

# 使用 MolAICal 的深度学习模型和分子对接程序进行药物的虚拟筛选

作者：MolAICal （update 2020-07-05）

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/cntutorial.html>

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

## 1. 简介

一种新药的研发大概需要耗费 26 亿美元。即使有大量资金的投入，90%的新药仍会在临床试验和获批上市过程中夭折[1]。本教程介绍了 MolAICal 基于人工智能和分子对接进行药物虚拟筛选的流程，其中 model.pdb 是优化的蛋白质模型文件，你可以替换成自己的蛋白质模型。此方法将帮助药物学家、化学家及其它领域的科学家根据靶点的活性口袋合理设计药物。

## 2. 工具

### 2.1. 所需软件下载地址

- 1) MolAICal (win64 or linux64): <https://molaical.github.io>
- 2) UCSF Chimera: <https://www.cgl.ucsf.edu/chimera/>
- 3) MGLTools: <https://ccsb.scripps.edu/mgltools/downloads/>
- 4) Python: <https://www.python.org/>
- 5) Pymol: <http://www.lfd.uci.edu/~gohlke/pythonlibs>

前四个软件的安装相对简单，可根据提示轻松完成安装。Pymol 的安装需要 numpy, pmw, pymol\_launcher 和 pymol 组件，上述四个组件的版本需要与你操作系统上选择的 Python 版本保持一致，这些组件可在下面的网站下载：

<https://www.lfd.uci.edu/~gohlke/pythonlibs/#numpy>

<https://www.lfd.uci.edu/~gohlke/pythonlibs/#pymol-open-source>

下载后把对应版本的 numpy, pmw, pymol\_launcher 和 pymol 组件放到同一个文件中，使用以下命令安装 Pymol:

```
#> pip install --no-index --find-links="%CD%" pymol_launcher
```

Pymol 软件名称为“pymol.exe”将安装在 Python 目录下的“Scripts”文件夹中。你可以在桌面创

建此软件的快捷方式。确保所有的软件正确安装。

## 2.2. 操作示例文件

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/002-AIVS>

## 3. 操作流程

这一步是在分子对接前的蛋白质结构处理。如果你熟悉 Autodock vina, 请自动忽略该步。你可以在 <https://youtu.be/-GVZP0X0Tg8> 网站上观看或者在 <http://vina.scripps.edu/tutorial.html> 网站上下下载该步的教程视频进行学习。为了使本教程更加完善, 处理蛋白结构的步骤陈述如下:

### 3.1. 使用 UCSF Chimera 将蛋白质和配体结构分开

1) 首先, 在 UCSF Chimera 中载入复合物结构。File→Open→model.pdb (如图 1 所示)

