合肥耀淼生物科技有限公司
HEFEI YAOMIAO BIOTECHNOLOGY CO., LTD

MD 分子动力学模拟项目报告

订单号：_____

客户：_____

日期：_____

MD 分子动力学模拟项目报告

一、实验目的：

本次实验旨在通过 Gromacs 分子动力学模拟的方法，探究 A 蛋白质、B 蛋白分别与小分子化合物 B 的结合情况，从而评估蛋白质与小分子之间的相互作用。

二、实验方法：

1. 本次实验所使用的体系为一种蛋白质与小分子复合物，包含了一个蛋白质分子和一个小分子。我们使用 Gromacs 软件进行模拟，具体步骤如下：

2. 分子准备：我们将蛋白质和小分子的结构文件分别导入 Gromacs，通过 pdb2gmx 和 gmx editconf 命令生成拓扑文件和模拟盒子。

3. 能量最小化：我们对蛋白质和小分子的结构进行了最小化处理，使用 gmx grompp 和 gmx mdrun 命令进行模拟。

4. 分子动力学模拟：我们使用 gmx grompp 和 gmx mdrun 命令对蛋白质和小分子进行了长达 100 纳秒的分子动力学模拟，记录了模拟期间的构象信息。

5. 数据分析：我们使用 gmx rms 命令计算了蛋白质与小分子之间的 RMSD、RMSF、Gyrate、SASA 和 HBond 等指标，并通过画图和统计表格的方式对结果进行了展示和分析。

三、实验步骤及参数：

本次分子动力学模拟采用 Gromacs2022.3 版本软件进行分子动力学模拟。小分子预处理使用 **AmberTools22** 对小分子添加 **GAFF 力场**同时使用 **Gaussian16W** 对小分子进行加氢和计算 RESP 电势操作，电势数据会添加到分子动力学体系拓扑文件中。模拟条件在静态温度 300K 和常压（1Bar）下进行，力场采用 **Amber99sb-ildn**，溶剂为水分子（Tip3p 水模型）并通过添加适当数量的 Na+ 离子来中和模拟系统的总电荷。分子动力学模拟体系先采用最陡下降法进行能量最小化，此后分别进行 100000 步的等温等容系综（NVT）平衡 和等温等压系综（NPT）平衡，耦合常数为 0.1ps、持续时间为 100ps。最后运行自由分子动力学模拟，该过程共 5000000 步，步长为 2fs，共计时长 100ns。在完成模拟后，

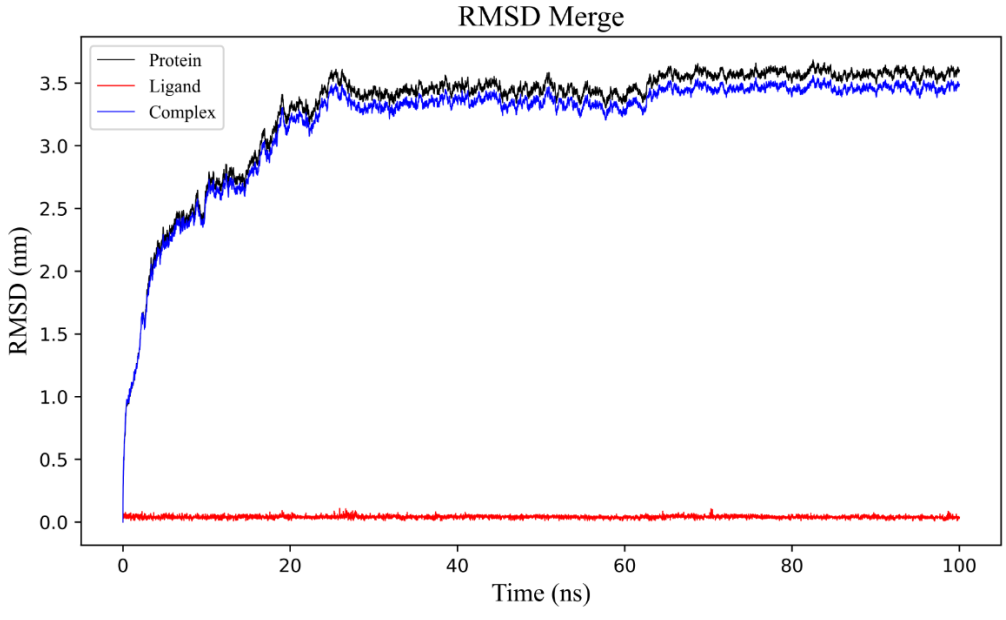
采用该软件的自带工具分析轨迹，计算各氨基酸运动轨迹的均方根方差（RMSD）、均方根涨落值（RMSF）和蛋白质回转半径，结合自由能（MMGBSA），自由能形貌图等数据。

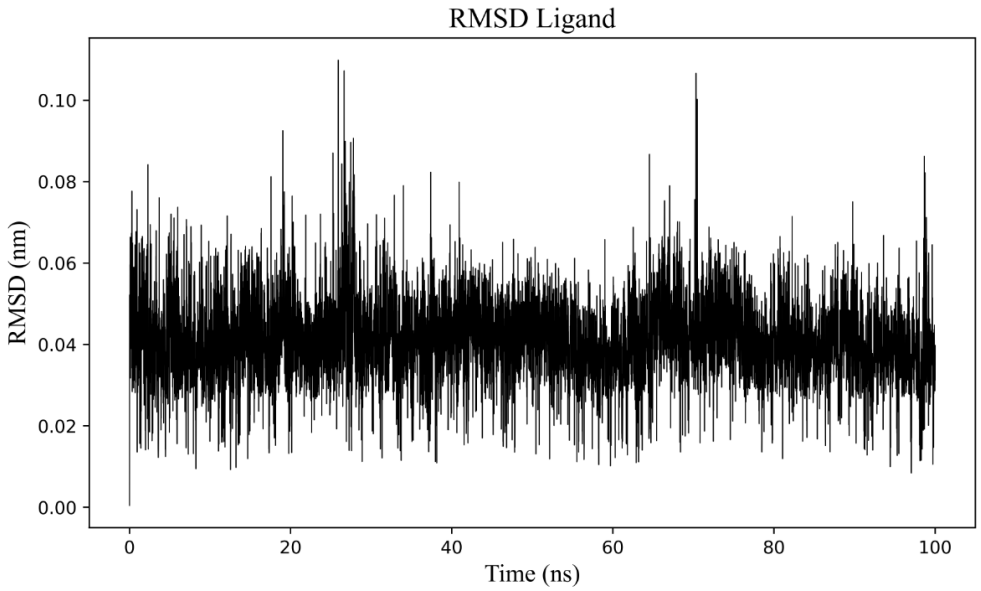
三、实验结果：

A 蛋白

1. RMSD

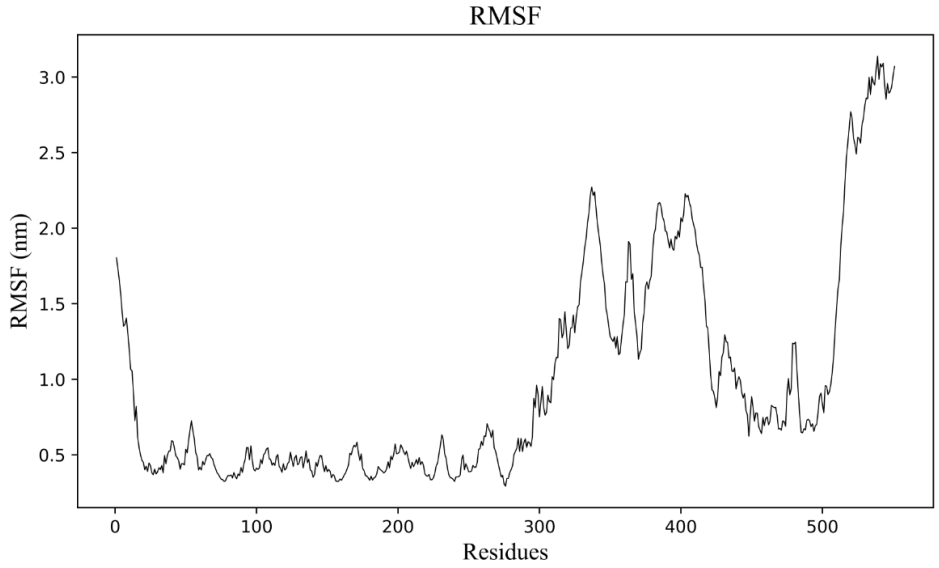
RMSD 是评估蛋白质结构变化的指标。较低的 RMSD 值表明蛋白质的高稳定性，而较高的 RMSD 则表明骨架在模拟时间内发生了结构构象变化。(一般变化范围小于 1 是正常的)。在本次模拟中，我们计算了蛋白质与小分子之间的 RMSD，并绘制了 RMSD 图表（见结果文件夹，RMSD-merge.tiff 文件把蛋白，小分子，蛋白-小分子复合物放在一张图里显示。 rmsd-pro.tiff 文件展示了蛋白 RMSD， rmsd-lig.tiff 文件展示了小分子 RMSD、RMSD-complex.tiff 表示的是蛋白小分子复合物的 RMSD）。结果显示，在前 25 纳秒的模拟过程中，蛋白质的 RMSD 值变化较大，之后逐渐趋于平稳，说明蛋白质构象相较开始有了较大变化，但最终稳定下来。小分子的 RMSD 值始终平稳保持在 0.04 左右，说明小分子已经是最好最稳定的构象。整体来看，RMSD 值经历波折后一直保持平稳，且平稳后变化范围小于 1，说明蛋白质与小分子的结合比较稳定。





