Tutorial 1: Manual fitting of ligand and quick evaluation

(Platform: Windows 10)

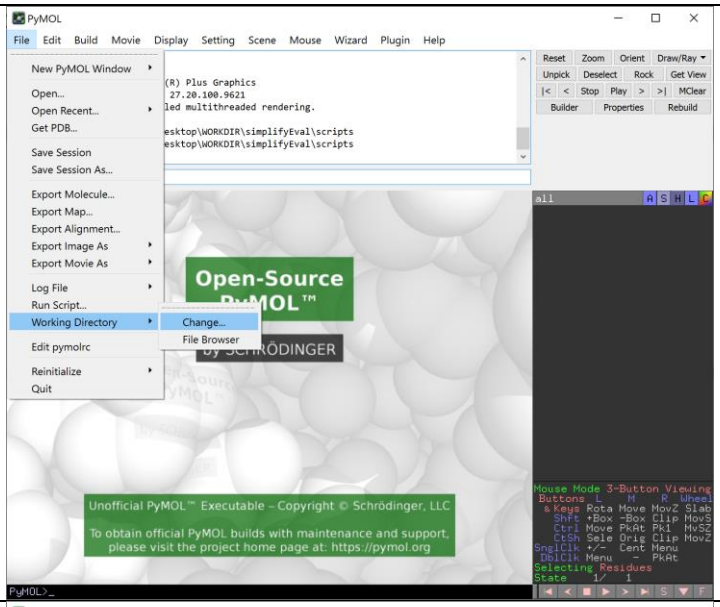
- All the scripts descripted here can be found in the sub-folder of your corresponding OS/conda installation within the “scripts” folder, or in the “box-local” sub-folder.
- You may need to copy the scipts to another folder for evaluating the binding pose. (See the instructions below)
- A spreadsheet viewer of your choice to work on .csv output would be useful.
- Currently limited support for non-windows users (only in the “summary” session). Inclusion of UNIX instructions in the detailed tutorial is still in progress.

Ex1: Warm-up exercise

Re-docking of ligand CCO10 in FTO and evaluation of new ligands modified form it, and limits of fast computer models (expt. ref: https://pubs.acs.org/doi/10.1021/acs.jmedchem.1c01204)

1. Assuming you have followed the installation procedures in the previous tutorial and closed PyMOL afterwards: start PyMOL by double-clicking "XXX-pymol.bat”, where XXX is anaconda-local or miniconda-local.

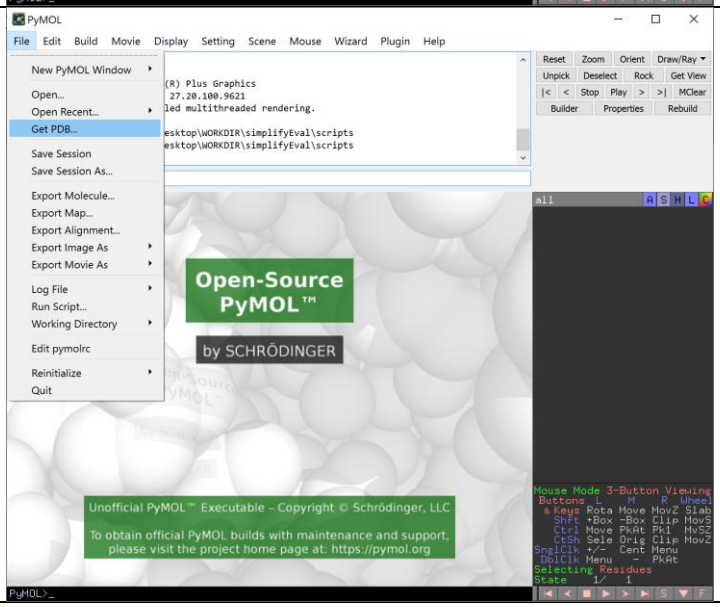
2. (*optional*) change the *working directory* make it easier to monitor all the files downloaded and generated. Create a new folder if you need.



3. Download the experimental PDB structure (requires internet connection)

   a. File -> Get PDB

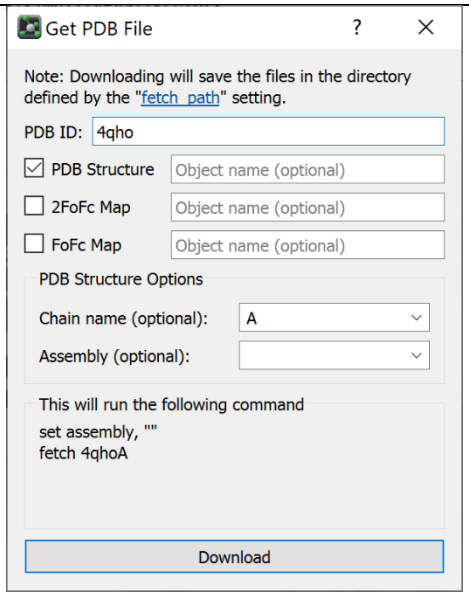
*Remember to check for **missing** or **mutated** residues in the targeted region! PDB2PQR server would only help to model missing atoms, but not a chain of missing residues with unmodelled backbone position, or restoring mutated sites..*



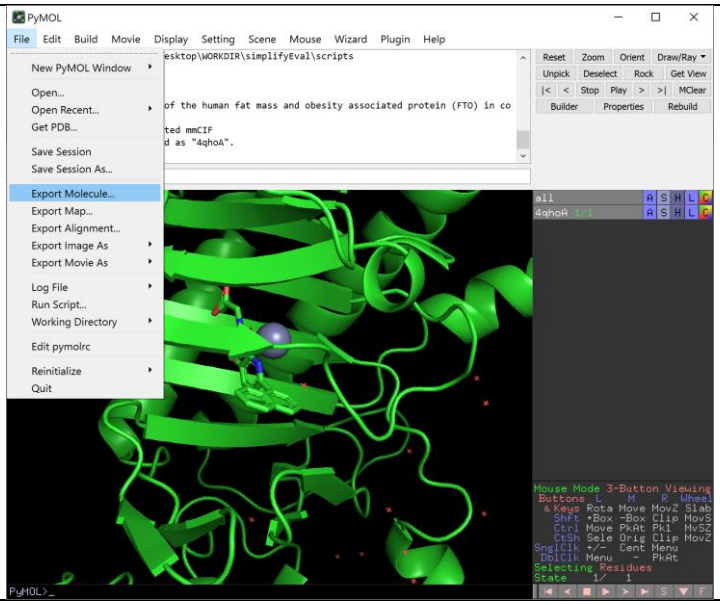
b. Fill the form according to the screenshot, then click download.

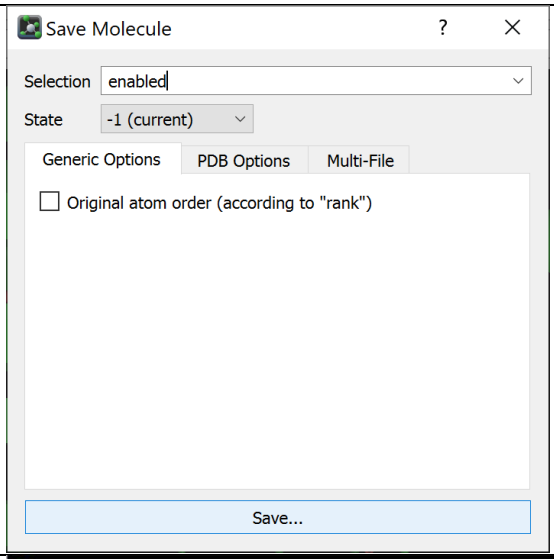
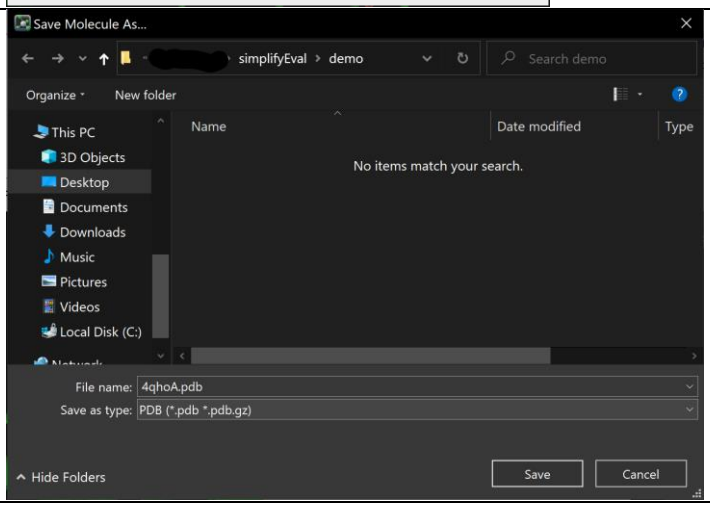
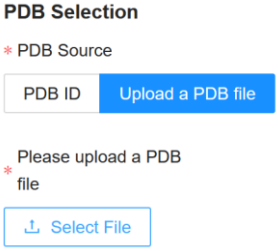
Before you decide on which PDB structure to use in your study, you can just visit the website of the database at https://www.rcsb.org/ and search for the deposited information of the protein, check the quality of the structure (e.g. resolution) and depends on your purpose what ligand is complexed with the protein.

