A protocol of 3D drug design in the protein pocket by artificial intelligence and virtual screening method

Qifeng Bai
Email: molaical@yeah.net
Homepage: https://molaical.github.io
School of Basic Medical Sciences
Lanzhou University
Lanzhou, Gansu 730000, P. R. China

1. Introduction

A new drug development may cost about 2.6 billion USD. However, about 90% of drugs are failure in the process of clinical trial and approval for marketing even though a lot of capital is used to drug development[1]. In this tutorial, the standard protocol of MolAICal soft package (https://doi.org/10.1093/bib/bbaa161) is introduced for the drug design of SARS-CoV-2 Mpro by artificial intelligence and molecular docking method. It will help the pharmacologist, chemists and other scientists design rational drugs according to the three-dimensional active pocket of proteins.

2. Materials

2.1. Software requirement

1) MolAICal (win64 or linux64): https://molaical.github.io
2) UCSF Chimera: https://www.cgl.ucsf.edu/chimera/

2.2. Example files

All the necessary tutorial files are downloaded from:
https://github.com/MolAICal/tutorials/tree/master/002-AIVS

3. Procedure

This step deal with protein structure for molecular docking. “model.pdb” is the optimized “6lu7.pdb”. Of course, users can choose to use “6lu7.pdb”, directly. Here, to let this tutorial completely, the dealt procedure is supplied as below:

3.1. Separate the protein and ligand structures by UCSF Chimera

1) Firstly, loading complex structures. File→Open→model.pdb (see Figure 1)

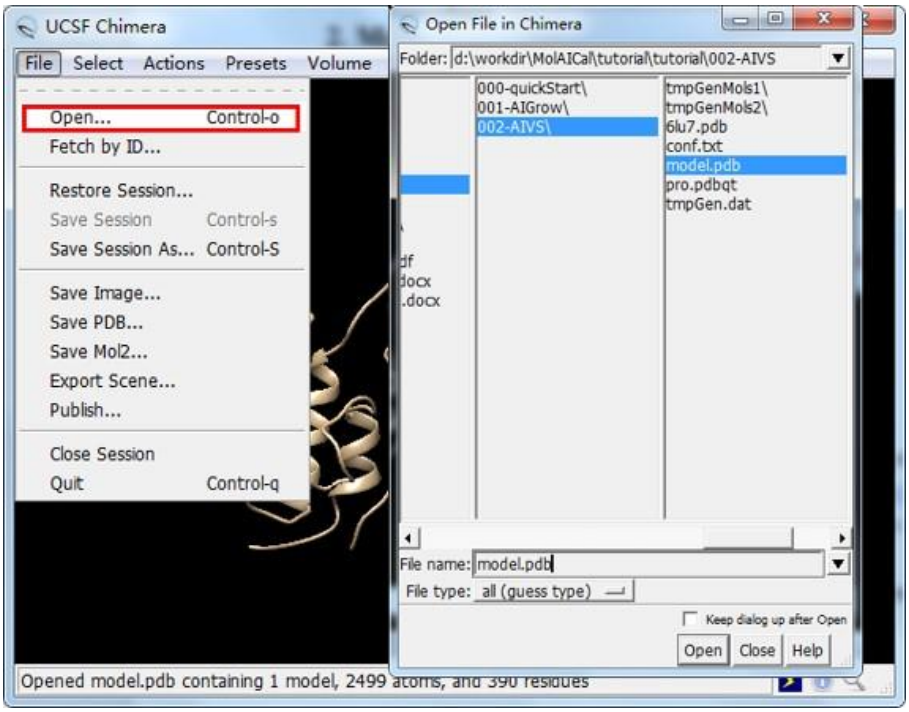

Figure 1. Load protein structure files.

2) Select ligand named LIG and delete it (see Figure 2). Using the same way in Figure 2, delete the water named HOH.

