

使用 MolAICal 进行受体和配体批量格式转换和续跑虚拟筛选任务教程

作者：Qifeng Bai (update 2024-03-10)

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>

QQ 讨论群： 210685059

1. 简介

虚拟筛选和评估任务需要对受体和蛋白质形式进行批量处理。此外，一些操作和意外事件可能会导致药物的虚拟筛选任务中断。因此，本教程介绍了使用 MolAICal 批量转换受体和配体形式以及续跑药物虚拟筛选任务的方法。有关 MolAICal 的更多详细信息，请访问 <https://molaical.github.io>。

2. 材料

2.1 软件要求

- 1) MolAICal: <https://molaical.github.io> 或 <https://molaical.gitee.io>
- 2) UCSF Chimera: <https://www.cgl.ucsf.edu/chimera>

注意：确保 MolAICal 和 Chimera 的安装正确！

2.2 示例文件

- 1) 所有必要的教程文件都可以从以下位置下载：

https://gitee.com/molaical/tutorials/tree/master/023-convert_rec_lig_batch

第一部分：受体和配体的批量转化

在这一部分中，MolAICal 提供了一个用于批量转换受体和配体文件的脚本。

1.受体文件的批量转换

1.1 转到虚拟环境

从 Github 或 Gitee 下载教程所需文件“023-convert_rec_lig_batch”，进入目录“023-convert_rec_lig_batch”，命令如下：

```
#> cd 023-convert_rec_lig_batch
```

然后：

在 Windows 系统中：

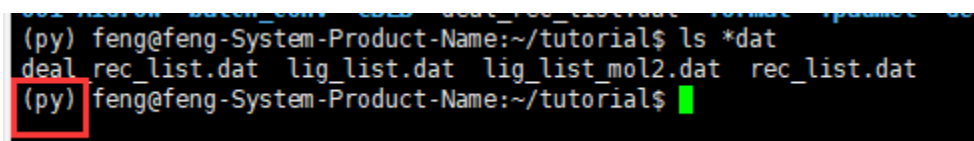
```
#> D:\MolAICal-win64\mtools\py\Scripts\activate.bat
```

在 Linux 系统中：

```
#> source /home/feng/tutorial/MolAICal-linux64/mtools/py/bin/activate
```

注意：请将“D:\MolAICal-win64”或“/home/feng/tutorial/MolAICal-linux64”替换为您的真实 MolAICal 路径。

如果您进入了虚拟环境，您将看到类似于以下的截图：



```
(py) feng@feng-System-Product-Name:~/tutorial$ ls *dat
deal_rec_list.dat  lig_list.dat  lig_list_mol2.dat  rec_list.dat
(py) feng@feng-System-Product-Name:~/tutorial$
```

1.2 获取受体和配体列表

在这里，以下命令将获得配体路径列表：

```
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -sd "batch_conv" -ds -ik "_ligand.sdf" > lig_list.dat
```

在这里，以下命令将获得受体路径列表：

```
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -sd "batch_conv" -ds -ik "_protein.pdb" -nk "_protein.pdbqt" > rec_list.dat
```

注：请将“MolAICal-xxx”替换为您的真实 MolAICal 路径。同时，应该在“-sd”之后设置“batch_conv”的实际路径。

-ds: 是否搜索指定字符的文件，如果只输入“-ds”，则表示搜索指定字符文件为真。

-sd: 用于搜索的目录。

-ik: 这个关键字是用来搜索的。

-nk: 该关键字已从搜索中排除。

有关命令的更详细输入参数，请参阅本文档中的“附录”。

1.3 批量转化受体文件

```
#####
```

选项

如果您认为您的受体文件不需要进一步处理，请跳过此步骤，进入下一步。有时，受体含有不必要的离子和水或错误的氢。因此，在这一部分，MolAICal 将用于批量处理受体文件。

```
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -m -dr -cp "D:\Chimera 1.15rc\bin\chimera.exe" -ir "rec_list.dat"
```

注意： 请将“MolAICal-xxx”或“D:\Chimera 1.15rc\bin\Chimera.exe”替换为您的 MolAICal 或 Chimera 的**真实路径**。“rec_list.dat”应替换为实际路径。