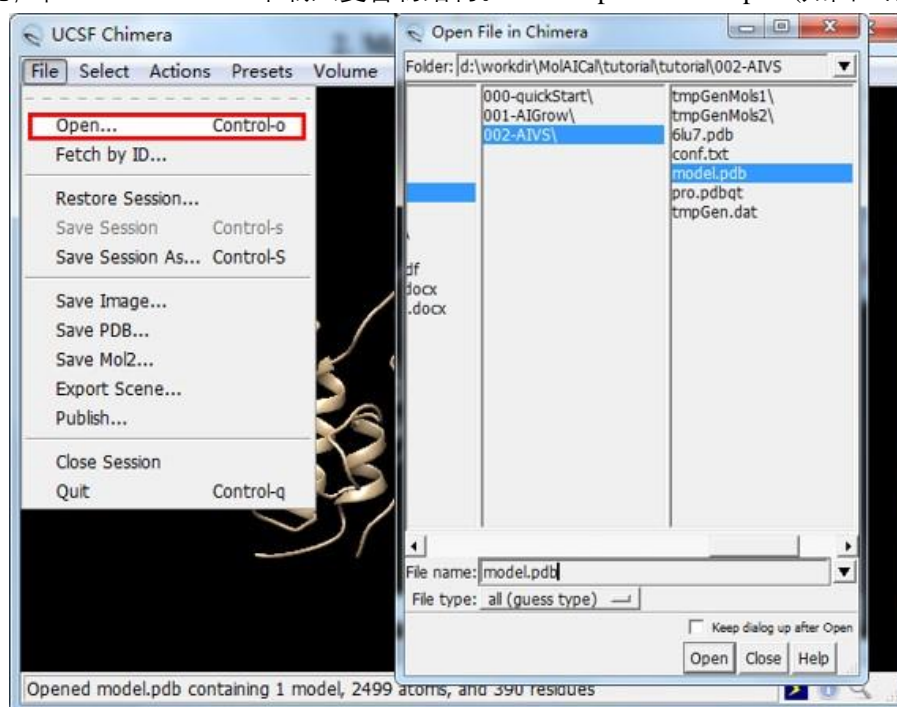


图 1. 载入蛋白结构文件

2) 选择配体 LIG 并将其删除 (如图 2 所示)。使用图 2 中相同的方法将水 HOH 选中并删除。

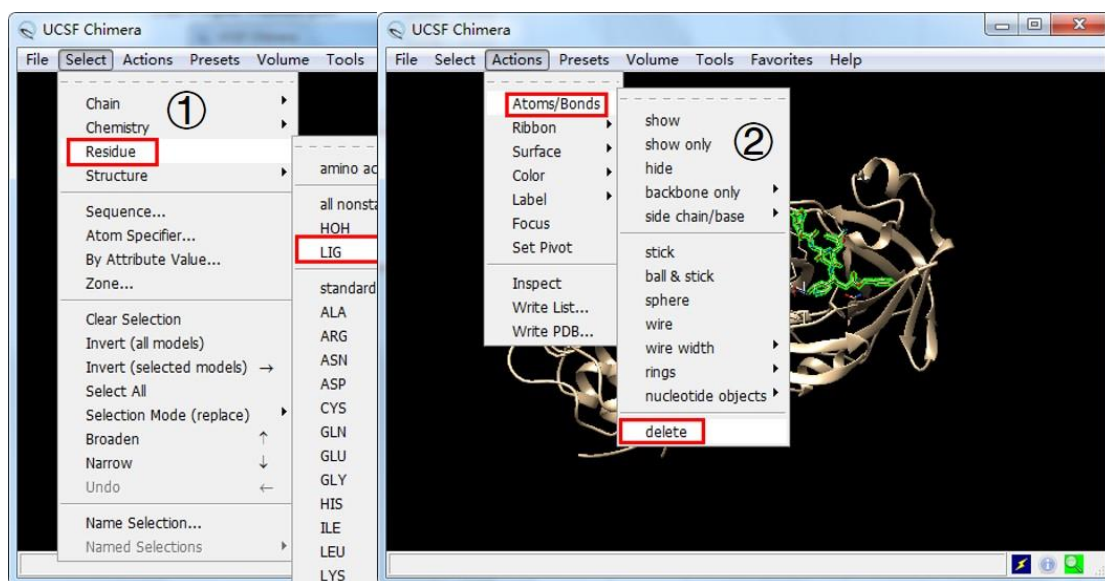


图 2. 选中并删除配体

- 3) 保存没有配体的蛋白质结构并且命名为“protein.pdb” (如图 3 所示)。

