Tutorial of Simple QSAR calculation by MolAICal

Qifeng Bai, Xiaojun Yao
Email: molaical@yeah.net
Homepage: https://molaical.github.io
Lanzhou University
Lanzhou, Gansu 730000, P. R. China

1. Introduction

The quantitative structure-activity relationship (QSAR) models are regression or classification models used in drug design. In this tutorial, the simple regression model of QSAR is introduced based on the ligands of signal transducer and activator of transcription 3 (STAT3) protein which is considered as a potential drug target of cancer.

2. Materials

2.1. Software requirement

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

2) DRAGON: http://www.talete.mi.it/index.htm

Note: You can use any molecular descriptor calculator besides DRAGON software for this tutorial.

2.2. Example files

1) All the necessary tutorial files are downloaded from:

https://github.com/MolAICal/tutorials/tree/master/006-QSAR

3. Procedure

3.1. Calculate molecular descriptor

1) Open DRAGON and load ligands in folder "006-QSAR/ligands" (see Figure 1). The .bin format of ligands is optimized and saved by HyperChem software. You can optimize the ligand structures and save them with sybyl mol2, etc for calculation.
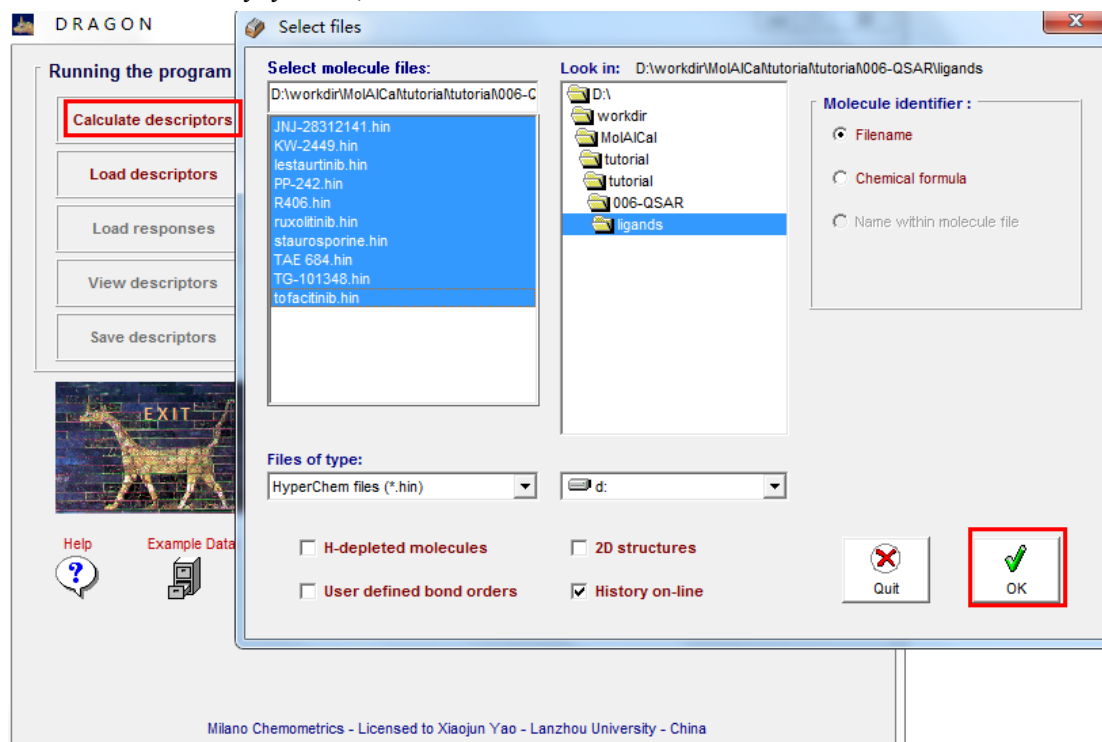

