

# Flexibility and Constraints in GOLD

2020.0 CSD Release

## Table of Contents

Purpose of Docking .....	3
GOLD's Evolutionary Algorithm.....	4
GOLD and Hermes .....	4
Handling Flexibility and Constraints in GOLD .....	5
Recap of the Related Pose Prediction Examples .....	5
Aims of the Example on Flexibility and Constraints .....	5
Your Challenge: Docking Two Inhibitors into the PI3Kdelta Apo Protein .....	6
Exploring Electron Density .....	7
Exploring Interactions in the PI3Kdelta APO Structure and Deriving a Flexibility Hypothesis for Our Docking .....	8
Setting Constraints Based on a Homologous Protein Structure .....	9
Overview of Flexibility and Constraint Options in GOLD .....	10
<i>Constraint Options</i> .....	10
<i>Protein Movement Options</i> .....	10
Edit Your Protein .....	11
Clearing the Hermes buffer .....	11
Visualisation .....	11
Prepare Your Active Site .....	12
Launch the GOLD Wizard .....	12
Load your protein .....	12
Define Your Binding Site from a Protein Atom.....	13
Specify Flexible Sidechains .....	14
Set a Protein HBond Constraint .....	15
Consider Alternative Protonation/Tautomeric States for Your Protein and Ligands ....	16
Enter Your Ligands.....	17

Set Fitness and Search Options .....	17
Specify Output, Save Your Setup and Invoke GOLD .....	17
Compare Ligand 1 Docking Results to the Co-Crystal Structure, 2x38.....	18
Compare Ligand 2 (Idelasilib) Docking Results to the Co-Crystal Structure, 4XE0.....	19

## Purpose of Docking

When you investigate potential drugs for a new protein target or try to improve existing medicines, it is important to know:

- if and how strongly a proposed molecule binds to the protein (binding affinity prediction)
- **how it binds to the protein (pose prediction)**

Docking 3D coordinates of a ligand into low energy states of a protein binding site (or active site) is commonly used for pose prediction. It approximates the process of a flexible ligand binding to an active site. Docking can be set up to make the most of a scientist's knowledge about:

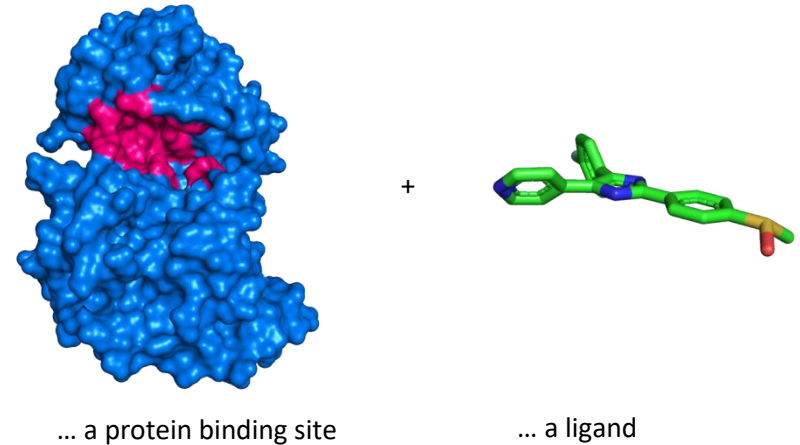
- the flexibility of the target
- structural waters
- metal binding sites
- preferred protein-ligand interactions.

Proposed docking poses give chemists insights and testable hypotheses about which parts of the ligand:

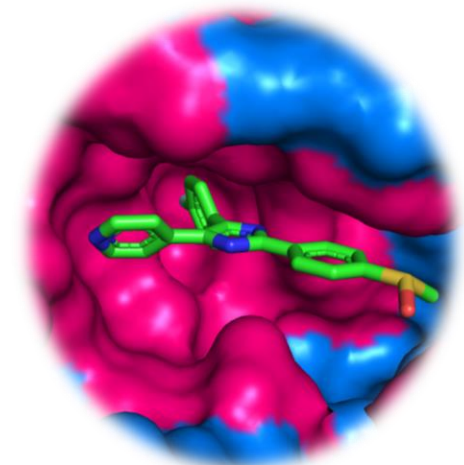
- are responsible for the binding to the protein
- should be optimised to improve the potency of the drug
- would cause clashes with the protein if changed
- can be replaced to improve pharmacokinetic properties and alleviate side effects.

The quality of an individual docking result is measured by a relative score. This is used to judge which poses generated for a particular ligand are the most likely ones. Docking scores cannot accurately predict whether a compound actually binds and what its affinity is. This needs to be confirmed by *experimental* assays.

Given ...

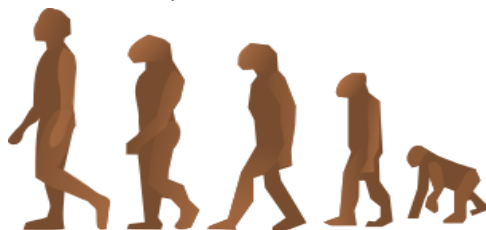


Docking calculates ...



## GOLD's Evolutionary Algorithm

GOLD stands for “**G**enetic **O**ptimisation for **L**igand **D**ocking”. It is one of the most widely used docking programs and is available as part of the CSD-Discovery and CSD-Enterprise Suites.



GOLD's evolutionary algorithm modifies the position, orientation and conformation of a ligand to fit into one or more low energy states of the protein active site. It maps ligand geometry parameters onto populations of chromosomes and then runs evolutionary rounds of **mutation**, **crossover**, **scoring** and **selection** to optimise protein-ligand interactions.

## GOLD and Hermes

GOLD itself is accessible from Windows, Linux and macOS command lines. In this guide, you will:

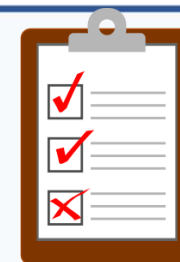
- prepare input for GOLD
- perform dockings
- analyse docking results.

To get started, you will use the Hermes visualiser to assist you with:

- setting up data in the correct format for GOLD
- running of GOLD and visualisation of docking output.

Understanding the difference between GOLD and Hermes will enable you to integrate working with GOLD into your own workflows and modelling environments.

## A Checklist for Docking



Have I got:

- 3D coordinates for my protein?
- access to electron density of my target?
- knowledge about where the protein active site is?
- knowledge on whether metal ions are involved in protein-ligand interactions?
  - 3D coordinates for such metal ions?
- 3D coordinates for my ligand, including the positions of hydrogens?
  - Are the bond lengths and angles in my ligand correct?
- an enumeration of all protonation/tautomeric states for my ligand?
- any expectations of which interactions are likely to take place within my active site?

Am I aware that:

- His side chains can adopt one of three protonation/tautomeric states?
- Glu and Asp side chains can interact with the ligand as carboxylates or carboxylic acids?
- crystallography at medium resolution is not able to provide certainty about the orientation of the amide groups in Asn and Gln side chains?
- protein side chains can move?

