

使用 MolAICal 进行分子对接

作者: MolAICal (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 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 Twitter: <https://twitter.com/MolAICal>

MolAICal QQ 学术讨论群: 1151656349

1. 简介

SARS-CoV-2 导致 2019 年冠状病毒病(COVID-19)在世界范围内迅速传播。在本教程中,选择在冠状病毒复制中起重要作用的 SARS-CoV-2 主蛋白酶(Mpro)作为示例目标。已经报道了 SARS-CoV-2Mpro 的晶体结构包括 PDB ID: 6LU7, 6Y2F 等^{[1][2]}。在本教程中,基于 MolAICal (<https://doi.org/10.1093/bib/bbaa161>)介绍了蛋白质和配体之间的分子对接。在 3130 复合物的实验结合亲和力的测试中, Autodock Vina 的 Pearson 和 Spearman 相关系数(rp/rs)为 0.5259 和 0.5421, 对于 MolAICal, 在与 Autodock Vina 相同测定条件下, rp/rs 分别为 0.5335 和 0.5489。这表明 MolAICal 比 Autodock Vina 具有较好的“对接”和“排名”能力。

2. 材料

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>

2.2 示例文件

1) All the necessary tutorial files are downloaded from:

<https://gitee.com/molaical/tutorials/tree/master/0000-docking>

3. 步骤

3.1 处理受体和配体

1. 打开与配体结合的 SARS-CoV-2 主蛋白酶文件(PDB ID: 6Y2F): File→Open (见图 1)。

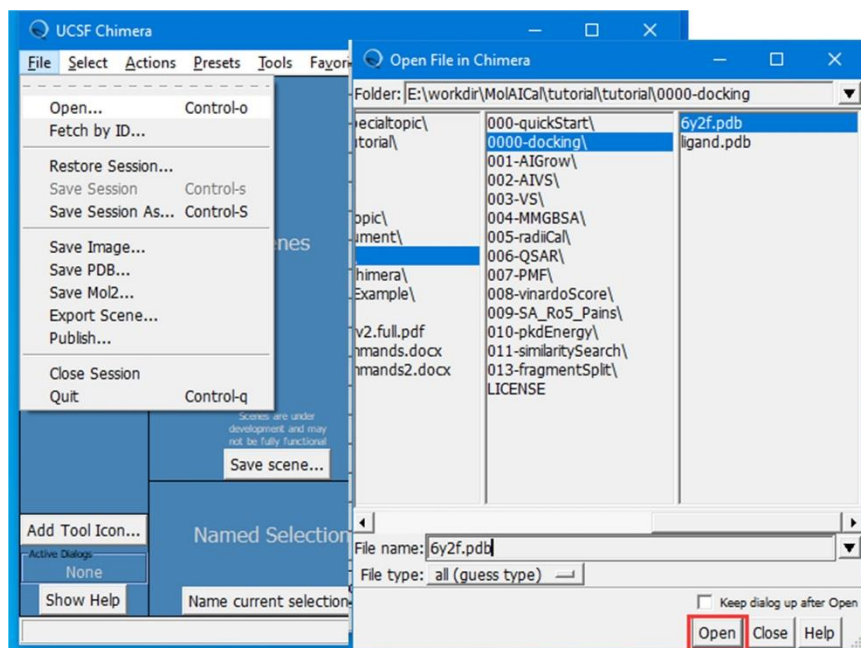


图 1

2. 准备 Mpro 受体并保存名为“protein.pdb”的 Mpro 受体。详细过程如图 2 所示。

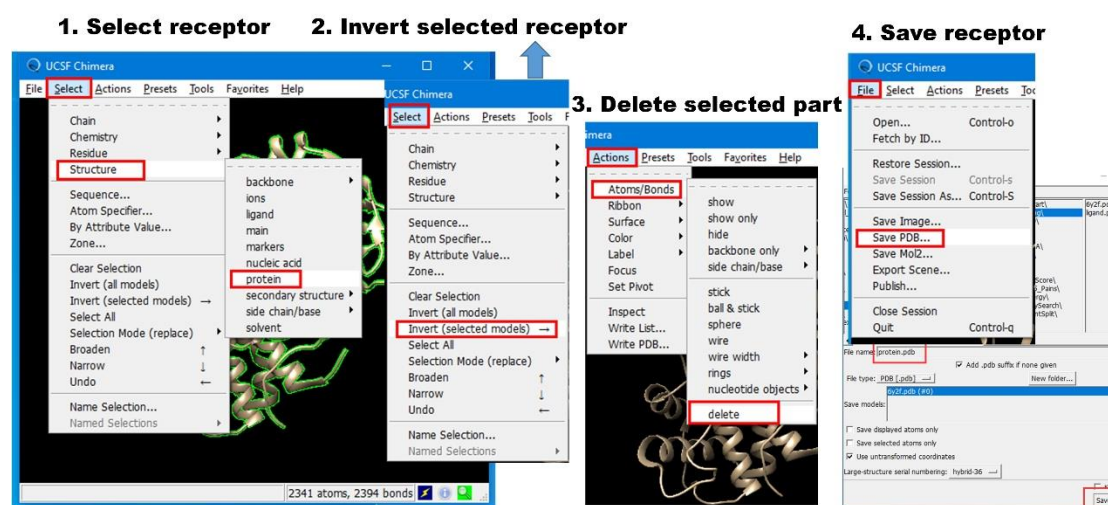


图 2

3. 准备配体并保存名为“ligand.pdb”的配体。详细过程如图 3 所示。

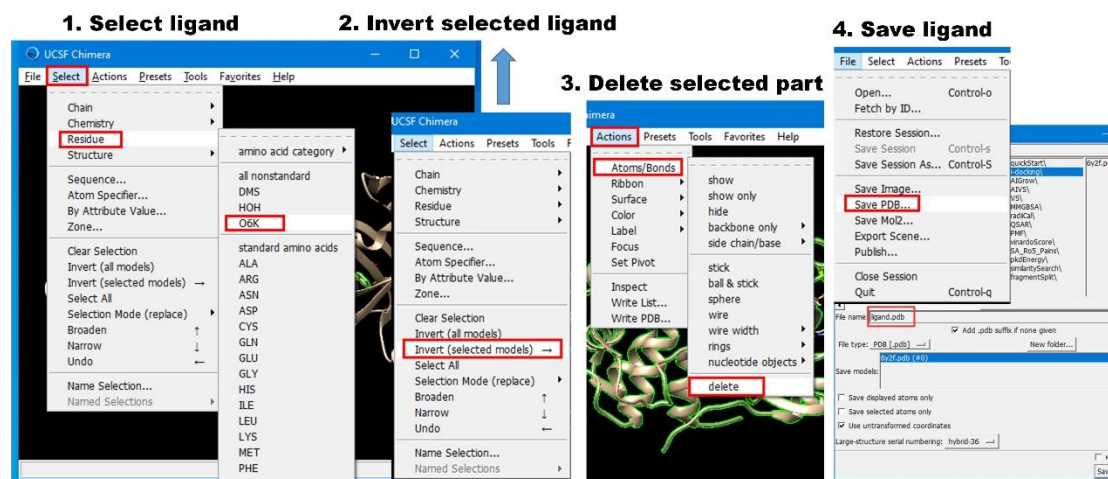


图 3

3.2 将受体和配体转换为 PQBQT 格式

1. 使用 `cd` 命令切换到 `protein.pdb` 所在的文件夹，并使用以下命令获取受体的 PDBQT 格式：

```
#> MolAIcal-xxx\molai cal.exe -dock receptor -i protein.pdb
```

注意：MolAIcal-xxx 是您下载的 MolAIcal 版本的目录。

它将生成名为“`protein.pdbqt`”的文件，该文件具有与“`protein.pdb`”相同的前缀名称。

2. 使用以下命令获取配体的 PDBQT 格式：

```
#> MolAIcal-xxx\molai cal.exe -dock ligand -i ligand.pdb
```

