

使用 MolAICal 进行药物 ADMET 的预测和分子属性的计算

作者: Qifeng Bai (update 2022-5-11)

更多教程 (含英文教程) 请见如下:

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

MolAICal 官方主页中国镜像: <https://molaical.gitee.io>

MolAICal 中文博客: <https://molaical.gitee.io/cntutorial.html>

1. 简介

由于毒性和不可接受的疗效, 导致大约有 40% 的候选药物研发失败。在这种情况下, 药物 ADMET (吸收, 分布, 代谢, 排泄和毒性) 的预测在候选药物研发中起到重要作用。在本教程中, 介绍了使用 MolAICal 和 FP-ADMET [1] 进行药物 ADMET 和分子属性的预测。有关更详细的 MolAICal 介绍, 请参考网址: <https://molaical.github.io>。

2. 方法和材料

2.1. 软件要求

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

注意: 请确保 MolAICal 安装正确!

2.2. 本教程所需文件

1) 所有在教程中的重要文件可以从该网址下载:

<https://gitee.com/molaical/tutorials/tree/master/018-admet>

Part I. 药物的 ADMET 计算

1. 安装 FP-ADMET

目前只支持 Linux 版本的 MolAICal 实现 Part I 的教程。

1) 点击并打开 FP-ADMET 的下载网址: [DownloadModel](#)
然后在打开的网址上依次打开 “[AImodels](#)→[ADMET](#)→[FPADMET](#)→[linux64](#)”, 并下载文件“fpadmet.tar.gz”

2) 将“fpadmet.tar.gz”移到文件夹“MolAICal-xxx/mtools”中。其中“MolAICal-xxx”是 MolAICal 的解压安装目录。而“mtools”是“fpadmet.tar.gz”要移入的目录。

3) 解压文件

```
#> tar -xzvf fpadmet.tar.gz
```

切换到目录 fpadmet:

```
#> cd fpadmet
```

4) 安装 fpadmet

```
#> chmod +x install.sh
```

```
#> ./install.sh
```

现在已成功安装 fpadmet。

2. 计算药物的 ADMET 步骤

2.1. ADMET 计算所需文件的格式

用户需要输入正确的文件格式, 如图 1 所示: 每一行包含一个分子的 SMILES 序列和分子名称, 其中第一列是分子的 SMILES 序列, 第二列是分子名称; 且第一列与第二列之间至少含有 1 个空格。

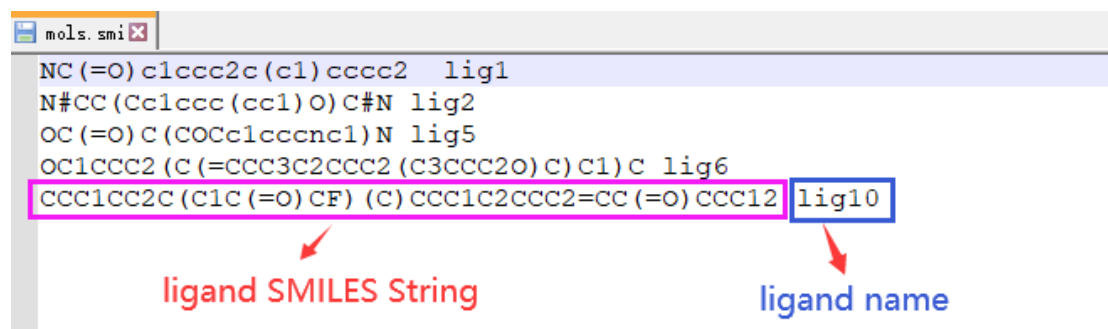


图 1. 输入文件的格式

2.2. 计算所有 SMILES 格式分子的 ADMET 条目

切换到目录 018-admet, 并运行如下命令:

```
#> molaical.exe -model fpadmet -s all -i mols.smi -o results.dat
```

结果的解释：命令运行完毕后将输出一个结果文件“**results.dat**”。打开文件“**results.dat**”，可以发现里面包含了所有分子的所有 ADMET 条目的计算结果。例如，预测 Blood-brain-barrier penetration (条目 2) 和 pKa dissociation constant (条目 50) 的结果如下：

#####

2: Blood-brain-barrier penetration

Predicted Confidence Credibility

lig1 Yes 0.93 0.32

.....