Have I checked that:

- there is crystallographic evidence for the position of long and surface exposed side chains, such as Arg and Lys?
- there are no unwanted water or buffer molecules in my active site?
- my assumptions on the positions of protein side chains are based on electron density data?

## Handling Flexibility and Constraints in GOLD

### Recap of the Related Pose Prediction Examples

In [GOLD Pose Prediction MLL1](#) we docked:

- a challenging large and flexible ligand
- into its “tailor-made” protein binding site.

In [GOLD Pose Prediction MLL1](#) we also learned how to use constraints to:

- reduce the time GOLD requires to find poses for your ligand
- make use of your expert knowledge
- improve the accuracy of your docking results.

### Aims of the Example on Flexibility and Constraints

In this example, *GOLD\_Flexibility\_and\_Constraints\_PI3KD*, we are facing the challenge to dock into an **apo** protein structure. Apo means the structure has been determined in the absence of any bound ligand in the cavity we are interested in. Please download the [example files](#) for this tutorial.

You will balance flexibility and constraints in two separate docking experiments trying to predict the poses of:

- a drug-like ligand
- the FDA approved drug idelasilib.

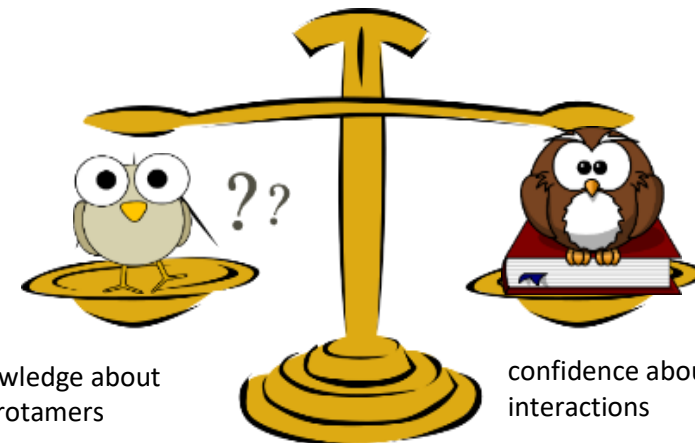
At the end of the tutorial, you will compare your predictions to the true structures of the protein-ligand complexes, determined by protein X-ray crystallography.

You will:

- carefully inspect your protein crystal structure
- use flexible sidechains in your sampling
- use a Protein HBond constraint to guide GOLD.



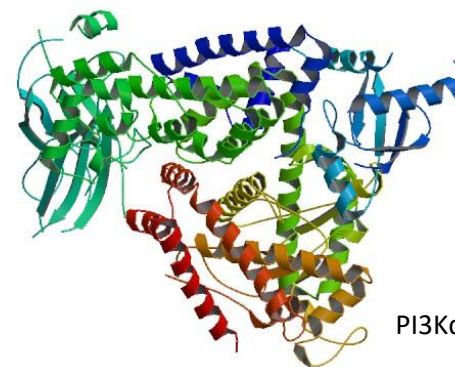
Download Input Files



## Your Challenge: Docking Two Inhibitors into the PI3Kdelta Apo Protein

### Protein kinases:

- transfer high energy phosphate groups from ATP to protein side chains or second messenger molecules
- are ubiquitous in regulatory processes
- are targets for intervening in cancer, cardiovascular and neurodegenerative diseases
- bind ATP in their hinge region in conserved patterns.



PI3Kdelta, deposited as PDB: 2WXR

### Your challenge is to:

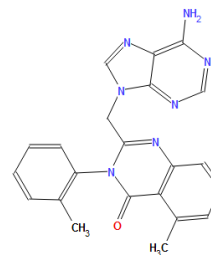
- time travel to 2009
- use the only known crystal structure of the leukaemia target Phosphoinositide kinase delta (PI3Kdelta), crystallised in the absence of any bound ligand (apo protein)
- to predict binding poses for two known inhibitors to PI3Kdelta.



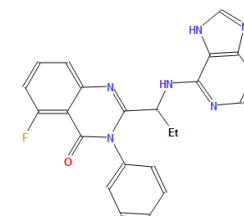
**2009**

### You are provided with the input coordinates for your docking experiments:

- two inhibitors:
  - ligand 1: 2-[(6-aminopurin-9-yl)methyl]-5-methyl-3-(2-methylphenyl)quinazolin-4-one
  - ligand 2: idelasilib, an approved drug for treating leukemia
- the coordinates for the PI3Kdelta apo structure, PDB: 2WXR.



Ligand 1: 2-[(6-aminopurin-9-yl)methyl]-5-methyl-3-(2-methylphenyl)quinazolin-4-one



Ligand 2: idelasilib

In 2009, a structure of PI3Kgamma complexed to an inhibitor (PDB: 2CHZ), was available and will be used in this tutorial.

At the end of this session, you will be allowed to time travel past 2009 and compare your predictions to the crystallographically determined binding conformations of Ligands 1, PDB: 2X38 and 2, PDB: 4XE0.

**2010, 2014**





## Exploring Electron Density

*What makes 2009 a good year to go to?*



In 2008, the Protein Data Bank (PDB) asked all depositors to submit structure factors along with the coordinates of their macromolecular models. [Structure factors](#) represent the reflection data from the crystallographic experiment and allow us to calculate the electron density that our model is based on.

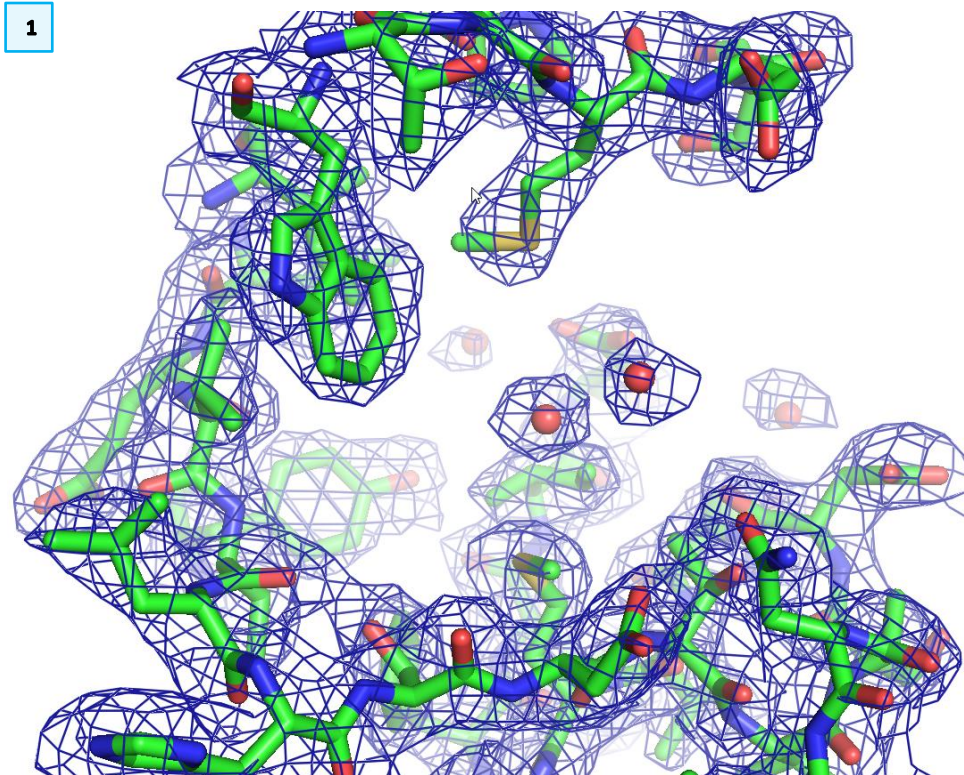
For ligand-bound structures, the PDB provides a density viewing option; for an example please study the density of [PI3Kgamma complexed with an inhibitor](#), which we will be referring to later in this tutorial.