2. RMSF

RMSF 是评估蛋白质动态性的指标。RMSF 为原子位置变化对于时间的平均，可以表征蛋白质氨基酸在整个模拟过程中的柔性和运动剧烈程度。蛋白的 RMSF 被认为是确定每个残基与参考位置的偏差。这个参数决定了配体-蛋白质相互作用在模拟时间的适用性。远离活性位点的氨基酸会表现出较高的 RMSF 值。同时，RMSF 和蛋白柔性也有关，越是柔性大的区域，RMSF 值越大。在本次模拟中，我们计算了蛋白质与小分子之间的 RMSF，并绘制了 RMSF 图表（见结果文件夹 RMSF.tiff）。**结果显示，蛋白的后半段及两端部分显示出较大的波动，这通常表明蛋白质的 N-端和 C-端较为灵活。残基 30 至 300 之间的波动性较低，这表明这部分结构相对较为稳定，可能包含蛋白质的核心结构域或活性位点周围的结构。**



3．Secondary-structure

Secondary-structure 图反映了分子动力学模拟过程中蛋白质的二级结构随时间的演化，该图通常用于研究蛋白质结构的稳定性和动态变化，可以帮助理解蛋白质在不同时间段的构象状态。图的横轴表示时间（以纳秒为单位），纵轴表示蛋白质的残基编号。图中不同颜色代表不同的二级结构类型，颜色对应的结构类型在图的右侧有一个图例。

具体解释如下：

红色 (B-Sheet)：表示 β-折叠片 (β-Sheet)，蛋白质二级结构的一种稳定形式。

黑色 (B-Bridge)：表示 β-桥 (β-Bridge)，一种短的 β-片段。

绿色 (Bend)：表示弯曲 (Bend)，一般是指发生在蛋白质主链中的急性转折。

黄色 (Turn)：表示转角 (Turn)，一种将链转回主链方向的结构。

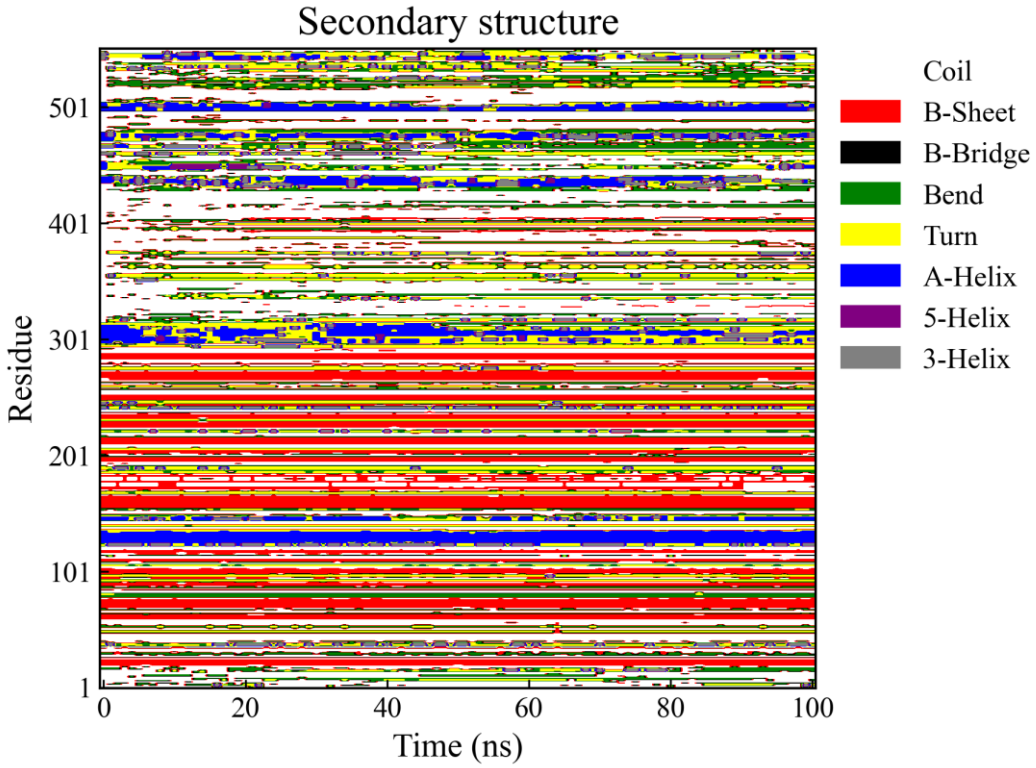
蓝色 (A-Helix)：表示 α-螺旋 (α-Helix)，另一种常见的二级结构，螺旋状。

紫色 (5-Helix)：表示 5 螺旋 (π-Helix)，较不常见的螺旋结构。

灰色 (3-Helix)：表示 3 螺旋 (310-Helix)，一种较窄的螺旋结构。

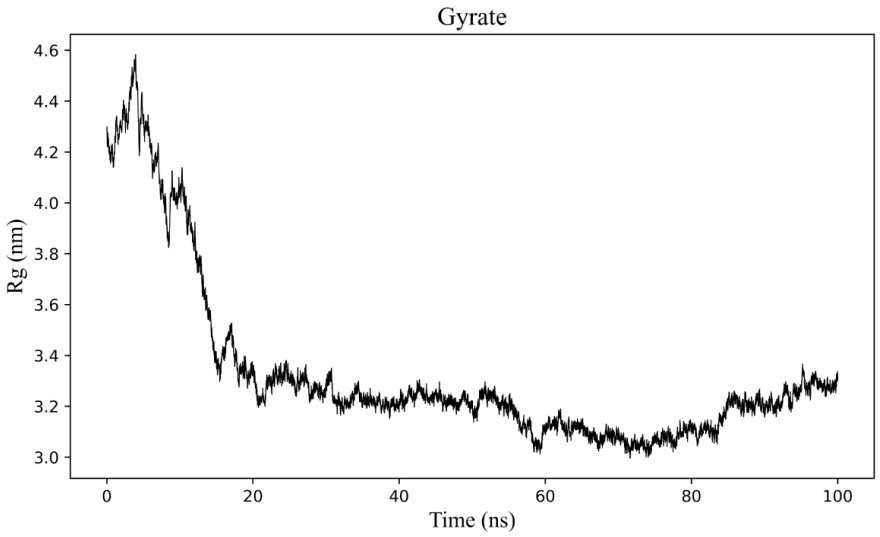
白色/无色 (Coil)：表示无规则卷曲 (Coil)，即没有明显二级结构的区域。

通过观察这个图，可以看到随着时间推移，蛋白质的每个残基二级结构如何变化。颜色的变化表示结构在时间上的变化。例如，如果一个残基的颜色从红色变为白色，则表示它从 β-折叠片转变为无规则卷曲，在本例中，蛋白质在 30-300 残基的二级结构比较稳定。



