



博士学位论文

蛋白质-配体结合自由能及其构象采样的方法发展与应用

作者姓名: 彭 诚

指导教师: 朱维良 研究员 中国科学院上海药物研究所

徐志建 副研究员 中国科学院上海药物研究所

学位类别: 理学博士

学科专业: 药物设计学

培养单位: 中国科学院上海药物研究所

2021 年 6 月

**The development and application of methods for calculating
protein-ligand binding free energy and conformational sampling**

**A dissertation submitted to
University of Chinese Academy of Sciences
in partial fulfillment of the requirement
for the degree of
Doctor
in Drug Design**

By

Peng Cheng

Supervisor

Professor Weiliang Zhu

Associate Professor Zhijian Xu

Shanghai Institutes of Materia Medica, CAS

June 2021

中国科学院大学
研究生学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：

日 期：

中国科学院大学
学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以按照学术研究公开原则和保护知识产权的原则公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名：

日 期：

导师签名：

日 期：

摘 要

蛋白质分子的构象变化与其生物学功能密切相关,配体的结合往往也会导致靶标蛋白的大规模构象变化。因此,研究蛋白质构象变化及与配体的相互作用对生物学机制和药物应用研究都具有重要意义。然而,常规的实验技术或分子理论模拟在研究靶标蛋白的大规模构象变化及蛋白-配体相互作用中面临着巨大的挑战,如常规理论计算的时间尺度远小于实际时间尺度,结合能计算准确性也有待提高等。为此,本论文紧密围绕高效准确模拟靶标蛋白大规模构象变化及精确模拟靶标蛋白与配体相互作用这两个方面开展研究工作。

论文第一部分旨在发展靶标蛋白构象采样及其与配体相互作用计算的新方法。为了在模拟靶标蛋白与配体的相互作用时精确考虑蛋白质-配体复合物受到的静电极化效应,我们发展了 SQMPC 方法来计算复合物体系的半经验量子化学 (SQM) 原子电荷,并将其与结合自由能计算方法 SIE 和 MM/GBSA 相结合。在 50 个测试体系中发现,结合 SQMPC 方法的结合自由能计算值与实验值的相关性系数 (R^2) 均高于常规力场 AMBER-ff03 电荷,其中 AM1 和 PM7 方法在四种测试的 SQM 方法中表现最佳。另外,SIE 方法得到的 R^2 总体比 MM/GBSA 方法略高,并且 SIE 方法不需要额外计算熵效应。因此,我们认为 SIE 与 SQMPC 方法的组合 (SIE-SQMPC) 将有助于蛋白质-配体结合自由能的快速准确预测。

近年来,非平衡分子动力学模拟如自适应拉伸分子动力学 (ASMD),通过配体结合或解离过程的平均力势 (PMF) 来预测配体结合能,能够兼顾焓变和熵变效应而受到关注,但这类方法通常也忽略了在配体解离过程中蛋白质受到配体及溶剂的静电极化作用。因此,为进一步扩展 SQMPC 的应用及提高 ASMD 方法的准确性,本论文将发展的 SQMPC 方法与 ASMD 结合,定期更新 ASMD 模拟中蛋白质-配体复合物的原子电荷,从而发展了 ASMD-SQMPC 方法。在 24 个 HIV-1 蛋白酶抑制剂体系中表明,ASMD-SQMPC 方法 ($R^2=0.86$) 能显著提高常规 ASMD ($R^2=0.52$) 方法的结合自由能预测准确性,并且对含氟的配体也有效 ($R^2=0.71$)。我们进一步分析发现随着配体的解离和靶标蛋白的构象变化,体系原子电荷发生着明显的电荷重分配。因此,ASMD-SQMPC 方法将有助于快速准确地研究配体的解离及相对结合自由能,同时也表明原子电荷的精确计算在理论模拟中的重要性。

靶标蛋白在功能发挥或结合配体时, 往往会发生大规模构象变化, 但在有限的计算机资源下, SIE-SQMPC 等方法中采用的常规分子动力学模拟由于受到其低采样效率的限制, 难以对这类构象变化进行充分采样。因此, 为提高靶标蛋白构象变化及其与配体相互作用的采样效率, 本论文的第三章研究工作致力于发展高效的增强型采样新方法。基于优化的副本交换分子动力学 (REMD) 框架和简正振动模式分析 (NMA) 作为集合变量, 我们发展了一种基于能量和几何二维空间同时加速的增强型采样方法, 称之为 ossPTMetaD。为评估 ossPTMetaD 的采样效率, 我们将其应用到 5 种功能各异的靶标蛋白体系, 包括 3 种不同的运动类型: AdK 体系的结构域相对运动、HIV-1 蛋白酶和 BACE1 体系的开放/闭合构象转变以及激酶体系的 DFG 模式翻转运动。结果表明, ossPTMetaD 能够高效准确地模拟这些蛋白质的功能性构象变化, 并且 ossPTMetaD 相比传统 REMD 方法的采样收敛速度更快, 所需计算资源也减少了约 70.0%。

论文第二部分旨在综合我们自主发展的新方法以及各种计算模拟手段解决具体科学研究问题。在第四章中, 为探究大环化合物的构象多样性与其透膜性、极性等药理性质的关系, 我们通过常规 REMD 模拟研究了大环药物劳拉替尼在不同极性溶剂环境 (纯水溶剂、DMSO/水混合溶剂和氯仿溶剂) 中的构象分布。结果表明, 在高极性的纯水和 DMSO/水混合溶剂中, 劳拉替尼与溶剂形成密集的氢键网络, 限制其构象变化, 因此只稳定呈现一种构象 1, 这种构象与劳拉替尼结合靶标蛋白的晶体构象高度相似。而在低极性氯仿溶剂中, 劳拉替尼除构象 1 外 (占比 80%), 还出现一类采样较少的构象 2 (占比 20%)。与构象 1 相比, 构象 2 的溶剂可及极性表面积和偶极矩都更小, 表明劳拉替尼可能通过构象 1 与构象 2 间的转变来实现它良好的透膜性。进一步实验验证表明, 劳拉替尼在不同极性溶剂下构象分布的模拟结果与核磁共振波谱 (NMR) 数据高度一致。

新型冠状病毒 (SARS-CoV-2) 自 2019 年末爆发后迅速蔓延全球, 目前尚未发现可以治疗 SARS-CoV-2 引发的新冠肺炎 (COVID-19) 的特效药。SARS-CoV-2 的刺突蛋白以三聚体形式嵌在病毒表面, 介导病毒与宿主细胞血管紧张素转换酶 2 (ACE2) 的结合, 从而帮助病毒入侵和感染宿主细胞, 是目前许多疫苗和抗体药物的重要靶点之一。研究表明, 刺突蛋白具有明显的构象动态, 其受体结合域 (RBD) 由 ‘向下’ 构象转变成 ‘向上’ 构象后才与 ACE2 受体结合。因此, 本论文通过课题组自主发展的高效采样方法 NUMD 和 vsREMD 探索了 SARS-

CoV-2 刺突蛋白的动态构象变化，并综合多种计算模拟手段（同源模建、结合自由能计算、自由能分解和可药性位点预测等）分析了刺突蛋白处于不同构象形态时的宿主 ACE2 结合能力，解析了刺突蛋白与宿主 ACE2 受体结合的关键作用残基，并预测了 5 个 SARS-CoV-2 刺突蛋白上的可药性结合位点。模拟结果发现随着刺突蛋白‘向上’的角度越大，SARS-CoV-2 刺突蛋白与宿主 ACE2 的结合能力也变得越强，并且在相同构象状态下强于非典病毒（SARS-CoV）。但是，相比 SARS-CoV，SARS-CoV-2 刺突蛋白呈现 RBD ‘向上’构象的比例更少，转变成 RBD ‘向上’状态的能垒也更高。这些结果表明尽管 SARS-CoV-2 刺突蛋白的 RBD 结构域具有比 SARS-CoV 更强的宿主 ACE2 结合能力，但 SARS-CoV-2 更不容易从‘向下’状态转变成‘向上’状态，因此其全长结构的 ACE2 结合能力与 SARS-CoV 相当甚至更弱。

随着 PDB 数据库中蛋白质三维结构的数量增加，蛋白质构象变化与其功能的相关性研究逐渐有了数据基础。在第五章中，为进一步帮助理解蛋白质运动的规律和影响因素，我们通过数据整理和可药性位点预测（D3Pockets）等方法，构建了一个基于 PDB 数据库的蛋白质运动数据库（D3PM）。我们发现蛋白质除了整体结构的大规模运动，其口袋残基的局部运动在蛋白质功能机制研究中同样不可忽视。但这种局部运动由于涉及氨基酸残基少，往往很难表征出来。因此，我们针对这类口袋残基的局部运动进行了系统分析并整理成 5 类，即口袋开拓型运动、口袋扩展型运动、口袋融合型运动、口袋闭合型运动以及其他类型运动。最终，D3PM 数据库中包含了 2319 对结合口袋残基局部运动和 5339 对蛋白质整体运动。基于 D3PM，我们比较了影响蛋白质运动的不同因素，结果表明配体诱导的蛋白质运动最为显著，但配体结合后会与残基形成相互作用，从而稳定结合口袋残基结构。此外，我们统计了 20 种氨基酸残基在结合口袋上的偏好性，以助于可药性结合位点的预测等。

关键词：结合自由能计算，增强采样，副本交换分子动力学，新冠病毒，蛋白质运动数据库

Abstract

The protein conformational changes are often related to its biological functions and its ligand-binding. Thus, understanding the molecular mechanisms of these motions and protein-ligand interactions is very important in both biology and applied medical research like drug design. However, due to the limitation of current experimental technologies and molecular modeling, it is still challenging to efficiently capture the dynamic process of protein large-scale conformational change, and to accurately predict protein-ligand binding free energy. This dissertation is focused on the development and application of novel theoretical methods to efficiently and accurately investigate the protein large-scale conformational change and protein-ligand interaction.

The first section is the development of novel theoretical methods for enhanced sampling of protein conformational space and predicting protein-ligand interactions. To deal well with the electrostatic polarization effect for studying protein-ligand interactions, we developed a novel method to calculate the atomic partial charges of protein-ligand complexes using semi-empirical quantum mechanics (SQM), namely SQMPC. Combining with SIE and MM/GBSA approaches, the test of SQMPC on 50 protein-ligand complexes achieved higher correlation coefficient (R^2) between the predicted and experimental binding free energy (ΔG) than AMBER-ff03 charge. In addition, AM1 and PM7 obtained much higher R^2 than other SQM methods, and SIE achieved more accurate and faster R^2 than MM/GBSA as no extra entropic calculation in SIE. Therefore, the combination of SIE and SQMPC, namely SIE-SQMPC, provided an efficient and accurate alternative to predict the ΔG of protein-ligand interactions.

Recently, non-equilibrium molecular dynamics (MD) simulation methods including the adaptive steered molecular dynamics (ASMD) have received increasing attentions to characterize protein-ligand interactions based on the potential of mean force (PMF). Moreover, such approaches contain both enthalpy and entropy of ligand binding. However, the electrostatic polarization effect is ignored in ASMD simulation of ligand unbinding process with normal force field. Therefore, to further expand the application of SQMPC and improve the accuracy of ASMD, we prosposed an innovative approach by combining the SQMPC and ASMD, namely ASMD-SQMPC, with regularly updated atomic partial charges. The R^2 between the PMF and experimentally determined ΔG is 0.86 for 24 inhibitors of HIV-1 PR, while that for the

conventional ASMD is only 0.52. Moreover, ASMD-SQMPC obtains satisfactory prediction for fluorine-containing molecules ($R^2 = 0.74$). Further investigation showed that the atomic charges of system continuously redistribute during the ligand unbinding and protein conformational changes. Therefore, this new method (ASMD-SQMPC) provides an efficient approach to study ligand unbinding process and its relative ΔG , and reveals the significance of precise atomic partial charges in theoretical simulations.

Protein often undergoes large-scale conformational changes when it functions or binds to a ligand. Due to its relatively low sampling efficiency, conventional MD used in SIE-SQMPC is difficult to reach the time-scale of important conformational changes with limited computational resources. In order to improve the sampling efficiency for protein conformational changes and protein-ligand interactions, we developed an enhanced sampling method (Chapter 3). A novel algorithm based on optimized replica exchange molecular dynamics (REMD) and using normal mode analysis (NMA) as collective variables, namely ossPTMetaD, is proposed with attempt to accelerate MD simulations simultaneously in temperature and geometrical spaces. For testing the performance of ossPTMetaD, five protein systems with three kinds of motion patterns were selected, including large-scale domain motion (AdK), open/closed movement (HIV-1 protease and BACE1), and DFG-motif flip in kinases (p38 α and c-Abl). The simulation results showed that the ossPTMetaD requires much less number of replicas than conventional REMD with a reduction of $\sim 70\%$, to achieve similar simulation results with conventional REMD and experimental data.

In the second section, we applied self-developed new methods and various computational techniques on specific scientific researches. For understanding the relationship between the conformational diversity and pharmacological properties of macrocycles, we predicted the conformation ensembles of macrocycles in polar and nonpolar environments using lorlatinib as an example (Chapter 4). The results showed that lorlatinib adopts the same conformer in H₂O/DMSO and H₂O solvent as when bound to its protein target (conformer 1), while an additional conformer 2 (20%) exists in CHCl₃. The hydrogen bond network formed between lorlatinib and H₂O is proposed to play an important role in constraining its conformational changes. The calculated solvent accessible polar surface area of conformer 2 is significantly smaller than that of 1. In addition, conformer 2 has a lower dipole moment compared to 1. The results suggested that the drug could switch its conformation while crossing membranes, thus, favoring its cell permeability. Furthermore, the simulation results are highly consistent

with that of solution NMR studies.

A new coronavirus known as SARS-CoV-2 has emerged at the end of 2019 and spreaded worldwide rapidly. However, very few therapies for COVID-19 have shown promising results so far. The spike (S) glycoprotein trimer of SARS-CoV-2 plays an essential role in viral infection, engaging human angiotensin-converting enzyme 2 (ACE2) as a receptor and mediating the fusion of SARS-CoV-2 and cellular membranes, making it an ideal target for anti-viral drug and vaccine. In order to recognize the ACE2, the receptor-binding domain (RBD) of S undergoes hinge-like conformational change from “down” to “up” conformation. Therefore, to explore the underlying mechanism of S-ACE2 binding, we simulated the conformational transition of S by using enhanced sampling methods (NUMD and vsREMD), and by using various computational methods (homology modelling, binding free energy calculation, free energy decomposition, druggable pockets prediction) we compared the ACE2-binding affinities of S in different states, identified key residues of S-ACE2 interactions, and predicted 5 potential ligand-binding pockets in S. The results showed that the ACE2-binding strength of the S gets stronger as the S becomes more “up”, and is much stronger than that of SARS-CoV S. However, the free energy barriers of the transition from RBD “down” to “up” of SARS-CoV-2 S are higher than that of SARS-CoV S. The predicted population of RBD “up” conformation of SARS-CoV-2 S is much less than that of SARS-CoV S. The results showed that while the ACE2-binding affinity of SARS-CoV-2 S is much stronger than that of SARS-CoV S, similar or even weaker binding affinity of SARS-CoV-2 S than SARS-CoV S could be identified if entire proteins are considered in experiments.

With more and more protein structures becoming available in PDB database, there is increasing data to relate structures to protein motions for studying their functions. Based on PDB, we constructed a comprehensive protein motion database (D3PM) by using data collection and pocket prediction (D3Pockets) for facilitating the analysis of protein motions with their influence factors (Chapter 5). We found that the local flips in pockets play an important role in protein functions, besides protein overall motions. However, it is difficult to present those local motions. Therefore, we collected and classified those pocket residues' motions into 5 types, viz. pocket-creating motion, pocket-expanding motion, pocket-fusing motion, pocket-shrinking motion and other motions. Finally, 5339 representative pairs with overall motions and 2319 representative pairs with local motions of pocket residues were collected. With the

D3PM, we compared the strength of different factors to drive protein conformational changes. The results showed that protein motions induced by ligands is significant, while the binding of ligands could rigidize the pocket for conformational adaptation. In addition, we analyzed the residues' preferences in the ligand binding site with hope to facilitate the prediction of druggable ligand-binding pockets.

Key Words: Binding free energy calculation, Enhanced sampling, Replica exchange molecular dynamics, SARS-CoV-2, Protein motion database

目 录

第 1 章 绪论.....	1
1.1 引言	1
1.2 蛋白质-配体结合自由能的准确预测	3
1.2.1 结合自由能计算方法.....	3
1.2.2 溶剂效应.....	5
1.3 副本交换分子动力学变形方法的蓬勃发展	5
1.3.1 基于温度的副本交换分子动力学.....	5
1.3.2 改善交换方式和提高交换成功率.....	7
1.3.3 缩减执行交换时的体系总势能以减少副本数目.....	8
1.3.4 基于体系哈密顿的 REMD 方法	10
1.3.5 副本交换引入其他增强采样方法.....	12
1.3.6 引入模拟过程物理量为增强采样变量.....	15
1.3.7 调整模拟流程提高采样效率.....	15
1.3.8 采样方法.....	17
1.3.9 适应分布和异构式计算环境解决资源需求问题.....	17
1.3.10 小结.....	18
1.4 论文总体安排	18
第 2 章 半经验量子化学电荷提高结合自由能计算精度	21
2.1 研究背景	21
2.2 材料与方法	23
2.2.1 体系准备.....	23
2.2.2 半经验量子化学电荷 SQMPC 计算	31
2.2.3 常规分子动力学模拟.....	31
2.2.4 结合自由能计算.....	32
2.2.5 自适应拉伸分子动力学模拟.....	33
2.3 结果与讨论	33
2.3.1 SQMPC 方法提高雌激素受体抑制剂的结合稳定性	33
2.3.2 雌激素受体与配体的结合自由能预测准确性评估	37
2.3.3 SQMPC 方法的体系可拓展性评估	38

2.3.4 MD 模拟时间对结合自由能相关性的影响	42
2.3.5 SQMPC 方法应用于非平衡分子动力学模拟	43
2.4 本章小结	51
第 3 章 蛋白质构象高效采样方法的发展及应用	53
3.1 研究背景	53
3.2 材料与方法	55
3.3 结果与讨论	58
3.3.1 ossPTMetaD 方法的计算效率与采样效率评估	58
3.3.2 ossPTMetaD 高效准确模拟 AdK 构象变化	59
3.3.3 ossPTMetaD 高效准确模拟蛋白 Flap 柔性变化	62
3.3.4 ossPTMetaD 高效准确模拟激酶 DFG 翻转运动	67
3.3.5 ossPTMetaD 与其他增强采样方法的比较	70
3.4 本章小结	73
第 4 章 高效采样模拟方法的实际应用	75
4.1 新冠病毒刺突蛋白与宿主 ACE2 结合机制的计算研究	75
4.1.1 研究背景	86
4.1.2 研究方法	87
4.1.3 结果与讨论	90
4.1.4 小结	108
4.2 大环化合物在不同极性溶剂下的构象分布	75
4.2.1 研究背景	75
4.2.2 研究方法	76
4.2.3 结果与讨论	78
4.2.4 小结	84
4.3 本章小结	108
第 5 章 蛋白质运动数据库的构建及其统计分析	111
5.1 研究背景	111
5.2 结果与讨论	112
5.2.1 构建蛋白质运动数据库 D3PM	112
5.2.2 蛋白质运动影响因素统计分析	116
5.2.3 结合位点残基偏好性研究	118
5.2.4 蛋白结构柔性及分子对接准确性研究	119

5.3 本章小结	121
第 6 章 总结与展望	123
参考文献.....	133
附录.....	127
致 谢.....	207
作者简历及攻读学位期间发表的学术论文与研究成果	209

图表目录

图 1.1	REMD 变形方法的代表性交换方式和控温流程图解。	16
表 2.1	SQMPC 方法测试体系及其结合自由能实验数据。	24
图 2.1	SQMPC 计算流程。	32
图 2.2	SQMPC 应用于 ER 体系的动力学分析。	35
表 2.2	SQMPC 与 Amber ff03 电荷模拟得到氢键平均距离比较。	36
表 2.3	不同 SQMPC 电荷的在不同 MD 时长下的平均 RMSD 值。	36
图 2.3	雌激素复合物测试体系的 RMSD 分析。	37
图 2.4	SQMPC 结合 SIE 和 MM/GBSA 方法得到的相关性分析。	38
图 2.5	SQMPC 在 4 个不同体系上的测试。	40
图 2.6	分子骨架差异性分析。	41
图 2.7	MD 模拟时长对相关性的影响。	42
图 2.8	MD 模拟时间对结合自由能相关性的影响。	43
图 2.9	ASMD-SQMPC 计算示意图。	44
图 2.10	ASMD-SQMPC 应用于 HIV-1 PR。	45
表 2.4	24 个 HIV-1 PR 抑制剂体系的<PMF>max。	46
图 2.11	ASMD-SQMPC 的准确性评估。	47
表 2.5	<PMF>max 的体积校正 (kcal/mol)。	48
图 2.12	含氟抑制剂的<PMF>max 与结合自由能实验值的相关性。	49
图 2.13	配体解离过程中的蛋白质电荷重分配。	50
图 2.14	配体解离时蛋白的 flap 动态变化。	51
表 3.1	ossPTMetaD 和 T-REMD 的模拟体系信息。	58
图 3.1	ossPTMetaD 在 AdK 体系上的采样能力评估。	60
图 3.2	ossPTMetaD 模拟预测的 AdK 构象转变机制。	61
图 3.3	半开放构象的分布比例随模拟时间的变化曲线。	63
图 3.4	ossPTMetaD 应用于 HIV-1 PR 体系的 RMSD 分析。	63
图 3.5	ossPTMetaD 模拟预测的 HIV-1 PR 构象转变。	64
图 3.6	BACE1 整体和 flap 结构的相对闭合构象重原子 RMSD。	65
图 3.7	ossPTMetaD 模拟预测 BACE1 构象变化。	66
图 3.8	BACE1 体系的作用分析。	67
图 3.9	ossPTMetaD 模拟预测 p38 α 构象变化。	69

图 3.10	ossPTMetaD 模拟预测 c-Abl 构象变化。	70
图 3.11	ossPTMetaD 与其他方法的丙氨酸二肽测试。	71
图 3.12	ossPTMetaD 与 REST2-MetaD 的 HIV-1 PR 测试。	72
图 4.1	劳拉替尼结构。	77
图 4.2	劳拉替尼的一维自由能图。	79
图 4.3	代表性低能构象。	79
图 4.4	劳拉替尼的模拟构象及自由能图景。	80
图 4.5	劳拉替尼的 NMR 构象分析。	81
图 4.6	劳拉替尼与溶剂的分子间相互作用展示。	82
图 4.7	劳拉替尼在水溶剂中的构象分布。	82
图 4.8	构象 P1 和构象 P2。	83
图 4.9	根赤壳菌素在 CHCl ₃ 溶剂中的 REMD 模拟。	84
图 4.10	同源模建结构的拉氏图分析。	88
表 4.1	CoV-2-S RBD 与 CoV-S RBD 的计算 ACE2 结合自由能。	91
表 4.2	CoV-2-S 与 CoV-S 三聚体的计算 ACE2 结合自由能。	92
图 4.11	CoV-2-S 和 CoV-S 的 ACE2 相互作用图解。	93
图 4.12	CoV-2-S 与 ACE2 的相互作用。	95
图 4.13	CoV-S 与 ACE2 的相互作用。	96
图 4.14	CoV-2-S 三聚体的构象转变。	97
图 4.15	CoV-2-S 三聚体与 ACE2 的计算结合自由能随 RBD 角度的变化。 ...	98
图 4.16	CoV-2-S 与 ACE2 作用关键残基的能量贡献变化。	99
图 4.17	CoV-S 三聚体与 ACE2 的计算结合自由能随 RBD 角度的变化。	100
图 4.18	CoV-2-S 三聚体构象转变中的残基均方根误差。	101
图 4.19	CoV-2-S 与 CoV-S 三聚体的构象分布。	101
图 4.20	CoV-2-S 三聚体的单体关联运动。	103
图 4.21	CoV-2-S 的潜在结合口袋预测。	104
表 4.3	靶向刺突蛋白阻止病毒侵入宿主的 8 个化合物。	105
图 4.22	8 个化合物针对 5 个预测口袋的分子对接分析。	106
表 4.4	糖基化 CoV-2-S 的计算结合自由能。	105
图 4.23	糖基化分析。	107
图 5.1	口袋残基运动类型。	114
表 5.1	D3PM 数据库总结。	115

图 5.2	D3PM 网站示意图。	116
图 5.3	蛋白运动类型分析。	117
图 5.4	20 种氨基酸在结合口袋和整体结构上的分布比例平均绝对误差。	119
图 5.5	20 种氨基酸在结合口袋和整体结构上的分布比例。	120

第1章 绪论

1.1 引言

现有的药物研发实践表明计算机辅助药物设计可以有效地减少药物研发成本,缩短药物研发的时间周期,已成为药物研发的一个有力工具。如神经氨酸酶抑制剂扎那米韦就是计算机辅助药物设计的成功典范之一^[1]。此外,重要的抗病毒药物奥司他韦的前期研发也使用了大量的计算机辅助药物设计手段^[2]。近年来,随着计算机硬件及药物设计理论的发展,基于结构的药物设计方法(如分子对接等)作为药物设计的重要工具极大地提高了药物先导化合物的发现效率,促进了药物设计的发展^[2-3]。

在基于结构的药物设计中,蛋白质和配体间的相互作用研究扮演着重要的角色,许多新型药物的发现也都来源于新颖的蛋白质和配体相互作用研究。其中,蛋白质-配体的结合自由能(ΔG_{bind})是表征它们相互作用的重要物理量,是评价其结合能力及配体生物活性的重要参数^[4]。因此,快速获得准确的结合自由能对药物设计至关重要。近年来,尽管结合自由能计算理论和新方法日就月将、层出不穷,但都难以真正应用到高通量药物筛选中。这些结合自由能计算方法受到诸多因素的制约,其中一个重要原因是预测结合自由能算法的不准确性,如不能精确描述溶剂化效应和静电极化效应等。目前快速的预测方法主要是基于经验方程,如广泛使用的定量构效关系和分子对接中的打分函数,但这类方法的准确性受限于拟合参数时训练集的选择,其普适性和准确性并不能令人满意。另一个制约结合自由能预测的重要原因是计算资源的高需求性以及计算机的运算能力,如目前结合自由能预测较为准确的自由能微扰和热力学积分等方法由于采样量的需求而需要消耗大量的计算资源。因此发展一种高效准确的蛋白质-配体结合自由能预测新方法对提高药物设计效率具有重要意义。

另一方面,配体或药物分子的结合往往会诱导靶标蛋白的构象变化,从而调节靶标蛋白的生物学功能。因此,研究蛋白质的构象变化对理解药物分子的作用机制,以及靶标蛋白结构功能之间的关系都具有重要意义,也能为药物分子的改造和优化提供不可或缺的帮助。例如重要药物靶标蛋白表皮生长因子受体 EGFR 在活化与非活化状态下具有截然不同的活化环构象^[5],而目前 EGFR 的一代和二

代抑制剂就分别作用于这两种不同的构象。另外，靶向蛋白质的特异性构象也是一些药物分子的高选择性来源^[6-10]。因此，研究蛋白质的构象变化对药物发现与设计至关重要。然而，现有的实验技术，如 X 射线晶体衍射、核磁共振等，难以捕捉蛋白质的动态行为。在计算机模拟方法中，基于牛顿力学原理的分子动力学（molecular dynamics, MD）模拟，能够从原子水平上观察蛋白质的动态变化，从而成为研究蛋白质构象变化的通用方法^[11-12]。近年来随着 MD 算法和计算机硬件的发展，MD 模拟尺度能够达到几十微秒甚至毫秒级别，但仍然小于蛋白质重要构象间转变的时间尺度（一般为毫秒到秒级别）。而且常规 MD 模拟方法采样效率低，易陷入局部极小值，因此极大地限制了其在大规模蛋白质构象变化研究中的广泛应用。

为进一步提高 MD 的采样效率，发展增强型采样算法方兴未艾，前赴后继的研究投入其中，各种增强型采样算法因而百花齐放。依据是否需要在集合变量（collective variables, CVs，也称为反应坐标）上施加限制势，这些增强型采样算法可归为两类：依赖反应坐标的限制性增强采样算法和不依赖反应坐标的非限制性增强采样算法。限制性增强采样算法是在预定的 CVs 上如原子距离、二面角、均方根偏差（root mean square deviation, RMSD）等对模拟体系施加额外势能，从而加强在 CVs 空间内的采样，典型方法有伞形采样（umbrella sampling, US）、拉伸分子动力学（steer MD）、靶向分子动力学（target MD）、metadynamics (MTD) 等。这类方法需要具备对研究体系的初步了解以满足 CVs 的预先设定，而对于复杂的蛋白质构象变化以及配体分子的结合、解离过程，我们较难选择恰当的 CVs。此外，对于势能面崎岖的模拟体系，反应能垒很可能不存在选定的 CVs 空间内，而存在其他 CVs 空间，即出现隐藏能垒。因此，反应坐标的恰当与否直接决定了此类方法的采样效果，使其较难应用于复杂的蛋白质构象变化。而非限制性增强采样算法基于一些广义的物理量（如温度和体系势能等）作为增强型采样的参数，并不依赖 CVs，因此能够避免 CVs 选取困难的问题。典型方法有模拟回火和副本交换分子动力学（replica exchange MD, REMD）等，其中 REMD 自 1999 年应用于甲硫啡肽链以来，目前广泛用于生物大分子的构象采样研究，如腺苷酸激酶，GPCR 等蛋白在溶液中的构象变化^[13-15]。

综上所述，靶标-药物体系的结合自由能计算和构象采样是目前计算机辅助药物设计面临的两大重要挑战，也是药物设计领域亟待解决的难题。所幸，科研

工作者们持之以恒的努力为攻克这两大难题做出了巨大贡献，正不断改善靶标-药物体系结合自由能计算和构象采样的理论和方法。在结合自由能计算上，高精度方法的效率提高和可极化力场的开发等为高通量药物虚拟筛选提供有力工具；在蛋白质构象采样上，各种增强型采样算法层出不穷，其中 REMD 方法在其形态和算法上得到了充分改进。

1.2 蛋白质-配体结合自由能的准确预测

蛋白质-配体相互作用的准确预测对于理解生物体中大多数生物学过程至关重要^[16]，同时设计具有选择性的蛋白质-配体相互作用是细胞和分子机制研究的有力工具，也是新药发现与设计的有效策略^[16]。在新药发现与设计过程中，蛋白质-配体相互作用的研究通常关注三个方面，首先通过结构建模或氨基酸突变等方式发现关键作用残基，即热点（hot-spot）；第二是通过虚拟筛选等方式发现活性化合物；并最终希望了解蛋白质-配体结合的结构信息和具体机制，其中蛋白质-配体结合自由能是一个重要参数^[17-18]。

Guvench 和 MacKerell Jr.将目前的蛋白质-配体结合自由能预测方法分为两类^[19]。其中第一类方法原理严格，能够再现实验上的绝对结合自由能，但这类方法需要耗费大量的计算资源，难以进行高通量虚拟筛选。第二类方法计算速度快，基于建模将小分子与蛋白质对接在一起，可以在几天内筛选数千到数百万种化合物，但这类方法基于一些假设的极端简化，在某种程度上降低了预测准确性^[20-22]。蛋白质-配体的结合自由能信息不仅对揭示真正的结合机制非常重要^[23]，对药物设计也是非常重要的^[24-25]。因此，如果有一个快速的方法能够准确预测蛋白质-配体的结合自由能，包括对柔性配体的构象搜索，那么蛋白-配体相互作用研究关注的上述三个方面都可以得到解决。

在实际计算中，使用常规分子动力学模拟对配体进行构象采样从而计算结合自由能，需要消耗大量的计算成本，尤其是对于具有许多构象自由度的蛋白质体系。另外预测配体的结合模式也是一项重要研究问题，对于两个具有相似三维结构的配体，它们的结合模式也可能不同。在这种情况下，如果基于结合模式相似的假设，那么预测的结合自由能与精确值会相差很大。

1.2.1 结合自由能计算方法

自由能微扰 (free energy perturbation, FEP) [26] 是目前应用最广泛、最准确的结合自由能计算方法。根据 Zwanzig 提出的公式 [27], 状态 X 和状态 Y 的自由能差异 ΔA 由式 1.1 表示。

$$\Delta A = k_B T \ln \langle \exp[-\frac{H_X - H_Y}{k_B T}] \rangle_Y \quad \dots (1.1)$$

其中, H_X 和 H_Y 分别表示状态 X 和 Y 的哈密顿量。 k_B 是玻尔兹曼常数。 $\langle \rangle_Y$ 表示状态 Y 的系综平均。在 FEP 原理中, 通过引入 λ 参数将状态 X 转变成状态 Y 过程中的哈密顿量变化分解成多个差异非常小的瞬时状态的叠加, 因此总结合自由能可计算为多个瞬时状态的自由能差总和。

$$H_\lambda = \lambda H_X + (1-\lambda) H_Y \quad \dots (1.2)$$

自 Singh 等人利用分子动力学模拟方法, 首次将自由能微扰理论应用于计算自由能变化 [28], 目前 FEP 已被广泛应用。FEP 对每一个瞬时哈密顿量都需要消耗大量计算资源去充分采样, 因此属于第一类方法。以 FKBP (FK506 结合蛋白) 体系为例, FEP 计算的结合自由能值与实验估算值的相关系数为 0.96, 平均绝对误差仅为 0.40 kcal/mol [29]。因此, FEP 能够获得非常精确的自由能结果, 但它需要长时间分子动力学计算 [29]。

热力学积分法 (thermodynamics integration, TI) 也是一种目前广泛应用的结合能计算方法 (公式 1.3), 属于第一类方法。FEP 和 TI 都适合于预测准确的绝对和相对结合自由能, 但不适用于筛选大量的化合物。这两种方法也被称为炼金术自由能计算方法 [30]。

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \langle \frac{\partial H}{\partial \lambda} \rangle_\lambda d\lambda \quad \dots (1.3)$$

另一种计算结合自由能的方法是线性相互作用能 (LIE) 方法。这种方法是基于式 1.1 的级数展开 [31]。在 LIE 中, H_X 和 H_Y 分别表示配体的自由状态和结合状态。蛋白质-配体结合自由能根据公式 1.4 计算。其中, $\langle V_{vdW} \rangle_{lig-pro}$ 和 $\langle V_{elec} \rangle_{lig-pro}$ 分别表示配体在结合状态时的范德华作用和静电作用系综平均; $\langle V_{vdW} \rangle_{lig-sol}$ 和 $\langle V_{elec} \rangle_{lig-sol}$ 分别表示配体在溶液自由状态时的范德华作用和静电作用系综平均。 α , β 和 γ 是实际操作中需要仔细调整的参数。在 LIE 中, 只需要对配体与蛋白质的结合状态和配体在溶剂中自由状态进行系综平均计算, 并不需要计算配体结合过程的瞬时状态自由能。并且, LIE 可以获得较为准确的结合自由能计算结果。Guvench 和 MacKerell Jr. [19] 认为, 这种方法介于第一类和第二类之间。然而, LIE

的预测准确性取决于确定参数时采用的训练数据集^[32]。以 FKBP 体系为例, LIE 方法给出的计算值与实验值的相关系数为 0.81, 与实验值的平均绝对误差为 0.44 kcal/mol^[33]。因此, LIE 方法计算结果良好, 计算成本较低, 但需要对公式 1.4 中的三个参数进行仔细的调整。

$$\Delta G_{bind} = \alpha(\langle V_{vdW} \rangle_{lig-pro} - \langle V_{vdW} \rangle_{lig-sol}) + \beta(\langle V_{elec} \rangle_{lig-pro} - \langle V_{elec} \rangle_{lig-sol}) + \gamma \quad \dots (1.4)$$

1.2.2 溶剂效应

溶剂效应对蛋白质-配体结合自由能的准确计算具有重要作用。通常情况下在模拟中采用显式溶剂分子(水分子)是最简单的处理方法, 但需要消耗大量的计算资源。因此, 显式溶剂模拟并不适用于高通量虚拟筛选。

连续介质溶剂模型被发展用来解决溶剂效应。经典方法包括 MM/GBSA 和 MM/PBSA (molecular mechanics/Poisson-Boltzmann (or generalized Born) surface area) 方法。在 MM/GBSA 方法中, 根据分子力学势能函数计算原子间相互作用, 基于广义泊松-玻耳兹曼方程或玻耳兹曼方程计算极性溶剂化自由能。如公式 1.5 所示, 非极性部分 (G_{SA}) 通过溶剂可及表面积来计算, 其中 $SASA$ 表示溶剂可及表面积, β 和 γ 都是需要调整的参数。

$$G_{SA} = \beta \cdot SASA + \gamma \quad \dots (1.5)$$

定义分子表面是 MM/GBSA 和 MM/PBSA 方法的关键之处。定义分子表面的简单方法是分子中原子范德华球的重叠表面, 基于溶剂可及性思想, 通过考虑一个沿着分子的范德华表面滚动的球的中心所绘制的表面来计算 $SASA$ 。这一想法最初是由 Lee 和 Richards^[34]、Richerd^[35]提出的, 一些计算方法也由多位作者提出来了^[36-38]。MM/GBSA 方法应用于 FKBP 体系的计算结果与实验结果的相关系数为 0.81, 平均绝对误差为 3.89 kcal/mol^[33]。因此 MM/GBSA 方法的相关系数与 LIE 方法相似, 但在 FKBP 体系^[33]中的平均绝对误差比 LIE 更大。

1.3 副本交换分子动力学变形方法的蓬勃发展

1.3.1 基于温度的副本交换分子动力学

温度型 REMD (T-REMD) 也被称为平行回火, 即 N 个平行副本在梯度温度 $T_n (n=1,2,\dots,N)$ 下模拟, 并定期交换副本间的模拟条件, 从而使得每个副本均能反

复游历最高温和最低温。在模拟过程中，体系能量 E 可视为由原子坐标 (x) 决定的体系势能 E_{pot} 和由原子速度 (v) 决定的体系动能 E_{kin} 之和 (公式 1.6)。每隔一段模拟时间，相邻温度副本 (T_a 和 T_b) 之间根据 Metropolis 准则计算交换概率 (P_{ab})，从而保证交换行为服从细致平衡条件以及马尔科夫链蒙特卡罗状态转换条件。其中 $\beta=1/k_B T$ 为温度因子， k_B 为玻尔兹曼常数。

$$E(x, v) = E_{pot}(x) + E_{kin}(v) \quad \dots(1.6)$$

$$P_{ab} = \min(1, \exp(\Delta\beta\Delta E_{pot} + \Delta\beta\Delta E_{kin})) \quad \dots(1.7)$$

为保证模拟副本的原子速度在发生交换后仍然服从 Maxwell 分布，T-REMD 对其动能进行调节 (公式 1.8)，其中 $E_{kin}^{(a \rightarrow b)}$ 为副本 a 交换到副本 b 之后的动能， $E_{kin}^{(b \rightarrow a)}$ 为副本 b 交换到副本 a 之后的动能。

$$\left\{ E_{kin}^{(a \rightarrow b)} = \frac{T_b}{T_a} E_{kin}^{(a)} ; E_{kin}^{(b \rightarrow a)} = \frac{T_a}{T_b} E_{kin}^{(b)} \right\} \quad \dots(1.8)$$

通过体系动能的调整， P_{ab} 中的动能差为零， P_{ab} 只由体系势能决定。

$$P_{ab} = \min(1, \exp(\Delta\beta\Delta E_{pot})) \quad \dots(1.9)$$

在公式 1.9 中， P_{ab} 与模拟温度密切相关，相邻副本间的温度差 ΔT 越小，其交换成功率越高。

在 T-REMD 中，副本最低温度 (T_{min}) 设置为关注体系变化的温度 (通常为室温 300 K)，而最高温度 (T_{max}) 设置为能帮助体系跨过反应能垒的最低温度。在实际应用时，副本交换的平均成功率需要保证在 20% 到 50% 之间以获得较好的采样效果，并且模拟副本需保证能够反复遍历温度梯度。为此，我们需要控制所有相邻副本之间的势能差 (ΔE) 与其自身能量波动 (δE) 的比值都近似一致。在温度 T 下，副本的 δE 与 $\sqrt{f}T$ 成正比，其中 f 为模拟体系总自由度，而 ΔE 与 $f\Delta T$ 成正比。因此， $\Delta E/\delta E$ 的比值正比于 $\Delta T\sqrt{f}/T$ ，即为保证其比值一致，相邻副本间 ΔT 与 $1/\sqrt{f}$ 成正比。当模拟体系 f 增加时， ΔT 需要减小，那么对于需包含特定温度区间 (T_{min} 和 T_{max} 确定) 的 T-REMD 模拟，所需的副本数目相应增多。对于生物大分子体系，体系 f 本身很大，再加上显式溶剂，普通计算平台很难承受 T-REMD 的巨大计算成本。因此，减少 T-REMD 计算资源以及进一步提高其采样效率成为 T-REMD 优化完善的持续目标。近年来，通过改进 T-REMD 采样算法以及与其他增强型采样算法整合，一系列 REMD 新变形方法陆续显露头角，在生物大分子构象采样中展现其高效性。

1.3.2 改进交换方式和提高交换成功率

在常规 T-REMD 中, 副本交换只考虑相邻副本, 且奇次交换只计算奇副本对, 偶次交换只计算偶副本对。例如, 对于含 $2N$ 个模拟副本的 T-REMD, 奇次交换计算相邻副本 1 和 2, 3 和 4, ..., $2N-1$ 和 $2N$ 之间的交换概率, 偶次交换计算相邻副本 2 和 3, 4 和 5, ..., $2N-2$ 和 $2N-1$, 及 $2N$ 和 1 之间的交换概率。当判断可以发生交换时, 相邻副本交换原子坐标或模拟条件, 从而促进模拟体系跨越反应能垒。因此, 通过提高交换成功率可以改善 T-REMD 的采样效率, 这也是近年来 T-REMD 优化的一个重要方向。例如, 通过速度调节的方式 Kouza 等人实现了零拒绝率的副本交换, 尽管测试发现这种强制交换的方法并未显著提高其采样效率^[39]。Suwa 和 Todo 提出的 Suwa-Todo 算法在求解细致平衡条件时引入了一种权重分配图形化方法, 能够使交换拒绝率最小化并服从全局平衡条件, 但 Suwa-Todo 算法的优势在常规 T-REMD 的有限次副本交换中难以充分体现, 而在副本扰动法 (Replica-permutation method, RPM) 的全局副本交换方式中得以体现^[40-41]。在 RPM 中, 对于含 N 个平行副本的模拟, 每次交换后的副本状态有 $N!-1$ 种可能, 这种方式显著提高了副本交换成功率, 因此 RPM 策略也逐渐被引入至其他 REMD 变形方法中^[42]。然而利弊相生, 当 RPM 副本数目增加时, 计算交换概率的时间成本就会显著增加。因此, 为了减少副本交换组合数且不对采样效率造成明显损失, Okumura 等人发展了副本子扰动法 (replica sub-permutation method, RSPM)^[43], 即只考虑副本与前后相邻副本之间的交换, 从而节省了 RPM 在计算交换概率上的时间成本。

此外, 增加副本交换频率也是提高 T-REMD 采样效率的一种方式^[44]。例如, 吉布斯采样副本交换 (Gibbs sampling replica exchange)^[45-46]方法能够实现副本间全局交换; 通过改进 Metropolis 交换概率算法, Brenner 等人提出了基于 Calvo 的全副本对交换策略 (Calvo's all-pairs exchange), 从而提高采样效率^[47]; 多尺度无限交换 REMD (Multiscale implementation of infinite-swap REMD)^[48]通过异构多尺度方法增加交换频率, 实现所谓的无限交换极限; 而另一种无限交换 REMD (infinite swapping REMD)^{[49]-[50]}方法是通过无限交换极限分析求解极限方程来增加交换频率。然而, 由于极限方程需要对副本的所有排列组合计算系数, 计算成本增加, 收益有时很小, 还可能提高交换拒绝率, 因此在实践应用中需要对极限方程计算进行近似。Urano 等人提出的 designed-walk 副本交换方法

(DEWREM)^[51], 执行交换遵循奇-偶顺序, 即先进行全奇相邻副本交换 (1 和 2, 3 和 4, ..., $2N-1$ 和 $2N$); 当副本对成功交换后, 这两个副本都停止交换, 直到所有奇副本对均完成一次交换, 之后以同种方式进行全偶相邻副本交换, 反复循环。这种方法强制周期性地使副本在高温和低温之间游走。总体而言, 由于不同模拟温度下的副本势能差较大, 因此, 尽管各种通过改变副本交换方式的算法能适当提高交换频率, 但其交换效率并没有显著提高。

1.3.3 缩减执行交换时的体系总势能以减少副本数目

由于副本数目与体系 f 的平方根成正比, 如跨膜蛋白 GPCR 等大蛋白质体系在显式溶剂模拟中需要使用大量的副本数目。因此, 缩减副本数目是当下 REMD 优化的迫切需要和改进目标。

生物大分子模拟体系中的溶剂原子占比往往超过 60%, 这大大增加了模拟体系的总自由度, 导致副本数目剧增。因此, 处理显式溶剂以减少副本数目是优化 REMD 的重要方向之一。隐式溶剂模型 (包括 PB 和 GB 模型) 相比显式溶剂模型计算更加简单快速, 也更能较精确反映体系溶剂化效应。因此, 隐式溶剂模型在早期分子动力学模拟中得到了广泛应用, 但是隐式溶剂模型模拟得到的自由能面与显式溶剂模型的模拟结果有一定偏差。因此, 如何在 REMD 模拟中使用隐式溶剂的同时又能得到较准确的模拟结果是 REMD 进一步优化需要解决的问题。

因为显式溶剂的使用会使体系 f 大幅增加, 许多 REMD 优化方法采用显隐式混合模拟 (hybrid REMD), 即采用显式溶剂模拟体系, 但在副本交换时使用隐式溶剂或混合溶剂模型计算能量, 从而减少模拟副本数目。由于这类方法中体系仍然在真实的显式溶剂中进行模拟, 因此交换时隐式溶剂模型对模拟结果的影响较小。例如 kernel REMD (kREMD)^[52] 方法通过显式溶剂进行分子采样, 而使用连续溶剂化模型 (GB) 获得的能量作为评分函数将构象进行排序, 并使用 kernel 交换概率代替 Metropolis 概率。目前 hybrid REMD 方法中可以使用两种不同的连续溶剂化模型 (PB 和 GB 模型)。例如 Mu 等人使用 PB 隐式溶剂模型对溶质的溶剂化能进行计算^[53], 而 Chaudhury 等人利用 GBMV2 隐式溶剂模型使得 PB 计算量进一步减少^[54]。另外对溶剂化能进行一定的调整, 也可以继续降低溶剂的自由度, 代表方法有 REMshH (replica-exchange molecular dynamics simulation with scaled hybrid Hamiltonians)^[55]。在 REMshH 中, 以调节因子 0.2 调整 GB 模型计

算得到的溶剂化能来减少副本数目。在简单的模拟体系中，这种方法可以得到更准确的结果，但由于隐式水模型不能准确考虑水分子的氢键作用，以及隐式水模型高估了静电作用，此方法在较复杂的模拟体系中的应用受到了极大的限制。基于此方法，人们进一步发展了混合溶剂模型，即在发生交换的时候，将与溶质分子密切相关的少量溶剂分子使用显式水溶剂模型，而外层溶剂使用隐式溶剂 GB 模型，从而考虑显式溶剂的氢键及其他特殊的溶剂效应，提高模拟的准确性^[56]。

hybrid REMD 方法同时采用了两种不同的溶剂模型，即在模拟时采用显式溶剂模型交换时采用隐式溶剂模型，但副本交换前后使用这两种溶剂模型得到的体系能量并没有一一对应起来，因此很可能会使模拟体系获得不正确的系综。为了解决这一问题，我们开发了基于速度优化的显隐式溶剂模型 REMD 方法，通过调节速度的方式（式 1.10 和 1.11，其中 P_{is} 是通过隐式溶剂模型计算的溶剂相关能量）将这两种溶剂模型获得的能量之间的差异相互关联，从而获得准确的模拟结果^[57-58]。为进一步改善隐式溶剂模型的缺陷，我们采用显式的水溶剂化能从而获得更加精确的模拟结果，继而开发了 vsREMD（Velocity-scaling optimized hREMD）方法¹¹。

$$v^{(b)} \leftrightarrow v^{(a)} = v^{(b)} \sqrt{\frac{E_{kin}^{(a)} - \Delta(P_{ww} + P_{pw} + P_{is})}{E_{kin}^{(b)}}} \dots (1.10)$$

$$v^{(a)} \leftrightarrow v^{(b)} = v^{(a)} \sqrt{\frac{E_{kin}^{(b)} - \Delta(P_{ww} + P_{pw} + P_{is})}{E_{kin}^{(a)}}} \dots (1.11)$$

另一方面是处理溶质的自由度。在 PREMD（Partial replica exchange molecular dynamics）方法中^[59]，整个模拟体系被分成关键区域（central group）和周围区域（surrounding group）两个区域，因此体系哈密顿量可分成 3 个部分，即周围域内部能量（ E_{ss} ），周围域-关键域能量（ E_{sc} ）和关键域内部能量（ E_{cc} ）。使用双热浴耦合控温系统，PREMD 只对模拟体系中的关键区域交换模拟温度，而周围区域保持基线温度，因此 PREMD 的交换概率可近似取决于（ $E_{sc} + E_{cc}$ ）。LREMD（Local replica exchange molecular dynamics）则进一步设置周围域共享一套原子坐标，那么 LREMD 的交换概率则完全取决于（ $E_{sc} + E_{cc}$ ）⁸⁰。总而言之，PREMD 与 LREMD 都是基于关键域和周围域之间弱耦合的假设，并且周围域的构象在模拟过程中几乎没有变化。但通常在许多重要的生物过程中这一假设并不成立，因

此这类方法并不多见于大分子生物模拟体系的应用研究。

1.3.4 基于体系哈密顿的 REMD 方法

在常规 REMD 方法中，模拟温度可视为增强采样变量。尽管温度是一个全局变量，且操作简便，但对于温度不敏感型构象变化的增强采样效率极不明显。因此，基于体系哈密顿量的 REMD (H-REMD) 方法^[60]采用特定哈密顿量组分作为增强采样变量，例如在蛋白质折叠及蛋白质-配体结合过程中起重要作用的非键相互作用能等。

通常这类方法是以体系势能为增强采样变量，体系势能又可分解为多项能量组分。因此，在实际应用中可以将不同能量组分作为模拟变量以满足特定模拟需求。根据分子组成，体系势能可分解成三部分，即蛋白质内部作用能 (E_{pp})，蛋白质-溶剂作用能 (E_{pw}) 以及溶剂内部作用能 (E_{ww})。因此，如果我们只对 E_{pp} 和 E_{pw} 进行能量调节（公式 1.12，调节因子 s 相当于有效温度），相当于控制 E_{pp} 和 E_{pw} 部分原子的模拟温度，而将 E_{ww} 保持在基线温度，那么能够有效减少 REMD 所需的副本数目（公式 1.13）。

$$E = s(E_{pp} + E_{pw}) + E_{ww} \quad \dots(1.12)$$

$$P_{ab} = \min(1, \exp(\beta \Delta s \Delta(E_{pp} + E_{pw}))) \quad \dots(1.13)$$

另外在分子力场中，体系势能被分为成键项 E_{bonded} 和非键项 $E_{non-bonded}$ ，通常成键项包括键伸缩项 E_{bonds} ，键角弯曲项 E_{angles} ，二面角扭转项 $E_{dihedrals}$ 以及 impropers 项 $E_{impropers}$ ，非键项由库伦作用 E_{ele} 和范德华作用 E_{vdW} 两部分组成（公式 1.14）。

$$\begin{aligned} E &= E_{bonded} + E_{non-bonded} \\ &= (E_{bonds} + E_{angles} + E_{dihedrals} + E_{impropers}) + (E_{ele} + E_{vdW}) \quad \dots(1.14) \end{aligned}$$

范德华副本交换（van der Waals REM, vWREM）^[61]方法利用范德华半径调节因子的交换来控制模拟体系的范德华相互作用（公式 1.15）。库伦副本交换（Coulomb Replica-Exchange Method, CREM）^[62]则是利用原子电荷调节因子的交换来改变体系静电项在总势能中的占比，从而对体系进行充分采样（公式 1.16）。Itoh 等人通过 CREM 成功预测了二聚化 β -淀粉样蛋白的热稳定构象^[63]。

$$P_{ab} = \min(1, \exp(\beta [E_{vdW}^a(x^a) + E_{vdW}^b(x^b) - E_{vdW}^b(x^a) - E_{vdW}^a(x^b)])) \quad \dots(1.15)$$

$$P_{ab} = \min(1, \exp(\beta [E_{ele}^a(x^a) + E_{ele}^b(x^b) - E_{ele}^b(x^a) - E_{ele}^a(x^b)])) \quad \dots(1.16)$$

在 H-REMD 方法中，成键项能量也能够用作增强采样变量。通常蛋白质的功能性构象变化由原子扭转自由度变化主导，因此，在扭转角空间内的 MD 模拟能采样到最相关的自由度。扭转角副本交换方法（Replica exchange torsion angle MD, REX-TAMD）正是利用扭转角项调节因子的交换来增强采样^[64-65]。此外，Okumura 等人也通过交换 α 螺旋和 β 折叠的骨架二面角调节参数来提高采样效率^[66]。Q-REMD^[67]利用 β 和 q 参数对体系势能进行非线性调节（公式 1.17），其中 q 值从第一个副本的 1 逐步增加，从而通过势能函数的修改来降低反应能垒。

$$U_q(\varepsilon) = E_{bond} + E_{angle} + \frac{1}{\beta(q-1)} \ln\{1 + \beta(q-1) * [E_{dihed} + E_{nonbond} + E_{solv} + \varepsilon]\} \dots (1.17)$$

基于 H-REMD 原理，Berne 等人提出了 REST 方法^[68]，将体系势能根据原子组成分解为三部分，即溶质内部作用能（ E_{pp} ），溶质-溶剂作用能（ E_{pw} ）以及溶剂内部作用能（ E_{ww} ）。目前 REST2 方法受到广泛应用，根据公式 1.18，REST2 对溶质相关势能，即 E_{pp} 和 E_{pw} 进行能量调节，相当于控制溶质原子在温度梯度下模拟（ T_m ），而其余原子保持在基线温度（ T_0 ）。因此，在计算交换概率时 E_{ww} 能量差为零，从而避免了显式溶剂对副本数目的影响（公式 1.19），但这种处理方式呈现热溶质、冷溶剂的模拟状况，某些情况下会得到有偏的构象分布。

$$E = \frac{\beta_m}{\beta_0} E_{pp} + \sqrt{\frac{\beta_m}{\beta_0}} E_{pw} + E_{ww} \dots (1.18)$$

$$P_{ab} = \min(1, \exp\{(\beta_b - \beta_a)(E_{pp}(x^b) - E_{pp}(x^a)) + \sqrt{\beta_0} (\beta_b - \beta_a) (E_{pw}(x^b) - E_{pw}(x^a))\}) \dots (1.19)$$

基于 REST 方法原理，REFT（Replica exchange with flexible tempering）^[69]方法将蛋白质原子分成刚性区域和柔性区域，并把柔性区域定义为溶质部分，刚性区域和溶剂原子定义为溶剂部分。gREST（Generalized REST）方法^[70]根据公式 1.20 进一步将溶质部分定义为势能组成部分，如范德华能量或二面角能量项等。其中 l_i 是第 i 个溶剂-溶质相互作用所涉及的最多原子数，例如以键能、范德华或库伦作用为溶质部分时 i 为 2，以键角为溶质部分时为 3，而以二面角为溶质部分时为 4； k_i 是参与第 i 个溶剂-溶质相互作用的溶质原子数，例如以范德华或库伦作用为溶质部分时为 1。gREST 方法比较适用于对体系关键区域的局部运动模拟。

$$E = \frac{\beta_m}{\beta_0} E_{pp} + \sum_i \left(\frac{\beta_m}{\beta_0} \right)^{k_i/l_i} E_{pw} + E_{ww} \dots (1.20)$$

FREM (Fragment replica-exchange method) [71] 方法将蛋白质分成多个结构片段并分配给不同副本。因此, 在每个副本中, 体系总能量由片段能量和非片段能量组成 (1.21), 其中片段能量通过参数 s 调节, 交换概率根据公式 1.22 计算。MFREM (The multiple fragment-size replica-exchange) 方法 [71] 是在 FREM 方法的基础上同时考虑温度和 s 参数作为增强采样变量。ResEx (Resolution exchange) 方法 [72] 则是在常规 REMD 模拟中插入粗粒度模型副本来帮助体系跨越能垒。

$$E = sE_f + E_{non-f} \dots (1.21)$$

$$P_{ab} = \min(1, \exp\{\beta(s_b - s_a)(E_f(x^b) - E_f(x^a))\}) \dots (1.22)$$

1.3.5 副本交换引入其他增强采样方法

近年来, 许多其他增强型采样算法或蛋白质-配体自由能计算方法也都开始引入副本交换思想。这些新的组合方法通过交换模拟参数或 CVs 数值, 不仅进一步增加了采样效率, 而且也降低了参数设置或 CVs 选择的困难性。MTD 是一种应用广泛的限制性增强采样算法, MTD 对 CVs 周期性施加时间依赖的偏执势 (V) 从而在 CVs 空间内充分采样。Bias-exchange MTD (BE-MTD) [73] 在 MTD 中引入副本交换策略, 进一步提高了 MTD 的采样效率 (公式 1.23)。

$$P_{ab} = \min(1, \exp\{\beta[V^a(x^a) + V^b(x^b) - V^a(x^b) - V^b(x^a)]\}) \dots (1.23)$$

本质动力学 (essential dynamics) 是一种在子结构本质集合模式空间内增强采样的方法。Kubitzki 等人在本质动力学中引入副本交换策略 [74], 使得只有关键子结构能耦合到高温, 而其余结构均在基线温度模拟, 从而达到减少副本数的目的。另外, Sugita 等人通过整合 REMD 与 multicanonical 算法 [75-76] 也能够显著减少 REMD 计算资源 [77-78]。

加速分子动力学 (Accelerated Molecular Dynamics, AMD) 通过在体系势能表面增加非负升高势从而帮助体系跨越能垒 (公式 1.24), 其中 E_{cut} 是指预先设置的能量阈值, α 为加速因子。Lopes 等人在 AMD 中也引入副本交换策略, 交换副本能量阈值或 α 参数 [79]。相比 AMD, 高斯加速分子动力学 (Gaussian accelerated Molecular Dynamics, GaMD) 通过对模拟体系施加高斯分布偏执势以帮助体系跨越能垒 (公式 1.25), 其中 k 为简谐力常数, E_{max} 和 E_{min} 分别表示体系势能最大值和最小值。rex-GaMD [80] 是 REMD 与 GaMD 的组合, 目前有两种形

