

# 使用 MolAICal 和 Pafnucy 模型对药物亲和力进行预测

作者：Qifeng Bai (update 2021-10-16)

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

深度学习模型可以用来计算配体和受体之间的亲和力，将 Pafnucy 嵌入到 MolAICal 用以计算配体的亲和力。这个案例可以帮助使用者快速了解深度学习是如何计算配体和蛋白的亲和力，有关 MolAICal 的更多详细信息，请参考网址：<https://molaical.github.io>。

## 2. 软件和所需数据

### 2.1 软件需求

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

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

2) UCSF Chimera: <https://www.cgl.ucsf.edu/chimera/>

3) VMD: <https://www.ks.uiuc.edu/Research/vmd/>

### 2.2 案例文件

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/015-bindingaffinityPafnucy>

## 3. 步骤

### 3.1 安装 Pafnucy 模型

现阶段只支持 linux 系统

1) 点击并打开网页：**DownloadModel**

在已打开的网页中根据路径 `AIModels/BindingAffinity/pafnucy/linux`，下载文件名

是“pafnucy.tar.gz”。

2) 把 pafnucy.tar.gz 移动到 MolAICal-xxx/mtools 目录下，“MolAICal-xxx”是你解压并安装 MolAICal 的目录，“mtools”是指定目录。

3) 解压目录

```
#> tar -xzf pafnucy.tar.gz
```

进入到 pafnucy 目录下

```
#> cd pafnucy
```

4) 安装 Pafnucy 模型

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

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

现已成功安装 Pafnucy 模型

## 3.2 计算一个复合物的亲和力

### 3.2.1 准备蛋白文件

1) 从 PDB 数据库下载“3ui7.pdb”文件（假设已加氢和加电荷。其中，对于加电荷，在 UCSF Chimera 中，标准残基使用 Amber ff14SB，非标准残基和配体使用 AM1-BCC）。使用 UCSF Chimera 打开文件，把“3ui7.pdb”的 A 链保存为“3ui7\_A.mol2”（如图 1 所示）。

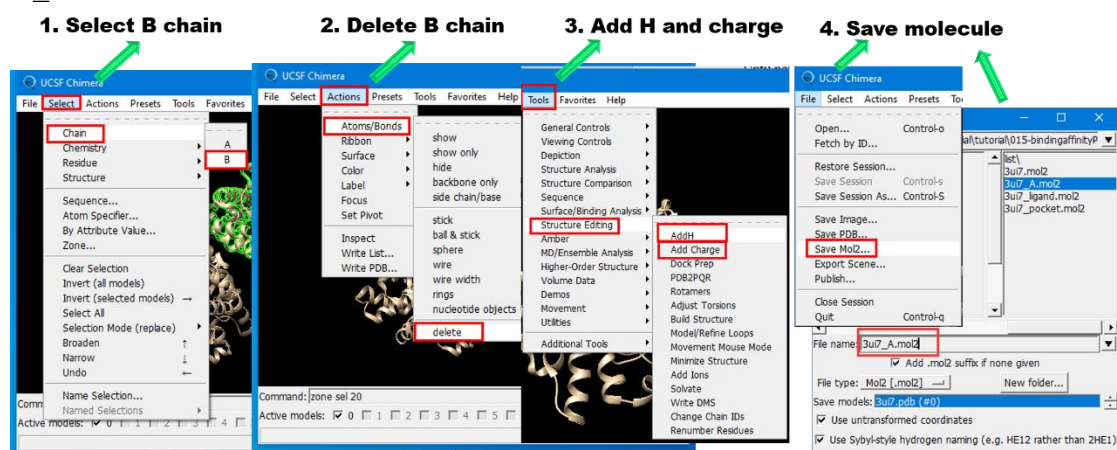
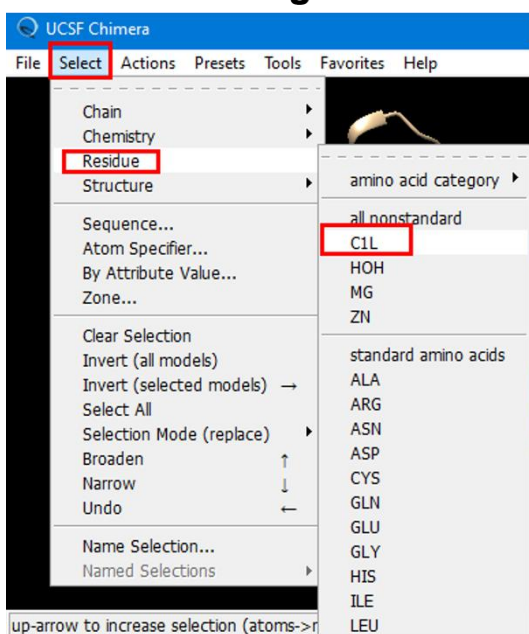


