

有机共轭小分子使用 Gromacs 拟合单个二面角的方法

0. 绘制分子结构

可以使用 ChemDraw 绘制，将绘制好的分子键线式复制到 Chem3D，并保存成 gjf 格式。用 GaussView 打开这个 gjf，重新保存成同名的 gjf，覆盖之前的文件，防止代表原子优化与否的额外数字 0 干扰 Multiwfn 等软件读取 gjf 的信息。如果体系有明显的点群对称性，建议调整结构让分子满足对应的点群。具体操作是在 GaussView 里往对称的方向调整结构后点 Tools->Point Group，勾上左上角的 Enable Point Group Symmetry，选择一个较大的 Tolerance 直到左边出现对应的点群，点右边的 Symmetrize，然后把 Constrain to subgroup 改成相同的点群，保存结构。

注意本文提到的都是拟合二面角的操作，为了排除其他基团的干扰，这里应该尽量去掉对电子结构影响不大，而基本只有空间位阻影响的取代基，并尽量简化模型，同时也可以降低计算量。之后直接用这个模型拟合的二面角参数，代入实际模拟的体系对应的二面角参数中即可。

1. 几何优化

用 Gaussian 完成，将上面调好的结构使用诸如 B3LYP/6-31G* EmpiricalDispersion=GD3BJ 级别优化，对于溶剂极性不大或者分子的离子性不强的体系，使用真空环境优化即可，否则需要考虑带隐式溶剂模型（建议 PCM 而不是 SMD，Gaussian 应该用 IEFPCM 这种实现而非 CPCM 或者 COSMO）。下同。将优化后的文件存成新的 gjf。

2. 单点计算，获取波函数

可以用 Gaussian 或 ORCA 完成，进行高精度单点计算。注意到如果使用 HF 方法或者双杂化泛函等方法，还需要考虑 HF 密度、自然轨道以及拟合静电势算法耗时增加等麻烦问题，建议只使用非双杂化的 DFT，诸如 B3LYP/def2-TZVP 级别下即可（D3 校正对给定结构的波函数无影响）。对于凝聚相模拟，可以考虑使用 RESP2 电荷（见 <http://sobereva.com/531>），这时候可以做一个气相单点和一个隐式溶剂模型下的单点计算，分别拟合 RESP 电荷，然后进行加权平均。

3. 拟合 RESP 电荷

使用 Multiwfn 载入上一步单点的 fch 文件，使用主功能 7 中主功能 18 进行计算（由于要考虑到具体点群对称性、对称性约束、电荷约束等问题，具体细节操作每次未必完全一样，请仔细阅读屏幕提示和参考下面提到的链接），并导出 .chg 文件。 .chg 文件最后一列即为 RESP 电荷。

建议使用文件包中的 CheckChargeSum.py 脚本检查拟合的 RESP 电荷和下文的 RESP2(0.5)电荷，看按照输出的有效数字，所有原子的电荷总和是否严格为 0。如果不是，可以手动微调小数点后最后一位，直到总和为 0。不过按照输出小数点后 10 位的精度来看，这步没什么必要。

对于拟合二面角的目的，应该用真空中计算 RESP，而实际模拟的时候可以用 RESP2(0.5)。

关于 RESP 和 RESP2(0.5)电荷的计算方法，为什么要用 RESP/RESP2(0.5)电荷，请参考 <http://sobereva.com/531>，<http://sobereva.com/441> 以及 <http://www.whxb.pku.edu.cn/CN/abstract/abstract27818.shtml>，其中尤其卢天老师的 441 号博文值得认真阅读。RESP2(0.5)电荷也可以通过文件包内的脚本 RESP2.sh 获得。

531 号博文提到，早期计算能力有限且没有隐式溶剂模型的时候，用 HF 方法高估的原子电荷抵消溶剂极化造成的原子电荷增加。但是根据较新的测试结果，首先该方法误差抵消程度好坏难以判断，另外对拟合静电势目的 6-31G*还是太小，误差较

大，同时 Amber 手册推荐的方法（HF/6-31G* Population=MK IOp(6/33=2,6/42=6,6/50=1)）里面的 IOp 在 Gaussian 手册中有问题，Gaussian 09 和 Gaussian 16 手册都相互矛盾，手动算出来的拟合点数量也和手册定义的不同，因此不予推荐。

4. 使用 ACPYPE 构建.gro 结构文件，.itp 和.top 力场文件

优化后的结构用 GaussView 修改形式键级，确保符合化学常识，诸如 C 都是 4 价，O 都是 2 价等，保存成.mol2 文件，假设叫 xxx.mol2。也可以用 OpenBabel 来自动判断键级，生成 mol2 文件，只要提供之前几何优化的或者计算 RESP 的 fch 文件，使用 obabel -i fch xxx.fch -o mol2 -O xxx.mol2 即可。