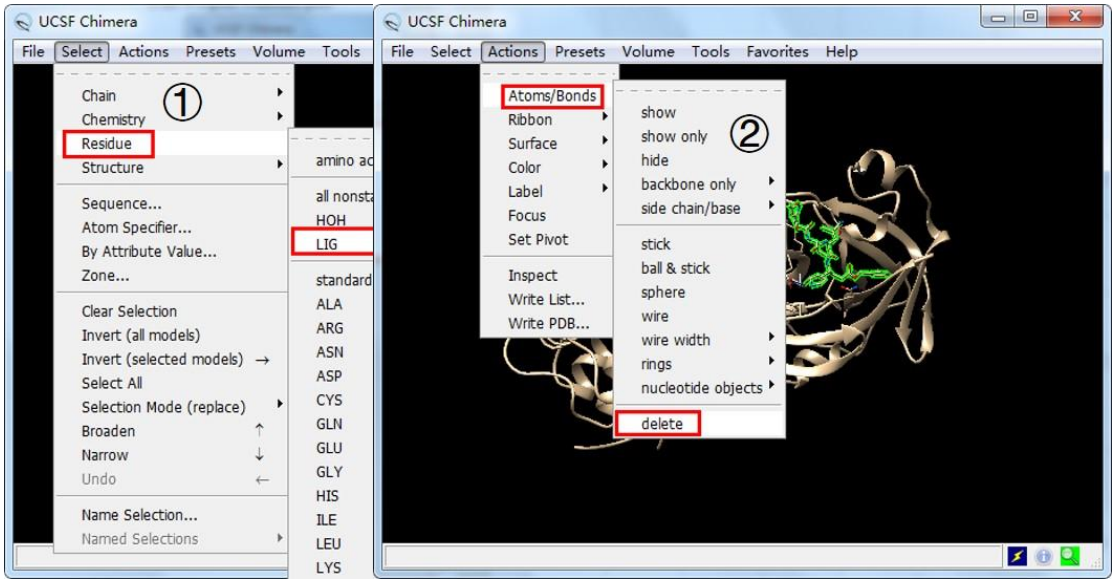

Figure 2. Select ligand and delete it

3) Save protein structure named “protein.pdb” without ligand (see Figure 3)

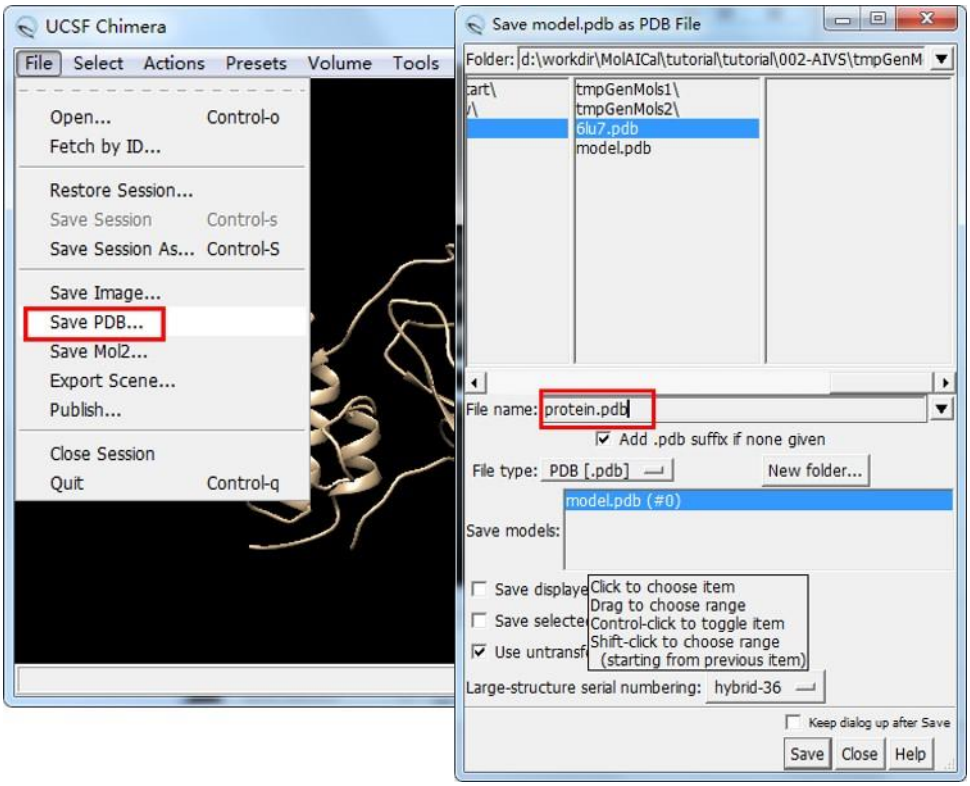

Figure 3. Save protein structure

4) Close Session, reload “model.pdb”, select ligand, invert (selected model), and delete protein (see orders in Figure 4).

