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 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/
3) MGLTools: https://ccsb.scripps.edu/mgltools/downloads/
4) Python: https://www.python.org/
5) Pymol: http://www.lfd.uci.edu/~gohlke/pythonlibs

It is easily to install the first four software. They can be easily installed by following step tips. For pymol install, it needs modules **numpy, pmw, pymol_launcher and pymol**. The numpy, pmw, pymol_launcher and pymol should choose the same version and correspond to your installed version of Python in your operating system. They can be downloaded from the below website:
https://www.lfd.uci.edu/~gohlke/pythonlibs/#numpy
https://www.lfd.uci.edu/~gohlke/pythonlibs/#pymol-open-source
Then install Pymol by following command:
#> pip install --no-index --find-links="%CD%" pymol_launcher

The Pymol named “pymol.exe” will be installed in the directory “Scripts” in your installed directory of Python. You can make a shortcut on your desktop of operating system.
Make sure all software is installed rightly.

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. If you are familiar with Autodock vina, you can skip this step. You can refer to the video: https://youtu.be/-GVZP0X0Tg8 or download video tutorial from http://vina.scripps.edu/tutorial.html. 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)

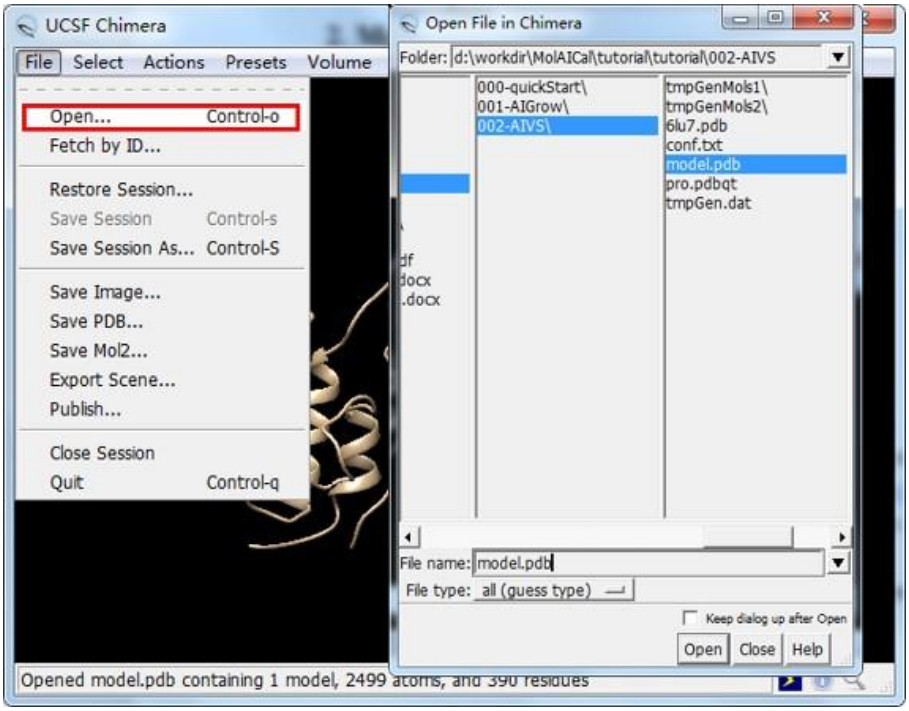

Figure 1. Load protein structure files.

