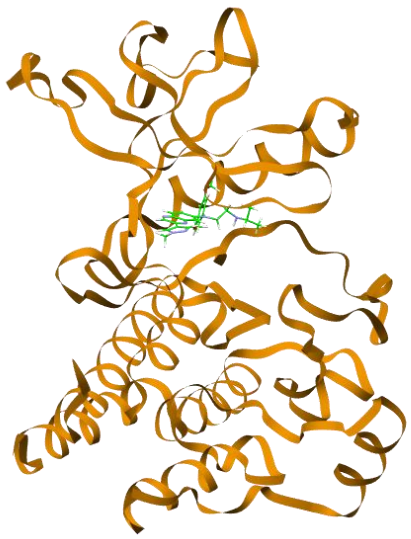
Protein-ligand docking 101 - running a simulation in GOLD Try

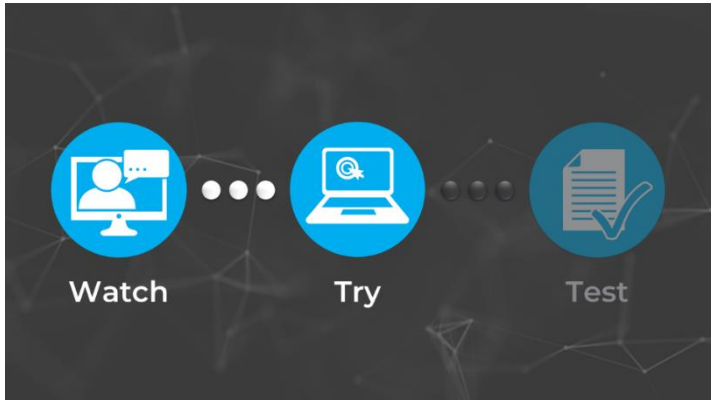
Developed using
2022.1 CSD Release



advancing structural science

Table of Contents

Try – it's your turn! 2
Objectives 2
Materials 2
Case Study 3
Introduction 3
Provided Input Files 4
Importing the protein-ligand complex from the PDB 5
Preparing Protein for Docking Experiment 6
Exploring Ligand for Docking Experiment 8
Defining the Binding Site 8
Selecting Ligand for Docking 9
Setting up the GOLD run 9
Run GOLD 11
Visualization and Analysis of the Results 11
Conclusions 14
Next steps 14
Glossary 15
Basics of Hermes Visualisation 18



Try – it's your turn!

In the second part of the *Protein-ligand docking 101 - running a simulation in GOLD* module, "U Try", you will review the concepts and features presented in the first part, "U Watch", and try them out yourself. If at any time you want to rewatch the videos, you can find them from the course webpage (https://www.ccdc.cam.ac.uk/Community/educationalresources/CSDU/gold-101/).

This tutorial will introduce you to how to perform protein-ligand docking in GOLD.

Objectives

In this workshop, you will investigate docking a ligand in the active site of a kinase receptor and observe the hydrogen bond interactions. In this workshop, you will:

- Learn how to set up a docking run using the GOLD wizard in Hermes
- Analyse docking results

This workshop will take approximately ***20-25 minutes*** to complete.

Note: The words in *Blue Italic* in the text are reported in the Glossary at the end of this handout.

Note: This workshop is for beginners; no entry skills are required, but you might find it useful to familiarise with Hermes interface. The basics visualisation options for Hermes are at the end of this handout.

Materials

For this workshop we will use the files that you can download from here.

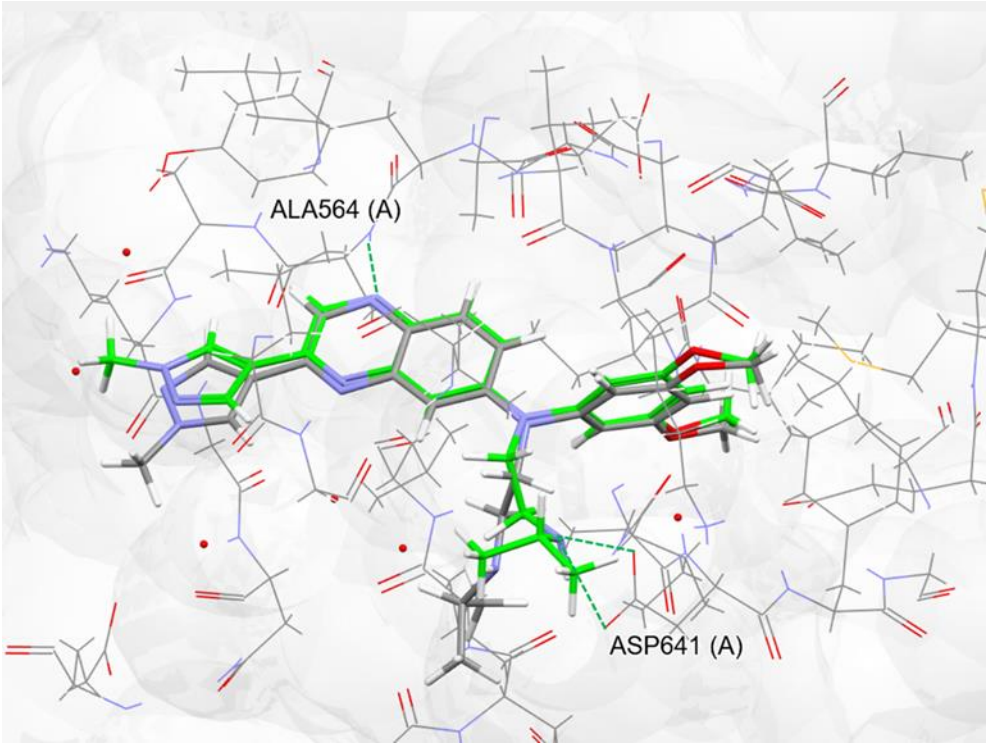


Figure 1. Docked complex of Erdafitinib with fibroblast growth factor receptor (FGFR) kinase (PDB code:5EW8). The docked ligand is coloured in green, native ligand is coloured in grey with capped representation, water molecules in red and the protein residues are coloured in grey. Erdafitinib docks in the binding site with an RMSD = 0.99 Å to the native structure.

Case Study

Introduction

Fibroblast growth factor receptors (FGFRs) are a family of receptor tyrosine kinases expressed on the cell membrane that play crucial roles in both developmental and adult cells. Frequent genetic alterations discovered in FGFRs and evidence implicating some as drivers in diverse tumors has been accompanied by rapid progress in targeting FGFRs for anticancer treatments.

(3,5-dimethoxyphenyl)-~{N}'-[3-(1-methylpyrazol-4-yl)quinoxalin-6-yl]-~{N}-propan-2-yl-ethane-1,2-diamine), aka Erdafitinib, is the first-ever fibroblast growth factor receptor (FGFR) kinase inhibitor indicated for patients with locally advanced or metastatic urothelial carcinoma.[2]

Your protein:

Fibroblast growth factor receptors (FGFRs) are a family of receptor tyrosine kinases (see Figure 1).

Your ligand:

(3,5-dimethoxyphenyl)-~{N}'-[3-(1-methylpyrazol-4-yl)quinoxalin-6-yl]-~{N}-propan-2-yl-ethane-1,2-diamine), aka Erdafitinib. The co-crystallised structure of FGFR-Erdafitinib (PDB code: 5EW8) reveals that the ligand binds in the hinge region of this kinase occupying the ATP binding site. (See Figure 3).

[2]Patani H., Bunney T. D., Thiyagarajan N., Norman R. A., Ogg D., Breed J., Ashford P., Potterton A., Edwards M., Williams S. V., Thomson G. S., Pang C. S.M., Knowles M. A., et al Landscape of activating cancer mutations in FGFR kinases and their differential responses to inhibitors in clinical use. Oncotarget. 2016; 7: 24252-24268.

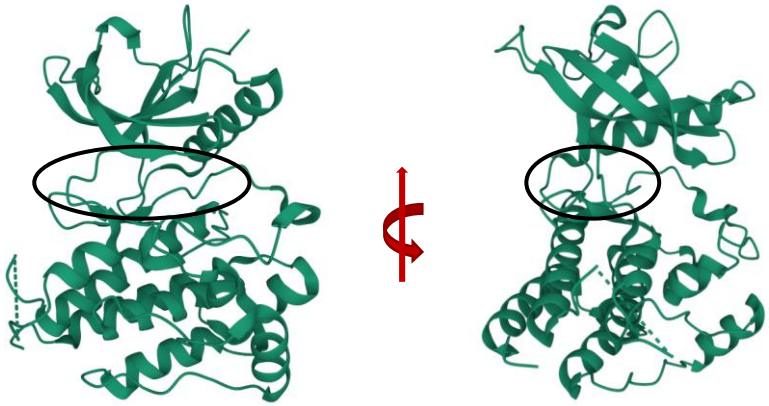


Figure 2. Crystal structure of the tyrosine kinase domain of fibroblast growth factor receptor 1 (PDB Code: 5EW8). The circled region shown in the figure is the hinge region. Hinge region of a kinase connects two lobes i.e., C and N-lobe and it is the site for ATP binding.

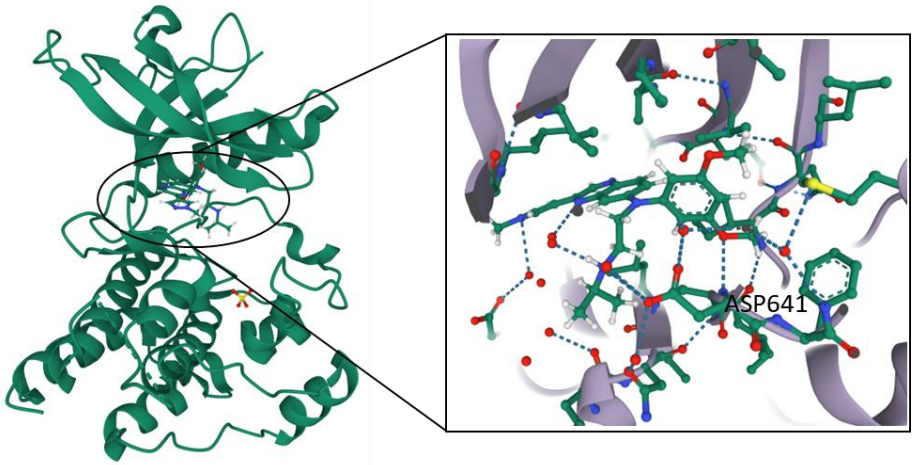


Figure 3. Co-crystal structure of the tyrosine kinase domain of fibroblast growth factor receptor bound with (3,5-dimethoxyphenyl)-~{N}'-[3-(1-methylpyrazol-4-yl)quinoxalin-6-yl]-~{N}-propan-2-yl-ethane-1,2-diamine), aka Erdafitinib (PDB Code: 5EW8).

The protein ligand interaction analysis of the PDB complex shows formation of hydrogen bond (bond length = 2.96 Å) between the ASP641 residue in the hinge region of the FGFR kinase and the dimethoxyphenyl ring of the Erdafitinib.

Challenges:

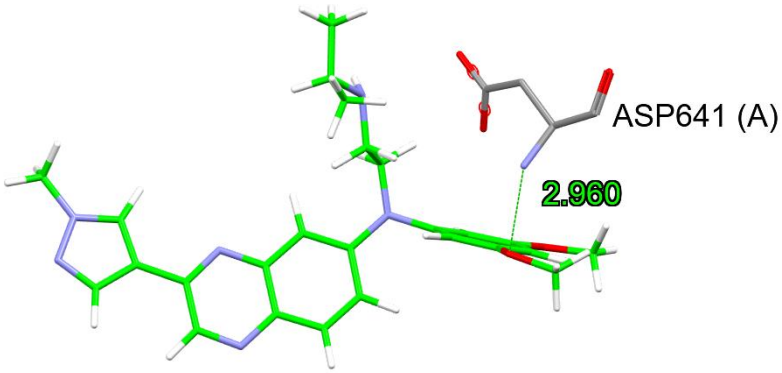
In this docking exercise we will try to dock the ligand Erdafitinib in its crystallographic or the native binding pose.

