Tutorial of binding affinity prediction by MolAICal and Pafnucy model

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

Deep learning model can be used to evaluate binding affinity between ligands and receptors. In this tutorial, Pafnucy [1] is introduced to calculate binding affinity of ligands by MolAICal. This example can be for users to know quickly how deep learning predicts binding affinity between ligands and proteins. This tutorial can only run on Linux system. For more detailed MolAICal, please read this paper (https://doi.org/10.1093/bib/bbaa161).

2. Materials

2.1. Software requirement

1) MolAICal: https://molaical.github.io
2) UCSF Chimera: https://www.cgl.ucsf.edu/chimera/
3) VMD: https://www.ks.uiuc.edu/Research/vmd/

2.2. Example files

1) All the necessary tutorial files are downloaded from:
https://github.com/MolAICal/tutorials/tree/master/015-bindingaffinityPafnucy

3. Procedure

3.1 Install Pafnucy model

Currently, only Linux version of MolAICal is supported.

1) Click and open website: **DownloadModel**
Then go to the folder "AImodels--BindingAffinity--pafnucy--linux" in the open website, download a file named "pafnucy.tar.gz"

2) Move "pafnucy.tar.gz" to the folder "MolAICal-xxx/mtools". Where "MolAICal-xxx" is your decompressed root directory of MolAICal. The folder "mtools" is the targeted directory.

3) Decompress file
#> tar -xzvf pafnucy.tar.gz

And go to folder named Pafnucy:
#> cd pafnucy

4) install Pafnucy model
#> chmod +x install.sh
#> ./install.sh
Until now, Pafnucy model is installed completely.

3.2 Calculate binding affinity for one complex

3.2.1 Prepare protein file

1) "3ui7.pdb" is a protein file that is downloaded from PDB database (**Here, assuming hydrogens and charge have been added, where Amber ff14SB is for standard residues and AM1-BCC is for non-standard residues and ligands in UCSF Chimera**). First of all, open UCSF Chimera and save A chain of "3ui7.pdb" as "3ui7_A.pdb".

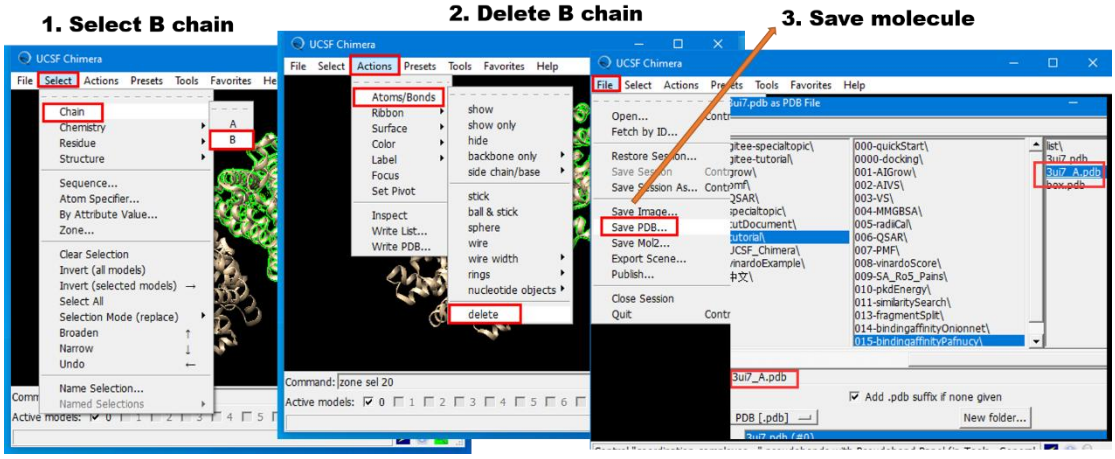

Figure 1. Deal with protein file

2) Get the geometric center of ligand (see Figure 2). Geometric center is 5.081, 12.632, 43.498