式:能量阈值 rex-GaMD 和力常数 rex-GaMD。其中能量阈值 rex-GaMD 交换副本的能量阈值参数, 而力常数 rex-GaMD 交换副本的简谐力常数。Oshima 等人进一步将 REMD、GaMD 和 US 组合成 GaREUS 增强采样算法, 副本交换同时考虑体系势能以及 GaMD 和 US 施加的偏执势 (公式 1.26), 其中 i 是指副本 i ^[81]。

$$\Delta V(x, E_{cut}, \alpha) = \frac{(E_{cut} - E(x))^2}{\alpha + (E_{cut} - E(x))} \quad \dots (1.24)$$

$$\Delta V_x = \frac{1}{2} k \frac{1}{E_{max} - E_{min}} (E_{cut} - V_x)^2 \quad \dots (1.25)$$

$$\Delta V(X) = \Delta V_i^{US}(x) + \Delta V_i^{GaMD}(x) \quad \dots (1.26)$$

伞形采样是在预先设置的采样窗口内施加简谐势能以实现窗口内的充分采样 (公式 1.27), 其中 ξ_i 是反应坐标。早在 20 年前, Sugita 等人就已经将 REMD 与 US 组合提出了 REUS 增强采样算法^[82]。

$$\Delta V_i^{US}(x) = k_i (\xi(x) - \xi_i)^2 \quad \dots (1.27)$$

近年来, 自由能计算方法也开始引入副本交换策略。包络分布采样 (Enveloping distribution sampling, EDS) 方法可以在单条模拟中同时计算多个终点态之间的转变能垒, 其中参考态 R 包括 $N(>=2)$ 种哈密顿终点态。根据公式 1.28, 将 EDS 与 REMD 组合可进一步提高 EDS 的准确性, 其中 s 表示势能平滑参数, E_R 表示能量偏移^[83]。

$$V_R(x, s, E_R) = -(\beta s)^{-1} \ln \left(\sum_{i=1}^N e^{-\beta s (V_i(x) - E_{(R,i)})} \right) \quad \dots (1.28)$$

BP-REXMS λ D (biasing potential replica exchange multisite λ -dynamics) 方法^[84]是 REMD 与 MS λ D 的整合, MS λ D 也是一种可以在单条模拟中同时计算多个化合物之间的相对结合自由能。根据公式 1.29, BP-REXMS λ D 对柔性基团进行扰动来计算相对结合自由能, 其中 X 和 x_i 分别表示配体周围原子坐标和配体 i 的原子坐标, M_{sites} 表示取代基位点 S 总数, N_s 表示某个取代位点上取代基 i 的总数, 此外 V 是添加的偏执势, λ 是扰动参数。

$$E(X, \{x\}, \{\lambda\}) = E(X) + \sum_{S=1}^{M_{sites}} \sum_{i=1}^{N_s} \lambda_{S,i} (E(X, x_{S,i}) - V_{S,i}) + \sum_{S=1}^{M_{sites}-1} \sum_{i=1}^{N_s} \sum_{T=S+1}^{M_{sites}} \sum_{j=1}^{N_T} \lambda_{S,i} \lambda_{T,j} (E(x_{S,i}, x_{T,j})) \quad \dots (1.29)$$

近年来, REMD 也常与其他增强采样方法组成多维 REMD 方法。例如, 将

BE-MTD 方法与 REMD 整合而成的 RSE-MTD 方法^[85-88]可进一步增强 BE-MTD 或 REMD 的采样效率^[89-90]。另外, REMD 与 AMD^[91-93]、REUS^[94-95]和 H-REMD^[96]等方法的组合也常见于生物大分子的构象变化研究。ABMD (Adaptively Biased Umbrella Sampling) 方法通过对模拟体系施加自适应性偏执势, 从而在窗口内充分采样。因此, ABMD 与 REMD 的组合能够在温度和 CVs 二维空间内同时进行增强采样, 其计算交换概率也同时考虑偏执势能 (公式 1.30) ^[97]。

$$P_{ab} = \min(1, \exp\{(\beta_b - \beta_a)(E(x^b) - E(x^a) + \beta_a[V^a(\xi^a) - V^a(\xi^b)] + \beta_b[V^b(\xi^b) - V^b(\xi^a)])\}) \dots (1.30)$$

另外, REMD 与自由能计算方法的整合也可以同时考虑扰动参数和温度两个变量。例如, 上述包络分布采样可以在引入多个平行副本后改变模拟 pH 来加速构象采样^[98-99]。自由能微扰 (Free Energy Perturbation, FEP) 是目前理论严格且计算准确的蛋白质-配体结合自由能算法之一。在 FEP 方法中, 两个配体间的转换被分解成一系列微小变化的窗口 λ (从 0 到 1), 总相对结合自由能是这些微小转变能垒的总和。FEP 与 REMD 方法的组合 (FEP/REST) 进一步提高了 FEP 的采样效率, 近年来受到广泛关注^[100-102]。

值得一提的是, Q-REMD 与 REMD 的组合 (Tq-REM) ^[103]在参数设置上与其他多维 REMD 方法截然不同。通常多维 REMD 方法的多维变量在单个副本均能发生交换, 温度梯度覆盖所有副本。但是 Tq-REM 方法只将部分高温副本设置为 Q-REMD, 即温度梯度只涉及副本 1 到 $N-M$ ($M \geq 1$), 这些副本的 q 参数均相同, 而副本 $N-M$ 到 N 的温度相同但 q 值不一样。另外, 为了降低多维 REMD 方法交换概率计算的复杂性和计算量, Cruzeiro 等人提出的多维副本交换方法 (Multidimensional Replica Exchange Simulations) ^[104]尽管同时采用温度、pH 和氧化还原势 (E_v) 多维变量, 但在计算交换概率时与众不同, 并不是同时考虑这三个变量, 而是每次交换只考虑一维变量, 并依次循环。例如, 进行模拟 pH 交换时利用公式 1.31, 其中 $N_a^{H^+}$ 代表副本质子数; 进行氧化还原势交换时利用公式 1.32, 其中 F 代表法拉第常数, E_{va} 代表副本 a 的氧化还原电势, 而 $N_a^{e^-}$ 代表副本 a 的电子数; 进行温度交换时利用计算公式 1.7。

$$P_{ab} = \min(1, \exp\{\ln 10(pH_a - pH_b)(N_a^{H^+} - N_b^{H^+})\}) \dots (1.31)$$

$$P_{ab} = \min(1, \exp\{\frac{F}{k_b T}(E_{va} - E_{vb})(N_a^{e^-} - N_b^{e^-})\}) \dots (1.32)$$

1.3.6 引入模拟过程物理量为增强采样变量

除了对体系势能直接相关变量增强采样, 许多 MD 模拟涉及的物理量也都与其采样效率有关, 可以与模拟温度整成多维 REMD 方法, 从而提高 REMD 采样效率, 并且通常这些物理量不影响体系势能公式表达, 因此不需要对 REMD 源代码进行大量修改。例如, V-REMD^[105]方法可以在常规 REMD 模拟方法中以溶剂粘度作为增强采样变量。通常溶剂粘度越低, 蛋白质链的扩散速度就越快, 从而通过改变溶剂粘度可以提高采样速率。表面张力是一种存在于油/水、真空/水和膜/水等液体界面上的作用力。表面张力 REMD^[106]方法在 NP γ T 系综中引入表面张力作为增强采样变量, 从而进行生物膜系统的高效模拟。另外, 考虑到在常规 REMD 的高温副本中显式脂双分子层难以维持结构的整齐性, MSREMD^[107]方法根据副本的模拟温度调整粒子质量, 即副本温度越高, 粒子越重, 从而提高 MD 模拟的稳定性。而且这种方式还避免了副本交换后的速度调整问题, 从而减少了 REMD 的计算成本。体积-温度 REMD (volume-temperature REMD, VTREMD)^[108]交换副本的温度和密度, 比较适合模拟具有大量局部极小值的构象变化。VTREMD 的交换概率计算公式中还考虑了势能受到模拟体积 V 的影响 (公式 1.33), 其中 P 为体系压强, A 为模拟盒子中的分子数。

$$E_{a,b} \approx E_{a,a} - (P_a - \frac{A}{V_a \beta_a})(V_b - V_a) \dots (1.33)$$

1.3.7 调整模拟流程提高采样效率

常规 REMD 在一个温度梯度下进行副本模拟, 不同副本的能量分布决定其交换成功率。但 TIGER (Temperature intervals with global energy reassignment)^[109]都是在同一基线温度 T_0 下交换所有副本, 因此能够明显减少副本数目。TIGER 的副本交换包括三个循环步骤: (1) 各副本模拟温度从基线温度 (T_0) 加热至副本温度 (T_m), 并以 $\sqrt{T_m/T_0}$ 调节副本内的原子动量以及进行体系平衡, (2) 在 T_m 恒温采样, (3) 各副本通过能量优化被淬火至基线温度 T_0 , 并经过短时间的平衡, 再执行交换。基于一定假设 TIGER 才满足细致平衡条件, 即淬火消耗的热能不依赖体系状态以及淬火时体系构象不变 (如二面角等), 但在复杂体系中这些假设是不成立的, 如在隐式溶剂模型中的多肽折叠模拟。TIGER2^[110]方法对循环步骤 3 进行了优化, 即副本内的原子的动量通过 $\sqrt{T_0/T_m}$ 调节到 T_0 温度, 以满足弱平衡

条件。而且,近年来 TIGER2 方法不仅解决了由于溶剂能量波动大导致的溶质能量采样不敏感的问题(TIGER2A)^[111],而且通过显隐式混合溶剂减少计算资源(TIGER2H)^[112]以及通过增加显式溶剂壳层来考虑反应场能量、溶剂极化作用以及空间效应 (TIGER2HS)^[113]。在 TIGER3^[114]方法中,原子的范德华半径在高温副本中被减小,并于淬火前恢复原子半径至正常值再重新平衡,以消除链纠缠问题。

自适应调整副本温度也能够实现采样效率的提高。例如,通过模拟时温度重置的自适应 T-REMD (Adaptive Temperature-Based Replica Exchange)^[115],能使模拟达到最大温度范围 (CM-REMD) 和最高平均交换率 (EE-REMD)^[116]。另外借助自适应的权重因子扩大模拟温度范围进行采样的统计温度分子动力学模拟(statistical temperature molecular dynamics, STMD)^[117]能够显著减少 REMD 的副本数目。

Modified REMD 1^[118]方法将模拟副本分为活跃副本和非活跃副本,前者与常规 REMD 副本无异,但后者被冻结以维持体系在该温度下的结构特性。增加非活跃副本的方式能够增加温度设置范围,从而减小温度间隔,但相应也增加了 REMD 的计算成本。Modified REMD 2^[119]方法通过将活跃副本的体系势能记录在势能数据库中,从而在副本交换后从相应温度的势能数据库中随机赋予非活跃副本一个势能,从而避免模拟非活跃副本以节省计算成本。

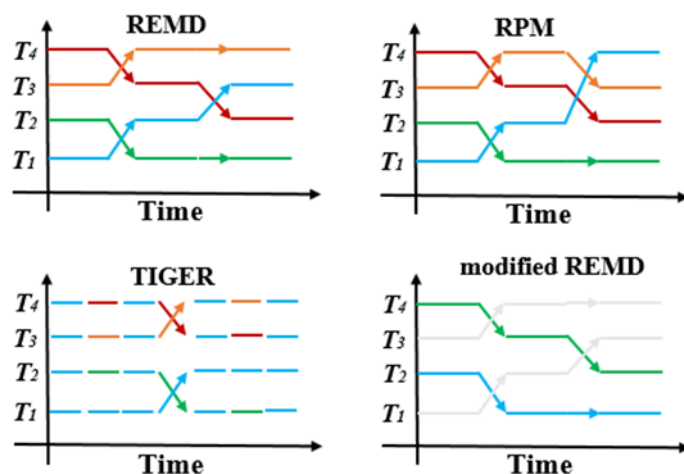


图 1.1 REMD 变形方法的代表性交换方式和控温流程图解。

Figure 1.1 The diagrams of representative REMD variations.

除了模拟温度不同,常规 REMD 模拟中平行副本的其余设置如控温控压方

法等均一致。而 Reservoir REMD^[120-122]将最高温副本设置为一个预先产生的构象库，其他副本正常运行，其优势在于最高温副本与相邻副本的交换能给相邻副本引入空间分布距离较远的构象，从而更好地帮助其跨越能垒，扩大采样范围，提高采样效率。为了提高交换成功率，REDS (Replica exchange with dynamical scaling)^[123]方法在副本中设置一些能够动态调节体系势能的副本，从而提高副本交换成功率。Convective REMD 方法则是在副本中设置一个逐步升温或降温的副本从而加速采样，另外在 REMD 模拟时，可以对一种感兴趣的蛋白质构象的骨架原子进行约束以保持其在一个空间势能阱中，约束力随着副本从完全强度缩减到零强度，从而使得最高温副本成为构象非特异性参考状态^[124]。

1.3.8 采样方法

在计算模拟中，各种采样方法各有千秋。上述中吉布斯采样副本交换方法通过吉布斯采样能够实现全局副本交换。与遵从吉布斯-玻尔兹曼权重的常规 MD 采样相比，Tsallis-REMD (combination of Tsallis weight sampling and REMD)^[125]采样遵从 Tsallis 权重，从而加速采样的收敛性。RXSGLD (Replica exchanging self-guided Langevin dynamics)^[126-127]借助自引导郎之万动力学优势，即不需要通过高温来加速构象运动，避免了过大的构象分布扰动，从而提高了副本交换成功率。QM/MM-REMD (Quantum Mechanical/Molecular Mechanics-REMD)^[128]方法借助 QM/MM 优势，采用 QM 精度处理模拟体系关键区域，适用于生物大分子体系的复杂化学过程。BOMD (Born-Oppenheimer MD)^[129]、SE-BOREMD^[130]以及 CPMD (Car-Parrinello MD)^[131]等采样方法与 REMD 的组合能适用于化学反应的计算模拟。而对于生物大分子体系的大规模构象变化，蒙特卡罗副本交换^[132-133]和 ICMD (Internal Coordinate MD)^[134]具有较好的采样效率优势。

1.3.9 适应分布和异构式计算环境解决资源需求问题

分布或异构式计算环境组成的网格计算网络以及同构高性能集群能够解决常规 REMD 的巨大模拟需求。greedy REMD (gREMD)^[135]方法通过预先设置的副本交换表来分散副本管理，从而使其适合于异构的计算网格。Platanina 等人提出的基于 Hadoop 的副本交换方法 (Hadoop-based replica exchange, HaRE)^[136]，能够在不同配置和大小的同构或异构分布式网络基础设施 (DCI) 上高效模拟。异步副本交换 (REMD simulations asynchronously, ASyncRE)^[137]是基于将副本

分配给正在运行或等待的计算列表，并允许等待列表中的一个副本子集独立于正在运行的列表中的其他副本进行交换。由于副本交换并不依赖于集中的同步步骤，因此 ASyncRE 方法能够扩展至任意数目处理器，并且避免维护静态处理器池。

1.3.10 小结

增强型采样算法通过各种方式克服常规分子动力学模拟的低采样效率问题，目前已成为蛋白质功能性构象变化研究的主流技术。REMD 是一种通过高温来提高 MD 采样效率的方法，由于其采样高效且原理简单，REMD 不仅广泛应用于研究蛋白质的构象变化，而且其算法也不断被改进。在常规 REMD 方法中，副本所需数目与体系总自由度的平方根成正比，因此，近年来 REMD 的改进方向主要在于减少其副本数目。其中，处理显式溶剂以及哈密顿副本交换是目前减少 REMD 计算资源的主流方向。此外，REMD 也可以借助其他增强型采样算法的优势，组合成多维 REMD 方法，这种方式不仅提高了 REMD 的采样效率，而且扩宽了 REMD 的应用范畴，还降低了选择集合变量的困难性。在 REMD 方法的继续改进中，除了考虑进一步减少计算资源和提高采样效率，我们需要考虑如何正确处理蛋白质在大规模构象变化时所发生的体系状态变化，例如质子化状态、二硫键断裂与生成等。总之，尽管各种 REMD 变形方法从各方面提高了 REMD 的采样效率，也丰富了实际应用的方法选择，但目前并没有一个完美的方法对实际模拟“放之四海而皆准”，因此在实际模拟时，我们需要兼顾采样效率和计算成本，选择适合研究体系的方法。

1.4 论文总体安排

本研究工作紧密围绕高效准确模拟靶标蛋白大规模构象变化及其与配体相互作用这一重要科学问题开展工作。在第二章中，为了使模拟研究蛋白质与配体相互作用时能精确考虑静电极化效应，我们通过半经验量子化学方法快速准确计算蛋白-配体体系原子电荷，从而发展了 SQMPC 方法。我们将 SQMPC 与结合自由能计算方法 MM/GBSA 和 SIE 相结合，并将其应用于 50 个靶标蛋白-配体体系。结果发现相比常规 Amber ff03 力场电荷，SQMPC 可以显著提高蛋白质-配体结合自由能计算值与实验值的相关性。为进一步提高配体解离时的模拟准确性，基于 SQMPC 方法，我们在自适应拉伸分子动力学 (ASMD) 模拟中定期更新蛋

白-配体复合物体系原子电荷, 发展了 ASMD-SQMPC 方法, 并将其应用于 24 个 HIV-1 蛋白酶与抑制剂复合物体系。结果发现随着 HIV-1 蛋白酶抑制剂的解离, 蛋白质本身的原子电荷会重新分配, 而 SQMPC 电荷比常规 Amber ff03 力场电荷能够更准确地体现这种电荷变化。并且, 我们发现 ASMD-SQMPC 得到的平均力势 (PMF) 与蛋白质-配体结合自由能具有良好的相关性 ($R^2=0.86$)。

配体的结合往往会诱导靶标蛋白质大规模的构象变化, 在有限的计算机资源下, SQMPC 方法中采用的常规 MD 模拟由于受到其低采样效率的限制, 难以对这类构象变化进行充分采样, 从而难以研究蛋白质构象间转变的自由能变化。因此, 为进一步研究蛋白质大规模的构象变化及其与配体的相互作用, 我的第三章研究工作主要致力于发展高效生物大分子构象采样新方法。基于副本交换分子动力学 (REMD), 我们发展了一种基于能量和几何二维空间同时加速的增强型采样方法, 称之为 ossPTMetaD, 并将其应用到 5 种大小及功能各异的靶标蛋白大规模构象变化的研究中。结果表明 ossPTMetaD 方法能够成功模拟不同运动类型的蛋白质模型体系, 并且能成功应用于极具挑战性的激酶 DFG 翻转运动, 而 ossPTMetaD 方法所需计算资源约为传统 REMD 方法的 30.0%。

本研究另一部分工作集中于综合各种计算模拟手段以及我们自主发展的新方法应用于靶标蛋白构象变化及结合自由能计算的具体研究问题。首先, 为探究大环化合物的构象多样性与其透膜性、极性等性质的关联机制, 我们通过副本交换方法模拟了大环药物劳拉替尼在不同极性溶剂 (二甲基亚砜 DMSO/水混合溶剂和氯仿) 中的构象分布。在高极性 DMSO/水混合溶剂中, 劳拉替尼与溶剂形成密集的氢键网络, 只稳定呈现一种构象 1, 这种构象与劳拉替尼结合 ALK/ROS1 蛋白的晶体构象高度相似。而在低极性氯仿溶剂中, 除构象 1 外 (占比 80%), 还出现一类采样较少的构象 2 (占比 20%)。与构象 1 相比, 构象 2 的溶剂可及表面积和极性都更小。我们进一步通过 NMR 实验验证了该模拟结果, 结果表明模拟预测的劳拉替尼在不同极性溶剂下的构象分布比例与 NMR 实验数据高度一致, 证实了模拟的可靠性。因此, 我们的模拟结果和 NMR 实验数据都表明劳拉替尼可能通过构象 1 与构象 2 的转变来实现它良好的透膜性。

通过高效副本交换分子动力学模拟我们探索了新冠病毒刺突蛋白的动态构象变化, 并综合多种计算模拟手段分析了刺突蛋白在处于不同构象形态时的宿主 ACE2 结合能力, 解析了刺突蛋白与宿主 ACE2 受体结合的关键作用残基, 并预

测了 5 个新冠病毒刺突蛋白上的可药性结合位点。结果发现随着刺突蛋白‘向上’的角度越大，刺突蛋白与宿主 ACE2 的结合能力也变得越强。另外，相比 SARS-CoV 病毒，新冠病毒刺突蛋白呈现‘向上’构象的比例更少，转变成‘向上’状态的能垒也更高，这一结果能够解释尽管新冠病毒刺突蛋白的 RBD 结构域具有比 SARS-CoV 病毒更强的宿主 ACE2 结合能力，但新冠病毒更不容易从‘向下’转变成‘向上’状态，因此其全长结构的 ACE2 结合能力与 SARS-CoV 病毒相当或者更弱，这一模拟结果与实验报道吻合。

在第五章中，为了从实验数据中总结蛋白质构象变化情况以帮助理解蛋白运动的规律和影响因素，我们基于 PDB 数据库构建了一个蛋白运动数据库 D3PM。我们发现除了蛋白质整体结构的大规模运动，口袋残基的局部运动也同样是蛋白质功能机制研究不可忽视的。但这种局部运动由于涉及氨基酸残基少，且运动幅度较小，往往在整体 RMSD 值上表征不出来。因此，我们针对这类口袋残基的局部运动进行了统计并整理分类。最终，D3PM 数据库中包含了 4 类蛋白质的整体运动和 5 类口袋残基的局部运动。基于 D3PM，我们比较了影响蛋白运动的不同因素，以及统计了 20 种氨基酸残基在结合口袋上的偏好性，以助于可药性结合位点的预测等。

第 2 章 半经验量子化学电荷提高结合自由能计算精度

2.1 研究背景

结合自由能是表征蛋白质与配体相互作用的一个重要物理量,也是计算机辅助药物设计领域的一个研究热点^[4, 138-139]。近年来,新的蛋白-配体结合自由能计算方法层出不穷,但尚未出现令人满意的适合高通量药物虚拟筛选的快速准确方法。目前较为快速的结合自由能预测方法主要基于经验方程,如定量构效关系以及在分子对接程序中广泛应用的打分函数等,但这些方法受限于拟合方程所用的数据集,其准确性并不高^[138],而能提供准确结果的自由能微扰^[140]和热力学积分^[141]等计算方法则需要消耗大量的计算资源,对计算体系也有严格的限制。因此,发展一种高效准确的蛋白质-配体结合自由能预测新方法对提高药物设计效率至关重要。近年来,相比于自由能微扰和热力学积分更为简便的基于分子动力学模拟的线性相互作用能^[142-143]、MM/GB(PB)SA^[144]和 Solvated Interaction Energy(SIE)^[145]等计算方法,由于具有操作方便,计算量更少,且计算结果与实验结果具有较好的相关性等优势而在药物设计实践中得到较好应用。但这些方法也存在一些缺陷,如计算结果与实验结果在数值上有较大偏差,依赖分子力场和模拟轨迹^[146]以及难以考虑静电极化效应^[147]等。

在基于分子动力学模拟计算结合自由能的这类方法中,分子力场的准确性是影响计算结果的关键因素^[146]。大量研究发现,静电极化效应可显著影响结合自由能的计算精度^[148-150]。但是目前常规分子力场,如 CHARMM, AMBER, GROMOS 和 OPLS 等均采用固定点电荷的描述方式,并未有效考虑原子间的静电极化效应,造成其结合自由能的计算精度不高。基于此,可精确考察原子间静电效应的可极化力场近年来得到广泛关注。如 Jiao 等人发现在研究胰蛋白酶与苯甲脒类抑制剂的相互作用时,AMOEBA 可极化力场可以获得更精确的计算结果,而且也发现静电极化作用是不可忽略的^[151]。然而,可极化力场尽管弥补了常规力场的固定点电荷缺陷,但其计算速度慢及难以参数化的缺陷大大限制了其在药物设计中的实际使用^[152-153]。另一方面,Cho 等人发现在分子对接时利用可以充分考虑静电极化效应的 QM/MM(量子力学/分子力学)方法重新计算配体小分子电荷能显著提高对接结果的准确性^[154]。因此,通过 QM 计算获得的原子电

荷，如果能与基于分子动力学模拟的结合自由能计算方法相结合有望提高蛋白-配体结合自由能计算的准确性，并发展成为高通量虚拟筛选的重要工具。但对于生物大分子体系而言，QM/MM 方法获得整个体系的极化电荷相当耗时。

为获得精确的蛋白-配体原子电荷，我们基于“分而治之”的策略发展了基于 QM/MM 的 QMPC 电荷计算方法。在 QMPC 电荷计算方法中，我们将蛋白体系以残基为单位划分为若干个 QM/MM 计算小体系，在雌激素受体-亲和素体系中测试发现，使用 QMPC 电荷的结合自由能计算值与实验结果更相符，而采用常规 Amber ff03 力场则难以区分不同活性的化合物^[155]。在 QMPC 计算中，每个 QM/MM 计算小体系都包括核心量化区、过渡量化区、分子力学区和溶剂电荷区，其中核心量化区采用高精度的 QM 方法计算原子电荷，过渡量化区是与核心量化区相邻的残基原子，同样采用高精度的 QM 方法计算原子电荷，而其它原子为分子力学区并通过低精度的 MM 方法计算原子电荷。为考虑溶剂化效应，使用 DELPHI^[156] 计算蛋白质的溶剂化电荷作为 QM/MM 计算的背景电荷。在每个分块区域中通过 QM/MM 方法迭代计算原子的精确静电势表面，进而拟合得到每个原子的点电荷。由于将生物大分子体系拆分成小的 QM/MM 体系，因此能在有限的资源内完成对整个生物体系的精确的量子化学电荷计算。尽管这种 QM/MM 方法提高了蛋白-配体复合物体系原子电荷的计算效率，但对于高通量药物筛选而言仍然相当耗时，如对于约 300 个残基的蛋白质体系至少 96 核时才能完成一次计算，因此极大地限制了 QMPC 电荷在药物虚拟筛选中的广泛应用。

近年来，许多研究表明半经验量子化学方法在非共价相互作用计算上能够媲美色散校正的密度泛函方法，而且其计算时间相比密度泛函方法能够缩减 3 个数量级^[157-158]。例如，有研究报道利用半经验量子化学 PM6 方法计算小分子原子电荷可以提高对接准确性^[159]。我们之前的研究也发现基于 PM6 的 QM/MM MD 模拟在处理氢键和极性作用方面的能力与 B3LYP 方法相当^[160]。因此，若能使用半经验量子化学方法代替我们之前 QM/MM 计算中相对耗时的 QM 方法(B3LYP/6-31G*)，则可大大减少蛋白-配体复合物原子电荷的计算时间，从而真正应用于高通量药物虚拟筛选中。为此，在本章中我们发展了基于半经验量子化学方法的原子电荷计算方法，称之为 SQMPC 方法。为测试 SQMPC 方法的准确性，我们测试了四种不同的半经验量子化学方法(AM1^[161]、PM3^[162]、PM6^[163]和 PM7^[164])，并将此 SQMPC 电荷与结合自由能计算方法(MM/GBSA 和 SIE 方法)相结合，

应用于 5 个靶标蛋白体系共计 50 个蛋白-配体复合物的结合自由能计算中。结果发现无论是 SIE 还是 MM/GBSA 方法, SQMPC 电荷基本都可以取得比普通 Amber ff03 电荷更准确的结合自由能计算结果, 而这种方法所需计算时间相比之前的 QMPC 电荷计算量显著减少。因此, 本章发展的 SQMPC 方法与基于 MD 的结合自由能计算方法 (MM/GB(PB)SA 和 SIE 方法等) 的结合, 通过精确处理体系极化电荷, 能够明显提高结合自由能预测的准确性。

另一方面, 考虑到 MM/GB(PB)SA 等方法尽管操作简便、计算量少, 然而没有考虑配体结合时的熵变效应, 而近年来发展的非平衡分子动力学方法, 如拉伸分子动力学 (SMD) 和自适应拉伸分子动力学 (ASMD) 等^[165], 在预测蛋白-配体结合自由能时能够综合考虑配体结合时的焓变和熵变效应, 有希望成为一个预测结合自由能的有力工具。相比 SMD, ASMD 通过将配体拉伸路径分成多个区间的方式既减少了模拟所需平行副本数目, 也提高了预测准确性。但 ASMD 等方法同样也受到分子力场的局限性, 即在模拟过程忽略了配体结合或解离过程中蛋白质受配体及溶剂的静电极化效应。因此, 为提高 ASMD 方法的准确性, 我们基于前面发展的 SQMPC 方法, 沿着配体的结合或解离路径, 定期更新 ASMD 模拟中蛋白-配体体系的原子电荷, 进一步发展了 ASMD-SQMPC 方法。我们将其应用于 24 个 HIV-1 蛋白酶抑制剂中发现, SQMPC 电荷更新的 ASMD 方法 (ASMD-SQMPC, $R^2=0.86$) 能显著提高常规 ASMD ($R^2=0.52$) 的 ΔG 预测准确性, 而且 ASMD-SQMPC 的相关性也高于基于 MD 的打分函数 (SIE-SQMPC, $R^2=0.54$)。因此, 基于 SQMPC 方法进一步发展而来的 ASMD-SQMPC 方法充分体现了考虑静电极化效应对蛋白-配体结合自由能预测的重要性, 有希望应用于实际的高通量药物筛选中。

2.2 材料与方法

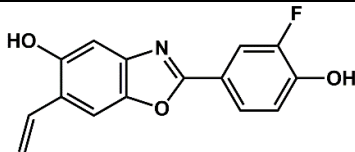
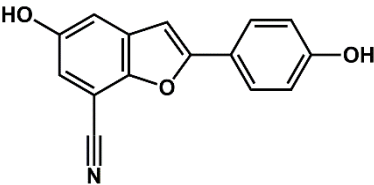
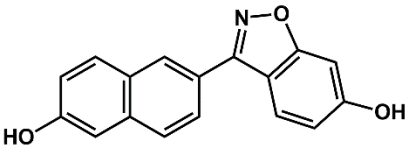
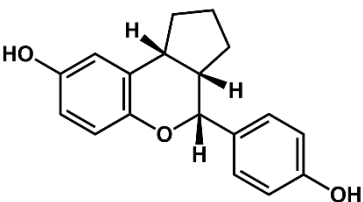
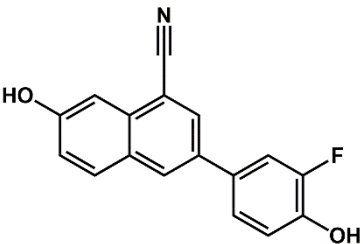
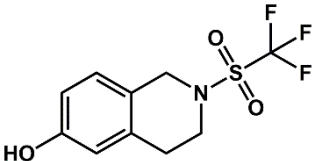
2.2.1 体系准备

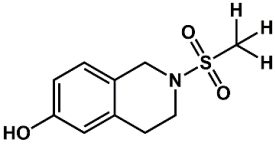
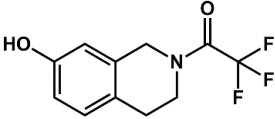
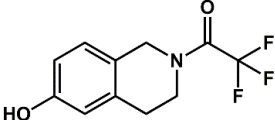
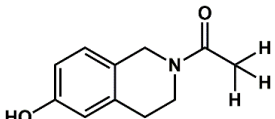
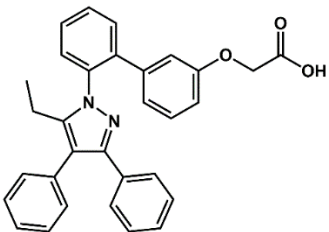
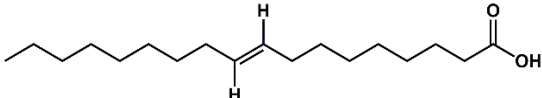
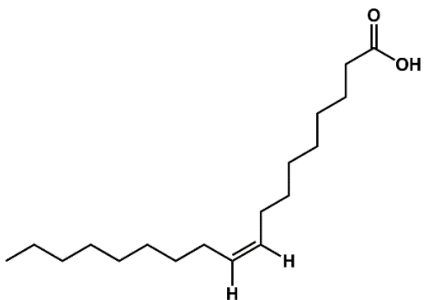
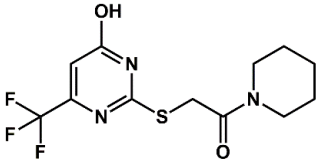
为评估 SQMPC 方法结合自由能预测的准确性, 我们挑选了五种不同的靶标蛋白模型 (ER、FABP4、BACE1、PDE4D 和 VEGFR2), 每种蛋白计算分析了 10 个有活性数据的配体分子。除 ER 体系中的 3OMQ_CH3 和 3OMO_CH3 复合物初始结构是在 3OMQ 晶体结构上进行配体修改得到, 其余 48 个蛋白-配体复合物的模拟初始结构均从 PDB 数据库中直接下载得到, 详细信息见表 2.1。所有复

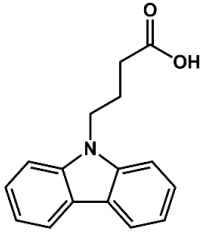
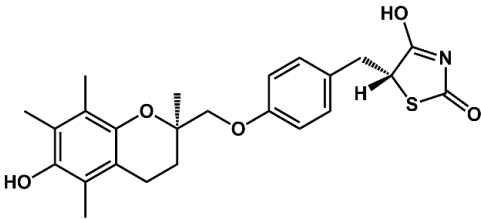
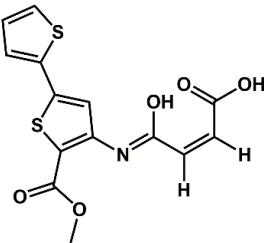
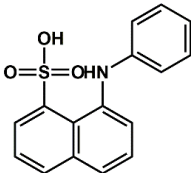
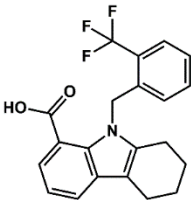
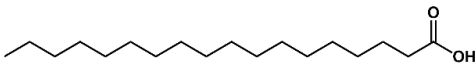
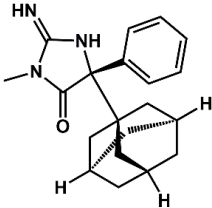
合物体系中的缺失链采用 Modeller 软件补齐。配体分子首先在 Gaussian 09 中进行结构优化, 采用 B3LYP/6-31G*基组, 计算静电势后拟合得到 RESP 电荷。体系的质子化状态通过 H^{++} 网站预测 (<http://biophysics.cs.vt.edu/>)。

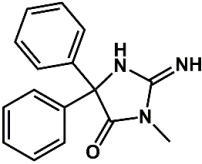
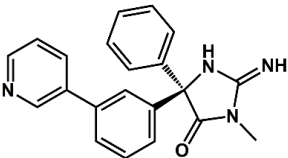
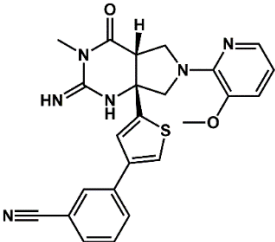
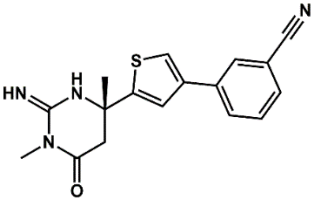
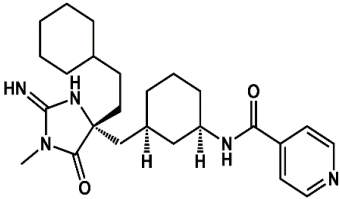
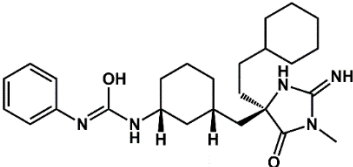
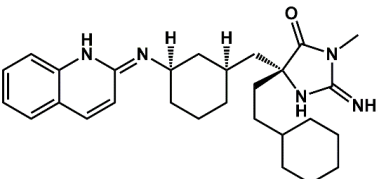
表 2.1 SQMPC 方法测试体系及其结合自由能实验数据*。

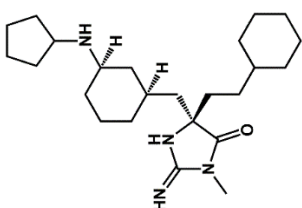
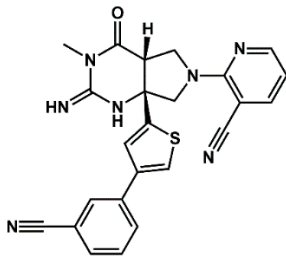
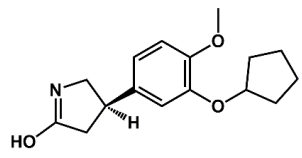
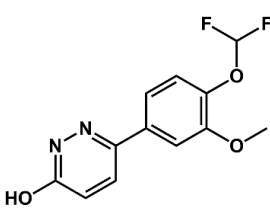
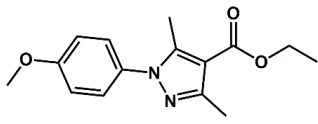
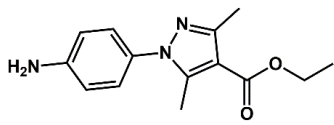
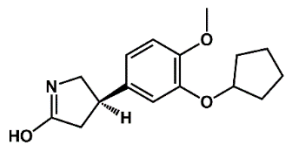
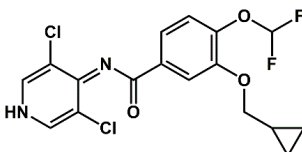
Table 2.1 The Experimental ΔG (kcal/mol) Data of the Tested System for SQMPC.

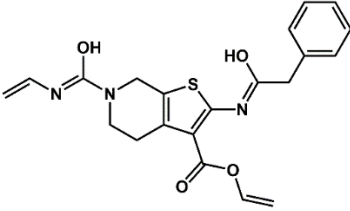
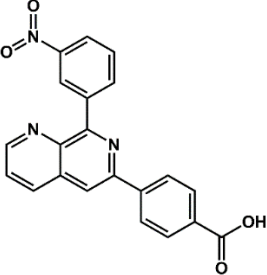
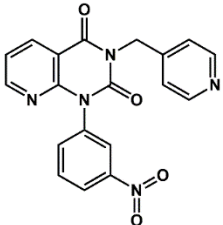
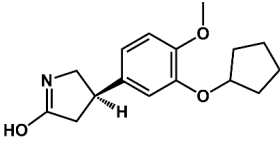
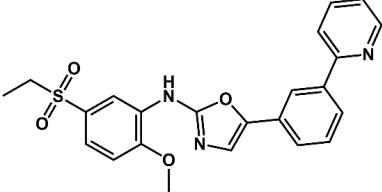
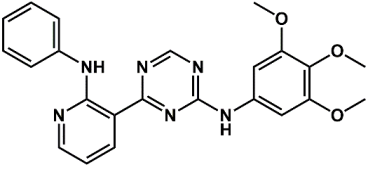
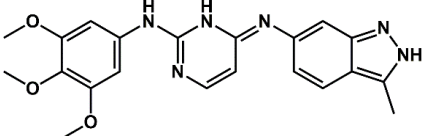
System	Ligand structure	Ligand charge	ΔG_{exp}
1X7B		0	- 11.44 ^[166]
1X76		0	- 11.90 ^[166]
1U3S		0	- 12.08 ^[167]
2I0G		0	- 13.26 ^[168]
1YYE		0	- 11.69 ^[169]
3OMQ		0	-8.49 ^[170]

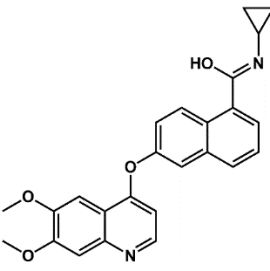
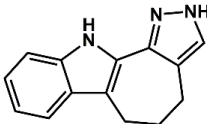
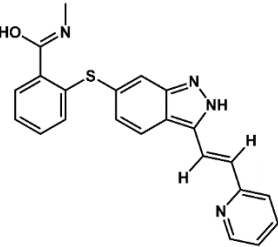
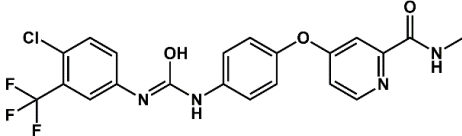
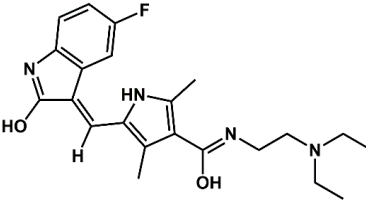
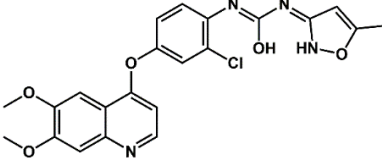
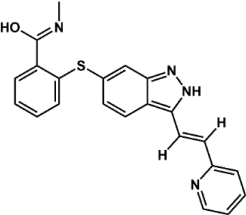
3OMQ_CH3		0	-6.41 ^[170]
3OMP		0	-7.39 ^[170]
3OMO		0	-8.25 ^[170]
3OMO_CH3		0	-4.97 ^[170]
2NNQ		-1	- 11.87 ^[171]
1HMR		-1	-8.94 ^[172]
1HMS		-1	-8.68 ^[172]
1TOU		0	-8.19 ^[173]

1TOW		-1	-8.52 ^[174]
2QM9		0	-10.6 ^[175]
3HK1		-1	-8.42 ^[176]
2ANS		0	-8.08 ^[177]
3FR4		-1	-8.44 ^[178]
1HMT		-1	-7.89 ^[172]
3IND		0	-7.93 ^[179]

4DJU		0	-7.42 ^[180]
4DJW		0	-8.56 ^[180]
4H3I		0	- 11.63 ^[181]
4HA5		0	-9.87 ^[181]
4R92		0	- 10.11 ^[182]
4R93		0	- 10.64 ^[182]
4R95		0	- 10.71 ^[182]

4R9I		0	-8.68 ^[182]
4H3G		0	- 11.22 ^[181]
1Q9M		0	-8.84 ^[183]
1XOR		0	-8.74 ^[184]
1Y2D		0	-7.77 ^[185]
1Y2E		0	-9.27 ^[185]
1OYN		0	-8.54 ^[184]
3G4L		0	- 11.24 ^[186]

3SL4		0	- 10.32 ^[187]
2QYN		-1	-8.52 ^[188]
3G4G		0	- 12.33 ^[186]
3G4K		0	-8.92 ^[186]
1Y6A		0	- 10.45 ^[189]
2P2H		0	-9.78 ^[190]
3CJF		0	- 11.18 ^[191]

3B8R		0	- 12.58 ^[192]
3VID		0	-8.47 ^[193]
4AG8		0	- 12.22 ^[194]
4ASD		0	- 12.66 ^[194]
4AGD		1	- 11.47 ^[194]
4ASE		0	- 15.01 ^[194]
4AGC		0	- 14.60 ^[194]

*: The binding free energy is evaluated using equation $\Delta G_{exp} = RT \ln IC_{50}$.

为了评估 ASMD-SQMPC 方法的准确性,我们从 PDB 数据库中选取 24 个 HIV-1 蛋白酶抑制剂体系作为模型测试体系 (4U7V^[195]、2QNQ^[196]、2QNP^[196]、2PSV^[197]、2QI0^[198]、2QHY^[198]、2FDD^[199]、1D4J^[200]、1D4Y^[201]、1W5X^[202]、1W5Y^[202]、1EC0^[202]、3OXC^[203]、3KDD^[204]、2UXZ^[205]、1OHR^[206]、3EKX^[207]、2Z4O^[208]、3NU3^[209]、3H5B^[210]、1HXW^[211]、3O9G^[212]、3O9F^[212]和 2BPX^[207]),其中包括 8 个含氟抑制剂体系(2QI0、2QHY、1D4J、1W5X、3KDD、1W5Y、1EC0 和 1D4Y)。

2.2.2 半经验量子化学电荷 SQMPC 计算

SQMPC 计算流程如图 2.1 所示,我们采用蛋白链分块的方法,将含有 N 个氨基酸残基的蛋白质链以残基为单位划分成 N 个 SQM/MM 计算体系。每个 SQM/MM 计算体系包括核心量化区,过渡量化区,分子力学区和溶剂化电荷区四个部分。其中核心量化区是由分块得到的第 i 个残基的原子组成,过渡量化区是指与核心量化区原子直接相连的残基,将这两个残基包括在量化区内可以减小量化计算的边界效应,让计算得到的残基静电势更加准确。然后将体系中的其他原子作为分子力学区。由于量化计算通常是在真空中进行的,而目前的研究表明溶剂化效应对体系的影响非常大。因此为了考虑溶剂化效应,我们使用软件 DELPHI 获得蛋白质的溶剂化电荷^[156],然后将这些电荷以背景电荷的方式放在 SQM/MM 计算中。所有的 SQM 量化计算均使用 MOPAC 软件,利用 SQM/MM 方法计算体系的 Coulson 电荷。最后判断此时得到的电荷和原来的电荷是否收敛,如果不收敛,利用更新的电荷重新进行 SQM/MM 计算直到收敛为止。为评估不同 SQM 方法计算的准确性,我们比较了 4 种不同的 SQM 方法: AM1、PM3、PM6 和 PM7。

2.2.3 常规分子动力学模拟

分子动力学模拟通过 Amber 14 软件来完成。首先在模拟初始结构中加入一定数量的 Na⁺或 Cl⁻中和体系电荷,并采用 TIP3P 水溶剂模型模拟溶剂环境,水盒子边缘距离蛋白体系至少 9 Å。GAFF^[213]和 Amber ff03^[214]力场分别用来参数化配体和蛋白体系。模拟中采用 SHAKE 算法限制含氢共价键的伸缩^[215],时间间隔为 2 fs,采用 PME 算法计算长程静电作用^[216]。为了消除原子间不合理接触,我们对蛋白质重原子施加强限制力 (1 kcal/mol/Å²) 并进行 5000 步的最速下降法和 5000 步的共轭梯度法优化,随后在 0.2 ns 内将体系从 0 K 逐渐升温到 300 K,

进而在 NPT 系综和 300 K 温度下平衡 0.1 ns。体系升温和平稳过程都对蛋白质重原子施加强限制力 ($1 \text{ kcal/mol/\AA}^2$)。最后, 我们通过 pmemd.cuda 在 GPU Tesla K40 上进行 5 ns 的常规分子动力学模拟。

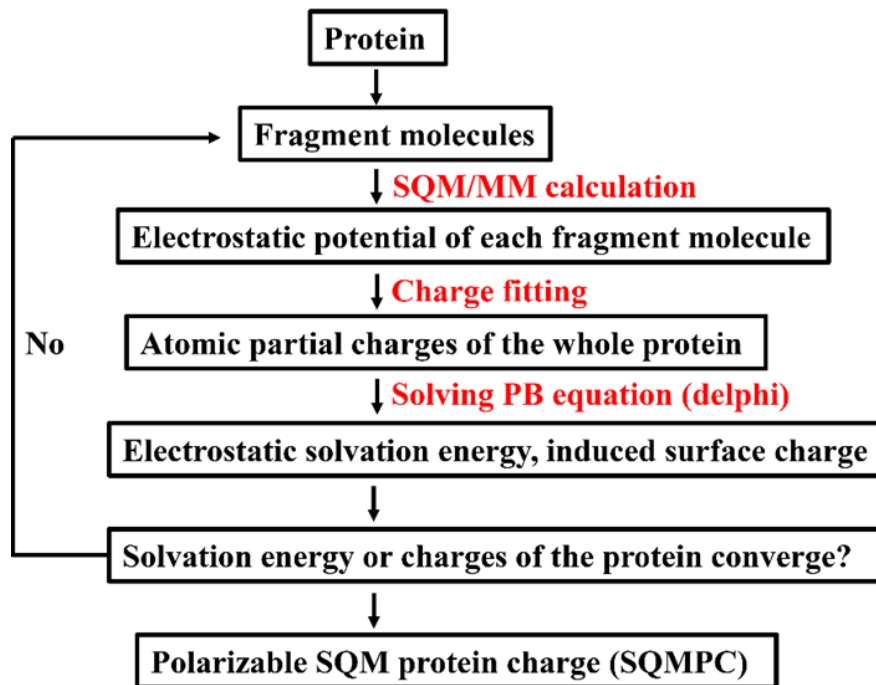


图 2.1 SQMPC 计算流程。

Figure 2.1 The workflow of SQMPC.

2.2.4 结合自由能计算

基于 0.5-5 ns MD 模拟轨迹以均一时间间隔方式取出 1000 个构象用于蛋白质-配体结合自由能计算。计算方法选取广泛使用的 MM/GBSA 和 SIE 方法。

对于 MM/GBSA 方法, 通过 Amber14 中 *mmpbsa.pl* 程序以公式 2.1 计算结合自由能:

$$\Delta G = \Delta H - T\Delta S = \Delta E_{ele} + \Delta E_{VDW} + \Delta G_{gb} + \Delta G_{np} - T\Delta S \dots (2.1)$$

其中 ΔE_{ele} 和 ΔE_{VDW} 分别代表静电和范德华相互作用能。使用 Onufriev 等人发展的 GB 模型计算极性溶剂化自由能 ΔG_{gb} , 而溶剂和溶质的介电常数分别设为 80 和 1。非极性溶剂化自由能 ΔG_{np} 基于可极化表面运用公式 2.2 计算:

$$\Delta G_{np} = \gamma SASA + \beta \dots (2.2)$$

其中 *SASA* 为体系溶剂可极化表面积, 使用 *MSMS* 程序计算得到。 γ 和 β 分别设置为 $0.0072 \text{ kcal/mol/\AA}^2$ 和 0.0 kcal/mol 。

结合过程中的熵变包括了转动、平动和振动三个部分，通过 Amber 14 中的 NMODE 模块计算得到。

对于 SIE 方法，通过 *sietraj* 程序运用公式 2.3 计算结合自由能：

$$\Delta G_{bind} = \alpha[E_C(D_{in}) + \Delta G_{bind}^R(\rho, D_{in}) + E_{VDW} + \gamma\Delta MSA(\rho)] + C \dots (2.3)$$

其中 E_C 和 E_{VDW} 分别代表采用分子力场计算的气相中静电和范德华相互作用能。本研究中采用 Amber ff03 分子力场。 ΔG_{bind}^R 表示配体结合时导致的反应场能变化，通过溶剂探针半径求解泊松方程得到。 ΔMSA 表示配体结合时诱导的溶剂可及表面积的变化。参数 α 是与熵变相关的比例系数， D_{in} 是溶质内部的介电常数， ρ 是溶剂可及表面积系数， γ 是线性标度的范德华半径系数， C 是能量常数。在本研究中，这些参数值均为 *sietraj* 程序默认值： $\alpha=0.1048$ ， $D_{in}=2.25$ ， $\rho=1.1$ ， $\gamma=0.0129 \text{ kcal/mol/\AA}^2$ ， $C=-2.89 \text{ kcal/mol}$ 。

2.2.5 自适应拉伸分子动力学模拟

在自适应拉伸分子动力学中，配体拉伸反应路径被分成 N 个模拟区间。在每个模拟区间内多条轨迹平行模拟，每个模拟副本中配体分子均在简谐力的驱动下沿着预先设置好的变化路径运动，并由 Jarzynski 公式 (2.4) 计算模拟平均力势 (PMF)。公式中 $\beta=1/(k_B T)$ ， $W_{\tau_t \leftarrow \tau_{t-1}}$ 是第 t 个模拟区间所做的功。在每个模拟区间内将最接近 PMF 的轨迹作为下一模拟区间的初始轨迹。

$$G(\tau_t) = G(\tau_{t-1}) - \frac{1}{\beta} \ln \langle e^{-\beta W_{\tau_t \leftarrow \tau_{t-1}}} \rangle_{t-1} \dots (2.4)$$

在本章中，反应路径设置为 HIV-1 蛋白酶活性位点 (R25A 和 R25B) 与抑制剂之间的距离，范围为 7 Å 到 32 Å，共分成 8 个模拟区间。研究报道配体被拉伸解离的速度越小，所做的功越接近 PMF，因此我们参考其它研究设置配体分子以 10 Å/ns 的速度离开结合位点^[217]。配体解离的前 4 个模拟区间简谐力弹簧系数 k 设置为 7.2 kcal/mol/Å²；后 4 个模拟区间设置为较小的弹簧系数($k=2 \text{ kcal/mol/\AA}^2$)^[217-219]。为评估弹簧系数对 ASMD 模拟结果的影响，我们同时测试了其他弹簧系数值 (2、5 和 10 kcal/mol/Å²)。

2.3 结果与讨论

2.3.1 SQMPC 方法提高雌激素受体抑制剂的结合稳定性

为评估 SQMPC 电荷对分子动力学模拟采样的影响，我们首先采用 10 个不

同的雌激素受体-配体复合物作为模型体系。由于不同的配体具有不同的极化效应，所以结合不同的配体时蛋白质的静电势是有区别的。如图 2.2A 所示，对于配体结构相似的 1X7B 和 1X76 复合物体系，通过 PM7 方法计算的 SQMPC 电荷与 Amber ff03 力场电荷的区别主要集中于配体附近的氨基酸残基，如 H212 和 M77 等；而距离配体较远的残基受配体静电极化的影响比较小。氢键是影响蛋白-配体相互作用的一个关键因素，因此，我们比较分析了通过 4 种 SQMPC 电荷（SQMPC_{AM1}、SQMPC_{PM3}、SQMPC_{PM6} 和 SQMPC_{PM7}）模拟的蛋白-配体间氢键的稳定性。以 1X7B 体系为例，如图 2.2B 所示，配体上的羟基氧原子（O3）与受体蛋白谷氨酸（E42）残基的羧基氧形成氢键，同时配体上的羟基氧原子（O2）与受体蛋白组氨酸（H212）残基上的氮原子（ND1）也形成氢键作用。我们通过 Amber 14 软件的 *Cpptraj* 模块统计分析 MD 模拟中这两个氢键的距离来分析氢键的稳定性。如图 2.2C&D 所示，E42 形成的氢键较为稳定，其氢键距离频率分布峰值都低于 2.7 Å。其中，PM6 和 PM7 方法计算得到的 SQMPC 电荷（SQMPC_{PM6} 和 SQMPC_{PM7}）与精度更高的 HF 和 B3LYP 方法计算得到的 QMPC 电荷（QMPC_{HF} 和 QMPC_{B3LYP}）所得到的氢键距离频率分布几乎重叠，均在 2.50 到 2.52 Å 之间。而 H122 形成的氢键比 E42 更不稳定，其氢键距离频率分布峰值都在 2.80-2.96 Å 之间。这些结果与 1X7B 复合物实验数据相符，E42 和 H122 在晶体结构中配体形成的氢键距离分别是 2.53 和 2.94 Å。此外，我们发现在 MD 模拟中，通过 SQMPC 电荷模拟的氢键都比直接采用 Amber ff03 电荷的模拟更稳定。

我们进一步计算了 10 个 ER-配体复合物体系中氢键作用的平均距离。如表 2.2 所示，通过 QMPC_{B3LYP}、QMPC_{HF}、SQMPC_{AM1}、SQMPC_{PM6}、SQMPC_{PM7} 和 SQMPC_{PM3} 电荷模拟的氢键平均距离分别为 3.10、3.06、3.12、3.32、3.41 和 3.56 Å，但直接采用 Amber ff03 电荷的平均距离为 4.08 Å。这一结果再次表明相比采用 Amber ff03 电荷，通过 SQMPC 电荷更能够稳定蛋白-配体复合物模拟的氢键作用。在这四种不同的 SQMPC 电荷中，SQMPC_{AM1}、SQMPC_{PM6} 和 SQMPC_{PM7} 电荷比 SQMPC_{PM3} 电荷得到氢键作用更稳定。

体系是否达到动力学平衡对结合自由能的准确计算至关重要，为了评估各体系是否达到动力学平衡，我们计算了 ER-配体体系主链重原子的均方根误差（RMSD）随模拟时间的变化曲线（图 2.3）。如图 2.3 所示，10 个体系的主链重原子 RMSD 值在 0.5 ns 后趋于平稳，因此我们使用 0.5-5 ns 的轨迹进行接下来的

结合自由能计算。同时我们发现对于一些模拟体系，如 3OMP 等，采用不同 SQMPC 电荷的 MD 模拟得到的 RMSD 值差异较大，而这些差异主要来自它们的不同电荷，这也说明了体系电荷对 MD 模拟过程中的构象变化具有重要影响。表 2.3 总结了通过各种 SQMPC 电荷模拟得到的体系 RMSD 变化情况，结果显示 B3LYP 具有最低 RMSD 值，AM1 方法在 5 ns 的 MD 轨迹中 RMSD 变化最小，而 Amber ff03 具有最大的 RMSD 值。

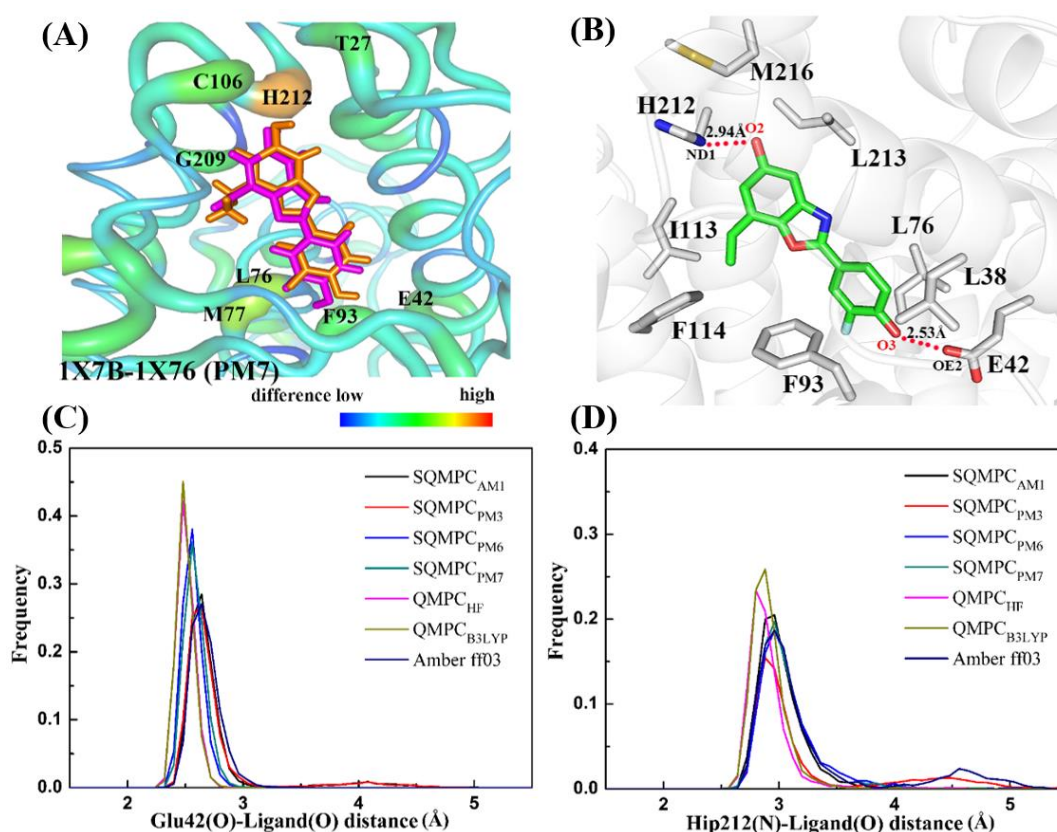


图 2.2 SQMPC 应用于 ER 体系的动力学分析。(A) 1X7B 与 1X76 的 SQMPC_{PM7} 电荷差异，(B) 1X7B 复合物结构中配体与蛋白形成的氢键展示，(C) E42 与配体形成的氢键距离频率分布，(D) H122 与配体形成的氢键距离频率分布。

Figure 2.2 The MD analysis of ER system with SQMPC. (A) The difference of SQMPC_{PM7} on 1X7B and 1X76, (B) The two hydrogen bonds formed by ER protein and its ligand in 1X7B, (C) The distance frequency analysis of the hydrogen bond formed by E42 and ligand, (D) The distance frequency analysis of the hydrogen bond formed by H122 and ligand.

