



GROMACS_INDEX

Brief Summary Junior Version



2025-3-7

V 1.0.0

Kuladmir Present

目录

1 分子动力学概述	3
1.1 分子动力学基础.....	3
1.2 其他概念.....	3
2 Gromacs 算法.....	4
2.1 非键作用算法	4
2.2 周期边界条件	4
2.2 组	5
2.4 温度耦合	6
2.5 压力耦合	6
2.6 约束算法.....	6
2.7 能量最小化	7
2.8 溶剂类型.....	7
3 键相互作用.....	8
3.1 成键相互作用	8
3.2 非键相互作用	9
3.3 力场.....	9
3.4 限制和极化	10
3.5 长程作用	10
4 Gromacs 必要文件.....	10
4.1 pdb 文件和 gro 文件.....	10
4.2 力场文件	11
4.2 拓扑文件.....	12
4.4 mdp 文件.....	12
5 Gromacs 运行.....	13
6 Gromacs 分析.....	13
6.1 组创建.....	14
6.2 选择.....	14
6.3 查看轨迹.....	14
6.4 计算能量.....	14
6.5 计算键角、键长、二面角.....	14
6.6 计算 RMSD	15
6.7 氢键分析.....	16

6.8 计算回旋半径	16
6.9 与蛋白质相关的	16
Supporting Information	18

资料仅供个人学习使用，请勿用于他用!!!

详细资料可参考：[GROMACS 中文手册|Jerkwin](#) 和 Gromacs 的 Manual 手册

本文中所举例子（含一些文件），可参考：[Protein-Ligand Complex](#)

1 分子动力学概述

1.1 分子动力学基础

分子动力学(Molecular Dynamics, MD)模拟是更通用的技术方法,可用于产生一个具有代表性的平衡系综。其可适用于研究:小分子、蛋白质、核酸、磷脂膜、多糖、金属、无机材料、界面等。对于非平衡系综的产生和动态事件的分析,只有第二种方法是适用的。与 MD 相比,蒙特卡洛模拟更简单(不需要计算力),然而,在一定的计算时间内,MC 并不能得到比 MD 明显更好的统计结果。

对含 N 个相互作用原子的系统,MD 模拟求解其牛顿运动方程:

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, i = 1, \dots, N$$

原子所受的力是一个势能函数导数的负值,同时,某原子核的受力就是能量对这个坐标的导数的负矢量:

$$\mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i}$$

模拟时,会以很小的时间步长同时求解这些方程,使系统演化一段时间,同时小心地使温度和压力维持在设定值,并以一定的时间间隔将原子坐标写入输出文件中。由于每个原子都有一个坐标,因此随着时间变化,这些坐标也会发生改变。随着时间变化的坐标代表了体系的一个 **轨迹**。经过初始的一些变化之后,体系通常会达到**平衡态**。通过对平衡后的轨迹进行平均,就可以从输出文件中提取出许多宏观性质。

MD 运行中,存在着多种近似,无法对所有内容都做到完全的精确。同时,执行一个完整的 MD 的过程中,也会经历很多必须的操作。

Gromacs 确定了一系列的基本物理量的单位:长度(nm),时间(ps),电荷(e),温度(K),质量(^{12}C 的 $1/12$)。所有的输入文件的数据,都以此作为基准;同时,Gromacs 也确定了导出单位:能量(kJ/mol),力(kJ/(mol*nm)),速度(nm/ps),偶极矩(e),压强(kJ/(mol*nm³))。在 gromacs 里,这些单位都是如此,和其他 MD 软件有所区别。

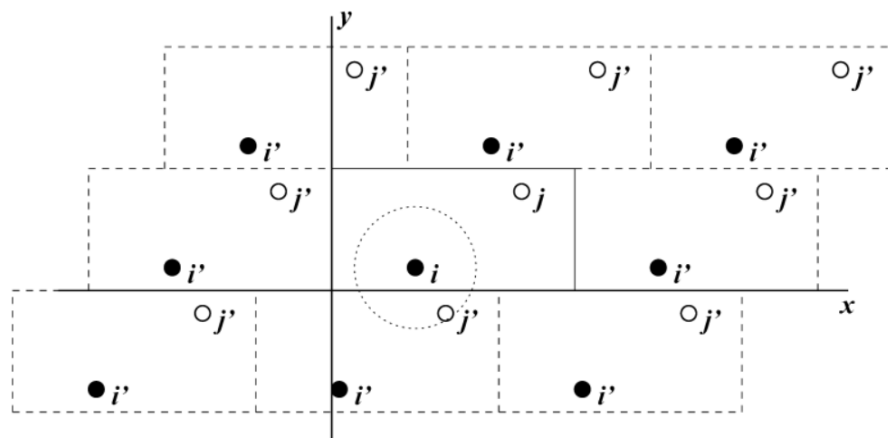
1.2 其他概念

系综(Ensemble)代表一定条件下一个体系的大量可能状态的集合。NVT 为正则系综,NPT 为等温等压系综,NVE 为微正则系综,NPH 为等焓等压系综。

2 Gromacs 算法

2.1 周期边界条件

周期边界条件 (Periodic Boundary Condition, PBC) 是模拟凝聚相体系的关键, 其假定真实宏观体系可由被模拟的盒子通过无线平移延展而产生。因此, 有时只需要模拟少量粒子就可以通过时间平均考察宏观性质。PBC 是对有限的体系, 减小边缘效应的经典方法。将待模拟体系的原子放到一个空间填充(space-filling)的盒子中, 周围是体系自身平移后的副本。



PBC 的盒子原理上无限制, 只要能够经过平移复制能够填满整个空间。用于研究溶液中近似球形的大分子, 菱形十二面体与截角八面体更适合, 这是由于对给定的大分子映像的最小距离, 填充这两种形状盒子需要的溶剂分子更少。同时, 由于菱形十二面体与截角八面体是三斜晶体的特殊情况, 而三斜晶胞又是最一般的空间填充元胞, 组成了所有可能的空间填充形状, 因此 Gromacs 是基于三斜晶胞的。

如果体系具有晶体结构, 这样的边界条件符合预期。如果希望模拟非周期性体系, 如液体或溶液, 周期性就会带来误差。如果对于生物大分子的模拟, 盒子必须恰当设置以避免大分子与其镜像发生相互作用, 因此需要注意: 大分子和边界不小于 1 nm, 至少不小于非键作用 cutoff 的一半; 要考虑大分子在模拟过程中可能出现的旋转。

另外, Gromacs 使用周期性边界条件, 并结合了**最小映像约定**: 对短程非键相互作用, 每个粒子只考虑它的一个最近的映像。然而, 对于长程非键相互作用, 使用**最小映像约定**不够好, Gromacs 会采取别的措施 (后文)。

2.2 非键作用算法

最简单算法是截断。通过设置一个 cutoff 值, 可以避免计算所有粒子间相互作用。每个粒子只会计算与它距离小 cutoff 的粒子的相互作用, 如果 cutoff 越大, 结果会越准, 但是会带来更大的计算量, 一般不小于 0.9 nm。当使用 cutoff