服务器上用 module load acpype/acpype 加载 acpype。执行 acpype.py -i xxx.mol2 -c user -k maxcyc=0。其中 -c user 代表不使用半经验的 AM1-BCC 方法计算原子电荷，而是用自己算的 RESP 等电荷。-k maxcyc=0 代表不用 SQM 的 AM1 半经验方法优化，因为结构已在更高级别的量化方法下优化过。如果有报错，考虑加上 -d 参数使用 debug 模式输出更多信息分析报错原因，通常是因为形式键级不够合理。

默认使用的是 GAFF 力场，也可以考虑 GAFF2 力场。据说 OPLS-AA 力场对凝聚相性质描述更好，但 ACPYPE 生成的 OPLS 力场参数有时候会有问题。新版本 acpype 可能兼容性更强，如果有兴趣欢迎尝试。输出目录里有 xxx_GMX.gro，xxx_GMX.itp 和 xxx_GMX.top，拷贝出来备用。这步可以和 2、3 步同时进行。

卢天老师最近开发了名为 sobtop 的软件，网址为 <http://sobereva.com/soft/Sobtop/>，据说功能很强大，有兴趣可以尝试。

还有钟成老师开发的名为 ztop 的软件，网址为 <http://bbs.keinsci.com/thread-22171-1-1.html>，同样看起来很强大，有兴趣可以尝试。

5. 用 RESP/RESP2 等电荷替换掉.itp 里面的原子电荷

在 itp 里搜索[atoms]字段，将 charge 列的电荷全部替换为.chg 文件的最后一列。可以用高级文本编辑器（Vim，Sublime，UltraEdit 等）的列模式快速替换。

6. 建立单分子的大盒子

建立一个足够大的盒子，只包含一个单分子。单分子两边要留出充足的距离（通常盒子长度在 100 Å 以上），防止分子自身与自身出现相互作用。原理上此时静电和范德华项最好用 cut-off 计算方法，但是如果盒子足够大，这个影响应该足够小。我们这里假设盒子边长是 150 Å，立方体盒子。注意 VMD、Packmol、Gaussian/GaussView（仅限于用户界面，程序内部变量等用的 Bohr）等程序的长度单位都是 Å，而 Gromacs 的单位是 nm。

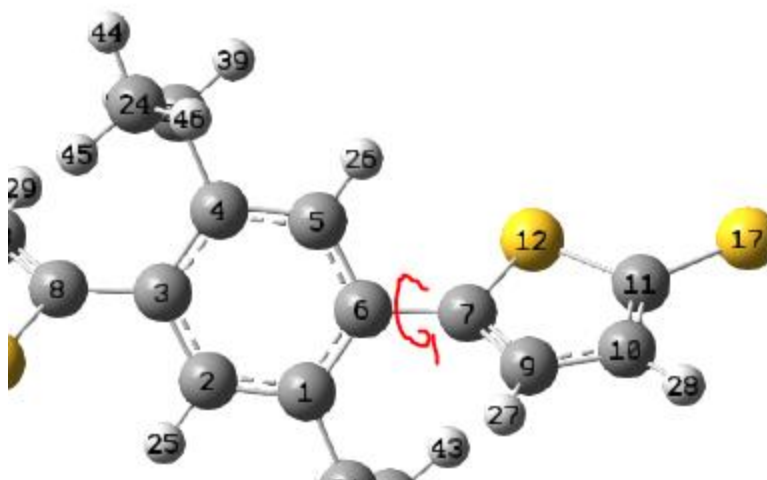
使用 gromacs 系列命令前先要加载集群的相关 module，目前主节点请使用 module load gromacs/2021.3-gcc7.5.0。

使用 gmx editconf -f xxx_GMX.gro -o xxx_box.gro -box 15 来构建盒子。默认分子会被平移到盒子正中心。之后用 VMD 载入 xxx_box.gro，命令行界面输入 pbc box -on -color white 即可显示盒子及分子。

7. 获得单分子 MD 的 tpr 文件，这里面包含了运行一个 MD 任务需要的全部信息

将本文文件包里的 scan.mdp 拷贝到工作目录。运行 gmx grompp -f scan.mdp -c xxx_box.gro -p xxx_GMX.top -o scan.tpr。如果报错 Too many warnings，检查 warnings 无关紧要后加上诸如 -maxwarn 5 参数忽略 warnings，重新运行。得到 scan.tpr。

8. 对要拟合的二面角进行 QM 下的柔性扫描



可以使用优化的级别 B3LYP/6-31G*

EmpiricalDispersion=GD3BJ，加上 Opt=Modredundant 关键词，最后空一行指定扫描的二面角。比如 1 6 7 12 S 36 10.0 代表从初始结构出发，扫描 1-6-7-12 号原子的二面角，每次增加 10° ，扫描 36 步（加上初始结构是 37 步），总共转 360° 。注意最后一个参数必须有小数点。习惯上我们用 cis 取代基之间的二面角。**要注意这个被扫描的二面角必须是 itp 文件中已经包含的项，且保持顺序一致。**