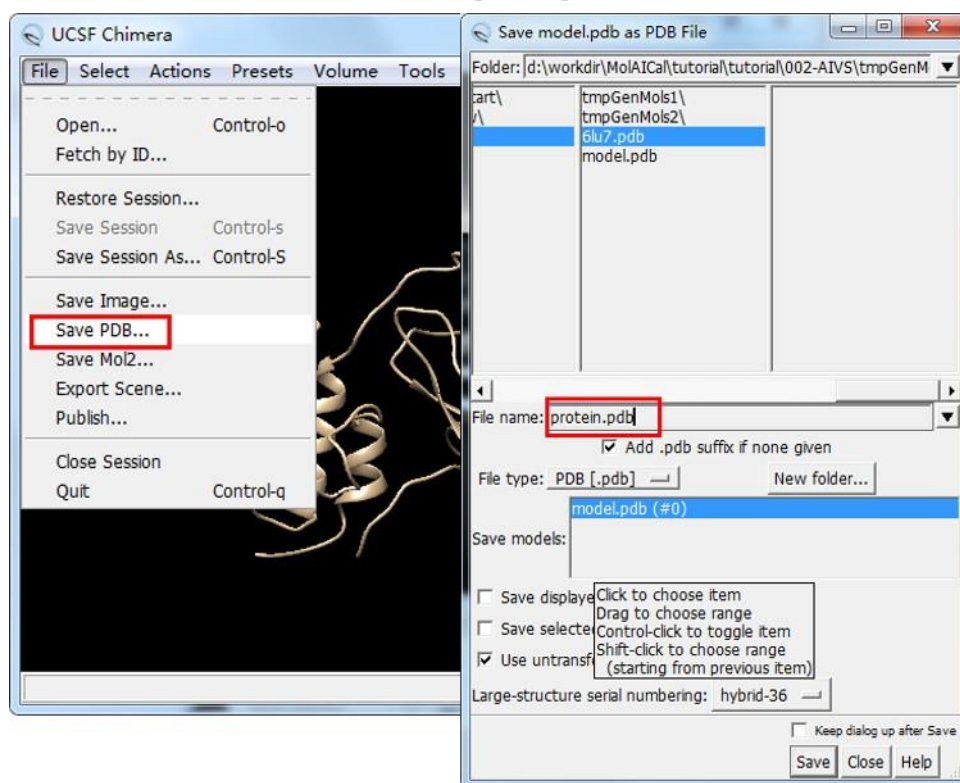


图 3. 保存蛋白质结构

- 4) 关闭本次会话，重新载入“model.pdb”，选择配体，反选并删除反选的蛋白 (如图 4 所示)。

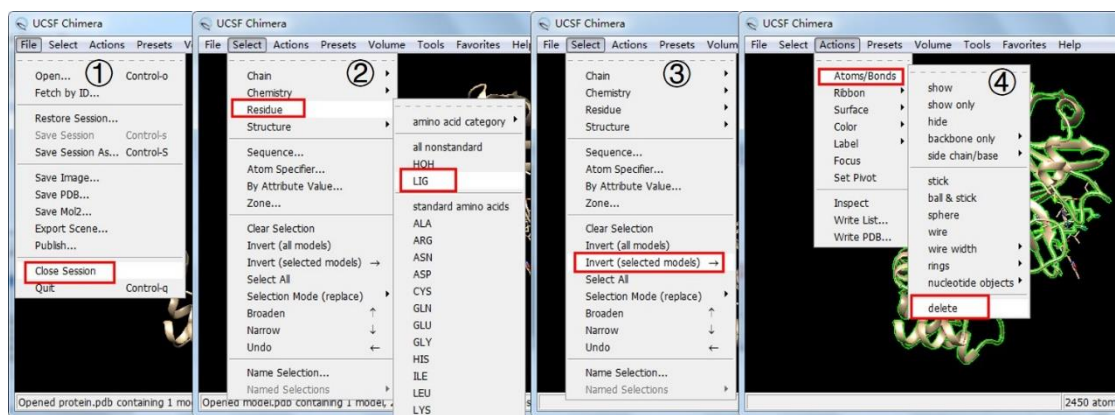


图 4. 选中并删除受体蛋白