50: pKa dissociation constant

Predicted quantile_0.025 quantile_0.975

lig1 5.19 -1.60 13.06

#####

“lig1”是药物或配体的名称。化合物“lig1”的标签“Yes”表明该化合物预测适用于血脑屏障穿透。置信值为 0.93 表示分类器非常确定，并且预测可能是单个标签。可信度相对较低的值 (0.32) 表明：像“lig1”这样的化合物在训练集中没有足够的代表性，用户应谨慎处理预测。对于第 50 项的回归示例，计算了 95% 的预测区间 (pKa 在 0.025 和 97.5% 的预测)，并给出了单个观察的预测范围。正如 FP-ADMET 的原始论文所说：狭窄的预测区间表明与预测相关的不确定性较低[1]。有关预测结果的更详细解释，请查看与本教程文档放在一起的文件夹“doc”中名为“ADMETModels-doc.pdf”、“FP-ADMET.pdf”和“FP-ADMET-SI.pdf”的文件。

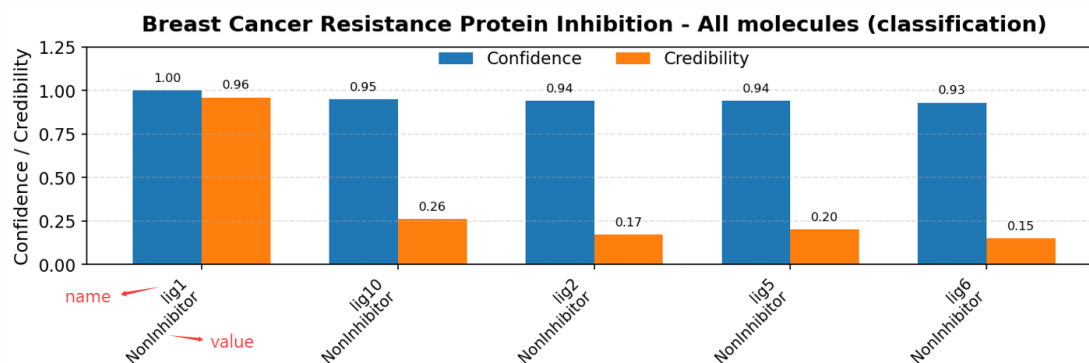
为可视化结果，以下命令可实现两种绘图方式：既可针对每个分子生成 58 个属性值的图表（每个分子拥有以其名称命名的独立文件夹），也可可为每个属性条目生成所有分子的综合图表（存储在“merge_lig”文件夹中）：

```
#> molaical.exe -call run -c sfile -i m_admetp.py -i results.dat --merge
```

该命令将执行以下操作：

- ✧ 在配体同名文件夹中（例如：lig1, lig2...）为每个分子生成 58 个属性值的独立图片；
- ✧ 在“merge_lig”文件夹中为每个属性条目生成所有分子的综合图片；
- ✧ 结果文件同时提供 PNG 图片和 PDF 文档两种格式，且两种格式的内容完全一致。

示例图如下所示。横坐标代表分子名称及其对应的 ADMET 值。请结合计算结果来解读此图：



2.3. 计算所有 SMILES 格式分子的单个 ADMET 条目

切换到 018-admet, 并使用如下命令:

```
#> molaical.exe -model fpadmet -s single -i mols.smi -n 1
```

结果的解释: 由于选择了任务项 1, 它将生成一个名为“1.dat”的文件。如果选择任务项 2, 它将生成一个名为“2.dat”的文件。如果选择任务项 3, 它将生成一个名为“3.dat”的文件.....因此生成的结果文件的前缀名称与所选任务项相同。打开文件“1.dat”, “1.dat”的内容如下:

```
#####  
Predicted Confidence Credibility  
lig1 Negative 0.77 0.31  
lig2 Negative 0.90 0.59  
lig5 Negative 0.80 0.35  
lig6 Negative 0.91 0.64  
lig10 Negative 0.96 0.86  
#####
```