4. Gyrate

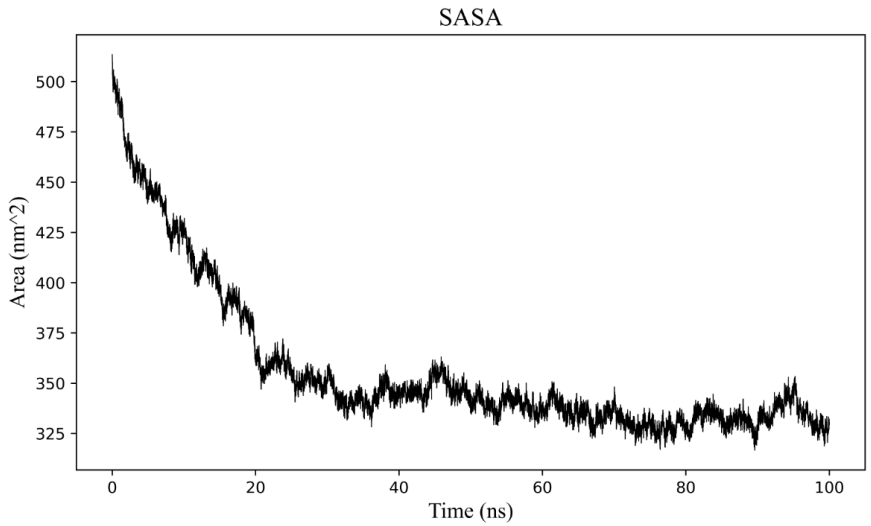
Gyrate 通常指的是半径矢量回转（radius of gyration），简称回转半径。这是一个描述分子形状和尺寸的参数，用于测量分子质量的分布相对于其质心的平均距离的平方，是评估蛋白质整体紧凑程度的指标。它计算蛋白质的平均距离中心点的距离，可用于评估蛋白质的紧密程度和折叠状态的变化。在本次模拟中，我们计算了蛋白质与小分子之间的 Gyrate，并绘制了 Gyrate 图表（见结果文件夹 gyrate.tiff），结果显示在最初的约 20 纳秒内，回转半径从接近 4.6 nm 快速下降到约 3.4 nm。这表明在模拟的初期阶段，蛋白质结构从一个较为松散的状态迅速转变为较为紧密的状态，这是由于模拟开始时蛋白质未能完全放在最稳定状态，经过初始的调整后趋于稳定。**在大约 20 到 100 纳秒的时间内，回转半径基本稳定在 3.0 nm 到 3.4 nm 之间，这表明蛋白质在这段时间内维持了较为稳定的空间构形，表示蛋白质达到了一个热力学相对稳定的状态，构象较为紧密。**



5. SASA

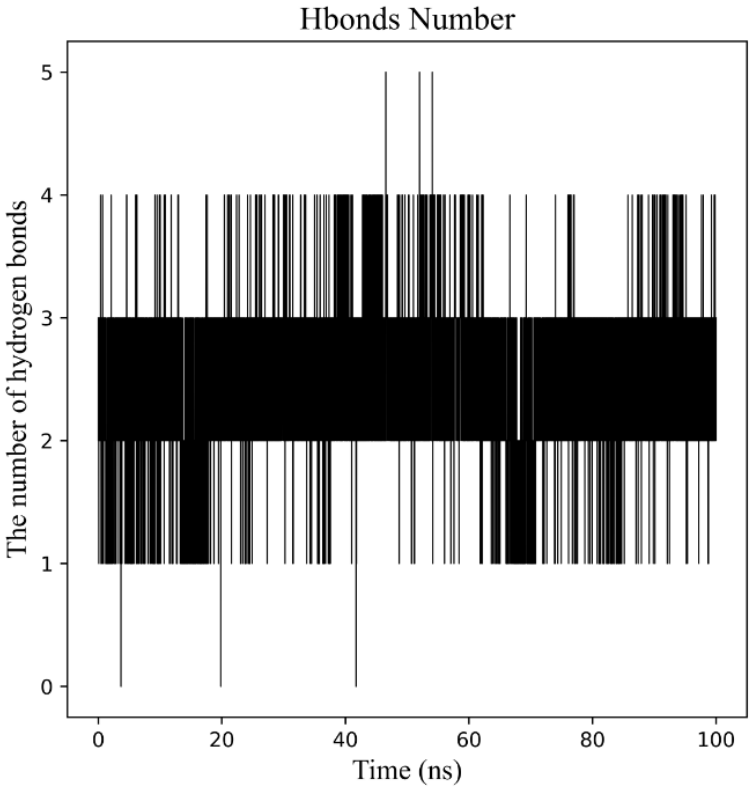
SASA 是评估蛋白质溶剂可及表面积的指标，溶剂可及表面积是指蛋白质表面可与溶剂（通常为水）接触的区域的面积，这个参数对于理解蛋白质的表面特性和相互作用非常重要。在本次模拟中，我们计算了蛋白质与小分子之间的 SASA，并绘制了 SASA 图表（见结果文件夹 SASA.tiff）。结果显示，从 0 至大约 20 纳秒，SASA 从接近 500 nm^2 快速下降至约 375 nm^2。这表明蛋白质在模拟初期迅速从一个较开放的状态转变为较为紧密的状态，减少了与溶剂的接触面积。在 20 纳秒之后，SASA 值在大约 325 nm^2至 360 nm^2之间波动，显示出在这个阶段蛋白质的外部构形相对稳定，与溶剂的接触面积保持在一个较稳定的范围。

结合之前的 Gyrate 分析，回转半径的减少与 SASA 的减少一致，**都表明蛋白质的整体结构在模拟初期变得更加紧凑。这种变化通常是由于蛋白质内部的非共价相互作用（如氢键、疏水相互作用等）的重新排列。**



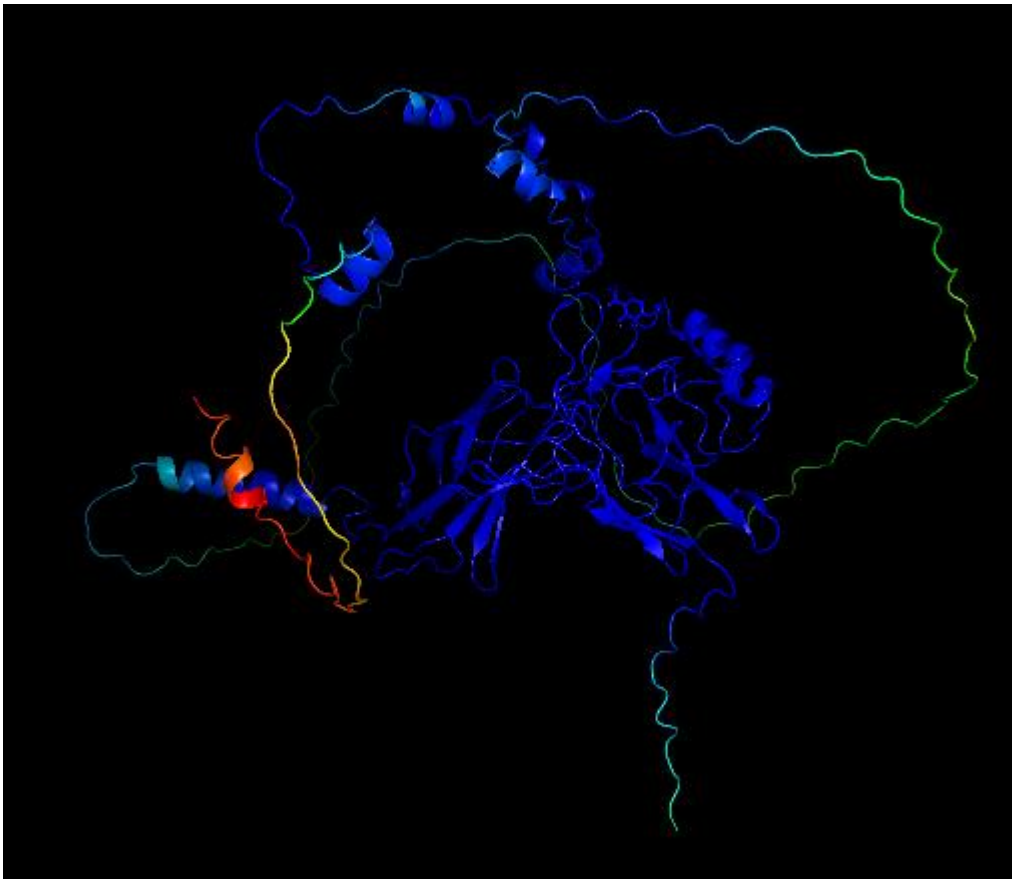
6. HBond

HBond 是评估蛋白质与小分子之间氢键的指标。在本次模拟中，我们计算了蛋白质与小分子之间的 HBond，并绘制了 HBond 图表（见结果文件夹 Hbond-number.tiff）。**结果显示，在模拟的过程中，氢键数量在 1 至 4 之间波动，绝大多数时间点上的氢键数目为 2 或 3，氢键数的波动表明蛋白质在模拟过程可能经历了多种构象变化，但总体上保持了一定程度的结构稳定性。**