- In this tutorial, we will evaluate if we are able to dock the provided ligand in the active site and achieve poses similar to the reference structure.

Provided Input Files

Please download and unzip the input files **from this link**. In the folder you will find:

- The c-crystal structure of the of FGFR-Erdafitinib (***5ew8.pdb***). You can use this protein to walk through the key steps required to prepare a protein for docking experiment or you can follow the option illustrated in the first step of this workshop.
- The ligand file (***5SF_ideal.sdf***). This has been set up in accordance with the guidelines for the preparation of input files that you can find in the GOLD User Guide (Setting Up the Protein(s) and Setting Up Ligands chapters).

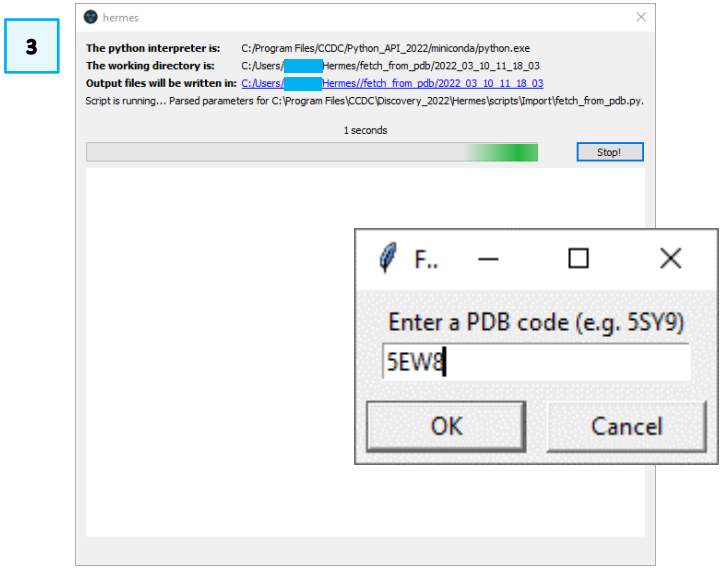


Importing the protein-ligand complex from the PDB

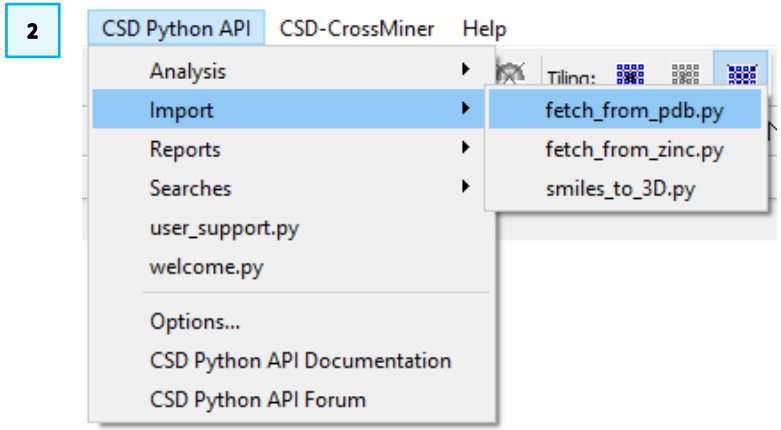
1. Launch Hermes by clicking its icon .

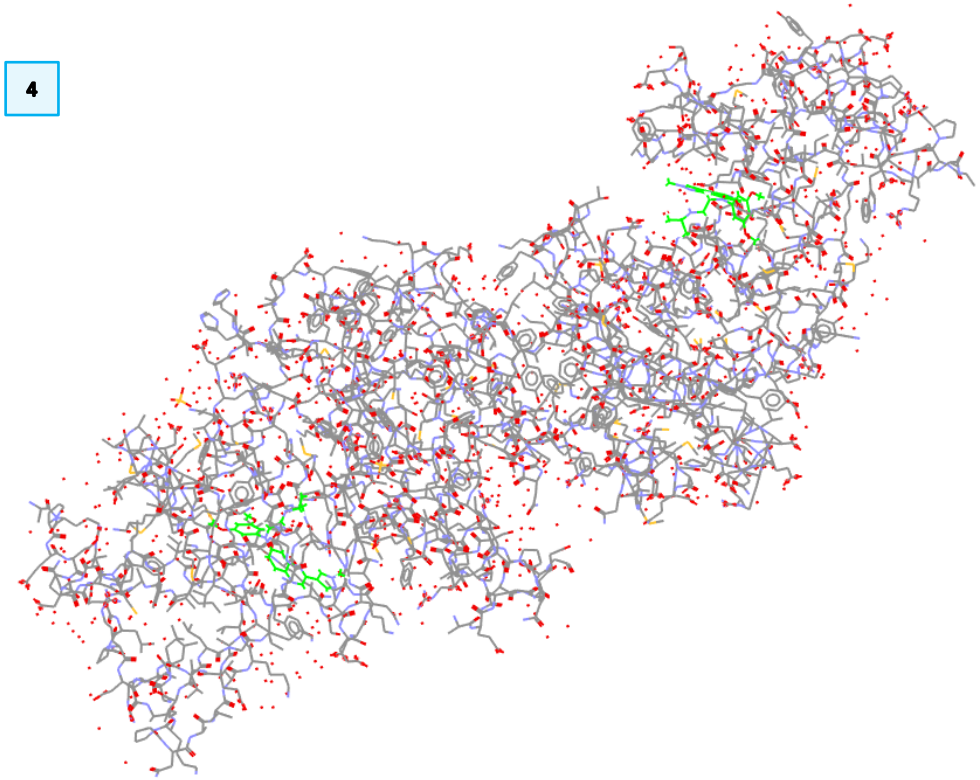
2. We are going to import the file of our protein and ligand directly from the PDB using the CSD Python API functionality. From the menus at the top, click on **CSD Python API**, then go to **Import**, and choose **fetch_from_pdb.py**.

3. This will bring up 2 windows: one displaying a loading bar, which show the script is running, and a smaller one, where we will insert the PDB code. For this case study we will type: 5EW8. Then click **OK**. It will take a few seconds to load the selected complex in Hermes. Once the structure is loaded, close the script window.



4. To better visualise the structure, you can zoom out using the **zoom-** button above the visualizer or right-clicking while moving the mouse down. The ligands are shown with green carbon atoms by default, the proteins with grey





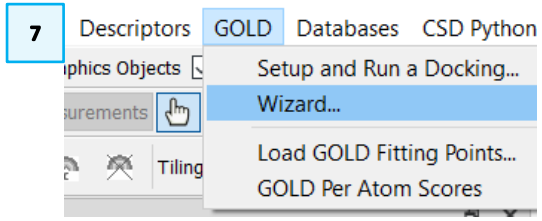
carbons, while the red dots around the structure are oxygen atoms of water molecules.

5. We can see we have two *homodimeric* chains. In the *Molecule Explorer* on the left of the visualiser, click on the arrow next to *5EW8*, then next to *Chains* and next to *Ligands* to expand and access all the components. If you don't have the Molecule Explorer displayed, see the instructions at the end of this document. You can try ticking / unticking each component to see that they will appear / disappear in the visualizer.

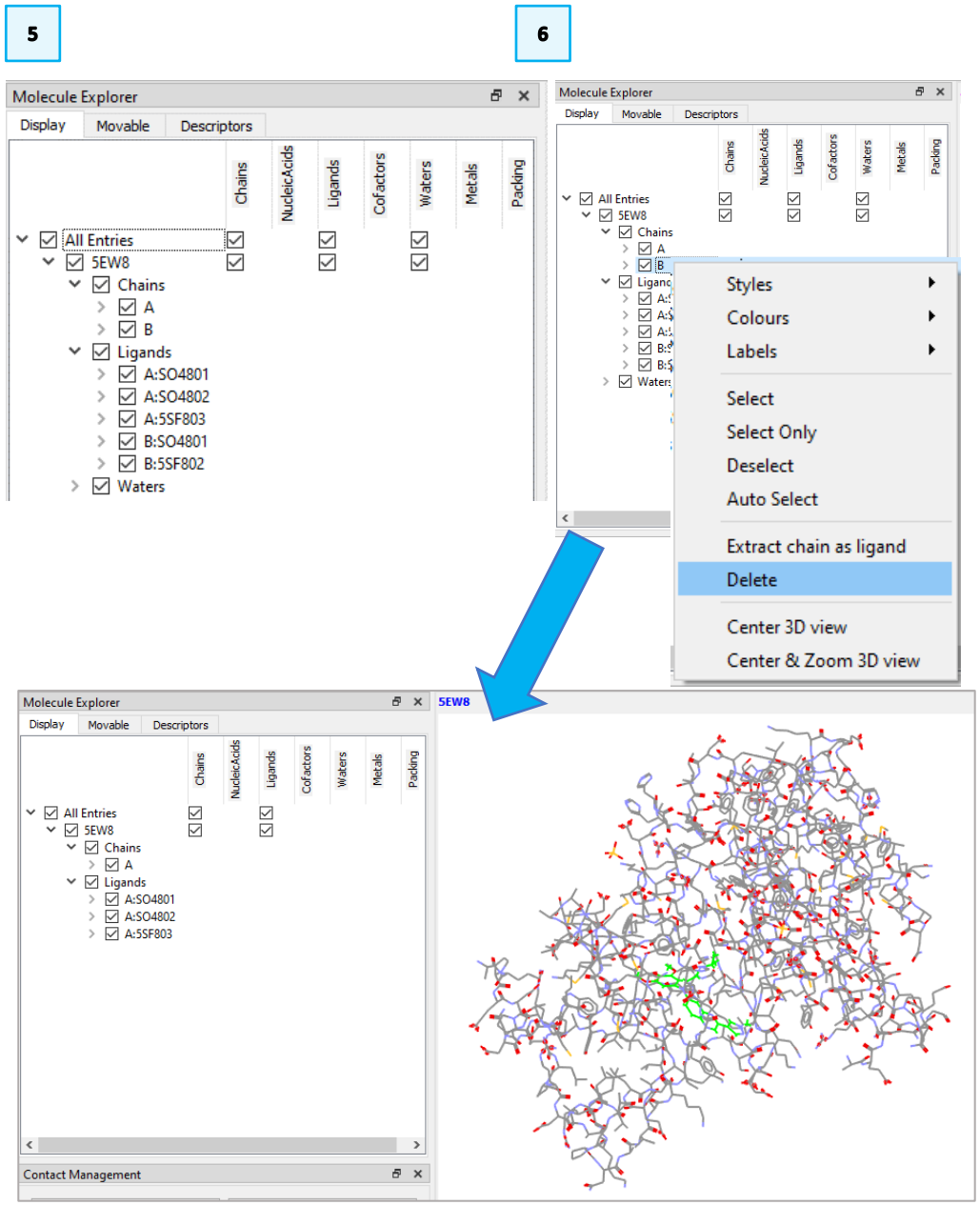
6. For our calculations we will use chain A. For sake of speed of the calculations we will delete chain B and its ligands. To do so, **right click** on B and select **Delete**. Repeat this operation for all the entries under *Ligands* marked with "B". For this example, there are no structural waters that need to be included, so we can delete them too. You do not need to expand the entry for *Waters* and delete the molecules on-by-one, but you can right-click directly on *Waters*.