表 2.2 SQMPC 与 Amber ff03 电荷模拟得到氢键平均距离比较*。

Table 2.2 Mean Value of the Hydrogen Bond Length (Å) in ER-inhibitor Systems with SQMPC and Amber ff03 Charge.

System	Hydrogen bond	Exp	AM1	PM3	PM6	PM7	HF	B3LYP	ff03
1X7B	E42(O)-LBD(O)	2.53	2.74	2.74	2.58	2.56	2.50	2.50	2.67
	H212(N)-LBD(O)	2.94	3.02	3.64	3.05	3.09	3.11	2.91	3.24
1X76	H212(N)-LBD(O)	3.28	3.01	4.31	3.04	3.29	4.21	3.28	4.20
1U3S	H212(N)-LBD(O)	2.91	2.96	4.08	3.36	4.60	3.29	4.52	4.62
	E42(O)-LBD(O)	2.61	2.62	3.10	2.57	2.58	2.48	2.50	5.30
	R83(N)-LBD(O)	3.40	4.25	3.54	4.23	4.18	3.43	3.44	4.31
2I0G	H212(N)-LBD(O)	3.42	2.98	3.05	3.00	2.98	2.88	2.90	5.28
	E42(O)-LBD(O)	2.46	2.63	3.83	2.61	2.64	2.49	2.52	5.64
	R83(N)-LBD(O)	2.99	3.61	3.47	3.50	3.78	3.43	3.51	3.45
1YYE	H212(N)-LBD(O)	3.18	3.10	5.56	5.68	5.20	3.03	2.93	5.26
	L76(O)-LBD(O)	2.77	3.67	3.74	3.53	3.49	2.71	3.60	4.57
3OMQ	L76(O)-LBD(O)	2.86	3.53	3.54	3.82	3.92	3.50	3.43	3.69
3OMQ _CH3	E42(O)-LBD(O)	2.88	3.49	3.52	3.69	3.65	3.67	3.26	3.74
3OMP	E42(O)-LBD(O)	2.60	2.67	2.68	2.63	2.66	2.62	2.63	2.66
3OMO	E42(O)-LBD(O)	3.03	2.59	2.65	2.59	2.53	2.54	2.62	2.63
mean		2.93	3.12	3.56	3.32	3.41	3.06	3.10	4.08

*: 统计误差根据 5 ns 的分子动力学模拟轨迹得到。

表 2.3 不同 SQMPC 电荷的在不同 MD 时长下的平均 RMSD 值。

Table 2.3 The mean RMSD values of different SQMPC in different MD simulation time.

Methods	0.5-1.0 ns	1.0-4.0 ns	4.0-5.0 ns	Δ RMSD
SQMPC _{HF}	1.24±0.18	1.51±0.24	1.69±0.24	0.45±0.16
SQMPC _{B3LYP}	1.23±0.17	1.42±0.30	1.66±0.38	0.43±0.26
SQMPC _{AM1}	1.43±0.18	1.62±0.18	1.71±0.18	0.28±0.17
SQMPC _{PM3}	1.54±0.21	1.93±0.25	2.01±0.26	0.47±0.20
SQMPC _{PM6}	1.42±0.28	1.63±0.36	1.89±0.34	0.47±0.14
SQMPC _{PM7}	1.52±0.45	1.70±0.53	1.88±0.51	0.36±0.23
Amber ff03	1.35±0.14	1.59±0.23	1.87±0.30	0.52±0.26

$$\Delta\text{RMSD} = \text{RMSD}(0.5-1.0 \text{ ns}) - \text{RMSD}(4.0-5.0 \text{ ns})$$

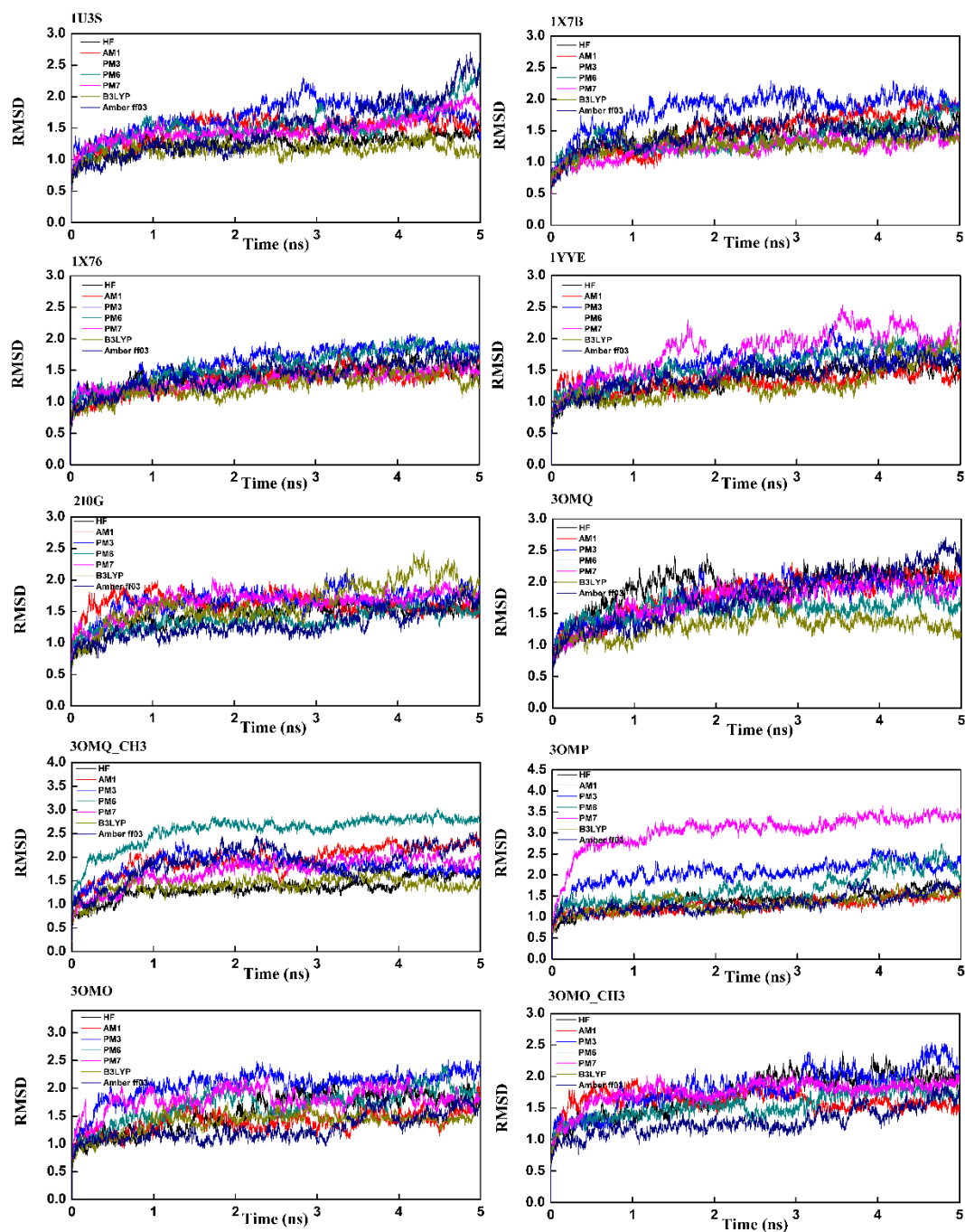


图 2.3 雌激素复合物测试体系的 RMSD 分析。

Figure 2.3 The RMSD (Å) analysis of tested ER-ligand complexes.

2.3.2 雌激素受体与配体的结合自由能预测准确性评估

结合自由能是评估蛋白质与配体间亲和力的重要参数，也是分子对接软件中评价配体活性的重要物理量。因此，能否准确预测结合自由能是评价打分函数的

关键指标。我们通过 MM/GBSA 和 SIE 两种基于 MD 的结合自由能计算方法计算了 ER-配体复合物的结合自由能。图 2.4 比较了通过 SQMPC 电荷和 Amber ff03 电荷得到的结合自由能计算值与实验值的相关性 (R^2)。对于 SIE 方法, B3LYP、HF、AM1、PM3、PM6 和 PM7 方法的相关性分别为 0.92 ± 0.01 、 0.83 ± 0.01 、 0.94 ± 0.01 、 0.69 ± 0.01 、 0.86 ± 0.01 和 0.98 ± 0.01 , 而直接采用 Amber ff03 电荷得到的 R^2 为 0.64 ± 0.01 , 因此通过各种不同的 SQMPC 电荷均能够明显提高 SIE 方法对蛋白-配体结合自由能的预测准确性。对于 MM/GBSA 方法, 通过各种不同的 SQMPC 电荷也能够获得更高的与结合自由能实验值的相关性。

从图 2.4 中我们发现通过 PM7 和 AM1 得到的 SQMPC 电荷对结合自由能预测准确性的提高最为明显, 甚至优于精度更高的密度泛函方法。而且, PM7 和 AM1 需要的 SQMPC 电荷计算时间 (1.5 核时) 仅为 HF 和 B3LYP (30 核时) 的 5%。另一方面, 我们发现 SIE 得到的相关性大都优于 MM/GBSA, 而且 SIE 不需要额外计算熵变化, 因此 SIE 在计算时间上比 MM/GBSA 具有明显优势。

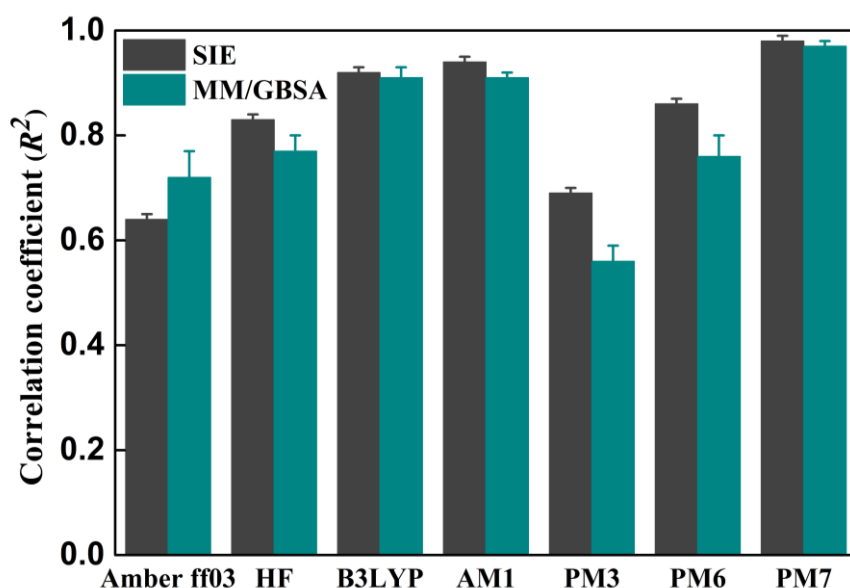


图 2.4 SQMPC 结合 SIE 和 MM/GBSA 方法得到的相关性分析。

Figure 2.4 Correlation coefficients of SQMPC calculated by SIE and MM/GBSA.

2.3.3 SQMPC 方法的体系可拓展性评估

为了研究 SQMPC 方法的体系可拓展性, 我们继续选用另外 4 种不同的靶标蛋白体系共计 40 个复合物进行测试, 这些蛋白包括 FABP4、BACE1、VEGFR2

和 PDE4D。这些体系的 MD 模拟以及自由能计算方法都与上述处理雌激素受体的方式一样。同样我们通过比较结合自由能计算值与实验值的相关性进行方法准确性评估。图 2.5 比较了这四种蛋白质体系通过 SQMPC 电荷与 MM/GBSA 和 SIE 方法得到的相关性。结果显示, 相比 Amber ff03 电荷, SQMPC 电荷都提高了结合自由能计算值与实验值相关性。

对于 FABP4 体系, SIE-AM1 的组合得到了最高的相关性, 其 R^2 为 0.71 ± 0.02 , 相比直接采用 Amber ff03 电荷 ($R^2 = 0.22 \pm 0.02$) 有显著的提高。其次 SIE-PM7 以及 SIE-PM6 的组合也得到了较好的相关性, 其 R^2 均为 0.59 ± 0.02 。此外, SQMPC 电荷与 MM/GBSA 方法的结合也能够提高采用常规 Amber ff03 电荷的 MM/GBSA 方法的结合自由能预测准确性。

对于 BACE1 体系, SIE-PM6 以及 SIE-AM1 的组合得到了最高的相关性, 其 R^2 分别为 0.83 ± 0.02 和 0.81 ± 0.02 。其次, 相比直接采用 Amber ff03 电荷 ($R^2 = 0.63 \pm 0.02$), SIE-PM7 的组合也明显提高了相关性, 其 R^2 为 0.76 ± 0.01 。在 4 种不同的 SQMPC 电荷中, PM3 方法得到的相关性 ($R^2 = 0.67 \pm 0.02$) 低于其他 3 种 SQM。另外我们也能发现, 与 FABP4 体系相似, SQMPC 电荷与 MM/GBSA 方法的结合也能够提高采用常规 Amber ff03 电荷的 MM/GBSA 方法的结合自由能预测准确性。

对于 PDE4D 体系, SIE-AM1、SIE-PM6 和 SIE-PM7 的组合均得到了比采用常规 Amber ff03 电荷更高的相关性 ($R^2 = 0.61 \pm 0.02$), 其 R^2 分别为 0.66 ± 0.02 、 0.74 ± 0.02 和 0.64 ± 0.02 。MM/GBSA-PM6 和 MM/GBSA-PM7 也得到了较好的结合自由能预测, 其 R^2 分别为 0.58 ± 0.05 和 0.66 ± 0.08 。

对于 VEGFR2 体系, 相比直接采用 Amber ff03 电荷 ($R^2 = 0.58 \pm 0.01$), SIE-AM1、SIE-PM3、SIE-PM6 和 SIE-PM7 的组合均得到了较高的结合自由能相关性, 其 R^2 分别为 0.69 ± 0.01 、 0.65 ± 0.01 、 0.64 ± 0.01 和 0.65 ± 0.01 。同样地, MM/GBSA-AM1、MM/GBSA-PM3、MM/GBSA-PM6 和 MM/GBSA-PM7 的组合也比直接采用 Amber ff03 电荷效果更好。

因此, 目前测试的 5 个靶标蛋白体系共计 50 个蛋白-配体复合物均表明 SQMPC 电荷能够明显提高结合自由能计算的准确性。在 4 种不同的 SQM 计算得到的 SQMPC 电荷中, PM6、PM7 和 AM1 的效果普遍优于 PM3。SIE 得到的相关性也普遍比 MM/GBSA 高。因此, 目前结果表示 SIE 与 SQMPC 方法结合

(SIE-SQMPC) 能够提高蛋白-配体结合自由能的预测准确性, 而且耗时低, 有希望成为高效准确的结合自由能计算方法。

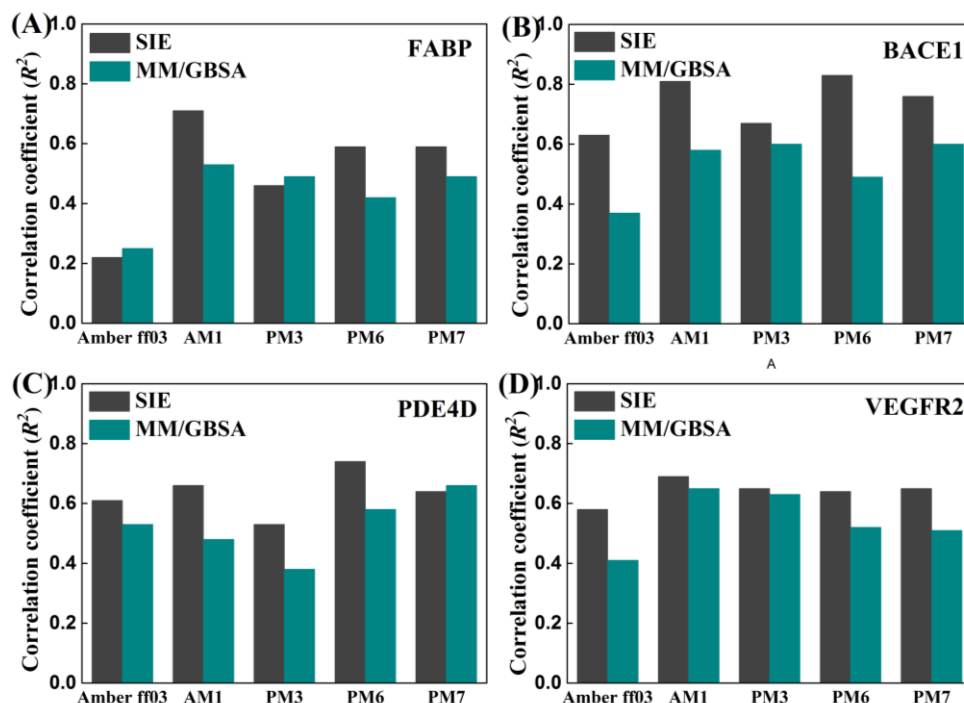


图 2.5 SQMPC 在 4 个不同体系上的测试。

Figure 2.5 SQMPC was tested on four different systems.

进一步我们计算了 5 个模型体系中配体的 Tanimoto 相似性系数 (图 2.6), 结果表明本研究中测试的配体分子骨架存在明显差异。例如, 在 FABP4 体系中, 1HMS、1HMT 和 1HMR 复合物之间的 Tanimoto 系数只有 0.1 到 0.3。这也说明 SQMPC 电荷也能区分骨架差异大的配体分子间的活性差异。

在相关性提高最明显的 FABP4 体系中, 我们发现该蛋白的 10 个测试配体中有 7 个配体是带净电荷的, 这些配体对蛋白结合口袋残基的极化效应应当更明显, 因此 SQMPC 在这些体系上的相关性与 Amber ff03 相比, 提高程度更大是可以理解的, 从而也进一步证明了 SQMPC 方法重新计算复合物体系极化电荷的优越性。

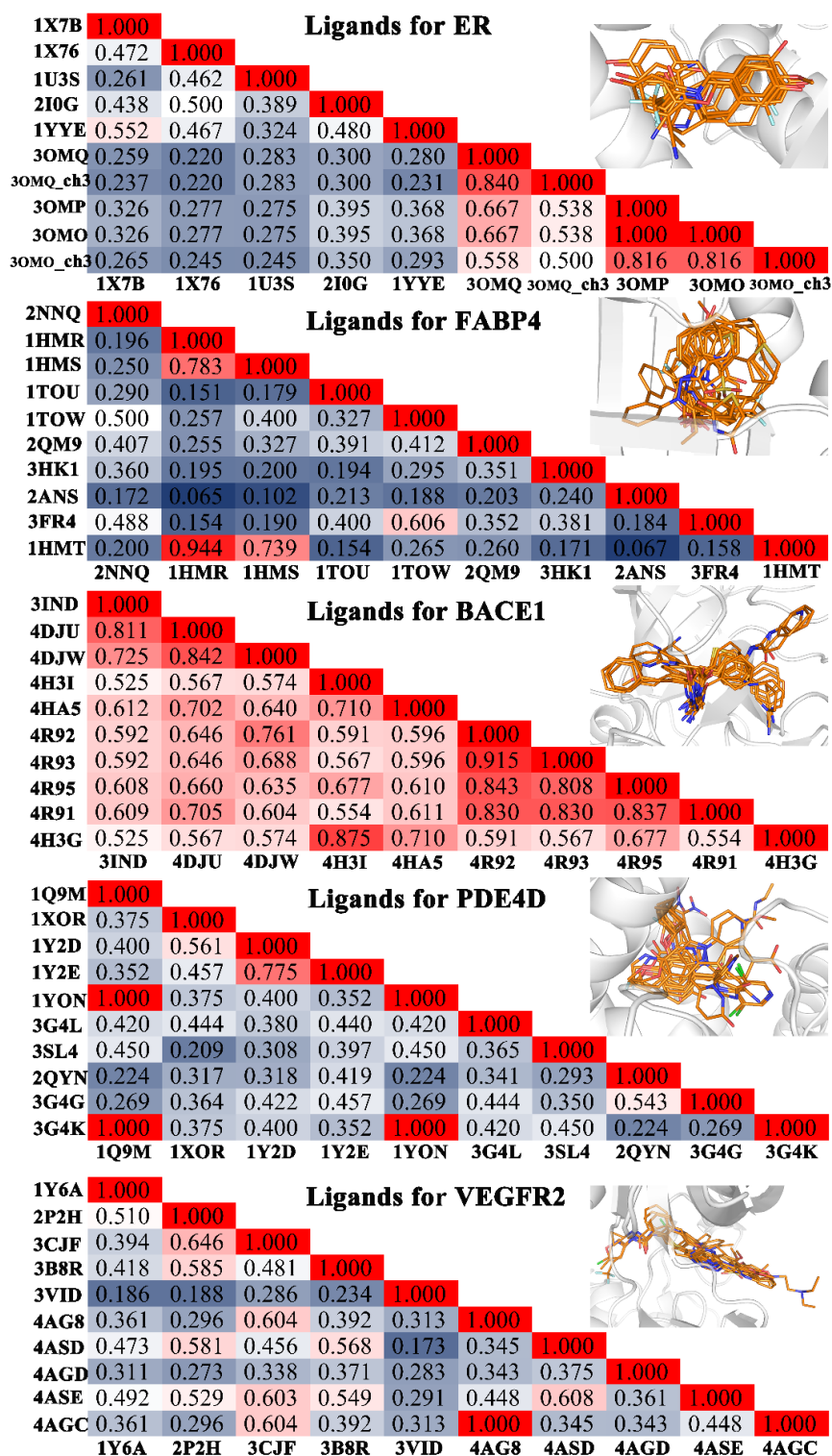


图 2.6 分子骨架差异性分析。分子间骨架差异性通过 Tanimoto 系数评估，指纹谱通过 MACCS 获得。

Figure 2.6 The analyses of molecular scaffold diversity. The pairwise scaffold diversity was evaluated by structural fingerprints using Tanimoto similarity coefficient. Fingerprints were obtained by Molecular ACCess System (MACCS) keys.

2.3.4 MD 模拟时间对结合自由能相关性的影响

为了评估 MD 模拟时间对 SIE-SQMPC 结合自由能预测准确性的影响，我们比较了通过不同时长 MD 轨迹计算得到的结合自由能与实验值的相关性，包括 0.5-1 ns、0.5-2 ns、0.5-3 ns、0.5-4 ns 和 0.5-5 ns 五个不同的 MD 模拟时长。如图 2.7 所示，MD 模拟时长能够明显影响结合自由能的预测相关性，但相关性并不是随着 MD 模拟时间的延长而提高，这一点与侯廷军等人的研究是保持一致的^[220-221]。例如，PM6 计算 FABP4 体系结果表明短时间的 MD 模拟能达到更好的预测效果，而计算 BACE1 和 PDE4D 体系则体现 MD 模拟时间越长结果越好。

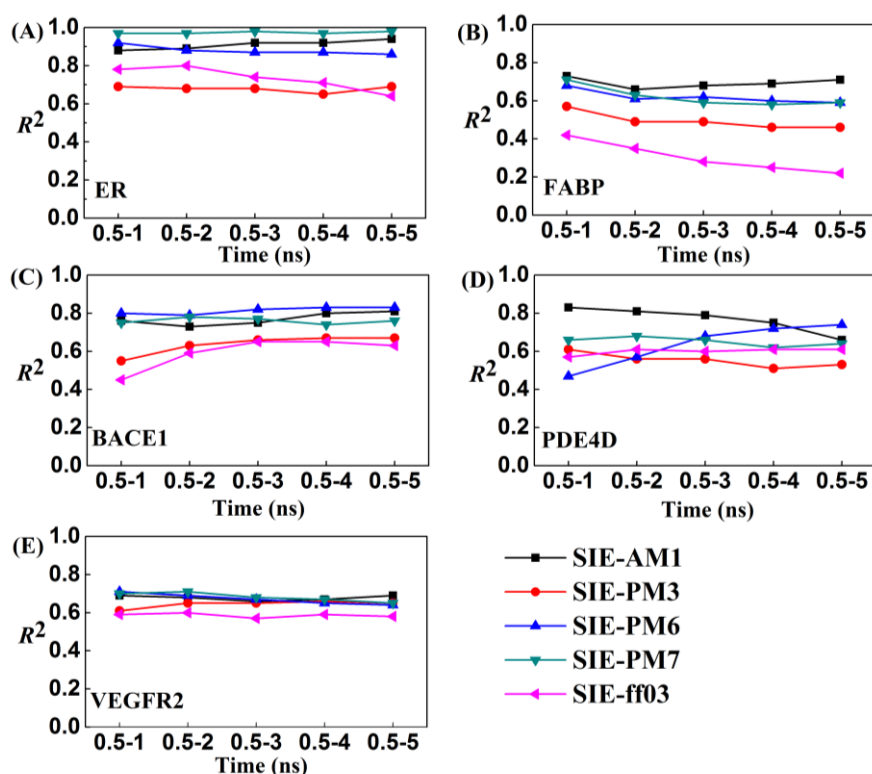


图 2.7 MD 模拟时长对相关性的影响。

Figure 2.7 the impact of MD simulation time on correlation coefficients.

由于不同体系中相关性最高的 MD 模拟时长不一致，难以获取通用的最佳 MD 模拟时长作为虚拟筛选参考。因此，我们以 4 种 SIE-SQMPC 方法为出发点，分析了它们各自最佳的 MD 模拟时长。如图 2.8 所示，对于测试的 5 个靶标体系，SIE-AM1 和 SIE-PM6 的组合在较短的 MD 模拟时间（0.5-1 ns）能获得更好的结合自由能预测准确性，而 SIE-PM7 在不同 MD 模拟时间内计算的相关性变化不大，尤其是对于 PDE4D 体系，这一结果表明了 SIE-PM7 方法的鲁棒性。

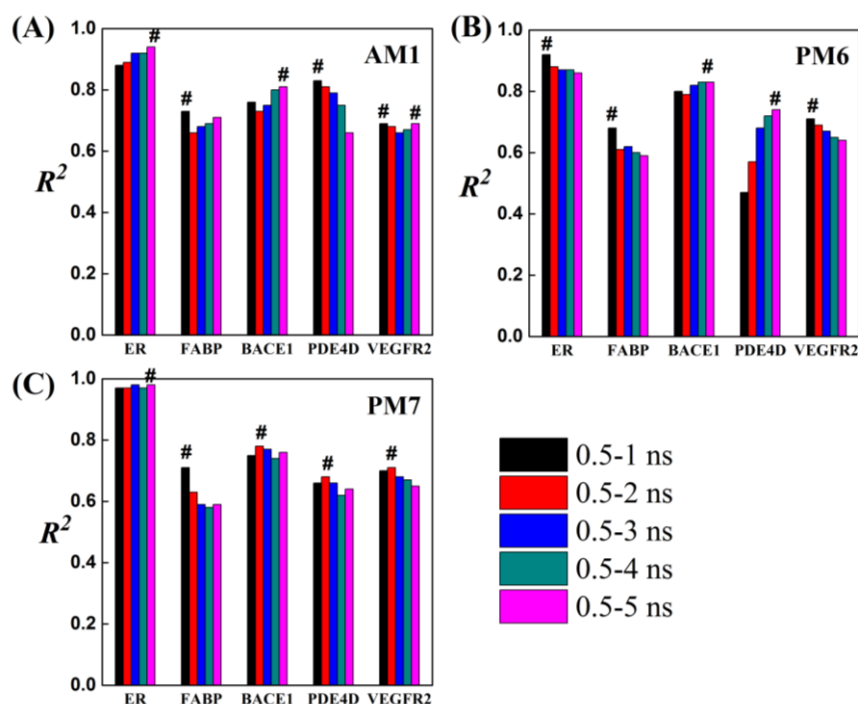


图 2.8 MD 模拟时间对结合自由能相关性的影响。在 5 个研究体系中，SIE-AM1(A)、SIE-PM6(B)和 SIE-PM7(C)方法在不同 MD 模拟时长下的结合自由能计算值与实验值相关性 (R^2)。“#”表示 R^2 最好的 MD 模拟时长。

Figure 2.8 The impact of simulation time on correlation coefficients. The correlation coefficients between experimentally determined and calculated ΔG with SIE-AM1 (A), SIE-PM6 (B) and SIE-PM7 (C) using different lengths of MD simulations for the 5 studied protein systems. “#” represents the best R^2 .

2.3.5 SQMPC 方法应用于非平衡分子动力学模拟

为进一步评估 SQMPC 电荷对其他结合自由能计算方法的影响，我们将 SQMPC 与自适应拉伸分子动力学 (ASMD) 相结合，发展了 ASMD-SQMPC 方法 (图 2.9)，并应用于 24 个 HIV-1 蛋白酶-抑制剂复合物体系。ASMD 是一种非平衡分子动力学模拟方法，相比 SIE 和 MM/GBSA 等基于 MD 的打分函数，ASMD 在预测蛋白-配体结合自由能时能够综合考虑配体结合过程中的焓变和熵变效应。然而，ASMD 同样受到分子力场的局限性，即不能考虑在配体解离过程中蛋白质受到的静电极化效应。因此，通过定期更新模拟体系的 SQMPC 电荷或可提高 ASMD 模拟的准确性。

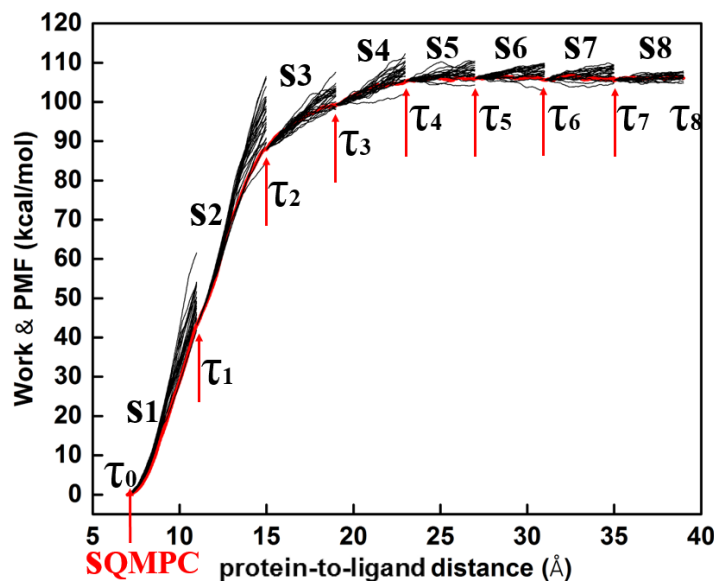


图 2.9 ASMD-SQMPC 计算示意图。在 ASMD-SQMPC 模拟中，配体拉伸反应路径以 $\tau_0, \tau_1, \dots, \tau_8$ 为界限被分成 8 个模拟区间，其中 S1 代表第一个模拟区间，每个模拟区间平行模拟 25 条轨迹。

Figure 2.9 The ASMD-SQMPC. Illustration of the ASMD-SQMPC applied to a system where the steering reaction path is divided into 8 segments bounded by $\tau_0, \tau_1, \dots, \tau_8$. S1 refers to the segment 1 simulation, and other designations (S2-S8) are in a similar way. 25 trajectories are performed independently in each segment.

在 ASMD-SQMPC 模拟中，我们以抑制剂与 HIV-1 PR 结合位点 D25A/B 之间的质心距离为反应坐标，距离范围设置为 7 到 32 Å（图 2.10A）。我们同时测试了弹簧系数为 2、5、7.2 和 10 kcal/mol/Å²，结果发现，当弹簧系数小于 7.2 kcal/mol/Å²，配体分子与蛋白的距离并不能很好的与预设距离相符。因此，本章研究的前四个 ASMD 模拟区间的弹簧系数设置为 7.2 kcal/mol/Å²（图 2.10B）。在偏执力的作用下，随着抑制剂逐渐从结合状态变成未结合状态，体系中每个原子定期通过 SQMPC 更新电荷。以 3OXC 为例，如图 2.10C 所示，模拟中的蛋白质-抑制剂距离与先前预定的距离非常吻合，表明抑制剂已按预期方式离开结合腔。另外，驱动力在前段中逐步增加，但在后四段中几乎在 0 左右波动，表明抑制剂从结合腔中被拉出后已完全进入水中。

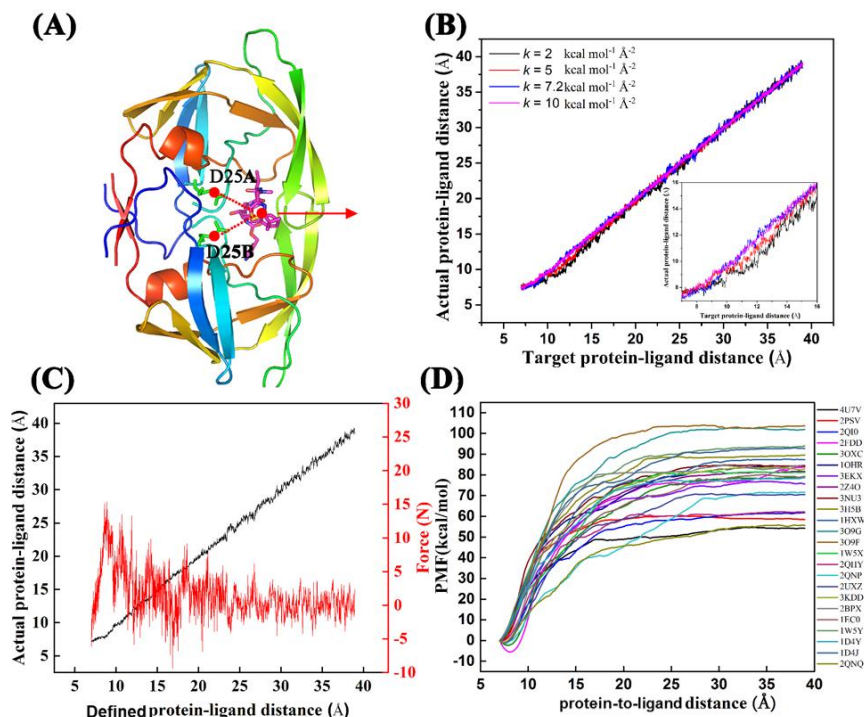


图 2.10 ASMD-SQMPC 应用于 HIV-1 PR。 (A) HIV-1 PR 抑制剂复合物。沿着 HIV-1 PR 结合腔 (D25A 和 D25B) 和抑制剂的质心之间的距离, 拉动路径用红色箭头标记。(B) 通过比较预定义和实际的蛋白质-配体距离来测试弹簧常数 $k=2$ 、 5 、 7.2 或 $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ 。(C) ASMD-SQMPC 模拟中 3OXC 复合物的配体运动轨迹 (黑线) 和偏执力曲线 (红线)。(D) 24 种 HIV-1 PR 抑制剂体系的 PMF 作为 HIV-1 PR 结合腔与抑制剂之间质心距离的函数。

Figure 2.10 The application of ASMD-SQMPC on HIV-1 PR. (A) Representation of the HIV-1 PR-inhibitor complex. The pulling pathway is labeled with a red arrow, along the distance between the mass center of HIV-1 PR binding cavity (D25A and D25B) and the inhibitor, which are labeled with a red dashed line. (B) comparison of the defined and performed protein-ligand distances with the spring constant $k = 2, 5, 7.2$ or $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. (C) The ligand trajectory of 3OXC complex in ASMD-SQMPC simulations (shown as black line), and the driving force profiles (shown as red line). (D) PMFs of the 24 HIV-1 PR-inhibitor systems as a function of the distance of between the mass center of HIV-1 PR binding cavity and the inhibitor.

有研究表明在拉伸分子动力学模拟中配体被拉伸至指定位置所需做的功越大, 则配体与蛋白的亲合力越高。因此, 本章中我们研究了 ASMD 模拟的平均力势 (PMF) 与配体亲和力的相关性。如图 2.10D 所示, 24 个 HIV-1 PR 抑制剂体系的 PMF 有明显的差别, 其 PMF 分布在 54.15 到 103.90 kcal/mol 之间。在蛋白-抑制剂距离为 $25 \text{ \AA}$ 左右时, 24 个 HIV-1 PR 抑制剂体系的 PMF 均达到最

大值区域。因此，我们计算了各体系 25 Å 之后的 PMF 平均值为 $\langle \text{PMF} \rangle_{\text{max}}$ (表 2.4)。其中 4U7V 体系的 $\langle \text{PMF} \rangle_{\text{max}}$ 最小，为 54.15 kcal/mol，是最大的 $\langle \text{PMF} \rangle_{\text{max}}$ 体系 (3O9F) 的 52%。这一结果与实验数据相符，4U7V 和 3O9F 的结合自由能实验值正好是 24 个体系中的最小值 (-7.50 kcal/mol) 和最大值 (-17.33 kcal/mol)，表明 $\langle \text{PMF} \rangle_{\text{max}}$ 与结合自由能可能存在一定的相关性。因此我们以 24 个体系为例，计算了 $\langle \text{PMF} \rangle_{\text{max}}$ 与结合自由能的皮尔森相关性 (R^2)。如图 2.11A 所示， $\langle \text{PMF} \rangle_{\text{max}}$ 与结合自由能实验值之间的相关系数 R^2 为 0.86，表明 ASMD-SQMPC 模拟得到的 $\langle \text{PMF} \rangle_{\text{max}}$ 与结合自由能具有很强的相关性。进一步考虑到配体在简谐力的作用下解离时，除了需要克服蛋白-配体的相互作用能，还涉及到配体在解离后结合口袋的变化，因此我们根据公式 2.5 对 ASMD 模拟得到的 $\langle \text{PMF} \rangle_{\text{max}}$ 进行了结合口袋的体积校正^[222]。在公式 2.5 中， k_B 和 T 分别为玻尔兹曼常数和温度； $G(r_0)$ 和 $G(r_\infty)$ 分别为能量最低值以及配体处于未结合状态时的能量值。由于配体处于 r_0 状态时的自由能面是各向异性的，因此我们将弹簧系数 k 分解为 k_x 、 k_y 和 k_z 。校正结果总结于表 2.5，最终相关系数 R^2 仍为 0.86。

$$\Delta G_{\text{correction}} = G(r_0) - G(r_\infty) - k_B T \ln \frac{\frac{\sqrt{\pi}}{\sqrt{\beta k_x/2}} \frac{\sqrt{\pi}}{\sqrt{\beta k_y/2}} \frac{\sqrt{\pi}}{\sqrt{\beta k_z/2}}}{1661} \quad (2.5)$$

表 2.4 24 个 HIV-1 PR 抑制剂体系的 $\langle \text{PMF} \rangle_{\text{max}}$ 。

Table 2.4 The $\langle \text{PMF} \rangle_{\text{max}}$ of 24 HIV-1 PR inhibitor complexes.

PBD ID	ΔG_{exp}	ASMD-SQMPC	ASMD-ff03	SIE-SQMPC	SIE-ff03
4U7V	-7.50	54.15 ±1.61	37.81 ±1.37	-10.09 ±0.04	-9.93 ±0.05
2QNQ	-8.34	55.71 ±1.93	53.50 ±1.43	-10.53 ±0.05	-9.93 ±0.06
2QNP	-8.75	71.61 ±2.21	41.86 ±2.50	-10.83 ±0.06	-10.15 ±0.06
2PSV	-9.88	60.23 ±1.42	53.78 ±1.32	-9.70 ±0.05	-9.14 ±0.06
2QI0	-10.07	61.44 ±1.44	45.25 ±1.61	-10.34 ±0.05	-10.62 ±0.06
2QHY	-10.21	62.11 ±1.89	50.28 ±1.42	-10.17 ±0.05	-9.91 ±0.05
2FDD	-11.35	78.15 ±1.29	50.23 ±1.77	-11.55 ±0.06	-11.94 ±0.06
1D4J	-11.41	78.51 ±1.42	59.81 ±1.86	-10.89 ±0.05	-10.35 ±0.05
1W5X	-11.46	81.08 ±1.39	58.55 ±1.45	-10.48 ±0.05	-10.95 ±0.05
3OXC	-11.48	78.51 ±1.42	62.55 ±1.26	-11.36 ±0.05	-10.03 ±0.05
3KDD	-11.55	82.45 ±1.32	51.01 ±1.42	-12.40 ±0.05	-11.9 ±0.06
1W5Y	-11.58	81.25 ±1.24	51.11 ±1.23	-11.41 ±0.05	-10.63 ±0.06

2UXZ	-11.58	70.88±1.28	62.72±1.28	-12.33±0.05	-11.80±0.05
1EC0	-11.60	79.23±1.31	56.62±1.41	-12.14±0.05	-11.44±0.05
1OHR	-11.87	84.21±2.09	57.45±1.57	-10.84±0.07	-10.42±0.05
3EKX	-12.84	77.42±1.42	54.05±1.32	-10.99±0.05	-10.55±0.06
2Z4O	-13.06	81.71±1.31	57.71±1.21	-11.40±0.06	-10.88±0.06
3NU3	-13.41	84.68±1.31	51.75±1.67	-10.98±0.05	-10.37±0.05
3H5B	-13.41	89.50±1.29	59.68±1.21	-11.52±0.05	-11.02±0.06
1HXW	-14.77	87.41±1.34	60.21±1.44	-12.88±0.05	-11.86±0.05
1D4Y	-15.15	92.82±1.23	61.76±1.23	-12.06±0.06	-10.99±0.05
3O9G	-16.79	101.94±1.27	72.24±1.41	-12.07±0.06	-11.32±0.06
3O9F	-17.33	103.96±1.27	58.55±1.34	-13.47±0.06	-11.02±0.06
2BPX	-12.83	83.55±1.32	59.87±1.42	-12.42±0.06	-11.08±0.06

*: The binding free energy is evaluated using equation $\Delta G_{exp} = RT \ln K_i$.

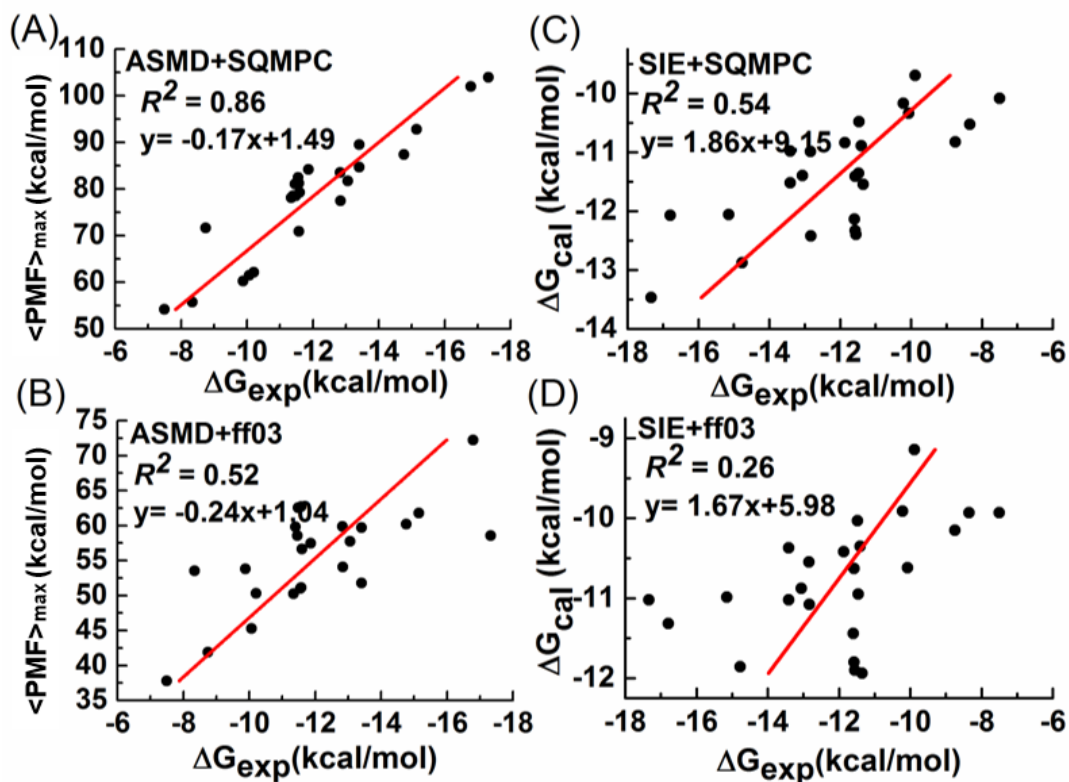


图 2.11 ASMD-SQMP 的准确性评估。ASMD 与 SIE 方法结合 SQMP 或 Amber ff03 电荷计算结果与实验结合自由能的相关性比较。

Figure 2.11 The evaluation on the accuracy of ASMD-SQMP. The correlation between the experimentally determined ΔG and calculated ΔG of ASMD-SQMP (A), SIE-SQMP (B), ASMD-ff03 (C) and SIE-ff03 (D).

表 2.5 $\langle \text{PMF} \rangle_{\text{max}}$ 的体积校正 (kcal/mol)。Table 2.5 The volume correlation of the $\langle \text{PMF} \rangle_{\text{max}}$ (kcal/mol).

PDB ID	ΔG_{exp}	$\langle \text{PMF} \rangle_{\text{max}}$	$G(r_0) - G(r_{\infty})$	Correction	$\Delta G_{\text{correction}}$	MUE*
4U7V	-7.50	54.15 $\pm$ 1.61	-54.15 $\pm$ 1.61	4.19	-49.96 $\pm$ 1.61	42.46
2QNP	-8.34	55.71 $\pm$ 1.93	-55.71 $\pm$ 1.93	4.12	-51.59 $\pm$ 1.93	43.25
2QNP	-8.75	71.61 $\pm$ 2.21	-71.61 $\pm$ 2.21	4.11	-67.50 $\pm$ 2.21	58.75
2PSV	-9.88	60.23 $\pm$ 1.42	-60.23 $\pm$ 1.42	4.21	-56.02 $\pm$ 1.42	46.14
2QIO	-10.07	61.44 $\pm$ 1.44	-61.44 $\pm$ 1.44	4.16	-57.28 $\pm$ 1.44	47.21
2QHY	-10.21	62.11 $\pm$ 1.89	-62.11 $\pm$ 1.89	4.10	-58.01 $\pm$ 1.89	47.8
2FDD	-11.35	78.15 $\pm$ 1.29	-78.15 $\pm$ 1.29	4.21	-73.94 $\pm$ 1.29	62.59
1D4J	-11.41	78.51 $\pm$ 1.42	-78.51 $\pm$ 1.42	4.18	-74.33 $\pm$ 1.42	62.92
1W5X	-11.46	81.08 $\pm$ 1.39	-81.08 $\pm$ 1.39	4.19	-76.89 $\pm$ 1.39	65.43
3OXC	-11.48	78.51 $\pm$ 1.42	-78.51 $\pm$ 1.42	4.15	-74.36 $\pm$ 1.42	62.88
3KDD	-11.55	82.45 $\pm$ 1.32	-82.45 $\pm$ 1.32	4.19	-78.26 $\pm$ 1.32	66.71
1W5Y	-11.58	81.25 $\pm$ 1.24	-81.25 $\pm$ 1.24	4.30	-76.95 $\pm$ 1.24	65.37
2UXZ	-11.58	70.88 $\pm$ 1.28	-70.88 $\pm$ 1.28	4.12	-66.76 $\pm$ 1.28	55.18
1EC0	-11.60	79.23 $\pm$ 1.31	-79.23 $\pm$ 1.31	4.21	-75.02 $\pm$ 1.31	63.42
1OHR	-11.87	84.21 $\pm$ 2.09	-84.21 $\pm$ 2.09	4.30	-79.91 $\pm$ 2.09	68.04
3EKX	-12.84	77.42 $\pm$ 1.42	-77.42 $\pm$ 1.42	4.15	-73.27 $\pm$ 1.42	60.43
2Z4O	-13.06	81.71 $\pm$ 1.31	-81.71 $\pm$ 1.31	4.15	-77.56 $\pm$ 1.31	64.50
3NU3	-13.41	84.68 $\pm$ 1.31	-84.68 $\pm$ 1.31	4.36	-80.32 $\pm$ 1.31	66.91
3H5B	-13.41	89.50 $\pm$ 1.29	-89.50 $\pm$ 1.29	4.30	-85.20 $\pm$ 1.29	71.79
1HXW	-14.77	87.41 $\pm$ 1.34	-87.41 $\pm$ 1.34	4.22	-83.19 $\pm$ 1.34	68.42
1D4Y	-15.15	92.82 $\pm$ 1.23	-92.82 $\pm$ 1.23	4.26	-95.56 $\pm$ 1.23	80.41
3O9G	-16.79	101.94 $\pm$ 1.27	-101.94 $\pm$ 1.27	4.34	-97.60 $\pm$ 1.27	80.81
3O9F	-17.33	103.96 $\pm$ 1.27	-103.96 $\pm$ 1.27	4.36	-99.60 $\pm$ 1.27	82.27
2BPX	-12.83	83.55 $\pm$ 1.32	-83.55 $\pm$ 1.32	4.26	-79.29 $\pm$ 1.32	66.46

*. The MUE is between the $\Delta G_{\text{correction}}$ and ΔG_{exp} (kcal/mol)

为进一步评估 ASMD-SQMPC 的预测效果，我们同时比较了另外三种方法：方法 1 是在常规 ff03 力场下进行 ASMD 模拟；方法 2 是在上述提到的 SQMPC 与基于 MD 的打分函数结合，即 SIE-SQMPC 方法；方法 3 是在常规 ff03 力场下进行 SIE 计算。如图 2.10 所示，这三种方法得到的 R^2 依次为 0.52、0.54 和 0.26。因此，ASMD-SQMPC 比这三种方法更有优势。此外我们也能发现 ASMD 方法得到的 R^2 高于基于 MD 的打分函数方法 SIE，而且 SQMPC 能够明显提高常规

方法（SIE 和 ASMD 等）的计算准确性。

在 24 个 HIV-1 PR 抑制剂体系中有 8 个含氟抑制剂（2QI0、2QHY、1D4J、1W5X、3KDD、1W5Y、1EC0 和 1D4Y）。在药物化学领域，氟原子作为最小的卤素原子有它独特的物理化学性质，常用来改造化合物活性^[223-224]。因此，我们评估了这 8 个含氟抑制剂的 ASMD-SQMPC 预测情况。如图 2.12 所示，ASMD-SQMPC 得到的 $\langle \text{PMF} \rangle_{\text{max}}$ 与结合自由能的 R^2 也达到了 0.74，而另外三种方法的 R^2 依次为 0.51，0.37 和 0.31。因此，结果表明 ASMD-SQMPC 也能够很好地区分含氟抑制剂的生物活性。

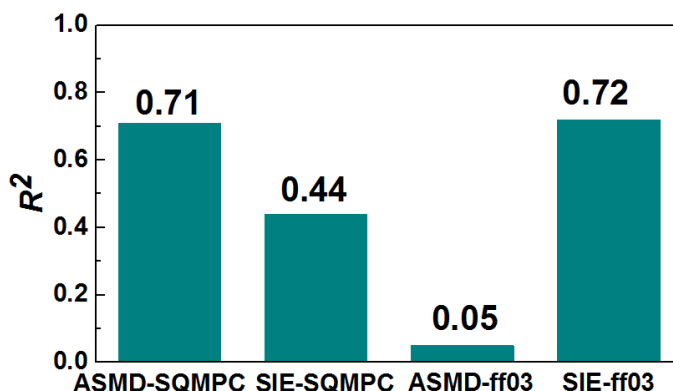


图 2.12 含氟抑制剂的 $\langle \text{PMF} \rangle_{\text{max}}$ 与结合自由能实验值的相关性。

Figure 2.12 The correlation between the experimentally determined ΔG and the $\langle \text{PMF} \rangle_{\text{max}}$ for the fluorine-containing inhibitors.

在 ER 体系中，我们发现配体周围的氨基酸残基受到的静电极化效应更明显，其 SQMPC 电荷具有明显差异。因此，为探究 ASMD 模拟中蛋白质的 SQMPC 电荷是如何变化的，我们计算并分析了 ASMD 模拟轨迹中蛋白质原子 SQMPC 电荷的变化情况，并将电荷变化明显的原子绘制成图。如图 2.13 所示，配体在解离过程中，会与蛋白发生短暂性的相互作用。HIV-1 PR 蛋白构象也随着配体解离发生改变，变化最为明显的地方位于蛋白酶活性位点外周的两端 flap 结构（残基 35-57），当配体即将离开结合位点时（蛋白-配体距离在 15-17 Å），该 flap 区域向外打开，而当配体解离后该 flap 区域向内闭合（图 2.14）。而且伴随着配体的离开，在解离路径上的氨基酸残基的原子电荷依次发生明显变化。例如，在 ASMD 模拟区间 1，配体位于 HIV-1 PR 活性位点处，主要与 D25A 和 D25B 形成氢键作

用,而此时也正是这两个残基在整个模拟阶段 SQMPC 电荷变化最显著的时期。在 ASMD 模拟区间 6,配体与蛋白活性位点距离 27 Å 左右,配体位于 HIV-1 PR 表面,此时位于蛋白表面的 K55B 发生着最明显的变化。这一结果表明,蛋白原子电荷随着配体的解离会相应地发生电荷重新分配,这就能够解释通过更新 SQMPC 电荷能够提高 ASMD 模拟预测蛋白-配体结合自由能计算准确性。

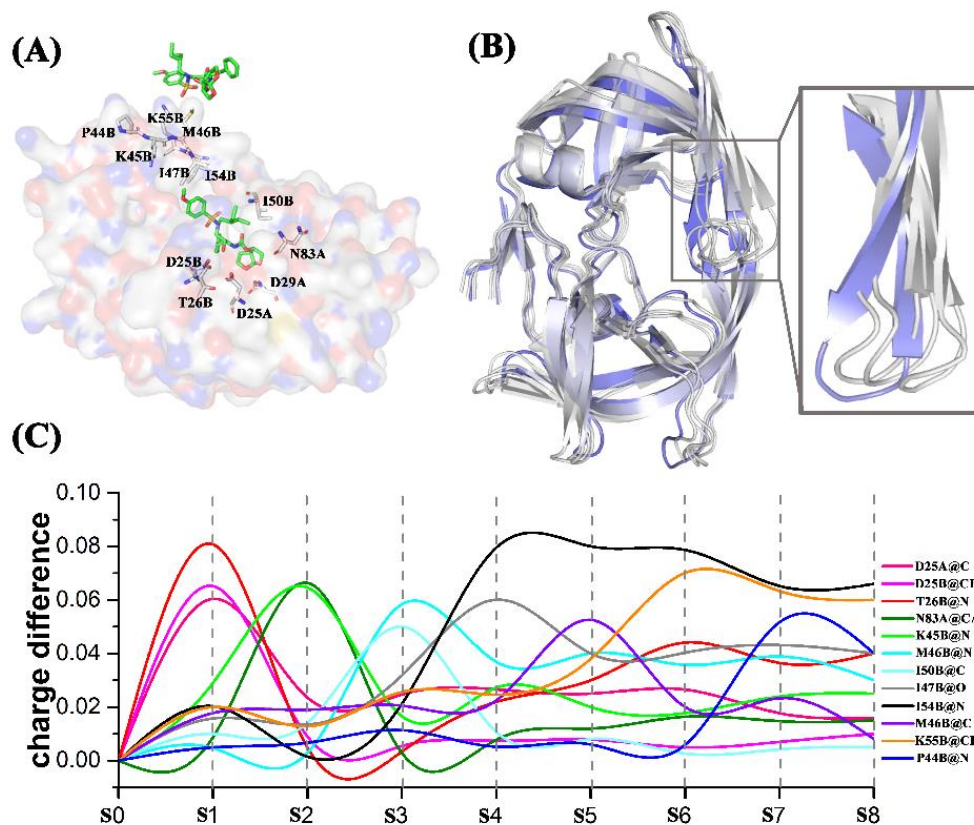


图 2.13 配体解离过程中的蛋白质电荷重分配。以 3O9F 为例, (A) HIV-1 PR 抑制剂解离路径附近的残基展示, (B) HIV-1 PR 蛋白的构象变化, (C) 在配体解离过程中, SQMPC 与 Amber ff03 电荷的差异, 其中 SQMPC 电荷为每个模拟区间更新的 SQMPC 电荷, “D25A@C”表示 A 链的 D25 的主链 C 原子 (Amber ff03 力场命名方式)。

Figure 2.13 The changes of partial atomic charges during the ligand dissociation process. Taking 3O9F as an example, (A) representation of residues along the HIV-1 PR inhibitor unbinding pathway, (B) representation of the conformational change of HIV-1 PR, (C) the difference between SQMPC and Amber ff03 charges of representative residues during the ligand dissociation process. The charge difference was calculated as a partial atomic charge difference between ff03 force field and updated SQMPC. For example, the symbol “D25A@C” refers to the C atom in D25 of chain A named by Amber ff03 force field.