7. B-factor

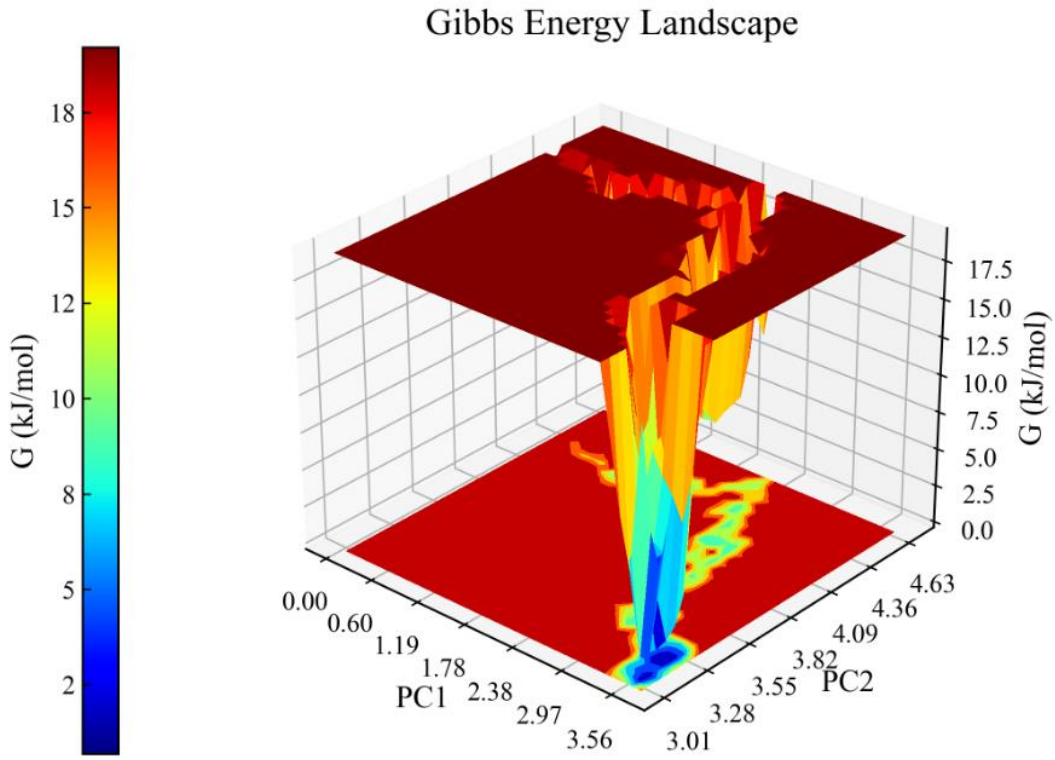
B-factor 是 x 射线做结晶得到的参数，表示蛋白部位的热稳定性，已经有研究表明，RMSF 和 bfactor 是正相关关系，本研究中的 B-factor 是通过跑分子动力学模拟的 RMSF 计算出来的，用 pymol 打开 bfactor-complex.pdb 文件，点击 C > spectrum > b-factors，显示出的图片中蛋白质颜色越红的部分柔性越大，越不稳定，越蓝的部分越稳定，通过计算可知蛋白模型的结构整体稳定。



8. Land-Scape

自由能形貌图（Free Energy Landscape）反映了体系中分子之间的相互作用和能量分布，可以帮助我们理解分子之间的相互作用、构象和稳定性等特征，对于设计分子作用机理和材料设计非常有帮助。其用于描述分子系统中稳定态和过渡态的图形表示，通常使用颜色深浅和等高线的方式表示自由能的大小和分布。热力学上更为稳定的构象通常对应于较低自由能区域，而较不稳定构象对应高自由能区域。图中颜色条代表了 Gibbs 自由能的大小（单位为 kJ/mol），颜色从深蓝色（低自由能）到红色（高自由能）逐渐增加，自由能图中的峰或低自由能区

域（如图中深蓝色和绿色区域）通常代表了能量稳定的构象。自由能形貌图见结果文件夹 LandScape-2d.tif、LandScape-3d.tif。在本次模拟中，图形由 RMSD（均方根偏差）和 Gyrate（半径回旋）两个指标制作而成。在模拟中，我们得到了一个低能量区域。这表明体系中存在一个比较稳定的构象。在这个低能量区域，分子结构可能与初始结构相似度较高，且结构相对紧凑。这表明在模拟过程中，分子可能会在这个状态下停留较长时间，显示出稳定性。根据 rmsd_gyrate_gibbs.xpm、rmsd_gyrate.log、bindex.ndx 中的索引找到最低自由能的构象帧数，提取出该构象保存在 low energy.pdb 中。



9. MM/PBSA

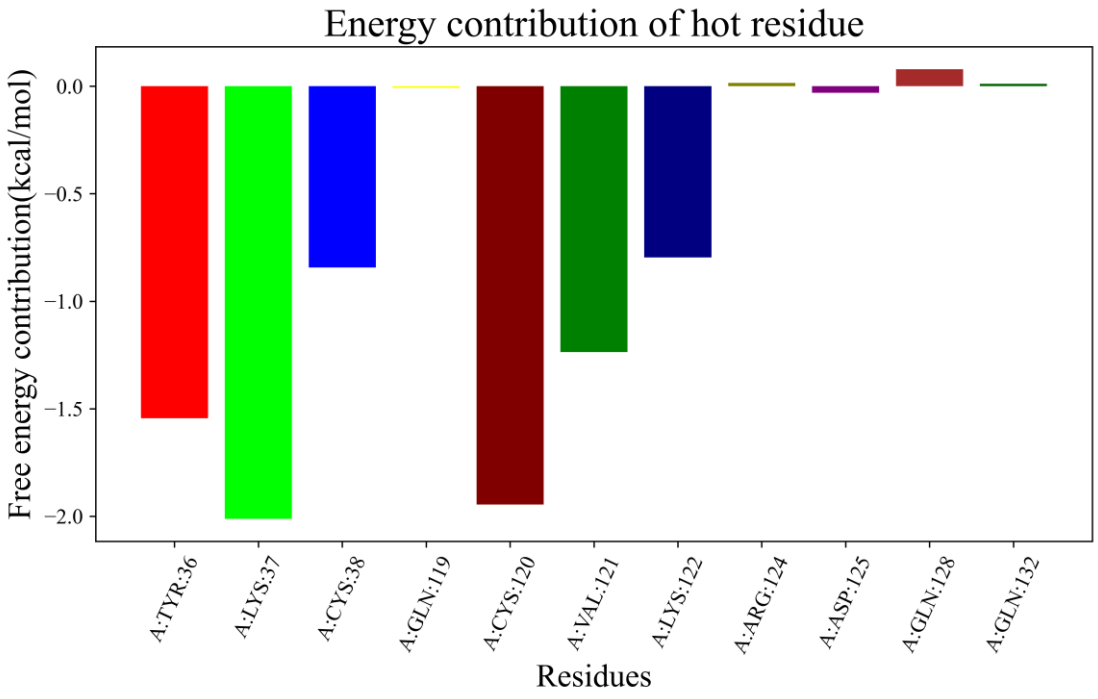
MM/PBSA（Molecular Mechanics Poisson-Boltzmann Surface Area）分析，通过对蛋白质与小分子的能量进行计算，我们可以得到复合物的总自由能，从而判断蛋白质与小分子之间的结合是否稳定。MM/PBSA 结合自由能计算，（见结果文件夹 FINAL_RESULTS_MMPBSA.dat，记事本打开）复合物的 Δ TOTAL 总结合自由能为-12110.56 kcal/mol，FINAL_DECOMP_MMPBSA.dat 展示 MM/PBSA 自由能分解数据。

10. Energy contribution of hot residue

Energy contribution of hot residue 描述了在生物分子，尤其是蛋白质中，某些关键残基（热残基）对分子整体结构和功能的能量贡献。在结构生物学和分子动力学研究中，这种分析有助于理解特定残基如何影响蛋白质的稳定性、活性或与其他分子的相互作用。