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

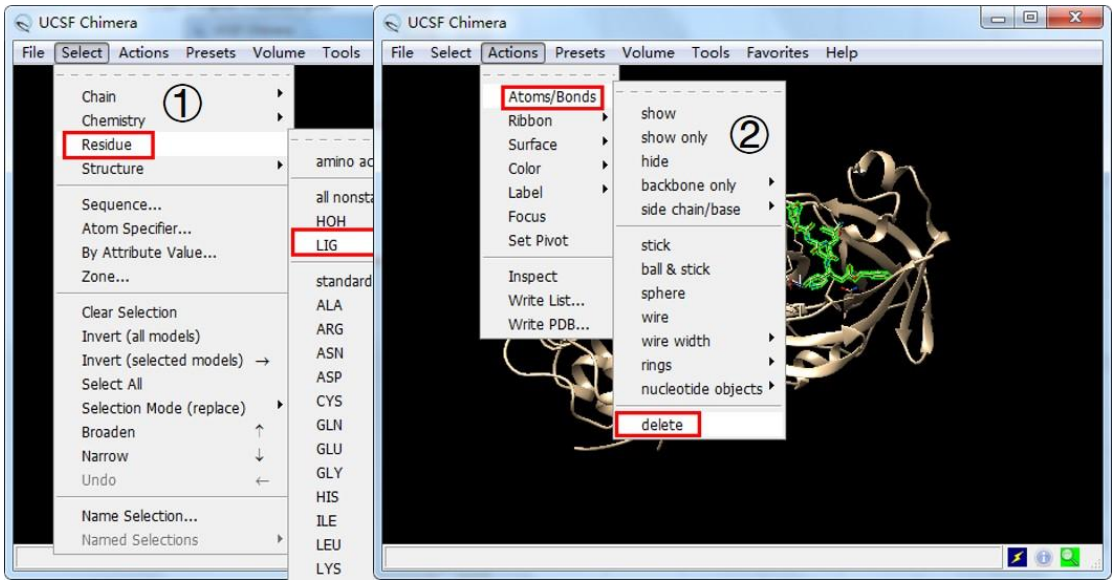

Figure 2. Select ligand and delete it

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

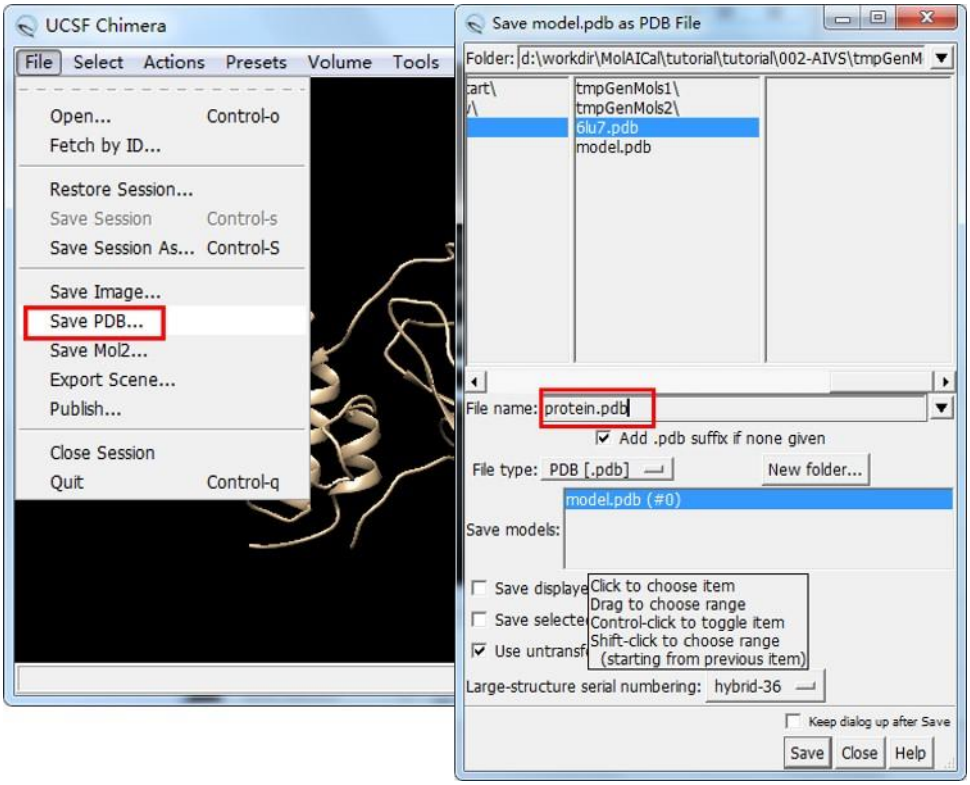

Figure 3. Save protein structure