As the apo structure of our target, PDB-code: 2WXR, was released in 2009, we have access to its underlying electron density and can judge the quality of the model.

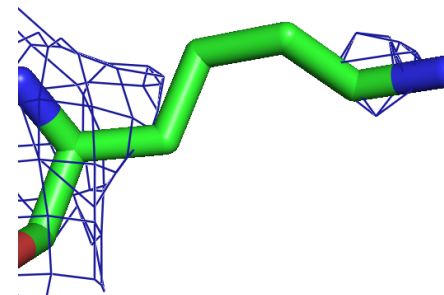
Coot and PyMOL are good free programs to visualise electron density. For this tutorial, we only illustrate the value of electron density inspection by looking at screenshots taken with PyMOL from the active site of 2WXR. We know where the active site is by homology with other kinases.

1. In 2WXR, one can find clear electron density for all sidechains in the binding site, no sidechains lack experimental density or are only added for model completion purposes.
2. For an example of a sidechain lacking experimental evidence, see the electron density of the surface exposed Lys841.

1



2



## Exploring Interactions in the PI3Kdelta APO Structure and Deriving a Flexibility Hypothesis for Our Docking

From the previous section, we know that all sidechains in our apo structure are in an experimentally confirmed conformation. Now we need to go further and think which side chains may move upon ligand binding.

An apo structure is likely to have solvent occupying the binding site. Let's see whether this is true for PI3Kdelta, how the solvent interacts with the sidechains and whether solvent molecules are likely to move upon ligand binding.

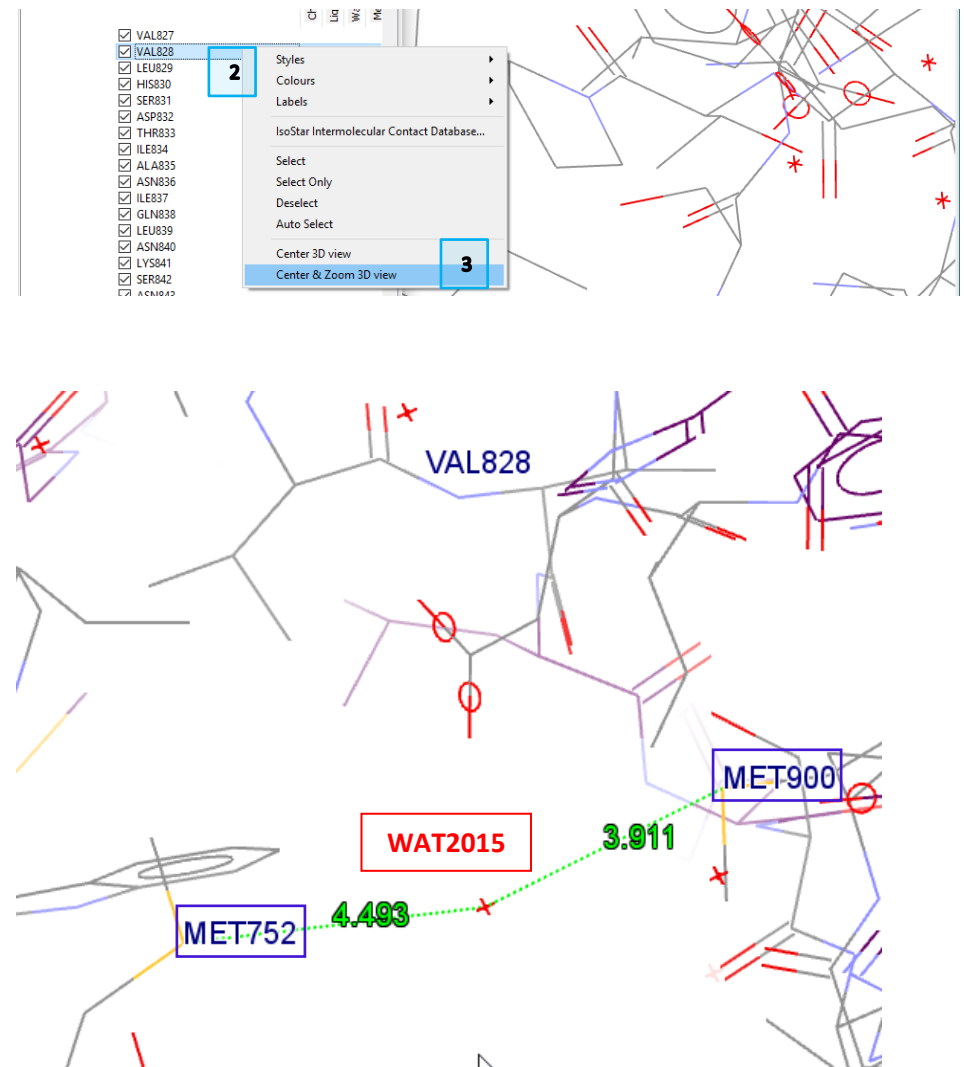
Load the PDB file *2wxr\_catalytic\_domain.pdb* into Hermes as explained in the [Introduction to Hermes](#) guide, page 3.

To focus in the active site:

1. Navigate to Val828 in *Molecule Explorer* (see *Introduction to Hermes*).
2. For clarity, **right** click on the residue to  
Select *Styles* -> *Ball and Stick*  
Select *Center and Zoom 3D view*.
3. Use distance measurements between the water molecules and protein sidechains to find out whether water molecules are tightly bound or just loosely attached.

The water molecule Wat2015 is loosely attached to Met752 and Met900. It is a valid assumption that this association will be disrupted once a ligand comes into the binding site. Thus, our docking will be based on the hypothesis that:

- Both Met752 and Met900 can move
- Wat2015 will move into the bulk solvent.





## Setting Constraints Based on a Homologous Protein Structure

### Balancing Conformational Uncertainty with Knowledge

In the previous section we have decided to treat two long methionine sidechains as flexible. This means that we have to:

- increase the torsional space for sampling
- perform long docking runs.

Exhaustive sampling bears the risk of GOLD “getting lost” in the multitude of potential combinations **of protein and ligand torsions**. To succeed with this docking, we need to **balance** the **uncertainty** about the conformation of sidechains within the active site with our **knowledge** about the target.