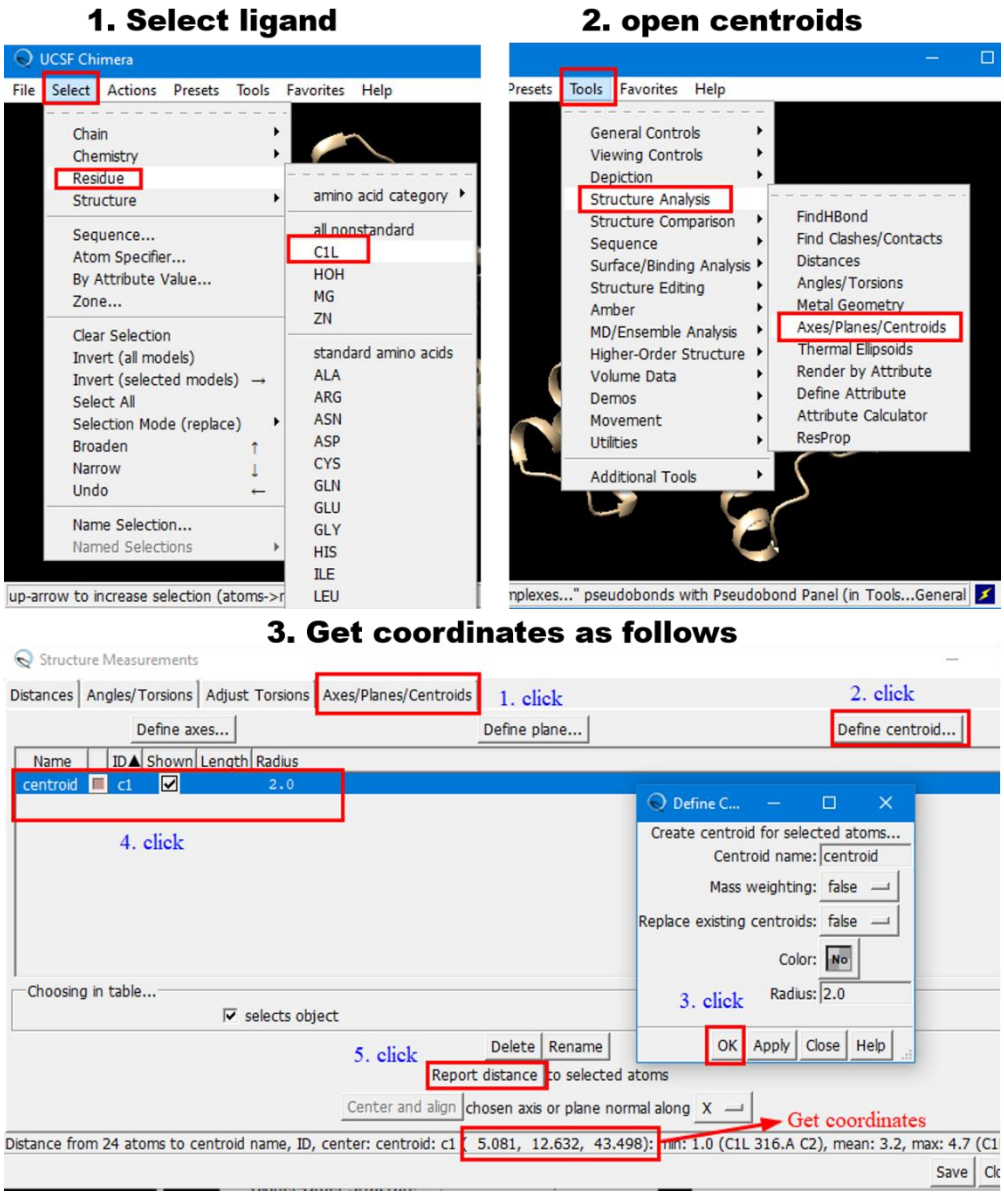

Figure 2

3) The original paper of Pafnucy [1] cropped the complex to a defined size of 20 Å cubic box around the geometric center of a ligand. Get maximum and minimum coordinates of 20 Å cubic box via MolAICal (**Notice: the double quotes are necessary for X, Y, Z coordinates. The interval distance among X, Y, Z coordinates and box lengths should be one space**):
#> molaical.exe -tool box -i "5.081 12.632 43.498" -l "20.0 20.0 20.0" -o box.bild

It will generate a file named "box.bild" which contains minimum and maximum coordinates in ".box" part. Open "box.bild", it can be find like this in ".box" part: -4.919 2.6319999999999997 33.498 15.081 22.631999999999998 53.498

4) load file "3ui7_A.pdb" and open Tk Console of VMD (see Figure 3).

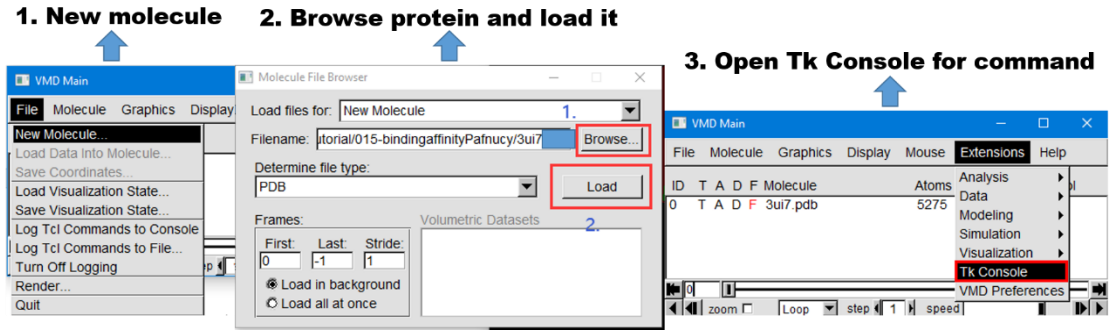

Figure 3

5) Using VMD to save the complex that is within 20 Å cubic box arround the geometric center of a ligand.