做周期性体系计算时，盒子最短的边必须大于 `cutoff` 的而被，否则会导致与另外粒子及其镜像产生作用。

2.3 组

GROMACS 的 MD 和分析程序可对用户自定义的原子组进行一些操作。组的最大数目为 256，但每个原子最多只能属于六类不同的组。这六类组如下：

温度耦合组

对每个温度耦合组可单独定义温度耦合参数(参考温度，时间常数，自由度数目，参见 3.4.4)。例如，在分子溶液中，相比高分子，溶剂分子(力和积分的误差使其容易产生更多热效应)可以使用更短的时间常数与热浴耦合，又如，可以将表面的温度维持得低于比吸附分子的温度。可以定义许多不同的温度耦合组。也请参看下面质心组。

冻结组

属于冻结组的原子在模拟过程中始终保持静止。在平衡体系的过程中这是很有用的，例如，可避免不当放置的溶剂分子对蛋白质的原子产生不合理的碰撞，尽管对受保护原子添加约束势可以得到同样的效果。如果需要，冻结选项可仅仅用于原子的一个或两个坐标，从而将原子冻结在一个平面或一条直线上。当一个原子被部分冻结时，它所受约束仍然可以使它移动，即便是在冻结方向上。一个被完全冻结的原子不能被它所受的约束移动。可定义许多冻结组。冻结坐标不受压力缩放影响；在某些情况下，这可能会产生无用的结果，尤其是与约束同用时(在这种情况下，你会得到非常大的压力)。因此，我们建议避免将冻结组与约束和压力耦合混合使用。为了平衡体系，可以先在使用冻结进行等体积模拟时，然后再使用位置约束与恒压模拟。

加速组

加速组的每个原子上会被加上一个加速度，这等同于受到一个外力。利用这个特性可驱使体系进入非平衡态，并进行非平衡 MD 模拟以计算输运性质。

能量监测组

在模拟中，所有能量监测组之间的交叉相互作用都会被考虑，且对 Lennard-Jones 项和库仑项的计算是分开进行的。原则上，最多可定义 256 组，但这样将会计算 256×256 项相互作用！最好保守地使用。能量监测组之间的所有非键相互作用都可以被排除在外。在能量监测组中被排除的粒子对不会被放入配对列表，当不需要计算体系中的某些相互作用时，这样做可以显著提高模拟速度。

质心组

GROMACS 可以移除质心(center of mass, COM)的运动，无论是整个体系的质心还是或原子组的质心。后者(移除原子组的质心)是有用的，比如，对于存在

有限的阻止质心运动的摩擦的体系(如气体体系)。对温度耦合与质心运动移除使用相同的组是合理的。

压缩位置输出组

为了进一步减小压缩轨迹文件(.xtc 或.tng)的大小,可以仅仅存储一部分粒子的轨迹。所有标记为压缩组的会被保存,其他则不会。如果没有指定这样的输出组,所有的原子坐标都将被保存到压缩轨迹文件。

2.4 温度耦合

温度耦合,也叫热浴或控温技术。相当于另体系中的粒子处在虚构的恒温环境中。这样可以吸收过剩的动能,也可以补充不足的动能。有多种温度耦合方法:随机碰撞、速度调节、Berendsen 控温技术、Velocity-rescale 控温技术、Nose-Hoover 技术。直接使用的分子动力学模拟对应于 NVE 系综。

Berendsen 温度耦合

Berendsen 算法模仿了与给定温度 T_0 的外部热浴相连,并具有一级动力学特征的弱耦合。该算法可以校正体系温度对 T_0 的偏差:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}$$

这意味着温度偏差指数衰减,其时间常数为 τ ,当这个常数越大,耦合效果越弱、温控越慢、温度波动越大,一般 τ 用 0.1-0.5 ps。对于以平衡为目的模拟可采取很小的时间常数(如 0.01 ps);但对于可靠的平衡模拟,可采用更大的时间常数(如 0.5 ps),在这种情况下,它几乎不会影响动力学的守恒性。

由于不断给所有粒子乘上统一系数 λ 使平均动能符合期望温度,到那时会导致粒子间的速度差异被拉大。

2.5 压力耦合

原子坐标和体系压力有直接关系,压力耦合通过对原子坐标乘上一定的因子,来达到调节压力的目的。压力耦合技术也有多种方法: Berendsen 控压技术、Parrinello-Rahman 控压技术等。

2.6 约束算法

限制: 一般是利用谐振势,用来将某些原子间的距离,分子内某些内坐标,或者某些原子的笛卡尔坐标限制在指定数值上。

冻结: 模拟中令某些原子的笛卡尔坐标一直固定在初值。

约束: 令分子的某些内坐标或某些原子间距离一直固定在特定值。

有一些算法可以实现这一功能: SHAKE、LINCS、RATTLE 等。

退火是指将体系温度升到比较高，维持此温度一段时间，再将其缓慢冷却至低温的过程。这可以帮助体系翻越较高势垒，从而比在常温模拟下更容易达到势能面的全局最小点附近。退火是通过简单地改变当前的参考温度实现的，因此实际的弛豫和耦合性质取决于你使用的热浴类型与耦合强度。由于参考温度会变化，所以体系温度不会瞬时达到参考值。所以需要考虑耦合算法内在的弛豫时间，如果退火参考温度的改变快于温度弛豫，当它们之间相差过大时，模拟可能会崩溃。

2.7 能量最小化

(大)分子体系的势能函数是多维空间中一个超曲面，具有非常复杂的形貌。它有一个最低点，全局极小点和大量的局部极小点。在这些点上，势能函数对所有坐标的导数为零，对所有坐标的二阶导数都是非负的。二阶导数组成的矩阵，即所谓的 Hessian 矩阵，不幸的是，构型空间的维数和局部极小点的数目非常多，我们不可能进行足够的采样来获得全部的信息。特别是，没有一种最小化方法能够保证在实际可承受的时间范围内找到全局极小点。然而，对于给定一个初始构型，则可以找到它的最近极小点。如果初始构型离平衡态非常远，体系的受力可能过大，以致 MD 模拟失败。因此需要先对体系进行稳健的能量最小化。

最速下降法

最速下降法 (steepest decent method, SD)，每一步沿着当前受力最大的方向，进行线搜索。其在远离极小点时效率很高，但在临近区域收敛很慢，容易震荡。

共轭梯度法

共轭梯度法 (Conjugated gradient method, CG)，是 SD 的改良：每步位移矢量除了考虑当前位置的梯度外还考虑前一步的位移矢量。

2.8 溶剂类型

显式溶剂模型：将溶剂分子直接加入到模拟体系中。水作为一个常用溶剂，有很多的模型被开发，根据不同的适用情况，已有了三点水、四点水、五点水模型等。

隐式溶剂模型：是指不把溶剂分子直接加入模拟体系，而是把溶剂作为具有一定介电常数的连续介质，能够减少计算开销、避免了溶剂摩擦效应、并可表现出溶剂的平均效果，适用于常 pH 模拟、REMD 等。并且在隐式溶剂模型下，也不需要周期边界条件。

溶剂化自由能 G_{solv} ：

$$\Delta G_{\text{solv}} = \Delta G_{\text{ele}} + \Delta G_{\text{cav}} + \Delta G_{\text{vdw}} + \Delta G_{\text{other}}$$