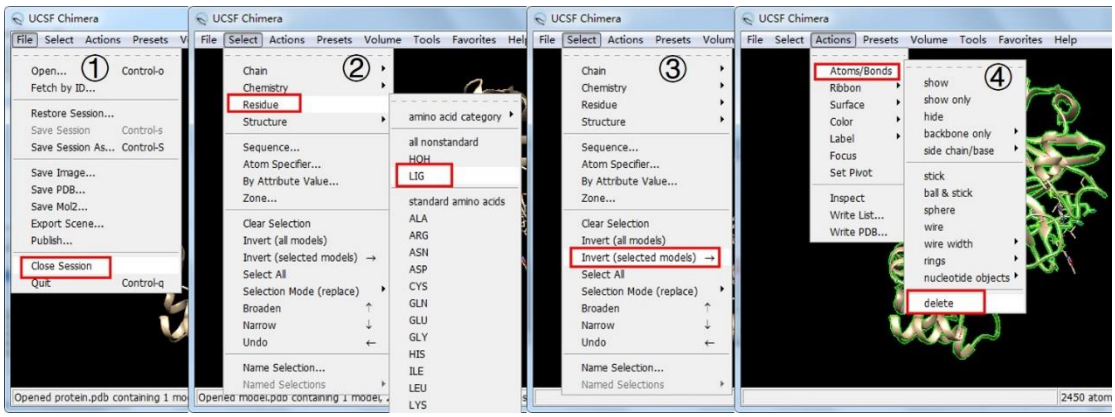

Figure 4. Save ligand without protein.

5) Save ligand file named “ligand.pdb” (see Figure 5).

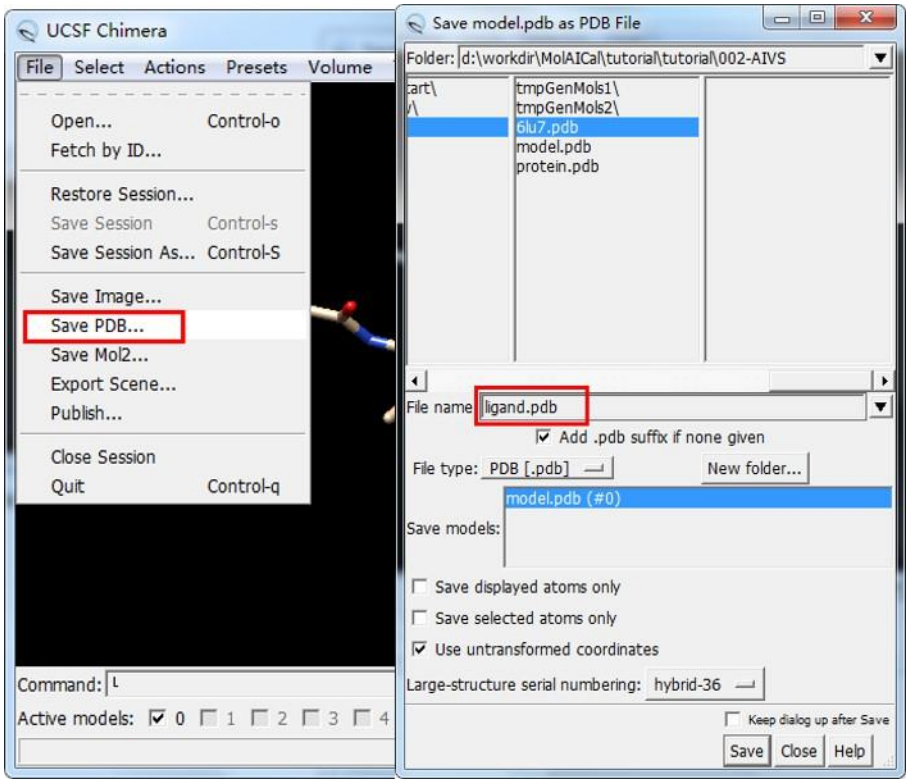

Figure 5. Save ligand file.

