Tutorial of ligand similarity search or comparsion 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 similarity search of ligands can help scientists to find potential compounds fastly. In this tutorial, the fingerprint similarity and 3D structural similarity searches are introduced for molecular comparison or simple virtual screening based on appointed ligand structure. Here, MolAICal (https://doi.org/10.1093/bib/bbaa161) 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/011-similaritySearch

3. Procedure

3.1. Fingerprint similarity search

Go to folder "011-similaritySearch/fingerprint"

```
#> cd 011-similaritySearch/fingerprint
```

1. Using the command "ls > listfiles.dat" in Linux console, or "dir /b > listfiles.dat" in DOS console of Window. Open generated file "listfiles.dat" and delete no useful characters. Make sure the file "listfiles.dat" only contains the ligand names.

2. Run command as below:

```
#> molaical.exe -tool finger -i listfiles.dat -o result.dat
```

It will generate the file "result.dat" which contains the values of structural similarities that correspond to names in "listfiles.dat". **Note:** the first line is the compared ligand in file "result.dat" except for the comment line which starts with "#".

3.2. 3D structural similarity search or comparsion

Go to folder "011-similaritySearch/3Dsimilarity"

```
#> cd 011-similaritySearch/3Dsimilarity
```

1. Using the command "ls > listfiles.dat" in Linux console, or "dir /b > listfiles.dat" in DOS console of Window. Open generated file "listfiles.dat" and delete no useful characters. Make sure the file "listfiles.dat" only contains the ligand names.

2. Run command as below:

```
#> molaical.exe -tool 3Dcompare -i listfiles.dat -s result.dat -f mol2list
```

It will generate the file “result.dat” which contains the values of structural similarities that correspond to names in “listfiles.dat” except the first line. **Note:** the first line is not the compared ligand in file “result.dat”. The first line of “result.dat” corresponds to the second line of “listfiles.dat”, the second line of “result.dat” corresponds to the third line of “listfiles.dat”, and so on.

In addition, MolAICal supports the 3D structural similarity comparsion between two ligands:
#> molaical.exe -tool 3Dcompare -i ligand.mol2 -s 1_10023_out.mol2

3.3. Rank results

The results of 3D structural similarity search are selected as an example. You can refer to this example for fingerprint similarity search.

If you work in Linux, first of all, the file “listfiles.dat” should correspond to the file “result.dat”. Here, the first line of “listfiles.dat” is deleted. And do it as below:
#> molaical.exe -tool col -f listfiles.dat -l result.dat -s " " -o merge.dat
#> sort -n -t ' ' -k 2r merge.dat > rank.dat

The generated file “rank.dat” contains the results ranks results.

If you work in Window, first of all, the file “listfiles.dat” should correspond to the file “result.dat”. Here, the first line of “listfiles.dat” is deleted. And do it as below:
Open Excel, and copy the content of “result.dat” into the 1st column, then copy the content of “listfiles.dat” into the 2nd column. Select all data in the first column and click “Sort Largest to Smallest” (see Figure 1).

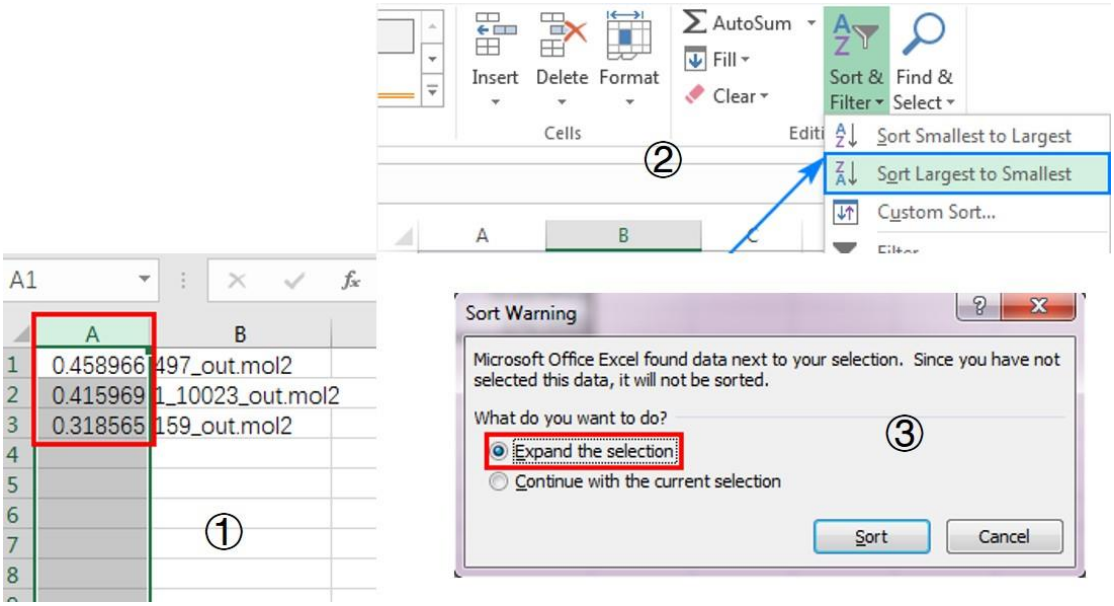

Figure 1. Rank steps of Excel