We are looking for interactions that are always made. Luckily, the [hinge region of protein kinases](#) (Figure 1) exhibits such strongly preserved interaction patterns.

### Picking one interaction pattern

To identify a likely protein-ligand interaction pattern for driving GOLD:

- Load the crystal structure of PI3Kgamma, *2chz-aligned.mol2* (dark grey, green ligand) and focus on the hinge region around residue Val828 in PI3Kdelta. Note that the PI3Kgamma was pre-aligned to the PI3Kdelta through **Calculate > Superimpose Proteins...** in Hermes.
- Val828 is conserved in PI3Kgamma (labelled VAL882) and forms 2 contacts to the ligand:
  - N-H...N-thiazol
  - C-O...N-amide.
- N-H group of Val828 points into a void at the back of the binding site. Water molecules are often “unhappy” at locations like this, **we strongly expect a ligand to form H-bonds with this group**.
- C=O group of Val828 already forms a hydrogen bond to Ser831, which is an interaction unique to PI3Kdelta (PI3Kgamma has Ala885 at the same spot). It is thus less likely to be involved in interactions with the ligand.

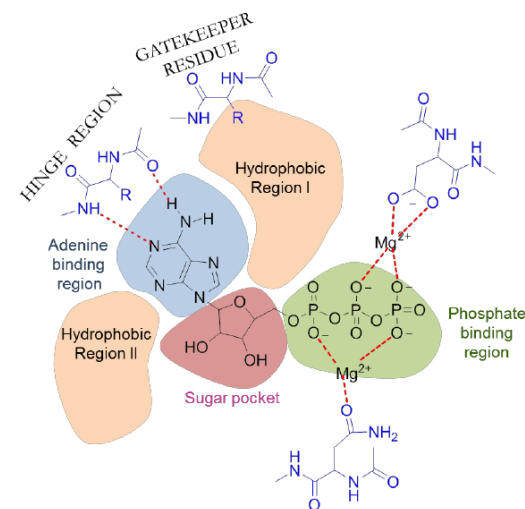
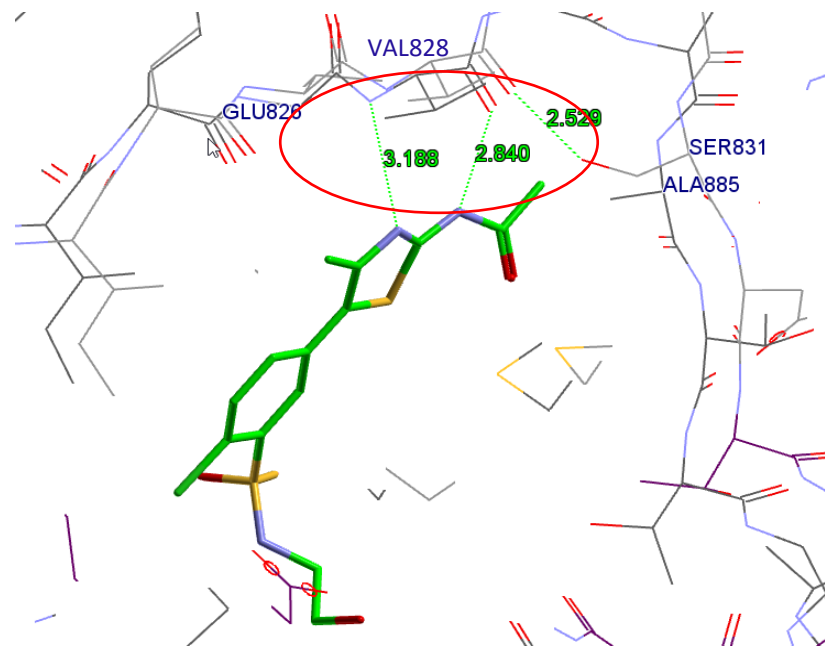


Figure1: Hinge region of protein kinases.



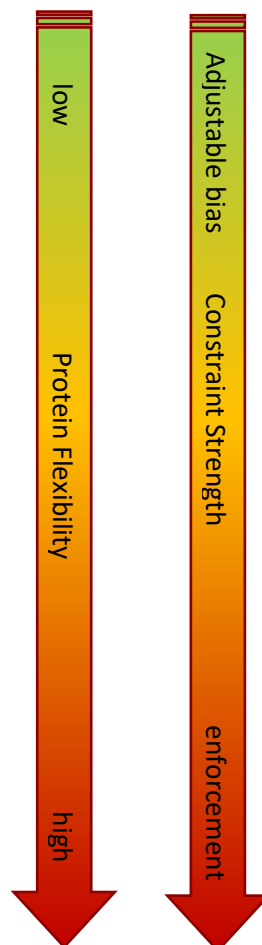
## Overview of Flexibility and Constraint Options in GOLD

Before we set up our docking, let's look at our chosen flexibility options in the context of all possible choices for flexibility and constraints within GOLD.

### Protein Movement Options

- Default:
  - results in: sampling of all conformational states of the ligand + terminal amino and hydroxyl groups in the protein
  - suitable for: docking into structures where we do not expect any major movements in the protein binding site
- Docking using soft potentials
  - requires: user to treat side chains with soft potentials
  - results in: minor clashes between docked ligand and protein side chains to be tolerated
- Flexible sidechains
  - requires: user to select flexible sidechains
  - results in: sampling of degrees of freedom
  - good to know: small main chain movements can be included
  - suitable for: sidechain rotations
- Ensemble docking:
  - requires: two or more protein conformers
  - results in: selection of the most favourable conformer
  - suitable for: large sidechain and/or backbone movements

**We will choose the flexible sidechain option in GOLD for our example and Protein HBond Constraints for our docking.**



### Constraint Options

- Default: poses determined by scoring function
- Protein H-bond: bias towards solutions involving specified protein donor/acceptor atom, individual weights can be set.
- Hydrophobic region: requires hydrophobic atoms to occupy specified 3D coordinates.
- Interaction motif: similar to Protein H-Bond constraints, but allowing C-H...O interactions to be specified as well.
- Distance: bias towards solutions satisfying a distance between protein and ligand atom, weight can be configured for each individual combination.
- Substructure based distance: similar to distance, using a ligand substructure match rather than a single atom.
- Hydrogen bond: constrain a protein-ligand atom-pair to selected protein atom to form an HBond.
- Similarity: biasing the docking towards the H-bond donor pattern, H-bond acceptor pattern or shape of a known ligand.
- Scaffold: enforce exact overlap to the position of a ligand fragment template.

## Edit Your Protein

**Note:** In Hermes 1.8.3 and previous versions of Hermes:

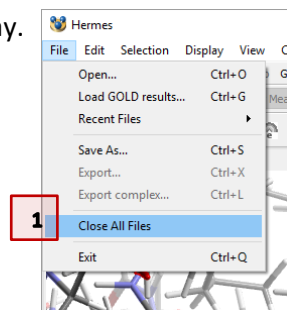
- Constraints can only be correctly set when starting the GOLD preparation process with a fully edited molecule in *.mol2* file format.