My workspace is in E: disk, so I use below commands to change my workspace in VMD TkConsole:
#> cd E:
#> cd workdir/MolAICal/tutorial/tutorial/015-bindingaffinityPafnucy/

Input command below in VMD TkConsole (**Note:** minimum and maximum x, y, z of box have been calculated above, users can replace their own real data in the below command):
#> set box [atomselect top "(not ((x < -4.919 or x > 15.081) or (y < 2.6319999999999997 or y > 22.631999999999998) or (z < 33.498 or z > 53.498)))"]

#> $box writepdb box.pdb

Figure 4 shows records in VMD Tk Console for this tutorial. Please check it. Users should use their own minimum and maximum data.

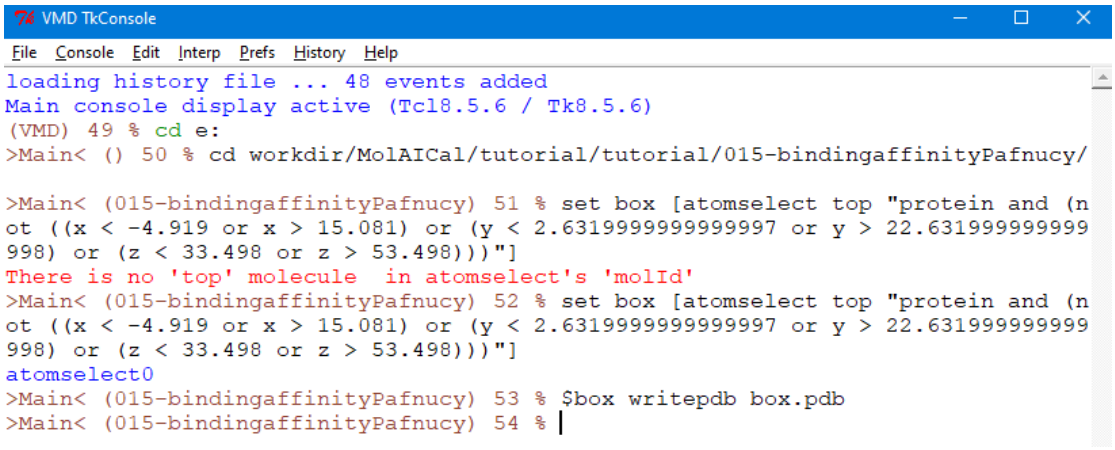

Figure 4

6) Save "box.pdb" to "3ui7_pocket.mol2" using UCSF Chimera (see Figure 5)
(**Discussion: Sometimes, I see somebodies employ full protein in mol2 for binding affinity prediction using Pafnucy model. If yes, users can omit above six steps.**)