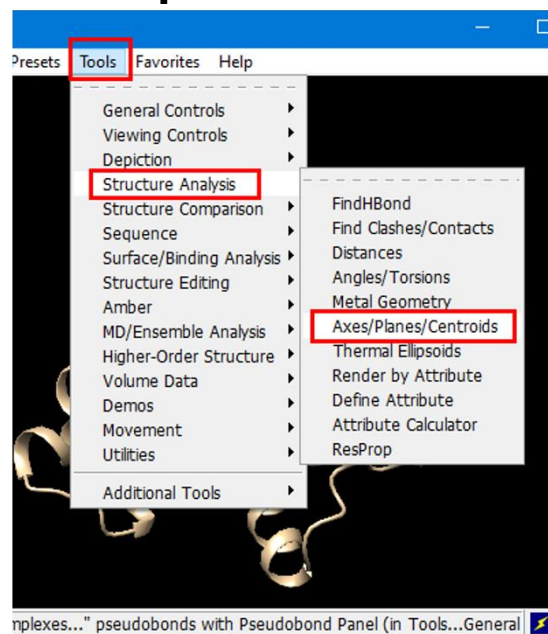
图 1. 处理蛋白文件

2) 得到配体几何中心（如图 2 所示），几何中心坐标为 5.411, 12.284, 43.559

## 1. Select ligand



## 2. open centroids



## 3. Get coordinates as follows

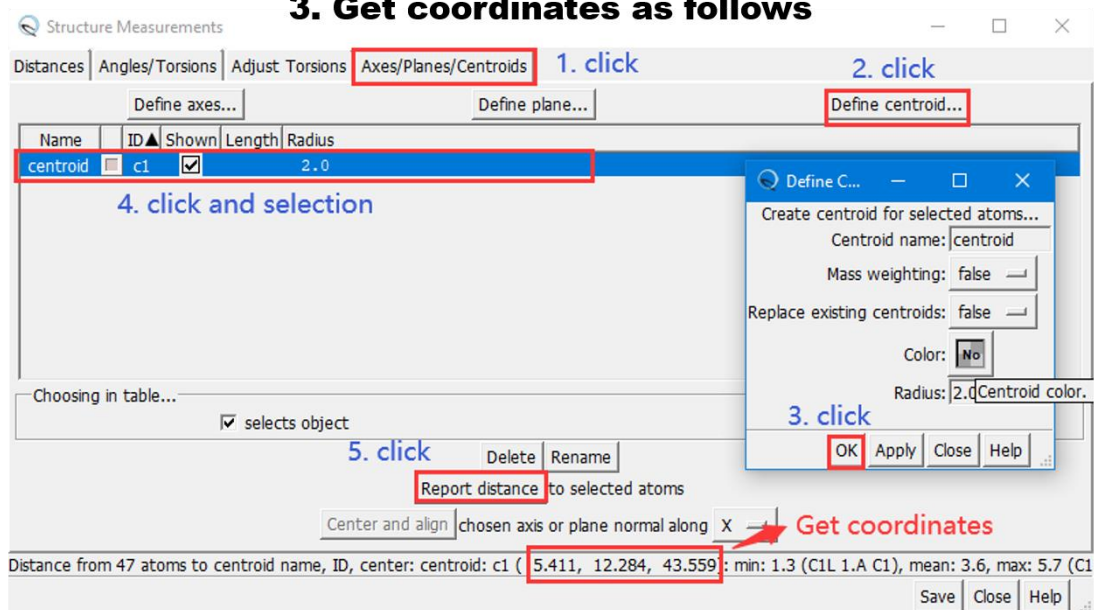


图 2

3) Pafnucy 的原文中是在配体几何中心周围剪裁了一个大小为  $20 \text{ \AA}$  的盒子。通过 MolAICal 获得  $20 \text{ \AA}$  盒子的最大和最小坐标（注意：在写坐标时，一定要加上双引号，X,Y,Z 和盒子的长度数值之间的间隔是一个空格）：

```
#> molaical.exe -tool box -i "5.411 12.284 43.559" -l "20.0 20.0 20.0" -o box.bild
```

