

Two steps for virtual screening in the pocket of SARS-CoV-2 Mpro by MolAICal

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

In this tutorial, we introduce the fast way for drug virtual screening of SARS-CoV-2 Mpro based on the known database such as ZINC database. The premise for this tutorial is that you can deal with the receptor and ligand structures by MolAICal (see “Quick Start 1” in <https://molaical.github.io>). If users want to know more detailed molecular preparation process, you can learn from the tutorial: <https://github.com/MolAICal/documents/tree/master/tutorials/002-AIVS>. Here, MolAICal soft package is employed for this tutorial.

2. Materials

2.1. Software requirement

1) MolAICal: <https://molaical.github.io>

2.2. Example files

1) All the necessary tutorial files are downloaded from:

<https://github.com/MolAICal/tutorials/tree/master/003-VS>

2) The file named “ligandSet.mol2” which contains 16 ligands obtained from ZINC database is chosen for demo. You can select your ligand database.

3) The protein file named “pro.pdbqt” that is PDBQT format structure of SARS-CoV-2 Mpro is used for molecular docking.

3. Procedure

Go to the tutorial directory:

```
#> cd 003-VS
```

1. Split ligands set into single molecule. If your drug database is built by single ligand with mol2 format, you can omit this step.

```
#> molaical.exe -tool mol2 -w split -n 1 -v true -m number -i ligandSet.mol2 -o splitdir
```

Note: when the second line content is valid in mol2 file of each ligand, you can also use the second line string of mol2 file as the file name by using the below command:

```
#> molaical.exe -tool mol2 -w split -n 1 -v true -i ligandSet.mol2 -o splitdir
```

Notice: Each molecular file in docking set file such as “ligandSet.mol2 must contain full structure. If molecular file in docking set file such as “ligandSet.mol2 is **lack of hydrogen**, MolAICal may not generate molecular file in PDBQT format. Users should firstly employ UCSF Chimera to add hydrogen on the molecule which is lack of hydrogen, or use MolAICal to convert molecule in PDBQT format by the following command (users should replace 1.mol2 into their own filename):

```
#> molaical.exe -tool format -i E:/1.mol2 -o E:/1.pdbqt
```

2. Running virtual screening. Before running the below command, please make sure the previous docking results are removed.

```
#> molaical.exe -dock vs -i splitdir -nc 3
```

Notice: If users use molecules in PDBQT format, directly, users can close format conversion for virtual screening as below:

```
#> molaical.exe -dock vs -i splitdir -b off -k pdbqt -nc 3
```

-nc: represents the number of used CPU cores.

-b: switch molecular format conversion. “on” means conversion. “off” means no conversion. The default value is “on”.

-k: the files contain the keyword for virtual screening. The default value is “.mol2”.

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 003-VS/splitdir

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 1 that indicates MolAICal obtains the suitable ligand from the known molecular database.

Notice: I omit the surface show process in Figure 1. If users want to obtain the result like Figure 1, “Actions→Surface→show” should be selected in UCSF Chimera.