Preparing Protein for Docking Experiment

7. To start our docking simulation, click **GOLD**, then pick **Wizard...** from the resultant pull-down menu.



8. In the *GOLD Setup* window, we see that we are at the **Wizard step 1: Select one or more proteins.** To select our protein, in the *Global Options* tab tick on the box next to the protein code, 5EW8. This will bring up a new tab labelled with the name of the protein file, *5EW8*. Click **Next** at the bottom of the



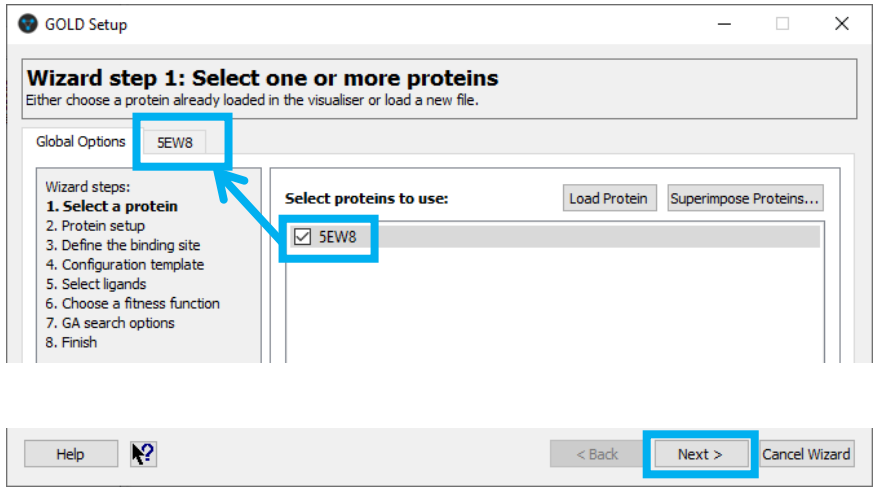
Wizard to move to **Wizard step 2: Protein setup**, then on the *5EW8* tab adjacent to the *Global Options* tab.

9. We are now going to add hydrogen atoms to the protein. From within the *5EW8* tab, at the *Protonation & Tautomers* option, click on **Add Hydrogens**. A pop-up window will inform you that 2347 hydrogens were added to the protein. Click **OK**. **Remember**: check the hydrogens in the binding site to ensure they are properly located! For this example, they are correct.

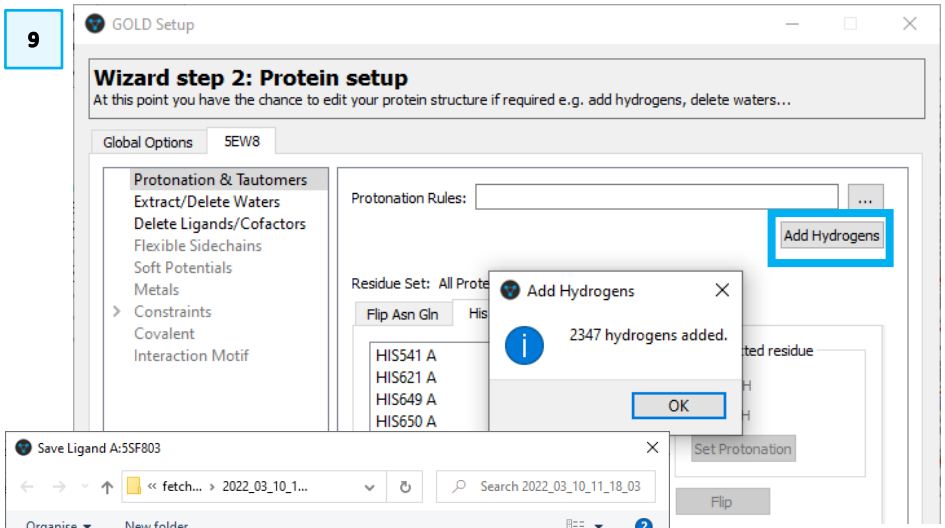
10. In this example we can skip the *Extract/Delete Waters* step, as we have already deleted all water molecules after importing the protein.

11. We move to the next part by clicking on *Delete Ligands/Cofactor* on the left-hand list. We select the ligand *A:5SF803*, then click **Extract**. The *Save Ligand A:5SF803* window pops up and we will insert a file name, for example *ligandA*, and click **Save** to save in the selected working directory.

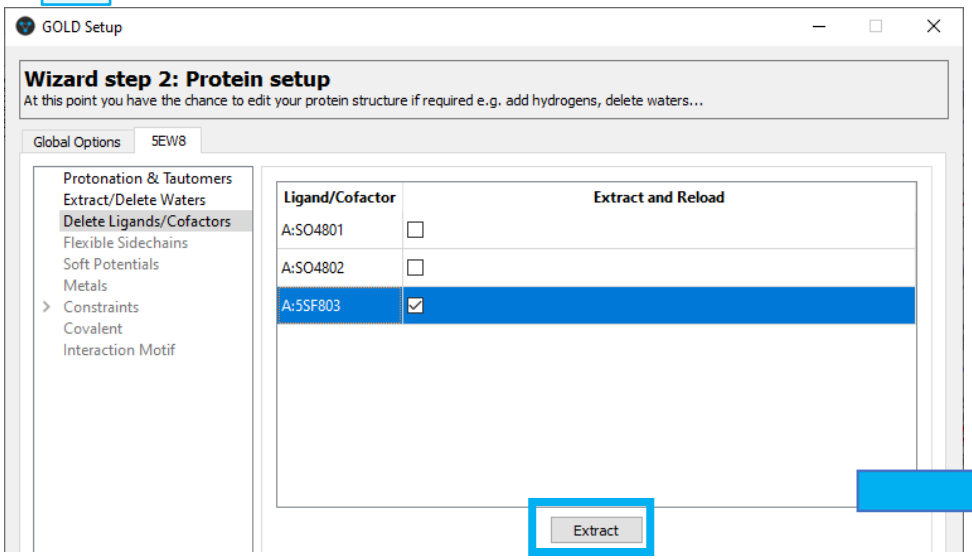
8

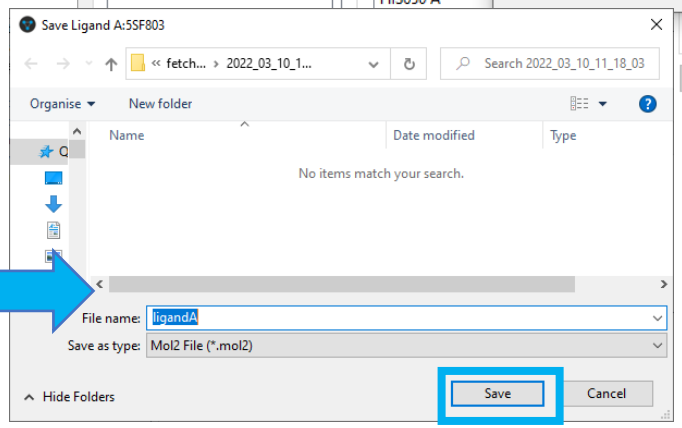


9



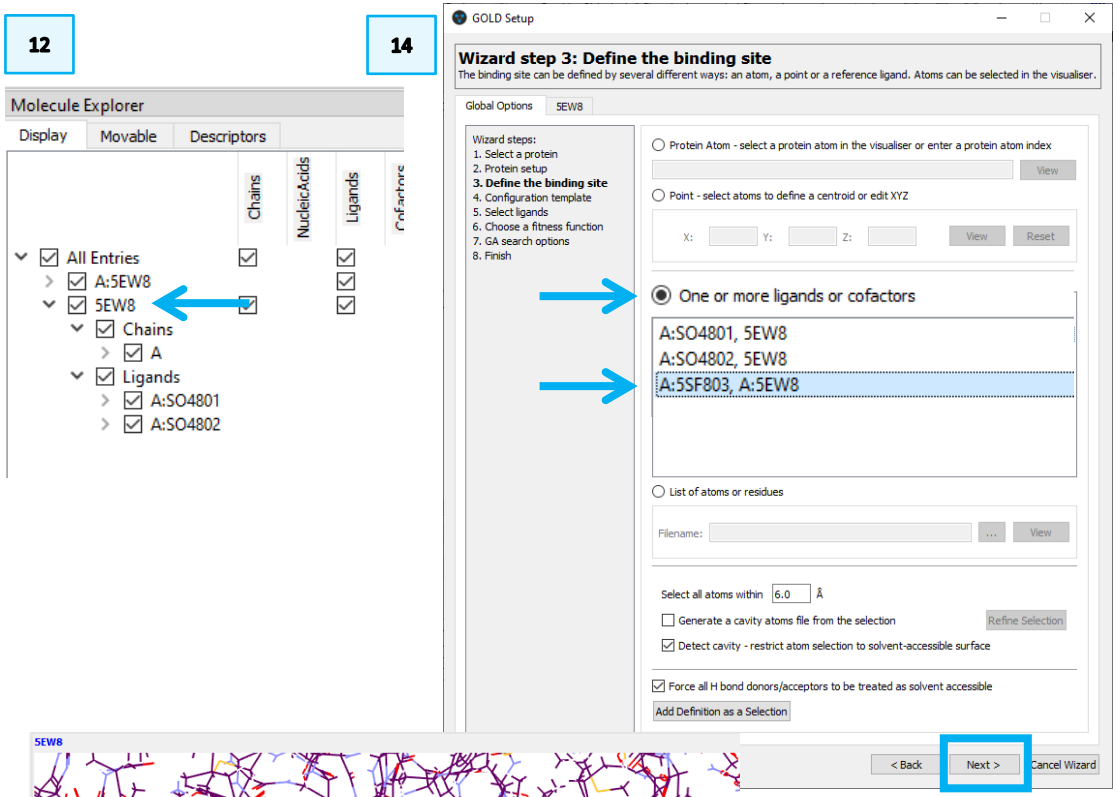
11





12. You see that the ligand is not listed anymore in the list of ligands under the protein, but it is now present in the *Molecule Explorer* as an independent object. We will leave the sulphate molecules as they are, as they are not close to the binding site.

13. The protein is now set-up for docking; click **Next** at the bottom of the Wizard to move to the following step, **Wizard step 3: Define the binding site**.



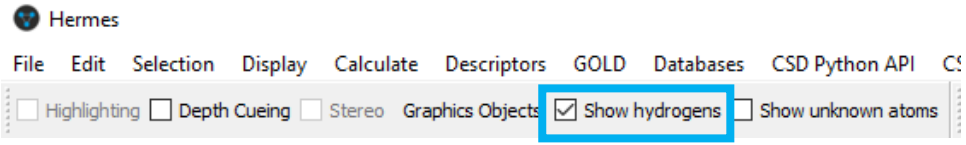
Exploring Ligand for Docking Experiment

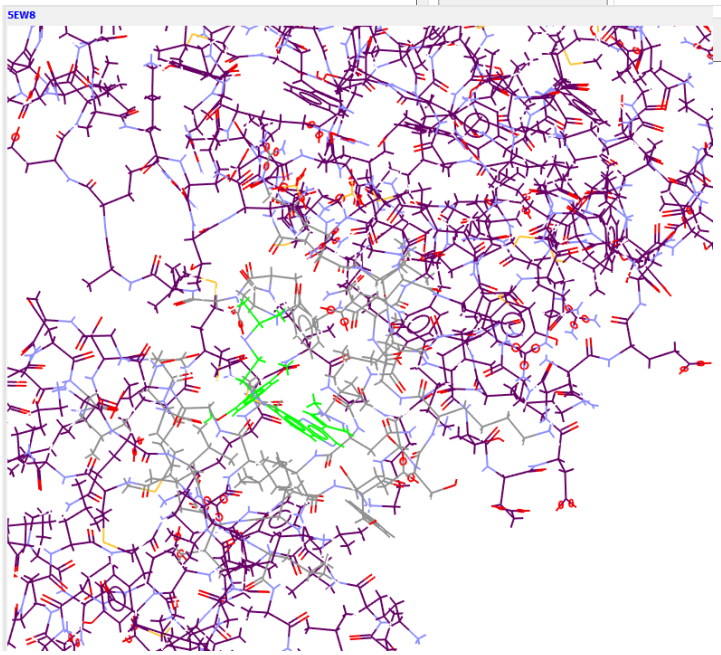
Defining the Binding Site

Defining the binding site for docking is a crucial step and can be done in several ways (see GOLD User Guide for further details).