将会生成“box.bild”文件，该文件中“.box”部分包含最大和最小坐标。打开该文件，你会发现“.box”部分类似于：-4.589 2.2840000000000007 33.559 15.411 22.284 53.559

4) 用 VMD 导入“3ui7\_A.mol2”文件，打开 Tk 终端（图 3）。

**1. New molecule    2. Browse protein and load it**

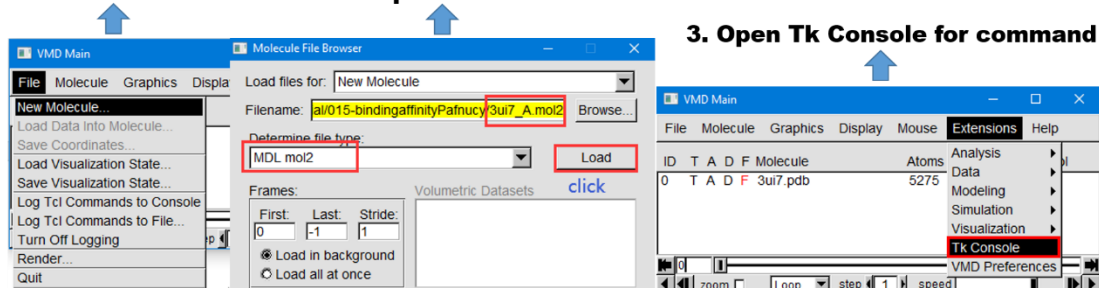


图 3

5) 用 VMD 保存距配体几何中心 20 Å 盒子的复合物。

本教程的工作目录时在 E 盘，因此使用下列命令改变 Tk 的工作目录。

```
#> cd E:
```

```
#> cd workdir/MolAICal/tutorial/tutorial/015-bindingaffinityPafnucy/
```

在 Tk 终端输入下面命令（由于先前的步骤已经计算出了盒子的 x,y,z 的最大和最小坐标，研究人员只需用自己真实数据代替下列坐标数据即可）

```
#> set box [atomselect top "(not ((x < -4.589 or x > 15.411) or (y < 2.2840000000000007 or y > 22.284) or (z < 33.559 or z > 53.559)))"]
```

```
#> $box writemol2 3ui7_pocket.mol2
```

图 4 展示了上述步骤的整个过程，请核对。最后将选取的分子文件保存为“3ui7\_pocket.mol2”。用户应该使用自己实际算出的最大最小 3D 坐标数据。

**(讨论: 某些情况下, 一些用户直接将整个蛋白用于亲和力预测, 如果这样的话, 用户可以跳过上面 5 个步骤。)**

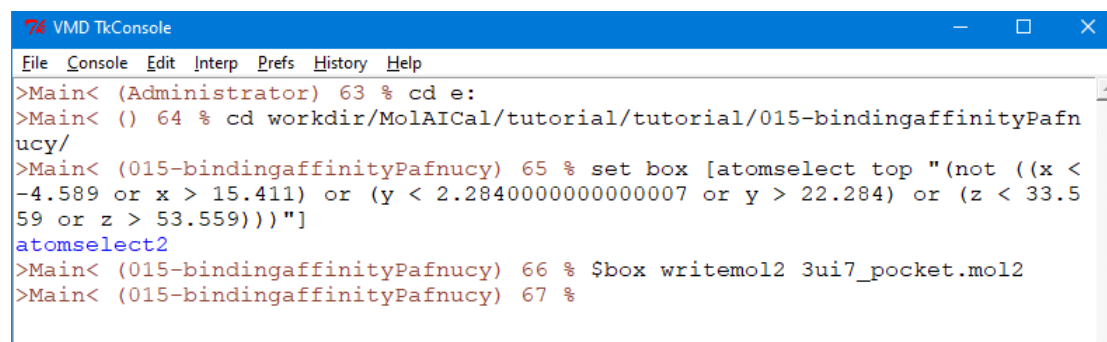


图 4

### 3.2.2 计算单个蛋白和配体之间的亲和力

到 015-bindingaffinityPafnucy 目录下，使用以下命令。

```
#> molaical.exe -model pafnucy -l 3ui7_ligand.mol2 -p 3ui7_pocket.mol2 -o results.csv
```