Figure 1. Molecular descriptor calculation by DRAGON

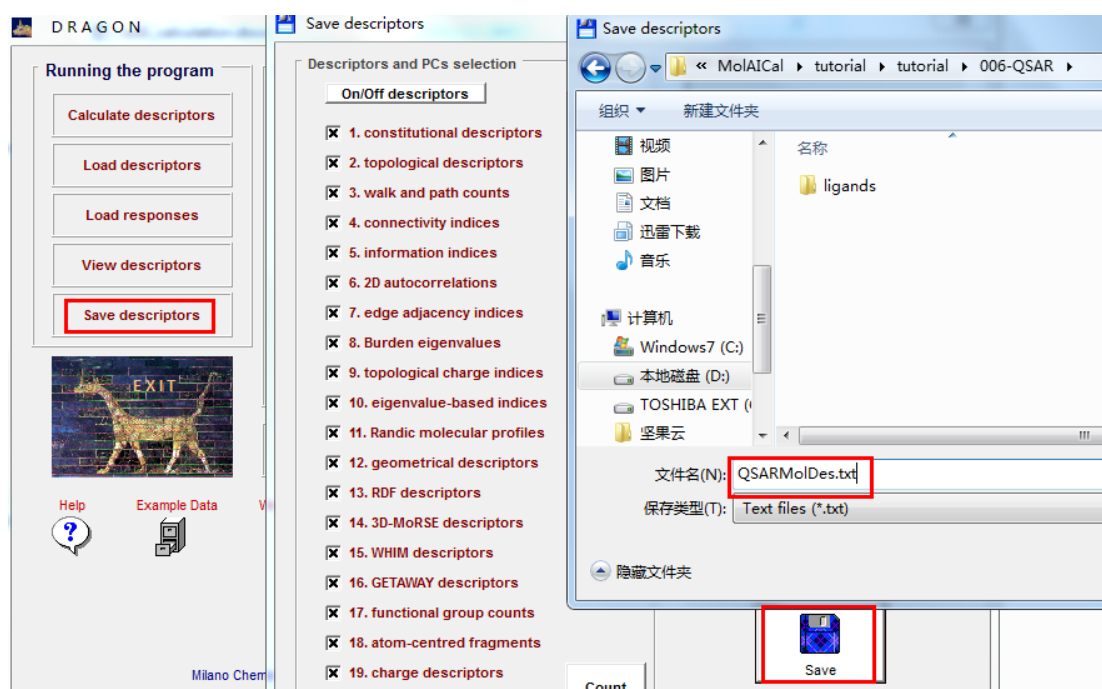2) Save the molecular descriptor named "QSARMolDes.txt" in TXT format file (see Figure 2).

Figure 2. Storing file named "QSARMolDes.txt"

3) Open "QSARMolDes.txt" with Excel and set parameters (see Figure 3).

Figure 3. Set parameter for QSAR

You must set the parameters in the "QSARMolDes.txt" strictly. The first number in the second line

must be the number of ligands for QSAR. The characters "on" means train and test sets are appointed. If "off" is chosen, it means LOO is used for QSAR calculation. The third and fourth lines for train and test sets can be omitted (See file "QSARMolDes_LOO.txt"). The numbers in the fourth line are train set. The numbers in the fifth line are the test set. The experimental values such as pKd should add in the third column.

3.2 QSAR calculations

Running command as below:

#> molaical.exe -qsar GA -i QSARMolDes.txt

Or

#> molaical.exe -qsar GA -i QSARMolDes_LOO.txt

This tutorial just contains 10 ligands. When Q2 is enough for your research, you can stop the QSAR running by "Ctrl + C". The results are stored in file "QSAROutFile.dat". Open "QSAROutFile.dat" and the information is as below

```
****** The 1th model ******
The Q^2-LOO is: 0.8542
R^2 fitting is: 0.9473
R^2 adjusted is: 0.9210
RSS is: 0.4042
The formula is: y = 0.68376 + (1.12498) * H0p + (2.45137) * Mor26e + (0.79399) * ESpm06d
The standard errors of b0 to b3 corresponding to formula is: 1.83351, 2.17332, 0.25011, 0.23398
The standard error of the regression (sigma) is: 0.2595
The experiment values, predicted values, calculated values by LOO validation and residuals:
8.0       8.1138      8.1743      -0.1138
7.12      7.2904      7.4440      -0.1704
8.43      8.5246      8.5705      -0.0946
7.96      7.7950      7.6441       0.1650
8.46      8.7477      8.8000      -0.2877
10.44    10.5084     10.7288      -0.0684
8.01      7.8877      7.8584       0.1223
7.8       7.9168      8.3305      -0.1168
8.96      8.5185      8.3828       0.4415
9.24      9.1171      9.0764       0.1229
```