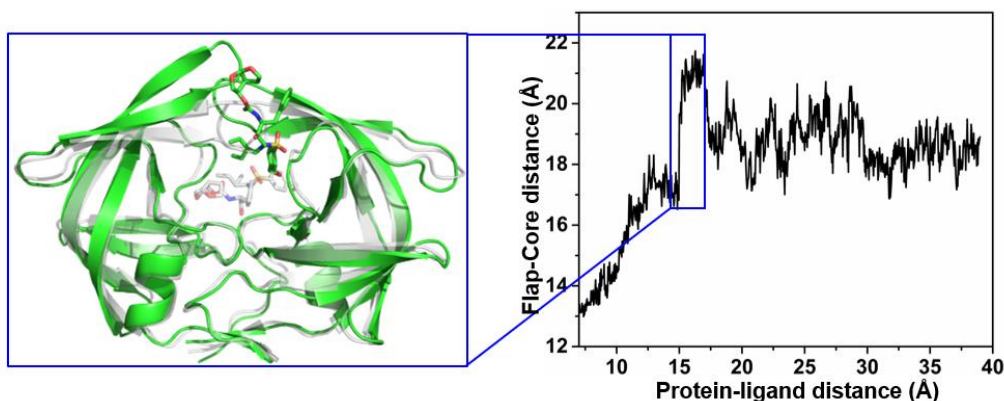


图 2.14 配体解离时蛋白的 flap 动态变化。蛋白-配体距离由 HIV-1 PR 活性残基 D25A 和 D25B 与配体的质心距离决定。Flap-蛋白中心距离由 HIV-1 PR 的 flap 顶端残基 (I50A 和 I50B) 与活性残基的质心距离决定。

Figure 2.14 Flaps' dynamics in response to ligand dissociation. Protein-ligand distance was determined by the distance between the mass centers of HIV-1 PR binding cavity (D25A and D25B) and the inhibitor. The flap-core distance was determined by the distance between the mass centers of HIV-1 PR flaps tips (I50A and I50B) and HIV-1 PR binding cavity.

在本章研究中，对于每个 HIV-1 PR 抑制剂体系，ASMD-SQMPC 模拟共分为 8 个模拟区间，每个模拟区间 25 条平行轨迹，共计 80 ns。24 个体系的 $\langle \text{PMF} \rangle_{\text{max}}$ 不确定性占其平均值的 1.8%，表明 ASMD-SQMPC 模拟具有结果稳定性。近年来，通过非平衡分子动力学模拟预测配体结合自由能的研究屡见报道。但目前报道的方法通常需要消耗较多的计算资源。例如 Li 等人运行了 6.5 μs 的 MD 模拟去预测 HIV-1 PR 抑制剂的亲和力；Ngo 等人也花费了 352 ns 的模拟时间去计算 HIV-1 PR 抑制剂的亲和力。因此，从目前的结果看 ASMD-SQMPC 有希望成为一个快速准确的蛋白-配体结合自由能计算的有力工具。

2.4 本章小结

在过去的几十年里，新的蛋白-配体结合自由能计算方法层出不穷，然而能够应用于高通量药物虚拟筛选的快速准确的计算方法并不多。近年来，基于力场的打分函数尽管计算速度快，但受限于其准确性和体系可拓展性；而自由能微扰和热力学积分等较为准确的方法在计算上往往相当耗时。因此，我们希望通过提高打分函数的准确性来达到准确性和计算效率上的平衡。在本章中，我们通过半

经验量子化学方法计算了蛋白-配体复合物体系的原子电荷来精确考虑蛋白质受到的静电极化效应,从而发展了 SQMPC 方法,并将其与结合自由能计算方法 SIE 与 MM/GBSA 相结合。应用于 50 个蛋白-配体复合物中发现, SQMPC 均可以取得比基于常规力场 AMBER-ff03 更准确的自由能计算结果,其中 AM1 和 PM7 在测试的四种 SQM 方法中表现最佳。甚至对于一系列具有不同骨架的配体, SQMPC 也能提高 SIE 和 MM/GBSA 方法的计算准确性。基于分子动力学模拟预测蛋白-配体结合自由能的方法,如自适应拉伸分子动力学 (ASMD),往往忽略了配体结合或解离过程中蛋白质受配体及溶剂的静电极化效应而影响其准确性。因此,为提高 ASMD 方法的准确性,我们将发展的 SQMPC 方法与 ASMD 结合,定期更新 ASMD 模拟中蛋白-配体体系的原子电荷,进一步发展了 ASMD-SQMPC 方法。将该方法应用于 24 个 HIV-1 蛋白酶抑制剂中发现, SQMPC 更新的 ASMD 方法 (ASMD-SQMPC, $R^2=0.86$) 能显著提高常规 ASMD ($R^2=0.52$) 的结合自由能预测准确性。因此,这项工作提出了 2 种能准确有效地预测蛋白-配体结合自由能的新策略 (SIE-SQMPC 和 ASMD-SQMPC),同时可为高通量处理分子对接和药物设计提供帮助。

第3章 蛋白质构象高效采样方法的发展及应用

3.1 研究背景

蛋白质的构象变化发生在诸多重要的生物过程中,例如蛋白质折叠^[225],蛋白-配体结合^[226]和蛋白-蛋白相互作用^[227]等,因而研究靶标蛋白的动态行为对理解其生物学功能至关重要,同时也为基于结构的药物发现与设计提供关键信息^[228-231]。分子动力学模拟(MD)能够在原子水平上观察物质的结构和动力学性质,已广泛应用于化学、物理、生物、材料等多种学科^[232]。随着计算硬件的迅猛发展(如 Anton 专用超级计算机和 GPU)以及计算算法的不断改进,MD 模拟能达到的时间尺度明显提高,能应对的生物体系也日益复杂。但由于生物体系的势能面复杂多样且崎岖不平,传统的 MD 模拟方法容易陷入势能面的局部极小值点,构象采样时需要花费巨额计算成本且收效可能甚微^[233]。因此,在有限的计算机资源下,为了克服 MD 模拟低采样效率问题,发展高效准确的分子模拟方法是药物设计领域中一个亟待解决的问题。

增强型采样算法旨在提高 MD 的采样效率,其原理可大致分成两类。一类是通过在集合坐标(collective variables, CVs)上施加偏执力,来帮助体系克服能垒,典型方法包括 metadynamics (MetaD)^[234]、伞形采样^[235]、靶向分子动力学^[236]、自适应偏执力^[237-238]、构象涨跃动力学^[239]和拉伸分子动力学^[240]等。其中 MetaD 方法是近年来生物大分子体系中备受关注和广泛使用的构象采样方法^[241-245]。在 MetaD 方法中,回溯性偏执力周期性施加于体系的 CVs 空间,相当于填充体系势能阱,从而使其克服能垒,探索更广阔的自由能景(free energy surface, FES)。但 MetaD 的缺陷在于 FES 的难以收敛^[246],因此, Parrinello 等人继续发展了 well-tempered metadynamics (WT-MetaD)^[247]方法来改善其收敛性问题。此外,对这类依赖 CVs 的增强采样方法而言,如何选择合适的 CVs 又是另一挑战。目前常用的 CVs 有二面角和原子距离等^[248],但在复杂的蛋白质构象变化以及配体结合、解离过程的模拟中适用性低。此外,施加的偏执力会将构象采样局限于 CVs 空间内,使反应路径预测受到限制,尤其是对于势能面崎岖的模拟体系,反应能垒很可能不存在于选定的 CVs 空间内,而存在于其他 CVs 空间,即容易出现隐藏能垒的问题。除此之外,诸多 MetaD 模拟应用研究表明,主成分分析(PCA)

[246, 249-251]和简正振动模式分析 (normal mode analysis, NMA) [252-256]也是有效的 CVs。前者可帮助分析蛋白质的运动趋势, 后者基于谐波理论常用于预测蛋白质构象转变路径。此外, NMA 相比 PCA 需要的计算资源明显更少^[257], 有望成为基于偏执力的这类增强型采样算法的合适 CVs^[258-259]。

温度型副本交换方法 (temperature replica exchange MD, T-REMD) 最早于 1999 年由 Okamoto 和 Hansmann 等人提出, 并应用于蛋白质折叠的蒙特卡罗模拟^[260]。T-REMD 的基本思想是一系列在不同温度下采样的平行副本, 通过在系统能量空间上的构象交换, 达到跨越能垒和充分采样的目的。T-REMD 目前广泛应用于蛋白质折叠^[260]、蛋白质功能性构象变化以及蛋白质-配体相互作用研究中^[261]。例如, Hansmann 等人利用 T-REMD 成功将含有 67 个残基的 Top67 蛋白折叠成与晶体结构非常接近的结构 (RMSD~3Å)。但 T-REMD 的缺陷在于其所需副本数目与体系自由度的平方根成正相关性, 对于显式溶剂的模拟尤其需要大量的计算资源 (对于蛋白原子数大于 5,000 体系, 副本数往往超过 100)^[262]。为此, 许多 T-REMD 优化方法通过处理溶剂自由度等方式来减少 T-REMD 所需的计算资源^[263-264]。例如, 在杂化 REMD 方法中, 体系在显式水溶剂中进行 MD 模拟, 在副本交换时使用隐式溶剂势能, 从而避免了体系中溶剂水的自由度对副本数目的影响, 显著减少了副本数目, 提高了交换效率^{[265],[266]}。另外, Berne 等人通过对模拟体系溶质和溶剂的哈密顿能量调整, 提出了 REST 方法 (replica exchange with solute tempering)^[267]。在 REST 中, 溶质原子相当于在温度梯度下模拟, 而溶剂原子相当于在基线温度下模拟, 从而使得副本交换概率计算中的溶剂内部作用能被消除。目前 REST 有多种变形方法, 其中 REST2^[268]是最常用的版本。

T-REMD 由于原理简单且采样效率可观, 因此也常与其他增强采样方法相结合, 组成多维 REMD, 进一步提高分子模拟的采样效率^[269-272]。例如, Yuko Okamoto 等人提出的副本交换伞形采样^[269-270]通过整合 T-REMD 与伞形采样, 使得平行副本在不同的谐振势和温度下模拟。PTMetaD 方法是一种整合 MetaD 和 T-REMD 的多维增强型采样算法^[241, 273-278], 广泛应用于构象动力学和蛋白质折叠等研究^[275-278]。例如, Marino 等人通过长时间 PTMetaD 模拟, 观察到了 B-Raf 激酶野生型和 V600E 突变体的活性/非活性构象转变并发现 V600E 突变体会增强酶活性^[278]。但 Parrinello 等人也指出, PTMetaD 方法的主要缺陷在于其有限大小的模拟体系^[277], 即在相同温度范围内, PTMetaD 方法所需副本数目与 T-REMD

相近。因此，为进一步减少 PTMetaD 方法所需计算资源以及提高其采样效率，我们对 PTMetaD 方法进行优化，将 NMA 作为 CVs 进行加速采样，并且在副本交换时去除溶剂内部作用能，发展了一种可以在能量空间和几何空间同时加速的二维增强型蛋白质构象采样方法，并称之为 ossPTMetaD (omit solvent-solvent interactions PTMetaD)。为了评估 ossPTMetaD 方法的可靠性和有效性，我们首先选取腺苷酸激酶 (AdK) 作为模型体系进行测试，结果表明 ossPTMetaD 模拟得到的 AdK 构象分布及转变自由能与目前的实验数据和常规 T-REMD 模拟结果相符，并且 ossPTMetaD 只需常规 T-REMD 计算资源的 30%。进一步地，我们将 ossPTMetaD 应用到另外四种具有代表性的靶标蛋白构象变化研究中，结果显示 ossPTMetaD 可在有限的计算资源内模拟这些蛋白的大规模构象变化，说明 ossPTMetaD 方法的可靠性、高效性和体系可拓展性。

3.2 材料与方法

基于温度的 REMD (T-REMD)。T-REMD 也称为平行回火 (parallel tempering)，即 N 个副本在梯度温度 T_n ($n=1, 2, \dots, N$) 下平行模拟，并定期交换各副本的模拟温度，使得每个副本均能在最高温和最低温间反复遍历，达到增强采样的效果。在 T-REMD 模拟过程中，体系哈密顿量 H (粒子数 n) 为依赖原子坐标 (x) 的体系势能 E_{pot} 和依赖原子速度 (v) 的体系动能 E_{kin} 之和 (公式 3.1)。

$$H(x, v) = E_{pot}(x) + E_{kin}(v) \quad \dots (3.1)$$

$$E_{kin}(v) = \frac{1}{2} \sum_i^n m_i v_i^2 \quad \dots (3.2)$$

根据 Metropolis 准则，T-REMD 定期计算相邻温度副本 (T_1 和 T_2) 之间的交换概率，以判断这两个副本能否发生交换 (公式 3.3)，其中 $\beta=1/k_B T_n$ 为温度因子 (k_B 为玻尔兹曼常数)。

$$w(1 \leftrightarrow 2) = \exp(\Delta\beta\Delta H) = \exp(\Delta\beta\Delta E_{pot} + \Delta\beta\Delta E_{kin}) \quad \dots (3.3)$$

为保证副本成功交换后原子速度依然服从麦克斯韦分布，T-REMD 通过公式 3.4 对体系动能进行调整，其中 $\hat{E}_{kin}^{(1)}$ 是副本 T_1 交换到 T_2 的动能， $\hat{E}_{kin}^{(2)}$ 是副本 T_2 交换到 T_1 的动能。

$$\left\{ \hat{E}_{kin}^{(1)} = \frac{T_2}{T_1} E_{kin}^{(1)} ; \hat{E}_{kin}^{(2)} = \frac{T_1}{T_2} E_{kin}^{(2)} \right\} \quad \dots (3.4)$$

经公式 3.4 的调整，公式 3.3 中相邻副本交换前后的动能差被抵消，因此 T-REMD 的副本交换概率只由体系势能决定（公式 3.5）。

$$\omega(1 \leftrightarrow 2) = \exp(\Delta\beta\Delta E_{pot}) \quad \dots(3.5)$$

PTMetaD 方法。PTMetaD 是一种整合 MetaD 和 T-REMD 的多维 REMD 方法。在 PTMetaD 中，平行副本在一个温度梯度和额外限制势 $V(s, t)$ 作用下模拟，其中限制势是在 t 时间内施加于 CVs 空间 (s) 的限制势总和。因此，体系哈密顿量可表达为体系势能、动能和额外限制势能之和（公式 3.6）。

$$H(x, v) = E_{pot}(x) + E_{kin}(v) + V(s, t) \quad \dots(3.6)$$

在 PTMetaD 中，额外限制势 $V(s, t)$ 以高斯函数形式施加于模拟体系（公式 3.7），其中 ΔT 和 ω 分别为温度和势能增长率。

$$V(s, t) = \Delta T \ln \frac{\omega \int_0^t \delta_{s, s(t')} dt'}{\Delta T} \quad \dots (3.7)$$

额外限制势的高斯函数高度 W 根据 $W = \omega e^{-\frac{V(s, t)}{\Delta T}} \tau_G$ 计算得到，其中时间间隔 τ_G 设置为 1 ps。自由能面 $F(s, t)$ 根据公式 3.8 计算，其中偏执因子 $(T + \Delta T)/T$ 设置为 6.0。

$$F(s, t) = -\frac{T + \Delta T}{\Delta T} V(s, t) \quad \dots(3.8)$$

最终，在 PTMetaD 方法中，相邻副本 (T_1 和 T_2) 之间的交换概率由体系势能和额外限制势决定，其中 $\hat{V}_{(s, t)}^{(1)}$ 和 $\hat{E}_{pot}^{(1)}$ 分别是副本 T_1 交换为副本 T_2 时的额外限制势和体系势能，同理 $\hat{V}_{(s, t)}^{(2)}$ 和 $\hat{E}_{pot}^{(2)}$ 分别是副本 T_2 交换为副本 T_1 时的额外限制势和体系势能。

$$w(1 \leftrightarrow 2) = \min(1, \exp(D)) \quad \dots (3.9)$$

$$\begin{aligned} D &= \Delta\beta\Delta E_{pot} + \Delta\beta\Delta V \\ &= -\beta_2 \hat{E}_{pot}^{(1)} - \beta_2 \hat{V}_{(s, t)}^{(1)} - \beta_1 \hat{E}_{pot}^{(2)} - \beta_1 \hat{V}_{(s, t)}^{(2)} + \beta_2 E_{pot}^{(2)} + \beta_2 V_{(s, t)}^{(2)} + \beta_1 E_{pot}^{(1)} + \\ &\quad \beta_1 V_{(s, t)}^{(1)} \quad \dots(3.10) \end{aligned}$$

ossPTMetaD 方法。体系势能根据原子组成可分成蛋白内部相互作用能 (P_{pp})、蛋白-溶剂相互作用能 (P_{pw}) 和溶剂内部相互作用能 (P_{ww})。

$$E_{pot} = P_{pp} + P_{pw} + P_{ww} \quad \dots (3.11)$$

公式 3.11 中的溶剂原子包括体系平衡离子（钠离子和氯离子）。我们之前发展的增强型采样算法 vsREMD^[266]表明，在副本交换时去除 P_{ww} 能够显著减少 T-

REMD 所需副本数目。因此，我们基于 PTMetaD 方法去除 P_{ww} 对副本交换率的影响，进一步减少 PTMetaD 的计算资源，并采用 NMA 作为 WT-MetaD 部分的 CVs 来加速采样效率，从而发展了 ossPTMetaD 方法。我们通过经 PLUMED 2.3 修复过的 GROMACS 5.1.4 软件中进行代码修改以实现 ossPTMetaD 算法。

在 ossPTMetaD 方法用到的 NMA，由本征向量 e 计算得到，其中 R 为 NMA 预测结构， R^0 是用来计算本征向量 e 的参考结构。

$$s = (R - R^0) \cdot e \quad \dots (3.12)$$

本研究中， e 是根据公式 3.13 得到的多个低频正则模式本征向量的线性组合，其中 u_i 是通过 *iMOD* 软件^[279] 计算的第 i 个低频正则模式的特征向量，步长因子 k 设置为 15，正则模式组合数 m 为 6^[280-281]。

$$e = k \sum_i^m u_i \quad \dots (3.13)$$

模拟体系准备。在本研究中，5 种代表不同蛋白运动类型的靶标体系被用于测试评估 ossPTMetaD 方法，它们分别是：具有结构域大规模相对运动的腺苷酸激酶（AdK）、具有 flap 开放/闭合构象变化的 HIV-1 蛋白酶（HIV-1 PR）和 β -位点 APP 剪切酶 1（BACE1），以及具有 DFG 模式翻转运动的丝裂原活化蛋白激酶（p38 α ）和酪氨酸激酶 c-Abl。这些蛋白都是重要的药物靶标蛋白，同时由于其发挥功能时伴随着明显的构象变化，因此也是基于结构的药物设计中备受关注兼具挑战的靶标体系。表 3.1 和表 S3.1 列出了这 5 个模拟体系的原子数和模拟温度等具体计算细节。RSFF2（Residue-Specific Force Field）力场^[282]和 GAFF^[283]力场分别用来参数化蛋白和小分子抑制剂。为模拟溶剂环境，模拟体系被置于 TIP3P 的立方体水盒子中，盒子边缘距离溶质最短距离为 10 Å。采用 LINCS 算法限制氢原子连接共价键^[284]，PME^[285]算法处理长程静电相互作用，截断值为 12 Å。每个模拟体系首先经最速下降法进行能量最小化，之后在 NPT 系综下升温至 300 K 并平衡 2 ns，升温和平衡过程均被施加大小为 10 kcal/mol/Å² 的弱限制力。ossPTMetaD 模拟在 GROMACS 5.1.4 软件中进行，并安装 PLUMED 2.3 插件^[286]，每隔 2 ps 进行副本交换尝试。对于 AdK、HIV-1 PR 和 BACE1 体系，每个副本模拟 100 ns；而对于 p38 α 和 c-Abl 激酶，每个副本模拟 500 ns 以获得收敛的自由能景。在所有的模拟体系中，温度梯度范围均设置为 300 K 到 450 K。为比较 ossPTMetaD 与其他增强采样方法，以 AdK 体系为例，我们在力场和模拟时间等参数均保持一致的情况下对比了 T-REMD 和 ossPTMetaD 的模拟结果。

表 3.1 ossPTMetaD 和 T-REMD 的模拟体系信息。

Table 3.1 System Information for ossPTMetaD and T-REMD simulations.

	AdK	AdK ^c	HIV-1 PR	BACE1 apo	BACE1 2QU3	BACE1 2WF1	p38 α	c-Abl
Amino acids	214	214	198	398	398	398	354	271
Total Atoms	27084	27084	26023	43663	43704	43956	41739	40743
Solute Atoms	3384	3384	3163	6042	6057	6243	5721	4407
N _{R,T-REMD} ^a	78*	80	75*	98*	98*	98*	96*	96*
N _{R,ossPTMetaD} ^b	24	24	16	30	30	30	28	28
Swap ratio	0.24	0.19	0.23	0.27	0.27	0.27	0.29	0.30

^aT-REMD 所需副本数。^bossPTMetaD 所需副本数。^cT-REMD 模拟实际所用副本数。*由网站 (<http://folding.bmc.uu.se/remd/>) 在副本交换率~0.25 时计算的所需副本数。

3.3 结果与讨论

3.3.1 ossPTMetaD 方法计算效率与采样效率评估

在常规 T-REMD 中, 由于溶剂原子数通常超过模拟体系总原子数的 85%, 为获得合理的总平均副本交换率 (~0.25), 在普遍模拟体系中 T-REMD 所需的副本数目往往超过 100 (如 BACE1、p38 α 和 c-Abl)。而 ossPTMetaD 方法去除了 P_{ww} , 所需副本数目仅为 T-REMD 的 ~30%, 显著减少了 T-REMD 的计算资源。

AdK 是一种在细胞信号转导和能量代谢中扮演重要角色的磷酸转移酶^[287], 由三个主要的结构域组成: CORE 结构域 (残基 1-29、68-117 和 161-214)、LID 结构域 (残基 118-167) 和 NMP 结构域 (残基 30-67)。AdK 具有两种主要的功能性构象, 即开放构象和闭合构象, 这两种构象间的转变由 LID 和 NMP 结构域的相对运动产生, 时间尺度为 μ s 到 ms 级别^[288-289]。以 AdK 闭合构象 (PDB ID: 1AKE^[290]) 为模拟初始结构, 我们通过 ossPTMetaD 模拟预测了 AdK 的开放/闭合构象变化, 并与常规 T-REMD 模拟结果进行对比。

为评估模拟副本是否在温度梯度中反复遍历, 我们分析了副本 T_1 的温度随模拟时间的变化曲线, 如图 3.1A-B 所示, ossPTMetaD 和 T-REMD 的副本 T_1 均在高温和低温 (300~450 K) 之间反复游走, 表明两种方法均能够在温度空间内充分采样。但值得注意的是, ossPTMetaD 的副本 T_1 遍历温度梯度的频率更快, 这

一现象也解释了 ossPTMetaD 的总平均副本交换率(0.24)比 T-REMD 更高(0.19), 说明 ossPTMetaD 在温度空间的采样效率比常规 T-REMD 高。

根据相对开放构象(PDB ID: 4AKE^[291])的分子骨架 RMSD 变化, ossPTMetaD 和 T-REMD 都成功模拟到了开放构象和闭合构象(RMSD 在 7.2 Å 左右)的相互转变, 但 ossPTMetaD 在 RMSD 大于 7.2 Å 的构象空间内采样更多(图 3.1C-D)表明 ossPTMetaD 比常规 T-REMD 能够探索更大的构象空间。

为进一步评估 ossPTMetaD 和 T-REMD 模拟的收敛性, 以 RMSD 为坐标, 我们绘制了 6 个时间梯度(50, 60, 70, 80, 90 和 100 ns)的 PMF。如图 3.1E-F 所示, T-REMD 模拟 70 ns 之后的 PMF 与 100 ns 高度一致, 表明 T-REMD 模拟在 70 ns 时已达到收敛状态, 而 ossPTMetaD 在~50 ns 时就已经达到收敛。因此, 对于 AdK 体系, ossPTMetaD 相对 T-REMD 能更快达到收敛状态。在 ossPTMetaD 模拟中, 开放构象、中间过渡构象和闭合构象的 PMF 分别为 1.45, 2.82 和 2.50 kcal/mol, 与常规 T-REMD (1.40, 2.48 和 2.73 kcal/mol) 相符, 表明 ossPTMetaD 的准确性。

3.3.2 ossPTMetaD 高效准确模拟 AdK 构象变化

以 AdK 三个结构域之间的相对距离为反应坐标, 即 LID 与 CORE 结构域的质心距离(LID-CORE)和 NMP 与 CORE 结构域的质心距离(NMP-CORE)^[292-294], 绘制 AdK 的构象变化自由能图景。在开放构象中, LID-CORE 和 NMP-CORE 距离分别为 30.8 Å 和 22.3 Å^[291]; 而闭合构象中这两个距离分别为 21.0 Å 和 18.5 Å^[290]。如图 3.2A 所示, ossPTMetaD 模拟采样了较大的构象空间, LID-CORE 和 NMP-CORE 距离分别涉及 17.5~37.0 Å 和 17.0~25.0 Å。从能量图中也能够看出, 开放构象相对闭合构象更稳定, 这与实验数据是吻合的^[291, 295]。为进一步评估 ossPTMetaD 采样的可靠性, 我们从 PDB 数据库中收集了 14 个 AdK 不同构象^[292, 296]并将它们映射在预测的自由能图景中。结果表明这 14 个实验构象分布在 AdK 体系的构象转变各个区域, 但都在 ossPTMetaD 和 T-REMD 的自由能图景的低能构象区域内(<1.80 kcal/mol)。ossPTMetaD 得到的开放和闭合构象间的自由能垒为~1.82 kcal/mol, 表明空 AdK 的柔性较高, 在溶液状态下容易发生构象转变, 并且该自由能垒与常规 T-REMD 相符(~1.75 kcal/mol), 表明 ossPTMetaD 和 T-REMD 模拟结果的相似性, 但在 NMP-CORE < 18 Å 等高能构象区域, 两者

得到的构象分布空间还是有差异的。

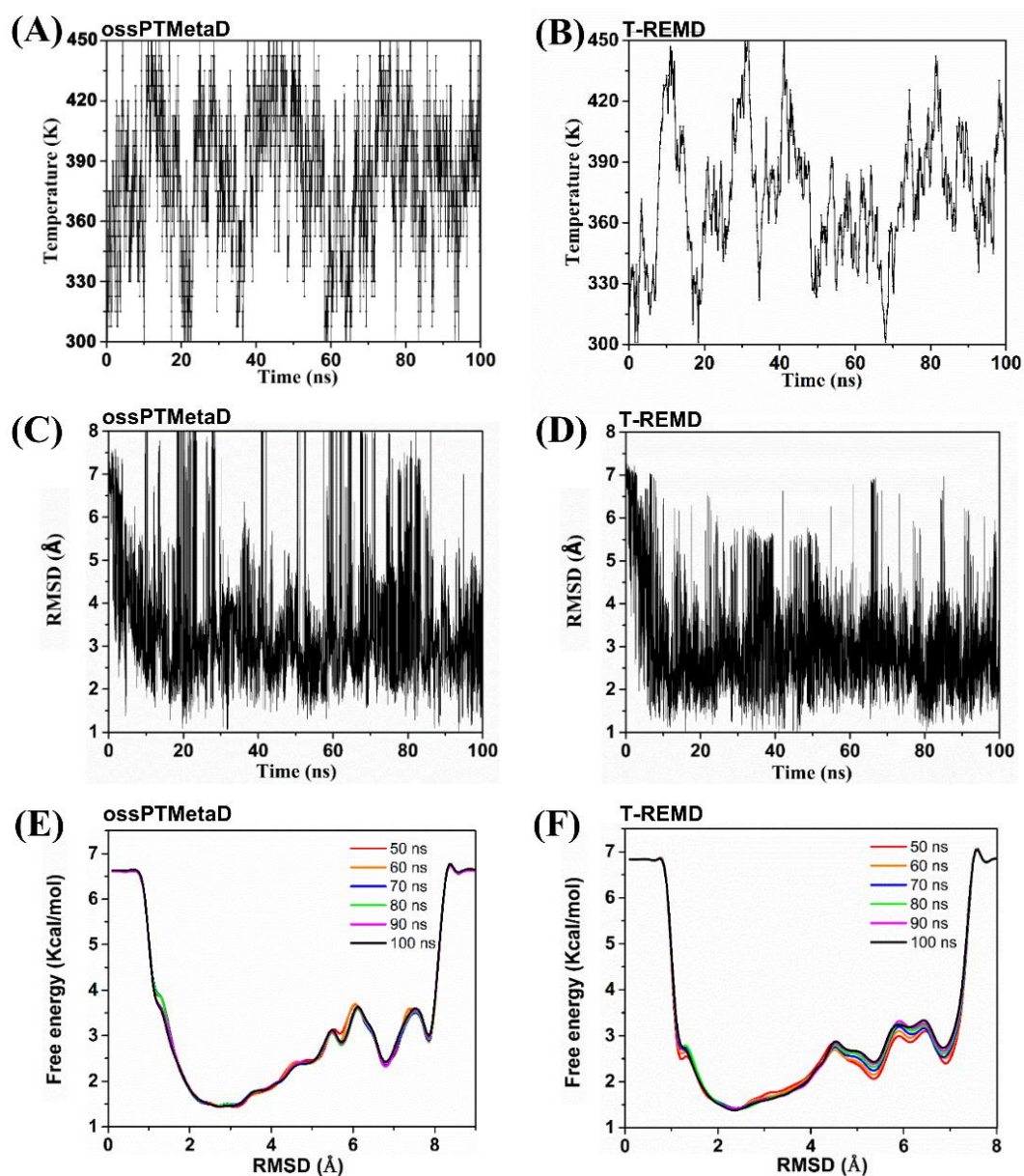


图 3.1 ossPTMetaD 在 AdK 体系上的采样能力评估。ossPTMetaD (A) 和 T-REMD (B) 模拟中副本 T_1 的温度变化。每 10 次交换采集一次数据。ossPTMetaD (C) 和 T-REMD (D) 模拟中相对开放构象的分子骨架 RMSD 变化曲线。ossPTMetaD (E) 和 T-REMD (F) 模拟中以 RMSD 绘制的 6 个不同时间模拟的一维 PMF。

Figure 3.1 The sampling capabilities of ossPTMetaD in AdK. Temperature histories of replica 1 during the ossPTMetaD (A) and T-REMD (B) simulations. The data was plotted every 10 exchanges. The backbone root-mean-square deviation (RMSD) relative to open structure in the ossPTMetaD (C) and T-REMD (D) simulations. One-dimensional potentials of mean force (PMF) as function of the RMSD of six different time periods for the ossPTMetaD (E) and T-REMD (F).

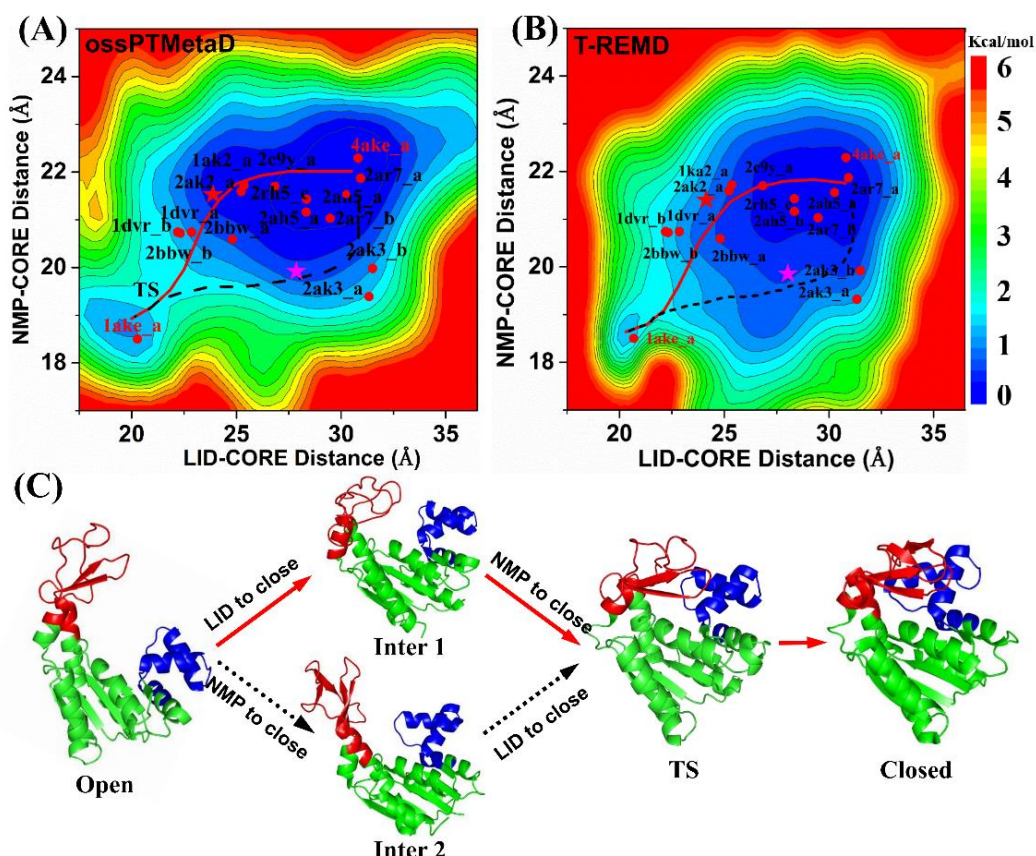


图 3.2 ossPTMetaD 模拟预测的 AdK 构象转变机制。以 LID-CORE 和 NMP-CORE 距离为坐标绘制的 ossPTMetaD (A) 和 T-REMD (B) 二维自由能图景。14 个实验结构以红色圆点映射入自由能图景中，中间构象 (Inter 1 和 Inter 2) 分别标记为红色和粉红色星形。最低能构象转变路径和第二条变化路径分别标记为红色实线和黑色虚线。AdK 的开放-闭合预测转变路径 (C)。

Figure 3.2 Conformational change mechanism of AdK predicted by ossPTMetaD. The two-dimensional free energy landscape as a function of LID-CORE distance and NMP-CORE distance calculated by ossPTMetaD (A) and T-REMD(B). 14 experimental intermediates are projected into the landscape as red dots, the Inter 1 and Inter 2 are labeled as red and pink stars, respectively. The minimum free-energy pathway is labeled as a red line, and the second predicted pathway is labeled as a black dash line. Two pathways of open-to-close transition of AdK (C).

通过 MEPSA 程序，我们计算了 AdK 的最低能构象转变路径 (MFEP) [297]，如图 3.2C 所示，“S”形的构象转变路径表明 AdK 蛋白从开放构象转变至闭合构象过程中，LID 结构域优先闭合，随后 NMP 闭合，这一结果与实验数据[295]和其

他模拟研究报道一致^[288, 292, 298-299]。并且, 我们发现上述 14 个不同的 AdK 实验结构除 2 个结构外 (2AK3 的 A、B 链) 均分布在 MFEP 附近。而 2AK3 的出现表明 AdK 可能存在另一种构象转变路径 (图 3.2 黑色虚线), 即 NMP 结构域优先闭合, 随后 LID 闭合, 这条路径在 Li 等人的长时间 MD 模拟中也被报道过^[288]。

为进一步探索 AdK 构象转变机制, 我们通过聚类分析得到了构象转变的代表性构象 (图 3.2C)。其中一种中间构象的 LID 和 NMP 结构域分别处于半闭合和半开放状态, 称为“Inter 1”, 在许多实验和模拟研究中均被报道^[292, 296, 300], 也是我们预测的 MFEP 路径上的构象。另一种中间构象的 LID 和 NMP 结构域分别处于开放和半闭合状态, 称为“Inter 2”, 位于预测的第二种转变路径附近。

3.3.3 ossPTMetaD 高效准确模拟蛋白 Flap 柔性变化

为评估 ossPTMetaD 方法在其他蛋白质运动类型上的适用性和采样效率, 我们以 HIV-1 PR 和 BACE1 为例, 进一步将其应用于药物设计中普遍出现的蛋白 flap 柔性变化研究。

HIV-1 PR 是 HIV 病毒复制和入侵过程中的关键蛋白酶, 因此是治疗获得性免疫缺陷综合症的重要靶标^[301-302]。大量的晶体结构表明, HIV-1 PR 以对称二聚体形式存在, 每个单聚体各有一段柔性的 β 发夹结构 (称为 flap, 残基 35-57)。当 HIV-1 PR 与底物或抑制剂结合时, 这两段 flap 同时紧闭并覆盖在活性位点上方, 形成闭合的结合空间。但 HIV-1 PR 的空蛋白表现出极为柔性的 flap 结构, 既有闭合状态也有开放状态^[303-305]。因此, HIV-1 PR 的 flap 构象变化对 HIV-1 PR 发挥功能以及药物设计至关重要。以闭合构象为起始结构 (PDB ID: 3EKV^[306]), 我们通过 ossPTMetaD 方法模拟预测了 HIV-1 PR 的构象变化。

通过分析半开放(semi-open)构象的分布比例随模拟时间的变化曲线(图 3.3), 我们发现其分布比例在 80 ns 后基本稳定, 表明 ossPTMetaD 模拟已经收敛。通过分析相对闭合构象的 RMSD 随模拟时间的变化曲线(图 3.4), 我们发现 RMSD 值在 1 到 8 Å 之间反复变换, 表明 ossPTMetaD 模拟采样了较大的 HIV-1 PR 构象空间。并且, HIV-1 PR 的 flap 结构 RMSD 波动明显高于其他结构, 表明 flap 结构更高的柔性, 这一结论与实验数据一致^[307]。

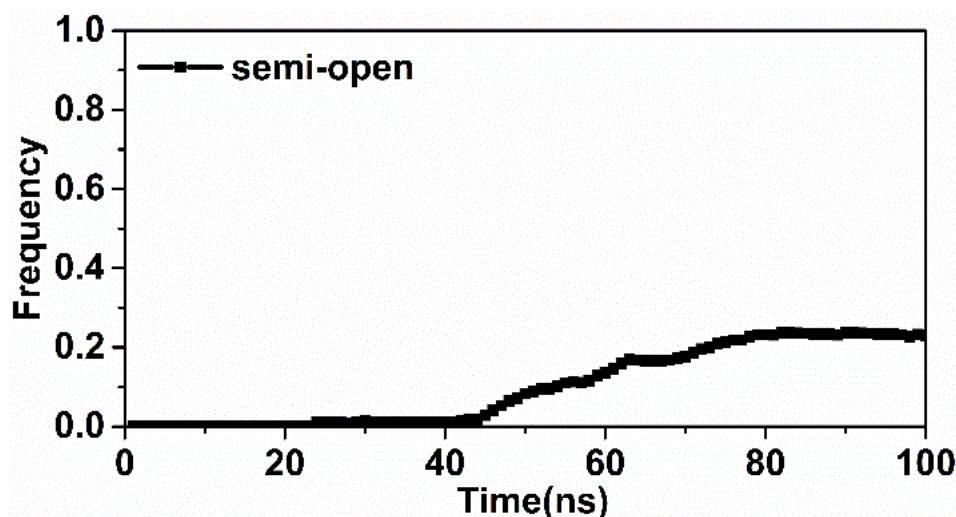


图 3.3 半开放构象的分布比例随模拟时间的变化曲线。相对 X 射线晶体结构（PDB ID: 1TW7, 半开放构象）的骨架 RMSD 小于 2 Å 的结构视为半开放模拟构象。

Figure 3.3 The fraction of semi-open structure along the trajectories. The structures with backbone RMSD from X-ray crystal semi-open structure (PDB ID: 1TW7) within 2 Å are defined as “semi-open” state.

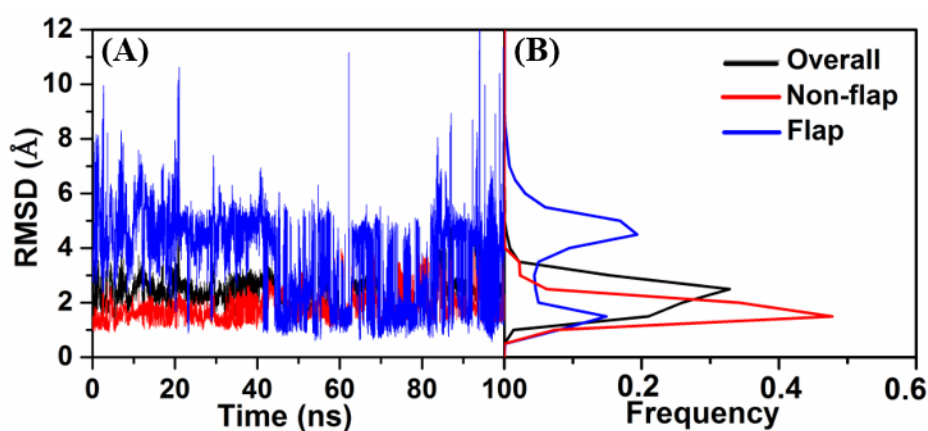


图 3.4 ossPTMetaD 应用于 HIV-1 PR 体系的 RMSD 分析。整体 RMSD、非 flap 结构 RMSD 和 flap 结构 RMSD 随模拟时间的变化曲线（A）及它们的频率分布分析（B）。

Figure 3.4 The RMSD analysis of ossPTMetaD simulation of HIV-1 PR. (A) RMSD of the overall RMSD, non-flap RMSD and flap RMSD from the closed structure of HIV-1 PR. (B) the frequency analysis of these RMSDs.

为分析 HIV-1 PR 的构象变化，我们以相对闭合构象的 HIV-1 PR 整体和 flap 结构 RMSD 为坐标，绘制其二维自由能图景（图 3.5）。结果表明，HIV-1 PR 主要有四种低能构象，即闭合、半开放、内卷曲（curled-in）和完全开放（fully open）

构象,与其他模拟方法的研究报道一致,如 hREMD 和 NMA-ITS^[292,308]。对于闭合构象,整体和 flap 结构的 RMSD 均小于 2 Å,闭合的 flap 结构与 80 号氨基酸残基所在区域形成疏水作用稳定此类构象。对于半开放构象,整体和 flap 结构的 RMSD 分别在 3-4 Å 和 4-6 Å 范围内波动,此时 flap 结构与 80 号氨基酸残基所在区域形成的疏水作用被破坏。与 NMR 实验数据一致^[303,309],二维自由能图景表明闭合构象和半开放构象是 HIV-1 PR 主要的两种低能构象。对于内卷曲构象,整体和 flap 结构的 RMSD 分别位于 4 Å 和 8 Å 附近,flap 结构的顶部以“L”形呈现,这类构象在其他 MD 模拟中也被报道过^[292,310]。在内卷曲构象中,二聚体的两段 flap 顶部(I50)形成内埋的疏水作用,能够稳定此类构象。完全开放构象相对闭合构象具有~4.8 kcal/mol 的自由能垒,因此其构象比例较少,只有 0.8%,但在其他模拟研究^[292,308]和 NMR 试验^[303,311]中也被报道过。

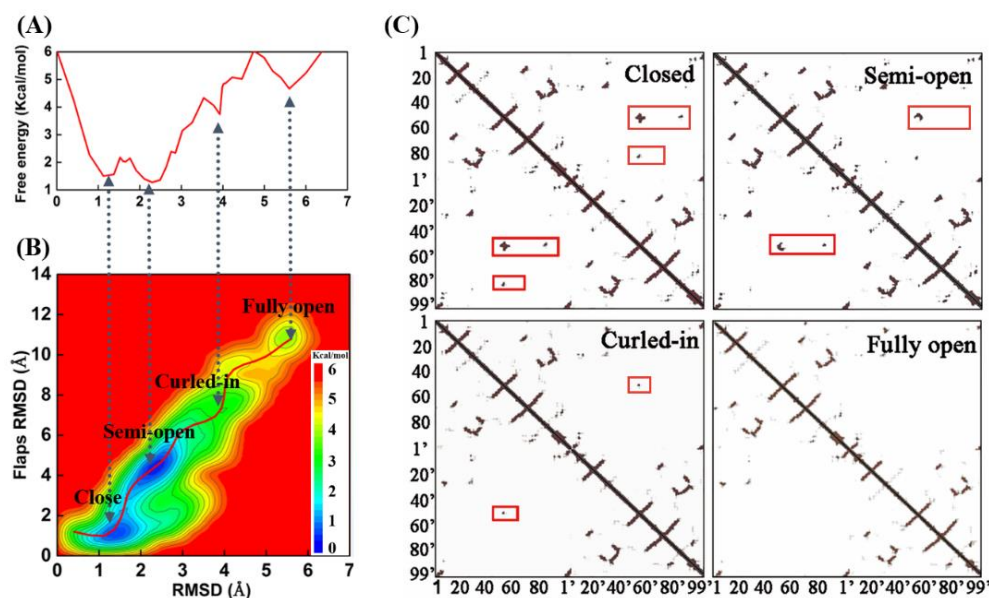


图 3.5 ossPTMetaD 模拟预测的 HIV-1 PR 构象转变。一维自由能曲线 (A) 和二维自由能图景 (B)。最低自由能转变路径标记为红色实线。(C) 闭合、半开放、内卷曲和完全开放代表性构象的残基接触分析,关键疏水作用标记在红色方框内。

Figure 3.5 Conformational change of HIV-1 PR predicted by ossPTMetaD. (A) The one-dimensional free energy profile. (B) and two-dimensional free energy landscape. The minimum free-energy pathway is labeled as a red line. (C) Contact maps of the closed, semi-open, curled-in, and fully open representative structures. Key hydrophobic contacts are highlighted by red squares.

BACE1 被认为是治疗阿尔茨海默病的重要靶点^[312]。晶体结构表明 BACE1

的催化活性残基 (D32 和 D228) 位于一段具有高度柔性的 β 发夹结构之下 (残基 67-75, 称之为 flap) [313-314]。在本章中, 通过 ossPTMetaD 方法, 我们模拟了 BACE1 空蛋白和抑制剂结合复合物的构象变化, 并探索了不同抑制剂对 flap 动态变化的影响。2QP8^[315] 为空蛋白模拟初始结构, 2WF1^[316] (抑制剂 ZY1) 和 2QU3^[317] (抑制剂 462) 为复合物模拟初始结构 (图 S3.1)。

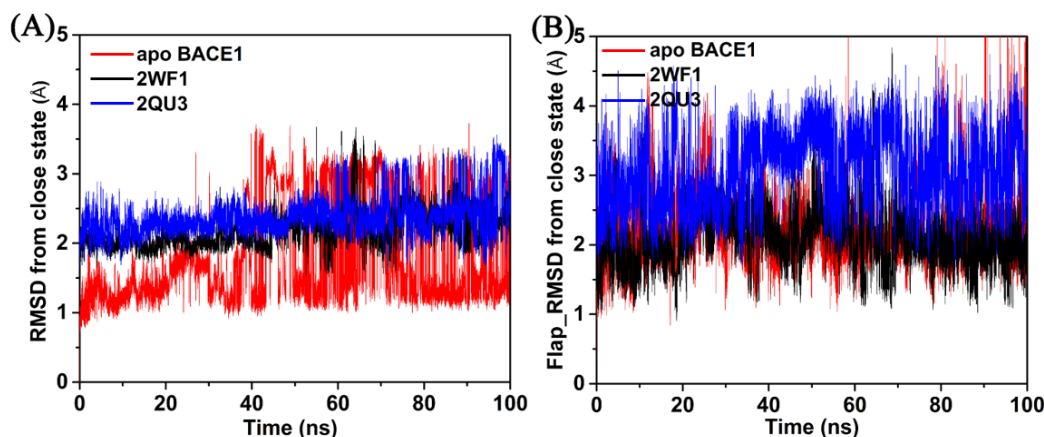


图 3.6 BACE1 整体 (A) 和 flap (B) 结构的相对闭合构象重原子 RMSD。空蛋白的 RMSD 值标记为红色, 两个配体结合复合物, 2WF1 和 2QU3 分别为黑色和蓝色。

Figure 3.6 The heavy-atom RMSD of overall BACE1 protein (A) and flap region (B) from closed structure. The apo simulation is colored red, and two holo simulations of 2WF1 and 2QU3 are colored black and blue.

相对闭合构象的 RMSD 计算结果表明 BACE1 空蛋白在 ossPTMetaD 模拟中经历了明显的构象变化, 尤其是 flap 结构 (图 3.6)。我们以 flap 顶部残基 (Y71) 与催化残基 (D32) 之间的距离 (Flap-Core 距离) 表征 flap 结构的开放/闭合构象变化^[318-321], 进而分析了 BACE1 蛋白 flap 结构的动态特性。一维自由能图景表明 ossPTMetaD 模拟在 70 ns 达到收敛 (图 S3.2)。以 Flap-Core 距离和 flap 结构的 RMSD 为坐标, 我们进一步分析了 BACE1 的二维自由能图景。如图 3.7A 所示, BACE1 空蛋白的 Flap-Core 距离在 6 到 19 Å 都有构象采样, 涵盖闭合构象和完全开放构象。flap 结构的 RMSD 和 Flap-Core 距离分别小于 2 和 9 Å 的构象, 称为闭合状态。在该状态中, flap 结构覆盖在催化位点上方, 从而阻碍了底物进入催化空腔, 因此也称之为自抑制状态^[313]。Flap-Core 距离位于 9 到 13 Å 之间的构象称为过渡构象, 这类构象在实验中经常看到^[322-324]。而当 Flap-Core 距

离增大至 13 Å 时, BACE1 表现为开放构象, 此时催化位点暴露在溶液中。我们从 PDB 数据库中收集了 16 个 BACE1 的不同构象, 结果表明它们都在 ossPTMetaD 采样低能区间, 且位于预测的最低能构象变化路径附近, 表明 ossPTMetaD 模拟的可靠性。另外值得注意的是, 我们在 Flap-Core 距离~15 Å 处发现一种低能构象, 相比开放构象, 该状态的 flap 提供了更大的通道帮助底物或抑制剂的结合。而在 2WF1 和 2QU3ossPTMetaD 模拟中, flap 的运动受到了明显的限制, 它们的 flap 分别被限制在闭合和开放两种不同的构象中 (图 3.7B-C)。

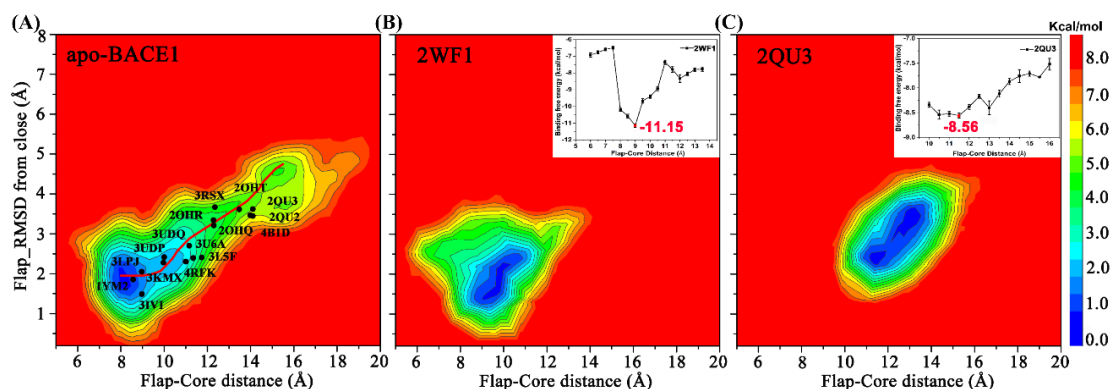


图 3.7 ossPTMetaD 模拟预测 BACE1 构象变化。A、B 和 C 分别为空蛋白、2WF1 和 2QU3 复合物。最低能构象转变路径标记为红色实线。在抑制剂结合复合物体系, 通过 sietraj 程序计算了配体结合自由能随 Flap-Core 距离的变化曲线。

Figure 3.7 Conformational change of BACE1 predicted by ossPTMetaD. The free energy landscape of conformational change in apo (A), and bound state of 2WF1 (B) and 2QU3 (C). The minimum free-energy pathway is labeled as a red line. In the two BACE1-inhibitor systems, the change of binding free energy along the conformational change in respect of the Flap-Core distance was calculated by sietraj program.

为分析两种不同结合模式的抑制剂 (2WF1 和 2QU3) 与 BACE1 的作用机制, 基于 ossPTMetaD 模拟轨迹, 我们计算了蛋白-配体结合自由能 (ΔG) 随 Flap-Core 距离的变化曲线 (图 3.7)。结果表明, 2WF1 和 2QU3 的最低计算 ΔG 值分别为 -11.15 和 -8.56 kcal/mol, 与它们各自的实验数据高度一致 (-11.98 和 -8.58 kcal/mol) [316-317], 表明 ossPTMetaD 能够采样准确可靠的复合物低能构象。通过 LIGPLOT^[325] 程序, 我们进一步分析了蛋白-配体间的氢键作用和疏水相互作用 (图 3.8)。对于 2QU3 体系, 抑制剂 **462** 与两个催化残基 (D32 和 D228) 形成了稳定的氢键作用, 与残基 Y71 和 F108 形成了较强的疏水作用。对于 2WF1 体

系，除了催化残基 D32 和 D228，抑制剂 **ZY1** 与残基 T232 和 N233 也形成了稳定的氢键作用，同时与 flap 结构上的其他残基，如 Y71、T72 和 Q73，也有不可忽视的疏水作用（图 3.8）。因此，催化残基、flap 区域残基共同与 BACE1 抑制剂形成的相互作用是稳定蛋白-配体结合的关键作用。

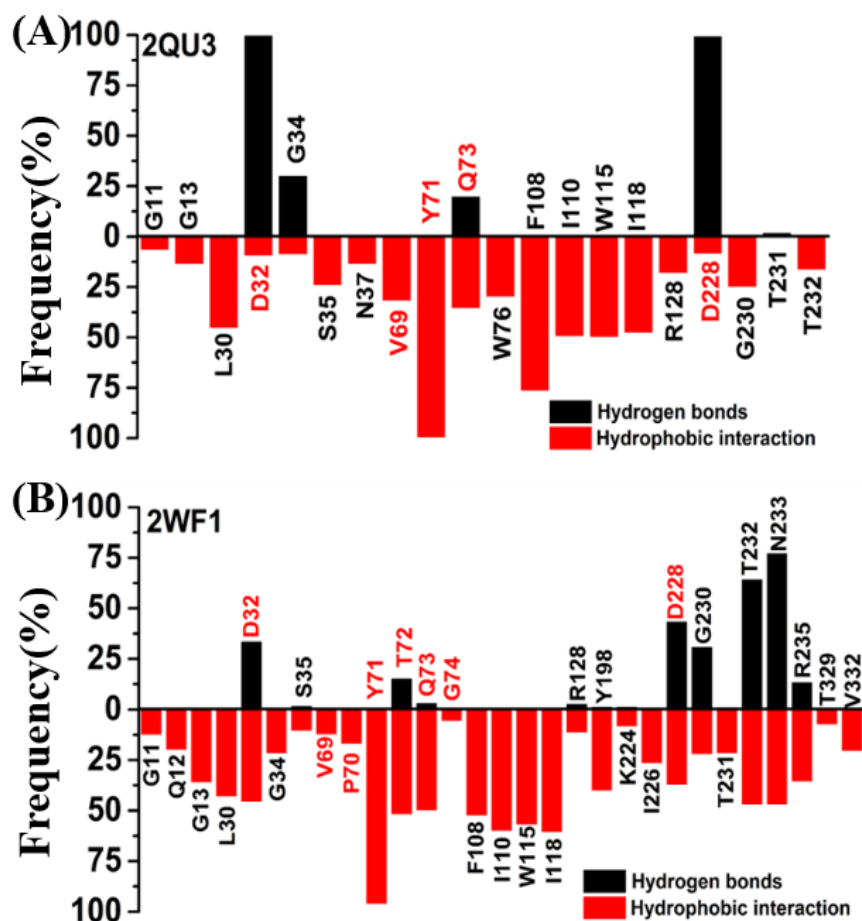


图 3.8 BACE1 体系的作用分析。2QU3 (A)和 2WF1(B)中抑制剂与蛋白形成的氢键作用（黑色）和疏水作用（红色）。催化残基和 flap 区域残基标记为红色。

Figure 3.8 The interactional analysis of BACE1. Hydrogen-bond (black) and hydrophobic (red) interactions between the inhibitors and the amino-acid residues for 2QU3 (A) and 2WF1(B). The residues in the flap and the two catalytic Asp residues are colored in red.