默认处理的受体名称为“rec_deal.pdb”。如果用户想要使用处理后的结果，请使用以下命令获取处理后的受体文件的路径：

```
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -sd "batch_conv" -ds -ik "rec_deal.pdb" > deal_rec_list.dat
```

注意： 如果使用此步骤，请在以下步骤中将“rec_list.dat”替换为“deal_rec_list.dat”。

```
#####
```

如果您省略了上述选项，请按如下方式批量转化受体：

```
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -rc -ir "rec_list.dat" -irf "pdb" -orf "pdbqt"
```

1.4 批量转化配体文件

```
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -lc -il "lig_list.dat" -ilf "sdf" -olf "pdbqt"
```

第二部分：通过 MolAICal 命令批量转化受体和配体

在这一部分中，它调用 MolAICal 命令来批量转换受体和配体的文件。然而，本部分教程对配体和受体文件的格式转换有限。例如，它不能支持配体和受体的 SDF 格式。你可以选择学习第一部分或第二部分。

2.1 使用 MolAICal 命令批量转换受体文件

```
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -m -rc -ir "rec_list.dat" -mp "D:\MolAICal-xxx\molaical.exe"
```

注： 请将“MolAICal-xxx”替换为您的**真实** MolAICal。“rec_list.dat”应替换为实际路径。

2.2 使用 MolAICal 命令批量转换配体文件

在本部分中，MolAICal 命令不支持 SDF 格式，可以选择 mol2 或 PDB 格式进行转换。本部分选取 mol2 格式的配体文件列表作为示例：

```
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -sd "batch_conv" -ds -ik
"_ligand.mol2" > lig_list_mol2.dat
```

然后，批量转化配体文件：

```
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -m -lc -il "lig_list_mol2.dat" -mp
"D:\MolAICal-xxx\molaical.exe"
```

注：请将“MolAICal-xxx”替换为您的真实 MolAICal 路径。“lig_list_mol2.dat”应替换为实际路径。同时，应该在“-sd”之后设置“batch_conv”的实际路径。

第三部分：使用 MolAICal 续跑虚拟筛选任务

请确保在虚拟环境中。具体[请参阅本教程文档第 1.1 部分，学习如何进入虚拟环境。](#)

进入材料目录，如下所示：

```
#> cd 023-convert_rec_lig_batch/vs_continue
```

首先，从文件夹中获得筛选结果，MolAICal 的筛选结果后缀为“_out.pdbqt”。因此，可以使用命令获得筛选结果列表，具体如下：

```
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -sd "F:\023-
convert_rec_lig_batch/vs_continue/vs_results" -ds -ik "_out.pdbqt" > out_list.dat
```

-ds: 是否搜索指定字符的文件，如果只输入“-ds”，则表示搜索指定字符文件为真。

-sd: 用于搜索的目录。

-ik: 这个关键字是用来搜索的。

然后，检查文件“out_list.dat”，如果没有错误，将结果移动到一个文件夹中，如下所示：

```
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -cvs -il F:\023-
convert_rec_lig_batch/vs_continue/out_list.dat -ol F:\023-
convert_rec_lig_batch/vs_continue/target
```

-cvs: 是否准备继续进行虚拟筛选。如果仅输入“-cvs”，则表示为 True，将续跑虚拟筛选任务。

-il: 配体列表的文件。

-ol: 在使用“-cvs”时继续虚拟筛选的部分，它是用于存储的文件夹。否则，它是输出配体的

名称

如果移动完已经虚拟筛选好的分子, 剩余的分子就可以继续按照 **MolAICal** 原来的虚拟筛选方式进行续跑。

注意: 请将“D:\MolAICal-xxx”或“F:\023-convert_rec_lig_batch”替换为用户系统的**真实路径**。

选项: 有时, 一个文件夹中可能有大量的分子。要减少一个文件夹中的分子数量, 可以按如下方式指定一个文件夹的分子数量:

```
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -cvs -dp 1 -il F:\023-convert_rec_lig_batch\vs_continue\out_list.dat -ol F:\023-convert_rec_lig_batch\vs_continue\target
```