它将生成名为“`ligand.pdbqt`”的文件，该文件具有与“`ligand.pdb`”相同的前缀名称。

3.3. 获取对接盒的中心和长度

1. 依次打开 `protein.pdb` 和 `ligand.pdb`。然后打开 Chimera 的“命令行”：

Favorites→Command Line（见图 4）。

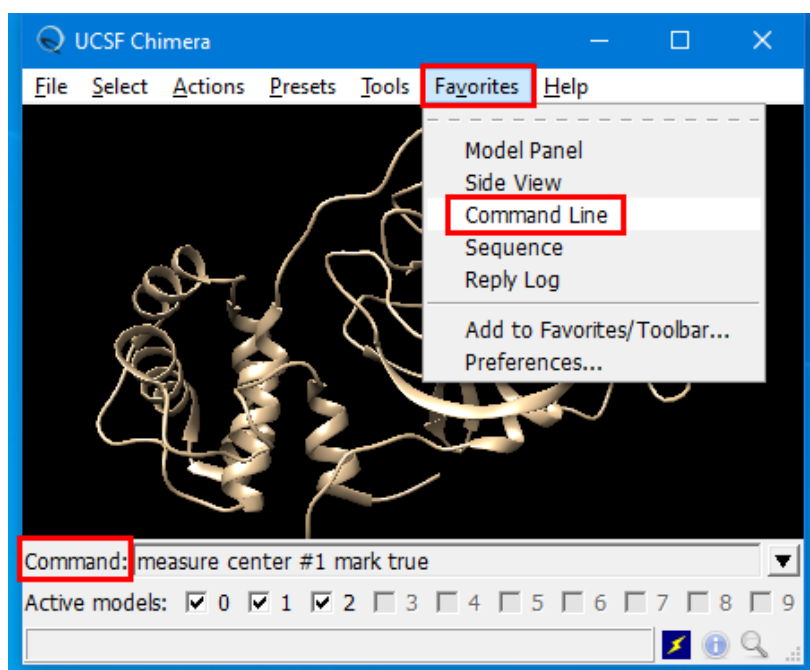


图 4

2. 确定配体的开放序列。如果蛋白质首先打开，它将对应于“Active models 0”。第二个打开对应于“Active models 1”，依此类推（见图 5）。在这里，配体是第二个打开的（“Active models 1”）。将以下命令放入命令行（见图 5）：

`define centroid mass false #1`

并按“Enter”键。然后，依次点击图 5 中的序号。它将显示配体的几何中心坐标 (x, y, z) 为 (10.879, -0.251, 20.754)。

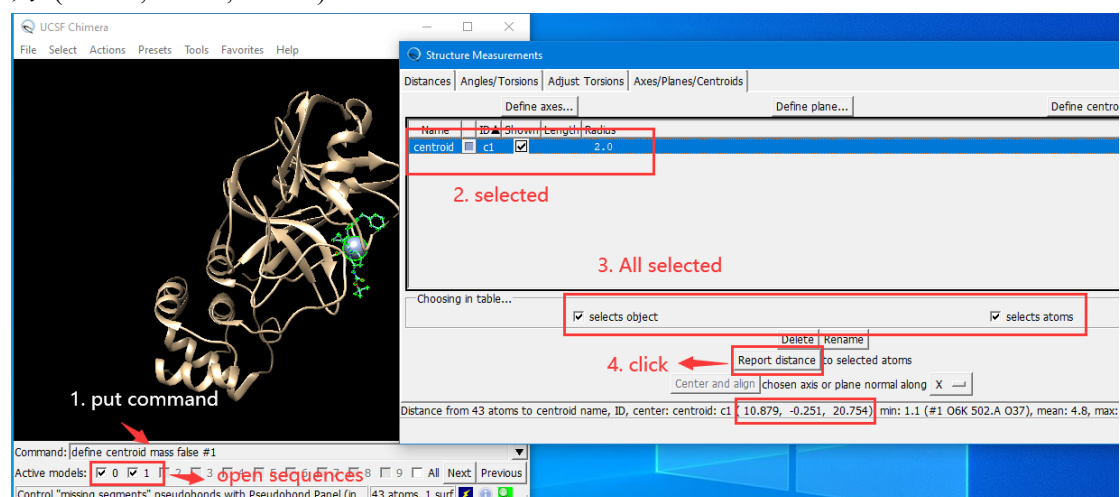


图 5

3. 通过 UCSF Chimera 确定框大小和中心。

打开盒子工具: Tools→Surface/Binding Analysis→Autodock Vina

选择盒子大小: 选择正确的受体（此处命名为“protein.pdb”）和配体（此处命名为“ligand.pdb”）（参见图 6）。将上述中心坐标“10.879, -0.251, 20.754”填入中心框内