3.3.4 ossPTMetaD 高效准确模拟激酶 DFG 翻转运动

在生物体内，蛋白质的磷酸化修饰是一种重要的翻译后修饰，直接或间接地参与细胞内的许多生物过程，而蛋白激酶（PKs）正是该过程的关键催化酶，因而成为重要的药物靶标。由于蛋白激酶家族的功能相似，即参与催化 ATP 的 γ -磷酸修饰丝氨酸，苏氨酸或酪氨酸的羟基过程，因此 PKs 之间也具有高度的结构

相似性。大部分 PKs 的功能调节受 DFG 模式运动控制。在 DFG-in 状态中, 激酶的 DFG 所在 loop 呈现开放状态, 以应对 ATP 或底物的结合, 因此是催化活性构象。而在 DFG-out 状态中, DFG 向活性位点翻转从而挡在了 ATP 的出入通道, 使激酶处于非活性状态。因此, 研究激酶的 DFG 翻转运动对激酶药物设计至关重要, 但同时也面临着高转变能垒的挑战。常规 MD 模拟由于其低采样效率问题, 研究激酶的 DFG 翻转运动相当困难^[326-327], 但即使用 REMD 方法也需要巨大的计算资源。考虑到本章中新发展的 ossPTMetaD 能够在 NMA 空间和温度空间同时增强采样, 因此我们希望采用 ossPTMetaD 观察此类激酶的构象变化。本章中, 我们选择了两种激酶, 即 p38 α 和 c-Abl 进行评估 ossPTMetaD 方法在 DFG 模式翻转运动上的有效性和准确性。

P38 α 激酶参与多种细胞生物过程和病理过程^[328-329]。以 DFG-out 结构为模拟初始构象 (PDB ID: 3S3I), 我们对 p38 α 进行了 ossPTMetaD 模拟。以 DFG 模式中的 F169 与 I141 的质心距离 (F169-I141 距离, 图 3.9A) 表征 DFG 的运动情况^[330], 结果表明 F169-I141 距离计算得到的自由能曲线在 460 ns 后达到收敛 (图 S3.2)。进一步以 F169-I141 距离和 F169-V38 距离为坐标绘制 P38 α 构象变化的二维自由能图景 (图 3.9B), 从中, 我们发现 P38 α 有两个低能构象, 即 DFG-out 和 DFG-in 结构, 其 F169-I141 和 F169-V38 距离分别为 (6.5, 16.0) 和 (16.0, 6.3)。从 PDB 数据库中, 我们收集了 213 个不同的 P38 α 实验结构 (表 S3.2) 并映射入自由能图景中, 结果发现这些结构都在预测最低能构象转变路径附近, 表明 ossPTMetaD 模拟的可靠性 (图 3.9B)。从 DFG-in 低能构象区域进行聚类分析得到的代表性构象与 DFG-in 实验结构 (PDB ID: 2BAJ) 的 RMSD 仅为 1.56 Å。并且, 在 DFG-in、过渡构象和 DFG-out 构象中, DFG 模式中的 D168 与 N155 形成的氢键作用始终存在。DFG-in 与 DFG-out 的转变能垒预测为~11.5 kcal/mol, 与其他激酶的模拟结果相符: CDK5 为~16-20 kcal/mol^[331]以及 c-Src 为~20 kcal/mol^[332]。

C-Abl 酪氨酸激酶被认为是治疗慢性粒细胞白血病的重要靶标。以 DFG-out 结构为模拟初始构象 (PDB ID: 3QRI), 我们对 c-Abl 进行了 ossPTMetaD 模拟。以 DFG 模式中的 F382 与 L354 的质心距离 (F382-L354 距离, 图 3.10A) 表征 DFG 的运动情况, 结果表明 F382-L354 距离计算得到的自由能曲线在 460 ns 后达到收敛 (图 S3.3)。进一步以 F382-L354 距离和 F382-V256 距离为坐标, 我们

分析了 c-Abl 构象变化的二维自由能图景 (图 3.10B)。从自由能图上, 我们发现 c-Abl 有两个低能构象, 即 DFG-out 和 DFG-in 结构, 其 F382-L354 和 F382-V256 距离分别为 (6.5, 15.0) 和 (15.0, 5.5)。从 PDB 数据库中, 我们收集了 31 个不同的 c-Abl 实验结构 (表 S3.3) 并映射入自由能图景中, 结果发现这些结构都在预测最低能构象转变路径附近 (图 3.9B)。从 DFG-in 低能构象区域进行聚类分析得到的代表性构象与 DFG-in 实验结构 (PDB ID: 4TWP) 的 RMSD 仅为 1.89 Å。DFG-in 与 DFG-out 的转变能垒计算为~3.4 kcal/mol, 与其他模拟结果相符^[275, 333]。

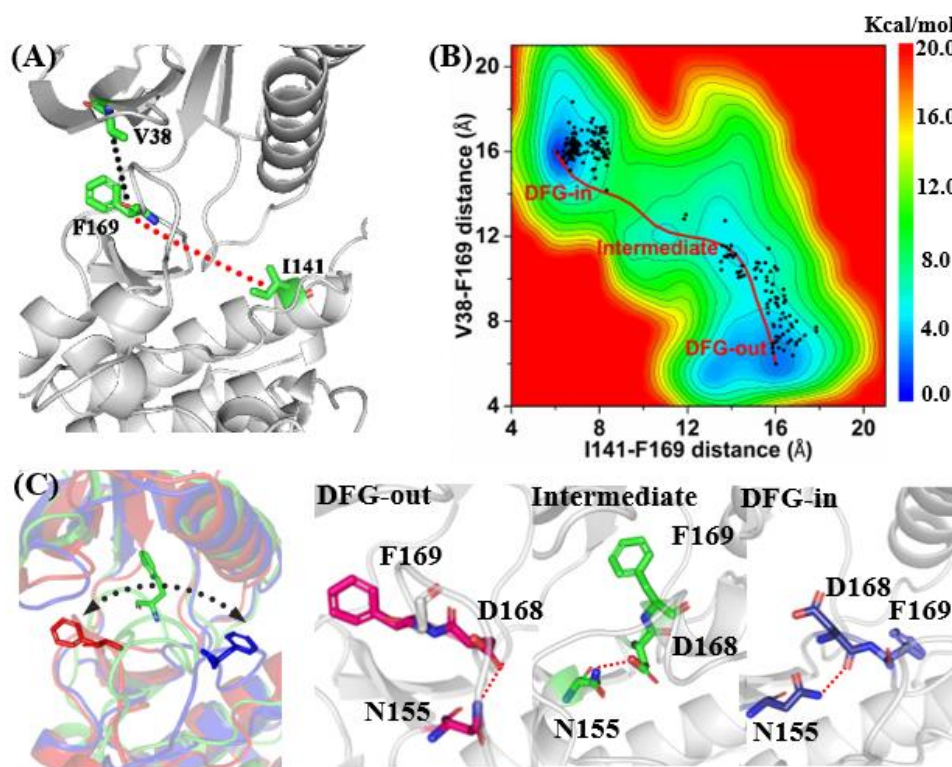


图 3.9 ossPTMetaD 模拟预测 p38 α 构象变化。(A) V38-F169 和 I141-F169 侧链距离示意图。(B) p38 α 构象变化二维自由能景图。最低能构象转变路径标记为红色实线。实验结构标记为黑色圆点。(C) DFG-out (红色)、过渡构象 (绿色) 和 DFG-in (蓝色) 的代表性构象。

Figure 3.9 Conformational change of p38 α predicted by ossPTMetaD. (A) a schematic diagram indicating the side-chain distances of the residue pairs of V38-F169 and I141-F169. (B) the two-dimension free energy landscape of conformational change. The minimum free-energy pathway is labeled as a red line. The experimental structures are projected into the landscape as black dots. (C) the representative structure of the DFG-out (red), the transition state (TS, green), and DFG-in (blue) conformations.

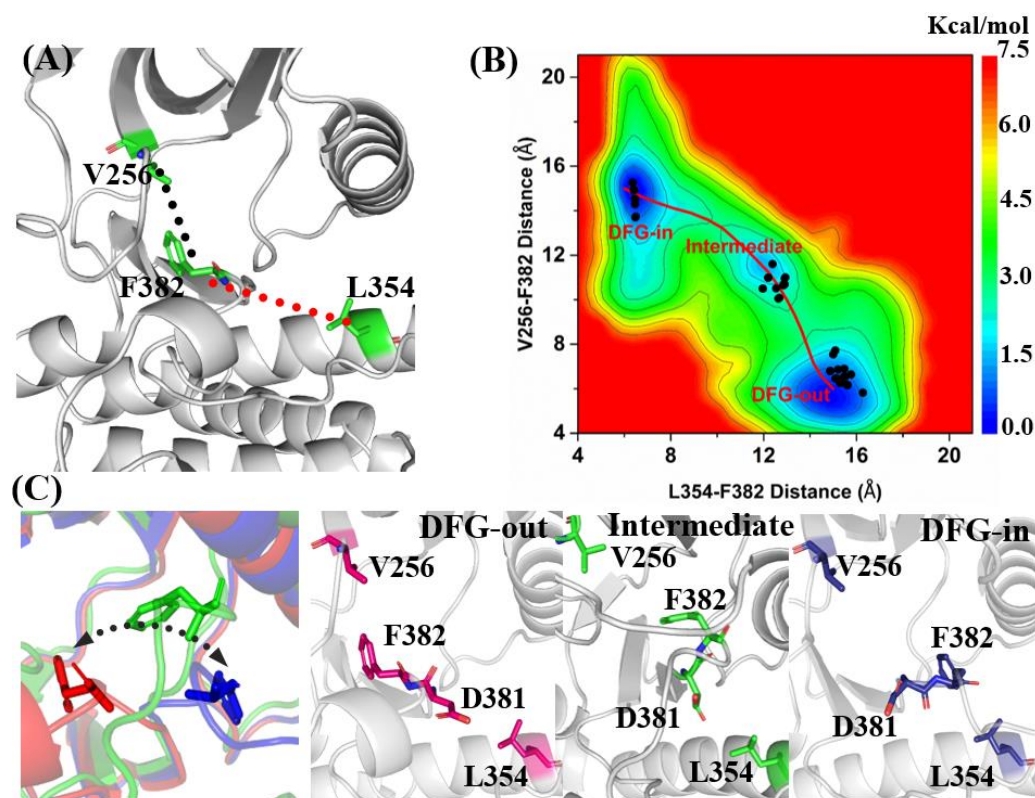


图 3.10 ossPTMetaD 模拟预测 c-Abl 构象变化。(A) V256-F382 和 F382-L354 侧链距离示意图。(B) c-Abl 构象变化二维自由能景图。最低能构象还变路径标记为红色实线。实验结构标记为黑色圆点。(C) DFG-out (红色)、过渡构象 (绿色) 和 DFG-in (蓝色) 的代表性构象。

Figure 3.10 Conformational change of c-Abl predicted by ossPTMetaD. (A) a schematic diagram indicating the side-chain distances of the residue pairs of V256-F382 and F382-L354. (B) the two-dimension free energy landscape of conformational change. The minimum free-energy pathway is labeled as a red line. The experimental structures are projected into the landscape as black dots. (C) the representative structure of the DFG-out (red), the intermediate (green), and DFG-in (blue) conformations.

3.3.5 ossPTMetaD 与其他增强采样方法的比较

近年来, 处理溶剂自由度是改进 T-REMD 采样效率和减少其计算资源的主流方式, 其中 REST2 方法通过对副本中溶剂和溶质的哈密顿量分开调节从而在副本交换时消除溶剂原子的影响。REST2 中溶剂原子始终在基线温度模拟, 因此 P_{ww} 的消除符合细致平衡条件, 但也有研究表明, 这种“热”溶质和“冷”溶剂的处理方式可能会使构象不符合玻尔兹曼分布^[70, 334]。因此, 为评估 ossPTMetaD 与

REST2 去除 P_{ww} 的不同方式, 基于 ossPTMetaD 方法思想, 我们将 REST2 与 WT-MetaD 组合, 称之为 REST2-MetaD, 并与 ossPTMetaD 进行体系应用评估。在 REST2-MetaD 方法中, 副本 m 的势能表达为公式 3.14。

$$E_{pot} = \frac{\beta_m}{\beta_0} P_{pp} + \sqrt{\frac{\beta_m}{\beta_0}} P_{pw} + P_{ww} + V(s, t) \quad (3.14)$$

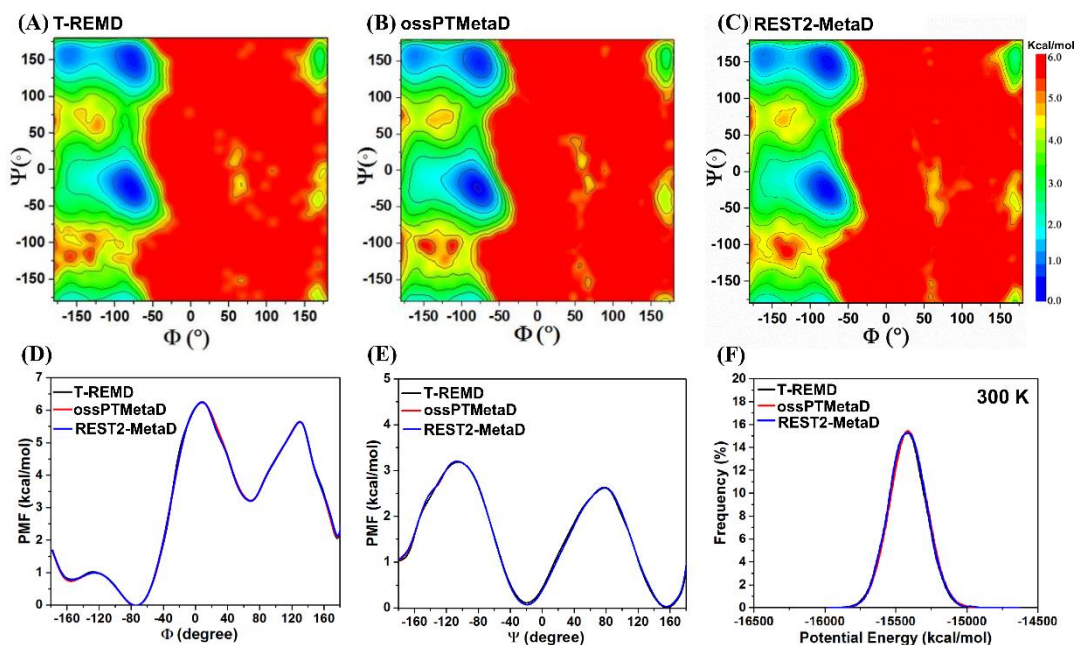


图 3.11 ossPTMetaD 与其他方法的丙氨酸二肽测试。以二面角 (Φ , Ψ) 为坐标绘制的 T-REMD (A), ossPTMetaD (B) 和 REST2-MetaD (C) 二维自由能图景, 以及以二面角 Φ (D) 和 Ψ (E) 绘制的一维自由能图, 和它们的势能分布 (F)。

Figure 3.11 The simulations for alanine dipeptide using T-REMD, ossPTMetaD and REST2-MetaD. The two-dimensional free energy landscape as a function of the backbone dihedral angles (Φ , Ψ) obtained from T-REMD (A), ossPTMetaD (B) and REST2-MetaD (C) simulations at 300 K. The comparison of one dimensional PMF as a function of two dihedral angles Φ (D) and Ψ (E), and potential energy distributions (F).

在增强采样方法发展中, 丙氨酸二肽是一个常用的成熟体系^[335-337], 因此, 我们首先以丙氨酸二肽为例, 以 T-REMD 为参照, 评估 REST2-MetaD 和 ossPTMetaD 的模拟结果。如图 3.11 所示, REST2-MetaD、ossPTMetaD 和 T-REMD 三者的 (Φ , Ψ) 构象分布一致, 而且它们之间的势能分布和自由能变化也是高度一致的 ($\Delta\text{PMF} < 0.1 \text{ kcal/mol}$)。因此, 丙氨酸二肽的测试结果表明 ossPTMetaD 能够获得与其他增强采样方法和常规 T-REMD 相一致的模拟结果。为进一步评估 REST2-MetaD 和 ossPTMetaD 方法, 我们通过 REST2-MetaD 方法模拟了 HIV-

1 PR 体系的构象分布。与 ossPTMetaD 相比, REST2-MetaD 在整体 RMSD 为 5 Å 左右区域采样相对更少,但它们都获得了相似的 4 个低能区域(图 3.12A)。对于 ossPTMetaD, 闭合构象、半开放、内卷曲和完全开放构象的 PMF 分别是 1.50、1.27、3.74 和 4.66 kcal/mol, 而 REST2-MetaD 为 1.54、1.29、3.61 和 4.66 kcal/mol (图 3.13B)。因此, ossPTMetaD 和 REST2-MetaD 方法得到的闭合构象到开放构象的自由能垒差别仅为 0.02 kcal/mol。并且, ossPTMetaD 和 REST2-MetaD 方法得到的这四种构象结构具有明显的相似性,其 RMSD 均小于 1.4 Å(图 3.12C)。因此,通过与 REST2-MetaD 方法的比较, ossPTMetaD 能够获得与服从细致平衡条件的方法相似的模拟结果,进一步表明其准确性。

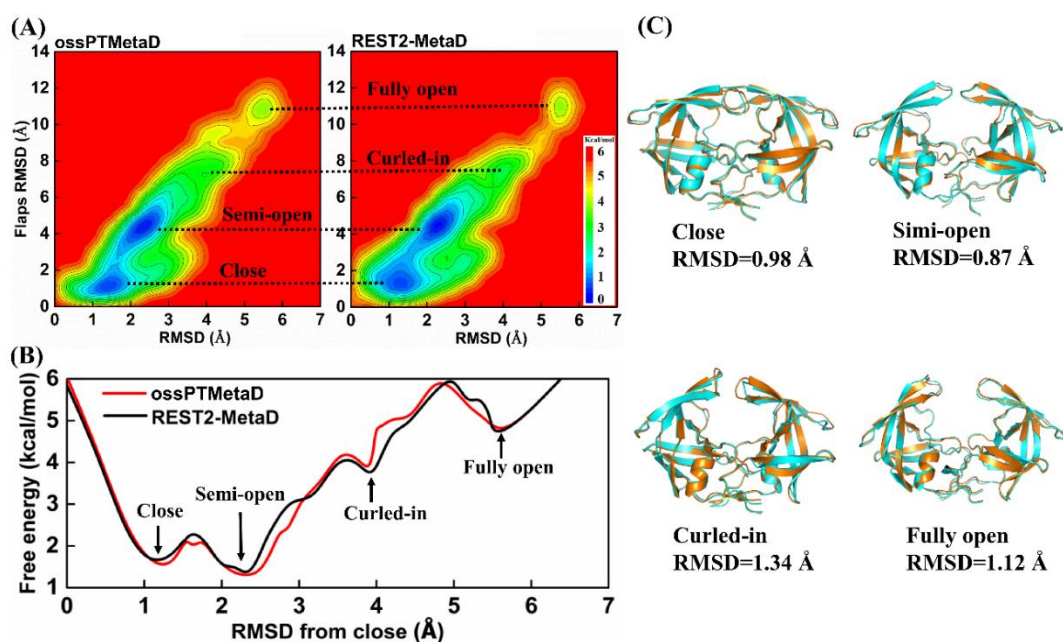


图 3.12 ossPTMetaD 与 REST2-MetaD 的 HIV-1 PR 测试。以整体和 flap 结构 RMSD 为坐标绘制的 ossPTMetaD (左边) REST2-MetaD (右边) 二维自由能图景 (A), 以及一维自由能图 (B)。闭合构象、半开放、内卷曲和完全开放构象的结构比对 (C)。

Figure 3.12 The simulations for flap dynamics of HIV-1 PR using ossPTMetaD and REST2-MetaD. (A) Two dimensional free energy landscape as a function of overall RMSD and flaps RMSD obtained from ossPTMetaD (left) and REST2-MetaD (right) simulations. (B) The one-dimensional PMF for HIV-1 PR conformational change pathway obtained from ossPTMetaD and REST2-MetaD simulations. (C) Alignment of the closed, semi-open, curled-in, and fully open representative structures obtained from ossPTMetaD (cyan) and REST2-MetaD (orange) simulations, respectively.

3.4 本章小结

蛋白质的功能发挥与其构象变化密切相关。因此,研究靶标蛋白的动态行为对理解其生物学功能以及药物发现与设计至关重要。分子动力学模拟能够从原子水平上监测蛋白质结构的动态变化,然而常规 MD 受限于其低采样效率,且易陷入局部极小值。增强采样方法通过各种原理提高 MD 采样效率,已成为研究蛋白质大规模构象变化的主流技术之一。其中, T-REMD 是一种通过多个平行模拟副本交互信息的增强采样方法,广泛应用于有机分子和生物大分子的构象采样及自由能的计算研究。但是 T-REMD 所需模拟副本数目与体系总自由度的平方根成正比,普通计算平台难以承受由此带来的巨大计算成本。本章工作通过整合 T-REMD 与另一种高效的增强采样方法 (Well-tempered Metadynamics), 并且采用 NMA 作为增强采样变量, 发展了 ossPTMetaD 方法。新方法在副本交换时也去除溶剂内部作用能的影响, 显著减少了体系的自由度及所需的副本数目。采用 ossPTMetaD 方法, 我们首先研究了 AdK 关闭与打开两种构象的动态转换过程。相比 T-REMD 方法需要 80 个模拟副本, ossPTMetaD 只需 24 个副本即可成功模拟此构象变化, 并且具有比 T-REMD 更高的副本交换率和更快的模拟收敛性。进一步将腺苷酸激酶不同状态的实验结构映射入自由能图景中, 我们发现这些实验结构均在 ossPTMetaD 模拟预测的构象转变路径附近, 表明 ossPTMetaD 模拟结果的合理性。

为评估 ossPTMetaD 方法的体系可拓展性, 我们将 ossPTMetaD 方法用于模拟 HIV-1 蛋白酶和 BACE1 蛋白的 flap 柔性运动以及激酶的 DFG 翻转运动。对于 HIV-1 蛋白酶, 我们通过 ossPTMetaD 方法获得了 4 种不同的 HIV-1 蛋白酶低能构象以及它们之间的相互转变能垒。其中, 除了目前实验已经观察到的闭合构象和半打开 (semiopen) 构象外, 我们还发现了另外两种低能构象: 内卷曲和完全打开构象。BACE1 蛋白被认为是一种治疗阿尔茨海默症的重要靶点, 我们通过 ossPTMetaD 方法不仅研究了 BACE1 空蛋白的 flap 柔性变化, 并且发现不同结构的 BACE1 抑制剂能够将蛋白结构限制在闭合和打开两种不同的状态。对于激酶体系, 我们通过 ossPTMetaD 方法成功模拟了 p38 α 和 c-Abl 两种激酶的 DFG 翻转运动, 并获得这类构象变化的自由能图景以及中间低能构象。同时需要注意的是, 不同于 T-REMD, ossPTMetaD 在副本交换时并未严格遵从细致平衡条件, 因此需要更多的体系测试以进一步验证该方法的准确性和普适性。

第 4 章 高效采样模拟方法的实际应用

4.1 大环化合物在不同极性溶剂下的构象分布

4.1.1 研究背景

大环化合物的发现犹如一缕春风拂过药物化学领域,大环药物研究应运而生,并逐渐遍地生花。半刚性的大环骨架不仅赋予这些大环化合物良好的亲和力和选择性,更使其具备良好的透膜性^[338-340]。尽管分子量往往超过 Lipinski 五规则^[341],大环化合物出其不意表现出来的类药性质差慰人意^[338-339, 342-343],诸如代谢稳定性和良好的口服生物利用度^[344-345]。迄今为止,已有超过 100 种大环药物处于上市或在研阶段^[338],其中大部分都是天然产物,例如红霉素(erythromycin)、环孢菌素 A(cyclosporin A)、利福平(rifampicin)、螺旋霉素(spiramycin)和雷帕霉素(sirolimus)等^[338, 346]。合成大环化合物表现出的分子多样性和复杂性也逐渐吸引科研工作者的兴趣和关注^[347]。近期一些合成大环化合物,如伐尼瑞韦(vaniprevir)^[348]、丹诺普韦(danoprevir)^[349]和劳拉替尼(lorlatinib, PF-06463922, Pfizer)^[350]等被批准上市,将大环化合物的成药潜力展露无遗^[351]。另外,大环化合物在一些极具挑战性的靶标蛋白中也颇有潜力,例如蛋白-蛋白相互作用(protein-protein interaction, PPI)^[352-353]。近年来,许多大环化合物被报道为 PPI 抑制剂靶向诸如 MDM2/p53^[354]、BCL-2^[355]和 IL-17A/IL-17RA^[356]等蛋白。

在溶液中能检测到大环化合物的靶标结合构象无疑为大环化合物的研究添光增彩^[357-362]。最近研究表明,大环化合物可通过键旋转而改变其构象变化以暴露或隐蔽其功能基团来应对不同极性环境^[363-365]。大环药物的这种构象应变能力使其具有良好的透膜性和生物利用度^[366]。虽然现有的实验技术包括 NMR 和 X 晶体衍射等能够用于准确测定大环化合物的溶液构象,但这些方法不仅需要真实的化合物样品,而且非常昂贵和耗时,因此对于研究大环化合物的构象变化仍极具挑战。

近年来,通过高效的理论计算模拟方法研究大环化合物的构象分布受到广泛关注^[367-372]。可用于大环化合物构象分布研究的算法包括几何类方法(inverse kinematics^[369])、蒙特卡罗扭转采样^[368]以及基于 MD 的采样方法(如 LowModeMD^[373]、加速 MD^[374]和 CoCo-MD^[375]等)。并且,合作课题组 Mate 教

授等人也已经比较研究过 MacroModel、LowModeMD 和 OMEGA^[367]方法在大环化合物构象采样上的表现^[376]。常规 MD (cMD) 模拟通常被认为能够在原子水平上观察分子的结构转变^[377], 然而 cMD 模拟往往局限于局部极小值点, 因而采样效率低。为了克服 cMD 的低采样效率问题, 副本交换分子动力学 (REMD) 通过高温克服能垒, 已经广泛应用于有机分子和生物大分子的构象采样, 如环糊精和环肽等^[378-379]。但目前的研究都关注于大环化合物在特定溶剂中的构象采样, 而大环化合物的构象分布与其所处的溶剂环境性质 (极性以及氢键供体/受体可容性等) 的关系尚不清楚。由于在生物体内, 体液是水溶剂而细胞膜的磷脂双分子层是非极性的, 因此准确预测大环化合物在不同极性溶剂下的构象分布对预测它们的药理性质有极大的帮助。

劳拉替尼分子被批准为治疗非小细胞肺癌的药物, 是含 12 环原子的大环化合物, 具有良好的选择性、有效性、安全性和生物利用度^[380-381]。由于大环骨架减少了劳拉替尼的可旋转键数目, 使得劳拉替尼结合靶标蛋白 (间变性淋巴瘤激酶 ALK 和 ROS1) 具有较为刚性的生物活性构象^[382]。X 射线衍射晶体结构表明劳拉替尼在这两种不同靶标蛋白中的结合构象相同^[380, 383]。但是, 劳拉替尼表现出的良好透膜性和血浆屏障穿透性暗示其潜在的构象应变性。因此, 劳拉替尼可以作为一个理想的模型以研究大环化合物在不同极性溶剂下的构象分布与其物理化学性质的关联。

在本章中, 通过 REMD 方法, 我们模拟了大环药物劳拉替尼在不同极性溶剂中的构象分布。水溶剂模型用以模拟体液环境, 氯仿溶剂模型 (介电常数 $\epsilon=4.8$) 用以模拟细胞膜的疏水内部 ($\epsilon=3.0$), 并且氯仿溶剂也是 NMR 实验常用的非极性溶剂。模拟结果显示不同极性的溶剂通过分子间相互作用对劳拉替尼的构象分布产生显著影响, 而且结果与 NMR 实验高度契合。

4.1.2 研究方法

体系准备与 REMD 模拟。通过 ChemDraw 软件获得劳拉替尼的模拟初始结构, 并通过 Gaussian 09 在 B3LYP/6-31G* 水平下优化。通过 *antechamber* 程序拟合得到体系原子 RESP 电荷。之后通过 Amber16 软件包的 *tleap* 模块将优化后的劳拉替尼置于立方体溶剂盒子中, 并在盒子中填充 CHCl₃ 或者 H₂O/DMSO (6:4) 溶剂。不同溶剂模拟体系总原子数目保持 1850 个左右。采用 GAFF 力场对劳拉替

尼进行参数化。

首先对劳拉替尼进行能量最小化, 包括 5000 步的最速下降法和 5000 步的共轭梯度法。之后在 0.5 ns 的 NPT 模拟中内将劳拉替尼体系升温至 300 K, 并平衡模拟 2 ns。最终, 进行 22 个平行副本的 REMD 模拟, 温度梯度包含 300 K 到 600 K, 每个副本模拟 100 ns, 每隔 2 ps 提取一次构象。通过 Amber16 软件包的 *cpttraj* 模块计算劳拉替尼相对初始构象的 RMSD 以监视其构象变化。

核磁共振 (NMR) 波谱。采用 D₂O/DMSO-*d*₆ (6:4) 和 CDCl₃ 溶剂, 在配备了 TCI 低温探针的 800 MHz Bruker Avance III HD NMR 光谱仪上于 25 °C 记录 NMR 波谱。使用一维 (1D, ¹H 和 ¹³C) 和二维 (2D, COSY, TOCSY, HSQC, HMBC 和 NOESY) 推导 NMR 谱图 (表 S4.1)。2D-NOESY 通过 7 个混合时间 (100, 200, 300, 400, 500, 600 以及 700 ms), *d*₁ 弛豫延迟设置为 2.5 s 获得。质子间的距离 (表 S4.2 和 S4.3) 是根据初始速率近似值 ($r^2 > 0.95$) 计算的, 使用邻位芳族质子 (2.51 Å) 和次甲基到甲基质子的距离 (2.43 Å) 作为内部距离参考。根据 $([\text{交叉峰 } 1 \times \text{交叉峰 } 2] / [\text{对角峰 } 1 \times \text{对角峰 } 2])^{0.5}$, 使用交叉峰和对角峰的归一化来计算 NOE 峰强度。使用方程 $r_{ij} = r_{ref}(\sigma_{ref}/\sigma_{ij})^{1/6}$ 将初始 NOE 增加速率转换为距离, 其中 r_{ij} 是质子 *i* 和 *j* 之间的距离 (Å), 而 σ_{ij} 是从 NOESY 实验获得的归一化强度。相关实验由 Mate 教授课题组完成。

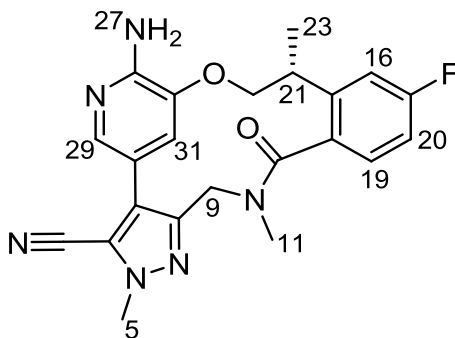


图 4.1 劳拉替尼结构。

Figure 4.1 Structure of Lorlatinib.

NAMFIS 分析。采用五种不同的分子力场 (OPLS, OPLS-2005, OPLS3e, AMBER* 和 MMFF) 进行无限制的蒙特卡罗构象分析, 确定 lorlatinib 的构象空间集合, 每个力场均采用氯仿和水的 GB/SA 溶剂化模型。使用蒙特卡罗算法进行了构象搜索, 该算法具有中间扭转采样, 5 万步蒙特卡罗, RMSD 截断值设置

为 2.0 Å。通过 Schrödinger 软件包的 Macromodel BatchMin V12.1, 使用 Polak-Ribière 型共轭梯度 (PRCG) 对每个构象进行分子力学能量最小化, 最多重复 5000 步。保存与能量全局最小值相差 42 kJ/mol 以内的所有构象 (表 S4.4), 并采用重原子 RMSD 截断值 1.0 Å 消除了多余的构象, 从而给出了用于 NAMFIS 分析的最终构象集。

根据之前研究中提出的方案, 通过将实验测量的距离与为计算预测的构型反算的距离相拟合来确定溶液构型^[384]。CH₂ 信号根据方程 $d = (((d_1^{-6}) + (d_2^{-6}))/2)^{-1/6}$ 处理, 以及根据 $d = (((d_1^{-6}) + (d_2^{-6}) + (d_3^{-6}))/3)^{-1/6}$ 处理甲基信号。使用标准流程验证 NAMFIS 构型分析, 即通过向实验数据中添加 10% 随机噪声, 随机去除单个约束以及通过比较实验来评估构象约束的可靠性。对于 D₂O/DMSO-*d*₆ (6:4) 和 CDCl₃, NAMFIS 分析的结果在表 S4.9 中给出。相关实验由 Mate 教授课题组完成。

使用 1.4 Å 的溶剂探针半径计算溶剂可及的 3D 极性表面积, 包括极性原子 (O, N 和相连的 H) 和部分电荷。基于分子亲脂性 (MLP) 的计算 logP 是一种定量描述分子亲脂性的参数, 通过 VEGA ZZ 软件计算。

X 射线晶体结构。 PDB 数据库中劳拉替尼在不同晶体结构中具有相似的构型, 本章中选择了 PDB ID: 4CLI 中的劳拉替尼结构进行分析。

4.1.3 结果与讨论

4.1.3.1 REMD 预测劳拉替尼在不同极性溶剂下的构象分布

在 REMD 模拟中, 我们将劳拉替尼置于两种不同极性溶剂中, 即 CHCl₃ 和 H₂O/DMSO (6:4) 混合溶剂。采用 H₂O/DMSO 混合溶剂是由于劳拉替尼的水溶解度低。以重原子 RMSD 为坐标, 我们计算劳拉替尼构象分布的一维自由能图, 如图 4.2 所示, 80 ns、90 ns 和 100 ns 的自由能曲线表明 100 ns 已经得到较好的收敛。在 H₂O/DMSO 混合溶剂中, 只有一种低能构象 P1 存在; 然而在 CHCl₃ 溶剂中有两种不同的低能构象 (P1 和 P2)。构象 P1 与劳拉替尼的靶标结合构象基本一致 (PDB ID: 4CLI; P_c, RMSD 0.27 Å), 而构象 P2 与结合构象 P_c 的 RMSD 为 1.2 Å (图 4.3), 说明劳拉替尼在其靶标蛋白 (ALK 和 ROS1) 中的结合构象就是其在水溶剂中的低能构象。

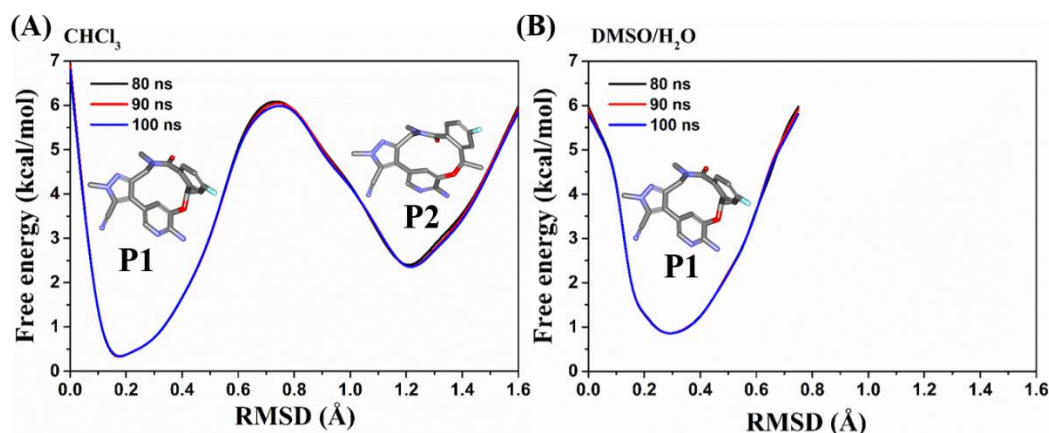


图 4.2 劳拉替尼的一维自由能图。CHCl₃ (A)和 DMSO/H₂O (4:6) (B)溶剂中以重原子 RMSD 计算的一维自由能图。代表性低能构象以棍棒模型展示。

Figure 4.2 One-dimensional free energy profiles of lorlatinib. One-dimensional free energy profiles as the function of heavy-atom RMSD in CHCl₃ (A) and DMSO/H₂O (4:6) (B). The representative low-energy conformers are shown as sticks.

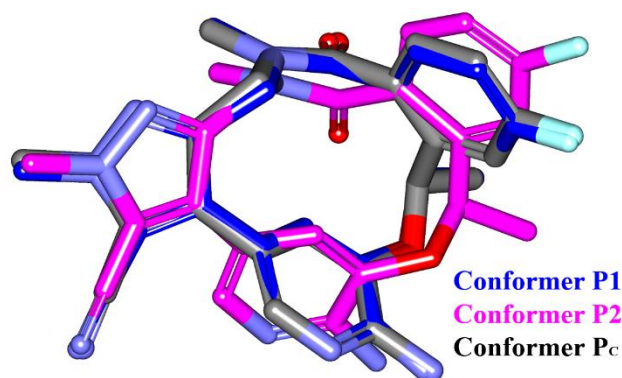


图 4.3 代表性低能构象。低能构象 P1（蓝色）和 P2（粉色）的代表构象展示。晶体结合构象 Pc 为灰色。

Figure 4.3 The representative low-energy conformations. The representative conformers P1 (blue colored) and P2 (pink colored) in CHCl₃. The crystal conformer Pc is colored by gray.

进一步我们将劳拉替尼分成两部分（Part1 和 Part2，图 4.4A）计算 RMSD 以绘制其二维自由能图景。其中 Part1 部分的 RMSD（RMSD_part1）主要表征吡啶环和吡唑环的翻转运动，Part 2 部分的 RMSD（RMSD_part2）主要表征苯环和酰胺键的翻转运动。如图 4.4B-D 所示，在 CHCl₃ 溶剂中，构象 P1 占比 80%，而构

象 P2 占比 20%。构象 P1 的 Part1 和 Part2 结构与结合构象 Pc 的 RMSD 均为 0.25 Å, 表明构象 P1 与 Pc 是高度一致的; 构象 P2 的 Part1 和 Part2 结构与结合构象 Pc 的 RMSD 分别为 1.50 Å 和 0.90 Å, 表明构象 P2 相对 Pc 的主要变化是 Part1 部分。并且, 构象 P1 相比 P2 更稳定, 他们之间的能量差为~11.3 kJ/mol, 但构象 P1 和构象 P2 之间的转变能垒为~22.6 kJ/mol, 表明构象 P1 和构象 P2 在 CHCl₃ 溶剂中可以相互转变。如图 4.4D 的低能转变路径所示, 构象 P1 转变成构象 P2 时, 首先发生 Part1 和 Part2 部分的同时内翻转, 之后 Part1 结构继续内翻转, 说明 Part1 部分的吡啶环的翻转运动对劳拉替尼的构象变化至关重要。然而, 在 H₂O/DMSO (6:4) 溶剂中, 只有一种低能构象 P1 出现 (图 4.4C)。

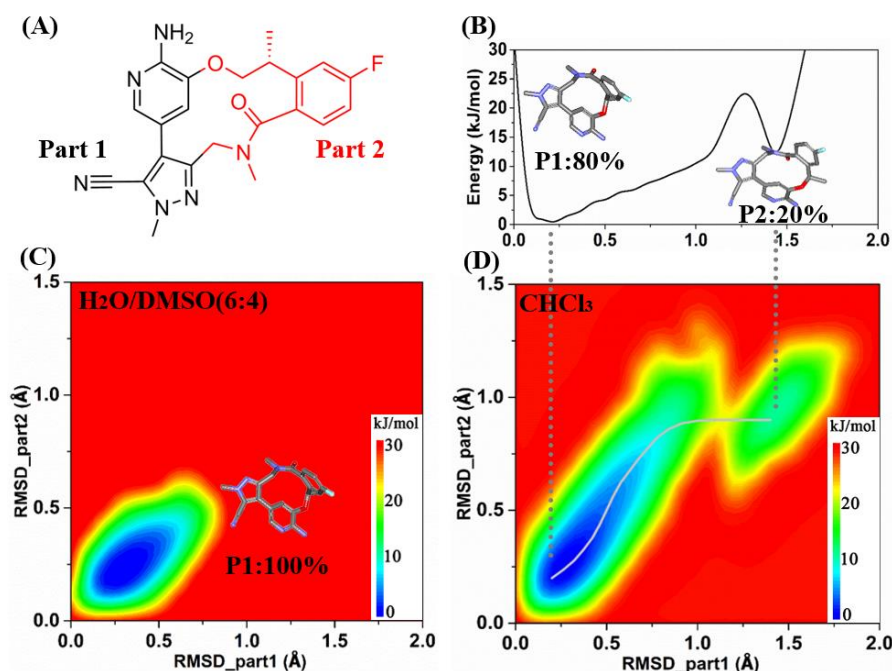


图 4.4 劳拉替尼的模拟构象及自由能图景。劳拉替尼的两部分 RMSD 值计算结构展示 (A)。CHCl₃ 溶剂中, 构象 P1 和 P2 之间结构转变的一维自由能图 (B)。在 H₂O/DMSO (C) 和 CHCl₃ (D) 溶剂中, 劳拉替尼构象分布的二维自由能图景。最低能量路径以灰线展示。

Figure 4.4 Simulated conformations and free energy profiles of lorlatinib. Lorlatinib and its two parts divided for RMSD analysis (A); One-dimensional free energy profile of the structural transition from P1 to P2 in CHCl₃ (B); Two-dimensional free energy landscapes of the conformation distribution of lorlatinib in H₂O/DMSO and in CHCl₃, respectively (C and D). The minimum free-energy pathway is shown as a gray line.

4.1.3.2 NMR 测定劳拉替尼在不同极性溶剂下的构象分布

通过 NMR 技术, 我们分别测定了劳拉替尼在 $D_2O/DMSO-d_6$ (6:4) 和 $CDCl_3$ 溶剂中构象分布 (表 S4.5)。结果表明劳拉替尼在 $D_2O/DMSO-d_6$ 溶剂中只有构象 **1**, 而在 $CDCl_3$ 溶剂中有两种构象 **1** 和 **2** (图 4.5)。在 $CDCl_3$ 中, 构象 **1** 是主要构象 (占比 76%) 而构象 **2** 占比 24% (表 S4.5)。如表 S4.6 所示, 构象 **1** 与劳拉替尼结合构象 Pc 的相似性 (RMSD 为 0.35 Å) 也明显高于构象 **2** (RMSD 为 1.02 Å)。因此, 劳拉替尼在这两种不同极性溶剂中的 NMR 数据也正好验证了我们 REMD 的模拟结果, 并且 NMR 构象 **1** 与模拟得到的低能构象 P1 之间的 RMSD 仅为 0.15 Å, NMR 构象 **2** 与构象 P2 之间的 RMSD 也仅为 0.17 Å (图 4.5)。并且目前的 NMR 实验数据以及 REMD 模拟结果均说明溶剂对劳拉替尼大环化合物的构象分布具有重要影响。



图 4.5 劳拉替尼的 NMR 构象分析。在 $D_2O/DMSO-d_6$ (蓝色) 和 $CDCl_3$ (粉色) 溶剂中的 NMR 构象, 并与预测构象 (灰色) 叠合比较。水溶剂中构象 **1** 与 P1 的比较 (A); 氯仿溶剂中构象 **1** 与 P1, 构象 **2** 与 P2 的比较 (B)。

Figure 4.5 The solution conformations of lorlatinib by NMR. In $D_2O/DMSO-d_6$ (blue) and $CDCl_3$ (pink) aligned with the predicted conformations (gray). (A) **1** vs P1 in aqueous solution; (B) **1** vs P1 (left) and **2** vs P2 (right) in chloroform.

4.1.3.3 劳拉替尼构象应变与其物理化学性质的关系

为进一步理解劳拉替尼的在不同极性溶剂下的构象应变性, 我们分析了 REMD 模拟中劳拉替尼与溶剂的分子间作用。在 $H_2O/DMSO$ 溶液中, 劳拉替尼在模拟中与水溶剂形成良好的氢键网络, 其中劳拉替尼和水分子均能扮演氢键供体和氢键受体 (图 4.6A)。这种紧密的氢键网络使得劳拉替尼整体结构变得刚性, 从而解释其在水溶剂中只呈现一种构象。然而, 在 $CHCl_3$ 中, 劳拉替尼与氯仿溶剂并没有形成强的分子间相互作用, 因此允许劳拉替尼分子内键的自由旋转 (图 4.6B)。

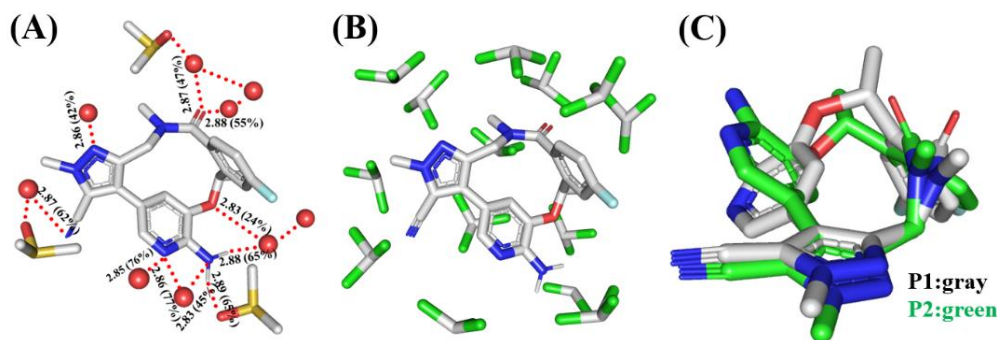


图 4.6 劳拉替尼与溶剂的分子间相互作用展示。在 $\text{H}_2\text{O}/\text{DMSO}$ 溶剂中，劳拉替尼与溶剂间存在氢键网络（红色虚线，A），但在 CHCl_3 溶剂中并没有强的分子间作用（B）。在 REMD 模拟中氢键供体与受体的平均距离以及占有率标注在虚线上。构象 P1 与构象 P2 的叠合。

Figure 4.6 Intermolecular interaction between lorlatinib and the solvents. There is a hydrogen bond network (red dash lines) in $\text{H}_2\text{O}/\text{DMSO}$ (A), while no strong intermolecular interaction in CHCl_3 (B). The average distance between the HB donor and acceptor during the REMD simulations is provided along the dash line (Å). The percentage time of HB during the REMD simulations is shown in the brackets. (C) Alignment of P1 and P2.

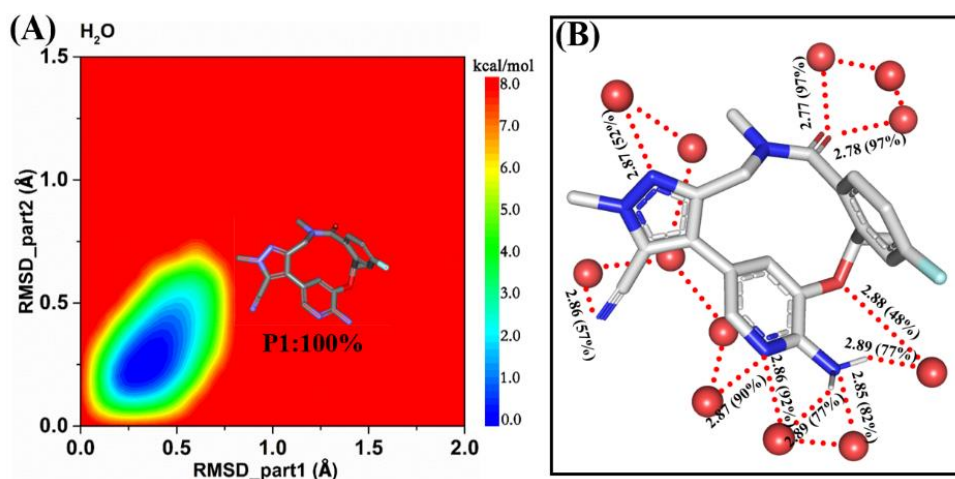


图 4.7 劳拉替尼在水溶剂中的构象分布。构象分布二维自由能图景（A）。劳拉替尼与溶剂水形成的氢键网络（B）。在 REMD 模拟中氢键供体与受体的平均距离以及占有率标注在虚线上。

Figure 4.7 The conformational distribution of lorlatinib in water. (A) Two-dimensional free-energy landscapes of the conformation distribution. (B) Schematic hydrogen bond network formed by lorlatinib and water. The average distance between the HB donor and acceptor during the REMD simulations is provided along the dash line (Å). The percentage time of HB during the REMD simulations is shown in the brackets.

为验证溶剂水对劳拉替尼构象变化的限制作用，我们在相同模拟条件下以纯水溶剂模型预测了劳拉替尼的构象动态特性。结果表明，在纯水溶剂中，劳拉替尼也只存在一种低能构象 P1（图 4.7），并且与 H₂O/DMSO 一样。劳拉替尼与水同样形成了紧密的氢键网络。因此，这一模拟结果再次说明了溶剂环境的氢键供体与受体可容性显著影响了大环化合物的构象分布。

通过观察劳拉替尼的构象 P1 和构象 P2，我们发现在水溶剂中构象 P1 的酰胺键羰基完全暴露于溶剂中，而构象 P2 的酰胺键羰基被吡啶基局部屏蔽于溶剂（图 4.6C & 4.8），暗示构象 P2 可能相对 P1 更具疏水性。因此，我们计算了这两种构象的溶剂可及 3D 表面积^[363]，结果表明构象 P2 的溶剂可及 3D 表面积（209.8 Å²）确实比构象 P1（216.5 Å²）小。而且，构象 P2 的偶极矩（ $\mu=3.6$ D）也比构象 P1 小（ $\mu=4.4$ D）。因此，构象 P2 更加疏水的分子表面也使其更适合于低极性溶剂环境。这也验证了我们之前的推测，即劳拉替尼在体液环境中以构象 P1 存在，而转变成构象 P2 帮助透膜，并且构象 P1 和 P2 之间的低转变能垒也保证了劳拉替尼快速的构象转变。

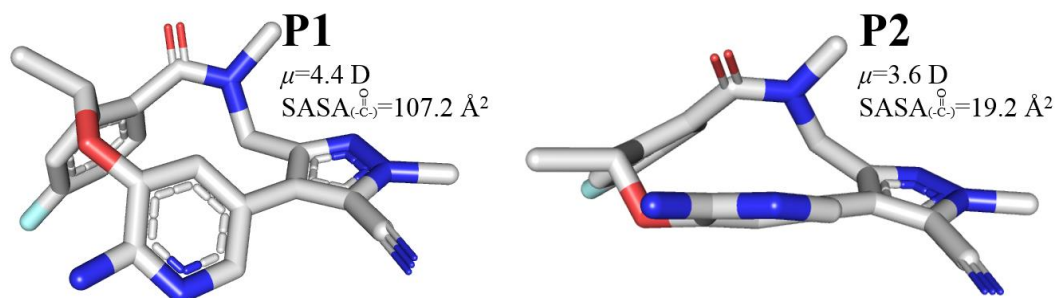


图 4.8 构象 P1 和构象 P2。羰基的溶剂可及表面积由 Pymol 软件计算，偶极矩由高斯 09 在 M06-2X/6-311G** 基组下计算得到。

Figure 4.8 The conformers P1 and P2. The solvent accessible surface area (SASA) of the carbonyl group calculated with Pymol and the overall dipole moment (μ) calculated with gaussian 09 at M06-2X/6-311G** level, are provided.

4.1.3.4 劳拉替尼构象应变与其物理化学性质的关系

为进一步验证 REMD 方法对大环化合物构象采样的准确性，我们将其应用于 Hsp90 抑制剂——大环化合物根赤壳菌素（radicicol）。在溶剂 CHCl₃ 中，根赤壳菌素存在 6 种不同的构象，其构象比例在 5%到 33%之间（图 4.9），其中构象占比排名第二的构象（28%）与赤壳菌素在 Hsp90 蛋白上的结合构象高度一致，

它们之间的RMSD仅为0.18 Å(PDB ID: 1BGQ)。并且,根据文献报道,在NAMFIS分析中,赤壳菌素的结合构象在 CDCl_3 溶剂中的占比是 21%^[385]。因此,在大环化合物赤壳菌素上再一次证明 REMD 模拟结果与实验数据的一致性。

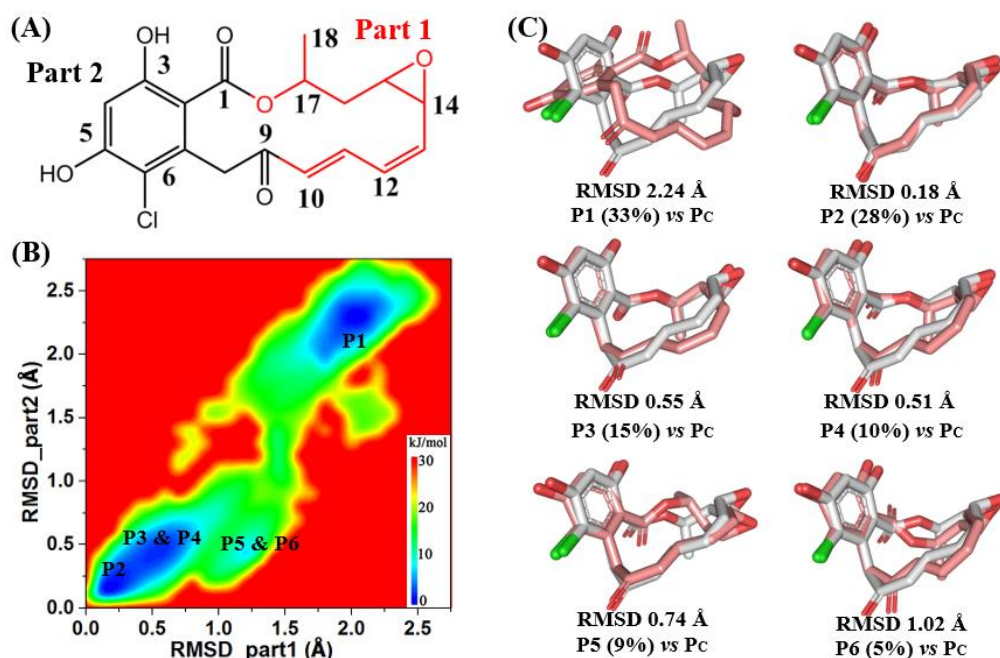


图 4.9 根赤壳菌素在 CHCl_3 溶剂中的 REMD 模拟。根赤壳菌素结构展示以及用于 RMSD 计算两部分结构 (A)。根赤壳菌素在 CHCl_3 溶剂中构象分布二维自由能图景 (B)。代表性低能构象 (P1-P6), 以及其各自的构象占比和与晶体构象 (Pc, 白色, PDB ID: 1BGQ) 的叠合。

Figure 4.9 The REMD simulation of radicicol in CHCl_3 . (A) Radicicol and the two parts it was divided into for RMSD analysis; (B) Two-dimensional free energy landscape of the conformational distribution of radicicol in CHCl_3 . (C) Representative low-energy conformers (P1-P6), with populations in %, from the REMD simulation compared to the crystal structure (Pc, colored in white) of radicicol (PDB ID: 1BGQ).

4.1.4 小结

以劳拉替尼为例,我们采用副本交换分子动力学方法,模拟了大环药物在不同极性溶剂中的构象分布,结果与 NMR 实验高度契合。结果表明,在 $\text{DMSO}/\text{H}_2\text{O}$ 混合溶剂以及纯水溶剂中,劳拉替尼与溶剂形成密集的氢键网络,因此只稳定呈现一种构象 1。这种构象与劳拉替尼结合 ALK/ROS1 蛋白的晶体构象非常相似 (RMSD=0.27 Å)。但在氯仿溶剂中,劳拉替尼除构象 1 外 (占比 80%),还呈现

另一类构象 2（占比 20%）。与构象 1 相比，构象 2 的极性表面积和偶极矩都更小，而 $\log P$ 更大，使其更适合于低极性溶剂环境。劳拉替尼构象 P1 与构象 P2 相互转变低能垒保证了其可以通过构象变化来达到良好的透膜性质。另外，在根赤壳菌素上的测试也再次验证 REMD 能够准确预测大环化合物的构象分布，为预测其透膜性等性质提供帮助。

4.2 新冠病毒刺突蛋白与宿主 ACE2 结合机制的计算研究

4.2.1 研究背景

新型冠状病毒 (SARS-CoV-2) 于 2019 年末爆发后迅速蔓延全球^[386-389], 但目前尚未发现可以治疗 SARS-CoV-2 引发的新冠肺炎 (COVID-19) 的特效药^[390-392]。SARS-CoV-2 的刺突 (Spike) 蛋白 (CoV-2-S) 以三聚体形式嵌在病毒表面, 介导病毒与宿主细胞人血管紧张素转换酶 2 (human angiotensin-converting enzyme 2, ACE2) 的结合, 从而帮助病毒入侵和感染宿主细胞^[393-394], 是目前许多疫苗和抗体药物的重要靶点之一。多项研究通过冷冻电子显微镜技术发现 CoV-2-S 在病毒表面具有灵活的构象动态, 尤其是 CoV-2-S 的受体结合域 (Receptor-binding domain, RBD) 存在明显的构象变化^[395-402]。当 RBD 处于闭合的‘向下’构象时, ACE2 受体在 RBD 上的结合空间被 CoV-2-S 三聚体中的其他单体占据, 因此该种构象无法与 ACE2 受体结合, 我们以残基 D405-V622-V991 定义此时的 RBD 角度在 30° 左右; 而当 RBD 结构由‘向下’构象转变成开放的‘向上’状态后, ACE2 受体在 RBD 上的结合空间暴露, CoV-2-S 三聚体的该种构象能够与 ACE2 受体结合, 此时 RBD 角度大于 50° 。

许多研究报道 CoV-2-S 的 RBD 与 ACE2 受体的结合能力比 SARS-CoV 病毒刺突蛋白 (CoV-S) 更强^[395, 403-408], 这或许是 SARS-CoV-2 传染性强的原因之一。但近期一项发表在 *Proc. Natl. Acad. Sci. U.S.A* 上的研究通过体外测试 (Pull down) 表明, CoV-2-S 全长结构的 ACE2 结合能力比 CoV-S 弱^[409]。另外一篇 *Cell* 文章通过 Blitz 实验表明 CoV-2-S 的 ACE2 结合能力与 CoV-S 相当^[410]。这两篇文章的实验结果与之前诸多 CoV-2-S RBD 上的实验数据相矛盾, 使得之前看似泾渭分明的 CoV-2-S 结合能力变得扑朔迷离, 因此研究 CoV-2-S 与 ACE2 的作用机制以理解这些矛盾的实验数据至关重要。

由于刺突蛋白需要发生大规模的构象变化才能与 ACE2 结合, 因此我们推测 CoV-2-S 和 CoV-S 在全长结构和 RBD 结构上不同的相对 ACE2 结合能力可能与其构象动态差异有关。然而, 我们难以通过现有的 CoV-2-S 三维结构实验数据得到准确的构象变化路径, 更加难以定量分析 CoV-2-S 和 CoV-S 各自的构象转变自由能垒。从 PDB 数据库中我们发现 CoV-2-S 和 CoV-S 结合 ACE2 时呈现出不同的 RBD 开放构象, 即 RBD 角度涵盖 54.8° 到 84.8° ^[395-402]。对于理解刺突蛋白与 ACE2 的作用机制, 了解刺突蛋白能够结合 ACE2 的最小 RBD 角度以及不

同 RBD 角度构象下的 ACE2 结合能力至关重要，然而基于目前的实验数据难以获悉这些问题的解答。

为深入研究 CoV-2-S 与 ACE2 的作用机制及其构象变化，我们通过综合同源模建、常规分子动力学模拟、增强采样方法、自由能计算、口袋探测和分子对接等多种计算模拟方法比较研究了 CoV-2-S 和 CoV-S 的大规模构象转变及其自由能变化；计算了它们处于不同 RBD 角度构象时与 ACE2 的结合能力；探索了 CoV-2-S 的可药性口袋及其动态变化；并评估了 CoV-2-S 的糖基化对其构象转变自由能和 ACE2 结合能力的影响。结果表明，CoV-2-S 的 RBD 结构需要从‘向下’构象开放至 RBD 角度 52.2° 时才能与宿主 ACE2 结合，并且随着 RBD 角度的增大，CoV-2-S 的 ACE2 结合能力也呈现增强趋势。在相同 RBD 角度构象下，CoV-2-S 的 ACE2 结合能力在全长结构和 RBD 上均比相同状态下的 CoV-S 强，其 RBD-ACE2 作用界面上的多个残基变异对 CoV-2-S 增强的 ACE2 结合能力起关键作用，包括残基 Q493 和 N487 等。但是，相比 CoV-S，CoV-2-S 在溶液中的 RBD ‘向上’构象明显更少，RBD 结构由‘向下’构象转变成‘向上’构象所需克服的自由能垒也更高。因此，这些结果表明尽管 CoV-2-S 的 RBD 结构具有比 CoV-S 更强的 ACE2 结合能力，但 CoV-2-S 更不容易从‘向下’构象转变成‘向上’构象，溶液中 RBD ‘向上’构象的比例也更少，因此在一些 CoV-2-S 全长结构的实验测试中出现比 CoV-S 更弱或者相当的 ACE2 结合能力。另外，通过研究 CoV-2-S 的构象转变路径及其口袋动态变化，我们发现了 5 个与 CoV-2-S 构象变化运动相关的口袋，可用于 CoV-2-S 的抑制剂虚拟筛选。

4.2.2 研究方法

体系准备：在本章中，我们比较研究了 CoV-2-S 和 CoV-S 的 RBD 和全长三聚体结构的 ACE2 结合能力。对于 CoV-S，RBD-ACE2 复合物(PDB ID: 2AJF^[411])和全长三聚体-ACE2 复合物(PDB ID: 6ACG^[394], 6ACK^[394], 6ACJ^[394], 6CS2^[412])的模拟初始结构均从 PDB 数据库中直接下载获取。对于 CoV-2-S，RBD-ACE2 复合物的模拟初始结构从 PDB 数据库中获取 (PDB ID: 6M0J^[413])，而 ACE2 结合的 CoV-2-S 三聚体实验结构尚未获得，因此我们通过 SWISS-MODEL 网站^[414]，以上述四个 ACE2 结合的 CoV-S 三聚体复合物为同源模建模板构建 ACE2 结合 CoV-2-S 三聚体复合物结构。为评估这四个 CoV-2-S 三聚体同源模建结构的可靠

性，我们计算了这四个建模结构的拉氏图。如图 4.10 所示，在 CoV-2-S 三聚体建模结构中低于 2%的残基位于二面角分布异常区域，表明其结构的可靠性。为比较 CoV-2-S 三聚体同源建模结构与目前能获取的 CoV-2-S 实验结构的差异，我们将 RBD-ACE2 复合物实验结构(PDB ID: 6M0J)与三聚体空蛋白实验结构(PDB ID: 6VSB)以两者的 RBD 为基础进行叠合，得到一个完整的 ACE2 结合的 CoV-2-S 三聚体结构，称之为叠合构象。结果表明，四个 CoV-2-S 三聚体建模结构与叠合构象的 RBD 闭合单体的 RMSD 均低于 0.70 Å，而由于 RBD 开放角度的不同，建模结构与叠合构象的 RBD 开放单体的 RMSD 较大，6ACG、6ACJ、6CS2 和 6ACK 的建模结构与叠合构象的 RBD 开放单体的 RMSD 依次为 5.81、3.50、4.77 和 6.18 Å。为了比较评估 CoV-2-S 与 CoV-2-S 的相对 ACE2 结合能力是否受 RBD 角度的影响，我们需保证 CoV-2-S 与 CoV-S 的模拟初始构象的 RBD 角度相近，因此，我们以这四个基于 CoV-S 的建模结构作为 CoV-2-S 模拟初始构象。

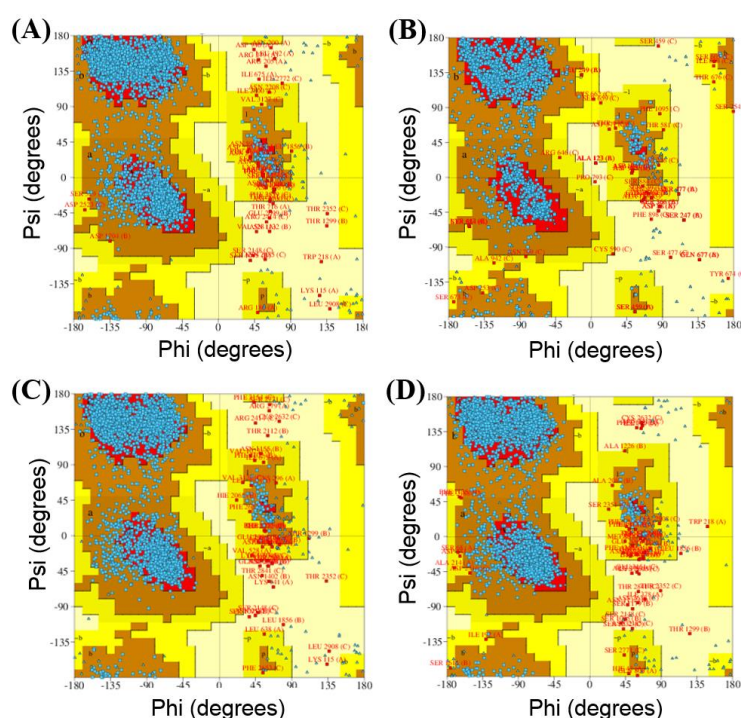


图 4.10 同源建模结构的拉氏图分析。以 (A) 6ACG, (B) 6ACK, (C) 6ACJ, (D) 6CS2 为模板。拉氏图通过 PDBsum (www.ebi.ac.uk/thornton-srv/databases/pdbsum) 网站获得。

Figure 4.10 Ramachandran plots of the modelled structures of (A) 6ACG, (B) 6ACK, (C) 6ACJ, (D) 6CS2, which were calculated by PDBsum.