5) 将配体文件保存为“ligand.pdb”(如图 5 所示)。

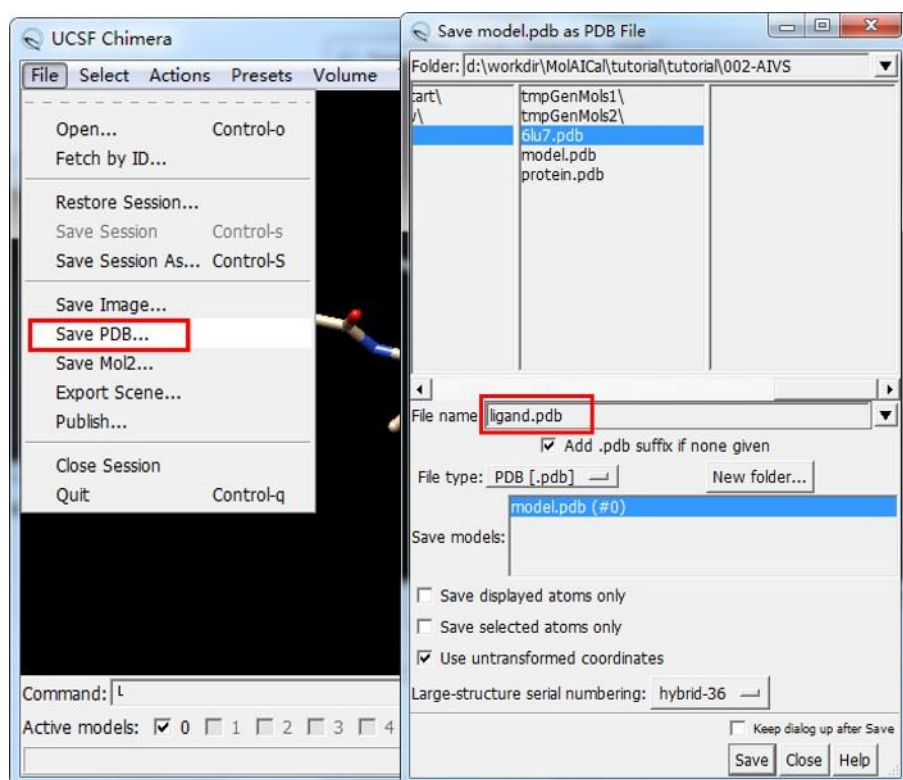


图 5. 保存配体文件

### 3.2. 计算盒子质心和长度

1. 参照上述步骤选择配体或者重新载入“ligand.pdb”并选择配体。然后选择距离工具：Tools→Structure Analysis→Distance (如图 6 所示)：