柔性扫描的细节详见 <http://sobereva.com/474> 的 5.1 节。

关于为什么要用柔性扫描而不是刚性扫描，个人认为是当体系中有诸如较大空间位阻排斥的时候，扫描过程中会得到维持高能的结构，而这些结构在实际体系 Boltzmann 分布或者动力学模拟中出现的几率非常低，即使对这些结构拟合很好，也不代表对能代表体系真实情况的那些占大多数分布比例的结构就能描述很好。

9. 提取柔性扫描的每一帧限制性优化后收敛的结构

本文文件包的 GauRelaxedScanSplit.exe 是笔者用 C 语言开发的一个小程序，用于将 Gaussian 柔性扫描的结果转成多帧 xyz。多帧 xyz 格式的介绍详见 <http://sobereva.com/477>。该格式可以被 VMD 等程序支持。

使用该程序载入 Gaussian 柔性扫描的.out/.log 文件，按提示进行，最终会得到柔性扫描结果的 RelaxedScanSplit.xyz 文件。将其改名为 traj.xyz。

如果不准备做第 10 步，用 GaussView 载入 Gaussian 柔性扫描的.out/.log 文件，点 Results->Scan，总能量上右键单击选 Save data，保存扫描各点能量到一个文本文件，文件名叫 scan_tot_ener.txt，然后在 Bash Shell 中使用本文文件包的 extract_energy_from_scan.sh 脚本将其转化为 qm_scan_energy.txt。

10. 对柔性扫描的每一个点做高精度单点计算（推荐）

由于能量比几何结构对计算精度更敏感，同时为了减少 def2-SVP 等基组在平缓势能面上可能存在的描述缺陷，推荐使用高精度单点能计算的电子能量来替代柔性扫描级别获得的电子能量，仅仅使用柔性扫描得到的几何结构，沿着柔性扫描的轨迹计算。具体可参考 <http://sobereva.com/387>。

这里强烈推荐使用 molclus 结合 ORCA，详见 <http://bbs.keinsci.com/thread-577-1-1.html> 和 <http://sobereva.com/490>。具体操作是用 Multiwfn 载入 traj.xyz，输入 oi 回车，选择合适计算级别（推荐双杂化泛函的 RI-PWPB95-D3(BJ)/def2-TZVPP 或更低一些的杂化泛函中算能量表现几乎最好的 RI-wB97M-V/def2-TZVP），保存文件名使用 template.inp。编辑 template.inp，把坐标删掉，坐标的位置只写一行[GEOMETRY]。把服务器的/opt/molclus-1.9.9.9/settings.ini 拷贝过来，iprogram 改成 3，itask 改成 1，建议 ibakout 改成 1。用 /home/scripts/SLURM/molclus/molclus-orca503.slurm 提交作业即可。会得到 isomers.xyz，其中每个结构后面 Energy = 之后就是该结构的能量。可以用本文文件包的 extract_energy.sh 脚本提取出来，保存在 qm_scan_energy.txt。

11. 从 traj.xyz 获得 traj.trr，用于 Gromacs 的 rerun 轨迹

使用本文附件的 xyz2trr_move.tcl 脚本，将第三行的 set bsize 后面的数值改为之前盒子的半径，注意单位是 Angstrom，例如 150。将 traj.xyz 和该脚本放到 VMD 所在目录，启动 VMD，将 xyz2trr_move.tcl 逐行依次复制粘贴到命令行界面。注意不要 vmd -e xyz2trr_move.tcl 或者进入 vmd 后 source xyz2trr_move.tcl，否则无法得到正确结果。

12. 使用 Gromacs，在 QM 柔性扫描的轨迹基础上用 MM 进行刚性扫描

将 scan.tpr 和 traj.trr 拷贝到一个目录，module load gromacs/2021.3-gcc7.5.0。使用 gmx mdrun -pin on -nt 1 -bonded cpu -nb cpu -pme cpu -v -deffnm scan -rerun traj.trr。运行正常的话，会得到 scan.edr 等文件。

13. 从.edr 中提取体系势能

gmx energy -f scan.edr -o mm_potential.svg，输入屏幕提示 Potential 对应的数字（每次可能不一样），回车，再输入 0，回车，生成 mm_potential.svg。这个文件是个纯文本文件，可以直接用 XmGrace 或者 QtGrace 打开，也可以处理后用 Origin 打开。其中#开头的行是注释，@开头的行是 Grace 的绘图参数。注意一定要用势能，不要用总能量。

