

# 使用 MolAICal 进行药物的 QSAR 计算

作者：MolAICal （update 2020-08-02）

更多教程（含英文教程）请见如下：

MolAICal 官方主页：<https://molaical.github.io>

MolAICal 文章介绍：<https://doi.org/10.1093/bib/bbaa161>

MolAICal 中文博客：<https://molaical.github.io/entutorial.html>

MolAICal blogspot：<https://qblab.blogspot.com>

## 1. 简介

药物的定量构象关系(QSAR)包含线性回归和分类，在本教程中选用 STAT3 蛋白靶点的药物分子作为研究对象；STAT3 是治疗癌症的一个重要蛋白靶点，研究 STAT3 药物的属性，有助于设计合理的抗癌药物。

## 2. 工具

### 2.1. 所需软件

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

2) DRAGON: <http://www.taletc.mi.it/index.htm>

**注意：**除了用 DRAGON 算药物分子的描述符外，DRAGON 属于商业软件，你可以使用任何合适的软件算分子的描述符。

### 2.2. 操作所需的示例文件

1) 本教程所需的教程文件可以从以下网址下载：

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

## 3. 步骤

### 3.1. 计算分子描述符

1) 打开 DRAGON 软件，然后在文件夹“006-QSAR/ligands”中导入配体文件（如图 1 所示），本教程的配体文件是.bin 格式的文件，.bin 格式的文件是经过 HyperChem 软件优化过后的默认文件格式。你也可以优化自己的配体分子，然后保存成 Sybyl Mol2 格式的文件用于进一步的计算。

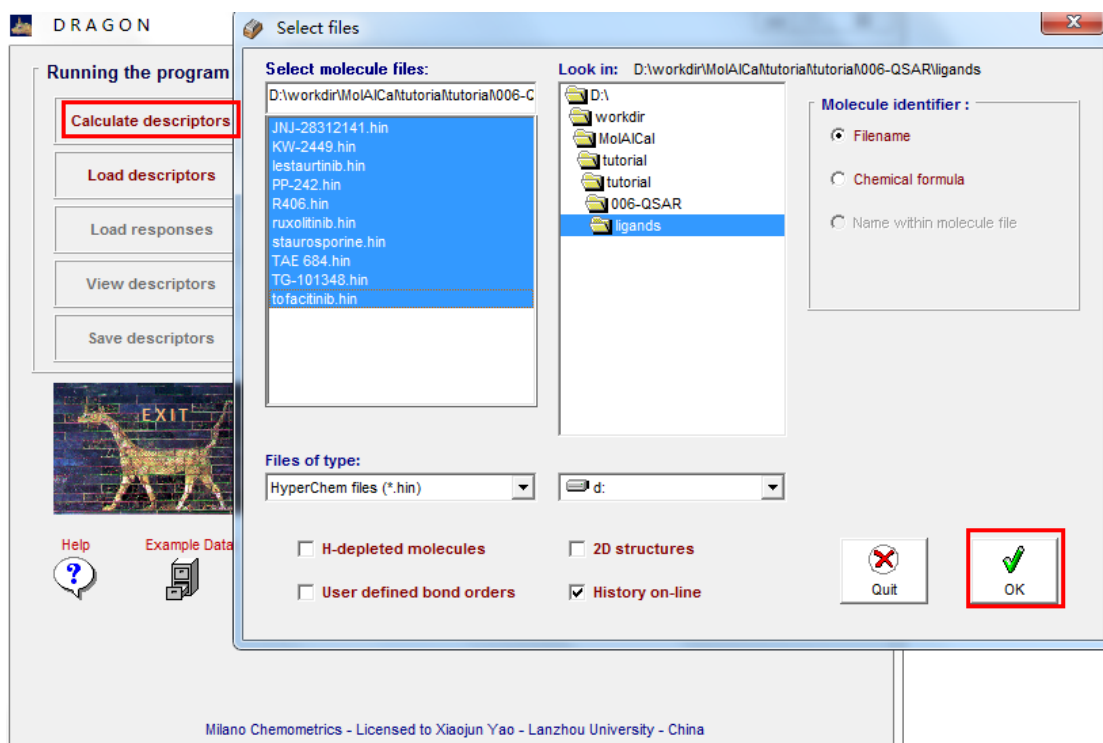


图 1. 使用 DRAGON 计算分子描述符

注意：你可以在这个数据库中检索蛋白受体的配体分子：[www.guidetopharmacology.org](http://www.guidetopharmacology.org) 等。