Use [Introduction to Hermes](#) to perform the following sequence of preparation steps:

1. If you have not yet done so, load *2wxr\_catalytic\_domain.pdb* into Hermes
2. **Edit > Add Hydrogens** (4166 H-atoms should be added)
3. Delete water molecules
4. Save your protein as *2wxr\_catalytic\_domain.mol2*

## Clearing the Hermes buffer

1. **Close All Files** in the Hermes display.



## Visualisation

Since the hydrogen atoms are hard to visualise on a white background,

- Click **Display > Display Options** and select a non-white background colour to control your work.

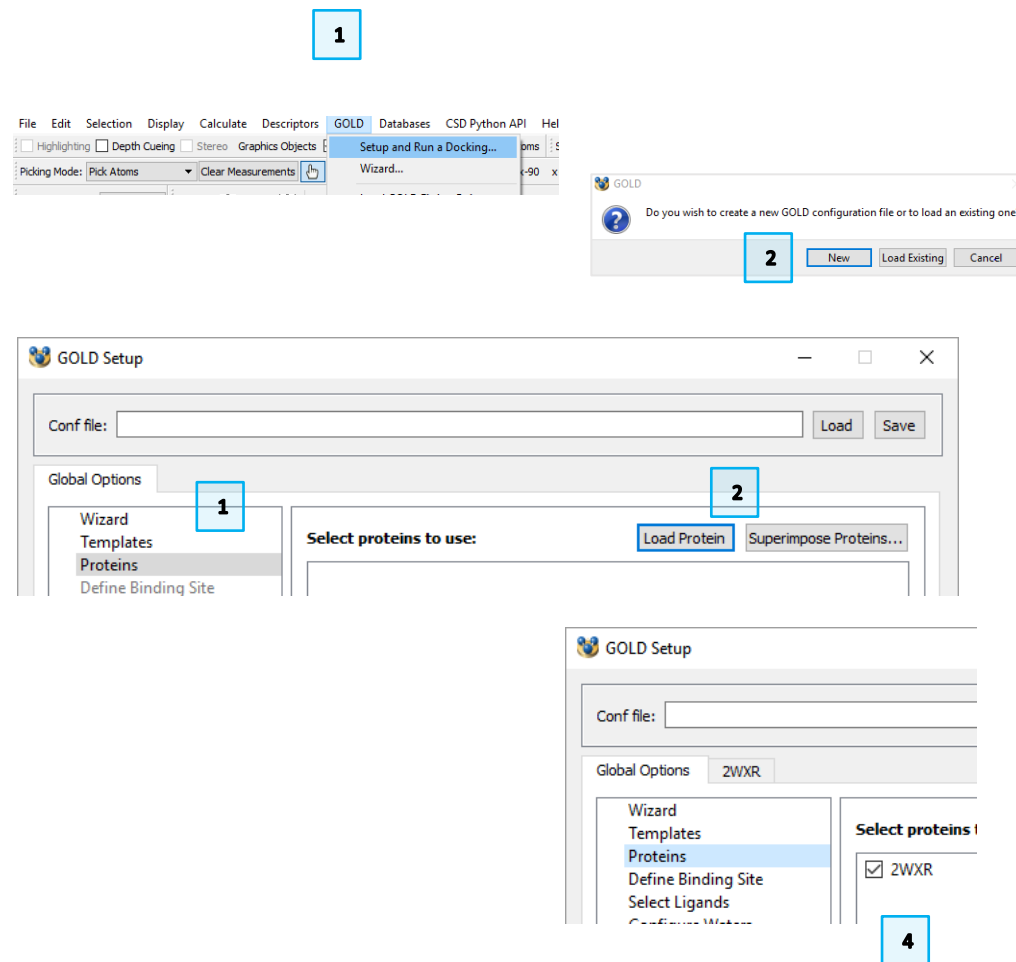
## Prepare Your Active Site

### Launch the GOLD Wizard

1. Click **GOLD > Setup and Run a Docking**.
2. Click **New** in the dialogue box to enter the *GOLD Setup* menu. You will enter the visual interface to GOLD.

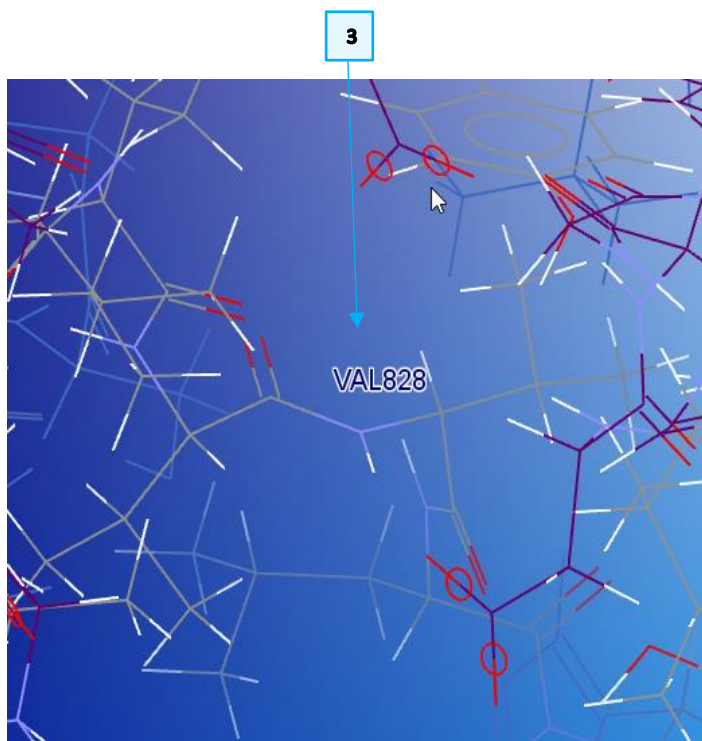
### Load your protein

1. Click on *Proteins* in the *Global Options* tab.
2. Click **Load Protein**.
3. Select *2wxr\_catalytic\_domain.mol2* from your example folder.
4. The protein will be automatically chosen as your input protein.



## Define Your Binding Site from a Protein Atom

1. In *Global Options* of the GOLD setup menu, click *Define Binding Site*.
2. Click the **Atom** radio button.
3. Click on the N atom of Val828 in the Hermes 3D view.
4. *N, VAL828* will appear in the text field.
5. Leave the selection radius at the default value of 10 Å.
6. Select the **Detect cavity** tick-box.