运行结束后，会生成一个包含 pKx 值的 results.csv 的文件，如果研究人员想要将 pKx 值转化为结合自由能，可以参照该网页(<https://molaical.github.io/tutorial.html>)，MolAICal 提供自由能转化模块。

**注意：**如果研究人员没有 mol2 格式的文件，可以使用 MolAICal 转换文件格式。例如，如果研究人员有 pdb 格式的文件，可以使用下面的命令来进行格式转换（注释：分子应该由正确的后缀，否则无法被 MolAICal 自动识别）。

```
#> molaical.exe -tool format -i ligand.pdb -o ligand.mol2
```

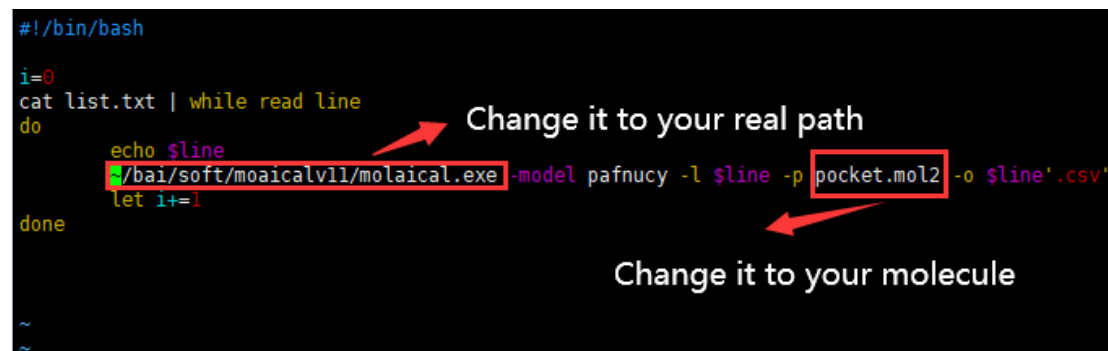
### 3.2.3 计算多个配体和蛋白之间的亲和力

某些情况下，研究人员想要计算多个配体的亲和力，基于以上需求，使用 linux 终端会很容易实现。

进入到 015-bindingaffinityPafnucy 目录，输入以下命令。

```
#> ls lig*.mol2 > list.txt
```

生成一个包含所有配体名字的“list.txt”文件，打开“run.sh”文件，修改 molaical.exe 的正确路径。（如图 5 所示）



```
#!/bin/bash
i=0
cat list.txt | while read line
do
    echo $line
    /bai/soft/maoicalv11/molaical.exe -model pafnucy -l $line -p pocket.mol2 -o $line'.csv'
    let i+=1
done
~
~
```

图 5

最终运行下列命令

```
#> bash run.sh
```

合并所有计算结果

```
#> cat *.csv > results.csv
```

```
#> sed -i -e '/name,prediction/d' results.csv
```

```
#> sed -i '1 i\name,prediction' results.csv
```

运行结束后，会生成一个包含预测值 pKx 的 results.csv 的文件，如果研究人员想要将 pKx 值转化为结合自由能，可以参照该网页(<https://molaical.github.io/tutorial.html>)，MolAICal 提供自由能转化模块。

### 3.2.4 在命令行中计算少数配体和蛋白的亲和力

如果研究人员不需要计算大量的配体，那么他们可以在命令行中使用一行命令来计算少量的配体和蛋白。

进入到 015-bindingaffinityPafnucy/list 目录下，输入下面的指令：

```
#> molaical.exe -model pafnucy -l "lig_1.mol2 lig_2.mol2" -p receptor.mol2 -o twoTest.csv
```

生成了“twoTest.csv”文件，lig\_1.mol2 和 lig\_2.mol2 两个配体文件所计算的亲和力包含在该文件中，请注意命令要有双引号括住这两个配体名“lig\_1.mol2 lig\_2.mol2”。更为详细的指令，请参考 MolAical 的文档。

## 附录：

Linux 系统下“run.sh”文件的内容：

```
#!/bin/bash

i=0
cat list.txt | while read line
do
    echo $line
    ~/bai/soft/maicalv11/molaical.exe -model pafnucy -l $line -p pocket.mol2 -o $line'.csv'
    let i+=1
done
```

## 参考文献：

1. Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P. Development and evaluation of a deep learning model for protein-ligand binding affinity prediction. *Bioinformatics*. 2018;34(21):3666-74.