其中， ΔG_{ele} 是静电部分，指溶质和溶剂之间的静电相互作用能，其数值为负。后三者是非静电部分， ΔG_{cav} 指构造容纳溶质分子的空洞所需要的能量，数值为正； ΔG_{vdw} 指溶质与溶剂分子间的范德华作用能。

表现溶剂效应最重要的是合理表现出溶质-溶剂之间的静电相互作用。广义 Born 溶剂模型 (Generalized Born model, GB) 对计算量的增加不大, Gromacs 支持。这一模型中, 存在一个重要的点: 有效 Born 半径的确定。可以依据此判断: 当某原子用有效 Born 半径算得的 GB 式中的自身项能量应当与真实分子中此原子与溶剂的静电相互作用能一致。

当质子在溶剂中的包埋程度越大, 其有效的 Born 半径越大, 对溶剂化自由能的贡献就约小。然而在 MD 过程中, 构象会改变, 有效 Born 半径也会改变, 因此需要每一步都重新计算, 通常计算基于 CFA 积分。

然而 GB 方法对于氢键、水桥等相互作用表现较差。

Poisson-Boltzmann(PB)模型比 GB 原理复杂, 且计算量消耗比较大。

3 键相互作用

体系的势能一般由成键项核非键相组成。

3.1 成键相互作用

成键项一般包含以下几项: 键伸缩项、键角弯曲项、二面角扭转项和 improper 项。

键伸缩项中, 最常用的是谐振势。

$$V_b(r_{ij}) = \frac{1}{2} k_{ij}^b (r_{ij} - b_{ij})^2$$

更为准确的是 Morse 势, 其与实际势能曲线更为相符, 能够在平衡位置更大范围准确描述实际势能曲线, 但是计算耗时更大。

键角弯曲项中, 最常用的是谐振势。

二面角扭转项中, 有多种势函数, 并且分为异常二面角和正常二面角。

异常二面角是为了保持平面基团 (例如芳环) 处于平面内, 或者为了防止分子翻转成其镜像分子。其描述方式有: 异常二面角简谐式、异常二面角周期型。

正常二面角的描述方式有: 正常二面角周期性、正常二面角 Ryckaert-Bellmans 势函数。

由于二面角中, 存在键的旋转, 因此势能曲线存在周期性, 因此使用谐振势描述不合理。

improper 项是用来维持共轭平面的平面性的。其与异常二面角有些类似, 尽管其与异常二面角的定义方式相似: 涉及四个原子, 但是 improper 项的目的是为了约束某些特定的几何结构。同时, 异常二面角的定义方式和二面角一致, 但是 improper 项只在一些特殊情况下定义和普通二面角类似。

除此之外, 还存在一些交叉项, 例如键键交叉项、键角交叉项等, 这只有在高精度立场中才会使用。在 CHARMM 等立场中, 存在 Urey-Bradley 项, 用以对键角项两端的原子间加入谐振势来校正。

$$V_a(\theta_{ijk}) = \frac{1}{2}k_{ijk}^{\theta}(\theta_{ijk} - \theta_{ijk}^0)^2 + \frac{1}{2}k_{ijk}^{UB}(r_{ik} - r_{ik}^0)^2$$

3.2 非键相互作用

非键项用来描述成键关系上非近邻的原子之间的相互作用, 主要可分为静电作用和范德华作用。

范德华项中, 通常通过 Lennard-Jones 12-6 函数表现。

$$V_{LJ}(r_{ij}) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6}$$

由于范德华作用会以 $1/r^6$ 的速度衰减, 因此一般计算最多 1.4 nm 以内的范德华作用即可。

Buckingham 势相比 L-J 势更为准确, 但是计算资源消耗更大。当 r 趋向于 0 的时候, 色散项趋向于负无穷, 这会导致原子间出现强烈吸引, 如果初期结构原子距离较近, 则在优化时会坍塌。

$$V_{bh}(r_{ij}) = A_{ij} \exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^6}$$

静电项中, 最常用的是库伦公式。

1-4 相互作用, 一个原子与其第三相邻原子之间的作用力, 由于扭转项的存在, 有时会忽略 1-4 相互作用, 有时也会将其乘以某个因子, 以期削弱, 在这一情况下, 扭转项和 1-4 非键作用能将共同描述二面角的扭转势, 因此 1-4 非键作用和扭转项的参数时密切关联的。

3.3 力场

立场一般有三种分类: (Class I), 其中键伸缩、键弯曲使用谐振势, 无耦合项, 并用 L-J 势描述范德华作用, 用以大体系长时间模拟。代表力场: CHARMM、AMBER、GROMOS 等。(Class II), 成键项包括了三次方甚至更高阶项, 有些还会用 Morse 势描述键伸缩, 并且成键项包含了一种或多种耦合项, 一些高精度立场: MMFF94、MM 系列, 和 CVFF、PCFF 等使用。(Class III), 原子参数考虑化学环境影响, 例如超共轭效应、电子极化等。

还有一种分类方式是，根据对待原子的方式。[全原子力场]，将每个原子视为一个粒子明确考虑。[联合原子力场]，将一些基团视为一个联合原子，联合原子有这些组成的粒子的全部质量和电荷。但是对于极性氢，由于其会形成氢键而需要额外考虑。[粗粒化立场]将原子粗粒化考虑，即将一些原子和分子片段并未一个粗粒化珠子，以减少原子计算量，提高模拟的时间和尺度。

3.4 限制

通过一些特殊的势能函数对体系施加限制，可以避免模拟中出现未知问题。限制是多样的，可以使用位置限制、角限制、二面角限制、距离限制、取向限制等。

位置限制，用于将粒子限制在参考位置，以避免体系关键部分产生剧烈变化。通过将粒子限制在围绕某个区域的壳层内，可以对区域进行细致的模拟。然而，由于缺失粒子引起的适当的相互作用，壳层只是近似的。

平底位置限制，可以将粒子限制在模拟体积的某一部分，使势能的平底区域中限制粒子不受力，但粒子处于平底区域外时会受到一个简谐力，并驱使其向平底区域运动。

角限制，用于限制两对粒子间或一对粒子与 z 轴之间的角度，其所用的函数类似于恰当二面角。二面角限制容易理解。

距离限制，当一对原子间的距离超过某一值后，距离限制会给势能施加一个限制项，通常这是用来施加来自实验的限制。例如 NMR 对体系运动的限制。

3.5 长程作用

长程作用包括长程静电作用、长程范德华相互作用。长程静电作用里，Ewald 方法能够精确计算。由于范德华作用里的色散作用会以 $1/r^6$ 的速度衰减，而静电相互作用则只以 $1/r$ 的速度衰减，因此范德华作用中的 cutoff 设 1.0 nm 足够，但对静电作用力不可。然而，绝大部分都会用 PME (Particle-mesh Ewald) 等方法。这种方法不直接对波矢进行加和，而是使用内插方法将电荷分配到网格上。使得计算速度远大于一般 Ewald 方法。SPME (Smooth PME) 方法作为 PME 的改良，其先使用 3D FFT 算法对格点进行傅里叶变换，在 k 空间中利用对格点的单个加和就可以得到倒易空间的能量项。