横轴：展示了不同的残基名称，包括它们的三字母缩写和位置编号，例如“A.TYR36”表示位于位置 36 的酪氨酸残基。

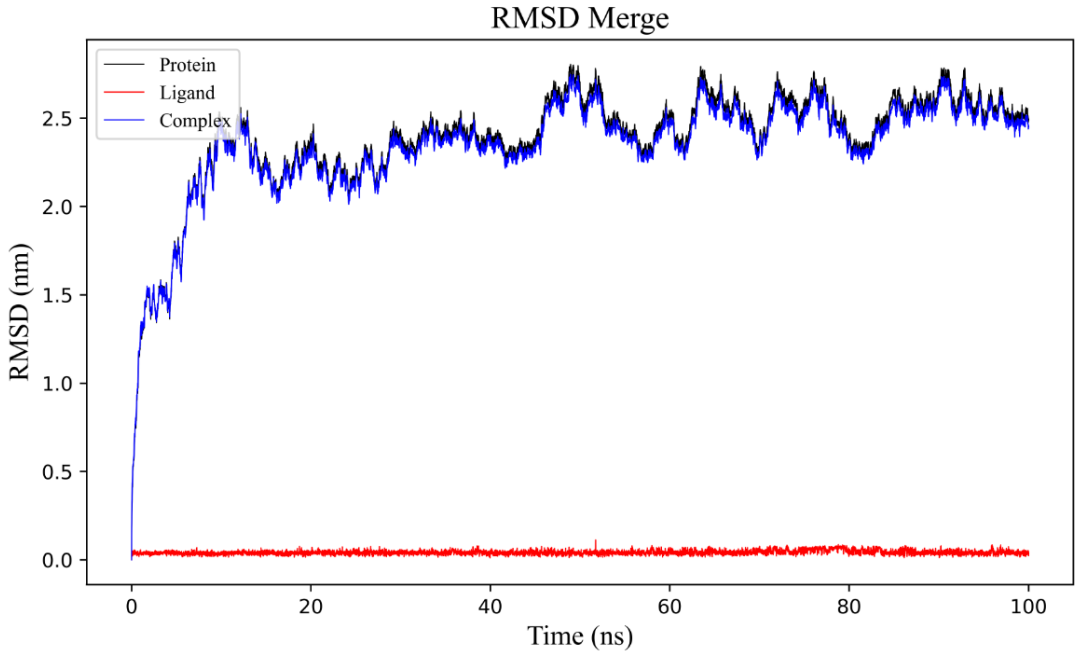
纵轴：表示自由能的贡献，负值表示这些残基对蛋白质结构的稳定性有贡献。负值越高，表示此类残基对蛋白质的结构稳定性贡献越大。

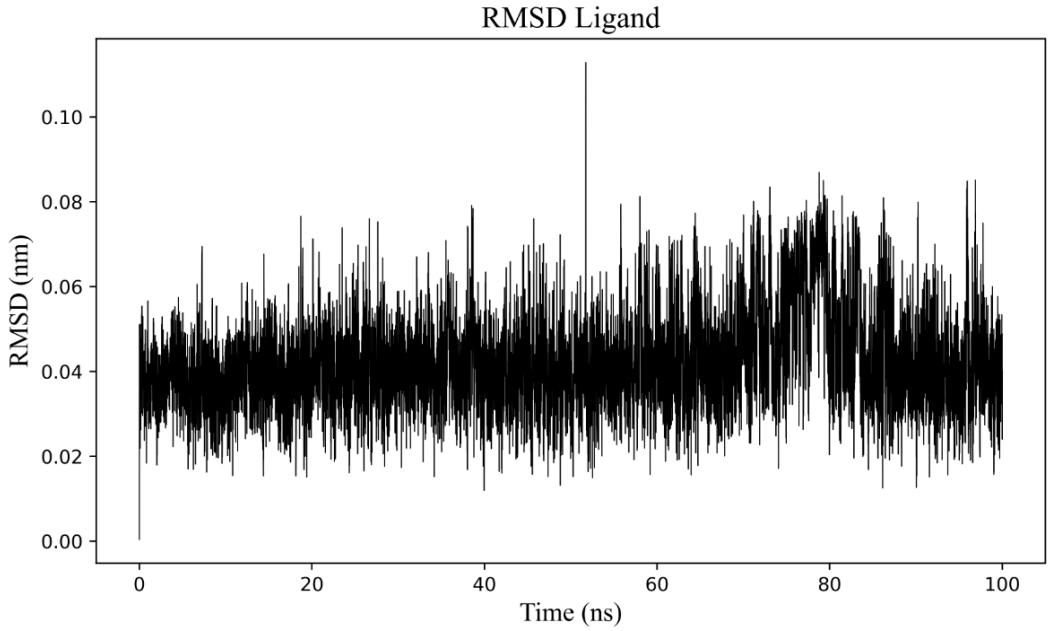


B 蛋白

1. RMSD

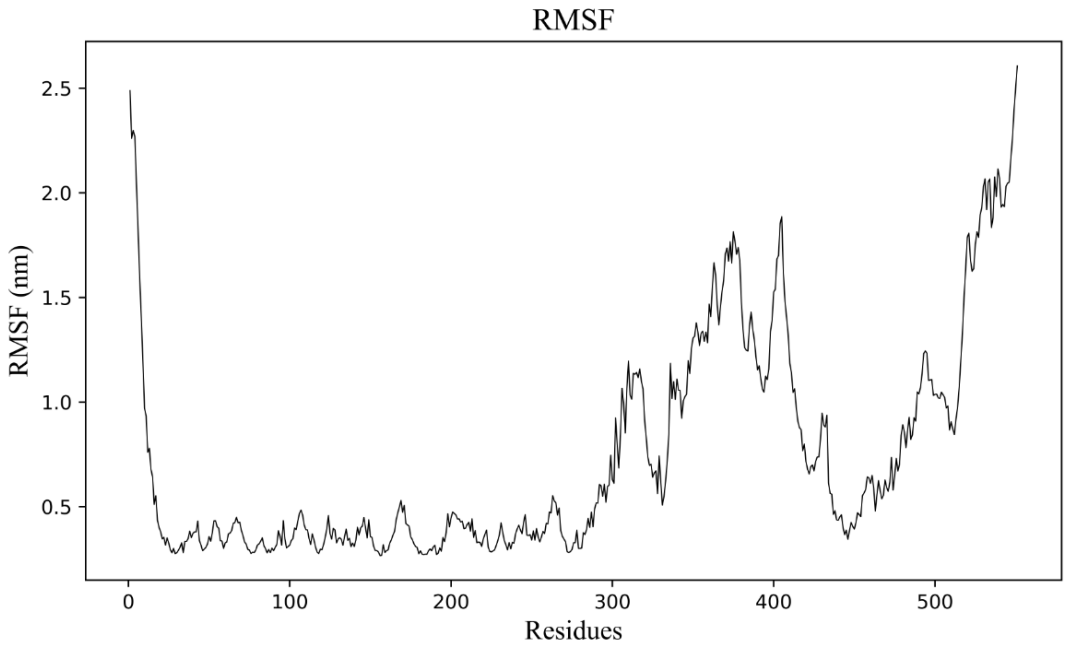
RMSD 是评估蛋白质结构变化的指标。较低的 RMSD 值表明蛋白质的高稳定性，而较高的 RMSD 则表明骨架在模拟时间内发生了结构构象变化。(一般变化范围小于 1 是正常的)。在本次模拟中，我们计算了蛋白质与小分子之间的 RMSD，并绘制了 RMSD 图表（见结果文件夹，RMSD-merge.tiff 文件把蛋白，小分子，蛋白-小分子复合物放在一张图里显示。 rmsd-pro.tiff 文件展示了蛋白 RMSD， rmsd-lig.tiff 文件展示了小分子 RMSD、RMSD-complex.tiff 表示的是蛋白小分子复合物的 RMSD）。结果显示，在前 20 纳秒的模拟过程中，蛋白质的 RMSD 值变化较大，之后逐渐趋于平稳，说明蛋白质构象相较开始有了较大变化，但最终稳定下来。小分子的 RMSD 值始终平稳保持在 0.04 左右，说明小分子已经是最好最稳定的构象。整体来看，RMSD 值经历波折后一直保持平稳，且平稳后变化范围小于 1，说明蛋白质与小分子的结合比较稳定。





2. RMSF

RMSF 是蛋白质中各个残基的均方根波动性，可以表征蛋白质氨基酸在整个模拟过程中的柔性和运动剧烈程度。蛋白的 RMSF 被认为是确定每个残基与参考位置的偏差。这个参数决定了配体-蛋白质相互作用在模拟时间的适用性。远离活性位点的氨基酸会表现出较高的 RMSF 值。同时，RMSF 和蛋白柔性也有关，越是柔性大的区域，RMSF 值越大。在本次模拟中，我们计算了蛋白质与小分子之间的 RMSF，并绘制了 RMSF 图表（见结果文件夹 RMSF.tiff）。**结果显示，蛋白的后半段及两端部分显示出较大的波动，这通常表明蛋白质的 N-端和 C-端较为灵活。残基 30 至 300 之间的波动性较低，这表明这部分结构相对较为稳定，可能包含蛋白质的核心结构域或活性位点周围的结构**。