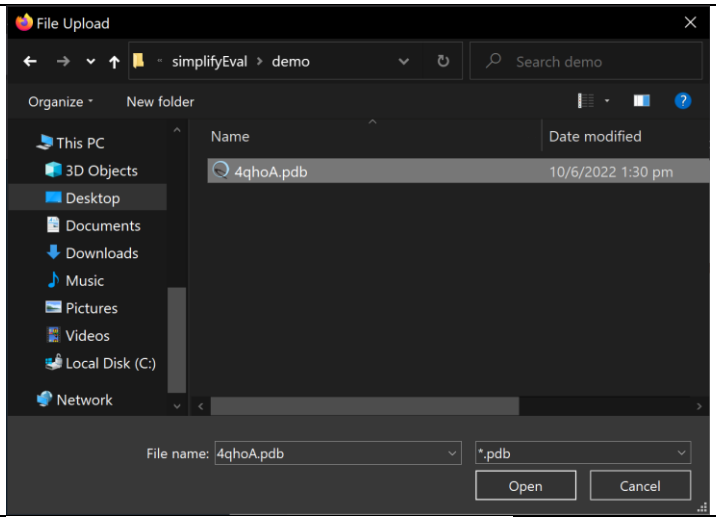
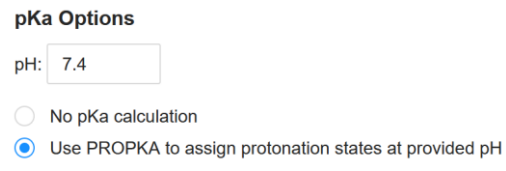
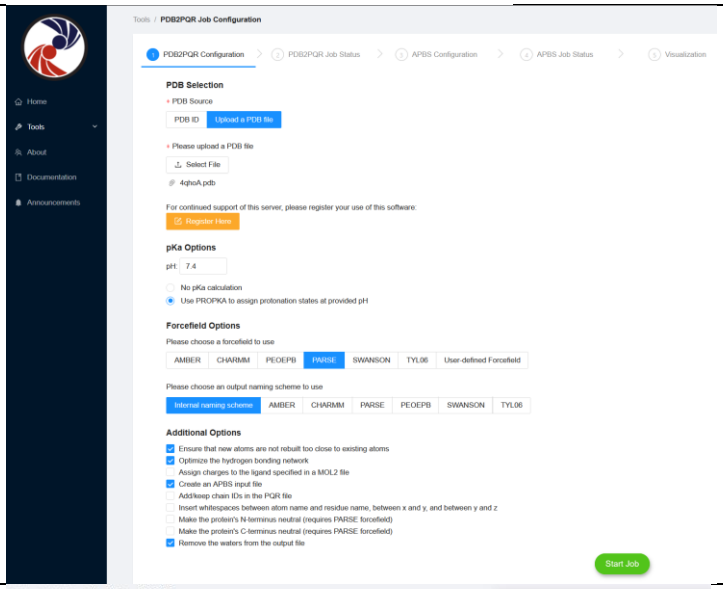
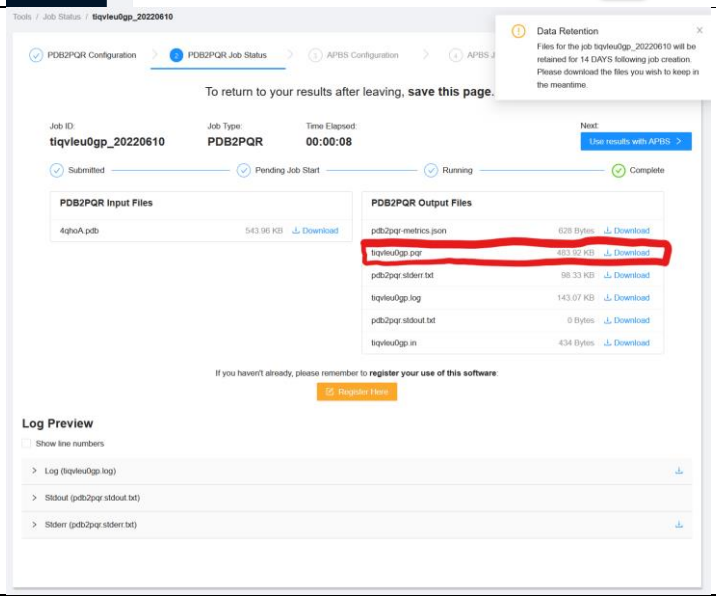
UniProt database https://www.uniprot.org/ is another useful resource, it contains a lot of background information of your protein of interest and provide links to the publicly available experimental/model structures.

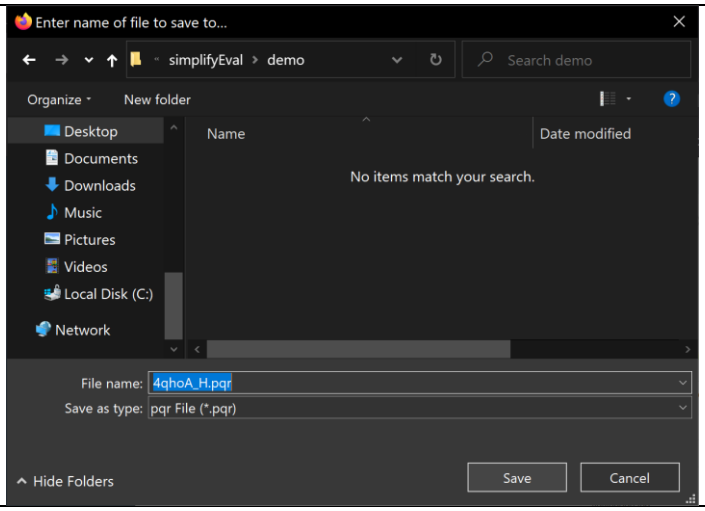
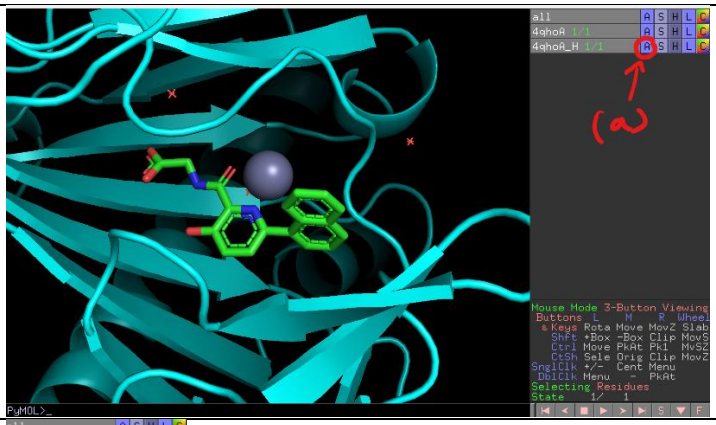
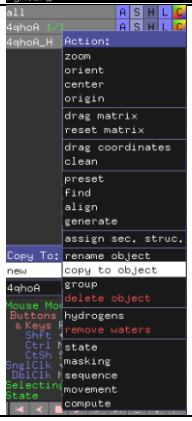


4. Save the file for later use.
   a. File -> Export Molecule
   b. You may also save the PyMOL session to avoid accidents.