对范德华作用也可以使用类似 Ewald 方法，进行精确考虑长程作用，然而由于其衰减速度较快，因此使用截断方式即可。然而当 cutoff 不够大时，在这个 cutoff 距离以外的粒子对体系的总范德华作用能和压力也是存在贡献。对于均相体系来说，这一贡献可以通过额外的修正近似考虑。

4 Gromacs 必要文件

4.1 pdb 文件和 gro 文件

pdb 文件遵循一定的格式。每一部分的功能和保存的信息都是事先规定好的。
详细介绍: <https://blog.sciencenet.cn/blog-548663-895916.html>。

```
404 ATOM      1  N   MET  A   1       5.561 -15.963 -8.933  1.00 16.19      N
405 ATOM      2  CA  AMET  A   1       7.044 -15.926 -8.982  0.50 15.51      C
406 ATOM      3  CA  BMET  A   1       7.052 -15.985 -8.910  0.50 15.98      C
407 ATOM      4  C   MET  A   1       7.565 -17.233 -9.569  1.00 15.36      C
408 ATOM      5  O   MET  A   1       6.935 -18.292 -9.473  1.00 14.94      O
409 ATOM      6  CB  AMET  A   1       7.633 -15.614 -7.579  0.50 15.40      C
410 ATOM      7  CB  BMET  A   1       7.591 -15.959 -7.473  0.50 16.50      C
411 ATOM      8  CG  AMET  A   1       9.121 -15.214 -7.502  0.50 15.40      C
412 ATOM      9  CG  BMET  A   1       7.587 -14.631 -6.772  0.50 17.53      C
413 ATOM     10  SD  AMET  A   1       9.709 -13.669 -8.227  0.50 14.04      S
414 ATOM     11  SD  BMET  A   1       8.595 -13.375 -7.571  0.50 24.67      S
415 ATOM     12  CE  AMET  A   1       8.304 -12.536 -8.057  0.50 10.74      C
416 ATOM     13  CE  BMET  A   1      10.121 -14.299 -7.810  0.50 22.50      C
417 ATOM     14  N   ASN  A   2       8.696 -17.109 -10.221  1.00 13.04      N
```

gro 文件中, 文件格式和 pdb 类似, 每一部分都有特定的意义。

```
1  LYSOZYME in water
2  33506
3
4  IMET      N   1   4.998  3.361  0.808 -0.5597 -0.0949  0.1346
5  IMET      H1  2   4.956  3.267  0.817 -1.1431  0.1750  0.2540
6  IMET      H2  3   5.002  3.407  0.901 -0.7062  0.0921  0.0493
7  IMET      H3  4   4.945  3.429  0.749  0.0618  0.1133 -0.2086
8  IMET      CA  5   5.134  3.355  0.739 -0.5674  0.0339 -0.3369
9  IMET      HA  6   5.169  3.457  0.732 -2.8664  0.9753  0.6392
10 IMET      CB  7   5.227  3.284  0.842  0.5171 -0.1000 -0.2291
11 IMET      HB1 8   5.201  3.178  0.863  2.2724 -1.0732 -2.5787
12 IMET      HB2 9   5.221  3.343  0.936  0.1349 -0.9522  0.2955
13 IMET      CG 10  5.379  3.296  0.813 -0.0679 -0.4234  0.6107
14 IMET      HG1 11  5.401  3.232  0.725  1.1982  0.1612  0.4732
15 IMET      HG2 12  5.437  3.257  0.899 -1.5632  0.1047  1.8923
16 IMET      SD 13  5.440  3.464  0.763 -0.1729 -0.0555  0.2491
17 IMET      CE 14  5.460  3.534  0.924  0.8572  0.1696  0.4986
18 IMET      HE1 15  5.522  3.625  0.920 -2.0301  2.2295 -0.3585
19 IMET      HE2 16  5.359  3.560  0.964  1.2806 -0.4261  2.0641
20 IMET      HE3 17  5.502  3.456  0.992 -1.7313 -1.6279  0.1920
21 IMET      C   18  5.137  3.301  0.596  0.1270  0.3850 -0.9321
22 IMET      O   19  5.122  3.184  0.576 -0.0859  0.1621  0.2274
23 2ASN      N   20  5.185  3.381  0.503  1.0081 -0.2703  0.3890
24 2ASN      HN  21  5.183  3.480  0.508 -0.6157 -0.2561  0.1160
```

坐标

原子速度 (nm/ps)

4.2 力场文件

下载好的力场文件夹内, 保存了多种必须的预设参数表。在 atomtypes.atp 中, 定义了不同类型原子, 和同种原子不同环境下的参数。一般可以从力场文件夹中的 atomtypes.atp 中找到。

```
388 ; from file toppar_c36_jul22/par_all36m_prot.prm
389 C      12.01100 ; carbonyl C, peptide backbone
390 C3     12.01100 ; cyclopropane carbon
391 CA     12.01100 ; aromatic C
392 CAI    12.01100 ; aromatic C next to CPT in trp
393 CC     12.01100 ; carbonyl C, asn, asp, gln, glu, cter, ct2
394 CD     12.01100 ; carbonyl C, pres aspp, glup, ctl
395 CE1    12.01100 ; for alkene; RHC=CR
396 CE2    12.01100 ; for alkene; H2C=CR
397 CP1    12.01100 ; tetrahedral C (proline CA)
398 CP2    12.01100 ; tetrahedral C (proline CB/CG)
399 CP3    12.01100 ; tetrahedral C (proline CD)
400 CPH1   12.01100 ; his CG and CD2 carbons
401 CPH2   12.01100 ; his CE1 carbon
```

在 rtp 文件中, 定义了电荷。在 ffnonbonded.itp 文件里, 定义了非键参数。

```
46 [ atomtypes ]
47
48 ; from file
49 ; type atnum      mass  charge ptype      sigma  epsilon
50
51 ; from file toppar_c36_jul22/metals/CHARMM_METAL/par_all22_prot_metals.inp
52 ; type atnum      mass  charge ptype      sigma  epsilon
53
54 ; from file toppar_c36_jul22/metals/CHARMM_METAL/top_all22_prot_metals.inp
55 ; type atnum      mass  charge ptype      sigma  epsilon
56 HB      1   1.008000  0.000  A  0.235197261589  0.0920480 ; backbone H
57 HE      2   4.002600  0.000  A  0.263706020570  0.0889937 ; helium
58 CAP     6  12.011000  0.000  A  0.355005321205  0.2928800 ; aromatic C for pyrimidines (see toppar_all22_prot_pyridines.str)
59 CN      6  12.011000  0.000  A  0.311914551349  0.8368000 ; C for cyano group (see toppar_all22_prot_pyridines.str)
60 COA     6  12.011000  0.000  A  0.356359487256  0.4602400 ; carbonyl C for pyrimidines (see toppar_all22_prot_pyridines.str)
61 CTIX    6  12.011000  0.000  A  0.405358916754  0.0836800 ; aliphatic sp3 C for CH, new LJ params (see toppar_all22_prot_aliphatic_c27.str)
62 CTXZ    6  12.011000  0.000  A  0.358141284692  0.2343040 ; aliphatic sp3 C for CH2, new LJ params (see toppar_all22_prot_aliphatic_c27.str)
63 CTX3    6  12.011000  0.000  A  0.363486677001  0.3263520 ; aliphatic sp3 C for CH3, new LJ params (see toppar_all22_prot_aliphatic_c27.str)
64 NC      7  14.007000  0.000  A  0.329632525712  2.5104000 ; N for cyano group (see toppar_all22_prot_pyridines.str)
```