3．Secondary-structure

Secondary-structure 图反映了分子动力学模拟过程中蛋白质的二级结构随时间的演化，该图通常用于研究蛋白质结构的稳定性和动态变化，可以帮助理解蛋白质在不同时间段的构象状态。图的横轴表示时间（以纳秒为单位），纵轴表示蛋白质的残基编号。图中不同颜色代表不同的二级结构类型，颜色对应的结构类型在图的右侧有一个图例。

具体解释如下：

红色 (B-Sheet)：表示 β-折叠片 (β-Sheet)，蛋白质二级结构的一种稳定形式。

黑色 (B-Bridge)：表示 β-桥 (β-Bridge)，一种短的 β-片段。

绿色 (Bend)：表示弯曲 (Bend)，一般是指发生在蛋白质主链中的急性转折。

黄色 (Turn)：表示转角 (Turn)，一种将链转回主链方向的结构。

蓝色 (A-Helix)：表示 α-螺旋 (α-Helix)，另一种常见的二级结构，螺旋状。

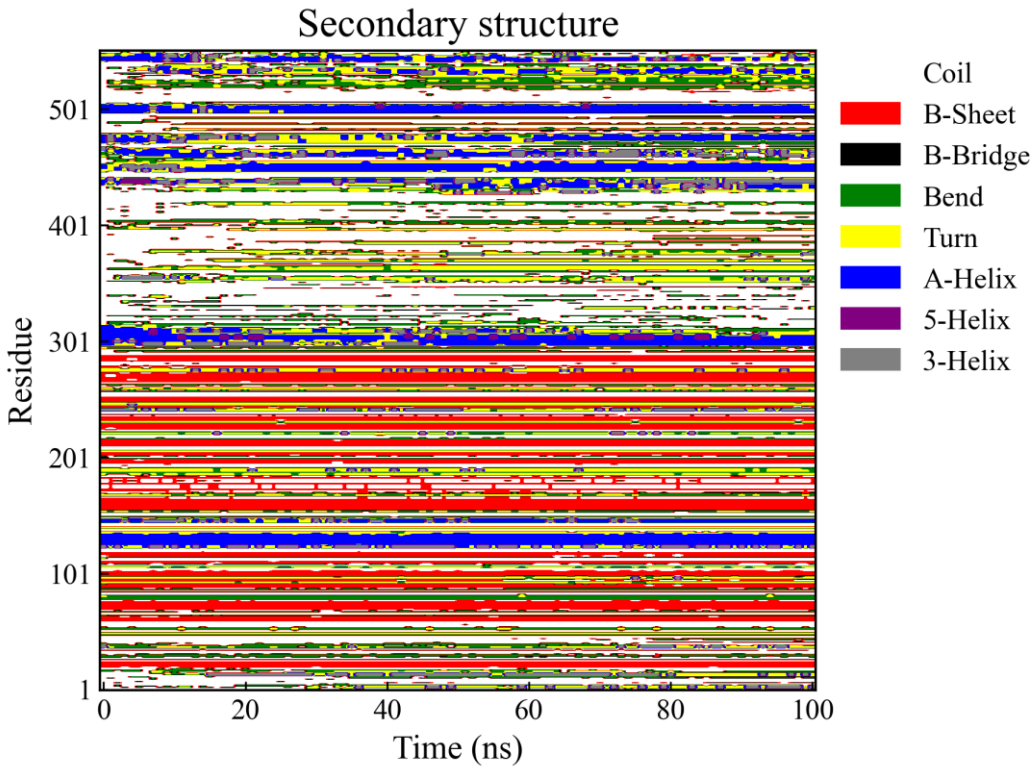
紫色 (5-Helix)：表示 5 螺旋 (π-Helix)，较不常见的螺旋结构。

灰色 (3-Helix)：表示 3 螺旋 (310-Helix)，一种较窄的螺旋结构。

白色/无色 (Coil)：表示无规则卷曲 (Coil)，即没有明显二级结构的区域。

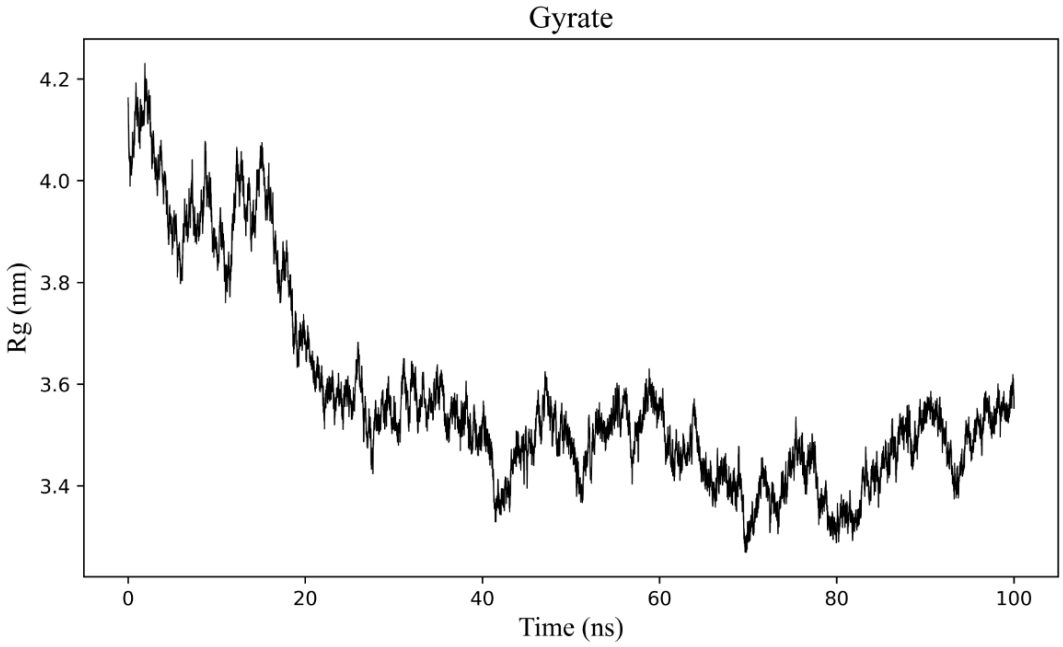
通过观察这个图，可以看到随着时间推移，蛋白质的每个残基二级结构如何变化。颜色的变化表示结构在时间上的变化。例如，如果一个残基的颜色从红色

变为白色，则表示它从 β-折叠片转变为无规则卷曲，在本例中，蛋白质在 30-300 残基的二级结构比较稳定。



4. Gyrate

Gyrate 是评估蛋白质整体紧凑程度的指标。它计算蛋白质的平均距离中心点的距离，可用于评估蛋白质的紧密程度和折叠状态的变化。在本次模拟中，我们计算了蛋白质与小分子之间的 Gyrate，并绘制了 Gyrate 图表（见结果文件夹 gyrate.tiff）。结果显示在最初的约 20 纳秒内，回转半径从接近 4.2 nm 快速下降到约 3.6 nm。这表明在模拟的初期阶段，蛋白质结构从一个较为松散的状态迅速转变为较为紧密的状态，这是由于模拟开始时蛋白质未能完全放在最稳定状态，经过初始的调整后趋于稳定。**在大约 20 到 100 纳秒的时间内，回转半径基本稳定在 3.4 nm 到 3.6 nm 之间，这表明蛋白质在这段时间内维持了较为稳定的空间构形，表示蛋白质达到了一个热力学相对稳定的状态，构象较为紧密。**