-cvs: 是否准备继续进行虚拟筛选。如果仅输入“-cvs”, 则表示为 True, 将续跑虚拟筛选任务。

-il: 配体列表的文件。

-ol: 在使用“-cvs”时继续虚拟筛选的部分, 它是用于存储的文件夹。否则, 它是输出配体的名称

-dp: 在使用“-cvs”时是续跑虚拟筛选的结果, 它是每个文件夹中存储筛选好的分子数量。默认值为“无”, 表示将所有筛选好的分子存储在一个文件夹中。它可以是 1、2 或 3...表示每个文件夹中存储的分子数量。

否则, 它表示搜索目录的深度, 默认值为“无”, 表示搜索所有目录。它可以是 0、1 或 2..., 表示当前、1 或 2...级别的目录。

附录

1. 高级选项 (选项):

如果用户想提高转换速度, 可以在 Linux 中通过“split”命令对输入文件进行分割, 然后

提交分割后的输出文件。split 命令如下所示：

```
#> split -a 3 -d -l 10 --additional-suffix=.txt lig_list.dat test-
```

-a 3: 使用“000”作为命名法，用户可以根据需要进行修改。

-d: 使用数字命名。

-l 2: 每 2 行分割一个文件，用户可以根据需要进行修改。

--additional suffix=.txt: 输出文件的后缀。

lig_list.dat: 输入的文件名。

test-: 输出文件的前缀。

对于 Windows:

用户可以在 Linux 中分割文件，并在 Windows 操作系统中下载输出的文件。

此外，用户可以通过安装 Gnuwin32 或 Git 在 Windows 操作系统上使用“split”命令：

<https://getgnuwin32.sourceforge.net> 或 <https://git-scm.com>

2. 详细参数:

MolAICal 中用于受体和配体文件的批量格式转换的参数如下：

-ir: 受体列表的文件。

-or: 受体的输出名称。

-irf: 输入受体的格式。

-orf: 输出受体的格式。

-il: 配体列表的文件。

-ol: 在使用“-cvs”时继续虚拟筛选的部分，它是用于存储的文件夹。否则，它是输出配体的名称。

-ilf: 输入配体的格式。它的值是分子名的后缀，例如，它可以是 sdf, pdb, mol2, pdbqt, mol 等。

-olf: 输出配体的格式。它的值是分子名的后缀，例如，它可以是 sdf, pdb, mol2, pdbqt, mol 等。

-rc: 是否转换受体格式，如果只输入“-rc”，则表示转换受体格式为真。

-lc: 是否转换配体格式，如果只输入“-lc”，则表示转换配体格式为真。

-lc3: 是否转换配体的 3D 格式，如果只输入“-lc”，则表示转换配体的 3D 格式为真。

-ds: 是否搜索指定字符的文件，如果只输入“-ds”，则表示搜索指定字符文件为真。

-sd: 用于搜索的目录。

-ik: 这个关键字是用来搜索的。

-nk: 该关键字已从搜索中排除。

-fn: 该文件包含不同的搜索关键字。

-dp: 在使用“-cvs”时是续跑虚拟筛选的结果，它是每个文件夹中存储筛选好的分子数量。默认值为“无”，表示将所有筛选好的分子存储在一个文件夹中。它可以是 1、2 或 3...表示每个文件夹中存储的分子数量。

否则，它表示搜索目录的深度，默认值为“无”，表示搜索所有目录。它可以是 0、1 或 2...，表示当前、1 或 2...级别的目录。

- mp:** MolAICal 可执行文件的绝对路径。
- cp:** UCSF Chimera 可执行文件的绝对路径。
- m:** 是否运行 MolAICal 命令进行分子格式转换，如果仅输入“-m”，则表示运行 MolAICal 命令进行分子形式转换为 True。
- dr:** 是否运行 MolAICAl 来处理受体文件。例如，删除离子和水，并修改氢。如果仅输入“-dr”，则表示运行 MolAICAl 处理受体文件为真。
- rn:** 创建的受体名称。
- ah:** 是否删除氢原子，它的值可以是“none”，“add”，“dah” or “del”：
 - “none” 是不做任何操作;
 - “add” 代表加氢;
 - “dah” 代表删掉氢之后再加氢;
 - “del” 代表删掉氢，这是默认是值。
- cvs:** 是否准备继续进行虚拟筛选。如果仅输入“-cvs”，则表示为 True，将续跑虚拟筛选任务。