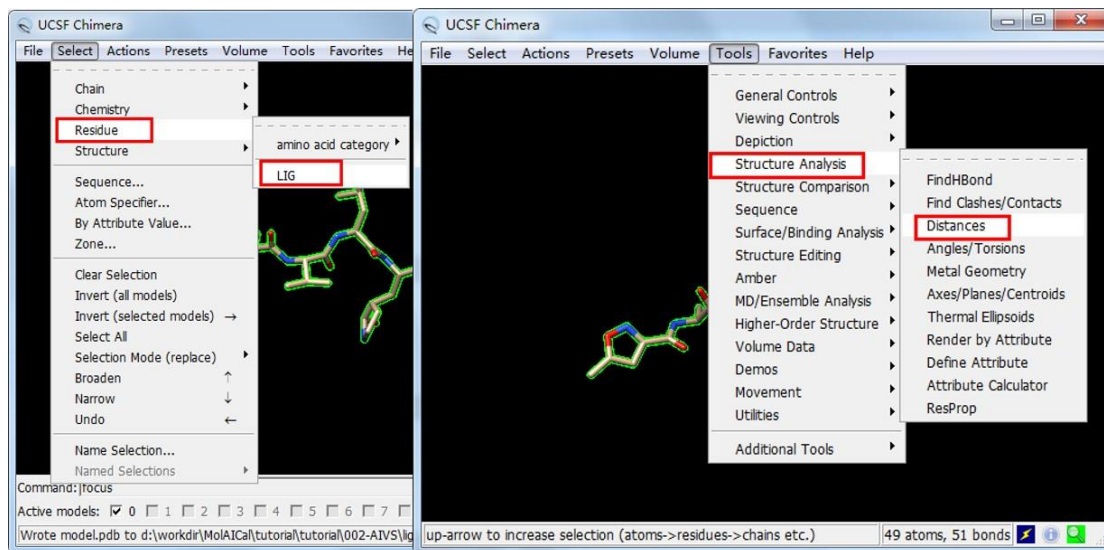


图 6. 选择距离工具

2. 根据配体计算蛋白质活性口袋的质心坐标 (如图 7 所示)。

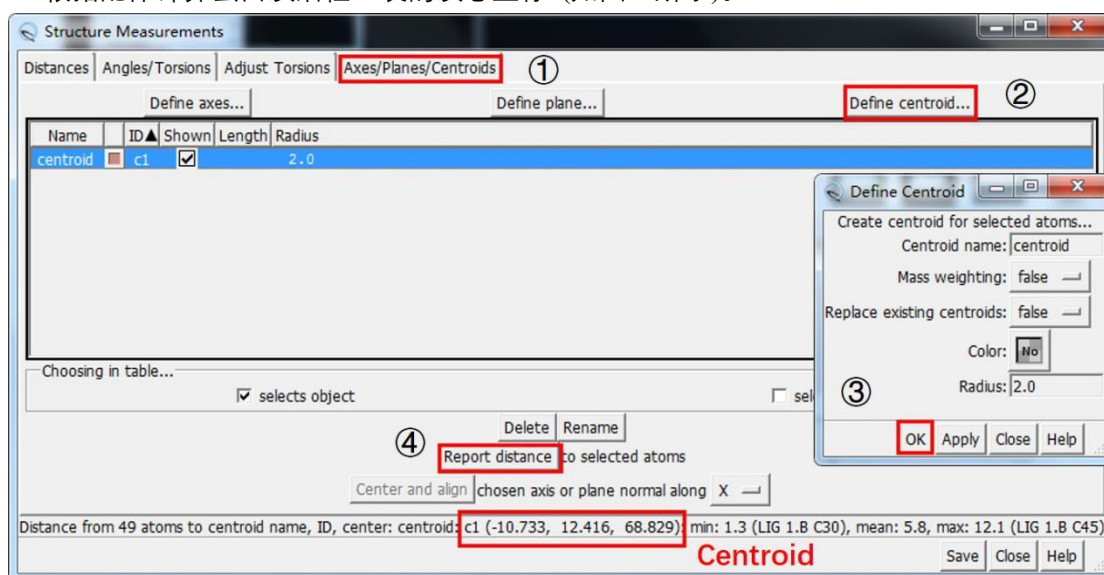


图 7. 获得质心坐标

创建 “conf.txt” 并将质心坐标写入该文件：

```
center_x = -10.733
center_y = 12.416
center_z = 68.829
```

注意：配置文件的文件名 “conf.txt” 是固定的。如果你使用其它字母创建文件名，MolAICal 将不能识别该文件名。用户可以根据自己的实际参数值，直接修改教程例子中的”conf.txt”文件。

3. 设置对接盒子的体积

计算最终盒子尺寸。你可以将 X, Y, Z 长度分别设置为 25, 30, 25。在 MolAICal 中使用下文



提到的命令生成“box.bild”(注意: X, Y, Z 坐标的双引号是必须添加的, X, Y, Z 坐标之间的间隔为一个空格。):

1) 执行以下命令, 获得“box.bild”:

```
#> molaical.exe -tool box -i "-10.733 12.416 68.829" -l "25.0 30.0 25.0" -o "box.bild"
```