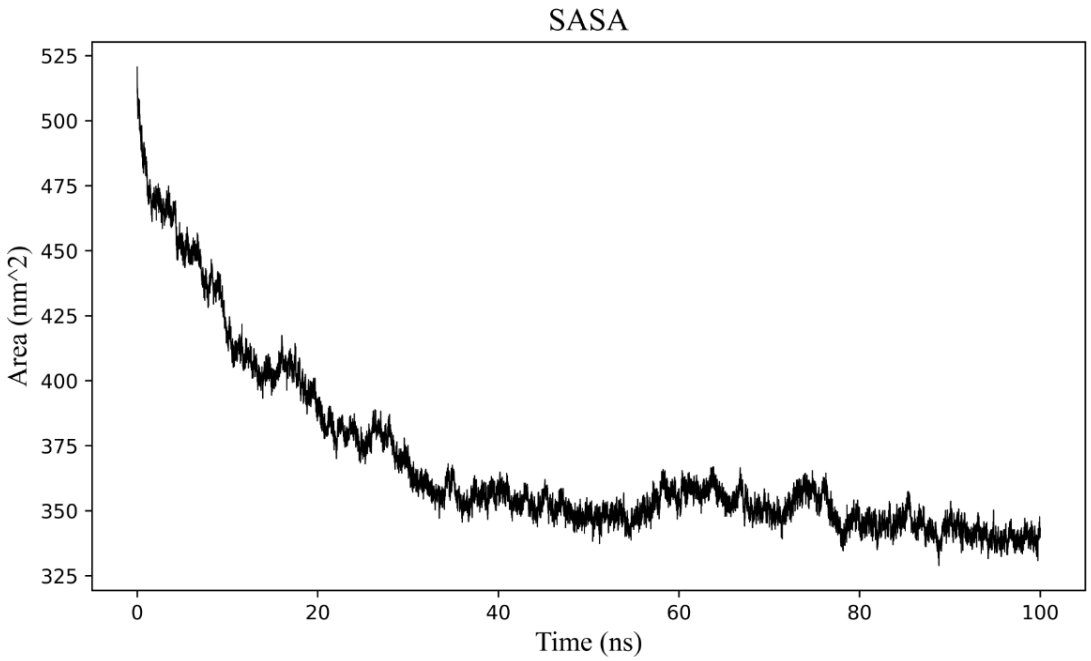


5. SASA

SASA 是评估蛋白质表面面积的指标。在本次模拟中，我们计算了蛋白质与小分子之间的 SASA，并绘制了 SASA 图表（见结果文件夹 SASA.tiff）。SASA 是评估蛋白质溶剂可及表面积的指标，溶剂可及表面积是指蛋白质表面可与溶剂（通常为水）接触的区域的面积，这个参数对于理解蛋白质的表面特性和相互作用非常重要。在本次模拟中，我们计算了蛋白质与小分子之间的 SASA，并绘制了 SASA 图表（见结果文件夹 SASA.tiff）。结果显示，从 0 至大约 20 纳秒，SASA 从接近 525 nm^2 快速下降至约 375 nm^2。这表明蛋白质在模拟初期迅速从一个较开放的状态转变为较为紧密的状态，减少了与溶剂的接触面积。

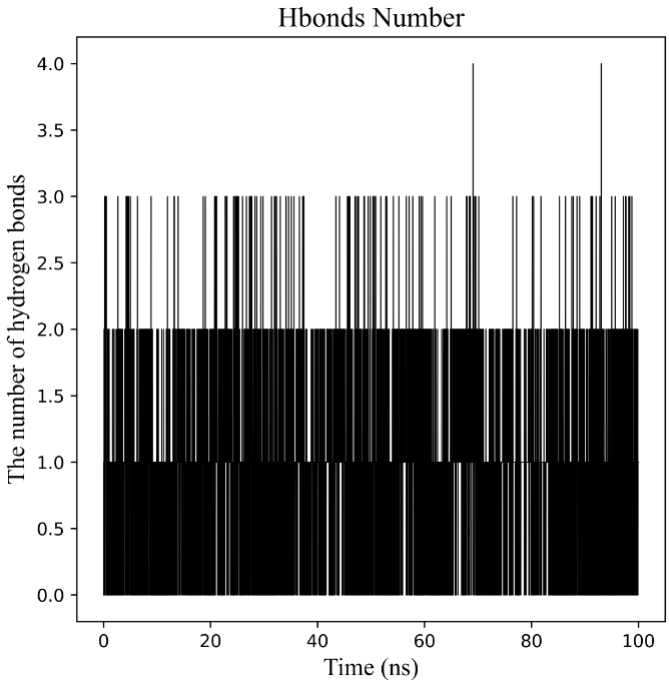
在 20 纳秒之后，SASA 值在大约 325 nm^2至 360 nm^2之间波动，显示出在这个阶段蛋白质的外部构形相对稳定，与溶剂的接触面积保持在一个较稳定的范围。

结合之前的 Gyrate 分析，回转半径的减少与 SASA 的减少一致，**都表明蛋白质的整体结构在模拟初期变得更加紧凑。这种变化通常是由于蛋白质内部的非共价相互作用（如氢键、疏水相互作用等）的重新排列。**



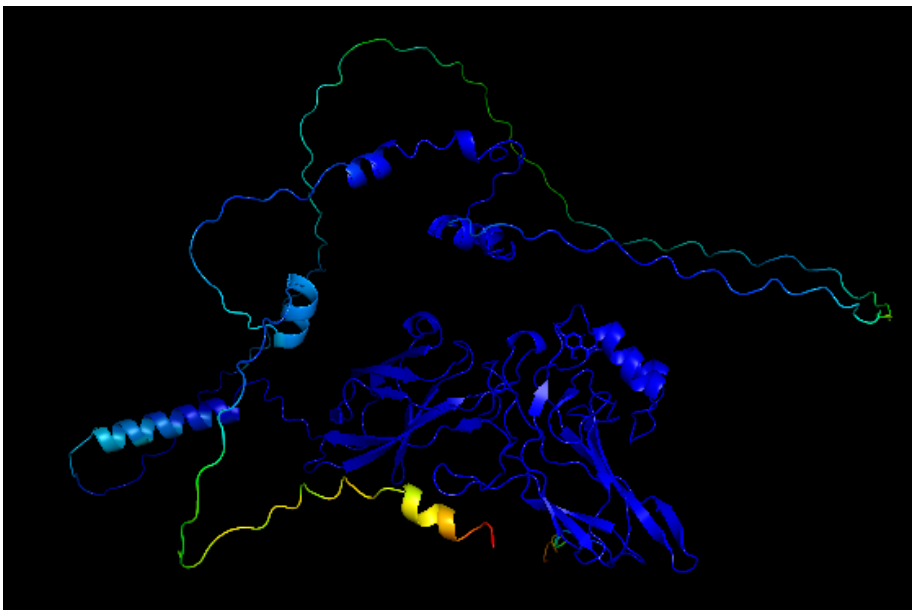
6. HBond

HBond 是评估蛋白质与小分子之间氢键的指标。在本次模拟中，我们计算了蛋白质与小分子之间的 HBond，并绘制了 HBond 图表（见结果文件夹 Hbond-number.tiff）。结果显示，在模拟的过程中，蛋白质与小分子之间形成了许多氢键，其中以蛋白质中的一些关键残基和小分子中的一些重要基团之间的氢键为主。**结果显示，在模拟的过程中，氢键数量在 0 至 3 之间波动，绝大多数时间点上的氢键数目为 1 或 2，氢键数的波动表明蛋白质在模拟过程可能经历了多种构象变化，但总体上保持了一定程度的结构稳定性。**



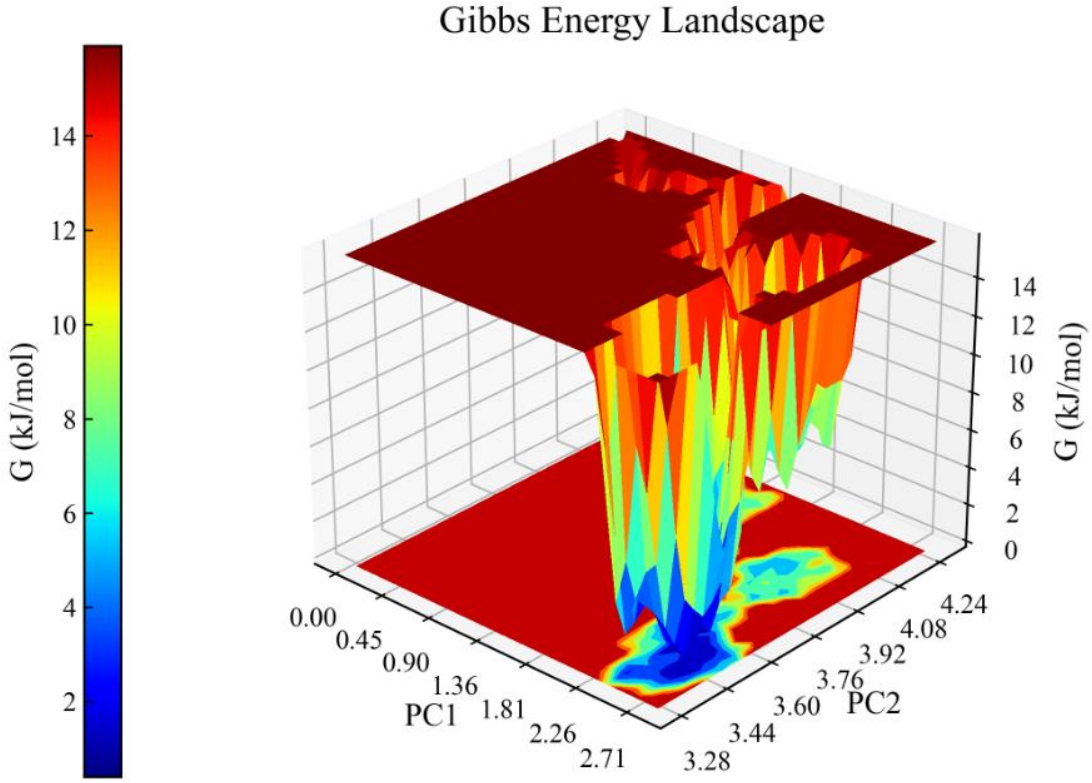
7. b-factor

B-factor 是 x 射线做结晶得到的参数，表示蛋白部位的热稳定性，已经有研究表明，RMSF 和 bfactor 是正相关关系，本研究中的 B-factor 是通过跑分子动力学模拟的 RMSF 计算出来的，用 pymol 打开 bfactor-complex.pdb 文件，点击 C > spectrum > b-factors，显示出的图片中蛋白质颜色越红的部分柔性越大，越不稳定，越蓝的部分越稳定。