In this tutorial, we will define the binding from a reference ligand.

14. From the GOLD Wizard, click the radio button next to **One or more ligands** and click to select the ligand that we have just extracted and reloaded, which is *A:5SF803, A:5EW8* to determine the binding site. Carbon atoms outside of the binding site will turn purple. Leave all other parameters to their defaults. For visualisation purposes, it can help here to switch off the display H-atoms using the **Show hydrogens** tick box in the top-level menu of Hermes. Click **Next** to proceed.





15. We are now at the **Wizard step 4: Optional – load a configuration template**. At this point you are given the option to load a configuration file template. These templates can be used to load recommended settings for a number of different types of docking protocols (see GOLD User Guide). These templates

are not relevant for this example, so we will specify all docking settings manually. We will skip this step and click **Next** to proceed to the **Wizard step 5: Select ligands**.

Selecting Ligand for Docking

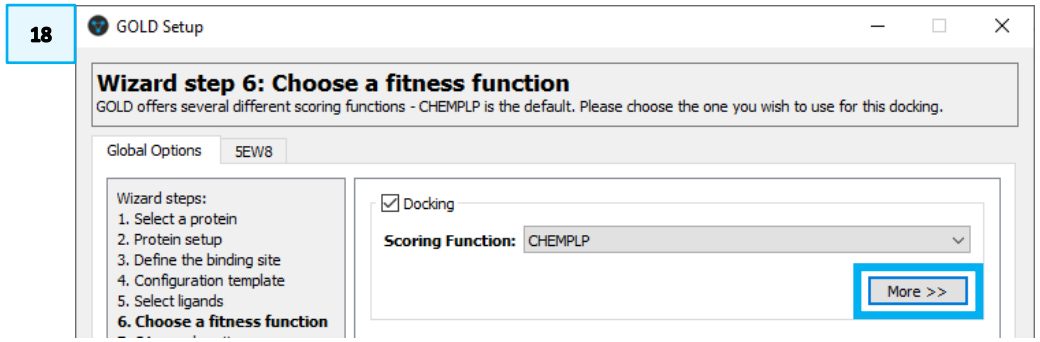
Here you have the option to supply the file containing the reference ligand (i.e., the ligand in the native *conformation*) that will be used to perform an automated RMSD calculation with respect to the reference ligand conformation. For each GOLD solution, the resultant RMSD value will be written out to the files containing the fitness function ranking, i.e., the ligand rank file (.rnk) and bestranking.lst file. First select the ligand you want to dock and the you have the option to load the reference crystallographic pose.

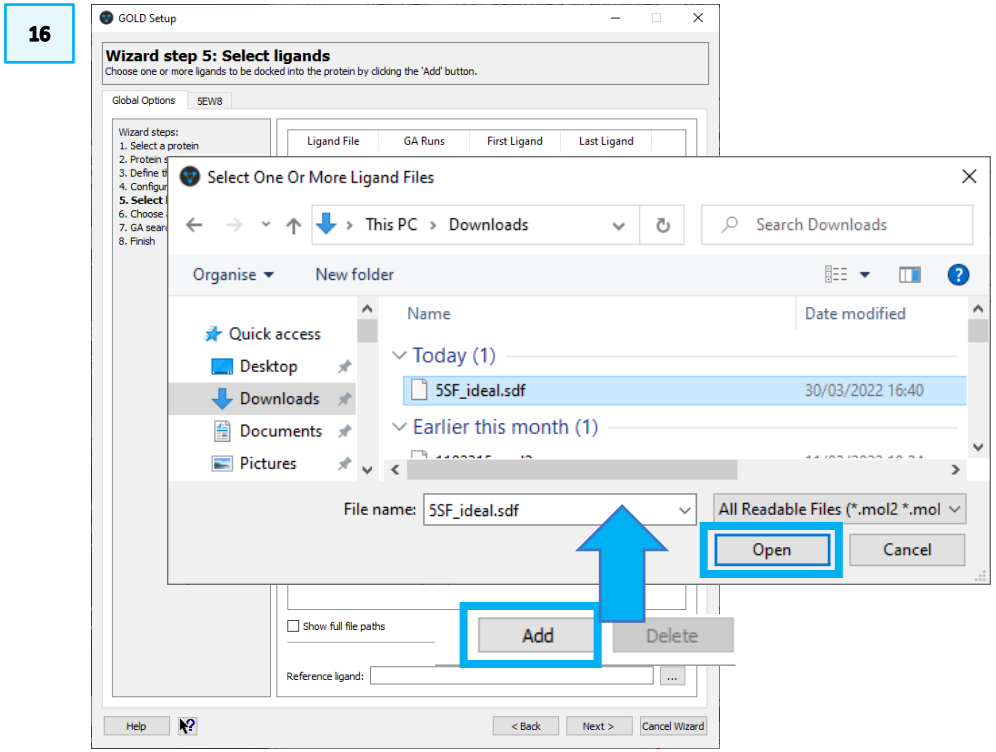
16. To proceed with the GOLD setup, we need to specify the ligand that we want to dock into the protein. For this example, we will use the file *5SF_ideal.sdf* (download it from Materials or from the Virtual Workshops webpage linked in the chat). To select the ligand, click **Add** at the bottom of the *GOLD Setup* window, navigate to the folder where the file is located, select the file *5SF_ideal.sdf* and click **Open**. Our ligand is now listed under *Ligand File.* Note that the number 10 means that 10 poses have been requested.

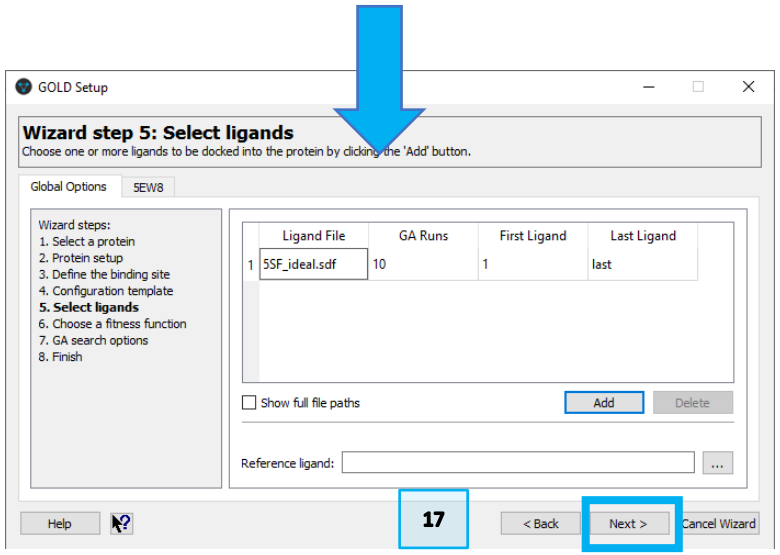
17. Once the ligand is selected, we can click **Next**.

Setting up the GOLD run

18. We are now at the **Wizard step 6: Choose a fitness function**. We will keep the **Scoring Function** as *CHEMPLP*. Click **More** to access more options.





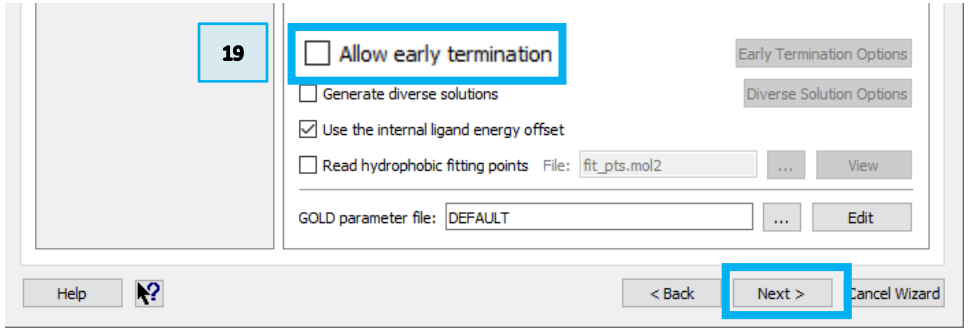


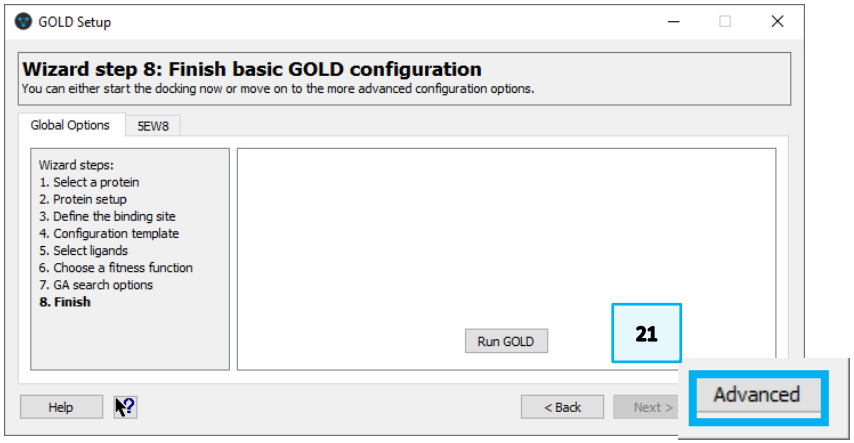
19. From the options at the bottom, we untick *Allow early termination*, then click **Next** to move to **Wizard step 7: Genetic Algorithm search options**.

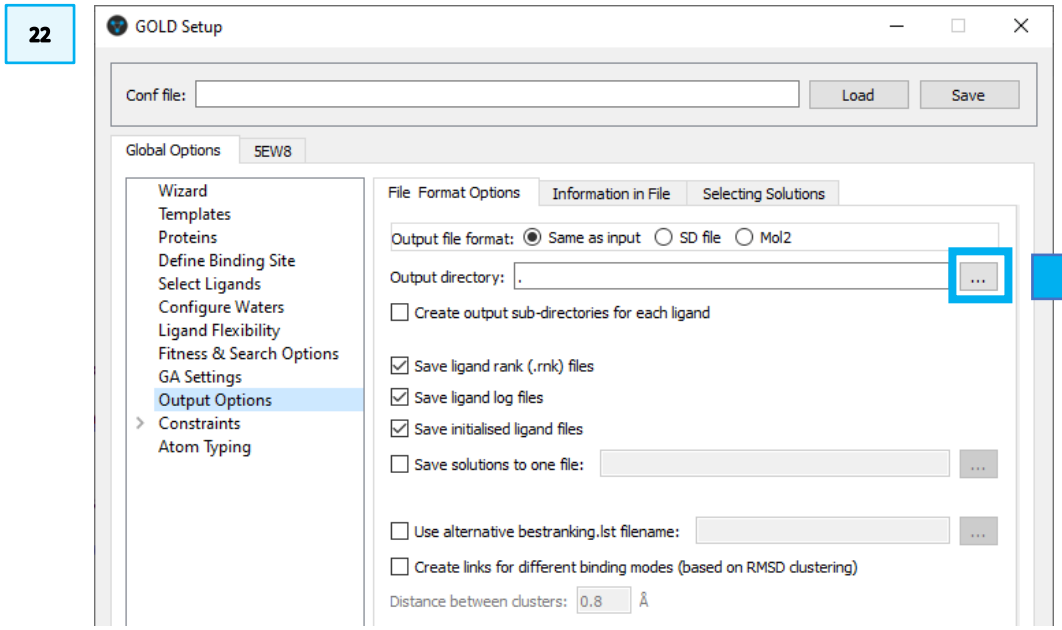
20. At this step we will keep the default options. If you want to explore more options, you can do so by clicking **More**. Click **Next** to move to the next step.