为模拟溶剂环境，每个模拟体系被置于 TIP3P 的立方体盒子中，盒子边缘距离溶质至少 12 Å，并加入适量平衡离子（Na⁺和 Cl⁻）中和体系电荷。我们通过 Amber ff03 力场^[415]对蛋白进行参数化。为优化模拟初始结构，消除体系中的错误连接，我们对模拟体系进行 10,000 步的能量最小化，其中包括 5,000 步最速下降法和 5,000 步共轭梯度法。之后每个模拟体系在 0.2 ns 内从 0 K 升温至 300 K，并且在 300 K 下进行 NPT 系综平衡 0.1 ns。体系的能量最小化、升温以及平衡过程都是在 Amber16 软件的 *sander* 模块下进行，并对体系重原子施加大小为 10 kcal/mol/Å² 的弱限制力。

常规分子动力学和 vsREMD 模拟：为分析 ACE2 与刺突蛋白 RBD 结构相互作用的动态变化，我们分别对 ACE2-RBD 复合物和 ACE2 结合的刺突蛋白三聚体复合物均通过 Amber 16 软件的 *pmemd.cuda* 模块进行了 100 ns 的常规 MD 模拟。模拟过程中，通过 Langevin 动力学控温，SHAKE 算法^[416]限制重原子氢键，范德华作用截断值为 12 Å，PME 算法处理长程静电相互作用^[417]。

通过高效分子动力学模拟方法 vsREMD，我们比较研究了 CoV-2-S 与 CoV-S 三聚体的构象变化。在 vsREMD 模拟中，一系列副本在温度梯度下平行模拟，每个副本均在显式溶剂中进行体系运动，但在计算副本交换概率时通过速度调节的方式忽略溶剂内部相互作用能，从而使得副本数目显著减少^[14]。在本章中，为保证总平均交换律在 0.2 到 0.5 之间，我们采用 48 个模拟副本，在 300 到 450 K 的温度梯度下进行（300.0, 302.6, 305.2, 307.9, 310.5, 313.2, 315.9, 318.7, 321.4, 324.2, 327.0, 329.8, 332.7, 335.6, 338.5, 341.4, 344.4, 347.4, 350.4, 353.4, 356.5, 359.6, 362.7, 365.8, 369.0, 372.2, 375.4, 378.7, 381.9, 385.3, 388.6, 392.0, 395.4, 398.8, 402.2, 405.7, 409.3, 412.8, 416.4, 420.0, 423.6, 427.3, 431.0, 434.7, 438.5, 442.3, 446.1, 450.0）。48 个模拟副本的初始构象分别从 CoV-2-S 和 CoV-S 的构象变化路径中以均一间隔提取得到。模拟过程中每 1000 步进行一次交换尝试，每个副本模拟了 100 ns。

结合自由能计算和能量分解：我们通过 MM/GBSA 方法比较评估了 CoV-2-S 和 CoV-S 的 ACE2 结合自由能。MM/GBSA 的方法原理在章节 2.1 部分已详细介绍。本章研究中，构象熵变通过 Amber16 软件的 *Nmode* 模块计算得到。MM/GBSA 采用 OBC 溶剂模型（igb = 5, mbondi2）^[418]，溶剂介电常数为 80。

由于溶质介电常数 (ϵ_{in}) 会影响 MM/GBSA 的计算结果, 因此我们测试了三种不同的 ϵ_{in} ($\epsilon_{in}=1, 2, 4$)。计算结果如表 S4.7 所示, 在这三种不同的溶质介电常数下, CoV-2-S 与 ACE2 的结合自由能均强于 CoV-S, 并且不受 RBD 角度的影响。另外, 在侯廷军等人的研究中发现, 当 ϵ_{in} 为 1.0 时, MM/GBSA 方法在蛋白-蛋白作用体系中的预测准确性最高^[419], 因此我们后续计算均设置 ϵ_{in} 为 1.0。为模拟生理环境, 我们在 MD 模拟以及 MM/GBSA 自由能计算时均采用 0.1M 盐离子浓度。为研究 CoV-2-S 与 ACE2 作用的关键残基, 我们将总结合自由能进一步分解至每个蛋白残基的贡献, 计算参数均采用默认值。

构象转变路径预测: 通过课题组发展的 NUMD 蛋白质构象变化路径预测方法, 我们预测了 CoV-2-S 蛋白由闭合构象转变成开放构象的运动路径, 模拟参数采用默认值^[420]。在 NUMD 方法中, 以闭合构象为起点, 开放构象作为参考构象, 通过多次迭代的简振模式分析 (NMA) 得到 CoV-2-S 蛋白的运动路径。根据公式 4.1, 中间构象 $R^{(k)}$ 在第 k 次迭代时, 由上一步迭代 ($k-1$) 的中间构象 $R^{(k-1)}$ 与多个低频本征模式组合得到。

$$R^{(k)} = R^{(k-1)} + v^{(k)} = R^{(k-1)} + S^{(k)} \sum_i^{m^{(k)}} (d^{(k-1)} \cdot u_i^{(k)}) u_i^{(k)} \dots (4.1)$$

公式 4.1 中 $v^{(k)}$ 是由 $m^{(k)}$ 个低频本征模式组合而成。对于第 i_{th} 个本征模式, 相对初始结构的位移正比于 $d^{(k-1)} u_i^{(k)}$, $d^{(k-1)}$ 在本征向量 $u_i^{(k)}$ 上以 $S^{(k)}$ 缩放的瞬时距离向量。在本章研究中, 缩放因子设置为 10.0^{21} 。初始构象和参考构象分别是以 CoV-S 三聚体 ‘向下’ 构象 (PDB ID: 5X58^[421], RBD 角度为 31.6°) 和 ‘向上’ 构象 (PDB ID: 5X5B^[421], RBD 角度为 84.8°) 为模板的同源建模结构。为了进一步探索 CoV-2-S 蛋白构象变化路径中的可药性结合位点, 以用于抗 CoV-2-S 蛋白抑制剂筛选, 通过 D3Pockets 网站, 我们分析了 CoV-2-S 蛋白潜在药物结合位点的动态变化 (<https://www.d3pharma.com/D3Pocket/index.php>)。

自适应拉伸分子动力学模拟 (ASMD)。 ASMD 的方法原理在章节 2.3 中已详细介绍^[422]。本章研究中, ASMD 模拟的每个区间采用 20 条平行非平衡模拟, 拉伸反应路径设置为 RBD 角度 30° 到 85° , 速度为 40 rad/ns。弹簧系数设置为 50 kcal/mol/rad²。

4.2.3 结果与讨论

4.2.3.1 CoV-2-S 具有比 CoV-S 更强的 ACE2 结合自由能

首先我们分别以 CoV-2-S 和 CoV-S 的 ACE2-RBD 复合物晶体结构为初始构象进行了 100 ns 的分子动力学模拟。根据它们的 RMSD 曲线判定复合物体系在 100 ns 内已达收敛状态，并基于 RMSD 稳定区间（50-100 ns）的轨迹，我们计算了 ACE2-RBD 的相互作用能 (ΔG)。结果如表 4.1 所示，CoV-2-S 与 ACE2 的结合自由能为 -24.86 ± 0.59 kcal/mol，明显强于 CoV-S (-10.04 ± 0.66 kcal/mol)。CoV-2-S 与 CoV-S 的总结合自由能差异最大部分为静电作用能 (E_{ele})，表明 CoV-2-S 更强的 ACE2 结合能力主要来自于 CoV-2-S 与 ACE2 增强的静电相互作用，这一结果与 Muhamed A. 等人利用 DELPHI 方法计算的结果一致，即 CoV-2-S 与 ACE2 更强的亲和力源自它们更强的静电相互作用^[405]。

表 4.1 CoV-2-S RBD 与 CoV-S RBD 的计算 ACE2 结合自由能。

Table 4.1 The predicted binding free energies (kcal/mol) of CoV-2-S RBD and CoV-S RBD to ACE2 *

Energy term	CoV-2-S RBD	CoV-S RBD
E_{vdw}	-86.91 ± 0.06	-80.73 ± 0.07
E_{ele}	-697.07 ± 0.56	-742.78 ± 0.71
E_{gb}	760.94 ± 0.51	812.86 ± 0.67
E_{np}	-12.05 ± 0.06	-10.34 ± 0.10
ΔH	-35.10 ± 0.62	-20.98 ± 0.64
$-T\Delta S$	-10.24 ± 0.56	-10.94 ± 0.69
ΔG	-24.86 ± 0.59	-10.04 ± 0.66

*: 统计误差根据 50-100 ns MD 模拟轨迹的计算结果获得。在 MM/GBSA 计算中，从 50-100 ns MD 模拟轨迹中均一提取 1000 个构象计算结合自由能，10 个构象计算熵效应。

另外，我们计算了 4 个 CoV-S 三聚体结构（6ACG、6ACJ、6ACK 和 6CS2）与 ACE2 的结合自由能。这 4 个结构的相同之处是都有两个未结合 ACE2 的‘向下’单体和一个结合 ACE2 的‘向上’单体，不同之处是‘向上’单体的 RBD 打开角度不同，跨度在 54.8° 到 84.6° ，这也表明了刺突蛋白的 RBD 并不需要完全向上打开就能与 ACE2 结合（表 S4.8）。由于 ACE2 结合的 CoV-S-2 三聚体实验结构目前尚不能直接获取，我们以这 4 个 CoV-S 三聚体结构为模板，通过同源模建构建 CoV-2-S 结构，在相同参数下进行 100 ns 的常规分子动力学模拟后，计算并比较了 CoV-S 和 CoV-S-2 三聚体结构与 ACE2 的结合能力。结果表明，

当 RBD 角度为 54.6° 时, CoV-2-S 三聚体与 ACE2 的计算 ΔG 为 -18.00 ± 0.84 kcal/mol, 而 CoV-S 三聚体的 ACE2 结合能力仅为 -10.59 ± 0.62 kcal/mol。并且, 当 RBD 角度为 68.4° (6ACJ)、 74.3° (6CS2) 和 84.6° (6ACK) 时, CoV-2-S 三聚体的 ACE2 能力均比 CoV-S 三聚体更强 (表 4.2)。

表 4.2 CoV-2-S 与 CoV-S 三聚体的计算 ACE2 结合自由能。

Table 4.2 The predicted binding free energy (kcal/mol) for CoV-S and CoV-2-S trimers binding to ACE2^a

	6ACG	6ACJ	6CS2	6ACK
RBD-angle ^b	$54.1^\circ \pm 1.24$	$68.0^\circ \pm 1.19$	$74.5^\circ \pm 0.98$	$84.4^\circ \pm 1.49$
CoV-S				
E_{vdw}	-74.67 ± 0.60	-84.99 ± 0.54	-81.48 ± 0.69	-86.15 ± 0.47
E_{ele}	18.69 ± 0.66	-53.59 ± 0.56	-109.67 ± 3.05	-120.10 ± 3.29
E_{gb}	34.91 ± 0.82	122.73 ± 0.68	173.18 ± 3.09	182.77 ± 3.26
E_{np}	-9.72 ± 0.08	-11.10 ± 0.07	-10.42 ± 0.10	-12.04 ± 0.06
ΔH	-24.79 ± 0.61	-26.96 ± 0.56	-28.39 ± 0.70	-35.52 ± 0.54
$-T\Delta S$	-14.20 ± 0.62	-14.57 ± 0.80	-16.85 ± 0.74	-14.89 ± 0.67
ΔG	-10.59 ± 0.62	-12.39 ± 0.68	-11.54 ± 0.72	-20.63 ± 0.60
CoV-2-S	$54.6^\circ \pm 1.02$	$68.4^\circ \pm 1.32$	$74.3^\circ \pm 1.27$	$84.6^\circ \pm 1.24$
E_{vdw}	-81.34 ± 0.47	-95.90 ± 0.53	-100.86 ± 0.72	-106.25 ± 0.56
E_{ele}	-816.09 ± 0.59	-796.20 ± 2.59	-763.73 ± 3.59	-763.10 ± 2.86
E_{gb}	875.10 ± 0.52	866.54 ± 2.59	830.25 ± 3.43	828.67 ± 2.66
E_{np}	-11.14 ± 0.07	-13.98 ± 0.06	-14.61 ± 0.08	-15.21 ± 0.05
ΔH	-33.47 ± 0.71	-39.55 ± 0.56	-48.95 ± 0.95	-55.89 ± 0.56
$-T\Delta S$	-15.47 ± 0.98	-14.45 ± 0.81	-16.55 ± 0.78	-16.37 ± 0.66
ΔG	-18.00 ± 0.84	-25.10 ± 0.68	-32.40 ± 0.86	-39.52 ± 0.61

^a: 统计误差根据 50-100 ns MD 模拟轨迹的计算结果获得。在 MM/GBSA 计算中, 从 50-100 ns MD 模拟轨迹中均一提取 1000 个构象计算结合自由能, 10 个构象计算熵效应。

^b: RBD 角度的平均值与标准差由模拟轨迹统计获得。

4.2.3.2 CoV-2-S 和 CoV-S 的 ACE2 相互作用关键残基

为探索 CoV-2-S 和 CoV-S 蛋白与 ACE2 相互作用时的关键作用残基, 我们将 MM/GBSA 计算的总结合自由能分解至每个残基上。如图 4.11 所示, 相比 CoV-S 蛋白, CoV-2-S 具有更多的氨基酸残基参与和 ACE2 的相互作用并且能量贡献大于 1.0 kcal/mol, 包括有残基 Y449, Q493 和 G496。其中残基 Q493 对 CoV-2-

S 与 ACE2 的相互作用具有 -2.64 kcal/mol 的能量贡献，而其对应的 CoV-S 上的残基 N479 并没有明显的能量贡献。这一发现与 Ghorbani 等人的 MM/PBSA 预测一致^[423]，同时也表明，相比 CoV-S，CoV-2-S 的 Q493 残基变异对于其增强的 ACE2 亲和力有重要作用。

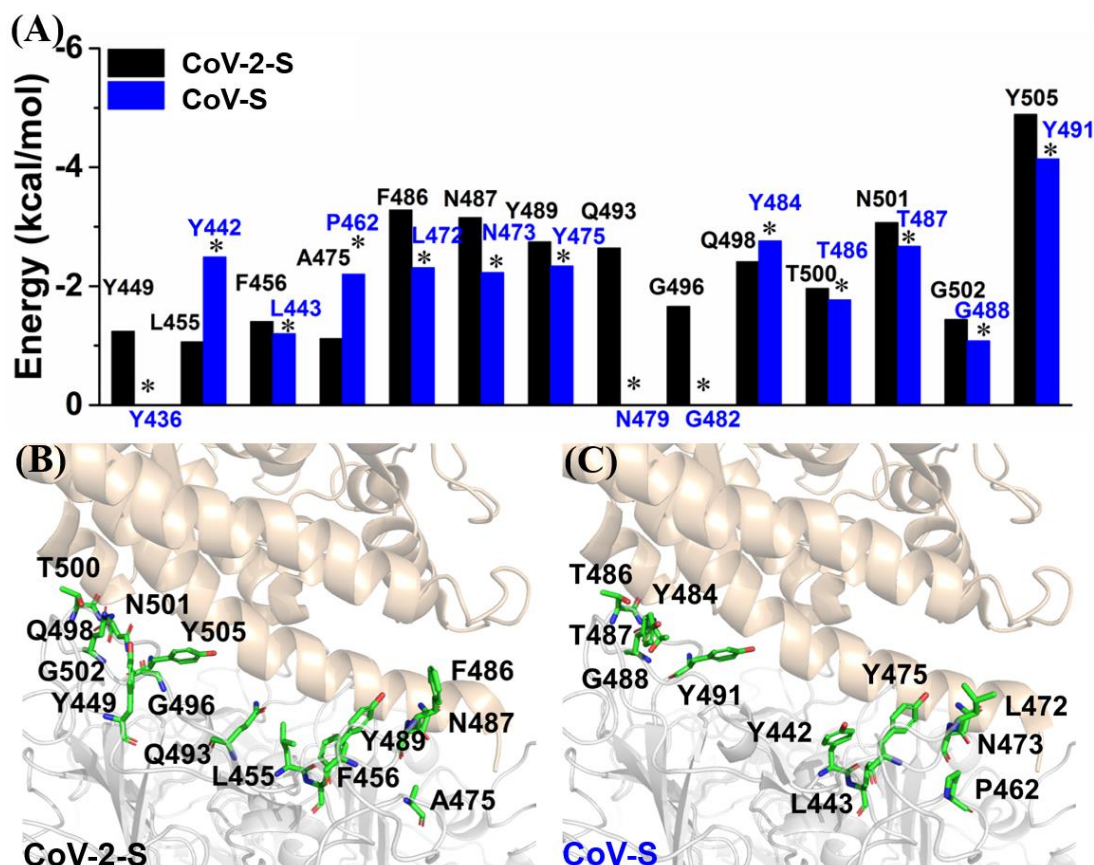


图 4.11 CoV-2-S 和 CoV-S 的 ACE2 相互作用图解。以 6ACG 为例，CoV-2-S 和 CoV-S 的 ACE2 相互作用关键残基能量贡献比较，统计上具有显著性差异的残基对用“*”标注 (A)。CoV-2-S (B) 和 CoV-S (C) 与 ACE2 相互作用的关键氨基酸残基结构展示。残基以绿色棍棒模型展示，其中每个残基对中结合自由能的贡献 ≤ -1.00 kcal/mol。

Figure 4.11 The spike-ACE2 interaction spectra of CoV-2-S and CoV-S. Taking 6ACG as an example, the comparison of key residues' energy contribution of CoV-2-S and CoV-S interacting with ACE2 (A). The difference that is statistically significant at the 1% level ($p < 0.01$) was labeled with “*”. Key residues of the CoV-2-S (B) and CoV-S (C) interacting with ACE2, were shown in sticks and colored green. Each residue contributed ≤ -1.00 kcal/mol to the overall binding free energy.

为研究这些关键作用残基参与 ACE2 相互作用的动态变化,我们计算了关键作用残基与 ACE2 形成的相互作用对的距离在模拟轨迹中的演变。如图 4.12 所示,对于 CoV-2-S 蛋白,残基 Y449、G496、T500、N501 和 G502 与 ACE2 形成了稳定的氢键作用;残基 F456 和 Y489 则与 ACE2 形成了稳定的疏水作用。另外,残基 Q493 也在模拟时间 44 ns 之后与 ACE2 蛋白上的残基 E35 形成了稳定的氢键作用;而残基 Y505 与 ACE2 蛋白上的残基 E37 形成的氢键作用相对不稳定。对于 CoV-S 蛋白,残基 T486、T487 和 G488 与 ACE2 蛋白形成了稳定的氢键作用(图 4.13)。为进一步研究 CoV-2-S 相对 CoV-S 更强的 ACE2 结合亲和力是否源自其 RBD-ACE2 作用界面的残基变异,我们将 CoV-S 蛋白 RBD-ACE2 作用界面上相对 CoV-2-S 的 18 个残基变异人为突变成 CoV-2-S 的相应残基,并计算突变后 CoV-S 蛋白的 ACE2 结合自由能。18 个残基突变分别为:K390R、R426N、S432V、T433G、Y442L、L443F、F460Y、P462A、D463G、K465T、L472F、Y476F、N479Q、D480S、Y484Q、T485P、T487N、I489V。以 6ACG 为例,突变 CoV-S 蛋白与 ACE2 蛋白的结合自由能为 -17.82 ± 0.59 kcal/mol,该结合自由能与 CoV-2-S 相当(-18.00 ± 0.84 kcal/mol),但远强于野生型 CoV-S 蛋白(-10.59 ± 0.62 kcal/mol)。因此,我们的计算结果表明,CoV-2-S 相对 CoV-S 更强的 ACE2 结合能力主要源自其 RBD-ACE2 作用界面的残基变异。

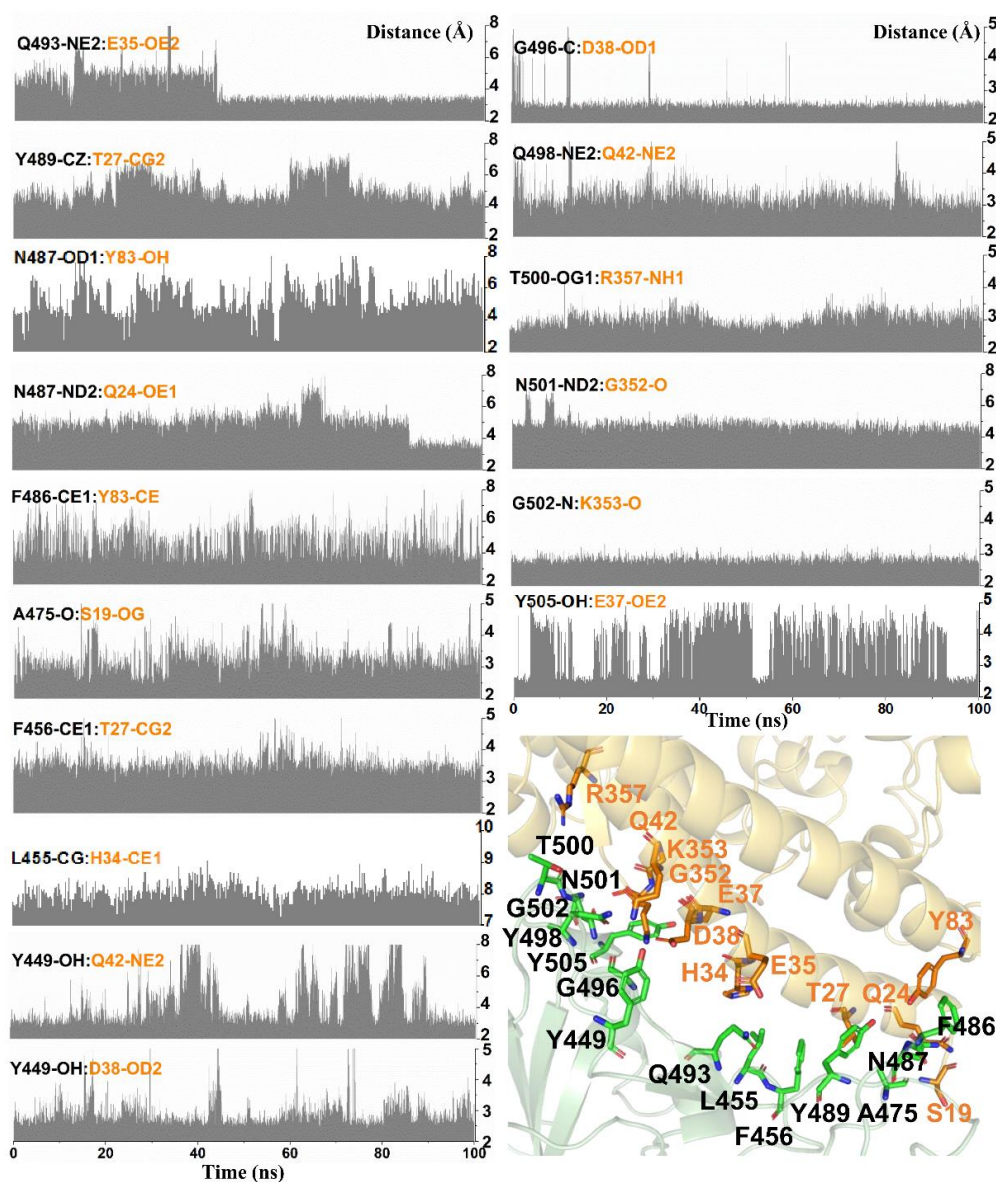


图 4.12 CoV-2-S 与 ACE2 的相互作用。关键残基作用对距离在 100 ns MD 轨迹中的演变（绿色棍棒模型为 CoV-2-S 残基，橙色棍棒模型为 ACE2 残基）。

Figure 4.12 The distance analyses of CoV-2-S and ACE2 interactions along the 100 ns MD trajectory. Key interactional residues of the CoV-2-S (green) and ACE2 (orange) were shown in sticks.

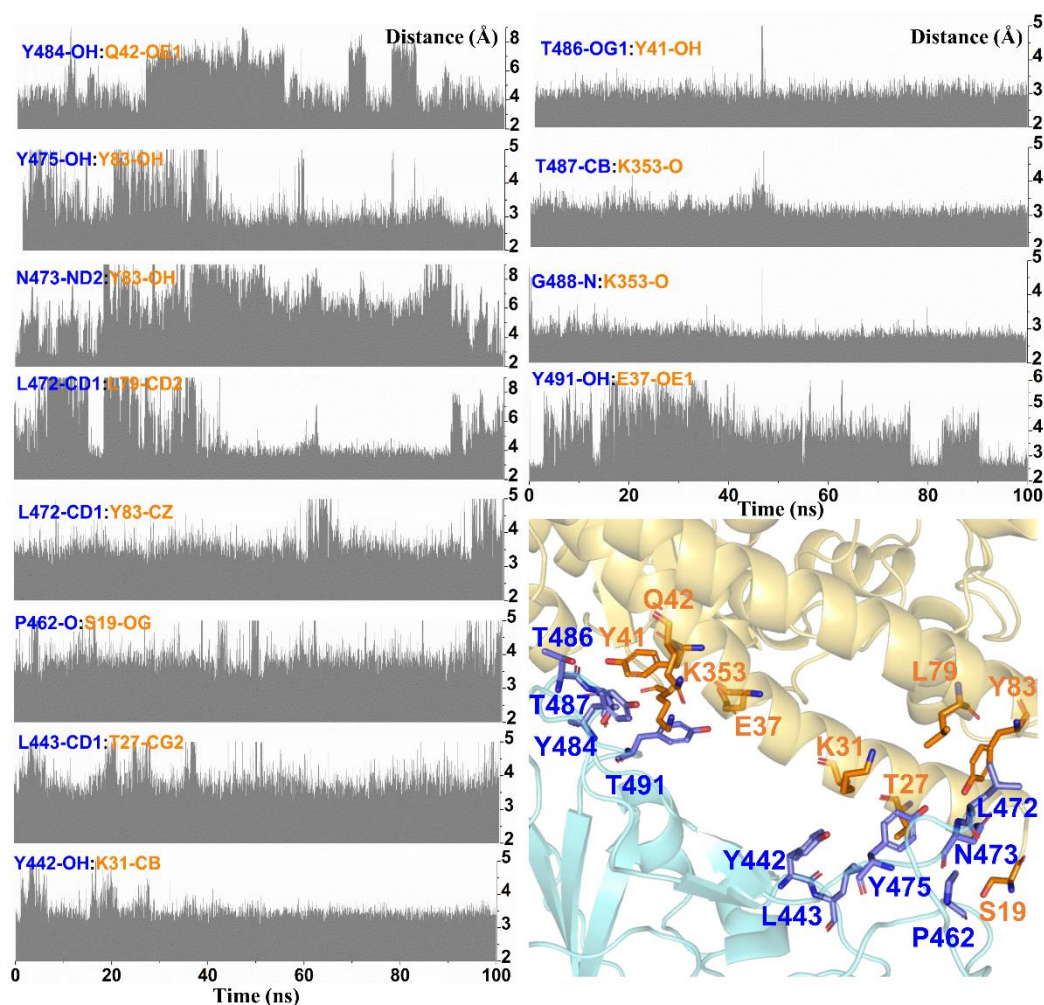


图 4.13 CoV-S 与 ACE2 的相互作用。关键残基作用对距离在 100 ns MD 轨迹中的演变（蓝色棍棒模型为 CoV-S 残基，橙色棍棒模型为 ACE2 残基）。

Figure 4.13 The distance analyses of CoV-S and ACE2 interactions along the 100 ns MD trajectory. Key interactional residues of the CoV-S (blue) and ACE2 (orange) were shown in sticks.

4.2.3.3 CoV-2-S 的构象变化以及 ACE2 蛋白可及性构象

CoV-2-S 的 RBD ‘向上’和‘向下’两种构象分别对应 ACE2 受体可及和不可及两种功能构象。如表 S4.8 所示，当 CoV-S 的 RBD 开放角度处于 54.8° 到 84.6° 之间，CoV-S 均能与 ACE2 结合。然而，通过尽管多但仍不完善的静态实验结构，我们仍然难以判定 ACE2 受体可及构象的最小 RBD 角度，以帮助基于结构的 CoV-2-S 虚拟筛选选择合适构象。为此，我们通过自主发展的 NUMD^[420]方法研究了 CoV-2-S 三聚体的构象转变路径，以预测 CoV-2-S 的 ACE2 可及构象的最小 RBD 角度。通过 RMSD 计算，我们发现 CoV-2-S 的 RBD 在构象转变路径

中的整体结构变化很小(RMSD 最大值小于 $1.30 \text{ \AA}$, 图 S4.1)。因此, 我们将 RBD-ACE2 的实验结构 (PDB ID: 6ACG) 以 RBD 结构为基础叠合至 CoV-2-S 三聚体的构象转变路径 (240 个中间构象, 图 4.14A)。结果表明, 当 CoV-2-S 三聚体的 RBD 角度从 31.6° 增加至 52.2° 后, ACE2 与 CoV-2-S 三聚体才开始没有原子碰撞。因此, 我们的模拟结果表明 CoV-2-S 三聚体的 ACE2 可及构象的最小 RBD 角度为 52.2° (图 4.14B), 这一角度也可用作后续区分 RBD ‘向上’ 和 ‘向下’ 构象的标准。另外, 得益于科研工作者的努力, 越来越多的 CoV-2-S 三聚体结构浮出水面, 而这些实验数据也正好验证了我们的模拟结果。例如, 目前 CoV-2-S 三聚体 ‘向下’ 构象实验结构中的 RBD 角度均小于 52.2° , 而 CoV-S 三聚体 ‘向上’ 构象实验结构的最小 RBD 角度为 54.8° (PDB ID: 6ACG), 大于 52.2° 。

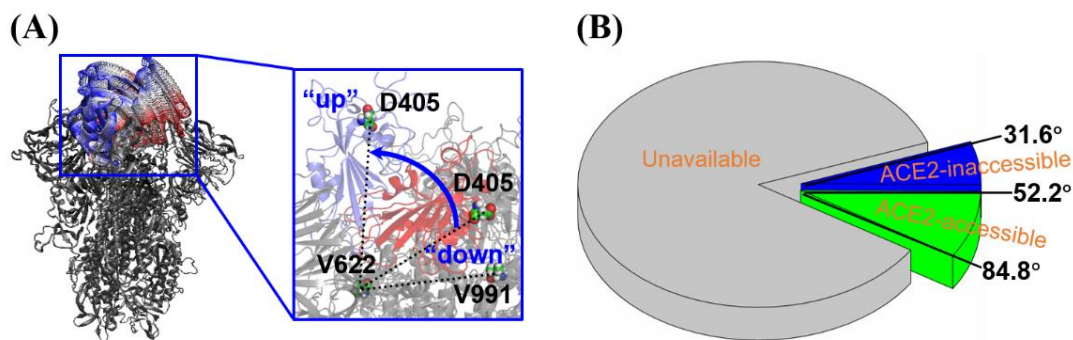


图 4.14 CoV-2-S 三聚体的构象转变。‘向下’ / ‘向上’ 构象转变路径 (20 个叠合构象为例, A)。(B) CoV-2-S 三聚体 ACE2 可及构象 RBD 角度区间 (蓝色)、ACE2 不可及构象 RBD 角度区间 (绿色) 和采样未获得构象 (灰色)。

Figure 4.14 The predicted conformational change pathway of CoV-2-S trimer. Twenty aligned conformations were extracted from between “down” and “up” states (A). The ACE2-inaccessible RBD-angle (blue), ACE2-accessible RBD-angle (green), and unavailable RBD-angle (gray) of CoV-2-S trimer (B).

4.2.3.4 CoV-2-S RBD 角度与其 ACE2 结合能力的相关性

基于 CoV-2-S 三聚体的构象转变路径, 我们进一步研究了 CoV-2-S 的 RBD 角度与其 ACE2 结合能力的相关性。如图 4.15 所示, 随着 RBD 角度的增加, CoV-2-S 与 ACE2 的结合自由能也增强, 并且它们之间的线性相关系数 (R^2) 为 0.64。通过结合模式分析, 我们发现尽管在不同的 RBD 角度构象中, CoV-2-S 与 ACE2 的结合模式相似, 但存在一些残基侧链的明显变化, 例如残基 N487、Q498

和 N501 在 RBD 角度增加时与 ACE2 形成的氢键作用也增强, 并且残基 Q493 在大 RBD 角度构象中除了与 ACE2 蛋白的残基 E35, 还额外与残基 K31 形成氢键作用 (图 4.16)。因此, 我们的模拟结果表明, 当 CoV-2-S 的 RBD 打开至 52.2° 时就能够与 ACE2 结合进而侵染宿主细胞, 而且其 ACE2 结合能力随 RBD 角度的增大而增强。当 CoV-2-S 处于完全 ‘向上’ 状态 (RBD 角度为 84.6°) 时, 计算得到的 ACE2 结合自由能为 -39.52 kcal/mol , 明显高于 RBD 角度 52.2° 时的结合自由能 (-19.86 kcal/mol)。另外, 在 CoV-S 三聚体上也有相同规律, 即 ACE2 结合自由能随着 RBD 角度的增加而呈现增强趋势, 其相关性系数为 0.58 (图 4.17)。

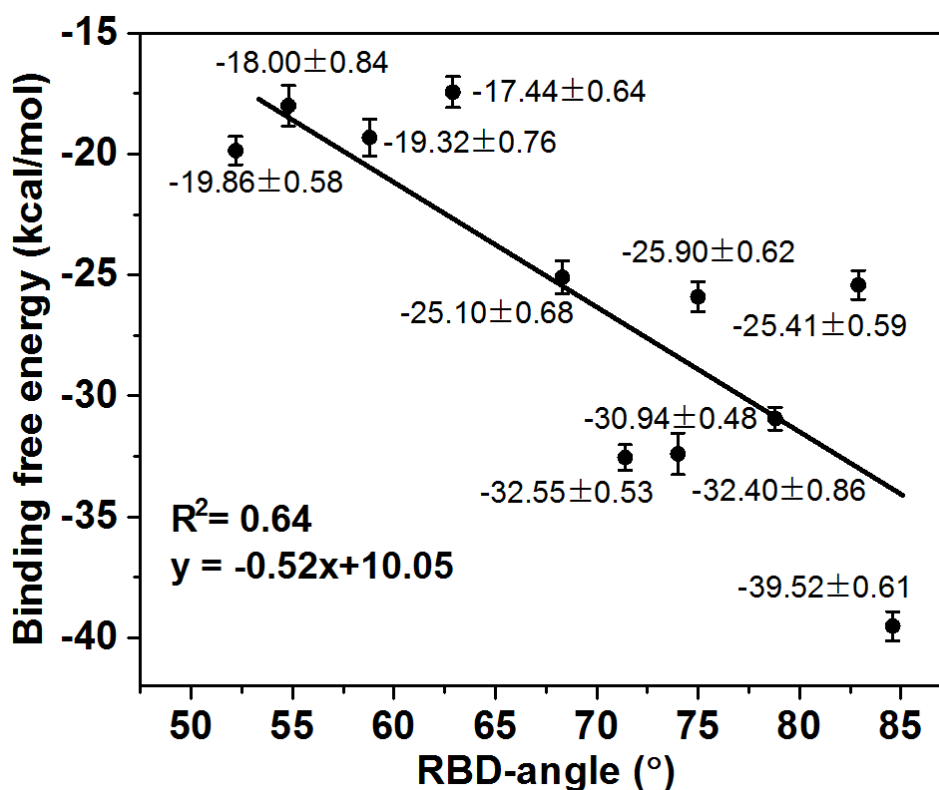


图 4.15 CoV-2-S 三聚体与 ACE2 的计算结合自由能随 RBD 角度的变化。

Figure 4.15 The calculated binding free energy of CoV-2-S trimer to ACE2 against the RBD-angle.

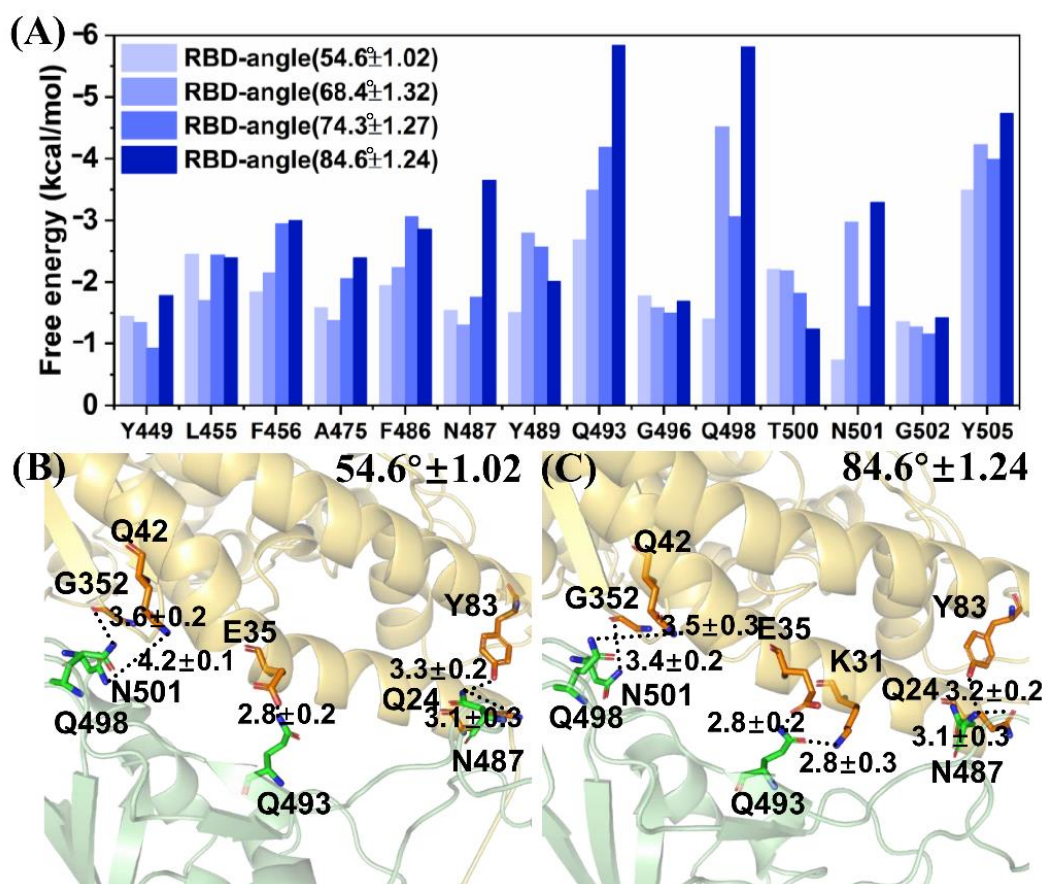


图 4.16 CoV-2-S 与 ACE2 作用关键残基的能量贡献变化。(A) 14 个关键残基的能量贡献，其中 4 个残基 (N487、Q498、Q493 和 N501, 绿色) 在小 RBD 角度 (B) 和大 RBD 角度 (C) 构象的 ACE2 相互作用模式 (ACE2 残基为橙色) 具有差异。

Figure 4.16 The ACE2-binding free energy decomposition into residues of CoV-2-S with four different RBD-angles. (A) The 14 key residues of CoV-2-S was labeled. Four residues (N487, Q498, Q493, and N501, colored green) that significantly form different interactions with residues in ACE2 (colored orange) were shown in sticks for CoV-2-S with small RBD-angle (B) and large RBD-angle (C).

4.2.3.5 CoV-2-S 与 CoV-S 的构象分布自由能图景比较

正如本章研究背景所述，近期在全长 CoV-2-S 的活性评估表明 CoV-2-S 的 ACE2 结合能力与 CoV-S 相当或者更弱，这与之前在 CoV-2-S RBD 上的实验结果截然相反^[409-410]。针对这一看似矛盾的结果，我们推测可能是因为在溶液中 CoV-2-S 的 ACE2 可结合状态构象比例少或者 CoV-2-S 的 RBD 从‘向下’构象转变成‘向上’构象所需的能垒更高。为了验证我们的推测，我们分别对 CoV-2-S 和 CoV-S 三聚体体系进行了 4.8 μ s 的 vsREMD 模拟^[14]，来比较它们的构象动态特性。

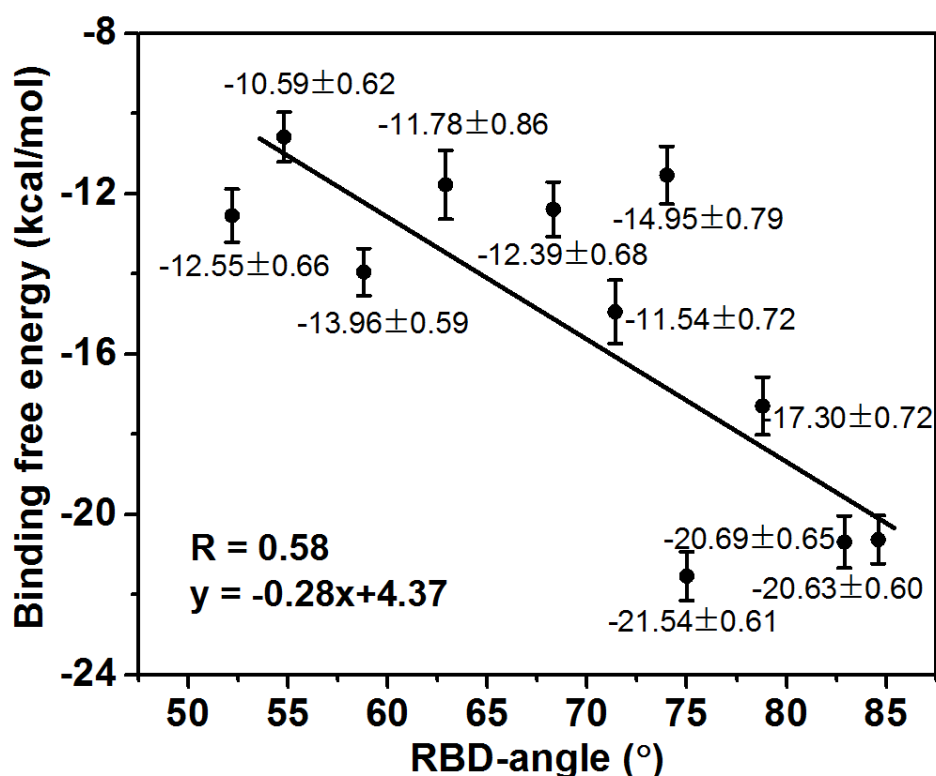


图 4.17 CoV-S 三聚体与 ACE2 的计算结合自由能随 RBD 角度的变化。

Figure 4.17 The calculated binding free energy of CoV-S to ACE2 against the RBD-angle.

通过比较 90 ns 和 100 ns 模拟轨迹的自由能图景差异，我们判定 vsREMD 模拟在 90 ns 后已达到收敛（图 S4.2&S4.3）。通过残基均方根误差分析（RMSF），我们发现定义 RBD 角度的三个残基中只有位于 RBD 结构上的残基 D405 随着 RBD 构象的改变具有较大的 RMSF（图 4.18）。因此，我们以 RBD 角度和 CoV-2-S 骨架原子 RMSD 为坐标绘制 CoV-2-S 蛋白的二维自由能图景。如图 4.19 所示，CoV-2-S 相对更稳定于 ACE2 不可及的 RBD ‘向下’ 构象，ACE2 可及的 RBD ‘向上’ 构象（RBD 角度 $\geq 52.2^\circ$ ）只占溶液中构象的 5.48%，而 CoV-S 的 RBD ‘向上’ 构象占比达到 22.7%，因此，我们的模拟结果说明在溶液中 CoV-2-S 的 ACE2 可结合构象比 CoV-S 少。近期，Lai 等人通过 350 ns 常规 MD 模拟也发现 CoV-2-S 蛋白的柔性比 CoV-S 弱^[424]；Ke 等人通过冷冻电子显微镜（cryo-EM）技术观察到新冠病毒表面 53% 的 CoV-2-S 三聚体预融合构象处于三个单体均 RBD ‘向下’ 构象^[425]，这些已报道的模拟结果和实验数据正好佐证 vsREMD

的模拟结果，即 CoV-2-S 的 RBD 结构更稳定于‘向下’构象，而且溶液中 CoV-2-S 的 RBD ‘向上’构象比 CoV-S 少。

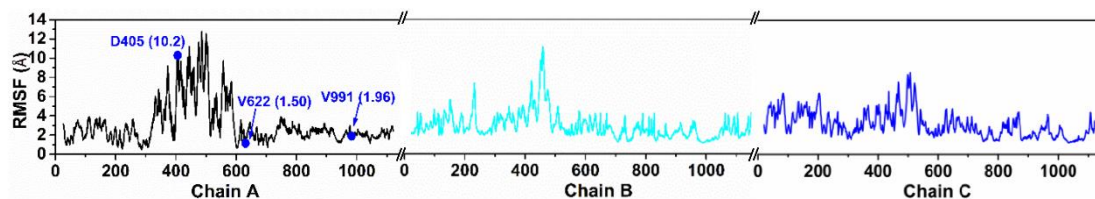


图 4.18 CoV-2-S 三聚体构象转变中的残基均方根误差。

Figure 4.18 The root mean square fluctuation (RMSF) of each residue in CoV-2-S trimer during its conformational change between the “down” and “up” states.

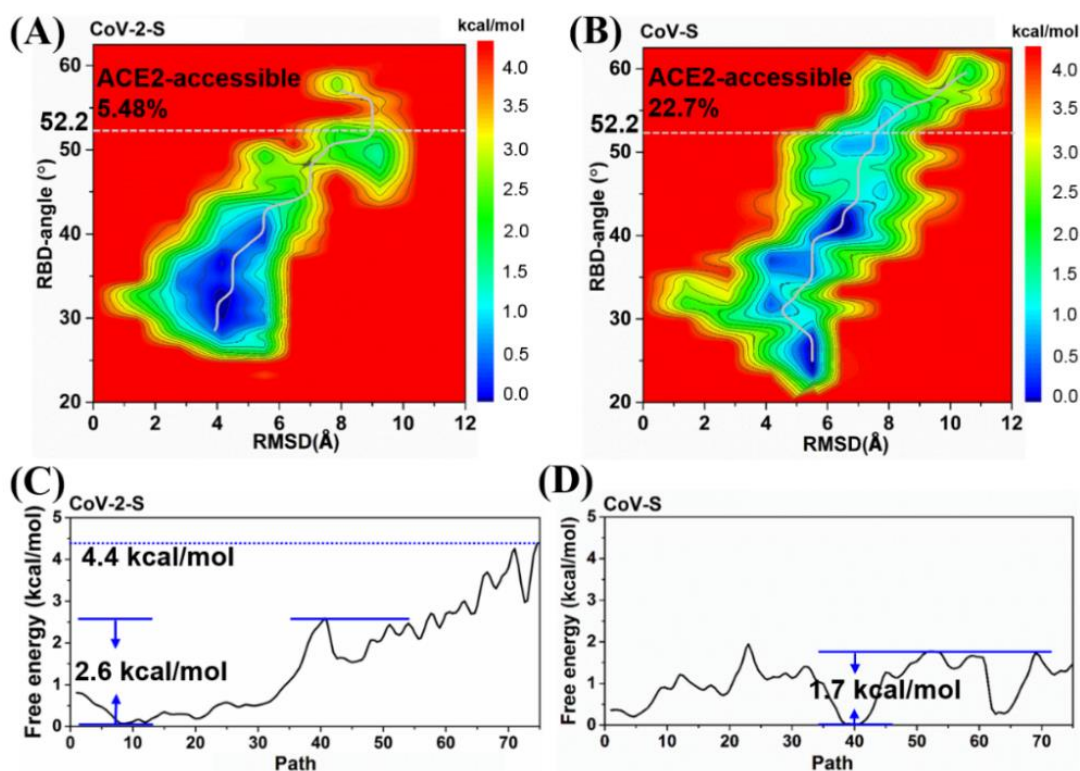


图 4.19 CoV-2-S 与 CoV-S 三聚体的构象分布。以相对初始构象的 RMSD 和 A 链结构的 RBD 角度为坐标绘制的 CoV-2-S (A) 和 CoV-S (B) 自由能图景，以及‘向下’构象转变至‘向上’构象的最低能量路径。CoV-2-S (C) 和 CoV-S (D) 构象变化路径的自由能变化。

Figure 4.19 The conformational distribution of CoV-2-S and CoV-S trimers. The free energy surface of CoV-2-S (A) and CoV-S (B) trimers as a function of RMSD against initial structures and RBD-angles of chain A, together with minimum energy pathway between “down” and “up” states. The free energy along the conformational change pathway from “down” to “up” states for CoV-2-S (C) and CoV-S (D) trimers.

我们进一步分析了 CoV-2-S 和 CoV-S 构象分布自由能图景中的 RBD‘向下’和‘向上’构象间的最低能量路径，结果发现 CoV-2-S 所需克服的能垒约为 2.6 kcal/mol，明显比 CoV-S 高 (~1.7 kcal/mol)，说明相比 CoV-S，CoV-2-S 更不容易转变成 ACE2 可结合状态（图 4.19C&D）。在图 4.19C 中，CoV-2-S 在 RBD 角度增大时，所需克服的能垒逐步增加，当转变成 RBD 完全‘向上’状态时至少需要克服 4.4 kcal/mol 的能垒，而 CoV-S 只需 1.7 kcal/mol（图 4.19D），说明 CoV-2-S 相比 CoV-S 更不容易变成 RBD 完全‘向上’状态。因此，目前的模拟结果说明全长 CoV-2-S 相对 CoV-S 表现出更弱的 ACE2 结合能力的原因之一可能是在溶液中其更少的 ACE2 可结合构象比例，以及更高的构象转变自由能垒。

CoV-2-S 三聚体结构具有单体结构多样性，其三个单体均为 RBD‘向下’时，CoV-2-S 不能与宿主 ACE2 结合，而只要有一个单体的 RBD 打开后，CoV-2-S 三聚体结构就能与 ACE2 结合。因此，ACE2 结合 CoV-2-S 三聚体除其中一个单体处于 RBD‘向上’构象外，也有两个或者三个单体同时 RBD‘向上’。为研究 CoV-2-S 三聚体中 RBD 的‘向上’打开运动是依次发生还是同时发生，我们对 vsREMD 模拟轨迹进行分析。结果发现，当 CoV-2-S 三聚体中的 A 链处于 RBD‘向上’构象时，C 链的 RBD 处于‘向下’/‘向上’转变的中间构象（图 4.20A），说明 CoV-2-S 三聚体的 RBD 开放运动是依次发生的。另外，我们发现当 A 链 RBD 处于‘向下’/‘向上’中间构象时，C 链的 RBD 角度明显比 B 链大（图 4.20B），说明当 A 链处于 RBD‘向上’构象时，C 链比 B 链更倾向于 RBD 开放运动。通过对三聚体的结构分析我们发现，当 CoV-2-S 三聚体的 A 链 RBD 从‘向下’转变成‘向上’构象时，A 链 NTD（N-端结构域）会靠近蛋白核心下移 ~2.4 Å，而其 NTD 结构毗邻 C 链 RBD，因此造成 C 链 RBD 相比 B 链更容易转变成‘向上’构象。

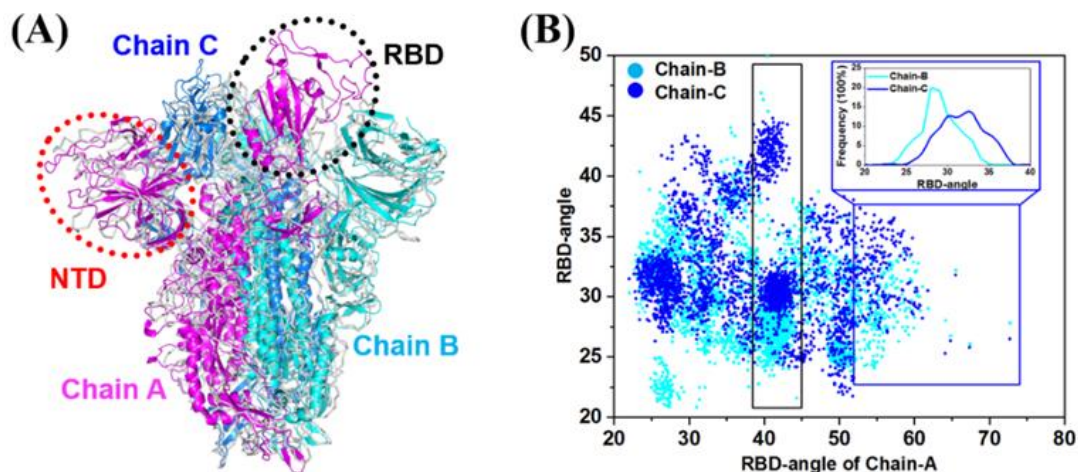


图 4.20 CoV-2-S 三聚体的单体关联运动。CoV-2-S 三聚体中 A 链处于 RBD ‘向上’ (粉色)、C 链为 RBD ‘向下’ / ‘向上’ 中间构象 (蓝色) 和 B 链 RBD ‘向下’ 构象 (灰色) 的代表构象 (A)。A 链与 B/C 链的 RBD 角度的关系 (B)。

Figure 4.20 The coordinated motions among the three chains of CoV-2-S trimer.

(A) a representative conformation of CoV-2-S trimer with one RBD in “open” state (chain A, colored pink) while another RBD in “closed/open” transition (chain C, colored blue), compared with three “closed” RBD conformation (colored gray). The correlation between the RBD-angles ($^{\circ}$) in chain A and chain B or chain C during the vsREMD simulations (B).