8. Land-Scape

自由能形貌图（Free Energy Landscape）反映了体系中分子之间的相互作用和能量分布，可以帮助我们理解分子之间的相互作用、构象和稳定性等特征，对于设计分子作用机理和材料设计非常有帮助。其用于描述分子系统中稳定态和过渡态的图形表示，通常使用颜色深浅和等高线的方式表示自由能的大小和分布。热力学上更为稳定的构象通常对应于较低自由能区域，而较不稳定构象对应高自由能区域。自由能图中的峰或低自由能区域通常代表了能量稳定的构象。自由能形貌图见结果文件夹 LandScape-2d.tif、LandScape-3d.tif。在本次模拟中，图形由 RMSD（均方根偏差）和 Gyrate（半径回旋）两个指标制作而成。在模拟中，我们得到了一个低能量区域。这表明你的体系中存在一个比较稳定的构象。在这个低能量区域，分子结构可能与初始结构相似度较高，且结构相对紧凑。这表明在模拟过程中，分子可能会在这个状态下停留较长时间，显示出稳定性。根据 rmsd_gyrate_gibbs.xpm、rmsd_gyrate.log、bindex.ndx 中的索引找到最低自由能的构象帧数，提取出该构象保存在 low energy.pdb 中。



9. MM/PBSA

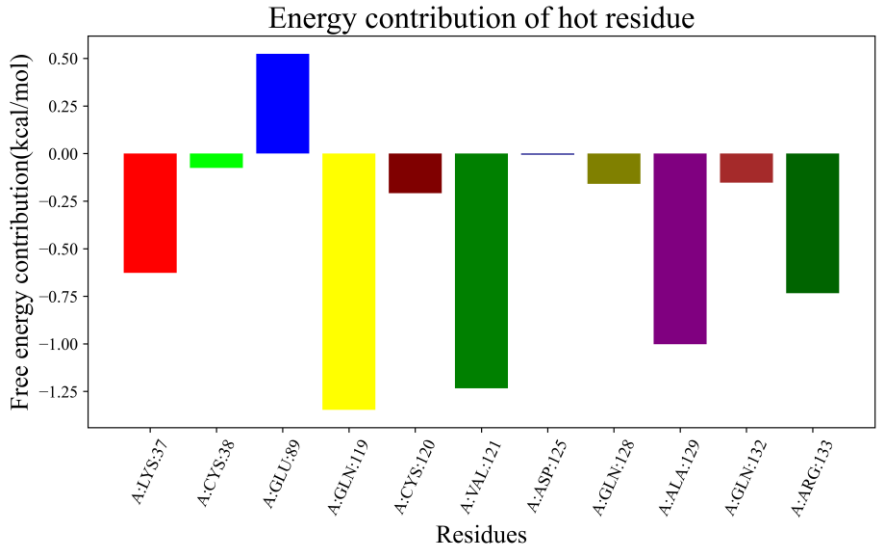
MM/PBSA（Molecular Mechanics Poisson-Boltzmann Surface Area）分析，通过对蛋白质与小分子的能量进行计算，我们可以得到复合物的总自由能，从而判断蛋白质与小分子之间的结合是否稳定。MM/PBSA 结合自由能计算，（见结果文件夹 FINAL_RESULTS_MMPBSA.dat，记事本打开）复合物的 Δ TOTAL 总结合自由能为-12136.30 kcal/mol，FINAL_DECOMP_MMPBSA.dat 展示了 MM/PBSA 自由能分解数据。

10. Energy contribution of hot residue

Energy contribution of hot residue 描述了在生物分子，尤其是蛋白质中，某些关键残基（热残基）对分子整体结构和功能的能量贡献。在结构生物学和分子动力学研究中，这种分析有助于理解特定残基如何影响蛋白质的稳定性、活性或与其他分子的相互作用。

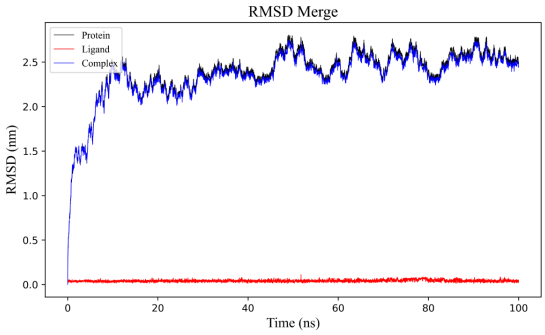
横轴：展示了不同的残基名称，包括它们的三字母缩写和位置编号，例如“A.TYR36”表示位于位置 36 的酪氨酸残基。

纵轴：表示自由能的贡献，负值表示这些残基对蛋白质结构的稳定性有贡献。负值越高，表示此类残基对蛋白质的结构稳定性贡献越大。图中的正值（即柱子向上的部分）表示该残基对系统的自由能有正贡献，这通常意味着这些残基不利于蛋白质-配体或蛋白质-蛋白质相互作用的稳定性，可能降低结合的稳定性或者增加系统的自由能。

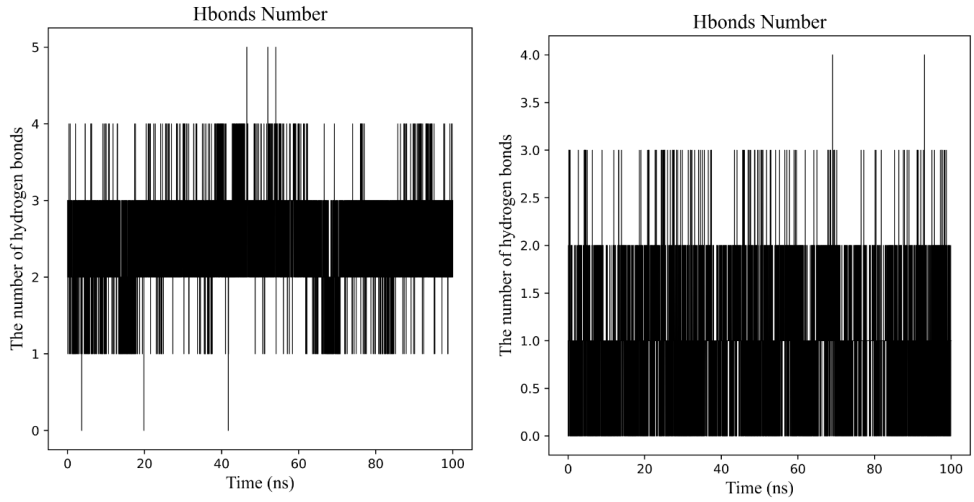


A 蛋白和 B 蛋白的对比

通过对比两个蛋白和小分子的模拟结果，A 蛋白的 RMSD 值比 B 蛋白波动幅度更小更稳定，A 蛋白的氢键数量也比 B 蛋白多，可以说明，A 蛋白相对 B 蛋白来说与小分子化合物结合更稳定。



左侧为 A 蛋白 RMSD 值，右侧为 B 蛋白 RMSD 值结果。



左侧为 A 蛋白氢键数量结果，右侧为 B 蛋白氢键数量结果

六、参考文献：

[1]Abraham M J, Murtola T, Schulz R, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers[J]. SoftwareX, 2015, 1: 19-25.
[2]Van Der Spoel D, Lindahl E, Hess B, et al. GROMACS: fast, flexible, and free[J]. Journal of computational chemistry, 2005, 26(16): 1701-1718.