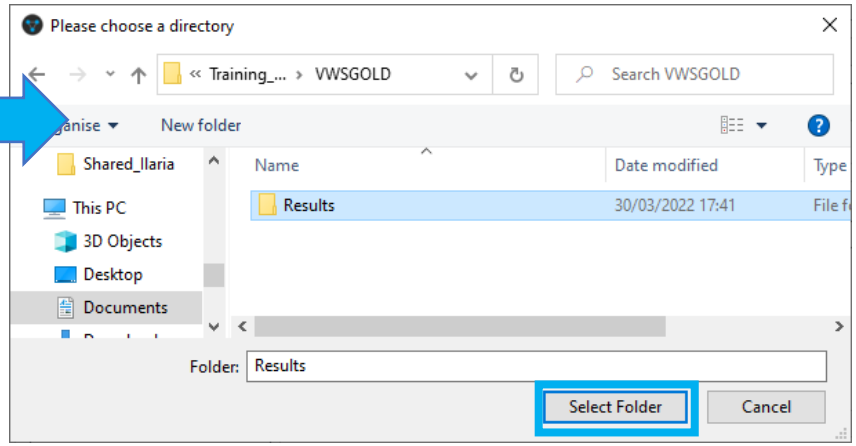
21. In **Wizard step 8: Finish basic GOLD configuration**, we click the **Advanced** button at the bottom right of the window.

22. From here, in the *File Format Options* tab we select *Output Options* from the list on the left-hand side, and select the directory where to store the results, by clicking on the **three dots** next to the **Output directory** bar. Navigate to the folder where you wish to store the results and/or create a new folder. Select the folder by clicking on it, then click **Select Folder**. You may also explore the other tabs, but for this example we will keep the other options at the default.



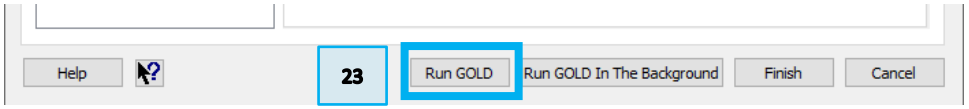




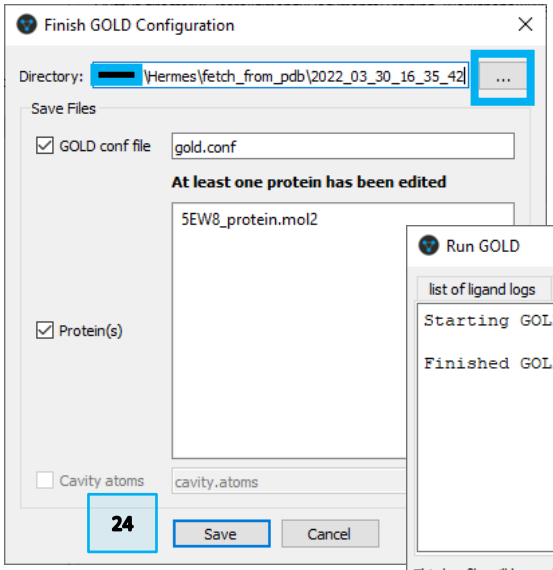


Run GOLD

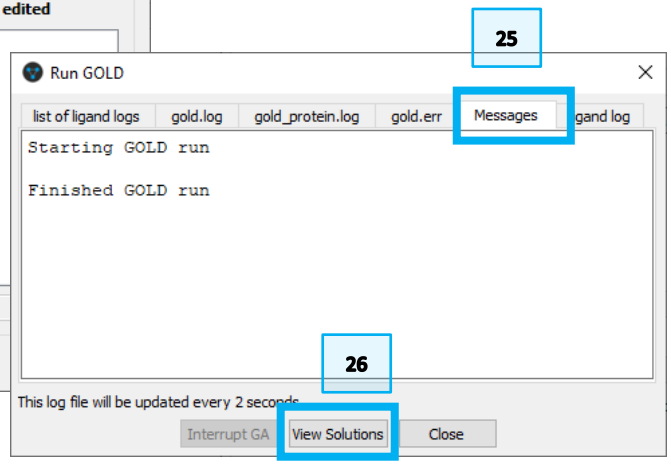
23. At the bottom of the *GOLD Setup* window click the **Run GOLD** button to start the GOLD run.



24. A pop-un window will appear, named *Finish GOLD Configuration*. You may change the folder where to store the files, for example to be next to the results, by clicking on the **three dots** next to **Directory**. When you are satisfied with your selection, click **Save**.

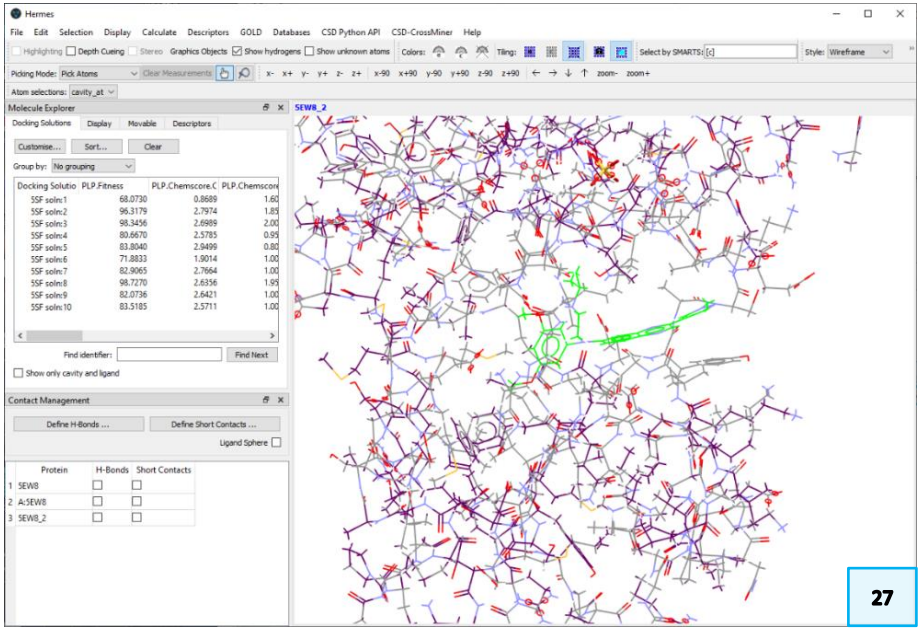


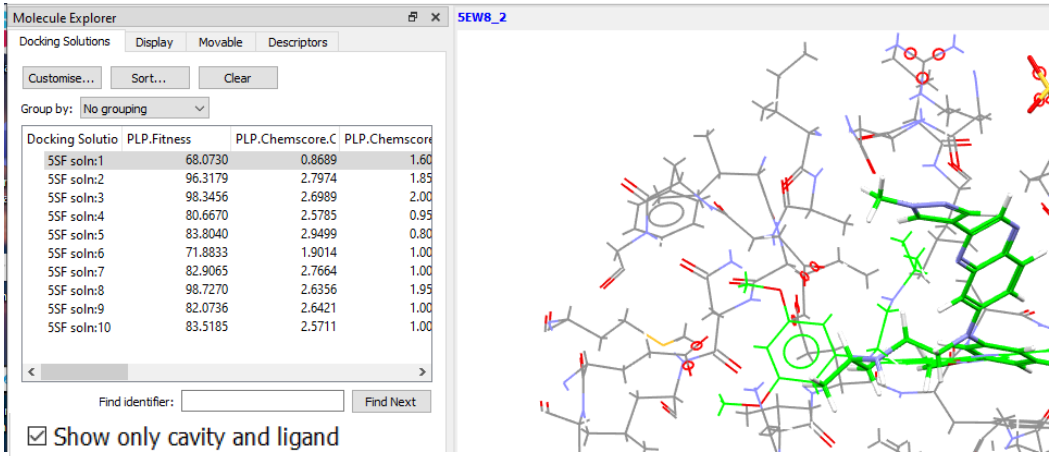
25. This will bring up a new window, *Run GOLD*, where you can follow the run, by going to the *Messages* tab.



Visualization and Analysis of the Results

26. Click **View Solutions** to visualise the results in Hermes. We can then click **Close** in the *Run GOLD* window and close the *GOLD Setup* window. The solutions are listed on the left-hand side in the *Docking Solutions* tab of the *Molecule Explorer*. This tab becomes available only after the GOLD run was completed.





27. If you click on each entry, you will see it displayed in the visualizer. You may tick the *Show only cavity and ligand* box for clarity of visualization.
    Note: GOLD is *stochastic*, so your results are likely to differ from the one displayed in the pictures in this example.

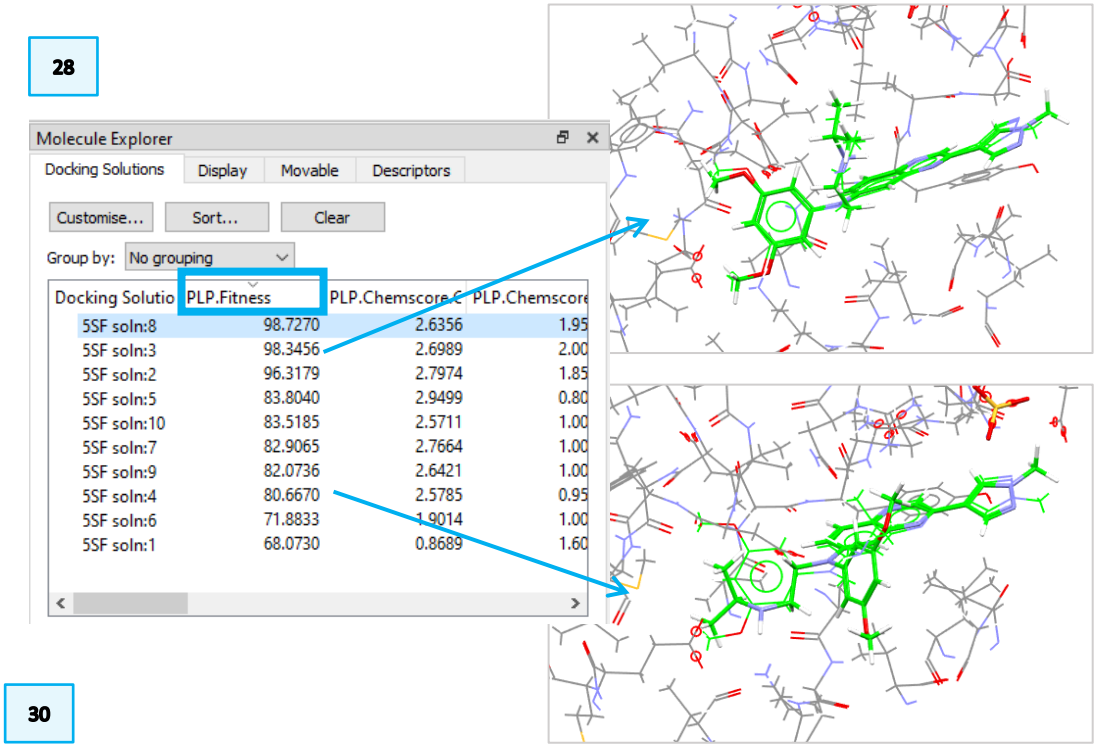
28. You can sort the solutions based on their PLP fitness score by clicking on **PLP.Fitness** in the *Docking Solutions* tab of the *Molecule Explorer* (click twice sort by high to small). You can now observe and compare the solutions, and sort them also based on the other scores.