它仅显示了 ADMET 中所有分子的“1: Anticommensal Effect on Human Gut Microbiota”项的计算结果。要了解 ADMET 的所有项目, 请查看 **MolAICal 手册**和本教程附录, 或与本教程文件放在一起的文件夹“doc”中名为“**ADMETModels-doc.pdf**”、“FP-ADMET.pdf”和“FP-ADMET-SI.pdf”的文件。

要查看可视化结果, 可在先前生成的“merge_lig”文件夹中根据属性条目名称查找对应图片。

Part II. 使用 MolAICal 计算 ADME 和分子属性

本部分教程适用 Linux 或 Windows 版的 MolAICal。

在 Part II 教程中, 将展示 MolAICal 计算分子 ADME 和分子属性的步骤, 具体如下:

1) 如果“-s”是 0, 将会输出药物分子的所有属性信息

```
#> molaical.exe -model adme -s 0 -i "O=C(C)Oc1ccccc1C(=O)O"
```

2) 如果“-s”是 1, 将会输出一个药物分子简短的 ADME 信息

```
#> molaical.exe -model adme -s 1 -i "O=C(C)Oc1ccccc1C(=O)O"
```

例如, 输出结果如下:

1. Pharmacokinetics:

GI absorption: High

BBB permeant: Yes

2. Lipinski Rule of 5 Violations:

No violations found

3. Medicinal Chemistry.

PAINS filter: False

Brenk filter: True

3) 如果“-s”是 2, 将会输出分子的一些基础信息, 如分子量、氢键供体数等。

#> molaical.exe -model admepr -s 2 -i "O=C(C)Oc1ccccc1C(=O)O"

更多属性预测的条目信息, 请参考 MolAICal 手册。

Reference

1. Venkatraman V. FP-ADMET: a compendium of fingerprint-based ADMET prediction models. J Cheminform. 2021;13(1):75.

附录

FP-ADMET 的 58 个 ADMET 条目如下:

- 1: Anticommensal Effect on Human Gut Microbiota
- 2: Blood-brain-barrier penetration
- 3: Oral Bioavailability
- 4: AMES Mutagenecity
- 5: Metabolic Stability
- 6: Rat Acute LD50
- 7: Drug-Induced Liver Inhibition
- 8: HERG Cardiotoxicity
- 9: Haemolytic Toxicity
- 10: Myelotoxicity
- 11: Urinary Toxicity
- 12: Human Intestinal Absorption
- 13: Hepatic Steatosis
- 14: Breast Cancer Resistance Protein Inhibition
- 15: Drug-Induced Choleostasis
- 16: Human multidrug and toxin extrusion Inhibition

17: Toxic Myopathy
18: Phospholipidosis
19: Human Bile Salt Export Pump Inhibition
20: Organic anion transporting polypeptide 1B1 binding
21: Organic anion transporting polypeptide 1B3 binding
22: Organic anion transporting polypeptide 2B1 binding
23: Phototoxicity human
24: Phototoxicity in vitro
25: Respiratory Toxicity
26: P-glycoprotein Inhibition
27: P-glycoprotein Substrate
28: Mitochondrial Toxicity
29: Carcinogenicity
30: DMSO Solubility
31: Human Liver Microsomal Stability
32: Human Plasma Protein Binding
33: hERG Liability
34: Organic Cation Transporter 2 Inhibition
35: Drug-induced Ototoxicity
36: Rhabdomyolysis
37: T_{1/2} Human
38: T_{1/2} Mouse
39: T_{1/2} Rat
40: Cytotoxicity HepG2 cell line
41: Cytotoxicity NIH 3T3 cell line
42: Cytotoxicity HEK 293 cell line
43: Cytotoxicity CRL-7250 cell line
44: Cytotoxicity HaCat cell line
45: CYP450 1A2 Inhibition
46: CYP450 2C19 Inhibition
47: CYP450 2C9 Inhibition
48: CYP450 2D6 Inhibition
49: CYP450 3A4 Inhibition
50: pK_a dissociation constant
51: logD Distribution coefficient (pH 7.4)
52: logS
53: Drug affinity to human serum albumin
54: MDCK permeability
55: 50% hemolytic dose
56: Skin penetration
57: CYP450 2C8 Inhibition
58: Aqueous Solubility (in phosphate saline buffer)