而在 `ffbonded.itp` 文件里，定义了成键项的参数，例如键伸缩项[bondtypes]等。

```
46 [ bondtypes ]
47
48 ; bond parameters from
49 ; i j func b0 kb
50
51 ; bond parameters from toppar_c36_jul22/metals/CHARMM_METAL/par_all22_prot_metals.inp
52 ; i j func b0 kb
53 HA CA 1 0.10830000 284512.00 ; ALLOW ARO
54 HA CC 1 0.11000000 265374.38 ; ALLOW POL
55 HA CP2 1 0.11110000 258571.20 ; ALLOW PRO
56 HA CP3 1 0.11110000 258571.20 ; ALLOW PRO
57 HA CS 1 0.11110000 251040.00 ; ALLOW SUL
58 HA CT1 1 0.11110000 258571.20 ; ALLOW ALI
59 HA CT2 1 0.11110000 258571.20 ; ALLOW ALI
60 HA CT3 1 0.11110000 269449.60 ; ALLOW ALI
61 HA CY 1 0.10800000 276144.00 ; ALLOW ARO
62 HB CP1 1 0.10800000 276144.00 ; ALLOW PRO
63 HB CT1 1 0.10800000 276144.00 ; ALLOW PEP
64 HB CT2 1 0.10800000 276144.00 ; ALLOW PEP
65 HB CT3 1 0.10800000 276144.00 ; ALLOW PEP
```

4.3 拓扑文件

拓扑文件记录了原子类型、电荷、成键和非键及参数等信息，一般 `top` 文件是体系的主拓扑文件，`itp` 文件是提供补充的拓扑文件，非必须。

`top` 文件里有不同的分段，每个分段都以 `[]` 作为开始，以 C 语言风格的 `#include "file_name"` 来引用其他文件，以 `;` 开头作为注释。

```
20 ; Include forcefield parameters
21 #include "../charmm36-jul2022.ff/forcefield.itp"
22
23 ; Include ligand parameters
24 #include "jz4.prm"
25
26 [ moleculetype ]
27 ; Name nrexcl
28 Protein_chain_A 3
29
```

在 Gromacs 里，共有三种类型粒子，分别为原子 (A)、壳层 (S)、虚拟相互作用位点 (V/D)。

在 `top` 文件中，可以使用 `#ifdef` 定义不同的参数，之后可以通过在 `top` 文件前使用 `#define` 语句或在 `mdp` 文件设定，可以实现特定功能。

4.4 mdp 文件

`mdp` 文件里包含了运行 MD 的必要的参数。不是所有的 `mdp` 文件都会包含最终运行 MD 的必要参数，一些 `mdp` 文件是在中间过程产生的，只和一些步骤有关系。

但是 `mdp` 文件里有一些共有的字段：

- **integrator**: 决定当前任务的目的，可以是 `steep`(做能量极小化)、`md`(运行分子动力学)等；

- **dt**: 动力学步长 (ps)；

- **nsteps**: 步数上线 (可以是能量最小化，也可以是 MD 运行次数)；

最终动力学模拟时间: $\text{time} = \text{dt} * \text{nsteps}$

- **vdwtype**: 设定范德华作用的计算方式，可选择 `cutoff`、`PME`、`User` (需要在 `table.vvg` 提前定义)；

- **vdw-modifier**: 设定如何修改范德华势，默认为 **potential-shift-verlet**（对范德华势进行 shift 使得 cutoff 处数值为 0），可选 **force-switch**、**potential-switch**，二者分别可以使范德华力/势从 **rvdw-switch** 到 **rvdw** 范围间平滑切换。

- **pbc**: 周期性设定，默认为 **xyz**（在 xyz 三个方向都用），可选择 **xy**、**no**（如此需要设置 cutoff 为 0）。

在能量极小化的 **mdp** 文件中，存在以下参数：

- **emtol**: 能量极小化时，最大受力小于多少时，认为收敛（kJ/mol）
- **emstep**: 最速下降法的最大步长（nm），默认 0.01；
- **nstcgsteep**: 每做多少步共轭梯度极小化时，做一次最速下降法极小化，默认 1000。

如果使用控温和控压，则存在 **t(p)coupl** 字段：**tcoupl** 是选择热浴，**pcoupl** 是选择压浴。**tc-grps** 字段控制了控温组，默认为 **system**，但是也可以对多个组单独控温。

在键约束部分，存在 **constraints** 字段，其控制了约束方式，默认为 **none**，可选 **h-bonds**（约束和 H 相连的键，如果还要约束与之相关的键角，应该使用 **h-angels**），**all-bonds**（约束所有的键）；**constraint-algorithm** 字段，控制了约束算法，默认为 **LINCS**，但其不能约束键角。

5 Gromacs 运行

一些操作步骤，可以查看 **gromacs_windows.md**。

6 Gromacs 分析和命令

gmx help selections 帮助命令。

注：本部分一些命令和参数，可能并不适用于 2020.6 版本，需要更高版本才能使用。（后期可能会出一个高版本的安装）

6.1 组创建

使用 **gmx make_ndx** 命令可以创建一个组，并写入索引文件内。自创组是必要的，在进行一些处理的时候很有用。

命令参数：

[-f file_name] 输入文件，可以为.gro 结构文件等；

[-n file_name.ndx] 设定索引参考文件，必须为.ndx 文件。

[-o output_file.ndx] 输出文件，因为是创建索引，所以必须为.ndx 文件。

在没有自定义组的时候，一般会有一些默认组可以选择使用（部分）：

System: 体系中所有的原子；

Protein: 所有的蛋白质原子

Protein-H: 除氢原子外的所有蛋白质原子

C-alpha: C α 原子

Backbone: 蛋白质骨架原子, N, C α 和 C

MainChain: 蛋白质主链上的原子: N, C α , C 和 O, 包括 C 端的氧原子

就像这样:

```
0 System      : 33506 atoms
1 Protein     : 2614 atoms
2 Protein-H   : 1301 atoms
3 C-alpha     : 163 atoms
4 Backbone    : 489 atoms
5 MainChain   : 651 atoms
6 MainChain+Cb : 803 atoms
7 MainChain+H : 813 atoms
8 SideChain   : 1801 atoms
9 SideChain-H : 650 atoms
```

6.2 选择

可以使用 **gmx select** 进行选择。这个命令和创建索引的功能很类似, 不过该命令可以根据定义, 生成动态选择结果,

命令参数 (部分):