29. When observing the solutions, we can notice that the conformational space has been explored during the run.

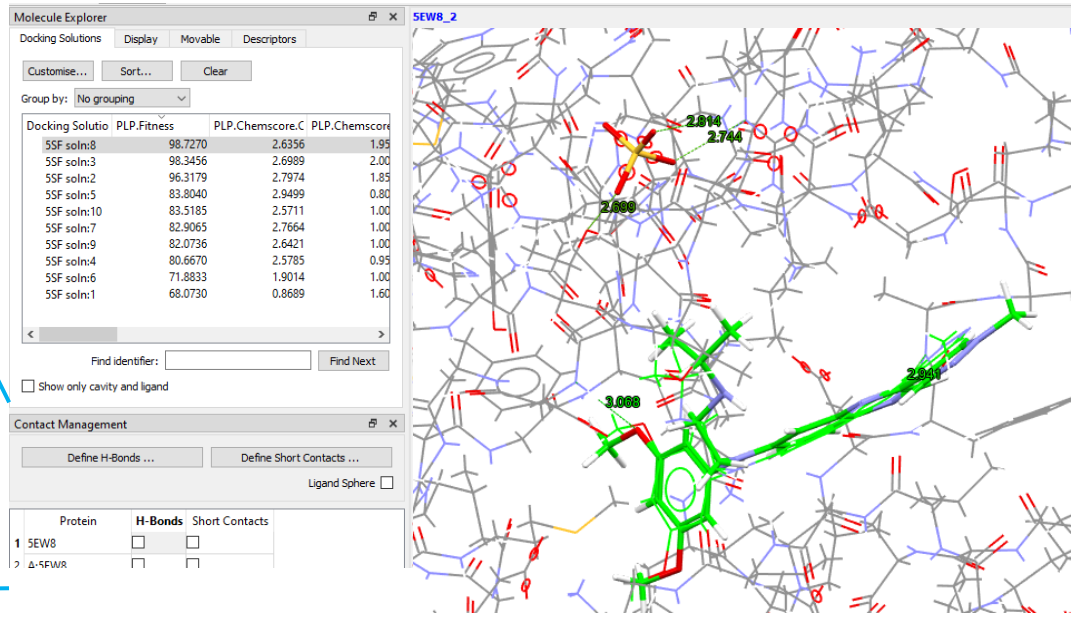
30. You may also explore the hydrogen bonds formed. To do so, in the *Contact Management* panel on the left-hand side, tick the box in the *H-bonds* column for the component of interest. The docked solutions correspond to entry 5EW8_2. If you don't see this tab, follow the instructions at the end of this handout to display it.

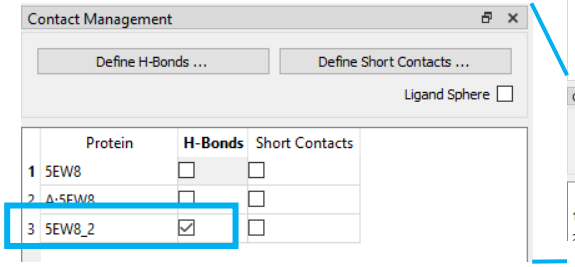
31. **Observe**:
    - What are the hydrogen bonds connecting?
    - How do the hydrogen bonds change in the different solutions?
    - Are the formed hydrogen bonds the same ones observed in the original crystallographic structure?
      Hint: you may wish to load the co-crystal structure provided in the Workshop folder (**5ew8.pdb).**

28



30

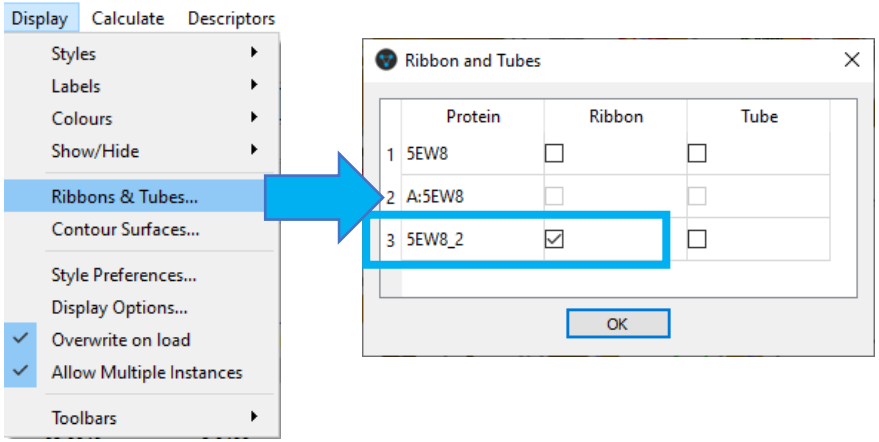




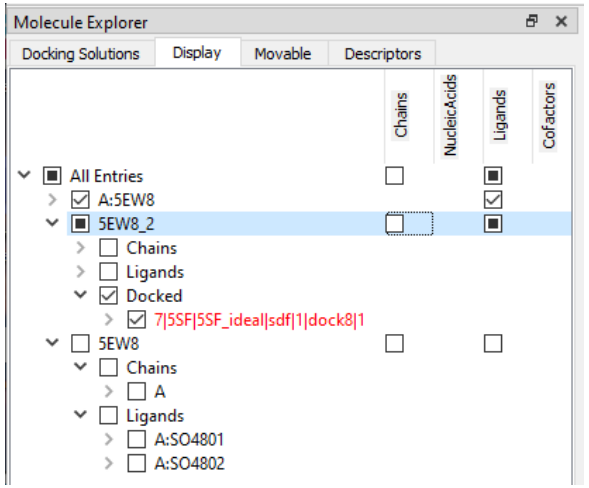
32. To enhance visualization, you may also use a ribbon display. Go to the **Display** menu at the top and select **Ribbons & Tubes**. In the *Ribbon and Tubes* menu that appears, tick the box in the *Ribbon* column next to *5EW8_2*. Click **Ok** to close the pop-up.
    At any stage, you may untick again the *Show only cavity and ligand* option that we ticked at **Step 27**.

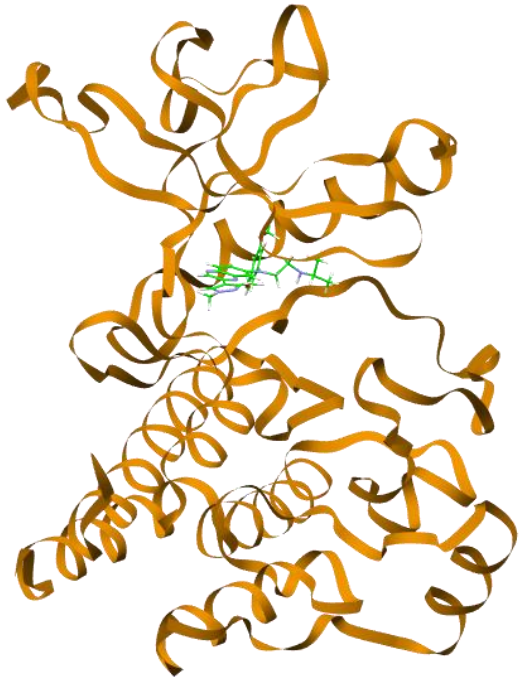
33. To improve visualization, you may wish to remove the underlying chain to the ribbon. To do so, go to the *Display* tab in the *Molecule Explorer* and:
    - Untick the protein entry *5EW8*.
    - Untick also the entry for the docked solutions *5EW8_2.*
    - Expand *the 5EW8_2* molecule and tick *Docked,* this will display the docked pose.

32



33





Conclusions

- The hydrogen bond interactions between active site residues and the docked ligand are retained, but in the case of highest scored docked molecule the NH group of the Asp641 forms hydrogen bond with the N26 on the propanyl-ethane-1,2-diamine chain of the ligand (Figure 4(C)) compared to the crystal structure where the Asp641 forms hydrogen bond with the methoxyphenyl group (Figure 4(B)).

- The Ala564 of the Erdafitinib retains its hydrogen bond interaction with the quinoxaline ring in the docking exercise (Figure 4(B), (C)).

- If we compare the results with the native ligand, we observe that the ligand, *5SF_Ideal.sdf* docks in the target binding site with low RMSD (*RMSD = 0.99 Å*) to the native ligand (Figure 4(A)). This suggests for good docking capability of GOLD.

After this workshop, you should be able to:

- Import a file in Hermes directly from the PDB using the CSD Python API functionality.
- Set up protein, ligands, and parameters for a GOLD run, and run it.
- Visualise and analyse the GOLD results with Hermes.

Next steps

Now that you have completed the Try part of this CSDU module, you can head to the final part: the quiz! Go back to the module webpage (https://www.ccdc.cam.ac.uk/Community/educationalresources/CSDU/gold-101/) and follow the instructions to complete the final test and earn a completion certificate.

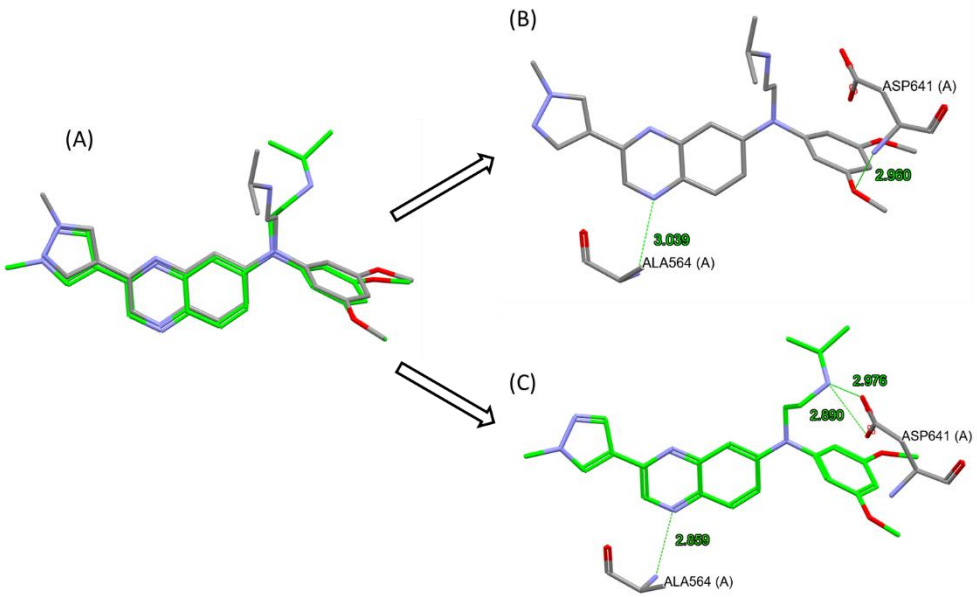
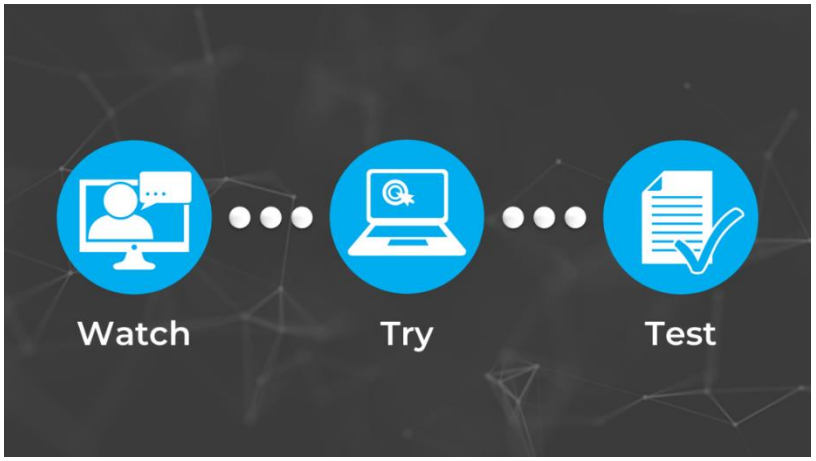


Figure 4. (A) Ligand overlay from the post docking analysis; it shows the docking pose and low RMSD of the docked ligand (Green) to the native ligand (Grey) (RMSD=0.99Å). (B) Protein-ligand interaction of Erdafitinib in the co-crystal structure. (C) Protein-ligand interaction of Erdafitinib (5SF_Ideal.sdf) docked in the active site of the receptor.