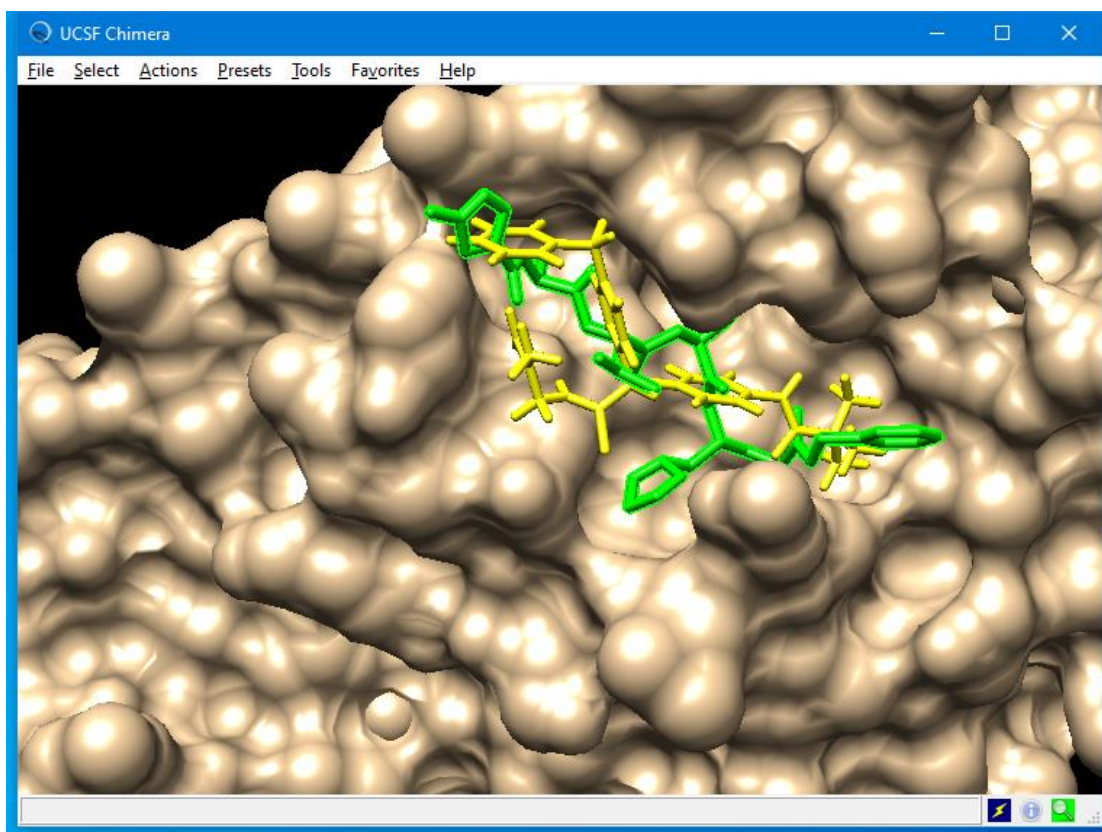


Figure 1. The ligand with green is the inhibitor N3 of SARS-CoV-2 Mpro. The ligand in yellow is obtained by virtual screening.

4.2 Ranking virtual screening results

Goto folder 003-VS, and run below command:

```
#> python /home/bai/MolAICal-xxx/scripts/printScore.py "splitdir/*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 "splitdir/*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 2).

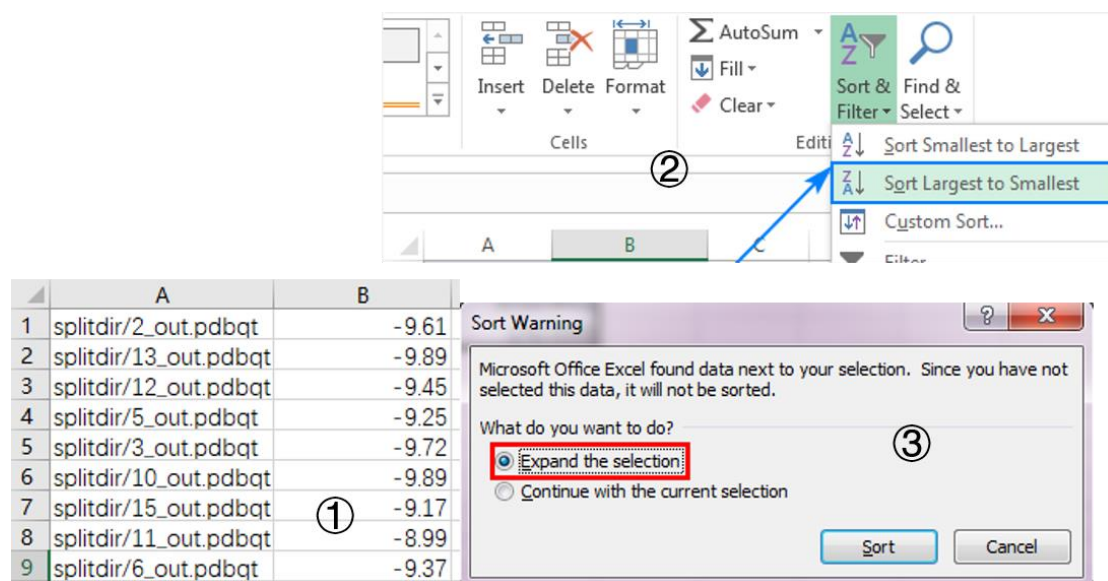


Figure 2. Ranking results

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 "splitdir/*out.pdbqt" 2 results
```

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