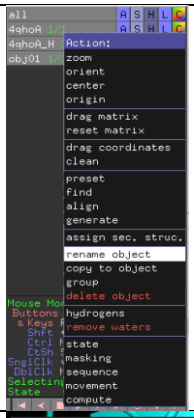
c. Sine we only have one object loaded in PyMOL, simply click “save”.	
d. Give it a name and save as type PDB. Then click “Save”. e. You are highly recommended to save the PyMOL session again too, either through the GUI or by pressing Ctrl+S.	
5. We will use the pdb2pqr server[1] to add hydrogens to our receptor. (requires internet connection) a. Open a browser and visit https://server.poissonboltzmann.org/pdb2pqr	
b. Under “PDB selection”, click “Upload a PDB file”. Then click “Select File”.	

c. Select the pdb file you just saved, then click “Open”.	
d. Under” pKa options”, type 7.4 in the pH box (or whatever pH you want). Other parameters can be kept as default.	
e. Under “Additional Options”, make sure “Ensure that new atoms are not rebuilt too close to existing atoms” is selected. (Should be the default settings) f. Click “Start Job” and wait.	
g. Once the job is finished, download the .pqr file generated.	

h. Give it a proper name and save it in your working directory.	
6. Return to PyMOL and open the .pqr you downloaded, e.g. by drag-and-drop to the viewer or via File -> Open. Now you have two objects loaded in pymol.	
7. Since the PDB2PQR server remove all HETATM entries from the PDB file to produce the .pqr file, you need to copy the non-protein residues you want back into the protonated structure. a. Make a copy of 4qhoA_H to be safe: Near the object name of "4qhoA_H" at the right panel, click "A".	
b. Select "copy to object", then "new".	

c. Rename the new “obj01”:

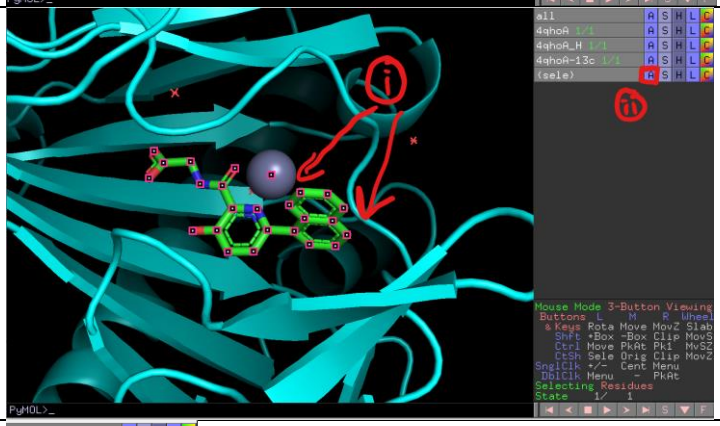
- click “A” next to “obj01”, then “rename object”



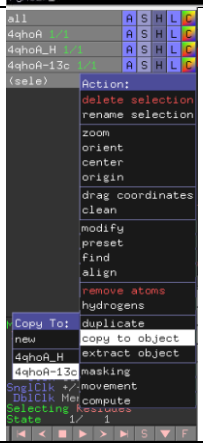
- use backspace of your keyboard to delete obj01 then type “4qhoA-13c”
- press Enter.

d. copy the metal ion and ligand to “4qhoA-13c”.

- Click the metal and the ligand
- Near (sele) object, click “A”



- “copy to object” -> “4qhoA-13c”

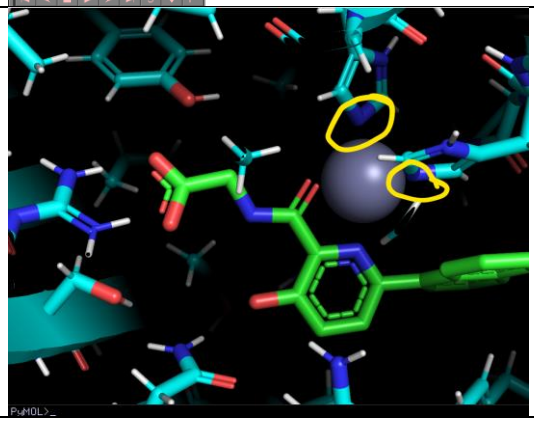


8. Object “4qhoA” should be disabled for now but not “4qhoA_H”. Click on “4qhoA_H” to disable it. Now you only have “4qhoA-13c” displayed in the viewport.

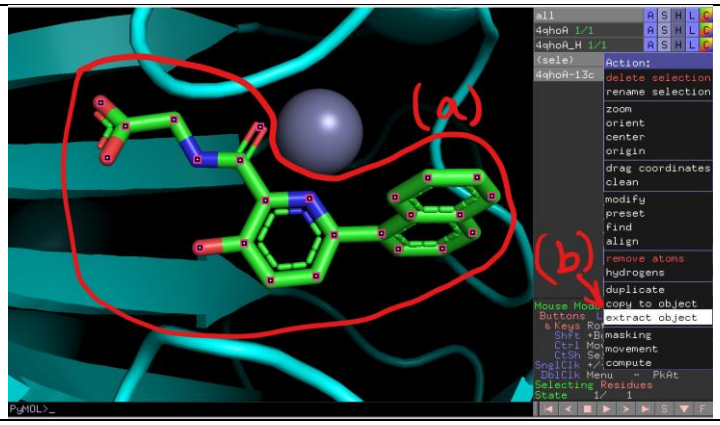
9. Show the atoms of the protonated receptor and check the hydrogens in the active site.
   a. Click “S”

   b. Select “licorice”

   c. Move around with your mouse. You should carefully investigate the protonation state of especially the two histidine residues (yellow circle) that coordinate with the metal ion.

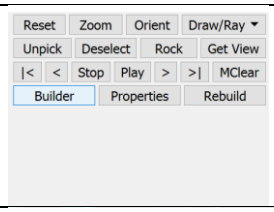


10. Now the receptor is prepared, we need to add hydrogens to the ligand.
    a. To be safe, extract the ligand first. Click on the ligand
    b. Near the (sele) object: click “A” -> extract object

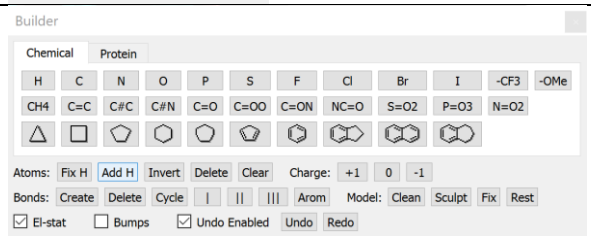


c. Rename obj01 to 13c_H similar to step (7c).

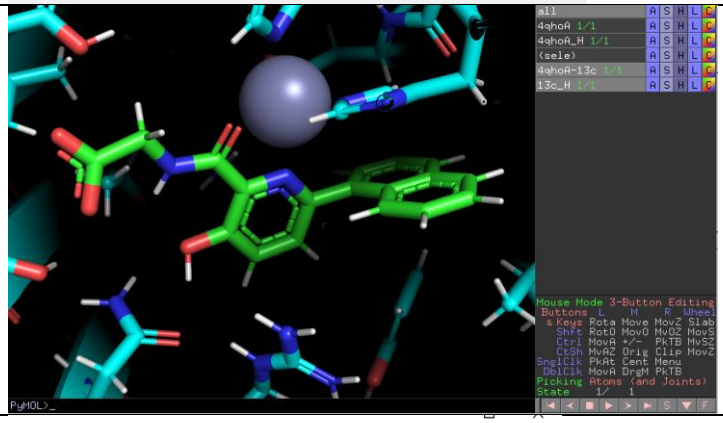
d. In the top-right panel, click “Builder”.



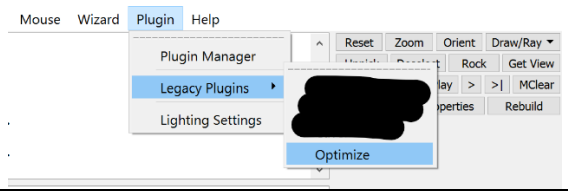
e. Click “Add H” in the Builder window.



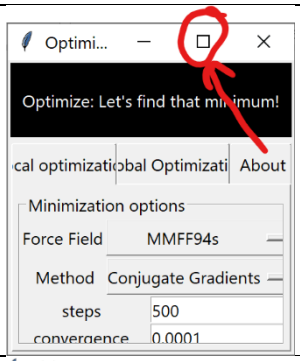
f. Then click on the ligand. Now you have a protonated ligand.
- Check carefully to see if the hydrogens were added correctly.
- You are highly recommended to save the PyMOL session again here, either through the GUI or by pressing Ctrl+S.



g. Now we need to optimize the ligand, i.e. fix its bond length and angles and remove steric clashes WITHIN the ligand itself.
- Plugin -> Legacy Plugins -> Optimize

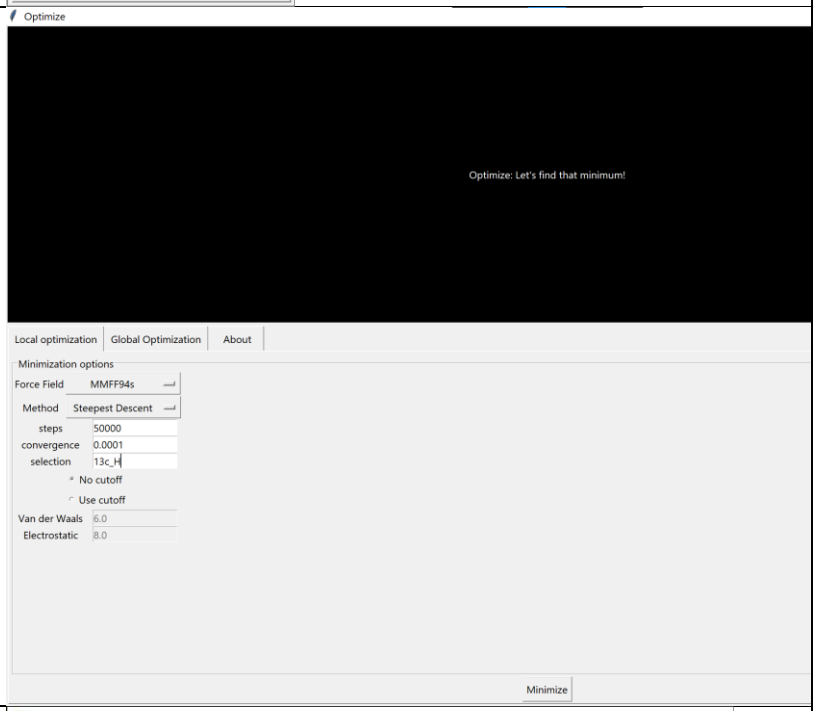


- Depends on your computer display, the Optimize window may be too small as shown in the screenshot. For windows users, DO NOT DRUG AND RESIZE IT, there is a bug that PyMOL would crash if you do so. Instead, maximize the window.