Glossary

Binding Site

A specific region (or atom) 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.

Conformation

The spatial arrangement of the atoms affording distinction between stereoisomers which can be interconverted by rotations about formally single bonds. Some authorities extend the term to include inversion at trigonal pyramidal centres and other polytopal rearrangements. *Sources*: PAC, 1994, 66, 1077. (Glossary of terms used in physical organic chemistry (IUPAC Recommendations 1994)) on page 1099.

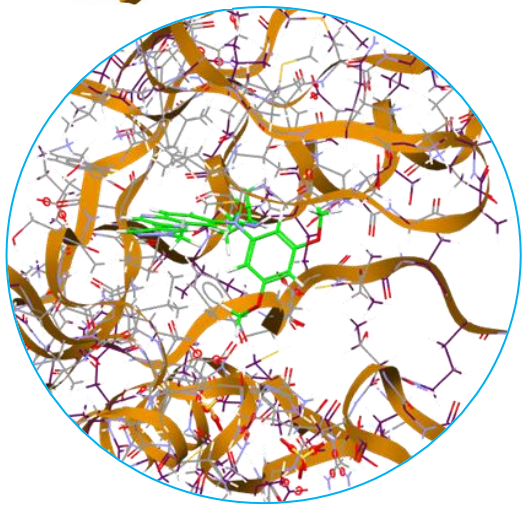
Docking

Docking studies are computational techniques for the expl oration of the possible binding modes of a substrate to a given receptor, enzyme or other binding site. *Source*: PAC, 1997, 69, 1137. (Glossary of terms used in computational drug design (IUPAC Recommendations 1997)) on page 1142.

Early Termination

The early termination option instructs GOLD to terminate docking runs on a given ligand as soon as a specified number of runs have given essentially the same answer. In this situation, it is probable that the answer is correct, and GOLD will just be wasting time if it performs more docking runs on that ligand. To switch early termination on, click on Fitness & Search Options from the list of Global Options given on the left of the GOLD Setup window, then enable the Allow early termination check box. To specify the early termination criterion, click on the Early Termination Options button.

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



Docking calculates a pose for the ligand and a score for the pose.

GA setting

A genetic algorithm is an optimization algorithm based on the mechanisms of Darwinian evolution which uses random mutation, crossover and selection procedures to breed better models or solutions from an originally random starting population or sample. *Source*: PAC, 1997, 69, 1137. (Glossary of terms used in computational drug design (IUPAC Recommendations 1997)) on page 1144.

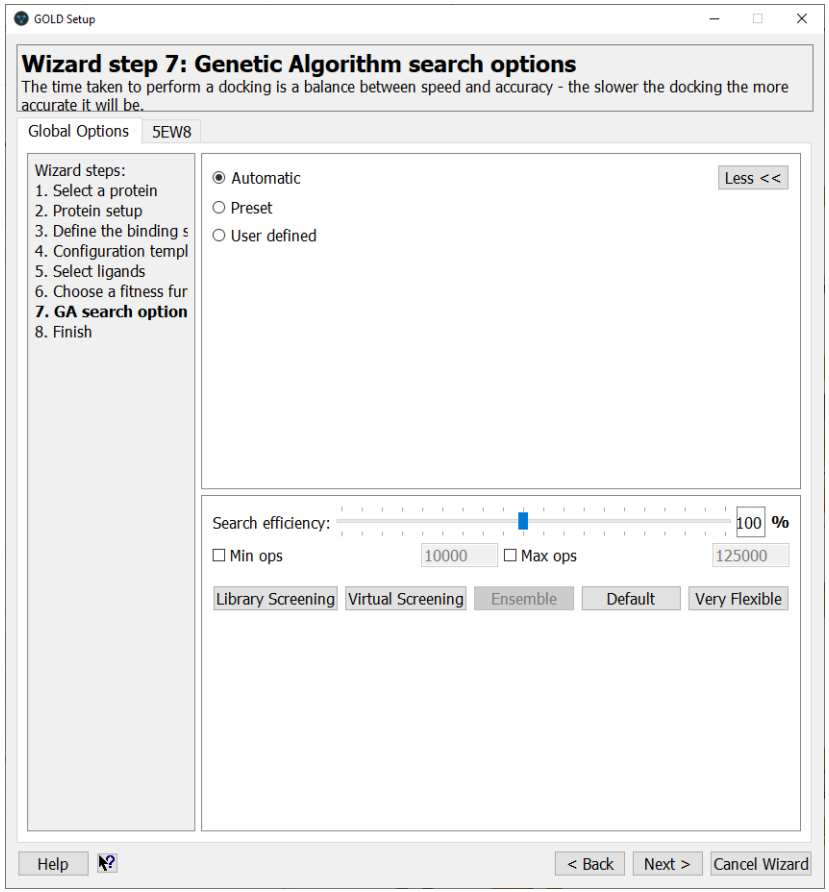
In GOLD, number of genetic operations performed (cros sover, migration, mutation) is the key parameter in determining how long a GOLD run will take (i.e. this parameter controls the coverage of the search space). GOLD can automatically calculate an optimal number of operations for a given ligand, thereby making the most efficient use of search time, e.g. small ligands containing only one or two rotatable bonds will generally require fewer genetic operations than larger, highly flexible ligands. The criteria used by GOLD to determine the optimal GA parameter settings for a given ligand include:

- the number of rotatable bonds in the ligand,
- ligand flexibility, i.e. number of flexible ring corners, flippable nitrogens, etc. (see Ligand Flexibility),
- the volume of the protein binding site, and
- the number of water molecules considered during docking (see Water Molecules).

The exact number of GA operations contributed, e.g. for each rotatable bond in the ligand, are defined in the gold.params file (see Altering GOLD Parameters: the gold.params File). For further information please refer the GOLD user manual.

GOLD

GOLD (Genetic Optimisation for Ligand Docking) is a genetic algorithm for docking flexible ligands into protein binding sites. GOLD has been extensively tested and has shown excellent performance for pose prediction and good results for virtual screening.

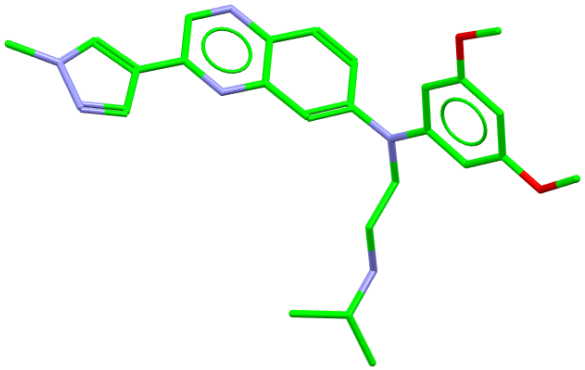


Homodimeric
In this context, homodimeric is used to refer to a protein composed by two identical parts.

Ligand
In this context, ligands are small molecules that bind to the protein and can change the protein function.

Scoring function
In the fields of computational chemistry and molecular modelling, scoring functions are mathematical functions used to approximately predict the binding affinity between two molecules after they have been docked. Most commonly one of the molecules is a small organic compound such as a drug and the second is the drug's biological target such as a protein receptor.

Stochastic sampling
Random sampling; random sample The sample so selected that any portion of the population has an equal (or known) chance of being chosen. Haphazard or arbitrary choice of units is generally insufficient to guarantee randomness. Source: PAC, 1990, 62, 1193 (Nomenclature for sampling in analytical chemistry (Recommendations 1990)) on page 1202.



5SF_Ideal.sdf aka Erdafitinib, is the first-ever fibroblast growth factor receptor (FGFR) kinase inhibitor indicated for patients with locally advanced or metastatic urothelial carcinoma. This ligand was docked in this workshop.

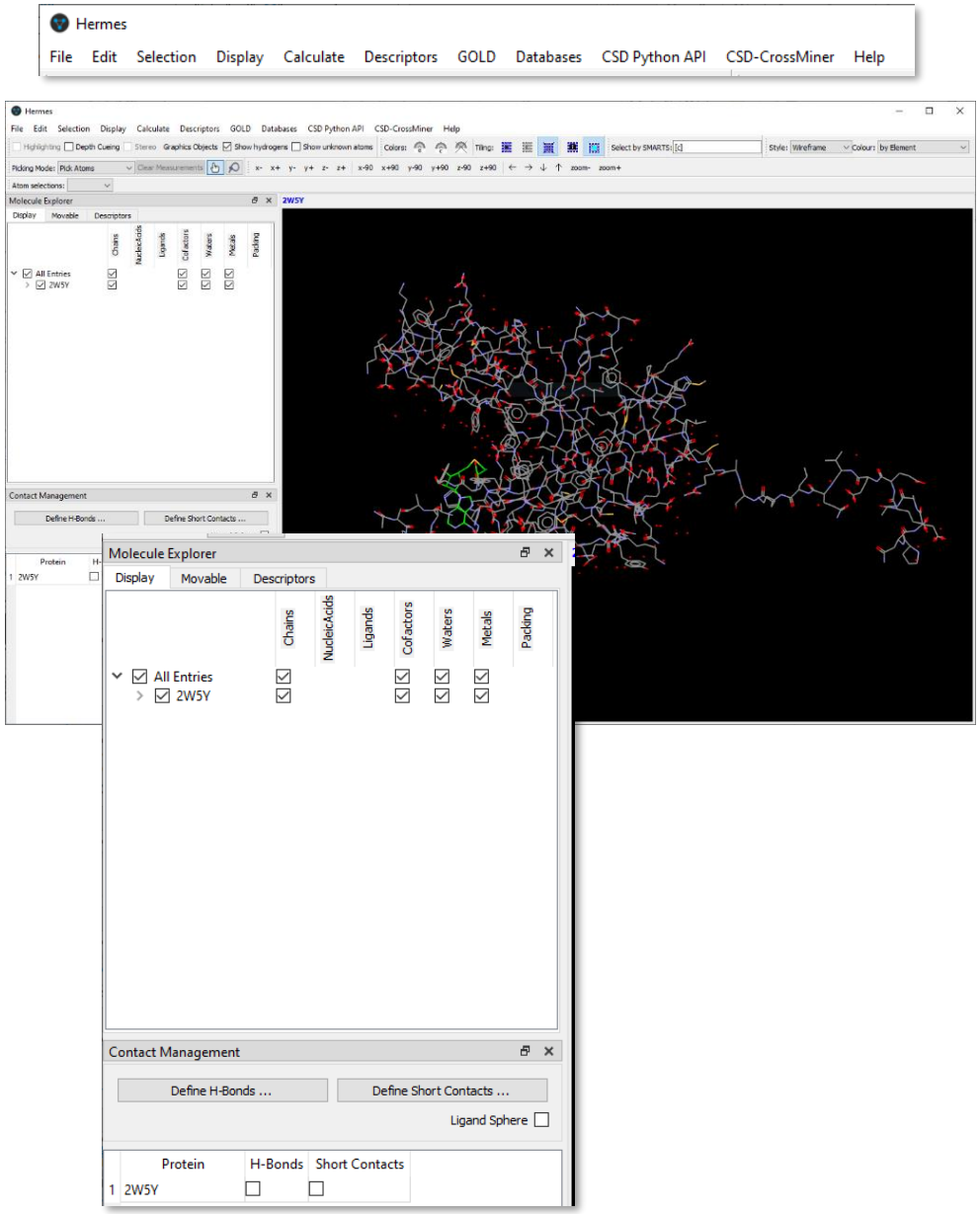
Basics of Hermes Visualisation

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**:

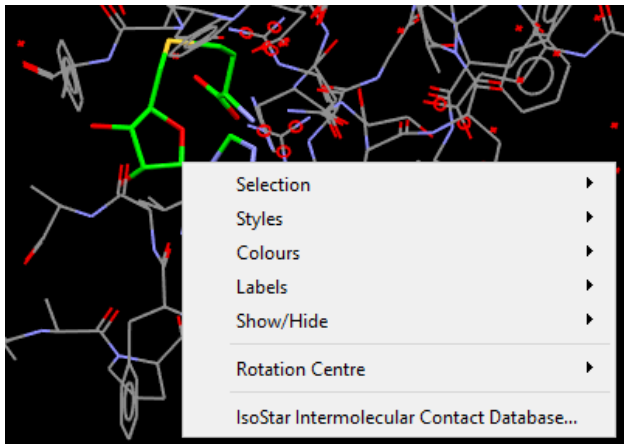
- Left mouse button and move – rotate molecules
- Middle Mouse wheel – move molecules up and down
- Right mouse button and move up and down – zoom in and out of molecules
- Shift + Left mouse button and move - rotate in the plane molecules
- Ctrl + Left mouse button and move - translate molecules

Right click:

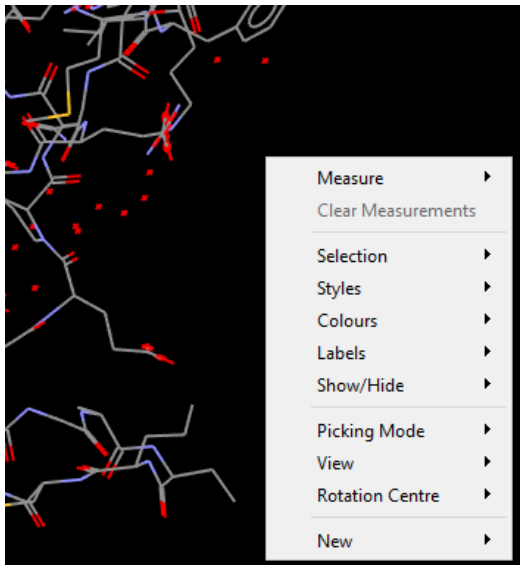
a) On a feature

b) Away from a feature

a



b



Creating high-resolution images

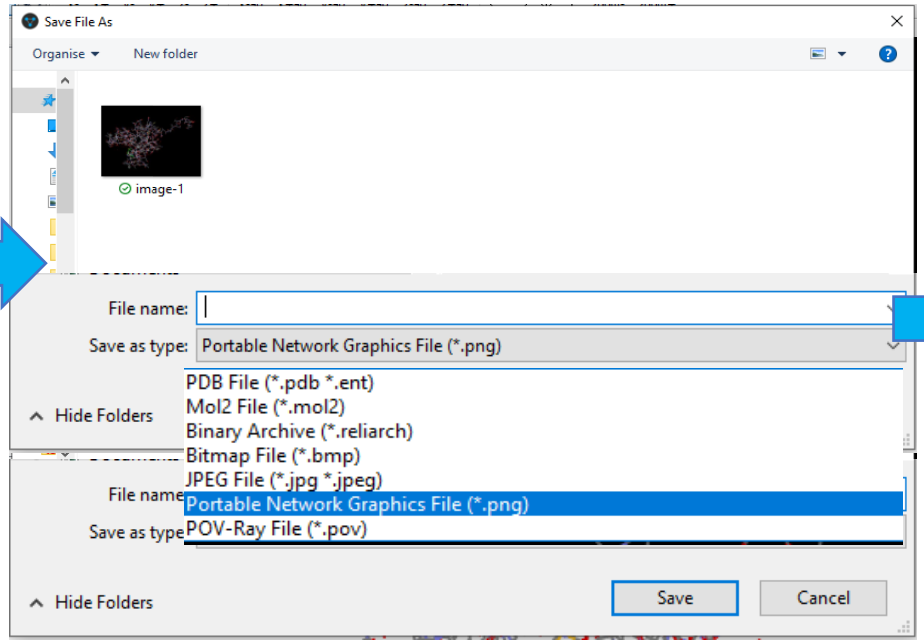
1. Go to the top menus and click on **File**, then **Save As**. Alternatively you can use the keyboard and do **Ctrl + S**.

2. In the *Save File As* window, select destination folder, file type and choose a name. Then click **Save**.

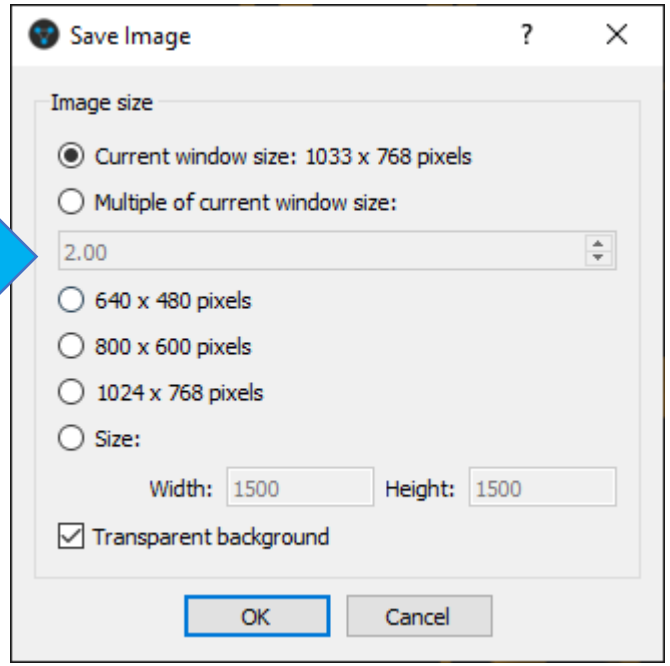
3. This will bring up the *Save Image* menu. Here you can select the resolution and tick whether you want a transparent background. Click **Ok**.

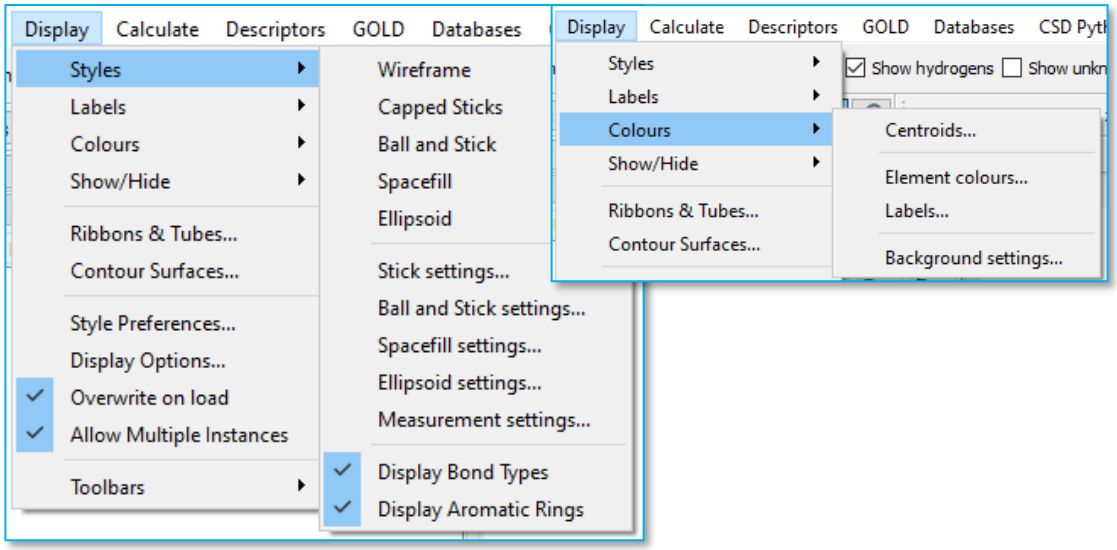
1

2



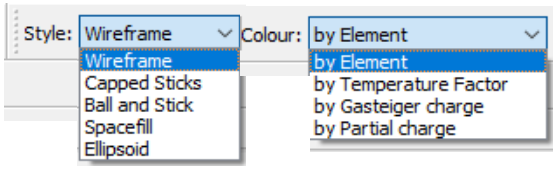
3





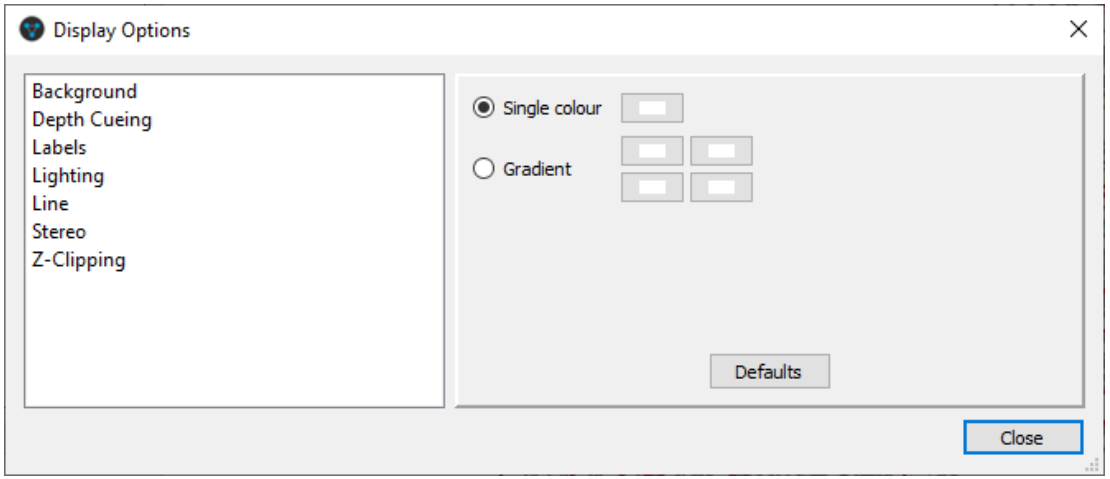
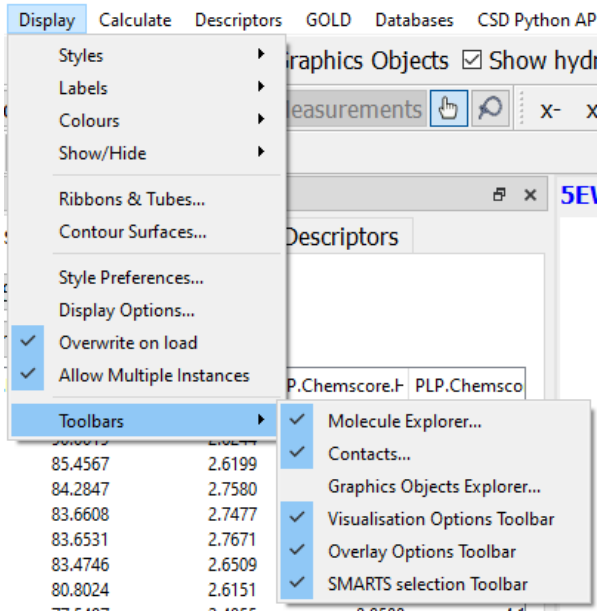
You can access visualization options from the **Display** menu.

- From **Display > Styles**, you can pick the representation style of atoms and bonds and change settings. **Note**: you can also access this feature from the dropdown menu at the top right of the 3D visualizer window.



- From **Display > Colours**, you can edit the colour for elements and object. **Note**: you can change the colouring style for the structure from the dropdown menu at the top right of the 3D visualizer window.

- Clicking on Display > Display Options brings up the Display Options menu, where you can edit settings for the 3D visualizer.



- If any of the toolbars at the top of the Hermes interface is not displayed and you wish to have it shown, or vice versa, you can do so from **Display > Toolbars**, where you can ensure to tick the ones you wish to have on.