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

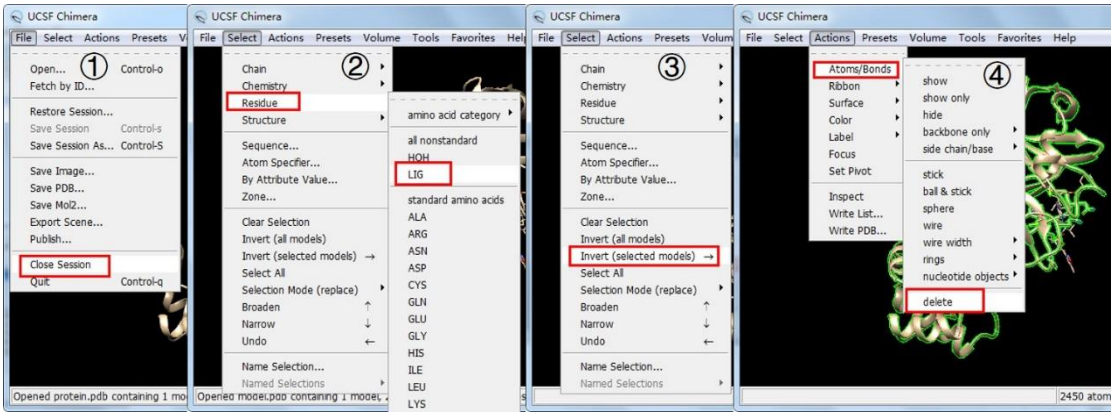

Figure 4. Save ligand without protein.

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

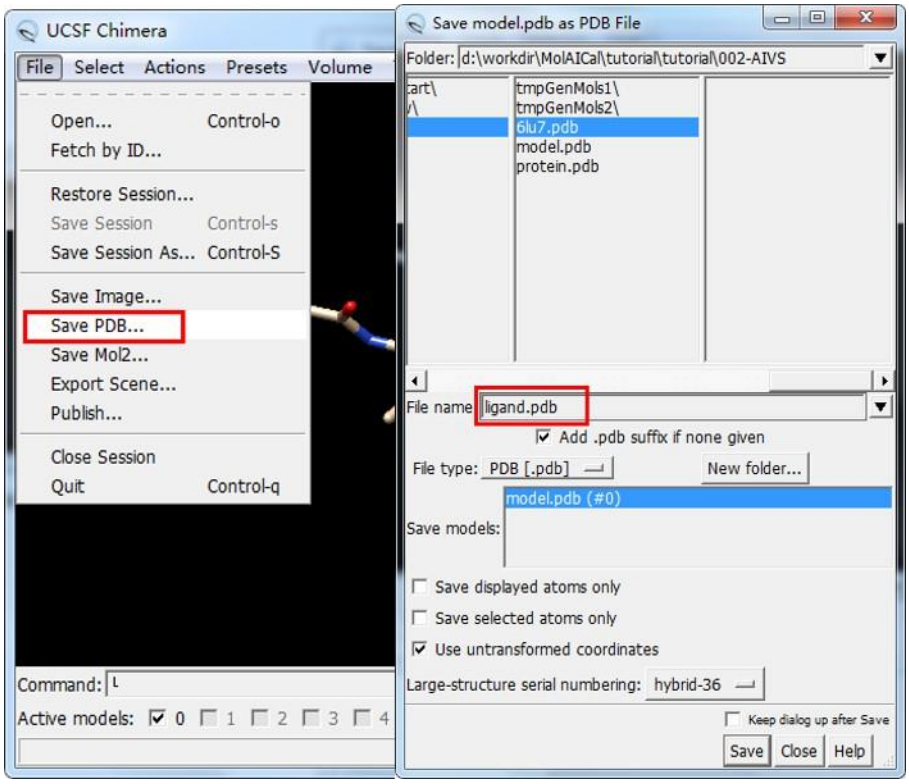

Figure 5. Save ligand file.

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):

