

使用 MolaIcal 和 MATCH 算法生成小分子 CHARMM 力场拓扑与参数文件

Qifeng Bai

邮箱: molaical@yeah.net

主页: <https://molaical.github.io> 或 <https://molaical.gitlab.io>

1. 引言

用于快速构建小分子 CHARMM 力场拓扑和参数文件的方法 (例如 CGENFF 程序) 通常依赖于模式匹配算法。即使使用力场工具包 (Force Field Toolkit, FFTK) [1], 原子类型分配仍需通过自动匹配或人工检查与指定来完成。MolaIcal 集成了一套升级并改进的 MATCH 算法 [2], 可自动化生成兼容 CHARMM 的小分子拓扑和参数文件。

MATCH 的核心算法 (基于 MIT 许可证) 基于图论和广度优先搜索, 将分子解析为有向图, 并将每个原子的局部环境扩展至最多 10 个化学键, 以捕捉化学上下文信息。通过与预定义的"超 SMILES"字符串库进行最大片段匹配, 实现自动化原子类型分配。部分电荷通过键电荷增量 (Bond Charge Increment, BCI) 规则的迭代反卷积方法推导, 而缺失的参数则通过基于化学相似性评分的替换方案智能生成。

该方法经过交叉验证和留一法测试验证, 电荷拟合的 R^2 值达到 0.94–1.0 (平均绝对误差为 0.0013–0.024 e), 溶剂化能预测的 $R^2 = 0.99$, 键角和二面角参数具有优异的可重复性, 充分证实了其可靠性和准确性。该方法非常适用于高通量虚拟筛选、现有力场的扩展与补缺, 以及跨力场验证, 支持包括 PDB、MOL、SDF 和 MOL2 在内的多种输入格式。它能够快速批量生成兼容 CHARMM 的文件, 以满足大规模初步筛选的需求; 然而, 对于需要更高精度的研究 (例如严格的自由能计算或生产级分子动力学模拟), 建议后续使用力场工具包 (FFTK) 或 FFParam [3] 进行参数优化, 结合量子力学计算和实验数据进行量子力学级别的精细调整。

2. 材料

2.1 软件要求

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

2.2 示例文件

所有必要的教程文件可从以下地址下载:

<https://gitee.com/molaical/tutorials/tree/master/032-gencharmmff>

3. 操作步骤

3.1 生成小分子的 CHARMM 拓扑和参数文件

进入文件夹 `032-gencharmmff`, 运行以下命令:

- 1) 确保文件中的原子名是唯一的。如果原子名相同, 将无法知道哪些原子是相连的, 从而导致 CHARMM 拓扑文件变形或出错。请使用以下命令:

```
#> molaical.exe -call run -c sfile -i u_atom_name.py ligand.pdb ligand.pdb
```

- ◆ 第一个参数是输入分子文件名;
- ◆ 第二个参数是将原子名改为唯一名称后输出的分子文件名, 坐标保持不变。

- 2) 然后为分子生成 CHARMM 力场:

```
#> molaical.exe -call charmm -i ligand.pdb
```

运行后输出的文件为 `.str` 后缀的文件, 其中包含输入分子的 CHARMM 拓扑和参数, 随后可用于构建 `.pdb` 和 `.psf` 文件, 以进行 NAMD 分子动力学 (MD) 模拟。

选项： 如果研究人员希望分别生成拓扑文件（后缀 .rtf）和参数文件（后缀 .prm），而非合并为 .str 后缀的单一文件，可添加参数 -div，并按如下方式运行命令：

```
#> molaical.exe -call charmm -i ligand.mol2 -div
```

注意： 输入分子也可以为 ligand.pdb。此处仅表示输入分子可以采用 PDB 或 MOL2 格式。

深入研究：

如果运行：

将分子中的原子名改为唯一的名称

```
#> molaical.exe -call run -c sfile -i u_atom_name.py final_lig_138.mol2  
final_lig_138.mol2
```

生成拓扑和参数文件

```
#> molaical.exe -call charmm -i final_lig_138.mol2
```

分子 final_lig_138.mol2 是在活性口袋内优化得到的，而非在自由状态下进行能量最小化。因此，某些原子间距离可能发生改变，这可能导致匹配算法在分配电荷或其他参数时失败，从而可能出现以下错误：

```
Could not determine a suitable solution for bond types for UNL1! be wary  
The Following Bonds Could Not be Charged!  
Name      Type      Name      Type      Bond Type  
C          CG2R51   C          CG311     1  
C          CG2D1   N          NG311     1  
C          CG2D1   N          NG311     1  
Could not Finish Charging of Molecule! at E:/test/log/MolAICal-  
win64//mtools/matchvx/MATCH/lib/MATCHer.pm line 174.
```

或

```
Allowed Prontation Not defined for C1 for C in <0> at E:/test/log/MolAICal-  
win64//mtools/matchvx/PerlChemistry/lib/Molecule.pm line 515.
```

解决方案：

用户可使用 UCSF Chimera 对单个分子进行能量最小化，然后将其保存为名为 138.pdb 的 PDB 文件（如下图 1 所示），再重新运行上述命令：

```
#> molaical.exe -call charmm -i 138.pdb
```