2) File→open, 然后打开“box.bild”, 检查生成的盒子大小是否合适 (如图 8 所示)。

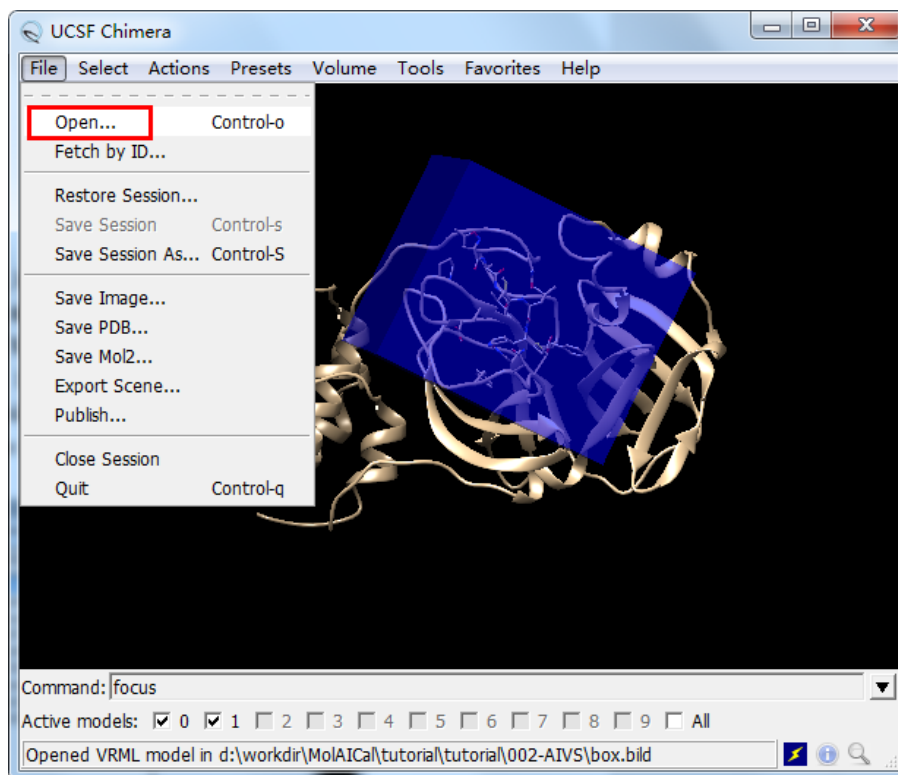


图 8. 使用 UCSF Chimera 打开 box.bild

如上图所示盒子大小 25, 30, 25 是合适的, 因此确定最终质心参数为 -10.733, 12.416, 68.829, 最终盒子沿 X, Y, Z 轴的长度为 25.0, 30.0, 25.0。