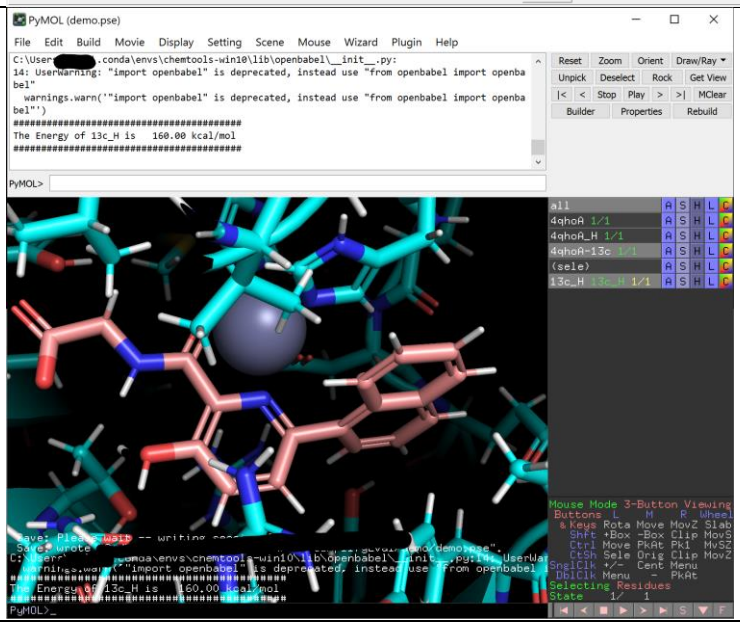
- Choice of minimization options really depends on user preference. For me I would use the "steepest descent" method and 50000 steps.
- IMPORTANT: choose the correct object every time you do minimization. This time you want to minimize "13c_H". Type that in the "selection" box.
- Click minimize. It will take a few second.

Q&A:
Q: *Why don't we just optimize the whole complex here?*
A: *The current plugin cannot handle that many atoms, plus doing so would be problematic for especially some older computers.*

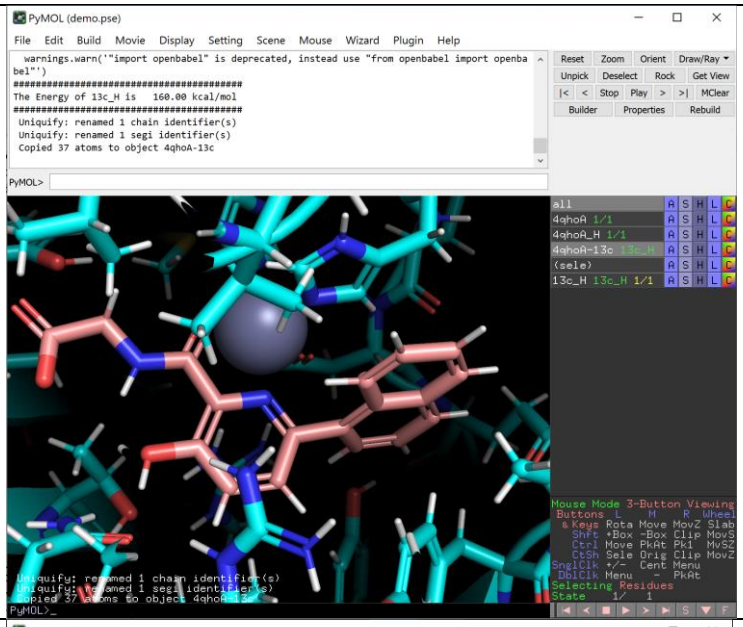


- Go back to the main PyMOL window. You will see the MMFF94s energy of the target object once the minimization is done. Don't worry if the ligand color is different, it is normal.

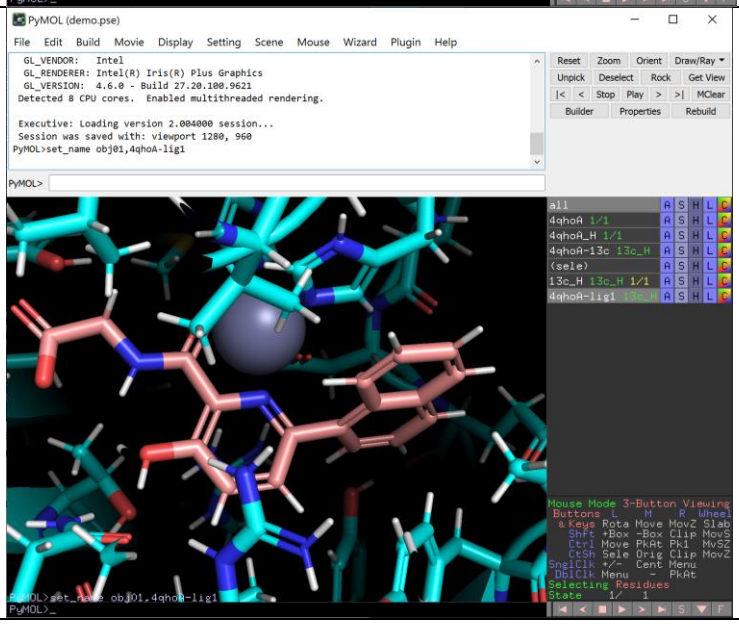
11. (the warning message above can be safely ignored)



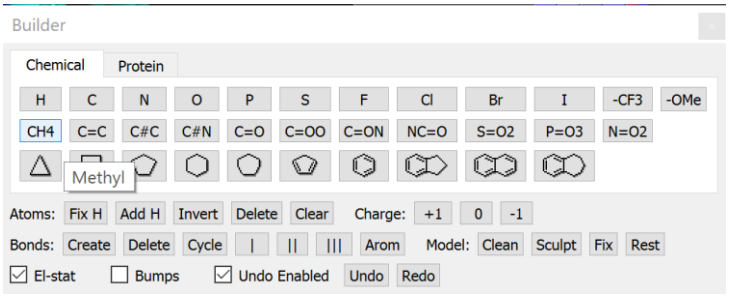
a. Copy the ligand back to object “4qho-13c”. now we have the protein-ligand complex in the same object, both protonated, and the protonated ligand has been optimized.
b. Optional but it is also good to save your PyMOL session again.

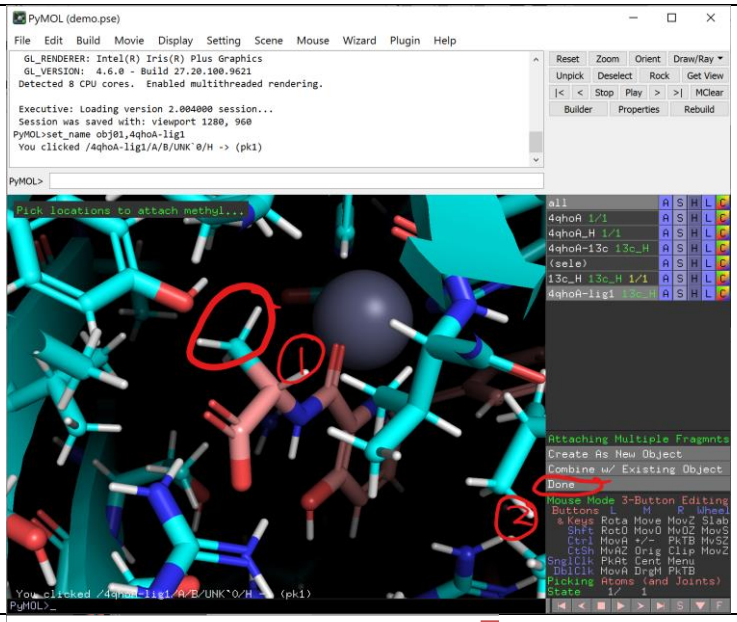


12. We now have our reference system. Lets move on to make a new ligand!
a. Make a copy of the “4qhoA-13c” object, and rename the new “obj01” to something like “4qhoA-lig1”.
b. Click on object “4qhoA-13c” to disable it to prevent accidental modifications to this object.