[-f file_name] 输入文件, 一般为.xtc 的轨迹文件。

[-s file_name] 输入文件, 一般为.gro 结构文件或.tpr 拓扑文件。

[-n file_name.ndx] 设定索引参考文件, 必须为.ndx 文件。

-select <> 查找条件, <>内为具体条件。

[-os file_name.xvg] 输出文件, 会将选择到的原子数写入文件。

[-oi file_name.dat] 输出文件, 会将选择的索引输出。

6.3 查看轨迹

可以使用 **gmx view** 来查看轨迹。

命令参数:

-f file_name 输入文件, 一般为.xtc 轨迹文件, 也可以为.gro 结构文件等。

-s file_name 输入文件, 必须为.tpr 拓扑文件。

6.4 计算能量

可以使用 **gmx energy** 来进行计算。

命令参数 (部分):

-f file_name.edr 输入文件, 必须为.edr 能量文件。

[-s file_name.tpr] 输入文件, 必须为.tpr 拓扑文件。

-o file_name.xvg 输出文件, 必须为.xvg 文件。

执行完该命令后, 会有能量体系的选择, 之后结束输入后, 会计算 RMSD、Average 等, 并输出。

6.5 计算键角、键长、二面角

计算键长，可以使用 **gmx distance** 来计算。

命令参数 (部分):

[-s file_name] 输入文件，一般为.tpr 文件或.gro 文件等结构文件。

[-f file_name] 输入文件，一般为.xtc 轨迹文件，或者.gro 文件。

[-select <>] 条件选择，条件里定义要计算哪些原子。

[-oall [file_name.xvg]] 设定了此为：距离随时间的变化。

[-oav [file_name.xvg]] 设定了此为：平均距离随时间的变化。

[-n file_name.ndx] 输入文件，必须为.ndx 索引文件。

计算键角，可以使用 **gmx angle** 来计算。

命令参数 (部分):

[-n file_name.ndx] 输入文件，必须为.ndx 索引文件。

[-f file_name] 输入文件，一般为.xtc 轨迹文件，或者.gro 文件。

[-ov file_name.xvg] 输出文件，必须为.xvg 文件，会输出键角平均值随时间的变化。

[-od file_name.xvg] 输出文件，必须为.xvg 文件，会输出二面角或键角的分布情况。

在使用 **gmx angle** 之前，需要先创建一个索引，其中包括了要计算几个原子，否则在使用 **gmx angle** 后，会出现索引组的选择，将可能导致无法计算。

gmx gangle 命令也可以计算键角和二面角，同时也可以计算两个矢量、矢量和平面等的角度。

命令参数 (部分):

[-f file_name] 输入文件，一般为.xtc 轨迹文件，或者.gro 文件。

[-s file_name.tpr] 输入文件，一般为.tpr 拓扑文件，也可以是.gro 文件。

[-oall [file_name.xvg]] 设定了此为：所有角度随时间的变化。

[-oav [file_name.xvg]] 设定了此为：平均角度随时间的变化。

6.6 计算 RMSD

RMSD，一个分子中的某些原子相对于参考结构的根均方偏差。可以使用 **gmx rms** 计算。

命令参数 (部分):

-f file_name 输入文件，一般为.xtc 文件，可以是.gro 文件。

-s file_name 输入文件，一般为.tpr 文件，可以是.gro 文件。

[-n file_name.ndx] 输入文件，必须为.ndx 索引文件。

-o file_name.xvg 输出文件，必须为.xvg 文件。

[-tu (?)] 设定参数，定义了时间单位，(?)可以为 ps, fs, ns, us 等。

其中，-s 参数引入的输入文件，是参考结构，-f 参数引入的轨迹，会和参考结构进行比较。

6.7 氢键分析

可以使用 **gmx hbond** 命令，进行氢键分析。以下参数需要在更高版本使用：

[**-f file_name**] 输入文件，一般为.xtc 文件，可以是.gro 文件。

[**-s file_name**] 输入文件，一般为.tpr 文件，可以是.gro 文件。

[**-n file_name.ndx**] 输入文件，必须为.ndx 索引文件。

[**-o file_name.ndx**] 输出文件，必须为.ndx 索引文件。

[**-tu (?)**] 设定参数，定义了时间单位，(?)可以为 ps, fs, ns, us 等。

[**-cutoff N**] 定义了计算供体和受体的最小距离范围，一般建议为 0.35 nm。

[**-hbr N**] 定义了氢键截止距离，一般建议为 0.35 nm。

[**-hba N**] 定义了 A-D-H 之间的截止键角度数为 30°。

在 2020.6 版本中，gmx hbond 命令具有仅有以下一些参数：

[**-f file_name**] 输入文件，一般为.xtc 文件，可以是.gro 文件。

[**-s file_name**] 输入文件，一般为.tpr 文件，可以是.gro 文件。

[**-n file_name.ndx**] 输入文件，必须为.ndx 索引文件。

[**-num file_name.xvg**] 输出文件，输出氢键数量随时间变化。

[**-dist file_name.xvg**] 输出文件，输出氢键距离分布。

[**-ang file_name.xvg**] 输出文件，输出氢键角度分布。

6.8 计算回旋半径

可以使用 **gmx gyrate** 计算。回转半径可以用以表征蛋白质结构的紧密程度，一般回转半径越小，说明蛋白的致密性越好，蛋白结构就越稳定。

命令参数（部分）：

[**-f file_name**] 输入文件，一般为.xtc 文件，可以是.gro 文件。

[**-s file_name**] 输入文件，一般为.tpr 文件，可以是.gro 文件。

[**-n file_name.ndx**] 输入文件，必须为.ndx 索引文件。

[**-o file_name.xvg**] 输出文件，可以输出回转图。

6.9 与蛋白质相关的

gmx do_dssp 命令，可以分析蛋白二级结构元素的变化。其具有的功能：识别蛋白中的二级结构、可视化与统计、溶剂可及表面积计算。

命令参数：

[**-f file_name.xtc**] 输入文件，必须为.xtc 轨迹文件。

[**-s file_name**] 输入文件，可以为.tpr 文件，也可以为.gro 文件。

[**-n file_name.ndx**] 输入文件，必须为.ndx 索引文件。

-o file_name.xpm 输出文件，其中包含了二级结构信息。

-sc file_name.xvg 输出文件，其中包含了二级结构残基的统计信息。

-ssdump file_name.dat 输出文件，其中包含了二级结构信息，可做后续分析。

gmx rama 命令，可以计算蛋白质 φ (phi) 和 ψ (psi) 二面角，并生成 Ramachandran 图。拉氏图 (Ramachandran plot) 是一种使蛋白质结构中，主链氨基酸残基的二面角 ψ 和 φ 可视化的方法。Phi 角是指氨基酸里氨基和 α -C 间的二面角，psi 是氨基酸 α -C 和羧酸间的二面角。肽键是 ω 。尽管如此，但也不是任意 ψ 和 φ 值都是可以的，由于位阻的原因。

在拉氏图里，存在多个区域，称为允许区，扩展一点后，成为最大允许区。一般稳定的蛋白构象里，绝大多数残基的 phi 和 psi 度数，都落在这些区域内。落在 α 区或 β 区时，则说明其处于 α 螺旋， β 折叠区域。有一些则落于 loop 区域。

命令参数 (部分):