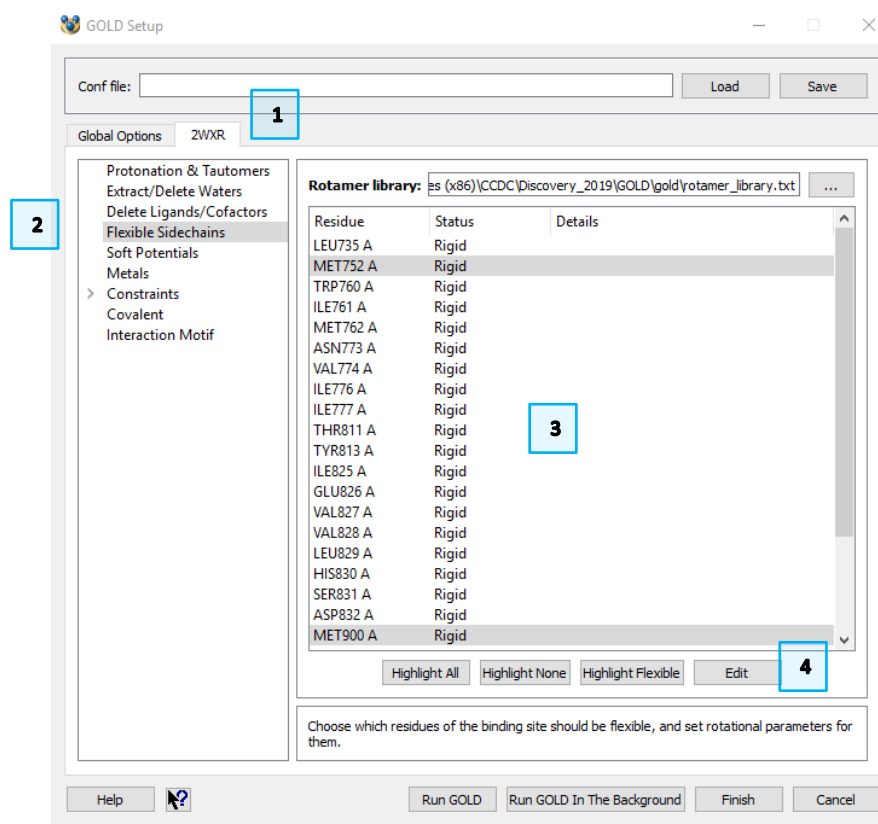
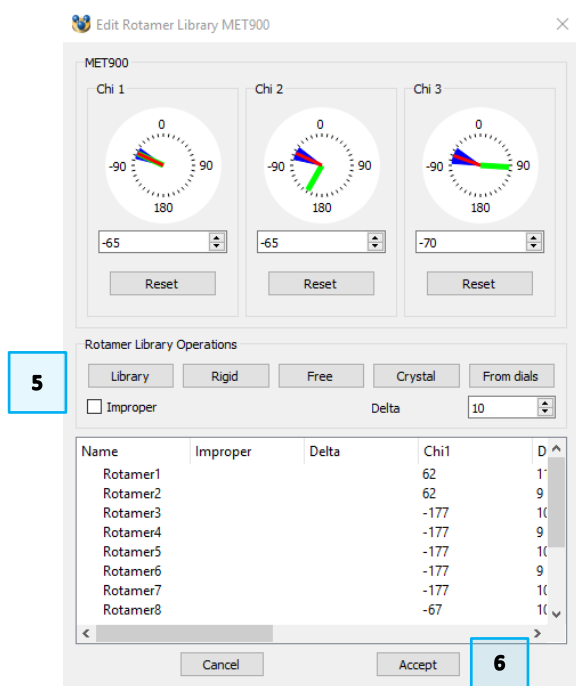


The GOLD Setup dialog box is shown with the 'Global Options' tab selected. The 'Define Binding Site' sub-tab is active. The 'Atom' radio button is selected, and 'N, VAL828' is entered in the text field. The 'Detect cavity' checkbox is checked. The 'Select all atoms within' field is set to 10 Å. The 'Generate a cavity atoms file from the selection' checkbox is unchecked. The 'Force all H bond donors/acceptors to be treated as solvent accessible' checkbox is checked. The 'Add Definition as a Selection' button is visible. The 'Run GOLD' button is highlighted.

## Specify Flexible Sidechains

To make Met752 and Met900 residues flexible:

1. Click on the 2WXR tab.
2. Click on *Flexible Sidechains*. A display of the sidechains in your active site will appear.
3. **Ctrl + Click** on *MET752 A* and *MET900 A* from the residues list.
4. Click on **Edit**.
5. In the *Edit Rotamer Library MET900* dialogue, click on **Library**. This will bring up a rotamer library with the option of further editing values from the dials.
6. Click **Accept** to set the docking to the selected rotamer values.
7. Repeat steps 5 and 6 for *Met752*.

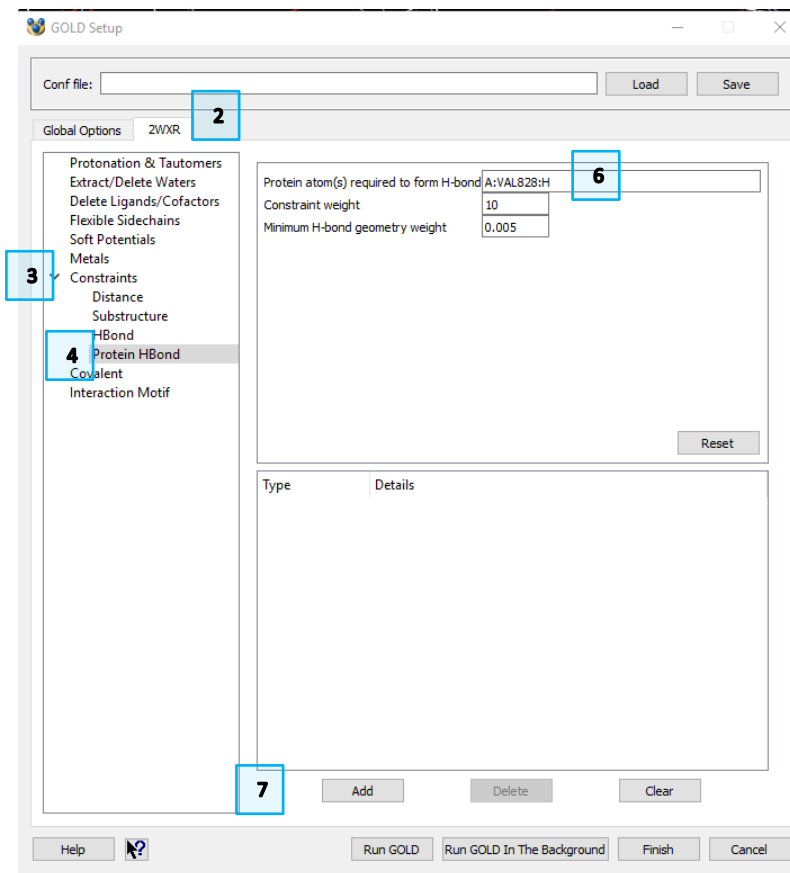




## Set a Protein HBond Constraint

We want GOLD to focus on solutions where the ligand makes the interaction to N-H of Val828 discussed in the “[Setting Constraints Based on a Homologous Protein Structure](#)” section of this tutorial. To achieve this:

1. Center on and label the Val828 in the 3D view.
2. Click on the 2WXR tab of *GOLD Setup* window.
3. Click on the “>” or “▶” symbol next to “Constraints” to open the list of available constraints.
4. Select *Protein HBond* from the list.
5. In the 3D view, click on the hydrogen atom (not N) of the Val828 NH group.
  - It will be highlighted in the 3D view.
  - A:VAL828:H will appear in **Protein atom(s) required to form H-bond** field box.
6. Leave the **Constraint weight** at the default value of 10.
7. Click on **Add** to enter the constraint to your configuration. Constraint details will then appear.



