Tutorials of the calculations of Synthetic Accessibility, Lipinski's rule of five and PAINS 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, the calculations of synthetic accessibility, Lipinski's rule of five and PAINS are introduced. The synthetic accessibility (SA) can be used to evaluate the synthesis difficulty of the compound. The Lipinski's rule of five also known as the rule of five (RO5) is a rule to estimate drug-like or determine if a chemical compound has pharmacological or biological activity that would be likely orally active drug[1]. Pan-assay interference compounds (PAINS)[2] are the compounds which often show the false positive results in the biological assay. Here, the MolAICal software (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/009-SA_Ro5_Pains

3. Procedure

3.1. SA calculation
You can calculate the SA by SMILES sequence directly.
#> molaical.exe -tool sa -i "FC(F)(F)c1cc(ccc1)N5CCN(CCc2nnc3[C@H]4CCC[C@H]4Cn23)CC5"

Or you can calculate SA of ligands in one file that contains many SMILES sequences.
#> cd "009-SA_Ro5_Pains/SA"
#> molaical.exe -tool sa -i SmilTest.smi

Note: the higher value of SA indicates the compound is easier to be synthesized.

3.2. RO5 calculation
The RO5 of single ligand can be calculated as below:
#> cd "009-SA_Ro5_Pains/ro5"
#> molaical.exe -tool ruleoffive -f mol2 -n zinc_1879871.mol2

If you want to calculate Lipinski's rule of five for many ligands in bulk, you can do it below steps:
1. Using command "ls > mol2List.dat" in Linux console, or "dir /b > mol2List.dat" in DOS console of Window. Open generated file "mol2List.dat" and delete no useful characters. Make sure the file "mol2List.dat" only contains the ligand names.

2. Run command as below:
#> molaical.exe -tool ruleoffive -f mol2list -i mol2List.dat -o result.dat
It will generate the file named "result.dat" which contains RO5 values in bulk. More detail about

RO5, please see the manual of MolAICal.

3.3. PAINS calculation
The single ligand with SMILES or mol2 format can be calculated as below:
```
#> cd "009-SA_Ro5_Pains/pains"
#> molaical.exe -tool pains -f smi -n "c1ccccc1N=Nc1ccccc1"
#> molaical.exe -tool pains -f mol2 -n ZINC00154323.mol2
```

If you want to calculate PAINS for many ligands in bulk, you can do it below steps:
For SMILES format:
```
#> molaical.exe -tool pains -f smilist -i painsTest.txt -o smiResult.txt
```

For Mol2 format:
```
#> cd "009-SA_Ro5_Pains/pains/mol"
```

1. Using the command "ls > mol2List.dat" in Linux console, or "dir /b > mol2List.dat" in DOS console of Window. Open generated file "mol2List.dat" and delete no useful characters. Make sure the file "mol2List.dat" only contains the ligand names.

2. run command as below:
```
#> molaical.exe -tool pains -f mol2list -i mol2List.dat -o mol2Result.txt
```

It will generate the file named "mol2Result.txt" which contains PAINS information in bulk.

Note: recommend the SMILES format for PAINS calculation.

References

1   Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **46**, 3-26 (2001).

2   Dahlin, J. L. *et al.* PAINS in the assay: chemical mechanisms of assay interference and promiscuous enzymatic inhibition observed during a sulfhydryl-scavenging HTS. *J Med Chem* **58**, 2091-2113 (2015).