3.2. Calculating box center and length

1. Select ligand following the previous step or reload “ligand.pdb” and select ligand. Then, select distance tool: Tools→Structure Analysis→Distance (see Figure 6):

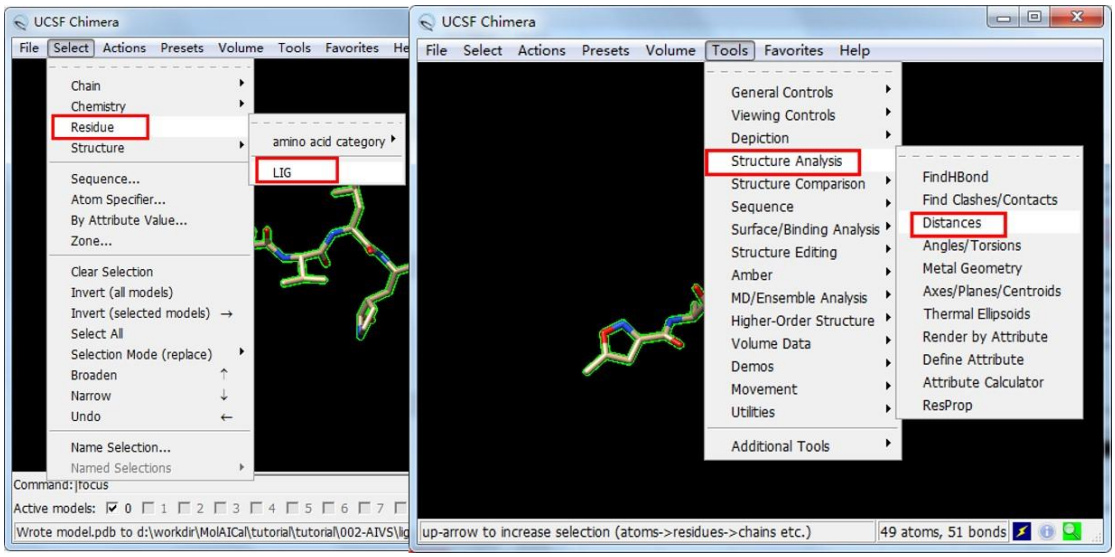

Figure 6. Select distance tool

2. Get centroid coordinate of protein pocket by ligand (see Figure 7)

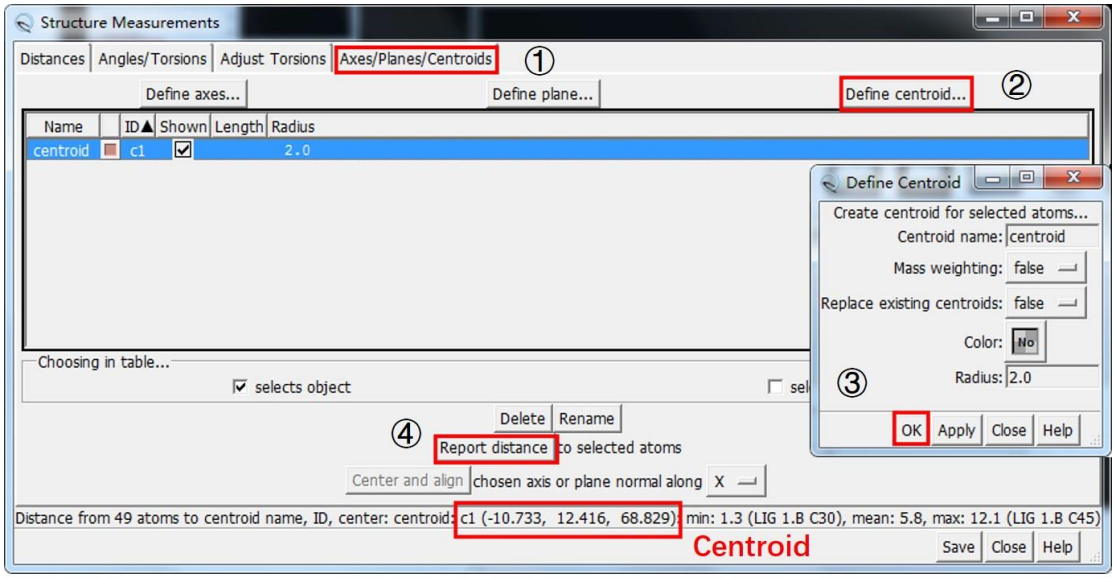

Figure 7. Get centroid coordinate

Create **"conf.txt"** and write the centroid coordinate to it as below:

---

center_x = -10.733

center_y = 12.416

center_z = 68.829

Notice: The name of **"conf.txt"** is default in MolAICal. If you create it with other words (for example, xxx.txt), you should add the parameter "-m xxx.txt" for command line of MolAICal. For more detailed commands, please check the manual of MolAICal. Besides, **you can directly modify the relative parameters of "conf.txt" for your work according to your real studies.**

3. Set the length of the dock box

Calculating the final box size. You can try X, Y, Z, lengths of 25, 30, 25. Generate the "box.bild" by using the command of MolAICal as below (note: **the double quotes are necessary for X, Y, Z coordinates. The interval distance between X, Y, Z coordinates should be one space**.):

1) To get "box.bild", input the command as below:
   #> molaical.exe -tool box -i "-10.733 12.416 68.829" -l "25.0 30.0 25.0" -o "box.bild"

2) File→open, then open "**box.bild**" (see Figure 8), and check whether the generated box is suitable (see Figure 8).

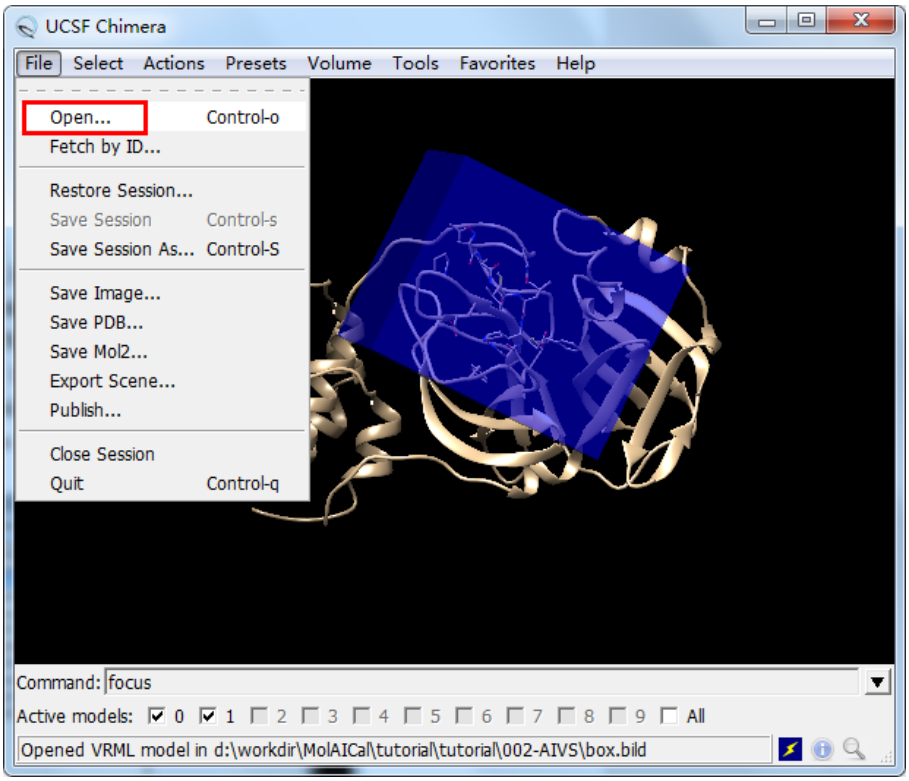

Figure 8. Open box.bild by UCSF Chimera.

The box size of 25, 30, 25 is suitable, so the final center parameter is -10.733, 12.416, 68.829 and the final box lengths of X, Y, Z are 25.0, 30.0, 25.0.

Notice: If you calculate the geometric center by VMD software, the final center parameter will be -10.86, 12.57, 68.82. They are all right.

3.3. Change protein to PDBQT format for virtual screening

Below command can prepare protein in PDBQT format:

#> MolAICal-xxx\molaical.exe -dock receptor -i protein.pdb

Note: MolAICal-xxx is your downloaded version of MolAICal.

Until now, all files are prepared.