13. Open the Builder again. Modify the ligand as you would like to. For example, add a methyl group to turn the glycine substructure into an *L*-alanine by clicking “CH4” in the builder then click on the position you would like to modify. Click “done” in the right panel once you are done.

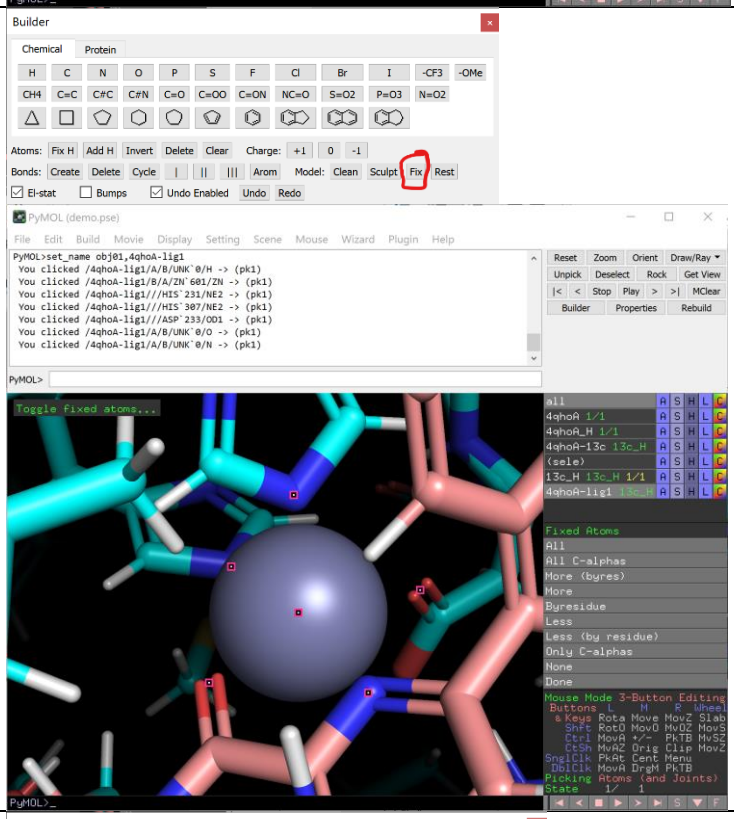




14. You may see that the methyl group is too close to the -OH of a nearby residue. Let's relax the structure a bit and see how it fits there!

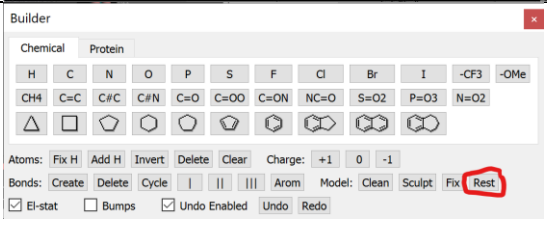
a. We need to keep some desired interactions, i.e. the metal coordination.

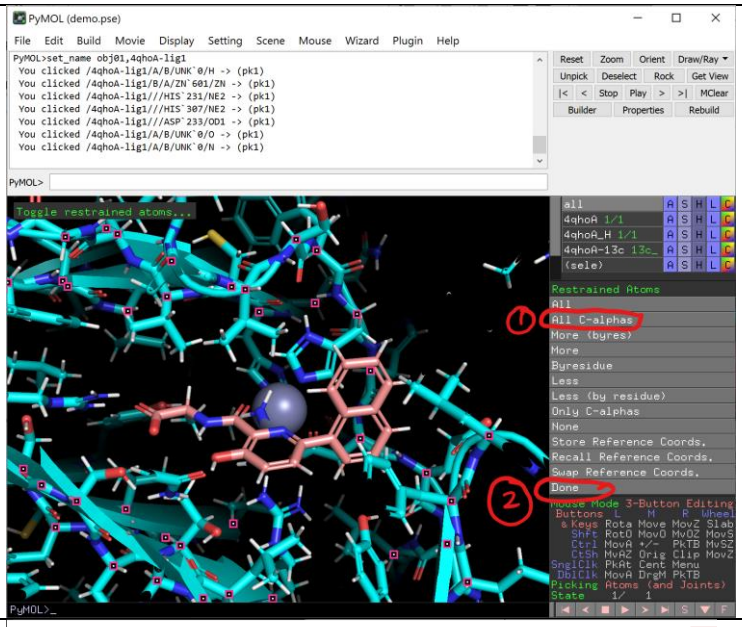
- In the builder window, click "fix". Then click on the 6 atoms at the metal center as shown in the screenshot.
- Click done when you are finished.



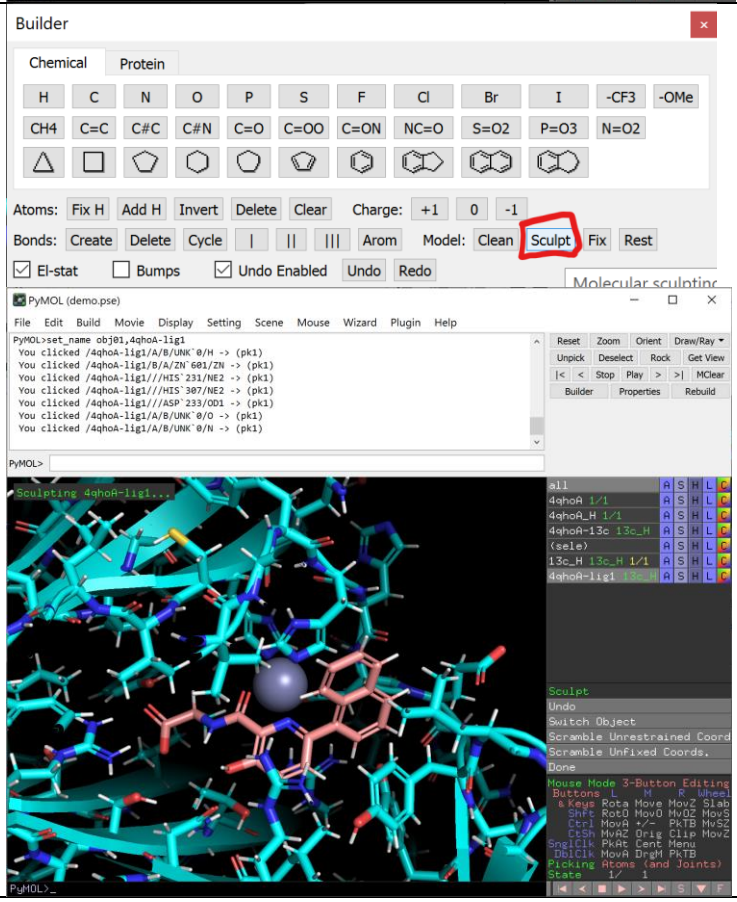
b. (Optional, depends on protein) To avoid the protein to move too far from the starting structure, it is also recommended to restrain the protein backbone.

- In the Builder window, click "Rest".
- Click "all C-alphas", then "Done".





c. Click “Sculpt”. Then wait until the structure stop moving.
d. (optional) if you want you can also drag the ligand with Ctrl+Left Click to let it adapt a pose you want.
e. Once you are happy with the roughly estimated pose, click “Done”.



15. This already give a rough estimation of the new ligand's binding pose.

16. However, the sculpting tools just use a rough consideration of steric factors, which we will discuss at the end of this demo.

17. Other than that, the tool has another issue, that the bond length of e.g. non-protein heteroaromatics estimated by the PyMOL sculpting tools is sometimes unrealistic.

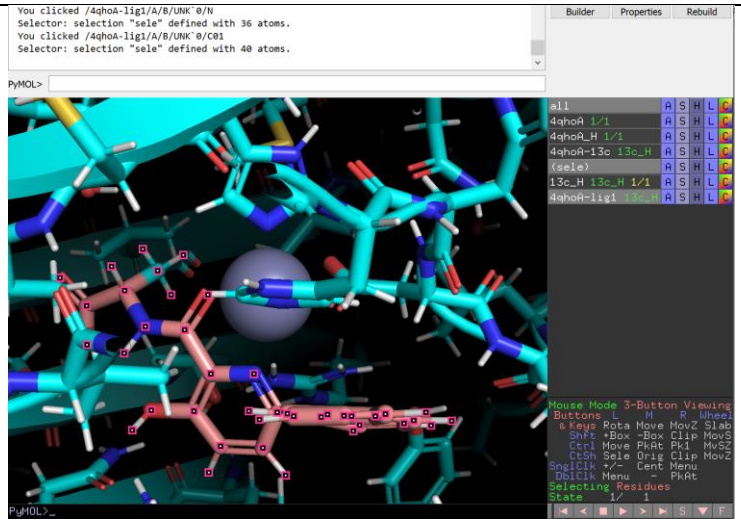
18. Thus, we better optimize the ligand again and use other tools to predict the binding pose.

a. Leave the "editing" mode by left-clicking on the bottom-right panel to enter the viewing mode. Then click on the ligand.

b. The methyl group may need to be selected separately. Make sure you have selected the whole ligand!