**注意:** 如果你用 VMD 软件计算了几何中心, 最终的中心点参数将是 -10.86, 12.57, 68.82。这两种方法得到的结果都可以用于本教程, 本教程暂使用 UCSF Chimera 算出来的质心坐标。

### 3.3. 虚拟筛选前将蛋白质结构转换为 PDBQT 格式

1. 打开“AutoDockTools”, File→Read Molecule→protein.pdb, 加上极性氢 (如图 9 所示)。

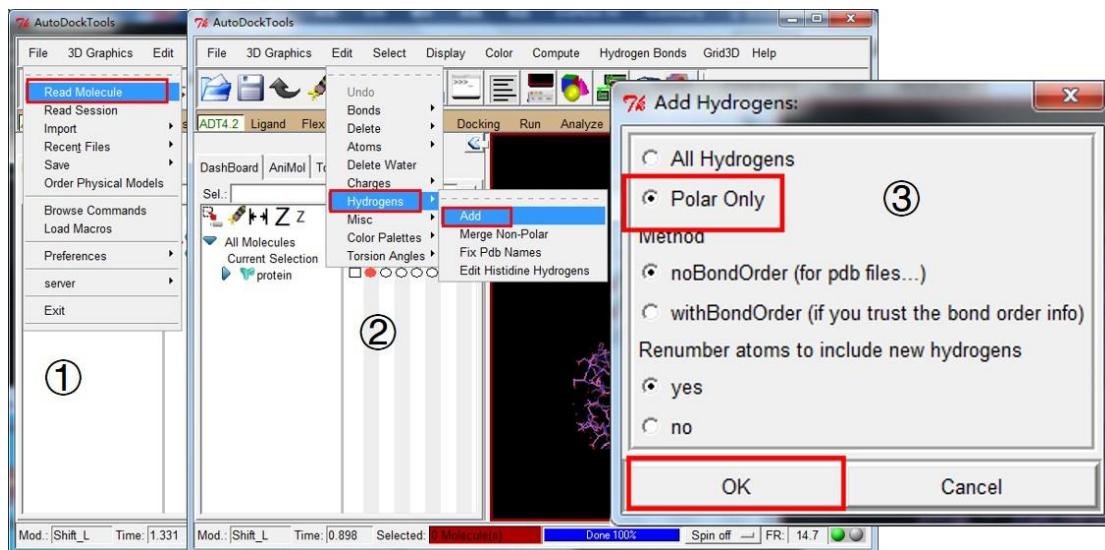


图 9. 加极性氢

2. 将蛋白保存为 PDBQT 格式。Grid→Macromolecule→Choose..., 然后点击“protein”选择“Select Molecule”, 将蛋白结构保存为“pro.pdbqt” (如图 10 所示)。

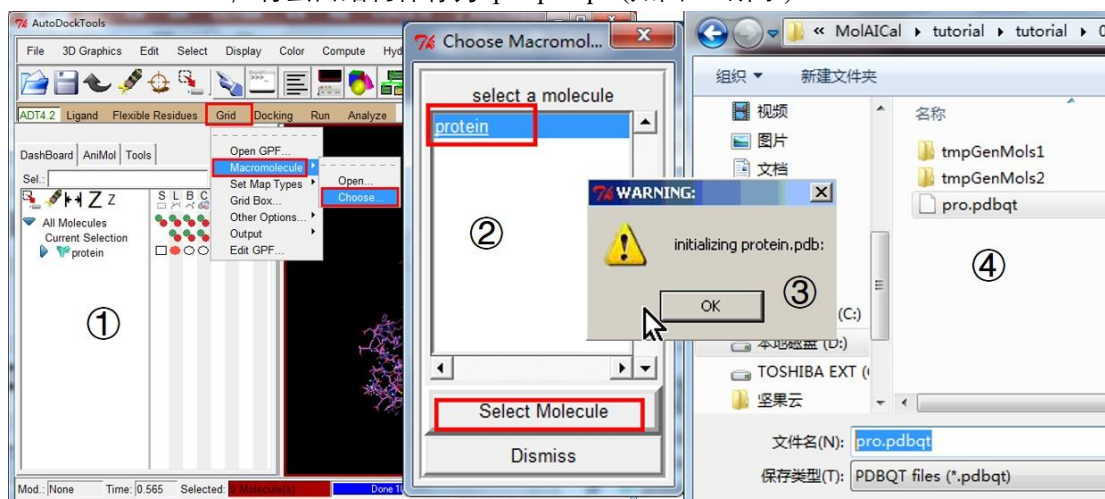


图 10. 用 PDBQT 格式保存蛋白

到处为止所有文件准备就绪。

### 3.4. 用深度学习模型和分子对接进行虚拟筛选

```
#> cd 002-AIVS
```

最后在后台运行以下命令：

**Linux 系统:**

```
#> molaical.exe -dock AI -s ZINCmol -n 6 -nf 3 -nc 3 >& vs.log &
```

**-n:** 代表对接产生的总分子数目

**-nf:** 单个文件夹中包含的分子数量

**-nc:** 执行命令使用的 CPU 数

### Windows 系统 (使用 PowerShell):

```
#> molaical.exe -dock AI -s ZINCMol -n 6 -nf 3 -nc 3
```

如果要在后台运行,请执行下面的命令:

```
#> powershell -windowstyle hidden -command "molaical.exe -dock AI -s ZINCMol -n 6 -nf 3 -nc 3"
```

如果你想依据已知的药物数据库进行经典的虚拟筛选,可以参考 MolAICal 教程中药物设计部分的第三部分(<https://molaical.github.io/tutorial.html>)。