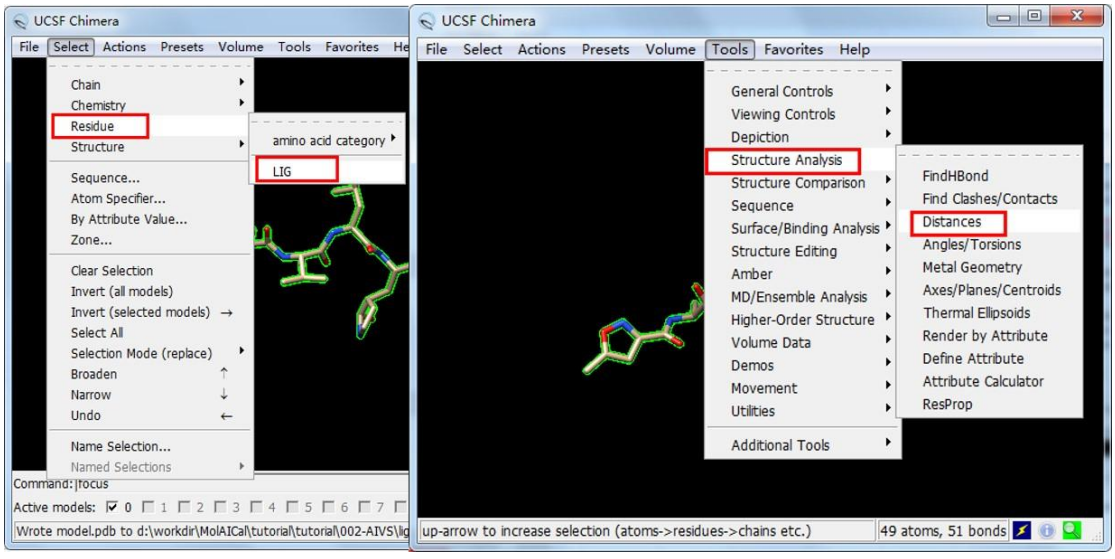

Figure 6. Select distance tool

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

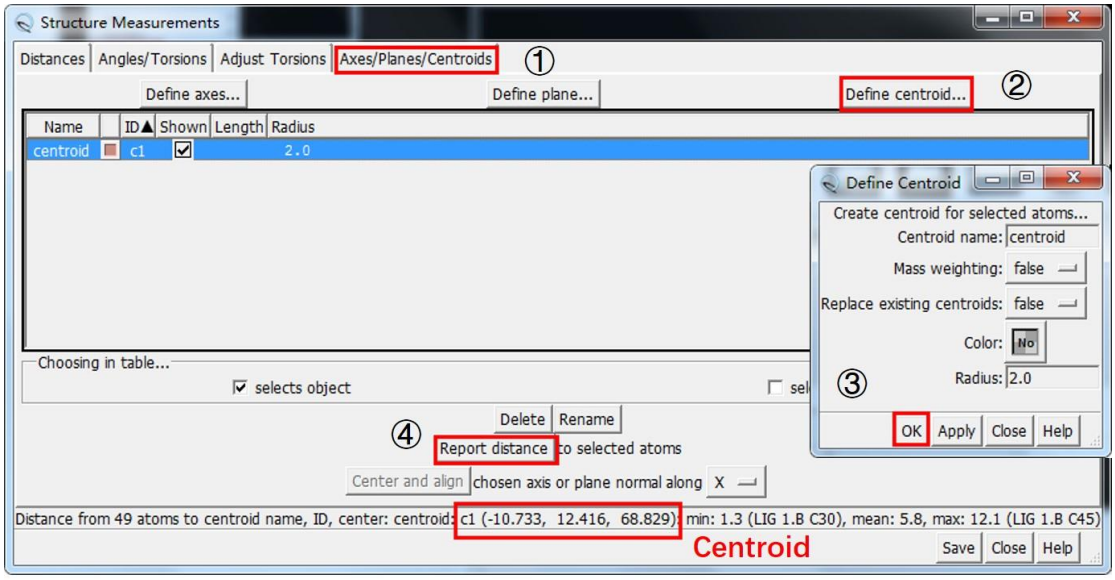

Figure 7. Get centroid coordinate

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 fixed in MolAICal. If you create it with other words, MolAICal cannot recognize it.

3. Set the length of the dock box

1. 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"