-f file_name.xtc 输入文件，必须为.xtc 轨迹文件。

-s file_name.tpr 输入文件，必须为.tpr 文件。

-o file_name.xvg 输出文件，可以绘制 Ramachandran 图。

gmx wheel 命令，可以绘制螺旋轮图。用以研究螺旋结构的亲疏水性。

命令参数 (部分):

-f file_name.dat 输入文件，必须为.dat 文件。

-o file_name.eps 输出文件，生成封装的 PostScript 文件，用以展示螺旋轮图。

-r0 N 定义第一个残基，N 必须为整数，默认为 1。

-rot0 N 定义了初始旋转角度，默认为 90° 。

-T string 定义了标题，但长度必须小于 10 字符。

-nn Yes/No 定义了是否显示编号，默认为 yes。

Supporting Information

: 对于 gromacs_windows.md 的.xvg 图像的展示。

能量极小化部分

执行命令: gmx energy -f em.edr -o em_test.xvg

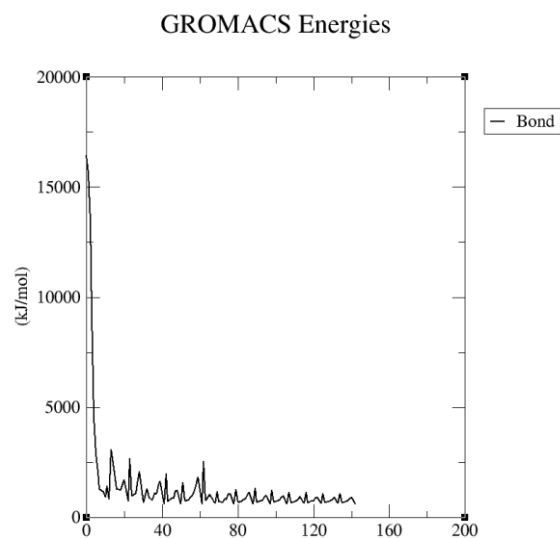


图 S1. Gromacs_mini_1

Energy	Average	Err.Est.	RMSD	Tot-Drift
Bond	1518.13	540	2456.88	-3438
(kJ/mol)				

温度耦合部分

执行命令: gmx energy -f nvt.edr -o nvt_test.xvg

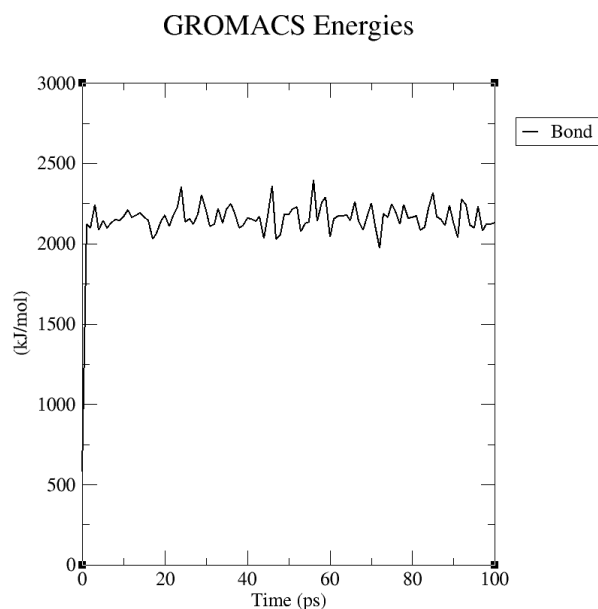


图 S2. Gromacs_nvt_1

Energy	Average	Err.Est.	RMSD	Tot-Drift
Bond	2155.29	12	108.272	42.4008
(kJ/mol)				

压力耦合部分

执行命令: `gmx energy -f npt.edr -o npt_test.xvg`

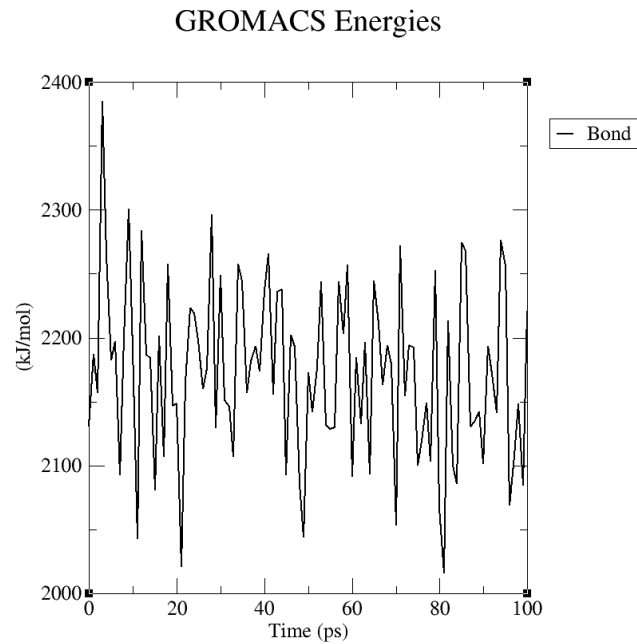


图 S3. Gromacs_npt_1

Energy	Average	Err.Est.	RMSD	Tot-Drift
Bond	2178.85	4.6	77.1134	-10.4802
(kJ/mol)				

MD 运行部分

命令: `gmx energy -f md_0_10.edr -o md_0_10_test.xvg`

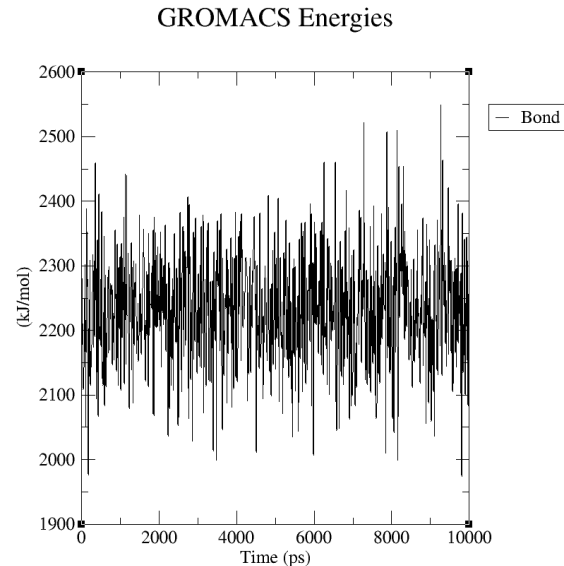


图 S4. Gromacs_md_1

Energy	Average	Err.Est.	RMSD	Tot-Drift
Bond	2227.16	0.4	80.2049	1.93877
(kJ/mol)				

RMSD 计算部分

命令: `gmx rms -s em.tpr -f md_0_10_center.xtc -n index.ndx -tu ns -o rmsd_jz4.xvg`