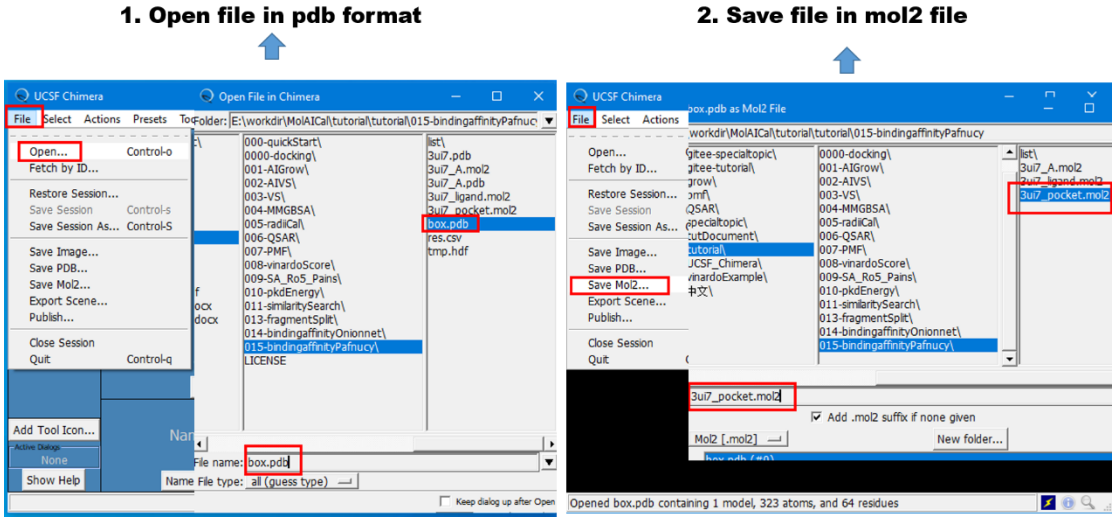

Figure 5

3.2.2 Calculate binding affinity of one protein and one ligand

Go to 015-bindingaffinityPafnucy, and use below command:

#> molaical.exe -model pafnucy -l 3ui7_ligand.mol2 -p 3ui7_pocket.mol2 -o results.csv

It will generate a file named results.csv that contains pKx (pKd or pKi) value of one complex. If users want to convert pKx to binding free energy, they can refer to "https://molaical.github.io/tutorial.html" and MolAICal manual which part is involved in "Binding free energy from pKd or pKi".

Notice: if users do not have mol2 format molecular file for ligands, they use the command of MolAICal to convert molecules in mol2 format. For instance, users have pdb format file, they can

use the below command for format conversion (**Note:** molecule should have correct suffix so that MolAICal can recognize them automatically):
#> molaical.exe -tool format -i ligand.pdb -o ligand.mol2

3.2.3 Calculate binding affinity of many ligands and one protein

Sometimes, users want to calculate binding affinities of many ligands, under these circumstances, it can employ Linux shell to realize it easily.

Go to 015-bindingaffinityPafnucy/list, and input below command:
#> ls lig*.mol2 > list.txt

It will generate a file named "list.txt" which contains ligand names. Then, open the file named "run.sh", and modify right path of molaical.exe (see Figure 6)

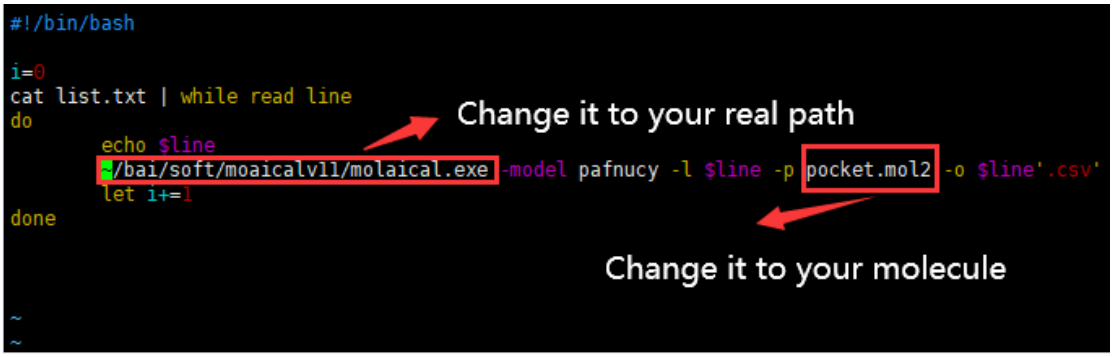

Figure 6

Finally run command as below:
#> bash run.sh

Merge all results:
#> cat *.csv > results.csv

It will generate a file named "results.csv" that contains the predicted pKx values. If users want to convert pKx to binding free energy, they can refer to "https://molaical.github.io/tutorial.html" and MolAICal manual which part is involved in "Binding free energy from pKd or pKi".

3.2.4 Calculate binding affinity of few ligands and one protein in command line

If users do not have huge of ligands, they can use one command to calculate binding affinity between few ligands and one protein in the command line:
Go to 015-bindingaffinityPafnucy/list, and input the below command:
#> molaical.exe -model pafnucy -l "lig_1.mol2 lig_2.mol2" -p receptor.mol2 -o twoTest.csv

It will generate "twoTest.csv" which contains two binding affinity values of ligand files called lig_1.mol2 and lig_2.mol2. Please note that "lig_1.mol2 lig_2.mol2" should have double quotation marks in the command line. For more detailed commands, please check MolAICal manual.

Appendix

The linux shell content of "run.sh":

```
#!/bin/bash

i=0
cat list.txt | while read line
do
        echo $line
        ~/bai/soft/moaicalv11/molaical.exe -model pafnucy -l $line -p pocket.mol2 -o $line'.csv'
        let i+=1
done
```

Reference

1. Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P. Development and evaluation of a deep learning model for protein-ligand binding affinity prediction. Bioinformatics. 2018;34(21):3666-74.