## Consider Alternative Protonation/Tautomeric States for Your Protein and Ligands

### Potential variations in protein and ligand protonation and tautomeric states

When a ligand binds to a protein, protein and ligand will find a protonation/tautomeric state that provides ideal complementarity. Your docking should provide sufficient coverage of protein and ligand protonation and tautomeric states to allow for an ideal match to be identified.

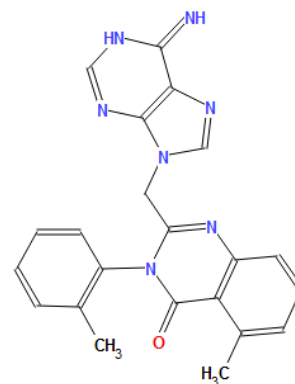
### Assumptions used in your example

- Glu826 in the active site is negatively charged.
- Only low energy tautomers of ligands are used.

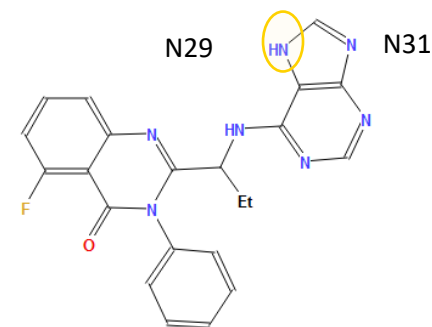
In the case of Ligand 1, tautomers corresponding to the imino form of adenine are much higher in energy than the input molecule. Therefore, we will not use them in this tutorial.

For Ligand 2, the hydrogen on N29 can move to N31. To generate this tautomer for docking:

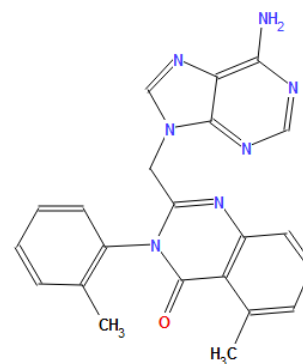
- Load *40L\_ideal.sdf* into Hermes.
- Use **Edit > Edit Structure** from the Hermes top-level menu to:
  - Click on **Atoms & Bonds** from the *Remove* panel of the *Edit Structure* window to remove the hydrogen from N29. Click on the hydrogen atom in the 3D view.
  - Click on **Hydrogen Atoms** from the *Add* panel of the *Edit Structure* window and then click on N31 in the 3D view to add a hydrogen (H18).
  - Set the **Bond Type** to **Single** in the *Edit* panel and click on the bond between C30 and N31.
  - Set the **Bond Type** to **Double** and click on the bond between N29 and C30.
- Save the molecule as *40L\_t2.mol2*



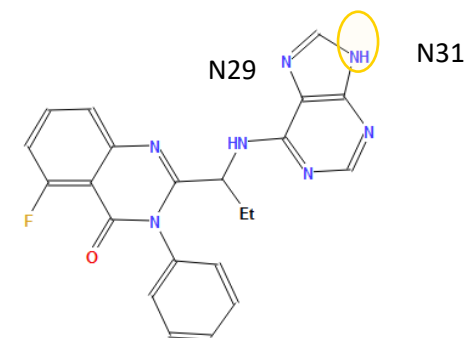
Ligand 1



Ligand 2 (idelasilib)



Ligand 1: rare imino form



Ligand 2 (idelasilib), alternative tautomer

## Enter Your Ligands

Use *Global Options > Select Ligands > Add* to pick Ligand 1 (IC8.mol2), and the two tautomers of Ligand 2 (40L\_t1.mol2, and 40L\_t2.mol2) as ligands for the docking.

## Set Fitness and Search Options

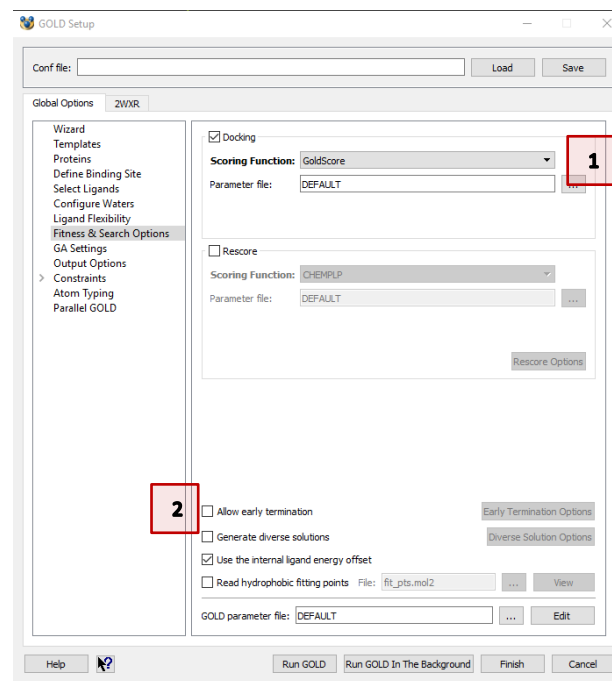
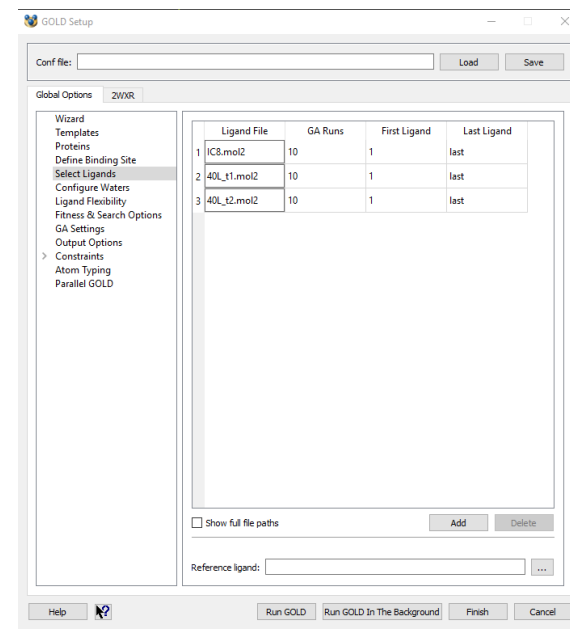
Click on the *Fitness & Search Options* to define the docking settings

1. Choose *GoldScore* as your docking function.
2. Unselect **Allow early termination** tick-box.

## Specify Output, Save Your Setup and Invoke GOLD

- In *Output Options*, specify the output folder, such as *flexibility*.
- Press **Run GOLD**: in the resulting dialogue, choose a name for your *.conf* file, such as *flexibility.conf*.
- Press **Save** to start your docking.