即可解决此问题。

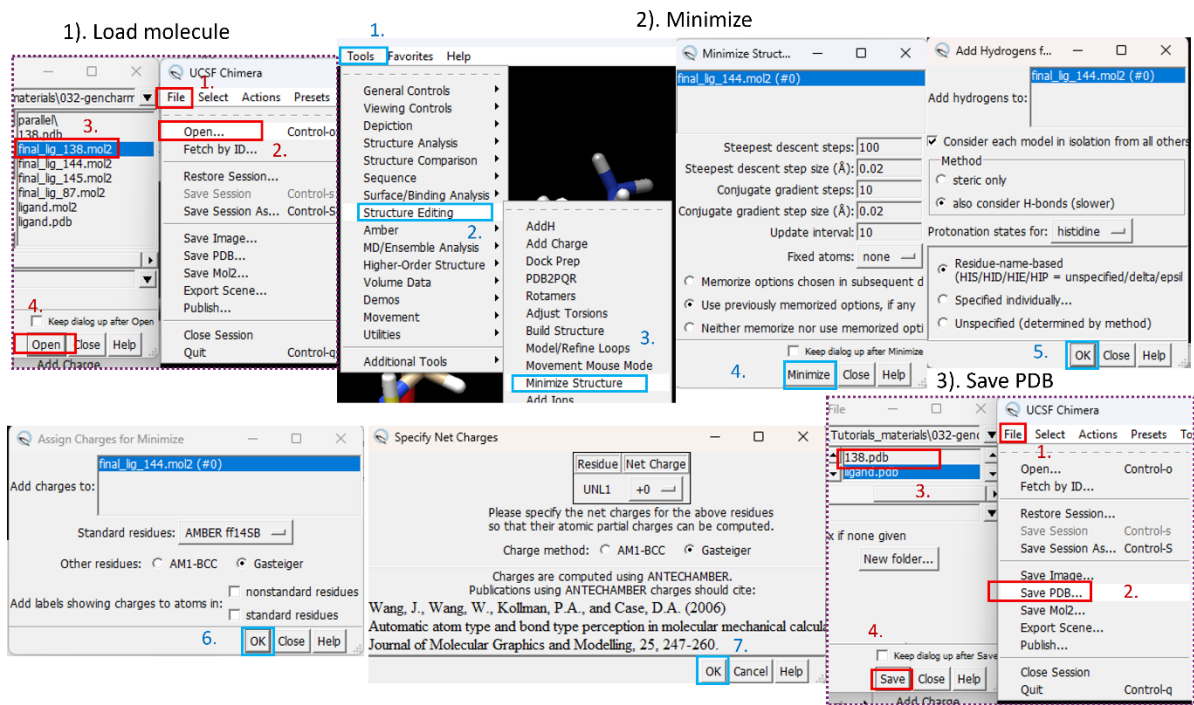


图 1. 分子能量最小化

3.2 使用多核 CPU 并行生成小分子的 CHARMM 拓扑和参数文件

进入文件夹 `032-gencharmff/parallel`，运行以下命令：

在并行模式下将分子中的原子名改为唯一的名称

```
#> molaical.exe -call run -c parallel -nc 3 -n "*.pdb" -i molaical.exe -
call run -c sfile -i u_atom_name.py marg marg
```

在并行模式下生成拓扑和参数文件

```
#> molaical.exe -call run -c parallel -nc 3 -n "*.pdb" -i molaical.exe -
call charmm -i marg
```

注意：

- ◆ `-call`：与外部程序或命令交互。`run` 表示调用外部程序，可从设置的环境文件中搜索程序路径。
- ◆ `-i`：外部程序的命令参数。包含外部程序的所有命令行参数，但 `-i` 参数必须最后指定，而 MolAICal 的其他所有标识符（例如 `-p` 等）必须位于 `-i` 之前。
- ◆ `-n`：指定的关键字（此处为 `/*.pdb`）用于在指定目录中查找包含该关键字的文件名；这些文件名将替换外部命令中的 `marg` 占位符，从而实现文件的批量并行处理。默认匹配规则为 SMART。
- ◆ `-nc <arg>`：调用的 CPU 核心数量。
- ◆ `-c`：若其值为 `parallel`，将以并行模式调用外部命令行。

运行流程：

首先在当前目录（包括所有子目录）中查找所有包含 `.pdb` 字符串的分子文件名；然后将这些文件名逐一替换命令行中的 `margin` 占位符（每条命令一个文件名，替换所有出现的 `margin`）；最后使用多个 CPU 核心（本例中为 `-nc 3` 指定的 3 个核心）进行并行计算。基于此规则，用户可以轻松扩展该方法，以并行方式运行其他外部命令程序。

参考文献

1. Mayne CG, Saam J, Schulten K, Tajkhorshid E, Gumbart JC. Rapid Parameterization of Small Molecules Using the Force Field Toolkit. *Journal of Computational Chemistry*. 2013;34(32):2757-70. doi: 10.1002/jcc.23422. PubMed PMID: WOS:000326466900001.
2. Yesselman JD, Price DJ, Knight JL, Brooks CL. MATCH: An Atom-Typing Toolset for Molecular Mechanics Force Fields. *J Comput Chem*. 2012;33(2):189-202. doi: 10.1002/jcc.21963. PubMed PMID: WOS:000297852600008.
3. Kumar A, Yoluk O, MacKerell A. FFParam: Standalone package for CHARMM additive and Drude polarizable force field parametrization of small molecules. *Journal of Computational Chemistry*. 2020;41(9):958-70. doi: 10.1002/jcc.26138. PubMed PMID: WOS:000504769500001.