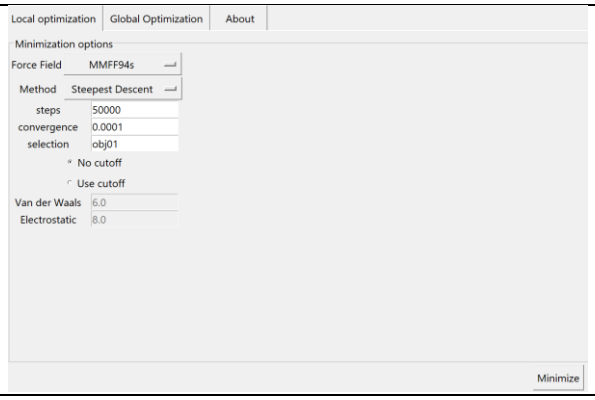
c. Next to the (sele) object, click "A", then "extract object".

d. (optional but recommended) Save the PyMOL session at this point.

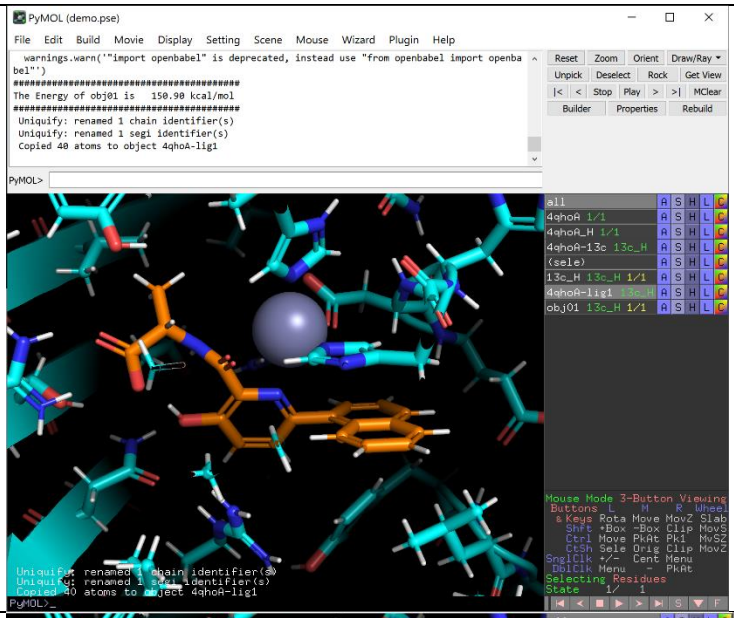


e. Start the Optimize plugin again if it is closed, then change the settings, as shown in the screenshot. Input the object name of extracted ligand in "Selection".

f. Click Minimize and wait for a few seconds.



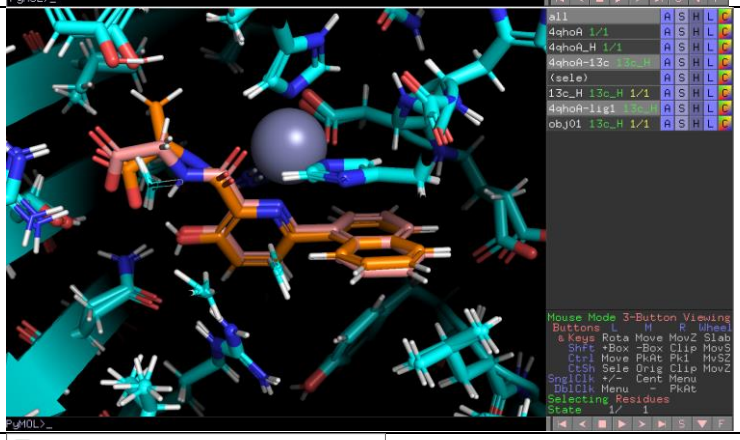
g. Copy obj01 back into “4qhoA-lig1”: “A” -> copy to object -> 4qhoA-lig1



19. Now we have a rough estimate of the coordinates of two protein-ligand complexes. Save them as PDB for docking/scoring!

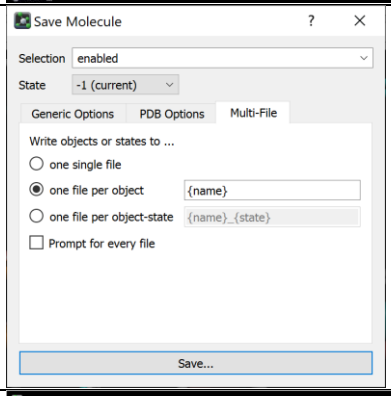
a. To be fast, enable only the complexes you want to save. This is particular useful when you have a lot of structures to save.

b. File -> export molecule.



c. Go to the “Multi-file” tab and do the same as the screenshot.

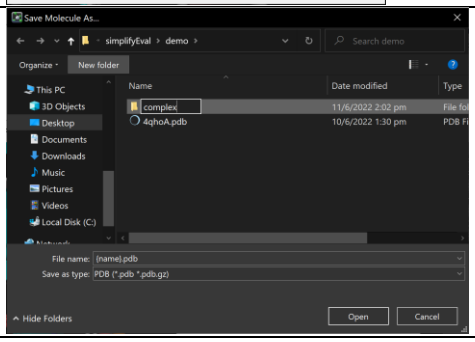
d. Click “Save…”

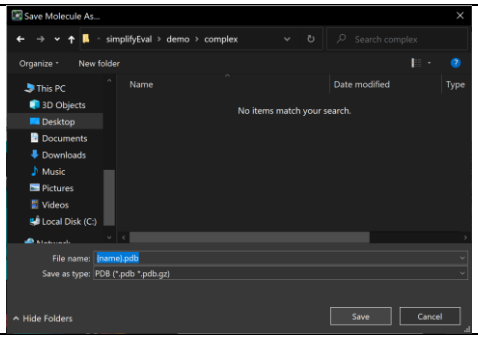
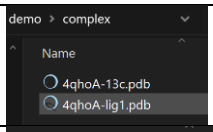
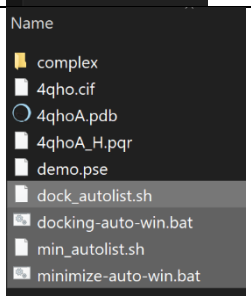


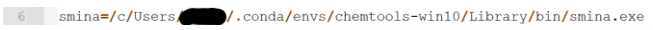
e. IMPORTANT: Create a new folder name “complex” and select it as the output location!!!

f. Make sure that the output folder does not have any other PDB files.

g. Click save.



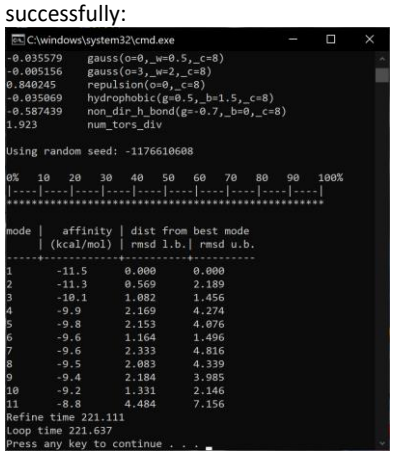
<table>
<tr><td></td><td> 
</td></tr>
<tr><td>h. Minimize or close PyMOL.</td><td></td></tr>
<tr><td>20. Open the folder you have just created. You should have the two complex you just saved.</td><td> 
</td></tr>
<tr><td>21. From the “scripts” directory, there is a “box-local” folder containing necessary scripts to do the calculation. Copy the four highlighted items in the screenshot to the folder contained the “complex” folder you just created. Now, your working directory should be like this.</td><td> 
</td></tr>
<tr><td>22. IMPORTANT: edit the PATH to smina executable in the .sh! Right-click to open them in a text editor. The .bat files may also need to be corrected.

See “demo0-install.pdf” for possible fix of common PATH issues in .bat scripts.</td><td> 
</td></tr>
<tr><td rowspan="2">23. Double-click on either minimize-auto-win.bat or docking-auto-win.bat and follow the instructions.

a. “docking-auto-win.bat” predict the ligand pose in the given site, while “minimize-auto-win.bat” use the provided pose and optimize the ligand conformation in the site and give the optimized pose a score.

b. To reduce computational time required for calculation, only the ligand would be allowed to be flexible, that the receptor would be kept rigid.

c. Depends on your hardware, “docking-auto-win.bat” will take a few minutes or more to finish.