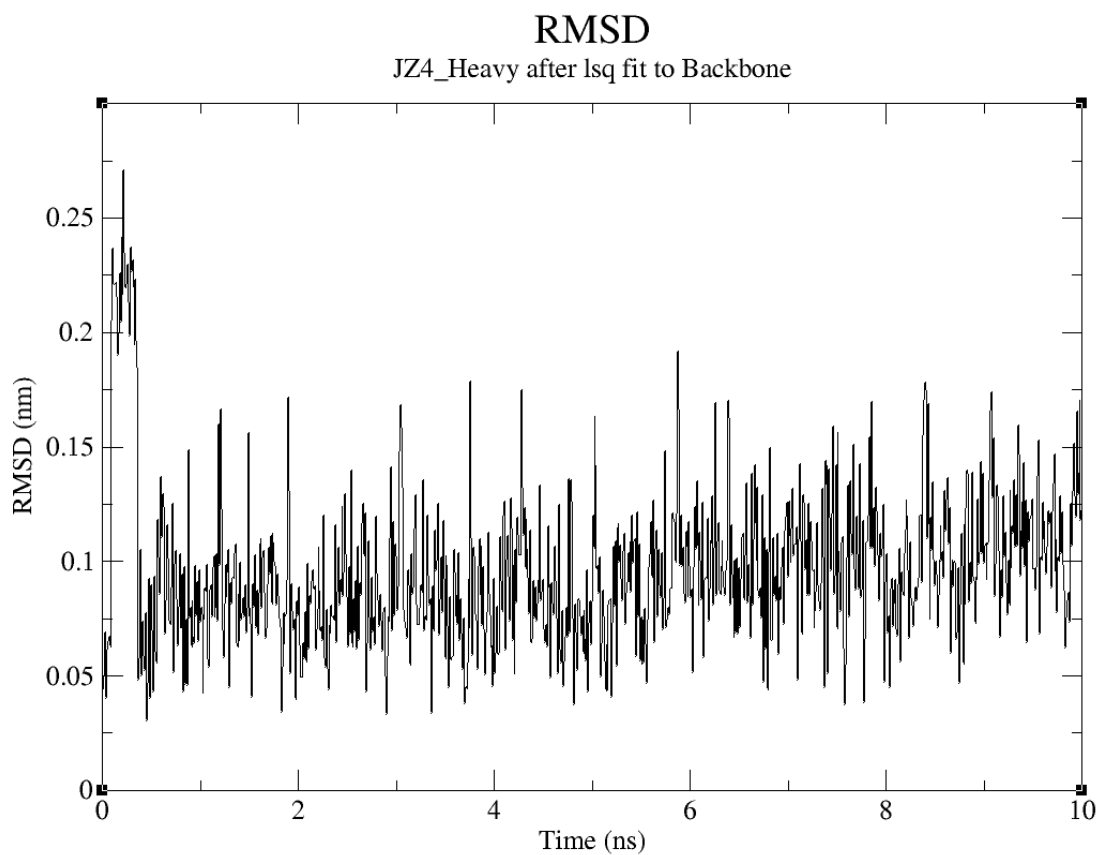


图 S5. Gromacs_rmsd_1

gmx gyrate 命令测试

命令: `gmx gyrate -f md_0_10.xtc -s md_0_10.tpr -n index.ndx -o md_test.xvg`

Radius of gyration (total and around axes)

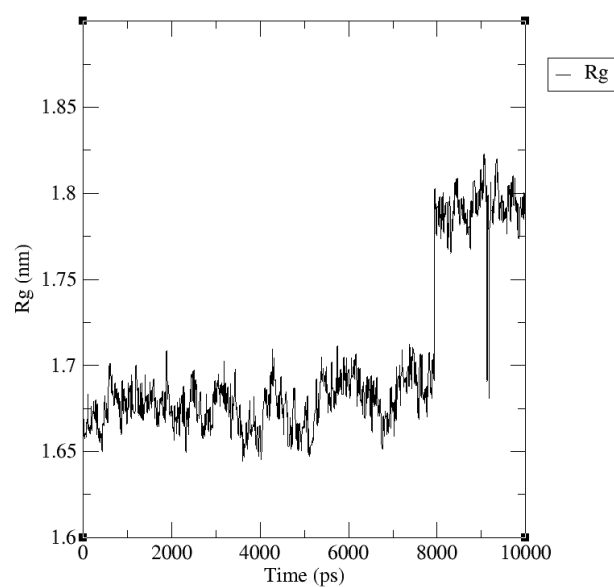


图 S6. Gromacs_gyrate_1

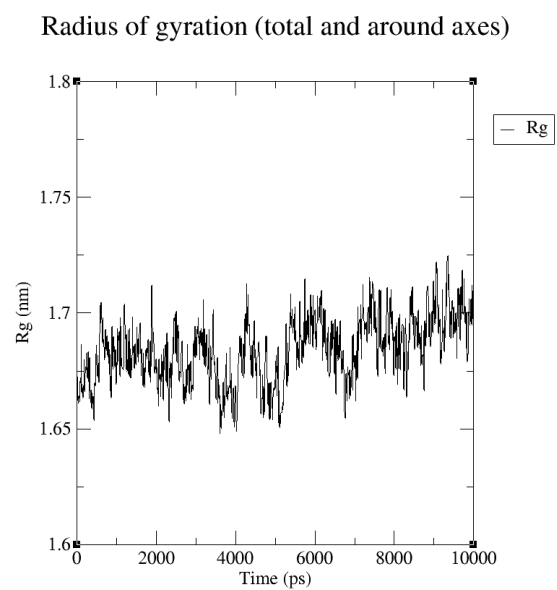


图 S7. Gromacs_gyrate_2

gmx rama 命令测试

命令: `gmx rama -f md_0_10.xtc -s md_0_10.tpr -o rama_test.xvg`

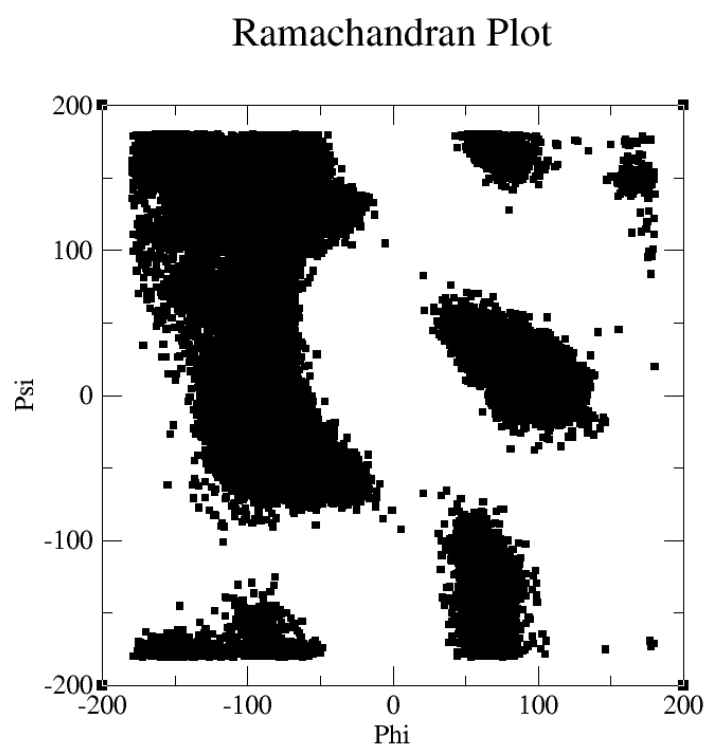


图 S. Gromacs_rama_1