

# 使用 MolAICal 两步实现对 Mpro 蛋白活性口袋的虚拟筛选

作者: MolAICal (update 2021-12-14)

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 文章介绍: <https://arxiv.org/abs/2006.09747> 和  
<https://doi.org/10.1093/bib/bbaa161>

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

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

MolAICal QQ 学术讨论群: 1151656349

## 1. 简介

在本教程中, 介绍了使用已知数据库 (如 ZINC 数据库等) 对 Mpro 活性口袋进行快速药物虚拟筛选的方法。本教程学习的前提是你会用 Autodock Vina 处理蛋白质结构。如果你对蛋白质结构处理不熟悉, 可以参考如下网站所提供的教程:

<https://github.com/MolAICal/documents/tree/master/tutorials/002-AIVS>.

## 2. 工具

### 2.1. 所需软件下载地址

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

国内镜像 MolAICal: <https://molaical.gitee.io>

### 2.2. 操作示例文件

1) 所有用到的操作教程文件均可在下面的网站下载:

<https://github.com/MolAICal/tutorials/tree/master/003-VS>

2) “ligandSet.mol2”文件包含 16 个从 ZINC 数据库中选择的配体作为示例。你也可以选择其它数据库进行练习。

3) 使用 PDBQT 格式的 Mpro 蛋白质文件“pro.pdbqt”用来做分子对接。

## 3. 操作流程

转至教程目录:

```
#> cd 003-VS
```

1. 将配体分子集“ligandSet.mol2”分割成单个分子文件。如果你所选的药物数据库已经被分割成了单个分子文件，这一步可以省略。

```
#> molaical.exe -tool mol2 -w split -n 1 -v true -m number -i ligandSet.mol2 -o splitdir
```

**另外选项：**当 mol2 文件的第二行是有效字符的时候，你也可以使用第二行的字符作为文件名，命令如下：

```
#> molaical.exe -tool mol2 -w split -n 1 -v true -i ligandSet.mol2 -o splitdir
```

**注意：**在 ligandSet.mol2 中的每个配体应该是完整的分子，如果配体缺氢原子，MolAICal 将可能不会生成对应配体的 PDBQT 文件，用户可以使用 UCSF Chimera 进行加氢，或使用 MolAICal 中的命令，如下（用户需要把 1.mol2 文件替换成自己的文件名）：

```
#> molaical.exe -tool format -i E:/1.mol2 -o E:/1.pdbqt
```

2. 运行虚拟筛选命令。

```
#> molaical.exe -dock vs -i splitdir -nc 3
```

**注意：**如果用户直接使用 PDBQT 格式的配体进行虚拟筛选，用户可以使用下面的命令：

```
#> molaical.exe -dock vs -i splitdir -b off -k pdbqt -nc 3
```

**-nc:** 代表 CPU 核心数。

**-b:** 是否进行分子格式转化，“on”代表进行转化，“off”代表不进行分子格式转化。默认值是“on”。

**-k:** 指定虚拟筛选的分子格式，默认值是 mol2 的分子格式。

## 4. 结果

### 4.1 查看结果

用户可以使用 Pymol (<http://www.lfd.uci.edu/~gohlke/pythonlibs>) 直接载入 PDBQT 格式的文件。这里使用 UCSF Chimera 来查看虚拟筛选的结果。

打开 003-VS/splitdir

1) 加氢 (可选)

```
#> molaical.exe -dock addh -i 1_out.pdbqt
```

2) 更改格式“pdbqt”成“pdb”

```
#> molaical.exe -dock pdbqt2pdb -i 1_out.pdbqt
```

上述命令将产生一个名为“1\_out.pdb”的文件。然后使用 UCSF Chimera 打开配体分子“1\_out.pdb”和受体分子“protein.pdb”。图 1 显示了对接筛选的结果，表明了 MolAICal 可以虚拟筛选到合适的配体分子。