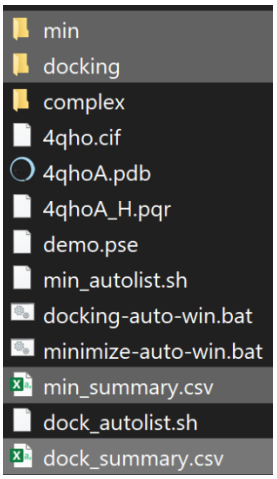
d. Once the calculation is completed, you will see something like the screenshot in the terminal output.</td><td>Sample terminal output when “minimize-auto-win.bat” finished its job successfully:
 
</td><td>Sample terminal output when “docking-auto-win.bat” finished its job successfully:
 
</td></tr>
<tr><td colspan="2">By default, the docking box defined in the scripts provided in “box-local” directory has been set to be 2A around the ligand to prevent the ligand to bind too far away from the manually fitted position. You may edit the value of the “autobox_add” options in *.sh scripts to change the settings.</td></tr>
</table>

e. Press any key to close the program.

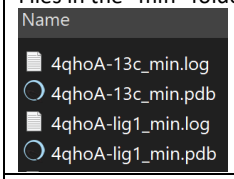
24. For each of the provided complexes, (each of) the programs would generate a .pdb output and a .log output in a new output folder called “min” and “docking” respectively.

25. Score of the top-ranked pose would also be summarized in a summary.csv file. A more negative score means a higher predicted affinity.

Your working directory should look like this after both calculations have finished. Files generated by the programs were highlighted in the screenshot.



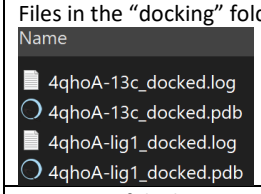
Files in the “min” folder.



Contents of min_summary.csv opened in excel.

	A	B
1	complex	score
2	4qhoA-13c	-11.1587
3	4qhoA-lig1	-11.4425

Files in the “docking” folder.



Contents of dock_summary.csv opened in excel.

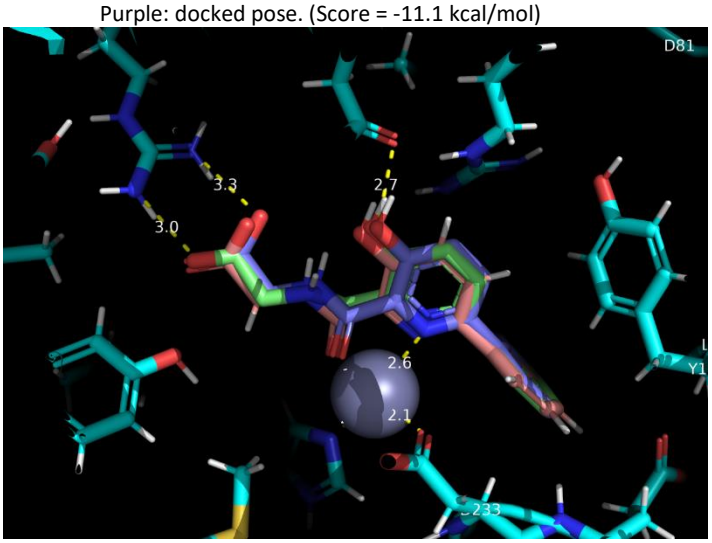
	A	B
1	complex	score
2	4qhoA-13c	-11.1
3	4qhoA-lig1	-11.5

26. Lets see if the program could reproduce the experimental structure of FTO-CCO10 (13c) first. Go back to PyMOL. If you have closed it already, start it again and open the previous session you saved. (File -> open and choose the .pse file of your previous session)

27. Open the output .pdb files located in “min” and “dock” folder.

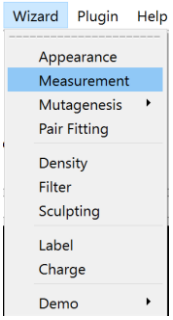
a. As we compare the binding pose from experiments and calculations, we can see that the three conformations shown in the figure are nearly identical.
Don’t worry if the non-polar hydrogens are missing in the docking and minimization output, the software use a “united atom” representation that merged $CH_{n=1-4}$ atoms to speed up calculations.

(a) **Re-docked pose of CCO10 (13c) in FTO**
Pink: experimental pose.
Lime green: minimized pose. (Score = -11.2 kcal/mol)
Purple: docked pose. (Score = -11.1 kcal/mol)

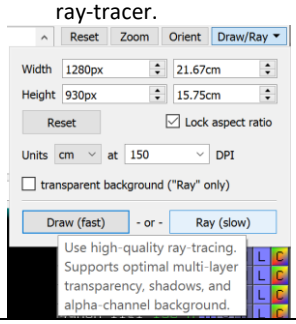


b. You can have a better idea of how the ligand interacts with the protein with PyMOL's measurement tool. (Wizard -> Measurement)
c. It would also be useful to label the protein residues for clearer visualization. (in the right panel of PyMOL, click "L" -> "residue (one letter)")
d. High-quality molecular graphics can be generated with PyMOL's internal ray-tracer

(b) The measurement tool in PyMOL.



(d) Generating high-quality image with PyMOL's internal ray-tracer.



28. Let's move on to check the binding pose of lig1. Open the corresponding .pdb files located both the "docking" folder and "min" folder. We can see that the poses predicted with both strategies were highly similar to our initial model from sculpting.

At the first glimpse, these conformations SEEMS TO BE reasonable, and the estimated affinities were higher than our lead-compound CCO10. Can we conclude here that lig1 is a better FTO inhibitor?

Taking a closer look to the predicted conformation, two major issues can be observed:

a. *Lig1's carboxylate being pushed away from the arginine that we want it to form hydrogen bond with.*
b. *As we introduce the methyl group and get the initial pose with sculpting, the hydroxyl group tyrosine residue having close contact with it must flip away from the experimental determined pose, towards surrounding hydrophobic residues to fit the ligand in there, which would be unfavorable.*

29. Accuracies of computational models varies with different molecular systems. Predictions were made with many assumptions. Hence human interpretation, visual inspection of modelled conformations, follow-up dynamical studies and experimental validation are critical to successful drug-design schemes.[2]

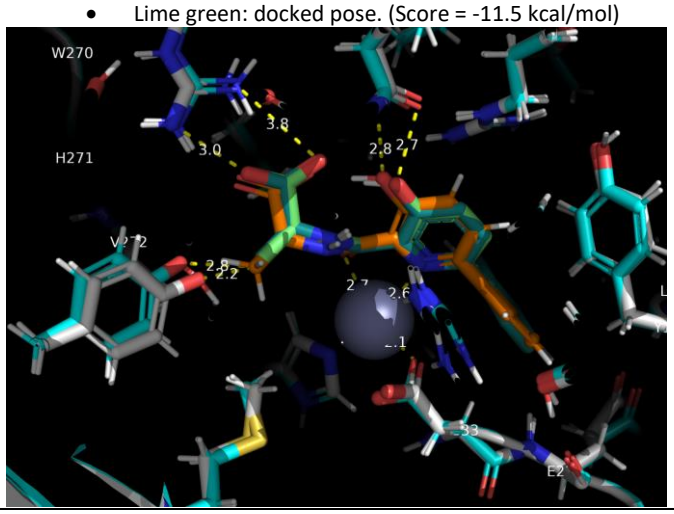
(a) Predicted pose of lig1 in FTO

Receptor:

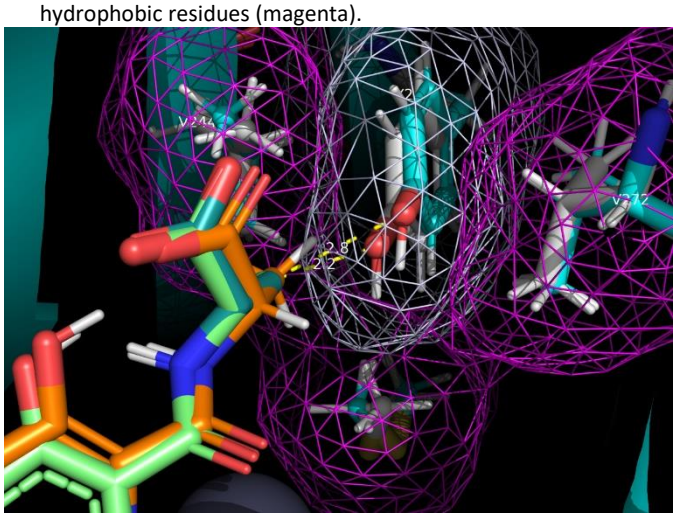
- White: Experimental conformation
- Cyan: Conformation after sculpting in PyMOL

Ligand:

- Orange: Pose from sculpting.
- Olive green: minimized pose. (Score = -11.4 kcal/mol)
- Lime green: docked pose. (Score = -11.5 kcal/mol)



(b) Molecular surface representation of the tyrosine that has close-contact with ligand methyl group (blue-white) and surrounding hydrophobic residues (magenta).