3.4. Run virtual screening with deep learning model and molecular docking

#> cd 002-AIVS

Finally, run the following command in the background:
For Linux:
#> molaical.exe -dock AI -s ZINCMol -n 6 -nf 3 -nc 3 >& vs.log &
-n: represents the total generated molecules for docking.
-nf: number of molecule in one folder
-nc: number of CPU cores for running job

For windows (using PowerShell):
#> molaical.exe -dock AI -s ZINCMol -n 6 -nf 3 -nc 3

If you want to run it background, you can run below command:
#> powershell -windowstyle hidden -command "molaical.exe -dock AI -s ZINCMol -n 6 -nf 3 -nc 3"

Of course, if you want to perform the classical virtual screening based on the known drug database, you can refer to the third item in the section of drug design tutorials of MolAICal (https://molaical.github.io/tutorial.html).

4. Results

4.1 checking results

Users can use Pymol software (http://www.lfd.uci.edu/~gohlke/pythonlibs) to load ligands with PDBQT format directly. Here, UCSF Chimera is introduced to show docking results.

Open 002-AIVS\tmpGenMols1

1) Adding hydrogen (option)
#> molaical.exe -dock addh -i 1_out.pdbqt

2) Changing "pdbqt" to "pdb" format
#> molaical.exe -dock pdbqt2pdb -i 1_out.pdbqt

It will get a molecular file named "1_out.pdb". And then use UCSF Chimera to load "1_out.pdb" and "protein.pdb". The docking result is shown in Figure 9 that indicates MolAICal obtains the suitable ligand from the molecular database generated by deep learning model.

Notice: I omit the surface show process in Figure 9. If users want to obtain the result like Figure 9, "Actions→Surface→show" should be selected in UCSF Chimera.

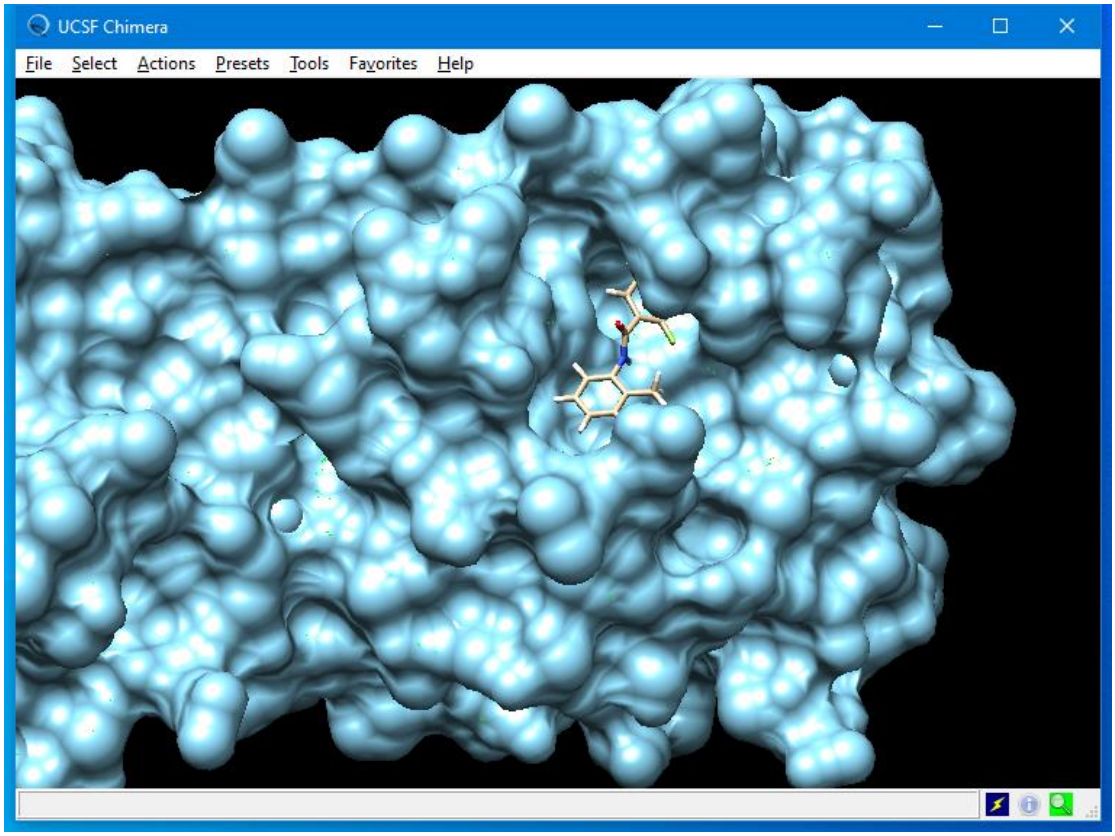

Figure 9. The screened ligand is in the pocket of SARS-CoV-2 Mpro.