Please refer to the [GOLD Pose Prediction MLL1](#) document, page 13, for more detailed information on those steps.

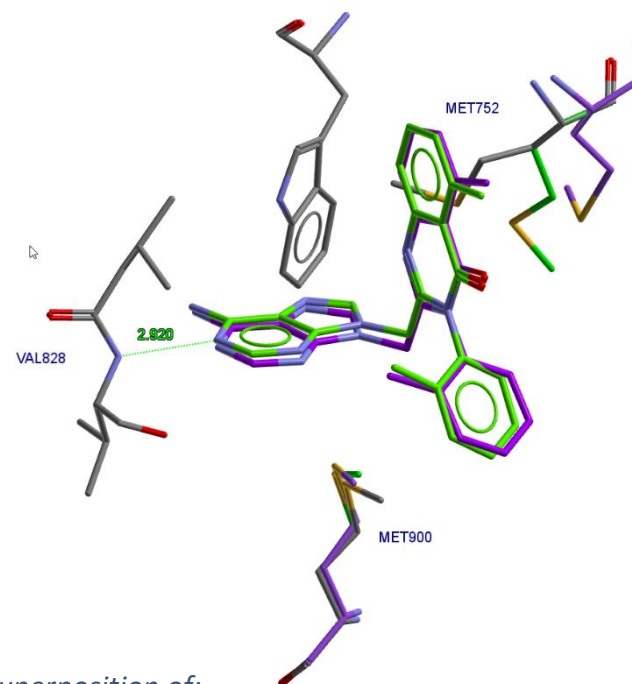




2010

### Compare Ligand 1 Docking Results to the Co-Crystal Structure, 2x38

- Load the results of your docking.
- Visualise the poses for ligand 1, IC8 in *Molecule Explorer*. Refer to the [GOLD Pose Prediction MLL1](#) document for more information on how to visualise docking results in *Hermes*.
- Load the co-crystal structure of Ligand 1 complexed to PI3Kdelta, provided as *2x38\_aligned\_to\_2wxr.mol2*. This structure was published in 2010.
- Due to the stochastic nature of GOLD's genetic algorithm, the output values differ between each GOLD run, including the exact values for the lengths of hydrogen bonds.
- You should be able to:
  - identify a good agreement of the top scoring pose with the experimental co-crystal structure
  - notice that Met752 has moved to accommodate Ligand 1
  - predominantly see solutions with Hbonds involving the –NH group of Val828.
- To compare with the protein starting conformation, load *2wxr\_catalytic\_domain.mol2*.



Superposition of:

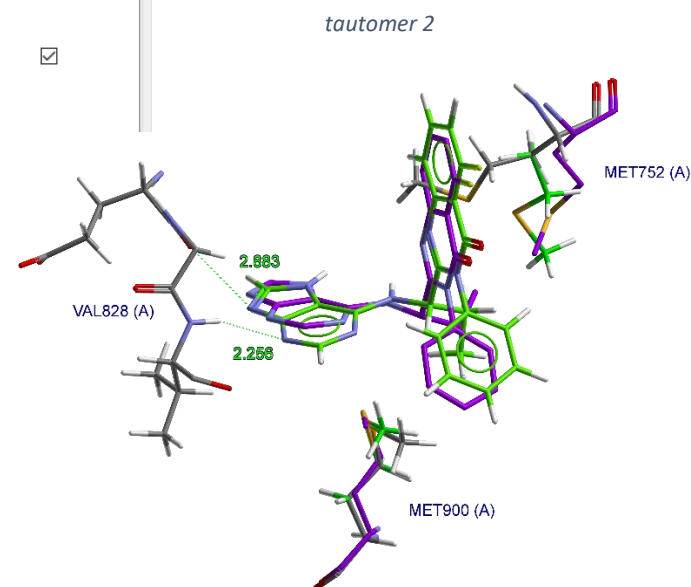
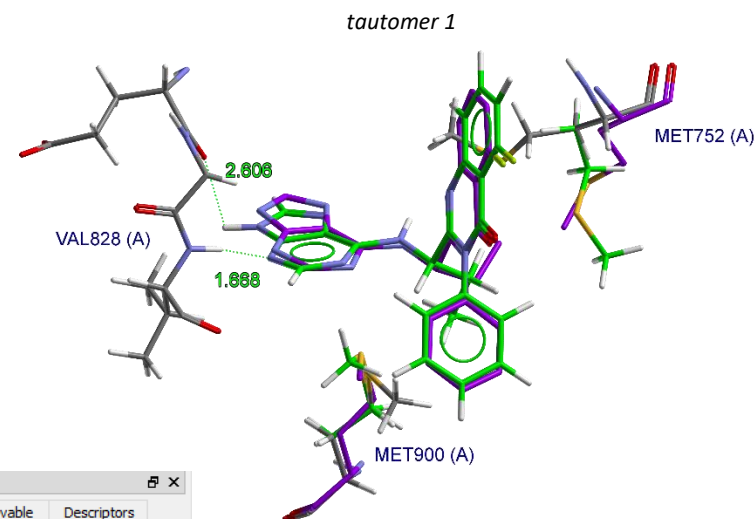
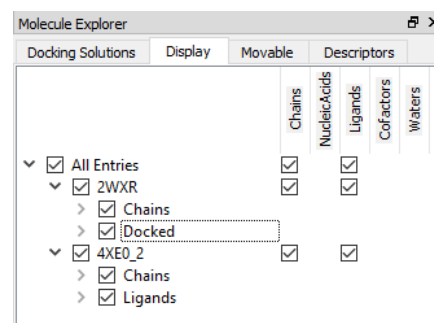
- Docked Ligand 1 (green): top scoring pose of ligand and flexible Met752 and Met900 sidechains
- 2WXR apo structure (grey): including starting positions for Met752 and Met900
- 2X38 co-crystal structure (purple)



## Compare Ligand 2 (Idelasilib) Docking Results to the Co-Crystal Structure, 4XE0

- First, clear all files from Hermes using **File > Close All Files**
- Load the docking results.
- Visualise the two sets of poses for the tautomers of ligand 2 (40L\_t1 soln\* and 40L soln\*, which are the solutions for 40L tautomer2).
- Load the co-crystal structure of ligand 1 complexed to PI3Kdelta, provided as *4xe0\_aligned\_to\_2wxr.mol2*. This structure was published in 2014.
- Restrict the display to structures relevant to ligand 2 (40L\*), using *Molecule Explorer*.
- Expect to see:
  - Near correct solutions to be identified for both tautomers
  - Similar agreement with the co-crystal structure for both tautomers, where tautomer 1 forms a weak Hbond between adenine N6 and the –C=O group of Glu826.

Congratulations! You have found the correct pose of the drug idelasilib starting from a protein apo structure!



### Superposition of:

- *Docked ligand 2 (green): top scoring pose of ligand and flexible Met752 and Met900 sidechains*
- *2WXR apo structure (grey): including starting positions for Met752 and Met900*
- *4xe0 co-crystal structure (purple)*