Tutorial of cluster results of virtual screening 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

Sometimes, the results of virtual screening have similar binding scores. The top-ranked molecules have very similar structures. Selecting the representative molecules for the further experimental assay can save money and time. Here, MolAICal is introduced to cluster virtual screening results based on binding scores and structural similarities. For more detailed MolAICal, please visit: https://molaical.github.io.

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/017-clusterVSResults

3. Procedure

3.1 calculate structural similarity

1) MolAICal supplies two ways for structural similarity calculation that are fingerprint similarity and 3D structural similarity (For more detail, please check MolAICal manual). Here, 3D structural similarity is selected for this tutorial.

Go to 017-clusterVSResults, and input command as below:

```
#> molaical.exe -tool 3Dcompare -i name.dat -s similarity.dat -f mol2list
```

It will generate a file named similarity.dat which consists of similarity data of ligands. Where "name.dat" is name list of ligands.

2) Now, merging "similarity.dat", "bindingScore.dat" (containing binding scores of ligands), and "name.dat". Input command as follows:

```
#> molaical.exe -tool col -f similarity.dat -l bindingScore.dat -s " " -o tmp.dat
```

Then, merge tmp.dat and name.dat into a file named "pc.dat".

```
#> molaical.exe -tool col -f tmp.dat -l name.dat -s " " -o pc.dat
```

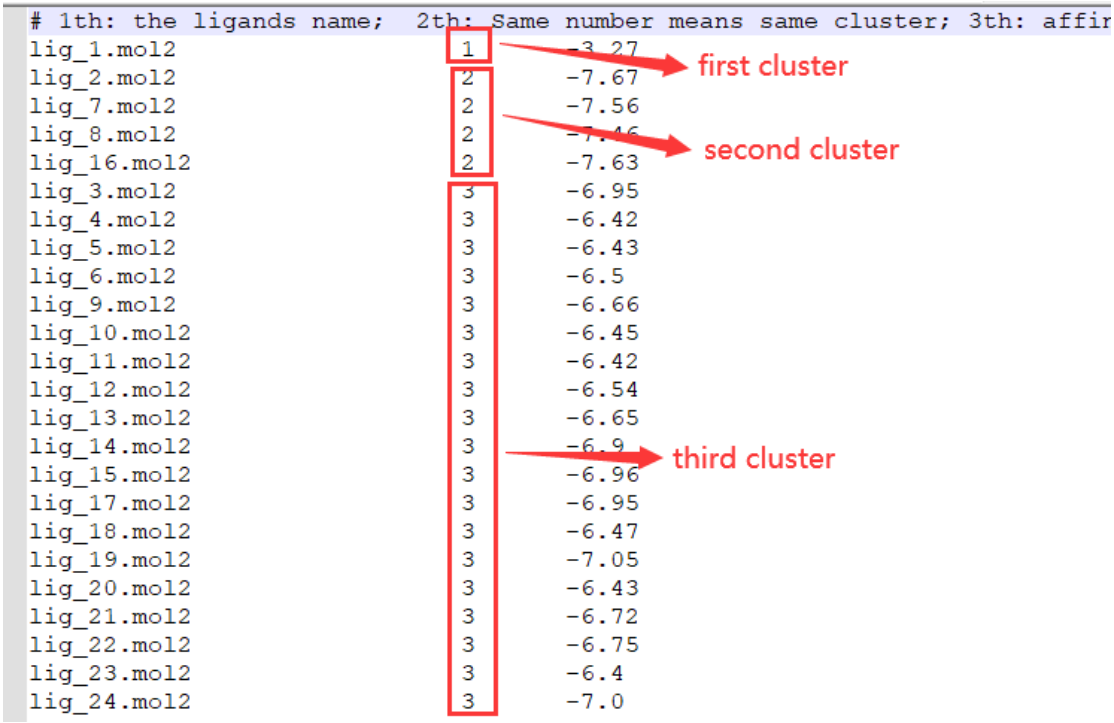
Notice: Please merge files in the above order. The first column is "similarity.dat", the second column is "bindingScore.dat", and the last column is "name.dat".

3.2 Cluster result

MolAICal employs k-means for clustering, input below command:

```
#> molaical.exe -tool kmeans -n 3 -i pc.dat -o results.dat
```

It will divide 3 clusters in the file “results.dat”. Open “results.dat”, it will be shown as below figure:




```
# 1th: the ligands name;  2th: Same number means same cluster; 3th: affir
lig_1.mol2                1          -3.27
lig_2.mol2                2          -7.67
lig_7.mol2                2          -7.56
lig_8.mol2                2          -7.46
lig_16.mol2               2          -7.63
lig_3.mol2                3          -6.95
lig_4.mol2                3          -6.42
lig_5.mol2                3          -6.43
lig_6.mol2                3          -6.5
lig_9.mol2                3          -6.66
lig_10.mol2               3          -6.45
lig_11.mol2               3          -6.42
lig_12.mol2               3          -6.54
lig_13.mol2               3          -6.65
lig_14.mol2               3          -6.9
lig_15.mol2               3          -6.96
lig_17.mol2               3          -6.95
lig_18.mol2               3          -6.47
lig_19.mol2               3          -7.05
lig_20.mol2               3          -6.43
lig_21.mol2               3          -6.72
lig_22.mol2               3          -6.75
lig_23.mol2               3          -6.4
lig_24.mol2               3          -7.0
```