4.2 Ranking virtual screening results

Goto folder 002-AIVS, and run below command:

#> python /home/bai/MolAICal-xxx/scripts/printScore.py 'tmpGenMols1/*out.pdbqt'

It will show the ligand name and corresponding binding score. If users want to save results into a file, they can use below command:

#> python /home/bai/MolAICal-xxx/scripts/printScore.py 'tmpGenMols1/*out.pdbqt' > results.log

Note: "/home/bai/MolAICal-xxx" is your installed directory path of MolAICal. All the useful scripts are supplied into the folder "scripts" of MolAICal.

If you work in Window:
Open "results.log" by using Excel. In "Separator", tick the space. Or you can copy all the content of "results.log" into Excel, directly. Select all data in the second column and click "Sort Largest to Smallest" or "Sort Smallest to Largest" according to your needs (see Figure 10).

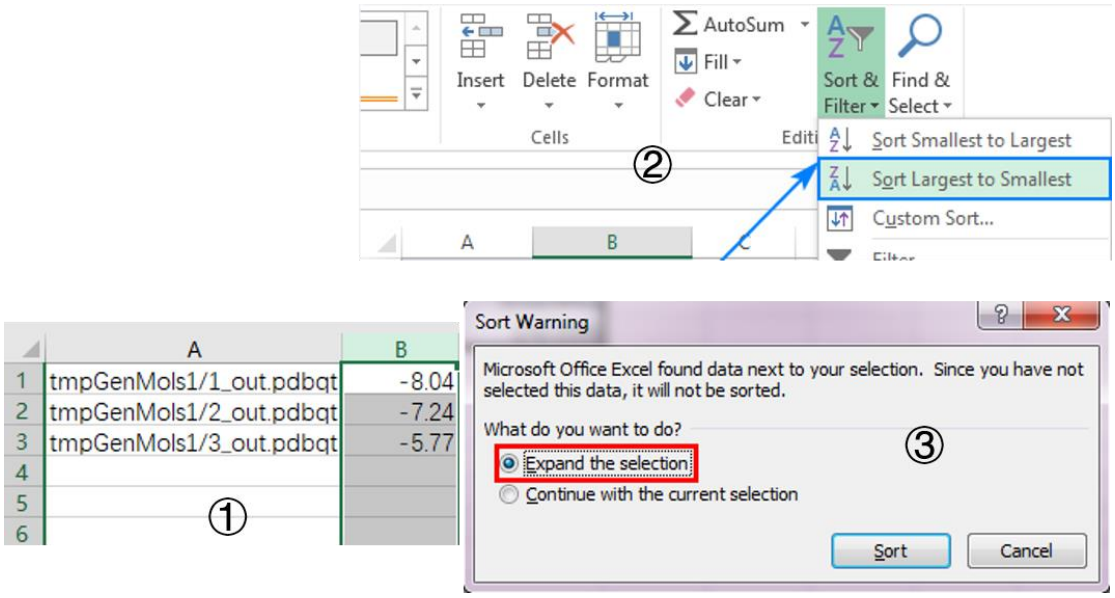

Figure 10. Ranking results

If you work in Linux, users can do it as below:

#> sort -n -t ' ' -k 2r results.log > rank.dat

Note: parameter "2r" is "Sort Smallest to Largest", while "1r" is "Sort Largest to Smallest".

4.3 Extract top-ranked molecule into the new folder

If users want to move top-ranked molecules into new folder, for example, in this tutorial, if users want to move 2 top-ranked molecules into new folder named "results", they can use the command as follows:

#> python /home/bai/MolAICal-xxx/scripts/molaicaldTopResults.py "tmpGenMols1/*out.pdbqt" 2 results

Two top-ranked molecules will be moved into folder "results".

References

1 Fleming, N. How artificial intelligence is changing drug discovery. *Nature* **557**, S55-S55 (2018).