2) 将药物分子描述符保存并命名为“QSARMolDes.txt” (如图 2 所示)。

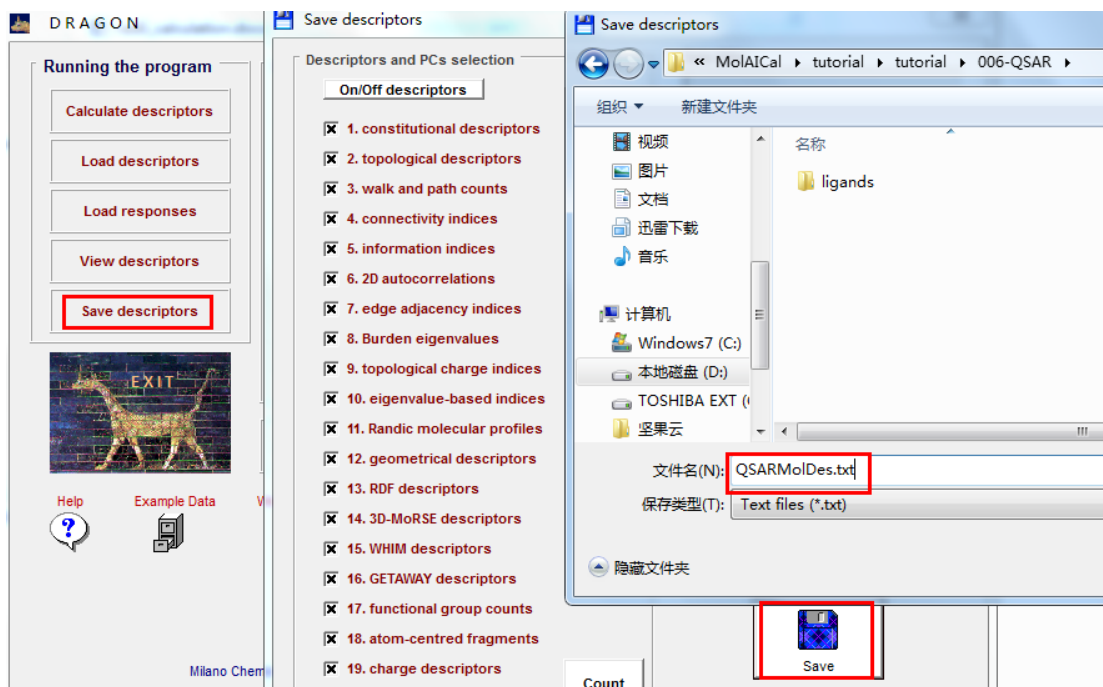


图 2. 保存并命名文件为“QSARMolDes.txt”

3) 使用 Excel 打开“QSARMolDes.txt”文件并设置相关参数 (如图 3 所示)。

Number of ligands

train set

on represents appointed train and validation sets, "off" means Loo, etc.

validation set

add pKd values of ligands

|    | A           | B            | C     | D      | E    | F     | G     | H     |    |
|----|-------------|--------------|-------|--------|------|-------|-------|-------|----|
| 1  | DRAGON data |              |       |        |      |       |       |       |    |
| 2  | 10          | 2            |       | 1364   | 0    | 0     |       |       |    |
| 3  | on          |              |       |        |      |       |       |       |    |
| 4  | 1           | 2            | 3     | 4      | 6    | 7     | 8     | 9     |    |
| 5  | 5           | 10           |       |        |      |       |       |       |    |
| 6  | No.         | MolID        | pKd   | MW     | AMW  | Sv    | Se    | Sp    | Ss |
| 7  | 1           | JNJ-28312    | 8     | 461.65 | 6.89 | 41.06 | 66.69 | 43.22 |    |
| 8  | 2           | KW-2449      | 7.12  | 333.45 | 7.25 | 29.57 | 45.75 | 30.95 |    |
| 9  | 3           | lestaurtinib | 8.43  | 439.5  | 8.14 | 36.41 | 54.57 | 37.69 |    |
| 10 | 4           | PP-242       | 7.96  | 308.38 | 7.91 | 25.46 | 39.36 | 26.3  |    |
| 11 | 5           | R406         | 8.46  | 470.51 | 8.25 | 36.01 | 58.71 | 37.1  |    |
| 12 | 6           | ruxolitinib  | 10.44 | 307.42 | 7.32 | 26.85 | 41.85 | 27.98 |    |
| 13 | 7           | staurospor   | 8.01  | 467.59 | 7.54 | 40.39 | 62.05 | 42.14 |    |
| 14 | 8           | TAE-684      | 7.8   | 615.3  | 7.41 | 50.74 | 83.06 | 54.23 |    |
| 15 | 9           | TG-101348    | 8.96  | 525.77 | 7.1  | 44.85 | 73.86 | 47.85 |    |
| 16 | 10          | tofacitinib  | 9.24  | 312.42 | 7.27 | 26.66 | 43.12 | 27.82 |    |
| 17 |             |              |       |        |      |       |       |       |    |
| 18 |             |              |       |        |      |       |       |       |    |

图 3. 在 Excel 中设置 QSAR 的参数

你必须在“QSARMolDes.txt”中严格按照格式设置参数, 在第二行的第一个数字是用于 QSAR 计算的配体分子数; 在第三行上的字符“on”代表指定了训练集和验证集, 第四行是训练集的序号, 第五行是验证集的序号, 此序号对应文件“QSARMolDes.txt”底下配体的序号 (如图 3 所示)。如果第三行是“off”, 则使用留一验证法 (LOO) 进行 QSAR 的计算, 在这种情况下, 第四、五行的数字可以省略, MolAICal 自动使用留一法指定训练集与验证集进行运算 (请参考示例文件: “QSARMolDes\_LOO.txt”)。除此之外, 实验值如 pKd 等应该加到第三列中 (如图 3 所示)

### 3.2. QSAR 计算

运行如下命令:

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

或

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

假如你想了解更多的 QSAR 参数, 请参考 MolAICal 的说明书。本教程仅仅包括 10 个配体。当 Q2 的运算值已经满足你的研究目的, 你可以通过“Ctrl + C”快捷键终止 MolAICal 的运行。最后的结果保存在“QSAROutFile.dat”文件中, 打开“QSAROutFile.dat”, 其具体运算结果的信息如下:

```

***** 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

```