14. 对比 QM 能量和 MM 能量

我们只关心能量的相对值。笔者在文件包里提供了一个 python 脚本 draw_compare_qm_mm.py，可以用来对比按照上面步骤生成的文件中 QM 和 MM 关于被扫描的二面角势能曲线。如果两者趋势高度一致，就不用做下面的步骤了，否则要重新拟合该二面角。

使用前应修改代码中的 start_value 等变量。同时应该加载包含 numpy, scipy, matplotlib 等库的 python3 环境，比如服务器

上 `module load anaconda/3`。运行结束后，会得到 `energy_diff.txt` 和 `compare_scan_qm_mm.png` 以及 `x_sort_seq.txt`。

之所以要拟合二面角参数而不是其他参数，个人认为主要是因为非键项原子间色散作用可以较好地用 L-J 势等描述，而静电项只要用的原子电荷对静电势重复较好即可较好描述。而成键项中的键长项和键角项相对刚性，即使力常数差一倍，只要平衡结构合理，对结果影响也很小。但二面角项本来就相对柔性，而且传统的 GAFF 力场没有考虑诸如两个相连环的共轭效应等，因此对这类二面角项描述不够合理甚至往往定性错误。更具体一些的内容可以参考钟成老师在 <http://bbs.keinsci.com/thread-24290-1-1.html> 中“背景知识”部分的讨论。

15. 拟合二面角

找到 `itp` 文件中 `[dihedrals]` 字段的 `proper` 项，找到所有对应之前拟合的二面角项，全部注释掉。比如上面图中的示例，1-6-7-12, 5-6-7-12, 1-6-7-9, 5-6-7-9 都要注释掉。然后重复步骤 7 得到关掉对应二面角的 `tpr`，再重复步骤 12-14，得到去掉对应二面角的 MM 能量曲线。

本文附带了 `fit_dih_template.py` 和 `prepare_fit_dih_py.sh` 脚本，Bash Shell 中执行诸如 `./prepare_fit_dih_py.sh 8` 就可以在当前目录下生成含有 8 项力常数拟合的 `fit_dih.py` 脚本，在包含 `energy_diff.txt` 的目录执行它，会得到 `fit_result.txt` 文件，里面记录了拟合的二面角参数。如果发现 `fit_diff.txt` 里面的拟合偏差明显太大，可以增加拟合项数，见 `fit_dih_template.py` 的注释。但是建议不要超过 8 项，尤其是 180° 拟合时，防止产生严重的过拟合。

如果执行 `fit_dih.py` 的目录里还有 `qm_scan_energy.txt`、去掉被拟合二面角项后得到的 `mm_potential.svg` 以及执行 `draw_compare_qm_mm.py` 脚本生成的 `x_sort_seq.txt`，那么该脚本还会自动尝试将拟合出的二面角项加入 MM 能量中，与 QM 结果对比。如果对比结果不满意，可以反复调节 `prepare_fit_dih_py.sh` 脚本使用时后面的命令行参数（即拟合使用

的力常数项），重新依次执行 `prepare_fit_dih_py.sh` 和 `fit_dih.py`，直到结果满意为止。

16. 将拟合好的二面角写入 itp 文件

这里我们只用其中一个二面角，即一开始扫描用的二面角的能量，来代替全部被关掉的二面角的能量。

按照下图格式填入之前拟合的参数。

```
;begin
; i      j      k      l      func      phase      kd      pn      ;      C1-      C6-      C7-      S12
1      6      7      12      9      0.746614      -6.051782      1 ;      C1-      C6-      C7-      S12
1      6      7      12      9      0.746614      -45.084589      2 ;      C1-      C6-      C7-      S12
1      6      7      12      9      0.746614      4.054993      3 ;      C1-      C6-      C7-      S12
1      6      7      12      9      0.746614      0.388132      4 ;      C1-      C6-      C7-      S12
1      6      7      12      9      0.746614      4.464024      5 ;      C1-      C6-      C7-      S12
1      6      7      12      9      0.746614      -0.790505      6 ;      C1-      C6-      C7-      S12
1      6      7      12      9      0.746614      -0.723759      7 ;      C1-      C6-      C7-      S12
1      6      7      12      9      0.746614      -0.774700      8 ;      C1-      C6-      C7-      S12
;end
```

其中 1 6 7 12 是二面角涉及的原子序号。9 是 `func`type，和 1 形式相同但是 1 只支持单项，9 支持多项，一行一项。`phase` 是拟合输出结果的 `phase`，`pn` 从 1 开始递增。

关于力场参数的函数形式和对应的 `function type`，见 Gromacs 2021 版手册 `functions/bonded-interactions` 和 `topologies/topology-file-formats` 部分。

17. 使用拟合好的参数重新对比

重复步骤 7、12、13、14。

如果结果合理，就可以用拟合好二面角参数的 `itp` 文件，继续跑动力学了。