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

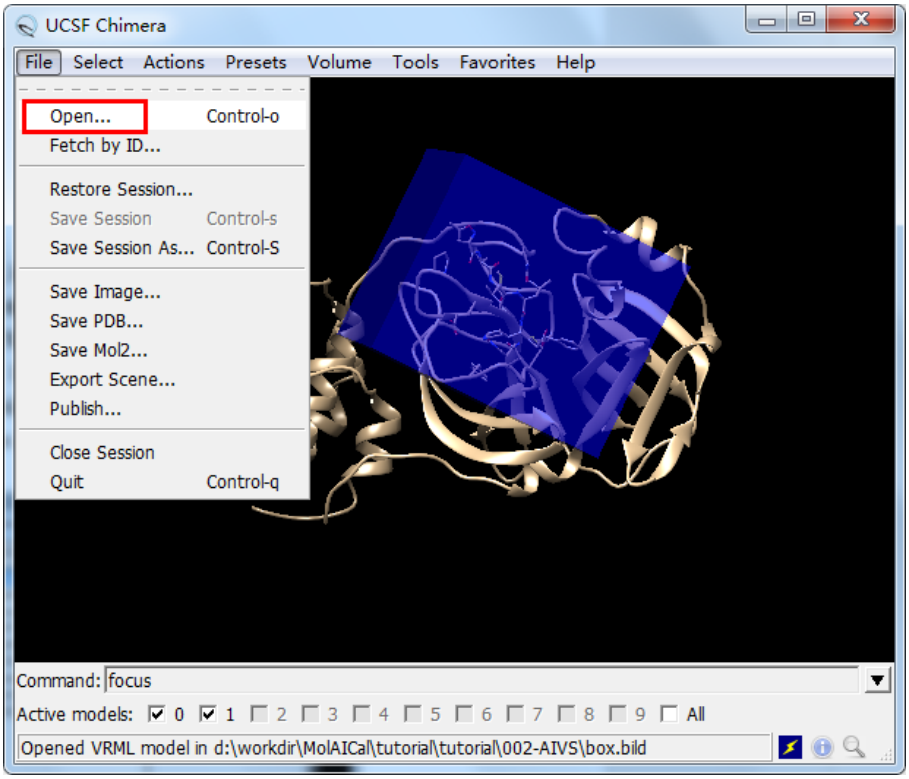

Figure 8. Open box.bild by UCSF Chimera.

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

1. Open "AutoDockTools", File→Read Molecule→protein.pdb, and add polar hydrogen (see Figure 9).

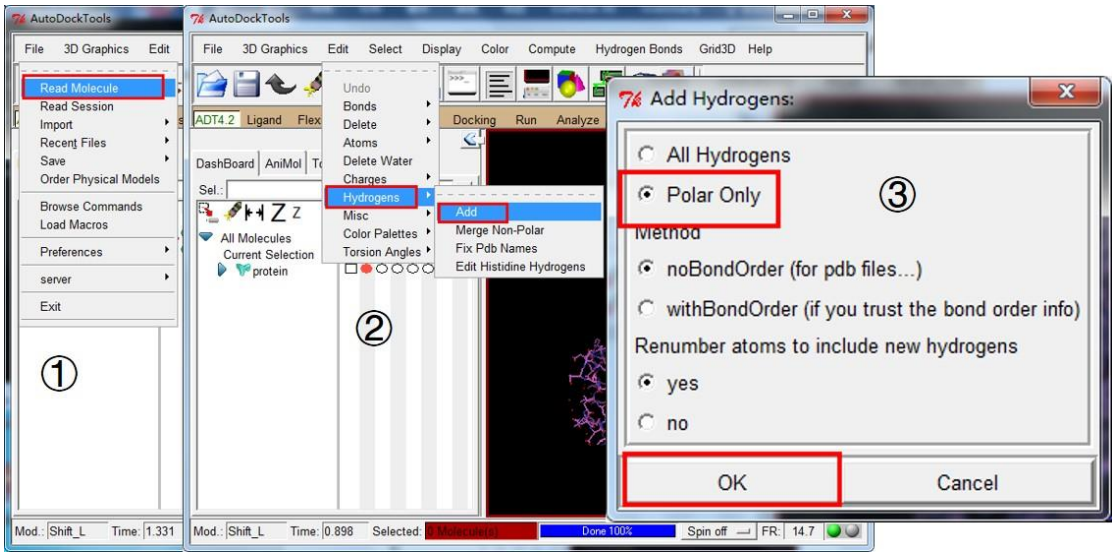

Figure 9. Add polar hydrogen.

2. Save protein with PDBQT format. Grid→Macromolecule→Choose…, then click the “protein” and press “Select Molecule” button, and save protein to “pro.pdbqt” (see Figure 10).