**注意：**图 1 的显示方面有些步骤省略了，假如用户想得到图 1 这样 surface 的图像，可以使用 UCSF Chimera 工具栏的选项：“Actions→Surface→show”

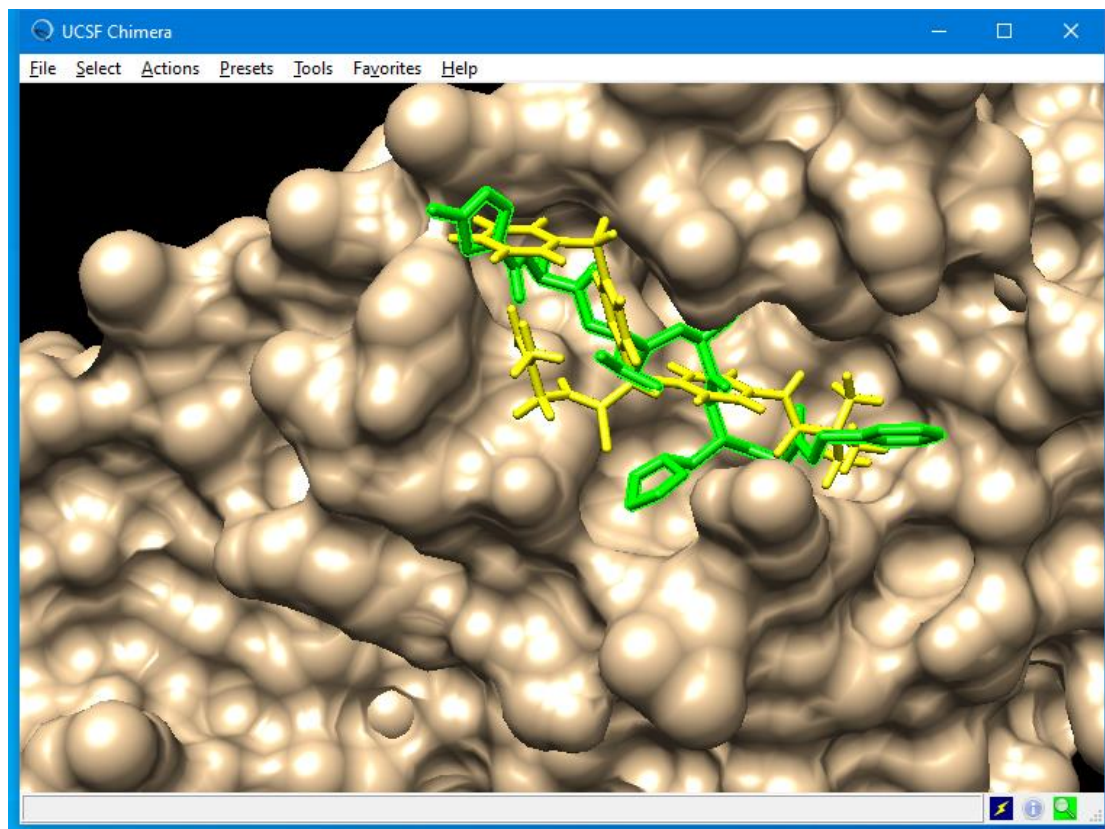


图 1. 绿色分子代表 SARS-CoV-2 Mpro 的抑制剂 N3，黄色分子代表筛选出的分子

#### 4.2 对虚拟筛选的结果进行排序

切换到 003-VS，然后运行如下命令：

```
#> python /home/bai/MolAICal-xxx/scripts/printScore.py 'splitdir/*out.pdbqt'
```

这个命令将显示分子名称和对应的打分，假如用户想保存输出的结果，可以使用如下命令：

```
#> python /home/bai/MolAICal-xxx/scripts/printScore.py 'splitdir/*out.pdbqt' > results.log
```

**说明：**“/home/bai/MolAICal-xxx”是实际 MolAICal 的安装目录。所有的脚本文件都保存在 MolAICal 子文件夹“scripts”中。

假如你在 **Windows** 环境下：

使用 Excel 打开“results.log”，其中分隔符“Separator”选择空格。或者直接将“results.log”的

内容复制到 Excel 中。在第二列，选择所有数据，并且根据需要选择工具栏中的“Sort Largest to Smallest”（降序）或“Sort Smallest to Largest”（升序）（见图 2）。

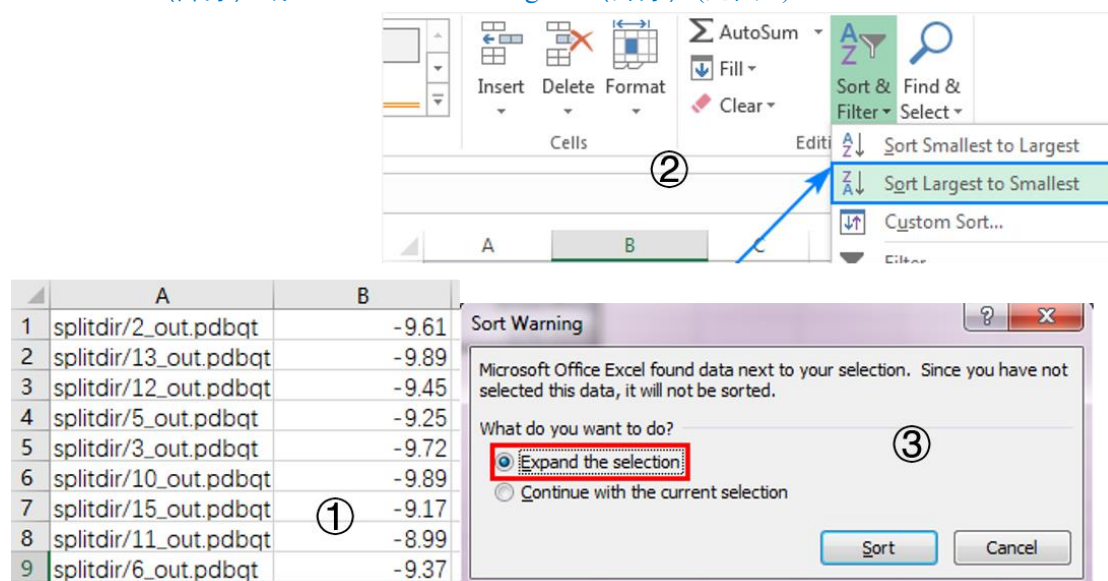


图 2. 排序结果

假如你在 Linux 环境下，用户可以使用如下命令：

```
#> sort -n -t ' ' -k 2r results.log > rank.dat
```

说明：参数“2r”是升序排列，而“1r”是降序排列。

### 4.3 提取排名靠前的分子到新建的文件夹中

假如用户想将排名靠前的分子移动到新建的文件夹中，以便于分析结果，本教程提供了一种方法，下面的命令是将 2 个排名靠前的分子移到名为“results”的文件夹中：

```
#> python /home/bai/MolAICal-xxx/scripts/molaicaldTopResults.py "splitdir/*out.pdbqt" 2 results
```

运行上述命令，2 个排名靠前的分子将被移动到“results”文件夹中