4.2.3.6 CoV-2-S 的可药性结合位点预测

将 CoV-2-S 构象限制在 ACE2 不可及状态能够达到抑制 CoV-2-S 结合 ACE2 的效果。例如，亚油酸被报道能够使 CoV-2-S 的 RBD 三聚体结构处于紧密的 ‘向下’ 状态，从而抑制 CoV-2-S 的活性^[426]。为了发现更多 CoV-2-S 上的可药性结合位点，通过 NUMD 方法我们预测了 CoV-2-S 的 RBD 构象转变路径，并通过 D3Pockets^[427] 程序预测了 CoV-2-S 的可药性结合位点及其动态变化。如图 4.21A&B 所示，我们发现 CoV-2-S 上存在 5 个口袋与 RBD 的构象变化动态相关。例如，口袋 2 与口袋 4 存在正相关性 ($R^2=0.88$)，即表明随着 RBD 角度的打开，口袋 2 与口袋 4 的体积同时增大 (图 4.21C)。因此通过小分子抑制口袋 2 或口袋 4 的体积增大能够起到阻止 CoV-2-S 变成 ACE2 可结合构象的作用。

另外通过文献检索，我们发现 8 个化合物报道具有刺突蛋白抑制活性，但它们的结合位点尚不清楚 (表 4.3)。我们将这 8 个化合物对接至预测的 5 个结合位点，结果发现，这 8 个化合物结合在这五个口袋的可能性极大。其中化合物槲皮素 (quercetin) 能够较好地结合在 RBD ‘向下’ 构象的口袋 5 中，其对接分数为

-9.04 kcal/mol, 在 CoV-2-S 表面的 39 个可结合位点中排名第一 (图 4.22A-C)。槲皮素能够与 CoV-2-S 的 RBD 形成氢键作用, 并且与该 RBD 毗邻的 NTD 上的残基也有相互作用 (S324A, F322A, R321A, N329A 和 T167B), 因此槲皮素的在口袋 5 的作用模式也说明槲皮素具有限制 CoV-2-S 于‘向下’构象的潜力 (图 4.22D)。

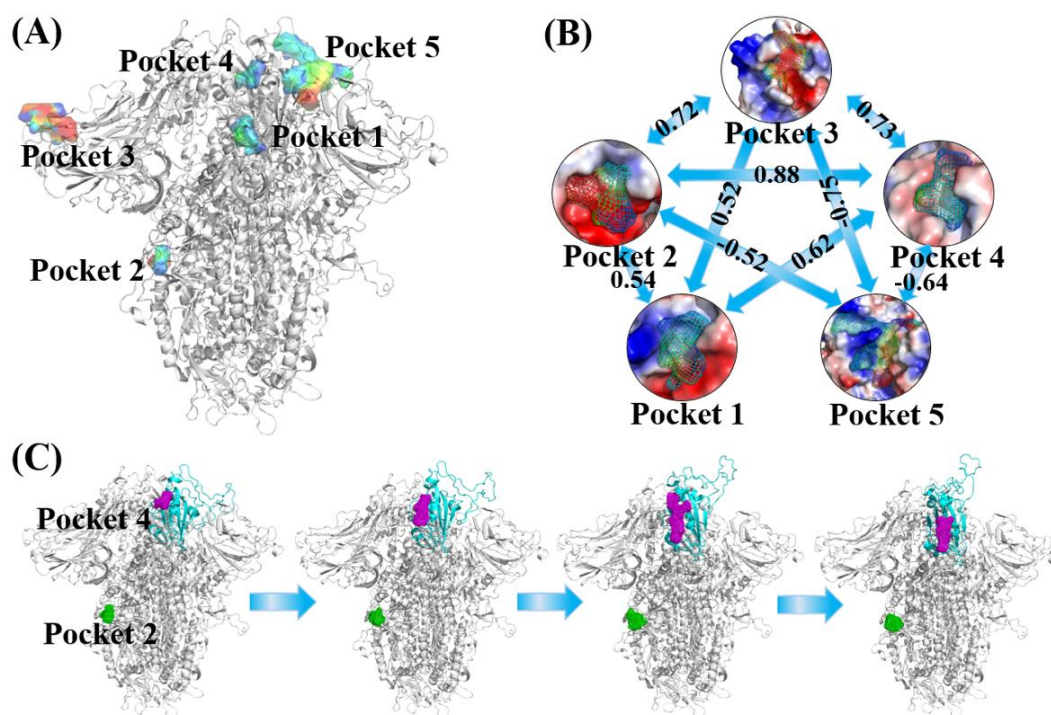
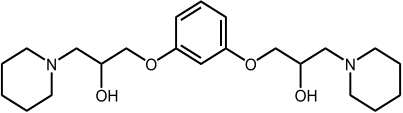
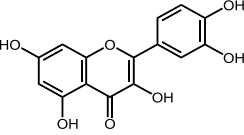
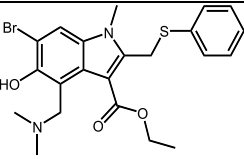
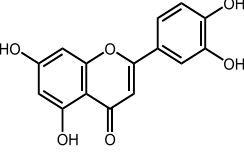
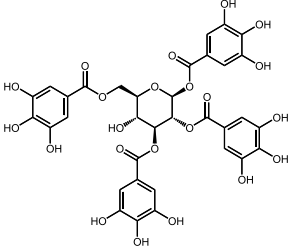
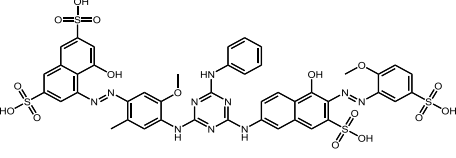
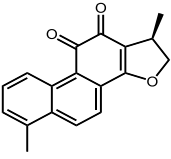
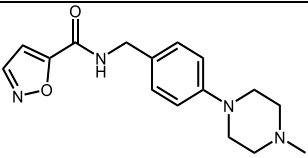


图 4.21 CoV-2-S 的潜在结合口袋预测。5 个潜在配体结合口袋展示 (A), 口袋格点颜色越红, 表明该位点在 CoV-2-S 的构象变化中越稳定存在。5 个潜在位点的运动相关性 (B), 相关性系数标注于口袋连接线上。口袋 2 与口袋 4 的相关运动展示 (C)。

Figure 4.21 The predicted potential ligand binding sites of CoV-2-S. Five potential ligand-binding pockets (A). The redder the pocket grids are, the more stable the subpocket regions throughout the pathway. Pocket correlation between the five predicted binding pockets (B). The numbers are the correlation coefficients between two pockets during conformational changes from “down” to “up” states. The pocket correlation between Pocket 2 and Pocket 4 (C).

表 4.3 靶向刺突蛋白阻止病毒侵入宿主的 8 个化合物。

Table 4.3 Eight compounds targeting spike protein to block the entry of coronavirus into host cells.

Name	Structure	Organism	Activity
VE607 ^[428]		SARS-CoV	EC ₅₀ = 1.6 μM
Quercetin ^[429]		SARS-CoV	IC ₅₀ = 83.4 μM
Arbido ^[430]		SARS-CoV-2	EC ₅₀ = 4.11 μM
Luteolin ^[429]		SARS-CoV	IC ₅₀ = 10.6 μM
TGG ^[429]		SARS-CoV	EC ₅₀ = 4.5 μM
ADS-J1 ^[431]		MERS-CoV	IC ₅₀ = 0.6 μM
Dihydrotanshinone ^[432]		MERS-CoV	/
SSAA09E2 ^[433]		SARS-CoV	EC ₅₀ = 3.1 μM

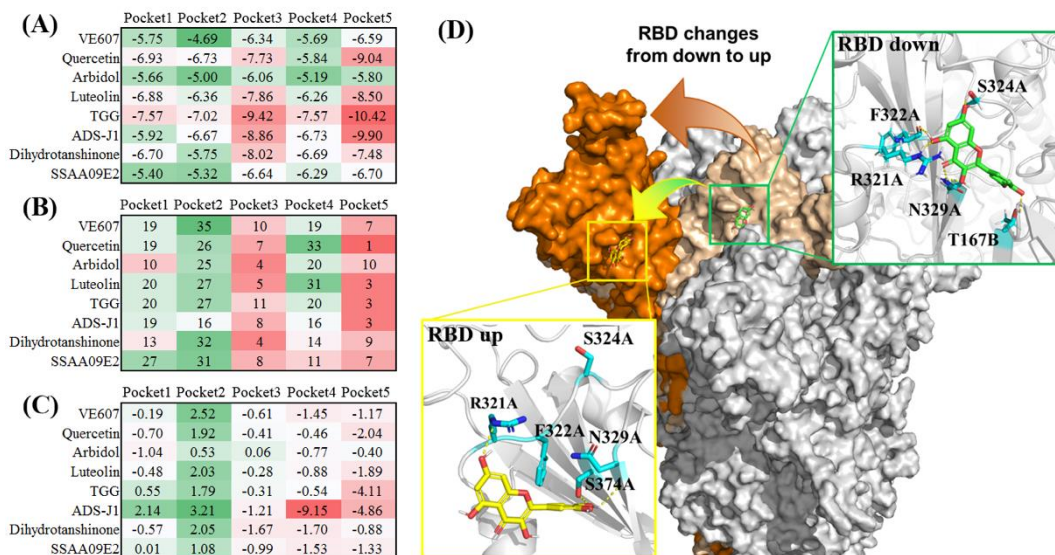


图 4.22 8 个化合物针对 5 个预测口袋的分子对接分析。对接软件为 smina。化合物对 CoV-2-S 的 RBD 闭合构象的 5 个口袋的对接打分 (A)。5 个预测口袋的对接打分在 CoV-2-S 表面 39 个口袋中的排序 (B)。RBD‘向下’和‘向上’构象的分子对接打分差异 (C)。槲皮素在口袋 5 的作用模式 (D)。

Figure 4.22 Molecular docking analysis of the eight compounds in the five suggested pockets by using smina software. The docking scores of the compounds for the five pockets in RBD-down CoV-2-S (A). The ranking of the eight compounds for the five pockets among 39 predicted druggable pockets by D3Pockets (B). The docking score difference between RBD-down and RBD-up CoV-2-S (C). The binding modes of quercetin in pocket 5(D).

4.2.3.7 糖基化对 CoV-2-S 的 ACE2 结合能力影响

研究报道 CoV-2-S 的糖基化对其构象变化至关重要^[434-435]。因此，为进一步研究 CoV-2-S 的糖基化对其 ACE2 结合能力的影响，根据糖基化位点实验数据我们构建了糖基化 CoV-2-S 和 CoV-S (表 S4.9-S4.10)。从糖基化 CoV-2-S 结构看，上述预测的 5 个结合口袋均暴露在糖基化位点之外 (图 4.23A)。当 RBD 角度为 54.5°时，糖基化 CoV-2-S 与 ACE2 的结合自由能为 -19.14 ± 0.72 kcal/mol (表 4.4)，与不加糖基修饰的 CoV-2-S 计算结合自由能相近 (-18.00 ± 0.84 kcal/mol)；当 RBD 角度增加至 84.6°时，糖基化同样对 CoV-2-S 的 ACE2 结合自由能影响甚微。进一步，我们通过自适应拉伸分子动力学模拟 (ASMD)^[422]方法比较研究了糖基化 CoV-2-S 与 CoV-S 的构象转变过程。结果表明，糖基化 CoV-2-S 从 RBD ‘向下’构象转变成 ‘向上’构象的 PMF 为 16.9 kcal/mol (图 4.23B)，而糖基化 CoV-S 只需要 12.0 kcal/mol，表明糖基化 CoV-2-S 更不容易发生构象转变，这与 vsREMD

模拟结果相一致。

表 4.4 糖基化 CoV-2-S 的计算结合自由能。

Table 4.4 The binding free energy (kcal/mol) calculated for glycosylated CoV-2-S*.

Energy term	6ACG (54.5 °±0.86)	6ACK (84.6 °±0.77)
E_{VDW}	-92.22 ±0.86	-100.78 ±0.91
E_{ele}	-915.87 ±2.95	-767.25 ±3.93
E_{gb}	983.72 ±2.90	824.33 ±3.82
E_{np}	-13.07 ±0.08	-14.60 ±0.10
ΔH	-37.46 ±0.70	-58.31 ±0.74
$-T\Delta S$	-18.32 ±0.74	-20.58 ±0.96
ΔG	-19.14 ±0.72	-37.73 ±0.85

*: 统计误差根据 50-100 ns MD 模拟轨迹的计算结果获得。在 MM/GBSA 计算中，从 50-100 ns MD 模拟轨迹中均一提取 1000 个构象计算结合自由能，10 个构象计算熵效应。

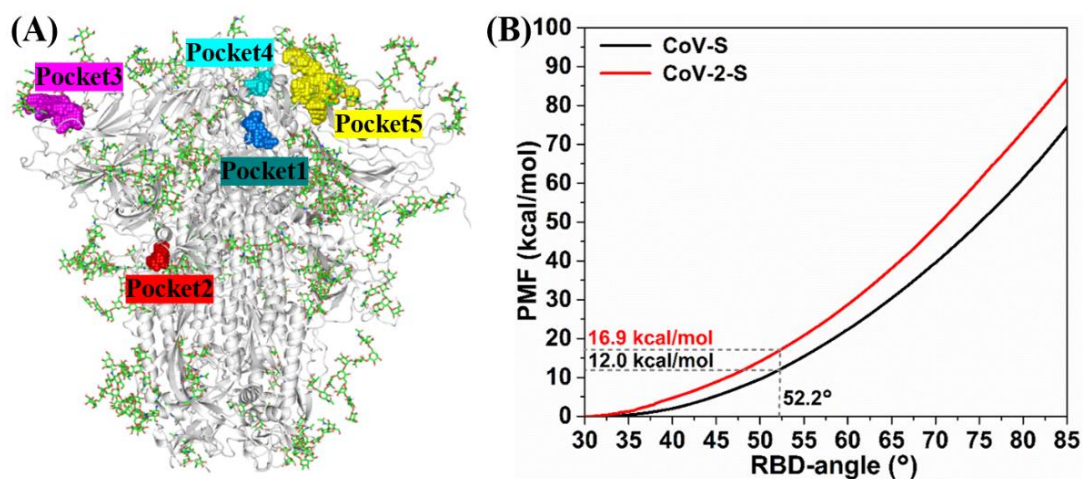


图 4.23 糖基化分析。5 个预测口袋在糖基化 CoV-2-S 上的展示 (A)，糖基分子展示为绿色棍棒模型。糖基化 CoV-2-S (红色) 和 CoV-S (黑色) 的平均力势 (B)，RBD 角度 30.0°到 52.2°变化的平均力势已标注。

Figure 4.23 The glycosylation analysis. The five predicted pockets in the structure of glycosylated CoV-2-S (A). Glycans are shown in green sticks. The potential mean of force (PMF) of the glycosylated CoV-2-S (red) and the glycosylated CoV-S (black) (B). The PMF from RBD-angle of 30.0° to 52.2° was labeled.

4.2.4 小结

新冠病毒刺突糖蛋白 (CoV-2-S) 能够帮助病毒附着并感染人类细胞。实验研究表明 CoV-2-S 的 RBD 与宿主 ACE2 蛋白的亲和力比 SARS-CoV 病毒强。但是在 CoV-2-S 全长结构上的实验却表明它的 ACE2 亲和力与 SARS-CoV 病毒相当或者更弱。因此为探究 CoV-2-S 与 ACE2 蛋白的作用机制, 我们通过多种计算模拟方法, 研究了 CoV-2-S 的动态构象变化, 计算了 CoV-2-S 处于不同构象形态时的 ACE2 结合能力, 揭示了其 RBD 结构域及全长蛋白与 ACE2 亲和力差异的分子机制, 并预测了多个潜在的药物结合位点。结果发现, RBD 结构域只需打开至 52.2° 即可结合 ACE2, 并且随着 RBD 角度的打开, 亲和力增强。但 CoV-2-S 更不容易由‘向下’的闭合构象转变到‘向上’的开放构象, 而且 CoV-2-S 开放构象的比例也比 SARS-CoV 病毒少。另外我们也探索了 CoV-2-S 与 ACE2 结合的关键作用残基的动态变化, 并发现了 CoV-2-S 上 5 个潜在的配体结合位点。通过比较糖基化 CoV-2-S 和非糖基化 CoV-2-S, 我们还发现 CoV-2-S 的糖基化对其 ACE2 结合自由能影响微小。

4.3 本章小结

本章旨在综合我们自主发展的新方法以及各种计算模拟手段应用于具体科学研究问题。为探究大环化合物的构象多样性与其透膜性、极性等性质的关系, 我们通过常规副本交换分子动力学模拟研究了大环药物劳拉替尼在不同极性溶剂 (纯水溶剂、DMSO/H₂O 混合溶剂和氯仿溶剂) 中的构象分布。结果表明, 在高极性的纯水和 DMSO/H₂O 混合溶剂中, 劳拉替尼与溶剂形成密集的氢键网络, 限制其构象变化, 因此只稳定呈现一种构象 **1**。而在低极性氯仿溶剂中, 劳拉替尼除构象 **1** 外 (占比 80%), 还出现一类适合低极性溶剂环境的构象 **2** (占比 20%)。进一步验证表明模拟预测的劳拉替尼在不同极性溶剂下的构象分布比例与 NMR 实验数据高度一致。

为分析新冠病毒刺突蛋白与宿主 ACE2 的作用机制, 本章通过高效采样方法 NUMD 和 vsREMD 探索了新冠病毒刺突蛋白的动态构象变化, 并综合多种计算模拟手段分析了刺突蛋白在处于不同构象形态时的宿主 ACE2 结合能力, 解析了刺突蛋白与宿主 ACE2 受体结合的关键作用残基, 并预测了 5 个新冠病毒刺突蛋白上的可药性结合位点。结果发现随着刺突蛋白‘向上’的角度越大, SARS-CoV-

2 刺突蛋白与宿主 ACE2 的结合能力也变得越强，并且在相同构象状态下强于非典病毒（SARS-CoV）。但是，相比 SARS-CoV，SARS-CoV-2 刺突蛋白呈现 RBD ‘向上’ 构象的比例更少，转变成 RBD ‘向上’ 状态的能垒也更高。这些结果表明尽管 SARS-CoV-2 刺突蛋白的 RBD 结构域具有比 SARS-CoV 更强的宿主 ACE2 结合能力，但 SARS-CoV-2 更不容易从 ‘向下’ 状态转变成 ‘向上’ 状态，因此其全长结构的 ACE2 结合能力与 SARS-CoV 相当甚至更弱。

第 5 章 蛋白质运动数据库的构建及其统计分析

5.1 研究背景

蛋白质的功能与其构象变化息息相关，而蛋白质构象的多样性又与其自身结构相关^[436-437]。因此，为了解蛋白质的功能发挥机制，我们需要掌握其结构的动态特征，这也有助于基于结构的药物设计。例如，在高通量虚拟筛选中，考虑蛋白质柔性的系综对接是一种比常规对接更容易发现活性化合物的策略^[438-439]。目前存在多种研究蛋白质构象多样性的方法，如用于解析生物分子结构的通用实验技术 X 射线晶体衍射和 NMR 光谱学^[440-441]。而在计算模拟中，简正振动模式分析和分子动力学模拟也被用来优化 X 射线晶体衍射和 NMR 光谱学所解析的蛋白结构，并且预测蛋白质的动态构象^[442]。随着越来越多的蛋白质结构被解析出来，蛋白质功能与其构象多样性之间的相关性研究也有了一定的数据基础，也逐渐引起了科研工作者的研究兴趣。

蛋白质的构象多样性是蛋白质在发挥功能时构象变化的必然结果。现有针对蛋白质运动的研究论文主要集中于蛋白质结构域铰链的相对运动和剪切运动^[443-446]。近年来，一系列预测蛋白质柔性结构域的计算方法已经被开发出来^[447-450]，例如差异距离法和变形图分析等，并对蛋白结构域的运动类型也有分类和整理。靶向蛋白质结构域运动的数据库近年来也逐渐浮出水面^[451-454]。另外，根据一些数据库中收集的蛋白质构象变化信息，蛋白质上的柔性结构牵涉甚广，可大可小，包括小的二级结构片段和两个亚基之间的相对运动。但是，在一些情况中，处于不同外界条件下（例如在配体结合之前和之后）的蛋白质运动幅度很小，从整体结构上很难观察出来，但其底物结合口袋周围残基或催化活性残基侧链却会发生显著的局部翻转运动^[455-456]，而且这些运动对蛋白质的底物结合或功能发挥有着至关重要的作用^[457-458]。例如，KAI2 空蛋白与其底物结合蛋白的构象差别仅仅在于底物结合位点附近的 F194 发生了约 90 度的翻转^[459]。此外，残基的侧链局部运动也可能间接影响相邻区域的构象^[460]。因此，为了解蛋白质功能发挥的机制，有必要研究配体结合口袋中残基的局部运动。

PDB 数据库^[461]已经收集了近 167,000 个蛋白结构（截至 2020 年 7 月），是

目前用来研究蛋白质运动特征的理想样本库。但由于蛋白质结构受到不同程度的外界环境影响, PDB 数据库会有冗余的蛋白质结构信息。因此, 为了剔除 PDB 数据库中存在的少量冗余数据, 研究人员基于 PDB 数据库整理出了一系列针对特定研究需求的子数据库。MolMovDB^[453]是目前比较完整的生物大分子构象变化数据库, 以相同蛋白的结构簇呈现, 并且包含蛋白质运动相关分析软件。此外也有一些数据库旨在研究配体诱导的蛋白质构象变化, 以成对的空蛋白和配体结合复合物结构呈现, 例如 ComSin^[462]、AH-DB^[463]、PDBFlex^[464]和 PSCDB^[465]等, 其中 AH-DB 数据库中收集的数据最多 (>700,000)。但众所周知, 蛋白质运动的形式和影响因素错综复杂, 除了受配体的诱导效应, 还包括本身的柔性和外界实验条件的影响 (例如温度和 pH) 等^[466-467]。PCDB^[468]和 CoDNaS^[469]收集了不同实验条件下的结构, 以相同蛋白的结构簇呈现, 但 PCDB 目前已经无法正常浏览。CCProf^[470]除了整理配体诱导的蛋白质构象变化, 还分析了 10 种生物特征用于研究配体结合位点。尽管如此, 依靠现有的数据库还是难以系统研究配体结合口袋中残基的局部运动, 而且也缺乏相关数据库。

因此, 我们在本章中构建了蛋白质运动数据库 D3PM, 其中包含蛋白质整体运动和结合口袋中残基的局部运动, 并且我们将这些运动分成了 4 类整体运动和 5 类口袋残基局部运动。考虑到蛋白结构对的呈现形式相比结构簇形式更便于分析蛋白质运动特征, 在 D3PM 中, 我们提供了所有蛋白质运动的结构对比较。我们希望 D3PM 能够为研究蛋白质运动与其功能发挥的关系提供重要信息, 帮助药物发现与发展。

5.2 结果与讨论

5.2.1 构建蛋白质运动数据库 D3PM

我们从 PDB 数据库中下载所有分辨率优于 3.0Å 的结构 (2018 年 10 月 25 日), 相同的 UniProt ID 视为同一种蛋白。为排除蛋白-蛋白作用的影响, 同一蛋白质按链分开, 多元聚合形式只限于同源多聚体。对于多配体复合物, 每个配体与蛋白质均保存成独立的单配体复合物结构。由于实验结构中许多结晶辅助物 (PEG 等, 表 S5.1) 也都结合在蛋白质上, 为排除这些分子的干扰, 我们人工去除这些含有结晶辅助物的蛋白-配体复合物体系。同一种蛋白质的所有结构成对计算 RMSD。最终, 我们从 PDB 数据库中获取了 7,730,788 个蛋白结构对, 并根

据整体 RMSD 阈值 2 Å 进行分类^[471-472]。同一种蛋白质的相同运动类型存在许多相似的结构对，为去除冗余信息，我们选择一对运动幅度最大且序列完整的结构对作为代表性结构收录于数据库中。

对于整体 RMSD 小于 2 Å 的结构对，我们发现其中不乏在配体结合位点周围具有明显残基局部变化的结构对，这也说明了整体 RMSD 容易隐藏某些局部运动。因此，我们计算了这一类结构对配体结合口袋的残基（配体周围 5 Å 的残基）RMSD，将含有残基对 RMSD 大于 2 Å 的结构对收集起来。为进一步对口袋残基运动进行分类，我们通过 D3Pockets 软件计算分析了配体结合口袋体积的变化。其中一类比较常见的残基运动负责调节结合口袋的开放和闭合，这些残基也称为门控氨基酸^[473]。例如，NF-κB 诱导激酶通过 R410 侧链胍基的翻转运动来打开或关闭腺苷酸结合位点^[474]。但大部分情况下，空蛋白可通过口袋残基的局部运动来扩大结合口袋体积，从而便于底物结合。体积扩大方式有两种：一种是口袋残基向外运动，使得原有结合口袋增大；另一种是口袋残基的运动导致两个小口袋融合成一个大口袋。另外，为了稳定配体结合或者参与配体结合后的催化反应，口袋周围残基也会在配体结合后向口袋内运动，使得口袋体积减小。例如，相比空蛋白结构，FOX-4 cephamycinase 的配体结合复合物中 F293 残基向结合口袋内移了 2.5Å，进而与配体形成 π - π 堆积相互作用^[475]。此外，也有一些结合口袋周围残基的局部运动对口袋体积无明显影响，也没有参与和配体的相互作用。综上所述，如图 5.1 所示，我们将所有结构对运动分成 5 类：PC 型，口袋开拓型运动，即残基的运动导致口袋的形成；PE 型，口袋扩展型运动，即残基的运动导致口袋体积增加；PF 型，口袋融合型运动，即残基的运动导致多个口袋合并为一个口袋；PS 型，口袋闭合型运动，即残基的运动导致口袋体积减小；以及其他类型运动 OM 型。

通过这种分类模式，我们对空蛋白-结合蛋白结构对、结合不同配体的结合蛋白-结合蛋白结构对和结合相同配体的结合蛋白-结合蛋白结构对均进行了计算和分类。对于空蛋白-结合蛋白结构对体系，最终获得 609,433 对结构，包含 568 种蛋白运动形式，每种蛋白我们挑选一对典型结构。对于结合不同配体的结合蛋白-结合蛋白结构对体系，共有 2,176,460 对结构的残基最大 RMSD 超过 2.0 Å，包含 967 种蛋白运动形式，其中 620 对结合在相同位点，另外 347 对结合在不同位点。这 347 对结构可视为空蛋白-结合蛋白对体系。因此，也按照图 5.1 的

分类模式进行分类。对于结合相同配体的结合蛋白-结合蛋白结构对体系，共有 1,297,976 对结构的残基最大 RMSD 超过 2.0 Å，包含 750 种运动模式。

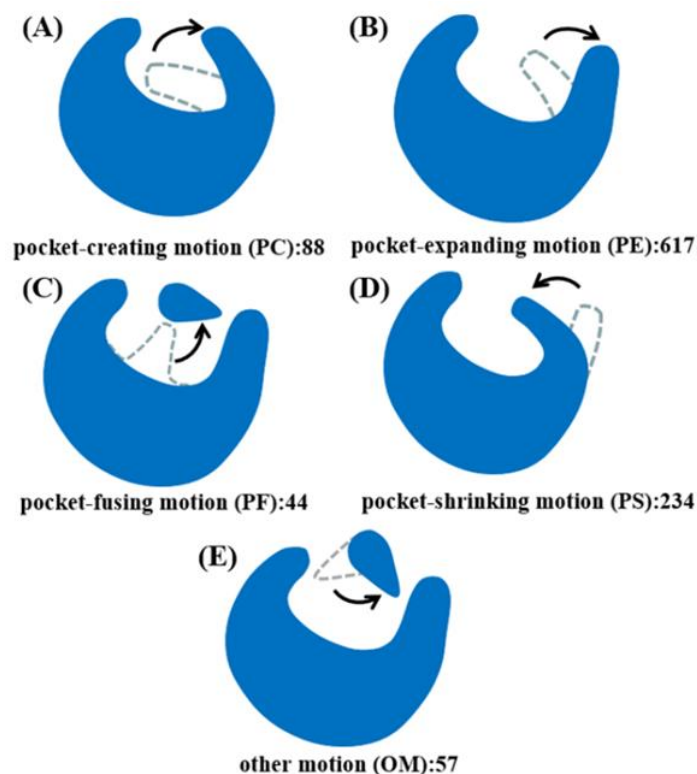


图 5.1 口袋残基运动类型。口袋开拓型运动 (A)，口袋扩展型运动 (B)，口袋融合型运动 (C)，口袋闭合型运动 (D)，其他类型运动 (E)。

Figure 5.1 Five classes of pocket residue motions. (A) pocket-creating motion (88), (B) pocket-expanding motion, (C) pocket-fusing motion, (D) pocket-shrinking motion and (E) other motion.

对于整体结构 RMSD 大于 2 Å 的结构对，我们分成四类：第一类：空蛋白-空蛋白结构对 (apo & apo)，由于已经排除了蛋白-配体和蛋白-蛋白相互作用的影响，这类结构对的 RMSD 表征空蛋白的柔性特征；第二类：空蛋白-结合蛋白结构对 (apo & holo)，表征配体的诱导运动；第三类：结合不同配体的结合蛋白-结合蛋白结构对(holo & holo (different ligands))，表征不同配体的诱导运动；第四类：结合相同配体的结合蛋白-结合蛋白结构对(holo & holo (the same ligand))，表征结合蛋白本身的柔性运动。对于每一种运动类型，挑选 RMSD 值前五的结构对中序列最完整的作为代表性结构对。最终，我们收集了 5,339 个代表性结构对。878 个结合相同配体的复合物结构对具有超过 2.0Å 的 RMSD 值，主要是源自蛋白-配体复合物的本身柔性。另外对于有明显结合口袋残基运动的体系，我

们对其进行分类并获得 125 个口袋残基运动类型的代表性结构对。

表 5.1 D3PM 数据库总结。

Table 5.1 Summary of the data available in the D3PM.*

The overall protein motions	Number of pairs	Overall RMSD (Å)	
		Mean	Maximum
apo & apo	2173	4.32	41.76
apo & holo	1282	4.34	39.53
holo & holo (different ligands)	1006	4.65	32.02
holo & holo (the same ligand)	878	4.39	26.50
The pocket residues motions		Pocket RMSD (Å)	
		Mean	Maximum
apo & holo	949	1.45	6.42
holo & holo (different ligands)	620	2.60	9.68
holo & holo (the same ligand)	750	1.33	2.50

*: “apo”代指空蛋白，“holo”代指配体结合蛋白。

为了给蛋白质的运动特征与药物发现关系研究更方便的信息，我们对每对蛋白质运动进行了 DrugBank 数据库注释和链接。DrugBank 数据库是目前整合了上市药物和相应靶标等信息的综合性免费数据库^[476]。例如，在 D3PM 中有 3 种运动类型含有药物靶标碳酸酐酶 2 (PDB ID: 3HS4)，包括由本身柔性和配体诱导的蛋白整体构象变化，以及 PE 类型口袋残基运动。

根据上述分类规则，我们共收集了 2319 对结合口袋残基局部运动，5339 对蛋白质整体运动，之后我们基于这些数据构建了一个蛋白质运动网站 www.d3pharma.com/D3PM/index.php，该网站可免费下载所有收集到的蛋白质运动数据（图 5.2）。D3PM 的界面设计兼顾了对整个数据库的浏览和蛋白质运动类型的具体搜索。用户可以通过蛋白名称、PDB ID、Uniprot ID、RMSD、残基名称或者配体等进行搜索。

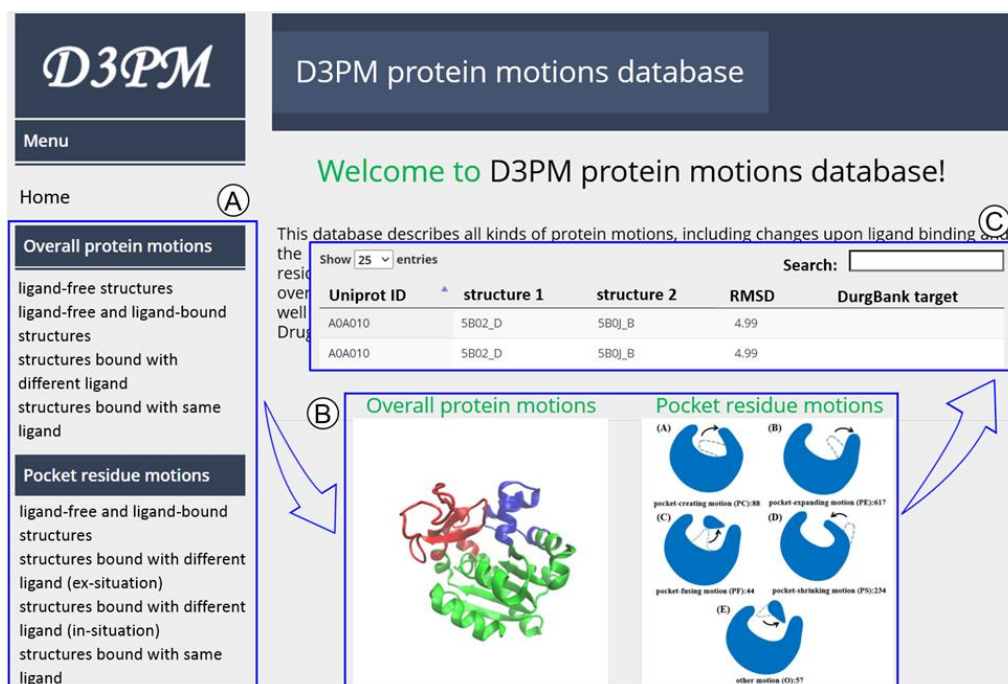


图 5.2 D3PM 网站示意图。D3PM 中所有蛋白质运动类型总览表 (A)，两种主要的运动类型示意图，即蛋白整体运动和结合口袋残基局部运动 (B)。蛋白运动结构对的具体信息 (C)。

Figure 5.2 The web page of the D3PM database. (A) the overview of types of protein motions included in the D3PM, (B) diagrams for the two main types of protein motions, viz, overall protein motions and pocket residue motions, (C) the information of each protein motion pair.

5.2.2 蛋白质运动影响因素统计分析

基于 D3PM 数据库，我们探索了蛋白质本质柔性、配体诱导效应等因素对蛋白质运动的具体影响。在 D3PM 中，7,730,788 个蛋白结构对被分成四类，即空蛋白-空蛋白结构对、空蛋白-结合蛋白结构对、结合不同配体的结合蛋白-结合蛋白结构对和结合相同配体的结合蛋白-结合蛋白结构对。为比较这四种运动类型，我们选择四种运动类型的蛋白种类交集进行分析，共计 3,990,497 个结构对，含 1,970 种蛋白。整体 RMSD 低于 $2.0 \text{ \AA}$ 的结构对被视为相似结构，因此 RMSD 低于 $2.0 \text{ \AA}$ 的结构对占比越小，则表明该类运动总体更加明显。

如图 5.3A 所示，我们发现结合相同配体的结合蛋白-结合蛋白结构对的整体 RMSD 的平均值最小，其 RMSD 低于 $2.0 \text{ \AA}$ 的结构对占比为 94.7%，比空蛋白-空蛋白结构对的比例 (93.2%) 还高，说明蛋白质结合配体后确实会倾向于减小结构柔性。值得一提的是，仍有超过 5% 的结合相同配体的结合蛋白-结合蛋白结

构对的 RMSD 超过 2.0 Å，其中大部分是由于配体结合无关的 loop 结构的柔性导致的。结合不同配体的结合蛋白-结合蛋白结构对中 RMSD 低于 2.0 Å 的结构对占比为 89.4%，比结合相同配体的结构对低 5.3%左右，表明配体诱导的蛋白构象变化与配体结构是有重要关系的。空蛋白-结合蛋白结构对的 RMSD 值最大，其 RMSD 低于 2.0 Å 的结构对占比为 85.5%，表明配体的诱导效应是不可忽略的。但尽管如此，我们注意到在这四种运动类型中，至少 85%的结构对的整体 RMSD 值小于 2.0 Å。

为进一步探索配体结合口袋残基的局部运动，将配体周围 5.0 Å 内的残基视为口袋残基，我们统计了这四类具有整体运动的结构对中口袋残基的运动情况（图 5.3B）。结果表明，结合相同配体的复合物结构对的 RMSD 低于 2.0 Å 的比例为 90.2%，表明配体结合后能稳定口袋残基的位置，使其柔性降低。但相比空蛋白结构，配体的诱导作用对口袋残基运动的影响非常明显，其 RMSD 低于 2.0 Å 的比例仅为 46.6%，而且比整体结构 RMSD 低于 2.0 Å 的比例低 22.8%，说明配体结合诱导的构象变化主要发生在结合口袋周围。超过一半的蛋白质体系在配体结合后口袋残基会发生明显的运动，说明我们在基于结构的药物设计中需要格外注意口袋残基的柔性。另外，统计结果也表明，配体结合会导致蛋白质发生明显的构象变化，尤其是在结合位点附近，但配体结合后也会使得蛋白质整体柔性降低。

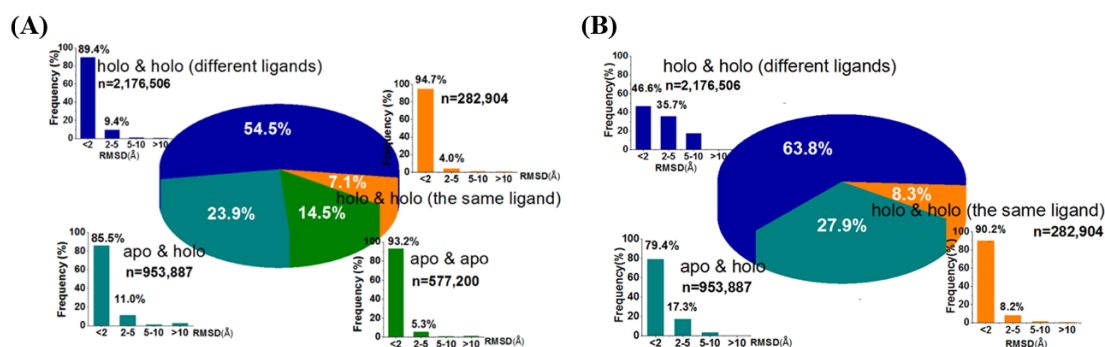


图 5.3 蛋白运动类型分析。四种蛋白质整体运动类型（A）和其中 3 类的口袋残基运动（B）比例分析。“apo”代指空蛋白，“holo”代指配体结合蛋白。

Figure 5.3 Protein motion analysis. The frequency of four types of protein overall motions (A) and three types of pocket residue motions (B). The “apo” referred to as ligand-free protein, and the “holo” referred to as ligand-bound protein.

5.2.3 结合位点残基偏好性研究

为研究氨基酸残基在结合位点上的偏好性，我们计算了 20 种常见氨基酸在结合口袋周围和蛋白质整体结构的分布比例。通常我们以距离配体 5.0 Å 为截断值定义口袋周围残基^[477]。本章中，以 178,778 个蛋白-配体复合物结构为统计样本，我们计算了 20 种氨基酸在不同截断值（2.0 到 6.0 Å）下的分布比例，并与它们各自在整体结构上的分布比例比较。如图 5.4 所示，我们计算了 20 种氨基酸在口袋周围和整体结构上的分布比例的平均绝对偏差。当截断值为 3.0 Å 时，口袋周围和蛋白整体结构的残基分布差异最大，这种差异随截断值的增加而减小，表明当截断值为 3.0 Å 时，以氨基酸的偏好性最能区分结合口袋和蛋白整体结构。因此，后续研究中我们分析了截断值为 3.0 和 5.0 Å 时的结合口袋氨基酸偏好性。如图 5.5A 所示，当截断值为 3 Å 时，残基 Arg、Asp、Ser、Glu、Thr、Lys、Tyr、Asn、His 和 Cys 在结合口袋周围的分布比例显著高于整体结构上的分布比例，表明这些残基更倾向于与配体形成短程相互作用，例如氢键作用和离子键作用等。当截断值为 5 Å 时，残基 Gly、Phe、Met 和 Trp 在结合口袋周围的分布比例显著高于整体结构上的分布比例，表明这些残基更倾向于与配体形成长程相互作用。上述 14 种残基会富集在结合口袋周围，与配体易形成短程或长程非键相互作用，我们称之为“亲口袋”残基（pocketphilic）。另外六种氨基酸残基，即 Leu、Ala、Val、Ile、Pro 和 Gln，在截断值为 3.0 或 5.0 Å 时的口袋周围分布比例都低于它们在整体结构上的分布比例，我们称这些残基为“疏口袋”残基（pocketphobic）。

为进一步研究 20 种氨基酸残基对运动类型的偏好性，我们统计了 D3PM 中口袋残基局部运动数据集中 RMSD 超过 2.0 Å 的残基的分布比例。如图 5.5B-C 所示，大部分残基的运动类型是拓展型运动（PE），占比超过 56%。而且亲口袋残基（Arg、Phe、Tyr、Lys、Glu 和 Asp）比疏口袋残基更容易运动。其中，精氨酸具有较长的侧链，容易发生侧链的翻转等运动。但侧链越长并不一定越容易运动，例如酪氨酸的侧链比苯丙氨酸长，但酪氨酸比苯丙氨酸具有超过 2.0 Å RMSD 的比例更少。而且，我们发现碱性氨基酸（精氨酸和赖氨酸）比酸性氨基酸（天冬氨酸和谷氨酸）更容易运动。

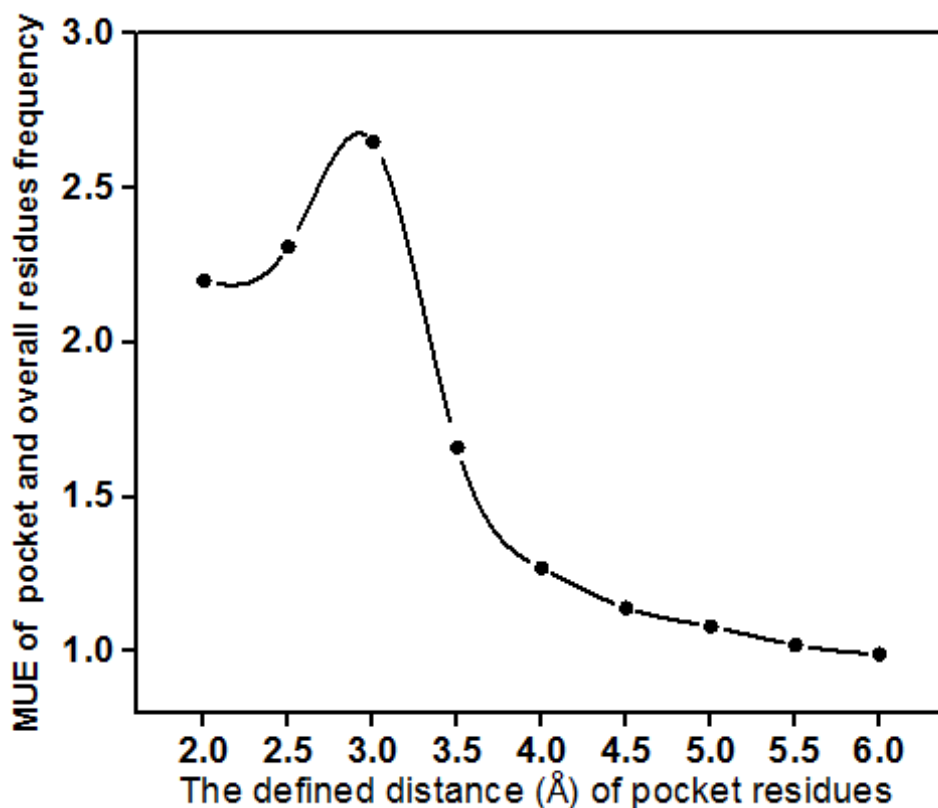


图 5.4 20 种氨基酸在结合口袋和整体结构上的分布比例平均绝对误差。口袋残基定义为距离配体 2.0 Å 到 6.0 Å 以内。

Figure 5.4 The mean unsigned error (MUE) of the frequencies of the 20 amino acids between the pocket and overall structure. The pocket residues were defined that around ligand by different distances, ranging from 2.0 Å to 6.0 Å.

5.2.4 蛋白结构柔性及分子对接准确性研究

通常高通量虚拟筛选以受体结构为基础，当空蛋白结构不可获取时，我们会选择结合配体的受体结构。但这种情况下，由于受体结构已经受到原有配体的诱导效应，可能会错失一些阳性化合物。因此，为了定量评估这种原有配体的诱导效应对分子对接的具体影响，我们以 D3PM 中 620 对结合不同配体的复合物结构对作为测试数据，比较它们的交叉对接打分差异，即比较配体对接在原受体结构和其他受体结构的打分。如表 S5.2 所示，共获得 1240 个对接打分数据，结果表明配体分子对接在原受体结构的打分平均值为-9.24，而对接到其他受体结构的打分平均值为-8.67。而且，23%的配体在其他受体结构上的对接打分比原有受体结构的打分少 1 kcal/mol 以上，这对于高通量筛选而言是一个不可忽视的比例。因此，我们在高通量虚拟筛选时需要关注受体结构的柔性。

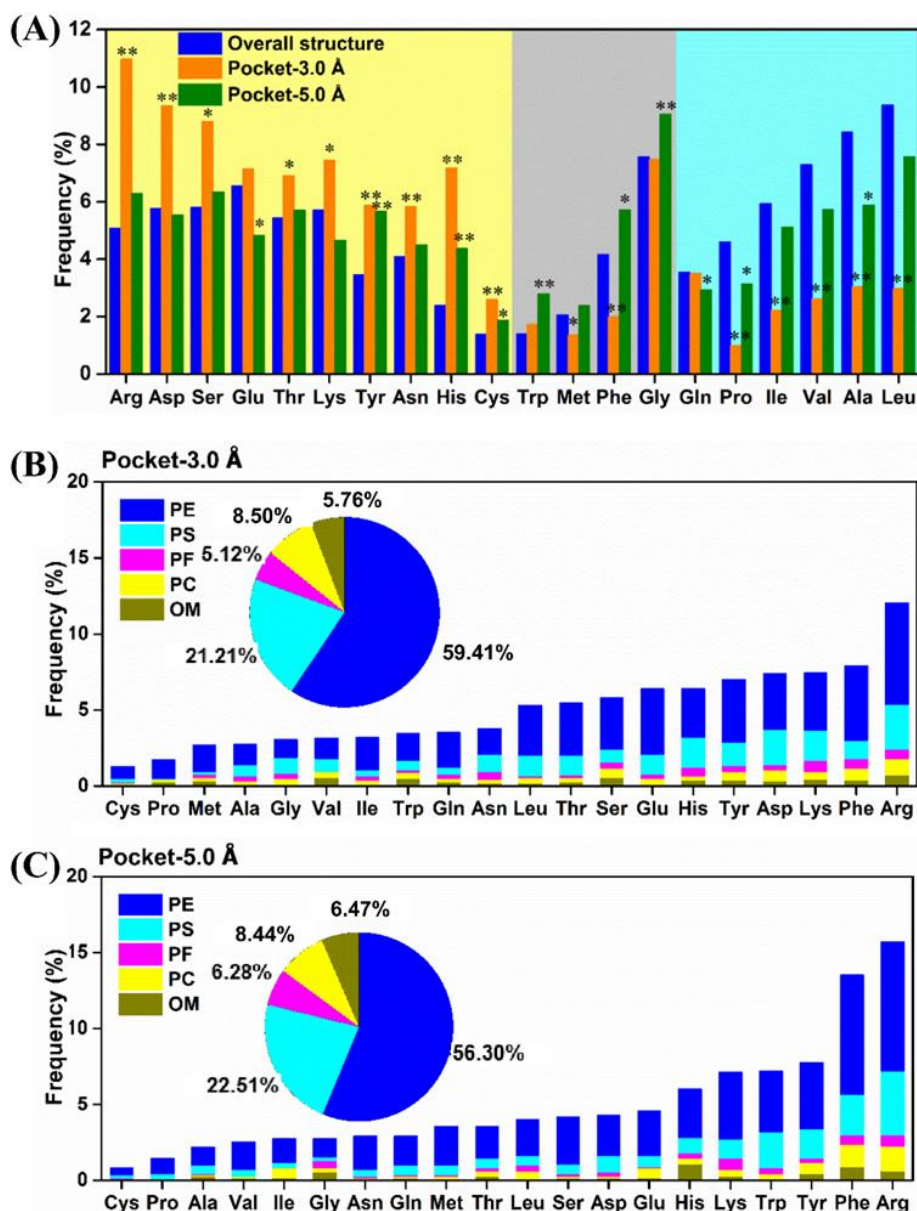


图 5.5 20 种氨基酸在结合口袋和整体结构上的分布比例。以 3.0 Å（橙色）和 5.0 Å（绿色）截断值。根据残基分布比例最大值，将残基分别排列在黄色区、灰色区和蓝色区，对应于最大值为截断值为 3.0 Å、5.0 Å 和整体结构。*表示结合口袋与整体结构的残基分布比例具有 5% 的显著差异 ($p < 0.05$)。**表示结合口袋与整体结构的残基分布比例具有 1% 的极显著差异 ($p < 0.01$)。以 3.0 Å (B) 和 5.0 Å (C) 截断值时，计算的容易发生运动的氨基酸分布比例。

Figure 5.5 (A) Frequencies of 20 amino acid residues in overall protein structure (blue) or binding sites around ligand by 3.0 Å (orange) and 5.0 Å (green). The residues are grouped in yellow, gray and cyan blocks, according to largest frequency belongs to pocket-3.0 Å, pocket-5.0 Å and overall structure, respectively. (*) the difference between the overall structure and binding site around ligand by 3.0 or 5.0 Å is statistically significant at the 5% level ($p < 0.05$).

(**) the difference between the overall structure and binding site around ligand by 3.0 or 5.0 Å is statistically very significant at the 1% level ($p < 0.01$). Frequencies of 20 amino acid residues easy to move in pocket, defined with the cutoff of 3.0 (B) or 5.0 Å (C)

5.3 本章小结

在本章中，我们构建了一个蛋白质运动数据库 D3PM，其中包含四种蛋白质整体运动和五种口袋残基局部运动。D3PM 主要以结构对的列表形式呈现。我们目前在发展一种基于结构对的蛋白质构象转变路径预测方法^[281]。因此，我们计划 D3PM 将来的发展包括 1) 提供蛋白质运动路径的动态图；2) 定期更新 D3PM 数据库。

基于 D3PM 数据库，我们比较了不同因素影响下的蛋白运动类型。结果表明配体诱导的蛋白质运动最为显著，其中 53.4% 的结构对在配体诱导作用下其结合口袋周围残基 RMSD 超过 2.0 Å。但配体结合后会与残基形成相互作用，从而稳定结合口袋残基结构，但由于蛋白质柔性结构的运动，仍有 5% 的结构对的整体 RMSD 超过 2.0 Å。在本章中，我们主要研究了蛋白质或复合物结构的本质柔性以及配体结合对蛋白质运动的影响，但其他因素，例如 pH、温度和残基突变等也会影响蛋白质的运动，是之后需要进一步研究的方向。此外，我们统计了 20 种氨基酸残基在结合口袋上的偏好性。结果表明，一些亲口袋残基容易富集在结合口袋，与配体形成短程相互作用。这些残基偏好性统计结果有帮助帮助结合口袋的预测等。目前该数据库可免费访问，网址为：
www.d3pharma.com/D3PM/index.php

第6章 总结与展望

蛋白质和配体间相互作用的研究在基于结构的药物设计中扮演着重要的角色,许多新型药物的发现也都来源于此。其中,蛋白质-配体的结合自由能是表征它们相互作用的重要物理量,是评价其结合能力及配体生物活性的重要参数。因此,快速获得准确的结合自由能对药物设计至关重要。然而,目前结合自由能的准确计算受到多因素的制约,其中一个重要原因是预测结合自由能算法的不准确性,如不能精确描述溶剂化效应和静电极化效应等。近年来,新的结合自由能计算方法屡被报道,但大多数难以真正应用到高通量药物筛选中。目前快速的预测方法主要是基于经验方程,如广泛使用的定量构效关系和分子对接中的打分函数,但这类方法的准确性受限于拟合参数时训练集的选择,其普适性和准确性并不能令人满意,而较为准确的自由能微扰和热力学积分等方法则需要消耗大量的计算资源。因此发展一种高效准确的蛋白质-配体结合自由能预测新方法对提高药物设计效率至关重要。另一方面,靶标蛋白在功能发挥和配体结合时往往会发生大规模构象变化。因此,研究蛋白质的构象变化对理解靶标蛋白结构和功能之间的关系及药物分子设计都具有重要意义。然而,常规的实验技术或理论模拟在靶标蛋白大规模构象变化中的广泛应用却面临巨大挑战,如常规理论计算的时间尺度远小于蛋白质运动的实际时间尺度。为此,发展高效准确的靶标蛋白大规模构象变化采样方法也一直是计算生物学及药物设计中的研究热点。

本论文工作紧密围绕高效准确模拟靶标蛋白大规模构象变化及其与配体相互作用这一重要科学问题开展工作。在第二章中,为了在模拟蛋白质与配体的相互作用时精确考虑蛋白受到的静电极化效应,我们基于“分而治之”的思想,将蛋白质-配体复合物体系以残基为单位划分成 SQM/MM 的小体系,发展了 SQMPC 方法来计算蛋白-配体复合物体系的半经验量子化学原子电荷。该方法的创新点在于实现了生物大分子体系上的量子化学原子电荷计算,能精确考虑蛋白质受到溶剂和配体的静电极化效应,相比课题组自主发展的 QMPC 方法在计算速率上有明显的提升。我们将 SQMPC 方法与结合自由能计算方法 SIE 与 MM/GBSA 相结合,应用于配体亲和力预测在 50 个蛋白-配体测试体系中发现, SQMPC 均取得了比基于常规力场 AMBER-ff03 电荷更准确的自由能计算结果,其中 AM1 和

PM7 方法在测试的四种 SQM 方法中表现最佳。而且对于一系列具有不同骨架的配体 (FABP4 体系), SQMPC 也能提高 SIE 和 MM/GBSA 方法的计算准确性。另外, 我们发现 MD 模拟时长是会影响 SIE 和 MM/GBSA 方法的准确性, 但并不是随着 MD 模拟时间的延长而准确性提高。在某些体系中, 短时间的 MD 模拟能达到更好的预测效果。在 50 个蛋白-配体测试体系中, SIE 方法预测的结合自由能与实验值之间的相关性比 MM/GBSA 方法更高, 而且考虑到 SIE 方法计算速度相对更快, 而 MM/GBSA 方法还需要额外计算熵效应, 因此, SQMPC 与 SIE 方法的组合 (SIE-SQMPC) 将有助于蛋白-配体结合自由能的快速准确预测。

此外, 基于分子动力学模拟预测蛋白-配体结合自由能的方法, 如自适应拉伸分子动力学 (ASMD), 往往也忽略了配体解离过程中蛋白质受到配体及溶剂的静电极化作用。因此, 为提高 ASMD 方法的准确性, 本论文将发展的 SQMPC 方法与 ASMD 结合, 定期更新 ASMD 模拟中蛋白-配体体系的原子电荷, 进一步发展了 ASMD-SQMPC。将该方法应用于 24 个 HIV-1 蛋白酶抑制剂中发现, ASMD-SQMPC 方法 ($R^2=0.86$) 能显著提高常规 ASMD ($R^2=0.52$) 的结合自由能预测准确性。并且, ASMD-SQMPC 方法对含氟配体也是有效的 ($R^2=0.71$)。因此, 本论文第二章提出了 2 种能准确有效地预测蛋白-配体结合自由能的新策略 (SIE-SQMPC 和 ASMD-SQMPC), 可为高通量处理分子对接和药物设计提供帮助。

靶标蛋白在发挥功能或配体结合时往往会发生大规模的构象变化, 在有限的计算机资源下, SQMPC 方法中采用的常规 MD 模拟由于受到其低采样效率的限制, 难以对这类构象变化进行充分采样, 从而难以研究蛋白质构象间转变的自由能变化。因此, 为进一步研究蛋白质大规模的构象变化及其与配体的相互作用, 本论文第三章研究工作主要致力于发展高效生物大分子构象采样新方法。基于优化的副本交换分子动力学 (REMD) 和简正振动模式分析 (NMA) 作为集合变量, 我们发展了一种基于能量和几何二维空间同时加速的增强型采样方法, 称之为 ossPTMetaD。该方法的创新点在于通过 REMD 和 WT-MetaD 的“强强联合”方式既减少了常规 REMD 的计算资源, 也利用“填坑”式增强采样提高了 REMD 的采样效率, 同时采用 NMA 作为集合变量不需要先验知识, 并且我们提出的这种组合方式具有良好的方法可移植性。为测试 ossPTMetaD 方法, 我们将其应用到 5 种不同的蛋白运动类型中, 包括 AdK 体系的结构域相对运动、HIV-1 蛋白酶和 BACE1 体系的开放闭合构象转变以及激酶的 DFG 模式翻转运动。结果表明

ossPTMetaD 方法能够成功模拟这些蛋白质运动, 并且 ossPTMetaD 相比传统 REMD 方法的采样收敛速度更快, 所需计算资源也减少 70.0%。

本论文另一部分工作集中于综合各种计算模拟手段以及我们自主发展的新方法应用于具体研究问题。首先对于小分子体系, 在第四章中, 为探究大环化合物的构象多样性与其透膜性、极性等性质的关系, 我们通过常规副本交换分子动力学模拟研究了大环药物劳拉替尼在不同极性溶剂(纯水溶剂、DMSO/H₂O 混合溶剂和氯仿溶剂)中的构象分布。结果表明, 在高极性的纯水和 DMSO/H₂O 混合溶剂中, 劳拉替尼与溶剂形成密集的氢键网络, 限制其构象变化, 因此只稳定呈现一种构象 1, 这种构象与劳拉替尼结合 ALK/ROS1 蛋白的晶体构象高度相似。而在低极性氯仿溶剂中, 劳拉替尼除构象 1 外(占比 80%), 还出现一类采样较少的构象 2(占比 20%)。与构象 1 相比, 构象 2 的溶剂可及表面积和极性都更小。我们进一步通过 NMR 实验验证了模拟的准确性, 表明模拟预测的劳拉替尼在不同极性溶剂下的构象分布比例与 NMR 实验数据高度一致。因此, 我们的模拟结果和 NMR 实验数据都表明劳拉替尼可能通过构象 1 与构象 2 间的转变来实现它良好的透膜性。

另外, 新型冠状病毒 SARS-CoV-2 自 2019 年末爆发后迅速蔓延全球, 目前尚未发现可以治疗 SARS-CoV-2 引发的新冠肺炎 COVID-19 的特效药。SARS-CoV-2 的刺突蛋白以三聚体形式嵌在病毒表面, 介导病毒与宿主细胞 ACE2 的结合, 从而帮助病毒入侵和感染宿主细胞, 是目前许多疫苗和抗体药物的重要靶点之一。因此, 本论文通过我们自主发展的高效采样方法 NUMD 和 vsREMD 探索了 SARS-CoV-2 刺突蛋白的动态构象变化, 并综合多种计算模拟手段(同源建模、结合自由能计算、自由能分解和可药性位点预测等)分析了刺突蛋白在处于不同构象形态时的宿主 ACE2 结合能力, 解析了刺突蛋白与宿主 ACE2 受体结合的关键作用残基, 并预测了 5 个 SARS-CoV-2 刺突蛋白上的可药性结合位点。结果发现随着刺突蛋白‘向上’的角度越大, 刺突蛋白与宿主 ACE2 的结合能力也变得越来越强。另外, 相比 SARS-CoV 病毒, 新冠病毒刺突蛋白呈现‘向上’构象的比例更少, 转变成‘向上’状态的能垒也更高, 这一结果说明尽管新冠病毒刺突蛋白的 RBD 结构域具有比 SARS-CoV 病毒更强的宿主 ACE2 结合能力, 但新冠病毒更不容易从‘向下’转变成‘向上’状态, 因此其全长结构的 ACE2 结合能力与 SARS-CoV 病毒相当或者更弱, 与实验结果相吻合。

在第五章中,为进一步帮助理解蛋白运动的规律和影响因素,我们基于 PDB 数据库构建了一个蛋白运动数据库 D3PM。我们发现除了蛋白质整体结构的大规模运动,口袋残基的局部运动也同样是蛋白质功能机制研究不可忽略的。但这种局部运动由于涉及氨基酸残基少,且运动幅度较小,往往在整体 RMSD 值上表征不出来。因此,我们针对这类口袋残基的局部运动进行了统计并整理分成了 5 类,即口袋开拓型运动、口袋扩展型运动、口袋融合型运动、口袋闭合型运动以及其他类型运动。最终,D3PM 数据库中包含了 2319 对结合口袋残基局部运动,5339 对蛋白质整体运动。基于 D3PM,我们比较了不同因素影响下的蛋白运动类型。结果表明配体诱导的蛋白质运动最为显著,但配体结合后会与残基形成相互作用,从而稳定结合口袋残基结构。此外,我们统计了 20 种氨基酸残基在结合口袋上的偏好性。结果表明,一些亲口袋残基容易富集在结合口袋,与配体形成短程相互作用。这些残基偏好性统计结果有望帮助结合口袋的预测等。

综上所述,本论文紧密围绕高效准确模拟靶标蛋白大规模构象变化及与配体相互作用开展工作。第一部分旨在发展蛋白质-配体结合自由能计算新方法(SIE-SQMPC 和 ASMD-SQMPC,第二章),为高通量药物筛选提供帮助。本论文提出的这两种新方法能获得相对常规方法有所改进的准确性,主要原因在于精确的原子电荷能够提高分子动力学模拟和结合自由能计算的准确性。对于抑制剂结合机制的探索问题,可进一步将 SQMPC 与 FEP 或者 TI 等高精度方法组合。另一方面,体系原子电荷在蛋白质运动过程中是不断重新分配的,因此采用定期更新体系原子电荷方式的 ASMD-SQMPC 方法能够获得比 SIE-SQMPC 更高的相关性。理论上,更频繁的电荷更新能获得更准确的计算结果,但伴随而来的是更大的计算成本。因此,在 SQMPC 之后的继续优化中需要在更新电荷时平衡计算时间和准确性。本论文第三章发展了一种蛋白质增强型采样新方法(ossPTMetaD),相比传统 REMD 方法显著减少了计算资源。ossPTMetaD 方法在副本交换时忽略溶剂内部作用能的方式理论上不服从细致平衡条件,尽管在 5 种经典靶标测试体系中获得了可靠的结果,但仍需在后续的应用研究中进一步验证。本论文第二部分综合各种计算模拟手段以及我们自主发展的新方法应用于具体研究问题。研究了大环化合物的构象多样性与其透膜性、极性等性质的关系,探索了新冠病毒刺突蛋白与宿主 ACE2 的作用机制。并且,为进一步帮助理解蛋白运动的规律和影响因素,基于 PDB 数据库构建了一个蛋白运动数据库 D3PM。

参考文献

- [1] Itzstein M. The war against influenza: discovery and development of sialidase inhibitors[J]. *Nat Rev Drug Discov*: 2007, 6: 967-974.
- [2] Varghese J N, Smith P W, Sollis S L, et al. Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase[J]. *Struct Fold Des*: 1998, 6: 735-746.
- [3] Valiente P A, Gil L A, Batista P R, et al. New parameterization approaches of the LIE method to improve free energy calculations of PlmII-Inhibitors complexes[J]. *J Comput Chem*: 2010, 31: 2723-2734.
- [4] Steinbrecher T, Cappel D, Lin T, et al. Accurate Binding Free Energy Predictions in Fragment Optimization[J]. *J Chem Inf Model*: 2015, 55: 2411-2420.
- [5] Yeung W, Ruan Z, Kannan N. Emerging roles of the alphaC-beta4 loop in protein kinase structure, function, evolution, and disease[J]. *IUBMB Life*: 2020, 72: 1189-1202.
- [6] Yang J, Talibov V O, Peintner S, et al. Macrocyclic Peptides Uncover a Novel Binding Mode for Reversible Inhibitors of LSD1[J]. *ACS Omega*: 2020, 5: 3979-3995.
- [7] Chen J, Yang J, Wei Q, et al. Identification of a selective inhibitor of IDH2/R140Q enzyme that induces cellular differentiation in leukemia cells[J]. *Cell Commun Signal*: 2020, 18: 55-65.
- [8] Zou Y, Ewalt J, Ng H L. Recent Insights from Molecular Dynamics Simulations for G Protein-Coupled Receptor Drug Discovery[J]. *Int J Mol Sci*: 2019, 20-27.
- [9] Calebiro D, Grimes J. G Protein-Coupled Receptor Pharmacology at the Single-Molecule Level[J]. *Annu Rev Pharmacol Toxicol*: 2020, 60: 73-87.
- [10] Ali A M, Atmaj J, Van Oosterwijk N, et al. Stapled Peptides Inhibitors: A New Window for Target Drug Discovery[J]. *Comput Struct Biotechnol J*: 2019, 17: 263-281.
- [11] Moldovean S N, Chis V. Molecular Dynamics Simulations Applied to Structural and Dynamical Transitions of the Huntingtin Protein: A Review[J]. *ACS Chem Neurosci*: 2020, 11: 105-120.
- [12] Orellana L. Large-Scale Conformational Changes and Protein Function: Breaking the in silico Barrier[J]. *Front Mol Biosci*: 2019, 6: 117.
- [13] Harpole T J, Delemotte L. Conformational landscapes of membrane proteins delineated by enhanced sampling molecular dynamics simulations[J]. *Biochim Biophys Acta Biomembr*: 2018, 1860: 909-926.
- [14] Wang J, Peng C, Yu Y, et al. Exploring Conformational Change of Adenylate Kinase by Replica Exchange Molecular Dynamic Simulation[J]. *Biophys J*: 2020, 118: 1009-1018.
- [15] Ngo S T, Nguyen P H, Derreumaux P. Impact of A2T and D23N Mutations on Tetrameric Abeta42 Barrel within a Dipalmitoylphosphatidylcholine Lipid Bilayer Membrane by Replica Exchange Molecular Dynamics[J]. *J Phys Chem B*: 2020, 124: 1175-1182.