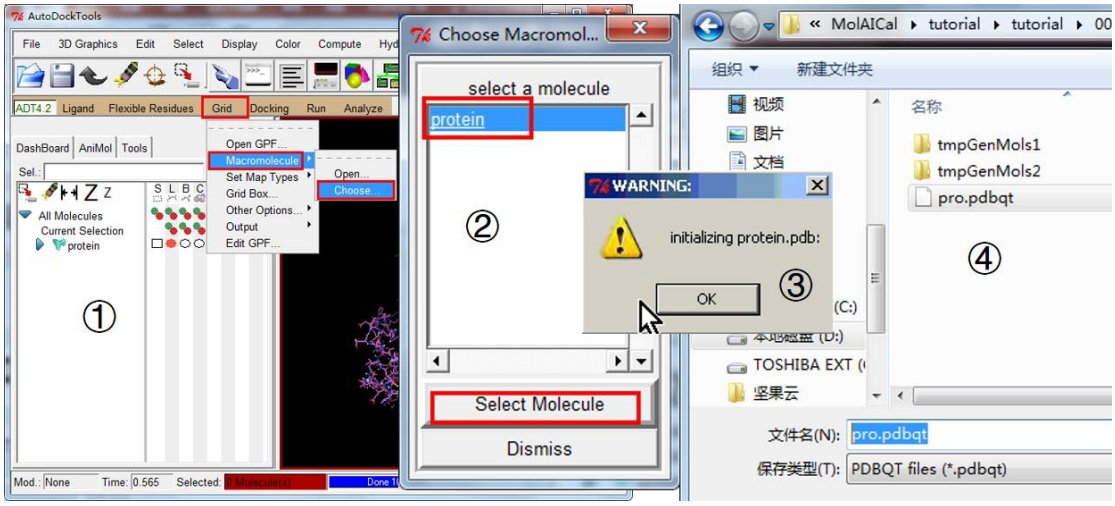

Figure 10. Save protein with PDBQT format.

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

You can convert PDBQT format of results to PDB format by Open Babel. Then loading it with UCSF Chimera. Here, the pymol software (http://www.lfd.uci.edu/~gohlke/pythonlibs) is used to

show results (see Figure 11). The results show MolAICal obtains the similar structural ligand from the ligands database generated by deep learning model.

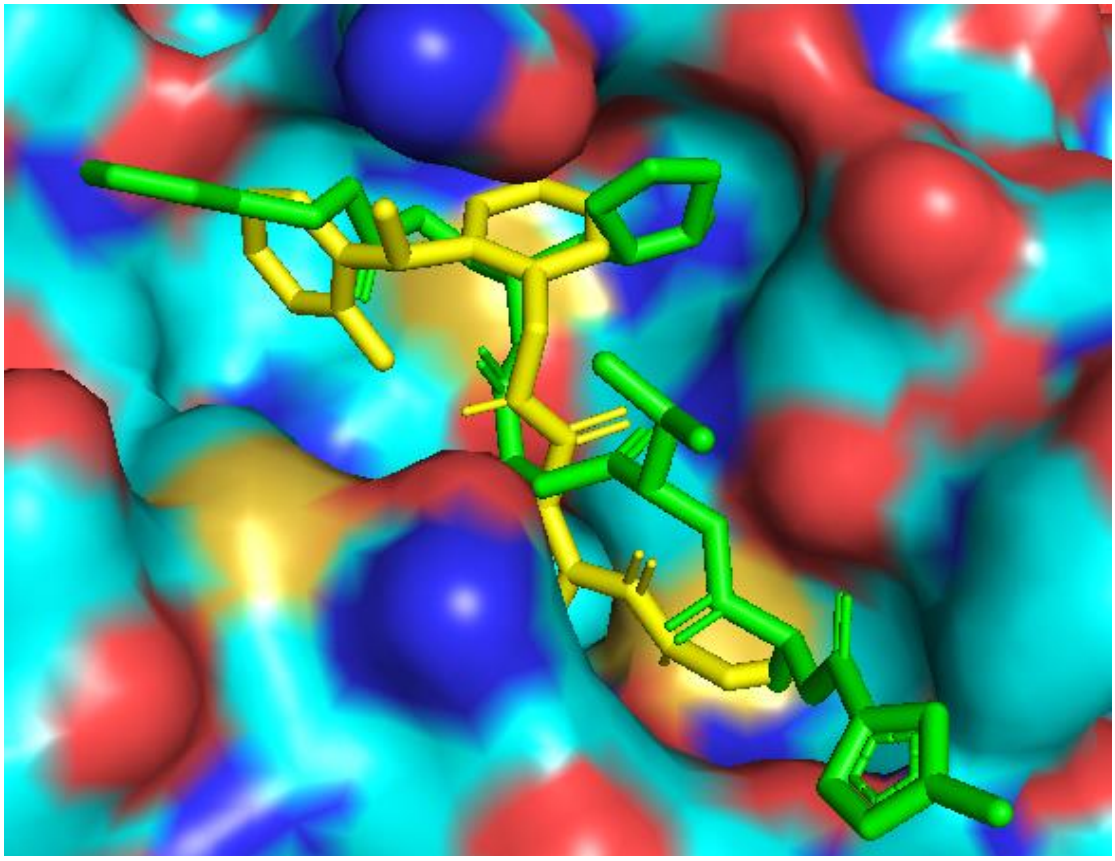

Figure 11. The ligand with green is the inhibitor N3 of SARS-CoV-2 Mpro. The ligand with yellow is obtained by AI and molecular docking.

References

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