## 4. 结果

你可使用 Open Babel 将分子的 PDBQT 格式转化为 PDB 格式,然后在使用 UCSF Chimera 打开查看结果。Pymol 可以直接载入 PDBQT 格式的分子。本教程使用 Pymol 软件(<http://www.lfd.uci.edu/~gohlke/pythonlibs>)展示结果(如图 11 所示)。结果显示 MolAICal 通过深度学习模型和分子对接程序获得了与现有配体相似的类似物。

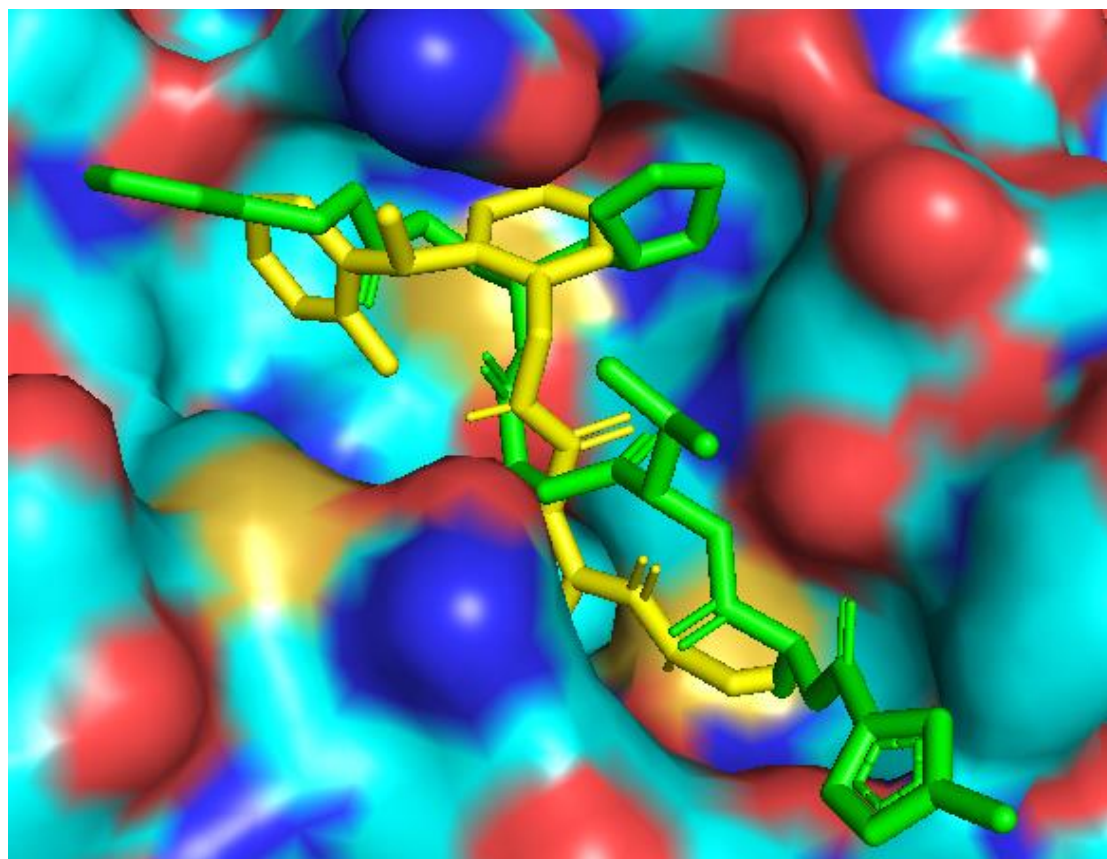


图 11. 绿色配体 N3 是原蛋白受体的抑制剂,黄色配体由深度学习模型和分子对接计算而得。



## 参考文献

1. Fleming N. How artificial intelligence is changing drug discovery, *Nature* 2018;557:S55-S55.