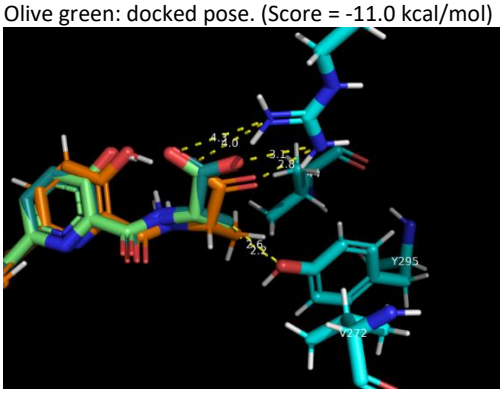
30. Try to evaluate the ligand in experimental receptor

Predicted pose of lig1 in FTO's experimental structure
Orange: Pose from sculpting.
Lime green: minimized pose. (Score = -10.9 kcal/mol)

conformation again and see the difference!
Hint: copy and rename 4qhoA-13c, remove the ligand, copy lig1 into it, export the complex as pdb, then redo the docking/minimization.

An example of the resulting binding pose is shown in the screenshot. With the unflipped tyrosine blocking the methyl-binding position, it is not surprising that a poorer affinity has been predicted this time.

Olive green: docked pose. (Score = -11.0 kcal/mol)



31. Is lig1 better or worse than CCO10 (13c)? We don't know yet, given that this is just an approximate model. More advanced computational technique and experimental validation will be required to provide additional information.
32. Still, we may interpret the results from this exercise this way:
    - Replacing the glycine substructure of CCO10 with an *L*-alanine may improve the binding by providing increased interaction surface area for VdW interaction with Met297, Val244 and Val272.
    - However, the methyl group is close to the sidechain of a Tyr295.
        - The hydroxyl group of that Tyr295 need to move towards a hydrophobic surface to fit the methyl group there. Given the close contact of the tyrosine's hydroxyl group and both the ligand's hydrophobic part and surrounding residue's apolar side chain, such interaction would be unfavorable.
        - To prevent the *L*-methyl group to clash with protein surface, the ligand's carboxyl group need to tilt away and have a less optimal orientation for hydrogen bond formation with Arg316, that the interaction would be less favorable.
    - Hence, we may assume that lig1 would have a lower tendency to adapt an ideal binding mode we want.

33. **Practical usage of this method:**
    - Recommended usage:
        - Minimization scripts
            - Provide a crude initial model with minimization (e.g. removal of clashes and unfavorable interactions) for a specific binding pose of ligand modelled from knowledge-based design
            - In cases that the "usual" method cannot provide a desirable interaction mode (e.g. ligands being too flexible).
        - Docking scripts
            - Evaluation of a small set of ligands in a certain receptor conformation.
            - Would be too slow for a slightly larger set of ligands! Use the scripts and workflows in later tutorials for some more efficient protocols!
    - Remarks:
        - The sculpting tool in PyMOL is fast but cannot provide a model good enough for direct scoring, e.g. having obvious issues with the bond length involving certain atoms. Thus, for practical usage, it is better to use the complex from **Step 30** instead of complexes from previous attempts!
        - To allow receptor movement, let the docking/minimization software decide instead, the tools would use a scoring function to evaluate the movement and wouldn't change the provided bond length (Try the scripts with "flex" keyword)
        - Optimizing the **ligand** before evaluation or prediction is important since the commonly used fast docking software would only alter the torsions, but not the lengths and angles of the bonds, in the fitting and evaluation process.

Cheat Sheet of the "Manual Fitting" Workflow

1) Select the receptor (and reference ligand, can be from the same PDB entry)
2) Start PyMOL
3) Download the specific chain(s) from the RCSB PDB (e.g. via PyMOL)
4) Check for any missing/mutated residues within the interested region
   a) PDB headers
   b) Visualize in 3D
   c) Corresponding journal article
5) Prepare receptor with the pdb2pqr server
6) Restore essential cofactors to the protonated receptor
7) Reference complex preparation
   a) Extract the reference ligand from co-crystal
   b) Protonate the reference ligand from co-crystal
   c) (optional) re-combine the receptor-ligand complex to remove steric clashes of the protonated receptor/ligand
      i) Combine the protonated ligand with the protonated receptor
      ii) In "builder" fix receptor coordinates
      iii) Sculpt to remove steric clashes due to the new hydrogens
      (Note the inaccuracies in ligand geometry!)
      iv) Extract the ligand again for geometry optimization
   d) Optimize the geometry of protonated reference ligand
   e) Combine the processed receptor and ligand to make the reference complex
   f) Save the reference pdb to "complex" folder
8) Preparing a model of the "new" complex
   a) Copy the reference complex to a new object (copy 1)
   b) Fix receptor coordinates
   c) Modify the ligand
   d) Use the sculpting tool to fit the new ligand by hand
   e) Extract the new ligand
   f) Optimize the geometry of the new ligand
   g) Make another copy of the reference complex (copy 2) and give it a name
   h) Delete the ligand in copy 2 of the reference complex
   i) Copy the optimized new ligand to the unliganded copy 2 of reference complex
   j) Save the complex as pdb in the "complex" folder
9) From the provided "scripts/PLATFORM/box-local" folder copy the scripts to the directory containing "complex" folder
   a) Quick minimization only: "min-autolist.sh" *(+ "minimize-auto-win.bat" for windows)*
   b) Quick minimization with flexible receptor: "min-flex-autolist.sh" *(+ "minimize-flex-auto-win.bat" for windows)*
   c) Docking with the gridbox defined by ligand: "dock-autolist.sh" *(+ "dock-auto-win.bat" for windows)*
10) check the directory structure
11) **Make sure there is no space in all the filename!!!**
12) *(**Windows only**, option: using double-click scripts):*
   a) Edit the path to conda environment in the .bat scripts
   b) Edit the smina variable in the .sh scripts
   c) double-click "XXX-auto-win.bat" and follow the instructions
13) (Non-windows users or command-line option for windows):
   a) open a terminal
   b) activate the conda environment with smina installed
   c) cd to the directory with the scripts, then type:

```
bash xxx-autolist.sh #where xxx is min, dock, min-flex etc depends on the method you are using
```

14) Open the results for visualization and analysis

Further Discussion

34. In this exercise, we have tried to predict the conformation of a FTO inhibitor by manually fitting it into the experimental structure, score it with computational methods and compare the results with that of a known inhibitor.
    - While the "manual fitting" with the current tools gives a quick estimation of the potential interaction mode of protein-ligand complexes, it is less efficient for evaluating the "penalties" of doing this move.
    - Speed of doing so would also be terribly slow with a large dataset to be evaluated.
35. Thus, using the rigid receptor from experiments and allowing the ligand to be flexible is one of the most common approaches for the high-throughput preliminary screenings of potential binders, that a "unrealistic" "undesirable" binding mode usually represent less-probable binders with potential steric clashes with receptor if it is "forced" to adapt an "ideal" interaction mode.

36. It should be noted that many important dynamical factors of the "reality" have been neglected as we use the rigid receptor for docking studies. Moreover, fast computational models make a lot of approximations that accuracy would be lowered to speed up the calculation. Quality of experimental models and the algorithm used in the calculation are also determining factors of model accuracy.
37. A more practical approach commonly applied to increase the success rate of drug-design schemes would be selecting a small set of compounds from docking-based ranking and follow-up with more advance but slower "dynamical" methods to re-score the complexes, which would be more efficient given a limited amount of time and computational resources.
38. To have a better model of protein-ligand interactions with the "fast" molecular docking, some people would apply multiple scoring and docking methods to evaluate the same set of compounds and judge based on the combined results. Others would allow a part of receptor to be flexible in the docking study. For those who are interested, scripts for docking/minimization with flexible receptor are also provided but be warned that flexible docking would be MUCH SLOWER compared to the current protocol.

39. Finally, as discussed previously visual inspection by human in assessing the quality of computer-modelled conformations are important to reduce false-positive and false-negative hits, regardless to the computational strategies applied in the preliminary design process.

40. We will introduce in the next demonstration the commonly applied docking protocol that let the software predict the binding pose of ligands into a pre-defined binding site, without the need of manual placement of ligand's initial coordinate, and how this could be applied to discover new inhibitors from an external molecular database.

Reference

[1] Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G.-W., Holst, M. J., McCammon, J. A., and Baker, N. A. (2017) Improvements to the APBS biomolecular solvation software suite. *Protein Science 27*, 112–128.

[2] Fischer, A., Smieško, M., Sellner, M., and Lill, M. A. (2021) Decision Making in Structure-Based Drug Discovery: Visual Inspection of Docking Results. *Journal of Medicinal Chemistry 64*, 2489–2500.