(见图 6)，用户可以尝试大小，直到找出合适的尺寸。

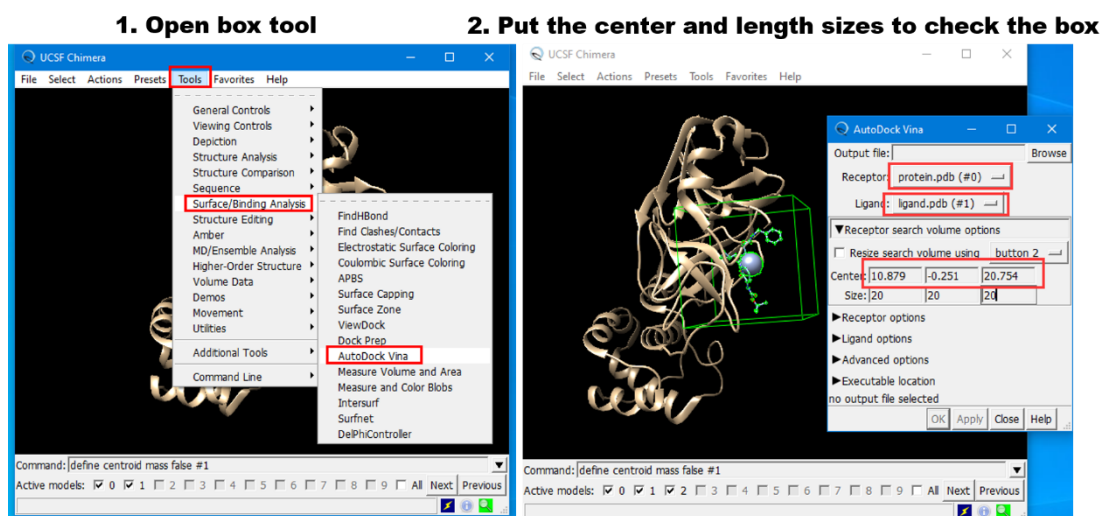


图 6

注意：用户可以勾选“Resize search volume using button 1, 2 or 3”。按钮 1、2 或 3 表示鼠标左键、中键或右键单击。如果用户选择此功能，他们可以通过鼠标调整框大小。如果你对它感兴趣，你可以试试这个功能。

4. 假设配置文件名为“conf.txt”，最终的配置文件可以写成：

```
out = all.pdbqt
cpu = 4
receptor = protein.pdbqt
center_x = 10.879
center_y = -0.251
center_z = 20.754
size_x = 20
size_y = 20
size_z = 20
num_modes = 3
```

其中“out”是输出文件名。“cpu”是使用 CPU 的数量。“receptor”代表受体名称。“num_modes”是生成对接构象的数量。如果“num_modes”为 3，它将生成 3 个配体的对接结构。

3.4 MolAICal 的分子对接

1. 现在 MolAICal 软包中的 MolAICalD 用于受体与配体的分子对接：

```
#> MolAICal-xxx\molaicald --config conf.txt --ligand ligand.pdbqt
```

注意：MolAICal-xxx 是您下载的 MolAICal 版本的目录。

在某些情况下，配体分子有很多旋转键，这样就需要多次试验才可以得到较好的对接结果，可以通过多次运行 MolAICal 找出合适的 random seed 来得到较好的对接结果，这个 random seed 可以进一步用来这个靶点的分子对接或虚拟筛选(见图 7)。

```
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: 555767984
Performing search ...
0% 10 20 30 40 50 60 70 80 90 100%
|----|----|----|----|----|----|----|----|----|----|
*****
done.
Refining results ... done.

mode | affinity | dist from best mode
      | (kcal/mol) | rmsd l.b. | rmsd u.b.
-----+-----+-----+-----
1      -8.53      0.000      0.000
2      -8.52      2.455      8.524
3      -8.47      2.506      6.349
Writing output ... done.
```

图 7

例如，本教程中，用户可以用筛选出的 random seed: 555767984 来重复本教程。请输入下面的命令：

```
#> MolAICal-xxx\molaicald --config conf.txt --ligand ligand.pdbqt --seed 555767984
```

2. 将结果拆分为单个分子

```
#> MolAICal-xxx\molaical.exe -tool pdbqt -i all.pdbqt -o ./
```

单个分子被命名为 1.pdbqt、2.pdbqt 或 3.pdbqt 等。“1.pdbqt”包含结合亲和力最好的对接构象，以此类推。

用户可以直接通过 Pymol 软件查看 1.pdbqt、2.pdbqt 或 3.pdbqt。在这里，UCSF Chimera 用于检查结果。它需要首先通过 MolAICal 使用以下命令将“pdbqt”转化为“pdb”格式：

1) 加氢（选项）

```
#> MolAICal-xxx\molaical.exe -dock addh -i 1.pdbqt
```

2) 将“pdbqt”更改为“pdb”格式

```
#> MolAICal-xxx\molaical.exe -dock pdbqt2pdb -i 1.pdbqt
```

用户可以在本教程中对 2.pdbqt 和 3.pdbqt 使用相同的方式。现在，打开 UCSF Chimera 并加载 protein.pdb、1.pdb、2.pdb 和 3.pdb：

3) 用户可以通过 Favorites→Model Panel 在所有分子加载时选择显示或隐藏分子（见图 8）

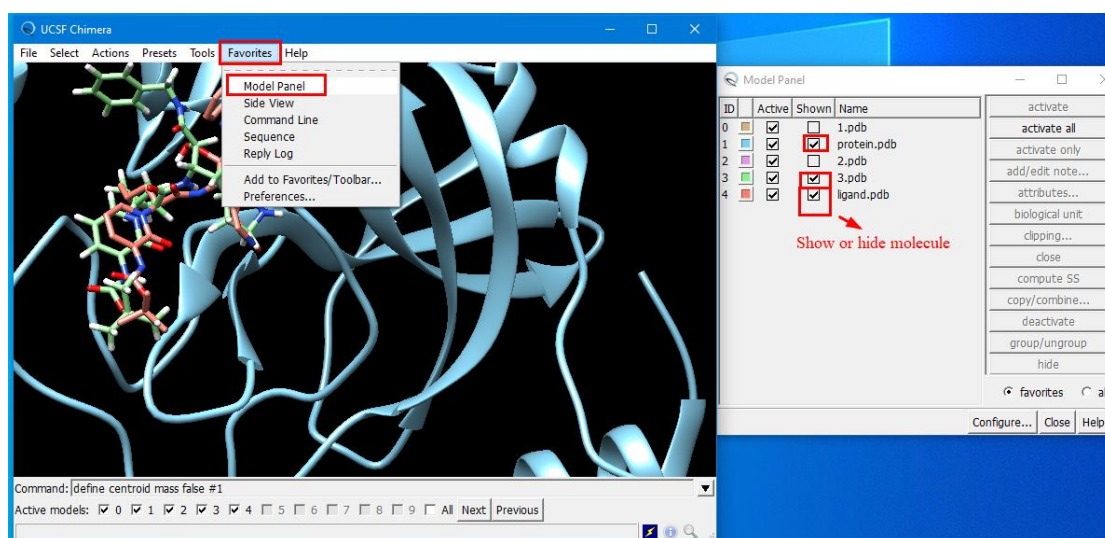


图 8

注意：如果用户想分析蛋白质和对接配体之间的相互作用，用户可以选择在 `protein.pdb` 上添加氢。

结果显示“1.pdb”有部分结构跟原始配体有叠合，而“3.pdb”与原始配体有相似的重叠部分。这个体系做过分子动力学模拟（见教程 MM/GBSA: <https://molaical.gitee.io/tutorial.html>）。MM/GBSA 教程的结果显示原始配体 N3 的 **Andricioaei 熵值是-84.70646297459386 (kcal/mol)**。这表明原始配体 N3 在 SARS-CoV-2 Mpro 的口袋中是不稳定的。

参考文献

- [1] Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, et al. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *bioRxiv*. 2020.
- [2] Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science*. 2020. doi: 10.1126/science.abb3405. PubMed PMID: 32198291