- [16] Karanicolas J, Kuhlman B. Computational design of affinity and specificity at protein-protein interfaces[J]. *Curr Opin Str Biol*: 2009, 19: 458-463.
- [17] Havranek J J, Harbury P B. Automated design of specificity in molecular recognition[J]. *Nat Str Biol*: 2003, 10: 45-52.
- [18] Sammond D W, Eletr Z M, Purbeck C, et al. Structure-based protocol for identifying mutations that enhance protein-protein binding affinities[J]. *J Mol Biol*: 2007, 371: 1392-1404.
- [19] Guvench O, MacKerell J. Computational evaluation of protein-small molecule binding[J]. *Curr Opin Str Biol*: 2009, 19: 56-61.
- [20] Brooijmans N, Kuntz I D. Molecular recognition and docking algorithms[J]. *Ann Rev Biophys Biomol Struct*: 2003, 32: 335-373.
- [21] Sousa S F, Fernandes P A. Protein-ligand docking: current status and future challenges[J]. *Proteins*: 2006, 65: 15-26.
- [22] Alonso H, Bliznyuk A A, Gready J E. Combining docking and molecular dynamic simulations in drug design[J]. *Med Res Rev*: 2006, 26: 531-568.
- [23] Schnell J R, Chou J J. Structure and mechanism of the M2 proton channel of influenza A virus[J]. *Nature*: 2008, 451: 591-595.
- [24] Zhou G P. The disposition of the LZCC protein residues in wenxiang diagram provides new insights into the protein-protein interaction mechanism[J]. *J Theor Biol*: 2011, 284: 142-148.
- [25] Chou K C, Jones D, Heinrikson R L. Prediction of the tertiary structure and substrate binding site of caspase-8[J]. *FEBS Lett*: 1997, 419: 49-54.
- [26] Leach A R. Molecular modelling. Principles and applications[J]. Pearson Education Ltd. Harlow: 2001.
- [27] Zwanzig R W. High-temperature equation of state by a perturbation method. I. Nonpolar gases[J]. *J Chem Phys*: 1954, 22: 1420-1426.
- [28] Singh U C, Brown F K, Bash P A, et al. An approach to the application of free energy perturbation methods using molecular dynamics[J]. *J Am Chem Soc*: 1987, 109: 1607-1614.
- [29] Jayachandran G, Shirts M R, Park S, et al. Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics[J]. *J Chem Phys*: 2006, 125: 084901.
- [30] Chodera J D, Mobley D L, Shirts R M, et al. Alchemical Free energy methods for drug discovery: Progress and challenges[J]. *Curr Opin Str Biol*: 2011, 21: 150-160.
- [31] Åqvist J, Medina C, Samuelsson J-E. A new method for predicting binding affinity in computer-aided drug design[J]. *Protein Eng*: 1994, 7: 385-391.
- [32] Wallnoefer H G, Liedl K R, Fox T. A challenging system: free energy prediction for factor Xa[J]. *J Comput Chem*: 2011, 32: 1743-1752.
- [33] Ashida T, Kikuchi T. Estimation of relative binding free energy based on a free energy variational principle for the FKBP-ligand system[J]. *J Comput Aided Mol Design*: 2013, 27: 479-490.
- [34] Lee B, Richards F M. The interpretation of protein structures: Estimation of static

- accessibility[J]. J Mol Biol: 1971, 55: 379-400.
- [35] Richards F M. Areas, volumes, packing and protein structure[J]. Annu Rev Biophys Bineng.: 1977, 6: 151-176.
- [36] Connolly M L. Solvent-accessible surfaces of proteins and nucleic acids[J]. Science: 1983, 221: 709-713.
- [37] Richmond T J. Solvent accessible surface area and excluded volume in proteins[J]. J Mol Biol: 1984, 178: 63-88.
- [38] Shrake A, Rupley J A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin[J]. J Mol Biol: 1973, 79: 351-371.
- [39] Kouza M, Hansmann U H. Velocity scaling for optimizing replica exchange molecular dynamics[J]. J Chem Phys: 2011, 134: 044124.
- [40] Itoh S G, Okumura H. Replica-Permutation Method with the Suwa-Todo Algorithm beyond the Replica-Exchange Method[J]. J Chem Theory Comput: 2013, 9: 570-581.
- [41] Itoh S G, Okumura H. Replica-permutation method to enhance sampling efficiency[J]. Mol Simulat: 2014, 41: 1021-1026.
- [42] Itoh S G, Okumura H. Oligomer Formation of Amyloid-beta(29-42) from Its Monomers Using the Hamiltonian Replica-Permutation Molecular Dynamics Simulation[J]. J Phys Chem B: 2016, 120: 6555-6561.
- [43] Yamauchi M, Okumura H. Replica sub-permutation method for molecular dynamics and monte carlo simulations[J]. J Comput Chem: 2019, 40: 2694-2711.
- [44] Zhang B W, Dai W, Gallicchio E, et al. Simulating Replica Exchange: Markov State Models, Proposal Schemes, and the Infinite Swapping Limit[J]. J Phys Chem B: 2016, 120: 8289-8301.
- [45] Chodera J D, Shirts M R. Replica exchange and expanded ensemble simulations as Gibbs sampling: simple improvements for enhanced mixing[J]. J Chem Phys: 2011, 135: 194110.
- [46] Vogel T, Perez D. Accelerating the Convergence of Replica Exchange Simulations Using Gibbs Sampling and Adaptive Temperature Sets[J]. Phys Procedia: 2015, 68: 125-129.
- [47] Brenner P, Sweet C R, VonHandorf D, et al. Accelerating the replica exchange method through an efficient all-pairs exchange[J]. J Chem Phys: 2007, 126: 074103.
- [48] Yua Q, Lub J, Abramse. C F, et al. Multiscale implementation of infinite-swap replica exchange molecular dynamics[J]. Proc Natl Acad Sci USA: 2016, 113: 11744-11749.
- [49] Nuria P, Doll J D, Dupuis P, et al. An infinite swapping approach to the rare-event sampling problem[J]. J Chem Phys: 2011, 135: 1-14.
- [50] Lu J, Vanden-Eijnden E. Infinite swapping replica exchange molecular dynamics leads to a simple simulation patch using mixture potentials[J]. J Chem Phys: 2013, 138: 1-9.
- [51] Urano R, Okamoto Y. Designed-walk replica-exchange method for simulations of complex systems[J]. Comput Phys Commun: 2015, 196: 380-383.
- [52] Mu Y, Yang Y, Xu W. a global optimization scheme: kernel replica exchange simulation

- method for protein folding[J]. *J Theor Comput Chem*: 2008, 7: 177-187.
- [53] Mu Y, Yang Y, Xu W. Hybrid Hamiltonian replica exchange molecular dynamics simulation method employing the Poisson-Boltzmann model[J]. *J Chem Phys*: 2007, 127: 084119.
- [54] Chaudhury S, Olson M A, Tawa G, et al. Efficient Conformational Sampling in Explicit Solvent Using a Hybrid Replica Exchange Molecular Dynamics Method[J]. *J Chem Theory Comput*: 2012, 8: 677-687.
- [55] Weixin X, Ye Y, Mu Y, et al. Global optimisation by replica exchange with scaled hybrid Hamiltonians[J]. *Mol Simulat*: 2008, 34: 575-590.
- [56] Asim O, Lauren W, Melinda L, et al. Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model[J]. *J Chem Theory Comput*: 2006, 2: 420-433.
- [57] Wang J, Zhu W, Li G, et al. Velocity-scaling optimized replica exchange molecular dynamics of proteins in a hybrid explicit/implicit solvent[J]. *J Chem Phys*: 2011, 135: 084115.
- [58] Yu Y, Wang J, Shao Q, et al. Increasing the sampling efficiency of protein conformational transition using velocity-scaling optimized hybrid explicit/implicit solvent REMD simulation[J]. *J Chem Phys*: 2015, 142: 125105.
- [59] Cheng X. Modified Replica Exchange Simulation Methods for Local Structure Refinement[J]. *J Phys Chem B*: 2005, 109: 8220-8230.
- [60] Fukunishi H, Watanabe O, Takada S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction[J]. *J Chem Phys*: 2002, 116: 9058-9067.
- [61] Itoh S G, Okumura H, Okamoto Y. Replica-exchange method in van der Waals radius space: Overcoming steric restrictions for biomolecules[J]. *J Chem Phys*: 2010, 132: 134105.
- [62] Itoh S G, Okumura H. Coulomb replica-exchange method: Handling electrostatic attractive and repulsive forces for biomolecules[J]. *J Comput Chem*: 2013, 34: 622-639.
- [63] Itoh S G, Okumura H. Dimerization process of amyloid-beta(29-42) studied by the Hamiltonian replica-permutation molecular dynamics simulations[J]. *J Phys Chem B*: 2014, 118: 11428-11436.
- [64] Rathinavelan T, Im W. A novel strategy to determine protein structures using exclusively residual dipolar coupling[J]. *J Comput Chem*: 2008, 29: 1640-1649.
- [65] Chen J, Im W, Brooks C L. Application of torsion angle molecular dynamics for efficient sampling of protein conformations[J]. *J Comput Chem*: 2005, 26: 1565-1578.
- [66] Okumura H, Itoh S G. Transformation of a design peptide between the α -helix and β -hairpin structures using a helix-strand replica-exchange molecular dynamics simulation[J]. *Phys Chem Chem Phys*: 2013, 15: 13852.
- [67] Jang S, Shin S, Pak Y. Replica-exchange method using the generalized effective potential[J]. *Phys Rev Lett*: 2003, 91: 058305.
- [68] Liu P, Kim B, Friesner R A, et al. Replica exchange with solute tempering: a method for

- sampling biological systems in explicit water[J]. *Proc Natl Acad Sci U S A*: 2005, 102: 13749-13754.
- [69] Samuel L, Servaas M, Ceulemans A. Improved Replica Exchange Method for Native-State Protein Sampling[J]. *J Chem Theory Comput*: 2011, 7: 231-237.
- [70] Kamiya M, Sugita Y. Flexible selection of the solute region in replica exchange with solute tempering: Application to protein-folding simulations[J]. *J Chem Phys*: 2018, 149: 072304.
- [71] Suzuki M, Okuda H. Fragment replica-exchange method for efficient protein conformation sampling[J]. *Mol Simulat*: 2008, 34: 267-275.
- [72] Lyman E, Zuckerman D M. Resolution Exchange Simulation with Incremental Coarsening[J]. *J Chem Theory Comput*: 2006, 2: 656-666.
- [73] Piana S, Laio A. A Bias-Exchange Approach to Protein Folding[J]. *J Phys Chem B*: 2007, 111: 4553-4559.
- [74] Kubitzki M B, de Groot B L. Molecular Dynamics Simulations Using Temperature-Enhanced Essential Dynamics Replica Exchange[J]. *Biophys J*: 2007, 92: 4262-4270.
- [75] Berg B A, Neuhaus T. Multicanonical algorithms for first order phase transitions[J]. *Phys Lett B*: 1991: 249-253.
- [76] Berg B A, Neuhaus T. Multicanonical ensemble: A new approach to simulate first-order phase transitions[J]. *Phys Rev Lett*: 1992, 68: 9-12.
- [77] Qi R, Wei G, Ma B. Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example[J]. *Methods Mol Biol*: 2018, 1777: 101.
- [78] Hayashi T, Okamoto Y. Efficient simulation protocol for determining the density of states: Combination of replica-exchange Wang-Landau method and multicanonical replica-exchange method[J]. *Phys Rev E*: 2019, 100.
- [79] Mikolai F, Donald H, McCammon J A. Replica-Exchange Accelerated Molecular Dynamics (REXAMD) Applied to Thermodynamic Integration[J]. *J Chem Theory Comput*: 2008, 4: 1565-1569.
- [80] Huang Y M, McCammon J A, Miao Y. Replica Exchange Gaussian Accelerated Molecular Dynamics: Improved Enhanced Sampling and Free Energy Calculation[J]. *J Chem Theory Comput*: 2018, 14: 1853-1864.
- [81] Oshima H, Re S, Sugita Y. Replica-Exchange Umbrella Sampling Combined with Gaussian Accelerated Molecular Dynamics for Free-Energy Calculation of Biomolecules[J]. *J Chem Theory Comput*: 2019, 15: 5199-5208.
- [82] Sugita Y, Kitao A, Okamoto Y. Multidimensional replica-exchange method for free-energy calculations[J]. *J Chem Phys*: 2000, 113: 6042-6051.
- [83] Sidler D, Schwaninger A, Riniker S. Replica exchange enveloping distribution sampling (RE-EDS): A robust method to estimate multiple free-energy differences from a single simulation[J]. *J Chem Phys*: 2016, 145: 154114.
- [84] Armacost K A, Goh G B, Brooks C L, 3rd. Biasing Potential Replica Exchange Multisite

- lambda-Dynamics for Efficient Free Energy Calculations[J]. *J Chem Theory Comput*: 2015, 11: 1267-1277.
- [85] Galvelis R, Re S, Sugita Y. Enhanced Conformational Sampling of N-Glycans in Solution with Replica State Exchange Metadynamics[J]. *J Chem Theory Comput*: 2017, 13: 1934-1942.
- [86] Galvelis R, Sugita Y. Replica state exchange metadynamics for improving the convergence of free energy estimates[J]. *J Comput Chem*: 2015, 36: 1446-1455.
- [87] Giovanni B, Francesco L, Alessandro L, et al. Free-Energy Landscape for Hairpin Folding from Combined Parallel Tempering and Metadynamics[J]. *J Am Chem Soc*: 2006, 128: 13435-13441.
- [88] Bonomi M, Parrinello M. Enhanced sampling in the well-tempered ensemble[J]. *Phys Rev Lett*: 2010, 104: 190601.
- [89] Jiang W, Roux B. Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations[J]. *J Chem Theory Comput*: 2010, 6: 2559-2565.
- [90] Zheng J, Alsamarah A, LaCuran A E, et al. Uncovering Molecular Bases Underlying Bone Morphogenetic Protein Receptor Inhibitor Selectivity[J]. *Plos One*: 2015, 10: e0132221.
- [91] Bergonzo C, Henriksen N M, Roe D R, et al. Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields[J]. *RNA*: 2015, 21: 1578-1590.
- [92] Roe D R, Bergonzo C, Cheatham T E, 3rd. Evaluation of enhanced sampling provided by accelerated molecular dynamics with Hamiltonian replica exchange methods[J]. *J Phys Chem B*: 2014, 118: 3543-3552.
- [93] Arrar M, de Oliveira C A, Fajer M, et al. w-REXAMD: A Hamiltonian Replica Exchange Approach to Improve Free Energy Calculations for Systems with Kinetically Trapped Conformations[J]. *J Chem Theory Comput*: 2013, 9: 18-23.
- [94] Okamoto Y, Kokubo H, Tanaka T. Prediction of Ligand Binding Affinity by the Combination of Replica-Exchange Method and Double-Decoupling Method[J]. *J Chem Theory Comput*: 2014, 10: 3563-3569.
- [95] Kokubo H, Tanaka T, Okamoto Y. Two-dimensional replica-exchange method for predicting protein-ligand binding structures[J]. *J Comput Chem*: 2013, 34: 2601-2614.
- [96] Ebrahimi P, Kaur S, Baronti L, et al. A two-dimensional replica-exchange molecular dynamics method for simulating RNA folding using sparse experimental restraints[J]. *Methods*: 2019, 162-163: 96-107.
- [97] Babin V, Karpusenka V, Moradi M, et al. Adaptively biased molecular dynamics: An umbrella sampling method with a time-dependent potential[J]. *Inter J Quantum Chem*: 2009, 109: 3666-3678.
- [98] Lee J, Miller B T, Damjanović A, et al. Enhancing Constant-pH Simulation in Explicit Solvent with a Two-Dimensional Replica Exchange Method[J]. *J Chem Theory Comput*: 2015, 11: 2560-2574.

- [99] Lee J, Miller B T, Brooks B R. Computational scheme for pH-dependent binding free energy calculation with explicit solvent[J]. *Protein Sci*: 2016, 25: 231-243.
- [100] Wang L, Deng Y, Knight J L, et al. Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors[J]. *J Chem Theory Comput*: 2013, 9: 1282-1293.
- [101] Wang L, Berne B J, Friesner R A. On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities[J]. *Proc Natl Acad Sci U S A*: 2012, 109: 1937-1942.
- [102] Jiang W, Thirman J, Jo S, et al. Reduced Free Energy Perturbation/Hamiltonian Replica Exchange Molecular Dynamics Method with Unbiased Alchemical Thermodynamic Axis[J]. *J Phys Chem B*: 2018, 122: 9435-9442.
- [103] Lee M, Yoon J, Jang S, et al. Conformational sampling of metastable states: Tq-REM as a novel replica exchange method[J]. *Phys Chem Chem Phys*: 2017, 19: 5454-5464.
- [104] Cruzeiro V W D, Roitberg A E. Multidimensional Replica Exchange Simulations for Efficient Constant pH and Redox Potential Molecular Dynamics[J]. *J Chem Theory Comput*: 2019, 15: 871-881.
- [105] Nguyen P H. Replica exchange simulation method using temperature and solvent viscosity[J]. *J Chem Phys*: 2010, 132: 144109.
- [106] Mori T, Jung J, Sugita Y. Surface-Tension Replica-Exchange Molecular Dynamics Method for Enhanced Sampling of Biological Membrane Systems[J]. *J Chem Theory Comput*: 2013, 9: 5629-5640.
- [107] Nagai T, Takahashi T. Mass-scaling replica-exchange molecular dynamics optimizes computational resources with simpler algorithm[J]. *J Chem Phys*: 2014, 141: 114111.
- [108] Liu L, Kuo J L. A LAMMPS implementation of volume-temperature replica exchange molecular dynamics[J]. *Comput Phys Commun*: 2015, 189: 119-127.
- [109] Li X, O'Brien C P, Collier G, et al. An improved replica-exchange sampling method: temperature intervals with global energy reassignment[J]. *J Chem Phys*: 2007, 127: 164116.
- [110] Li X, Latour R A, Stuart S J. TIGER2: an improved algorithm for temperature intervals with global exchange of replicas[J]. *J Chem Phys*: 2009, 130: 174106.
- [111] Li X, Snyder J A, Stuart S J, et al. TIGER2 with solvent energy averaging (TIGER2A): An accelerated sampling method for large molecular systems with explicit representation of solvent[J]. *J Chem Phys*: 2015, 143: 144105.
- [112] Kulke M, Geist N, Möller D, et al. Replica-Based Protein Structure Sampling Methods: Compromising between Explicit and Implicit Solvents[J]. *J Phys Chem B*: 2018, 122: 7295-7307.
- [113] Geist N, Kulke M, Schulig L, et al. Replica-Based Protein Structure Sampling Methods II: Advanced Hybrid Solvent TIGER2HS[J]. *J Phys Chem B*: 2019, 123: 5995-6006.
- [114] Li X, Murthy S, Latour R A. The Construction and Validation of All-Atom Bulk-Phase Models of Amorphous Polymers Using the TIGER2/TIGER3 Empirical Sampling Method[J].

Macromol: 2011, 44: 5452-5464.

- [115] Zhang W, Chen J. Efficiency of Adaptive Temperature-Based Replica Exchange for Sampling Large-Scale Protein Conformational Transitions[J]. J Chem Theory Comput: 2013, 9: 2849-2856.
- [116] Trebst S, Troyer M, Hansmann U H. Optimized parallel tempering simulations of proteins[J]. J Chem Phys: 2006, 124: 174903.
- [117] Kim J, Straub J E, Keyes T. Replica exchange statistical temperature molecular dynamics algorithm[J]. J Phys Chem B: 2012, 116: 8646-8653.
- [118] Chen C, Xiao Y. Improving the replica-exchange molecular-dynamics method for efficient sampling in the temperature space[J]. Phys Rev E: 2015, 91: 052708.
- [119] Chen C, Huang Y. Walking freely in the energy and temperature space by the modified replica exchange molecular dynamics method[J]. J Comput Chem: 2016, 37: 1565-1575.
- [120] Asim O, Daniel R. Roe, Guanglei Cui, et al. Improving Convergence of Replica-Exchange Simulations through Coupling to a High-Temperature Structure Reservoir[J]. J Chem Theory Comput: 2007, 3: 557-568.
- [121] Henriksen N M, Roe D R, Cheatham T E, 3rd. Reliable oligonucleotide conformational ensemble generation in explicit solvent for force field assessment using reservoir replica exchange molecular dynamics simulations[J]. J Phys Chem B: 2013, 117: 4014-4027.
- [122] Li H, Li G, Berg B A, et al. Finite reservoir replica exchange to enhance canonical sampling in rugged energy surfaces[J]. J Chem Phys: 2006, 125: 144902.
- [123] Rick S W. Replica exchange with dynamical scaling[J]. J Chem Phys: 2007, 126: 054102.
- [124] Shell M S. A replica-exchange approach to computing peptide conformational free energies[J]. Mol Simulat: 2010, 36: 505-515.
- [125] Kima J, Straub J E. Optimal replica exchange method combined with Tsallis weight sampling[J]. J Chem Phys: 2009, 130: 1-11.
- [126] Wu X, Hodoscek M, Brooks B R. Replica exchanging self-guided Langevin dynamics for efficient and accurate conformational sampling[J]. J Chem Phys: 2012, 137: 044106.
- [127] Olson M A, Legler P M, Goldman E R. Comparison of Replica Exchange Simulations of a Kinetically Trapped Protein Conformational State and its Native Form[J]. J Phys Chem B: 2016, 120: 2234-2240.
- [128] Dmitri G, Yuji S, Choi C H. Efficient Parallel Implementations of QM/MM-REMD (Quantum Mechanical/Molecular Mechanics-Replica-Exchange MD) and Umbrella Sampling: Isomerization of H₂O₂ in Aqueous Solution[J]. J Phys Chem B: 2013, 117: 7996-8002.
- [129] Moqadam M, Riccardi E, Trinh T T, et al. Rare event simulations reveal subtle key steps in aqueous silicate condensation[J]. Phys Chem Chem Phys: 2017, 19: 13361-13371.
- [130] Lin W, Paesani F. Systematic Study of Structural and Thermodynamic Properties of HCl(H₂O)_n Clusters from Semiempirical Replica Exchange Simulations[J]. J Phys Chem A: 2013, 117: 7131-7141.

- [131] Buló R E, Van Schoot H, Rohr D, et al. Bias-exchange metadynamics applied to the study of chemical reactivity[J]. *Inter J Quantum Chem*: 2010, 110: 2299-2307.
- [132] Zhou R. Exploring Protein Folding Free Energy Landscape: Replica Exchange Monte Carlo[J]. *J. Phys Chem B*:2003, 115: 8806-8812.
- [133] Suh D, Radak B K, Chipot C, et al. Enhanced configurational sampling with hybrid non-equilibrium molecular dynamics-Monte Carlo propagator[J]. *J Chem Phys*: 2018, 148: 014101.
- [134] Kandel S, Salomon-Ferrer R, Larsen A B, et al. Overcoming potential energy distortions in constrained internal coordinate molecular dynamics simulations[J]. *J Chem Phys*: 2016, 144: 044112.
- [135] Lockhart C, O'Connor J, Armentrout S, et al. Greedy replica exchange algorithm for heterogeneous computing grids[J]. *J Mol Model*: 2015, 21: 1-12.
- [136] Platania R, Shams S, Chiu C-H, et al. Hadoop-based replica exchange over heterogeneous distributed cyberinfrastructures[J]. *Concurrency Computat: Pract Exper* 2017, 29: e3878.
- [137] Xia J, Flynn W F, Gallicchio E, et al. Large-scale asynchronous and distributed multidimensional replica exchange molecular simulations and efficiency analysis[J]. *J Comput Chem*: 2015, 36: 1772-1785.
- [138] Ferreira L, Oliva G, Andricopulo, A D. Molecular docking and structure-based drug design strategies[J]. *Molecules*: 2015, 20: 13384-13421.
- [139] Ryde U, Soderhjelm P. Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods[J]. *Chem Rev*: 2016, 116: 5520-5566.
- [140] Jorgensen W L. Free energy calculations: a breakthrough for modeling organic chemistry in solution[J]. *Acc Chem Res*: 2002, 22: 184-189.
- [141] Chung W. Dynamics and design of enzymes and inhibitors[J]. *J Am Chem Soc*: 1986, 108: 3830-3832.
- [142] Jan A. A new method for predicting binding affinity in computer-aided drug design[J]. *Protein Eng*: 1994, 7: 385-391.
- [143] Tanfit H. Ligand binding affinity prediction by linear interaction energy methods[J]. *J. Comput Aided Mol Des.*: 1998, 12: 27-35.
- [144] Pu K. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models[J]. *Acc Chem Res*: 2000, 33: 889-897.
- [145] Manus N. Solvated interaction energy (SIE) for scoring protein–ligand binding affinities. 1. Exploring the parameter space[J]. *J Chem Inf Model*: 2007, 47: 122-133.
- [146] Xu L, Sun H, Li Y, et al. Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models[J]. *J Phys Chem B*: 2013, 117: 8408-8421.
- [147] Aaron Weis, Katebzade K. Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field[J]. *J Med Chem*: 2006, 49: 6596-6606.
- [148] Jay P. Wu C. Current Status of the AMOEBA Polarizable Force Field[J]. *J Phys Chem B*:

2010, 114: 2549-2564.

[149] Piquemal P, Jordan K D. From quantum mechanics to force fields: new methodologies for the classical simulation of complex systems[J]. *Theor Chem Acc*: 2012, 131.

[150] Khoruzhii O, Donchev A G, Galkin N, et al. Application of a polarizable force field to calculations of relative protein-ligand binding affinities[J]. *Proc Natl Acad Sci USA*: 2008, 105: 10378-10383.

[151] Jiao D, Golubkov P A, Darden T A, et al. Calculation of protein-ligand binding free energy by using a polarizable potential[J]. *Proc Natl Acad Sci USA*: 2008, 105: 6290-6295.

[152] Par S. Polarization effects in protein-ligand calculations extend farther than the actual induction energy[J]. *Theor Chem Acc*: 2012, 131: 1159.

[153] Cieplak P, Dupradeau F-Y, Duan Y, et al. Polarization effects in molecular mechanical force fields[J]. *J Phys: Condensed Matter*: 2009, 21: 333102.

[154] Cho A E, Guallar V, Berne B J, et al. Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach[J]. *J Comput Chem*: 2005, 26: 915-931.

[155] Wang J, Cossins B, Shi J, Chen K, Zhu W. Thermodynamics calculation of protein-ligand interactions by QM/MM polarizable charge parameters[J]. *J Biomol Struct Dyn*: 2016, 34: 163-176.

[156] Rocchia W, Sridharan S, Nicholls A, et al. Rapid grid - based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects[J]. *J Comput Chem*: 2002, 23: 128-137.

[157] Dral P, Wu X. Semiempirical quantum-chemical orthogonalization-corrected methods: benchmarks for ground-state properties[J]. *J Chem Theory Comput*: 2016, 12: 1097-1120.

[158] Yilmazer N D, Korth M. Enhanced semiempirical QM methods for biomolecular interactions[J]. *Comput Struct Biotech J*: 2015, 13: 169-175.

[159] Bikadi Z, Hazai E. Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock[J]. *J Cheminf*: 2009, 1: 15.

[160] Chen J, Wang J, Zhang Q, et al. A comparative study of trypsin specificity based on QM/MM molecular dynamics simulation and QM/MM GBSA calculation[J]. *J Biomol Struct Dyn*: 2015, 33: 2606-2618.

[161] Dewar M J, Zoebisch E G, Healy E F, et al. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model[J]. *J Ame Chem Soc*: 1985, 107: 3902-3909.

[162] Stewart J J. Optimization of parameters for semiempirical methods I. Method[J]. *J Comput Chem*: 1989, 10: 209-220.

[163] Stewart J J. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements[J]. *J Mol Model*: 2007, 13: 1173-1213.

[164] Stewart J J. Optimization of parameters for semiempirical methods VI: more

modifications to the NDDO approximations and re-optimization of parameters[J]. J Mol Model: 2013, 19: 1-32.

[165] Gungor O, Stephen Q, Hernandez R. Adaptive Steered Molecular Dynamics of the Long-Distance Unfolding of Neuropeptide Y[J]. J Chem Theor Comput: 2010, 6: 3026-3038.

[166] Eric M. Structure-based design of estrogen receptor- β selective ligand[J]. J. Ame Chem Soc: 2004, 126: 15106-15119.

[167] Malamas M, McDevitt R. Design and synthesis of aryl diphenolic azoles as potent and selective estrogen receptor- β ligands[J]. J Med Chem: 2004, 47: 5021-5040.

[168] Eric M. Benzopyrans are selective estrogen receptor β agonists with novel activity in models of benign prostatic hyperplasia[J]. J Med Chem: 2006, 49: 6155-6157.

[169] McDevitt R. ER β ligands. 3. Exploiting two binding orientations of the 2-phenylnaphthalene scaffold to achieve ER β selectivity[J]. J Med Chem: 2005, 48: 3953-3979.

[170] Mocklinghoff S, van Otterlo W A, Rose R, et al. Design and evaluation of fragment-like estrogen receptor tetrahydroisoquinoline ligands from a scaffold-detection approach[J]. J Med Chem: 2011, 54: 2005-2011.

[171] Sulsky R, Magnin D R, Huang Y, et al. Potent and selective biphenyl azole inhibitors of adipocyte fatty acid binding protein (aFABP)[J]. Bioorg Med Chem Lett: 2007, 17: 3511-3515.

[172] Young A C, Scapin, G., Kromminga, A., Patel, S. B., Veerkamp, J. H., & Sacchettini, J. C. Structural studies on human muscle fatty acid binding protein at 1.4 Å resolution: binding interactions with three C18 fatty acids[J]. Structure: 1994, 2: 523-534.

[173] Ringom R, Axen E, Uppenberg J, et al. Substituted benzylamino-6-(trifluoromethyl)pyrimidin-4(1H)-ones: a novel class of selective human A-FABP inhibitors[J]. Bioorg Med Chem Lett: 2004, 14: 4449-4452.

[174] Lehmann F, Haile S, Axen E, et al. Discovery of inhibitors of human adipocyte fatty acid-binding protein, a potential type 2 diabetes target[J]. Bioorg Med Chem Lett: 2004, 14: 4445-4448.

[175] Gillilan R E A S D, Noy N. Structural basis for activation of fatty acid-binding protein 4[J]. J Mol Biol: 2007, 372: 1246-1260.

[176] Hertz A V, Hellberg K, Reynolds J M, et al. Identification and characterization of a small molecule inhibitor of Fatty Acid binding proteins[J]. J Med Chem: 2009, 52: 6024-6031.

[177] McDevitt R. Studies of the ligand binding reaction of adipocyte lipid binding protein using the fluorescent probe 1, 8-anilinonaphthalene-8-sulfonate[J]. Biophys J: 1999, 77: 1107-1116.

[178] Barf T, Lehmann F, Hammer K, et al. N-Benzyl-indolo carboxylic acids: Design and synthesis of potent and selective adipocyte fatty-acid binding protein (A-FABP) inhibitors[J]. Bioorg Med Chem Lett: 2009, 19: 1745-1748.

[179] Malamas M S, Erdei J, Gunawan I, et al. Design and synthesis of 5,5'-disubstituted aminohydantoin as potent and selective human beta-secretase (BACE1) inhibitors[J]. J Med Chem: 2010, 53: 1146-1158.

[180] Cumming J N, Smith E M, Wang L, et al. Structure based design of iminohydantoin

BACE1 inhibitors: identification of an orally available, centrally active BACE1 inhibitor[J]. *Bioorg Med Chem Lett*: 2012, 22: 2444-2449.

[181] Mandal M, Zhu Z, Cumming J N, et al. Design and validation of bicyclic iminopyrimidinones as beta amyloid cleaving enzyme-1 (BACE1) inhibitors: conformational constraint to favor a bioactive conformation[J]. *J Med Chem*: 2012, 55: 9331-9345.

[182] Caldwell J P, Mazzola R D, Durkin J, et al. Discovery of potent iminoheterocycle BACE1 inhibitors[J]. *Bioorg Med Chem Lett*: 2014, 24: 5455-5459.

[183] Huai Q, Wang H, Sun Y, et al. Three-Dimensional Structures of PDE4D in Complex with Roliprams and Implication on Inhibitor Selectivity[J]. *Structure*: 2003, 11: 865-873.

[184] Card G L, England B P, Suzuki Y, et al. Structural basis for the activity of drugs that inhibit phosphodiesterases[J]. *Structure*: 2004, 12: 2233-2247.

[185] Card G L, Blasdel L, England B P, et al. A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design[J]. *Nat Biotechnol*: 2005, 23: 201-207.

[186] Burgin A B, Magnusson O T, Singh J, et al. Design of phosphodiesterase 4D (PDE4D) allosteric modulators for enhancing cognition with improved safety[J]. *Nat Biotechnol*: 2010, 28: 63-70.

[187] Nankervis J L, Feil S C, Hancock N C, et al. Thiophene inhibitors of PDE4: crystal structures show a second binding mode at the catalytic domain of PDE4D2[J]. *Bioorg Med Chem Lett*: 2011, 21: 7089-7093.

[188] Wang H, Peng M S, Chen Y, et al. Structures of the four subfamilies of phosphodiesterase-4 provide insight into the selectivity of their inhibitors[J]. *Biochem J*: 2007, 408: 193-201.

[189] Harris P A B A, Cheung M, et al. Discovery of 5-[[4-[(2, 3-dimethyl-2 H-indazol-6-yl) methylamino]-2-pyrimidinyl] amino]-2-methyl-benzenesulfonamide (Pazopanib), a novel and potent vascular endothelial growth factor receptor inhibitor[J]. *J Med Chem*: 2008, 51: 4632-4640.

[190] Brial H. Evolution of a highly selective and potent 2-(pyridin-2-yl)-1, 3, 5-triazine Tie-2 kinase inhibitor[J]. *J Med Chem*: 2007, 50: 611-626.

[191] Jeffrey S. Discovery and evaluation of 2-anilino-5-aryloxazoles as a novel class of VEGFR2 kinase inhibitors[J]. *J Med Chem*: 2005, 48: 1610-1619.

[192] Weiss M. Evaluation of a Series of Naphthamides as Potent, Orally Active Vascular Endothelial Growth Factor Receptor-2 Tyrosine Kinase Inhibitors[J]. *J Med Chem*: 2008, 51: 1668-1680.

[193] Iwata H, Oki H, Okada K, et al. A Back-to-Front Fragment-Based Drug Design Search Strategy Targeting the DFG-Out Pocket of Protein Tyrosine Kinases[J]. *ACS Med Chem Lett*: 2012, 3: 342-346.

[194] McTigue M, Murray B W, Chen J H, et al. Molecular conformations, interactions, and properties associated with drug efficiency and clinical performance among VEGFR TK inhibitors[J]. *Proc Natl Acad Sci USA*: 2012, 109: 18281-18289.

- [195] Bi C, Wang Q, Shao Y, et al. Non-wetting surface-driven high-aspect-ratio crystalline grain growth for efficient hybrid perovskite solar cells[J]. *Nat Commun*: 2015, 6: 7747.
- [196] Andreas Blum, Jark Böttcher, Andreas Heine, et al. Structure-Guided Design of C2-Symmetric HIV-1 Protease Inhibitors Based on a Pyrrolidine Scaffold[J]. *J Med Chem.*: 2008, 51: 2078-2087.
- [197] Chellappan S, Kiran Kumar Reddy G S, Ali A, et al. Design of mutation-resistant HIV protease inhibitors with the substrate envelope hypothesis[J]. *Chem Biol Drug Des*: 2007, 69: 298-313.
- [198] Altman; M D, Akbar Ali, G. S. Kiran Kumar Reddy, et al. HIV-1 Protease Inhibitors from Inverse Design in the Substrate Envelope Exhibit Subnanomolar Binding to Drug-Resistant Variants[J]. *J Am Chem Soc.*: 2008, 130: 6099-6113.
- [199] Miller J F, Andrews C W, Brieger M, et al. Ultra-potent P1 modified arylsulfonamide HIV protease inhibitors: the discovery of GW0385[J]. *Bioorg Med Chem Lett*: 2006, 16: 1788-1794.
- [200] Andersson H O, Fridborg K, Lowgren S, et al. Optimization of P1-P3 groups in symmetric and asymmetric HIV-1 protease inhibitors[J]. *European J Biochem*: 2003, 270: 1746-1758.
- [201] Suvit T, Harvey I, Steve R, et al. Structure-Based Design of HIV Protease Inhibitors: Sulfonamide-Containing 5,6-Dihydro-4-hydroxy-2-pyrones as Non-Peptidic Inhibitors[J]. *J Med Chem*: 1996, 39: 4349-4353.
- [202] Lindberg J, Pyring D, Lowgren S, et al. Symmetric fluoro-substituted diol-based HIV protease inhibitors. Ortho-fluorinated and meta-fluorinated P1/P1'-benzyloxy side groups significantly improve the antiviral activity and preserve binding efficacy[J]. *Eur J Biochem*: 2004, 271: 4594-4602.
- [203] Tie Y, Kovalevsky A Y, Boross P, et al. Atomic resolution crystal structures of HIV-1 protease and mutants V82A and I84V with saquinavir[J]. *Proteins*: 2007, 67: 232-242.
- [204] Kawasaki Y, Chufan E E, Lafont V, et al. How much binding affinity can be gained by filling a cavity?[J]. *Chem Biol Drug Des*: 2010, 75: 143-151.
- [205] Xiongyu W, Per Ö, Jenny K, et al. Two-Carbon-Elongated HIV-1 Protease Inhibitors with a Tertiary-Alcohol -Containing Transition-State Mimic[J]. *J Med Chem*: 2008, 51: 1053-1057.
- [206] Appelt J, Burgess K. Viracept (Nelfinavir Mesylate, AG1343): A Potent, Orally Bioavailable Inhibitor of HIV-1 Protease[J]. *J Med Chem*: 1997, 40: 3979-3985.
- [207] King N M, Prabu-Jeyabalan M, Bandaranayake R M, et al. Extreme entropy-enthalpy compensation in a drug-resistant variant of HIV-1 protease[J]. *ACS Chem Biol*: 2012, 7: 1536-1546.
- [208] Yuan-Fang W, Yunfeng T, Peter I B, et al. Potent New Antiviral Compound Shows Similar Inhibition and Structural Interactions with Drug Resistant Mutants and Wild Type HIV-1 Protease[J]. *J Med Chem*: 2007, 50: 4509-4515.
- [209] Shen C H, Wang Y F, Kovalevsky A Y, et al. Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters[J]. *FEBS J*: 2010, 277: 3699-3714.
- [210] Arun K, David D. Design of HIV-1 Protease Inhibitors with Pyrrolidinones and

Oxazolidinones as Novel P1'-Ligands To Enhance Backbone-Binding Interactions with Protease: Synthesis, Biological Evaluation, and Protein-Ligand X-ray Studies[J]. *J Med Chem*: 2009, 52: 3902-3914.

[211] Stewart A. ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans[J]. *Proc. Natl. Acad. Sci. USA*: 1995, 92: 2484-2488.

[212] Nalam M N, Ali A, Reddy G S, et al. Substrate envelope-designed potent HIV-1 protease inhibitors to avoid drug resistance[J]. *Chem Biol*: 2013, 20: 1116-1124.

[213] Wang J, Wolf R M, Caldwell J W, et al. Development and testing of a general amber force field[J]. *J Comput Chem*: 2004, 25: 1157-1174.

[214] Duan Y, Wu C, Chowdhury S, et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed - phase quantum mechanical calculations[J]. *J Comput Chem*: 2003, 24: 1999-2012.

[215] Ryckaert J-P, Ciccotti G, Berendsen H J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes[J]. *J. Comput. Phys.*: 1977, 23: 327-341.

[216] Essmann U, Perera L, Berkowitz M L, et al. A smooth particle mesh Ewald method[J]. *J Chem Phys*: 1995, 103: 8577-8593.

[217] Ozer G, Quirk S, Hernandez R. Adaptive steered molecular dynamics: validation of the selection criterion and benchmarking energetics in vacuum[J]. *J Chem Phys*: 2012, 136: 215104.

[218] Gungor O, Edward F V, Stephen Q, et al. Adaptive Steered Molecular Dynamics of the Long-Distance Unfolding of Neuropeptide Y[J]. *J Chem Theor Comput.*: 2010, 6: 3026-3038.

[219] Ozer G, Keyes T, Quirk S, et al. Multiple branched adaptive steered molecular dynamics[J]. *J Chem Phys*: 2014, 141: 064101.

[220] Hou T, Wang J, Li Y, et al. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations[J]. *J Chem Inf Model*: 2010, 51: 69-82.

[221] Xu L, Sun H, Li Y, et al. Assessing the performance of MM/PBSA and MM/GBSA methods. 3. the impact of force fields and ligand charge models[J]. *J Phys Chem B*: 2013, 117: 8408-8421.

[222] Yoshifumi F, Daisuke M, Nakamura H. Protein-Ligand Binding Free Energy Calculation by the Smooth Reaction Path Generation (SRPG) Method[J]. *J Chem Inf Model*: 2009, 49: 1944-1951.

[223] Ilardi E A, Vitaku E, Njardarson J T. Data-mining for sulfur and fluorine: an evaluation of pharmaceuticals to reveal opportunities for drug design and discovery[J]. *J Med Chem*: 2014, 57: 2832-2842.

[224] Bohm H J, Banner D, Bendels S, et al. Fluorine in medicinal chemistry[J]. *Chembiochem*: 2004, 5: 637-643.

[225] Englander S W, Mayne L. The nature of protein folding pathways[J]. *Proc Natl Acad Sci*

USA: 2014, 111: 15873-15880.

[226] Kappel K, Miao Y, McCammon J A. Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor[J]. *Q Rev Biophys*: 2015, 48: 479-487.

[227] Skwarczynska M, Ottmann C. Protein-protein interactions as drug targets [J]. *Fut Med Chem*: 2015, 7: 2195-2219.

[228] Doerr S, Harvey M J, Noe F, et al. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery[J]. *J Chem Theory Comput.* : 2016, 12: 1845-1852.

[229] Dror R O, Green H F, Valant C, et al. Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs[J]. *Nature*: 2013, 503: 295-299.

[230] Nolan T L, Geffert L M, Kolber B J, et al. Discovery of novel-scaffold monoamine transporter ligands via in silico screening with the S1 pocket of the serotonin transporter[J]. *ACS Chem Neurosci*: 2014, 5: 784-792.

[231] Shukla D, Meng Y, Roux B, et al. Activation pathway of Src kinase reveals intermediate states as targets for drug design[J]. *Nat Commun*: 2014, 5: 3397.

[232] Karplus M, McCammon J A. Molecular dynamics simulations of biomolecules[J]. *Nat Struct Mol Biol*: 2002, 9: 646.

[233] Dror R O, Dirks R M, Grossman J P, et al. Biomolecular simulation: a computational microscope for molecular biology[J]. *Annu Rev Biophys*: 2012, 41: 429-452.

[234] Laio A, Parrinello M. Escaping free-energy minima[J]. *Proc Natl Acad Sci USA*: 2002, 99: 12562-12566.

[235] Torrie G M, Valleau J P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling[J]. *J Comput Phys*: 1977, 23: 187-199.

[236] Schlitter J, Engels M, Krüger P. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions[J]. *J Mol Graphics*: 1994, 12: 84-89.

[237] Darve E, Rodríguez-Gómez D, Pohorille A. Adaptive biasing force method for scalar and vector free energy calculations[J]. *J Chem Phys Lett*: 2008, 128: 144120.

[238] Darve E, Pohorille A. Calculating free energies using average force[J]. *J Chem Phys Lett*: 2001, 115: 9169-9183.

[239] Grubmüller H. Predicting slow structural transitions in macromolecular systems: Conformational flooding[J]. *Phys Rev E*. 1995, 52: 2893-2906.

[240] Wriggers W, Schulten K. Investigating a Back Door Mechanism of Actin Phosphate Release by Steered Molecular Dynamics[J]. *Bioinformatics* 1999, 35: 262-273.

[241] Bussi G, Gervasio F L, Laio A, et al. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics[J]. *J Am Chem Soc*: 2006, 128: 13435-13441.

[242] Spiwok V, Oborsky P, Pazurikova J, et al. Nonlinear vs. linear biasing in Trp-cage folding simulations [J]. *J Chem Phys*: 2015, 142: 115101.

[243] Clark A J, Tiwary P, Borrelli K, et al. Prediction of Protein-Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations [J]. *J Chem Theory Comput*:

2016, 12: 2990-2998.

- [244] Lelimosin M, Limongelli V, Sansom M S. Conformational Changes in the Epidermal Growth Factor Receptor: Role of the Transmembrane Domain Investigated by Coarse-Grained MetaDynamics Free Energy Calculations [J]. *J Am Chem Soc*: 2016, 138: 10611-10622.
- [245] Gouron A, Milet A, Jamet H. Conformational flexibility of human casein kinase catalytic subunit explored by metadynamics[J]. *Biophys J*: 2014, 106: 1134-1141.
- [246] Spiwok V, Lipovová P, Králová B. Metadynamics in Essential Coordinates: Free Energy Simulation of Conformational Changes[J]. *J Phys Chem B*: 2007, 111: 3073-3076.
- [247] Barducci A, Bussi G, Parrinello M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method[J]. *Phys Rev Lett*: 2008, 100: 020603.
- [248] Halder S, Kuhrova P, Banas P, et al. Insights into Stability and Folding of GNRA and UNCG Tetraloops Revealed by Microsecond Molecular Dynamics and Well-Tempered Metadynamics [J]. *J Chem Theory Comput*: 2015, 11: 3866-3877.
- [249] Ludovico S, Marco D, Gervasio. F L. Comparing the Efficiency of Biased and Unbiased Molecular Dynamics in Reconstructing the Free Energy Landscape of Met-Enkephalin[J]. *J Chem Theory Comput*: 2010, 6: 3640-3646.
- [250] Spiwok V, Králová B, Tvaroska I. Continuous metadynamics in essential coordinates as a tool for free energy modelling of conformational changes[J]. *J Mol Model*: 2008, 14: 995-1002.
- [251] Amadei A, Linssen A B, Berendsen H J. Essential dynamics of proteins[J]. *Proteins*: 1993, 17: 412-425.
- [252] Petrone P, Pande V S. Can conformational change be described by only a few normal modes?[J]. *Biophys J*: 2006, 90: 1583-1593.
- [253] Zheng W, Brooks B R. Normal-modes-based prediction of protein conformational changes guided by distance constraints[J]. *Biophys J*: 2005, 88: 3109-3117.
- [254] Ahmed A, Rippmann F, Barnickel G, et al. A normal mode-based geometric simulation approach for exploring biologically relevant conformational transitions in proteins[J]. *J Chem Inf Model*: 2011, 51: 1604-1622.
- [255] Vashisth H, Brooks L. Conformational Sampling of Maltose-transporter Components in Cartesian Collective Variables is Governed by the Low-frequency Normal Modes [J]. *J Phys Chem Lett*: 2012, 3: 3379-3384.
- [256] Samson A O, Levitt M. Normal modes of prion proteins: from native to infectious particle[J]. *Biochemistry*: 2011, 50: 2243-2248.
- [257] Kitao A, Go N. Investigating protein dynamics in collective coordinate space[J]. *Curr Opin Struct Biol*: 199, 9: 164-169.
- [258] Costa M G, Batista P R, Bisch P M, et al. Exploring free energy landscapes of large conformational changes: molecular dynamics with excited normal modes[J]. *J Chem Theory Comput*: 2015, 11: 2755-2767.
- [259] Batista P R, Pandey G, Pascutti P G, et al. Free Energy Profiles along Consensus Normal

- Modes Provide Insight into HIV-1 Protease Flap Opening[J]. *J Chem Theory Comput*: 2011, 7: 2348-2352.
- [260] Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding[J]. *Chem Phys Lett*: 1999, 314: 141-151.
- [261] Bernardi R C, Melo M C R, Schulten K. Enhanced sampling techniques in molecular dynamics simulations of biological systems[J]. *Biochim Biophys Acta*: 2015, 1850: 872-877.
- [262] Yan C, Pattani V, Tunnell J W, et al. Temperature-induced unfolding of epidermal growth factor (EGF): Insight from molecular dynamics simulation [J]. *J Mol Graphics Model*: 2010, 29: 2-12.
- [263] Bernardi R C, Melo M C R, Schulten K. Enhanced sampling techniques in molecular dynamics simulations of biological systems[J]. *Biochim Biophys Acta*: 2015, 1850: 872-877.
- [264] Yang Y I, Shao Q, Zhang J, et al. Enhanced sampling in molecular dynamics[J]. *J Chem Phys*: 2019, 151: 070902.
- [265] Okur A, Wickstrom L, Layten M, et al. Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model[J]. *J Chem Theory Comput*: 2006, 2: 420-433.
- [266] Wang J, Peng C, Yu Y, et al. Exploring Conformational Change of Adenylate Kinase by Replica Exchange Molecular Dynamic Simulation [J]. *Biophys J*: 2020, 118: 1009-1018.
- [267] Liu P, Kim B, Friesner R A, et al. Replica exchange with solute tempering: a method for sampling biological systems in explicit water [J]. *Proc Natl Acad Sci USA* 2005, 102: 13749-13754.
- [268] Wang L, Friesner R A, Berne B J. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2) [J]. *J Phys Chem B*: 2011, 115: 9431-9438.
- [269] Sugita Y, Kitao A, Okamoto Y. Multidimensional replica-exchange method for free-energy calculations [J]. *J Phys Chem*: 2000, 113: 6042-6051.
- [270] Ito S, Fedorov D G, Okamoto Y, et al. Implementation of replica-exchange umbrella sampling in GAMESS [J]. *Comput Phys Commun*: 2018, 228: 152-162.
- [271] Bergonzo C, Henriksen N M, Roe D R, et al. Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide[J]. *J Chem Theory Comput*: 2014, 10: 492-499.
- [272] Moradi M, Babin V, Roland C, et al. Are long-range structural correlations behind the aggregation phenomena of polyglutamine diseases?[J]. *PLoS Comput Biol*: 2012, 8: e1002501.
- [273] Han M, Xu J, Ren Y, et al. Simulation of coupled folding and binding of an intrinsically disordered protein in explicit solvent with metadynamics[J]. *J Mol Graph Model*: 2016, 68: 114-127.
- [274] Bernetti M, Masetti M, Pietrucci F, et al. Structural and Kinetic Characterization of the Intrinsically Disordered Protein SeV NTAIL through Enhanced Sampling Simulations[J]. *J Phys Chem B*: 2017, 121: 9572-9582.

- [275] Lovera S, Sutto L, Boubeva R, et al. The Different Flexibility of c-Src and c-Abl Kinases Regulates the Accessibility of a Druggable Inactive Conformation[J]. *J Am Chem Soc*: 2012, 134: 2496-2499.
- [276] Kuzmanic A, Sutto L, Saladino G, et al. Changes in the free-energy landscape of p38alpha MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations[J]. *Elife*: 2017, 6.
- [277] Barducci A, Bonomi M, Parrinello M. Metadynamics[J]. *WIREs Comput Mol Sci*: 2011, 1: 826-843.
- [278] Marino K A, Sutto L, Gervasio F L. The effect of a widespread cancer-causing mutation on the inactive to active dynamics of the B-Raf kinase[J]. *J Am Chem Soc*: 2015, 137: 5280-5283.
- [279] Lopez-Blanco J R, Garzon J I, Chacon P. iMod: multipurpose normal mode analysis in internal coordinates [J]. *Bioinformatics* 2011, 27: 2843-2850.
- [280] Krebs W G, Alexandrov V, Wilson C A, et al. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic[J]. *Proteins*: 2002, 48: 682-695.
- [281] Wang J, Shao Q, Xu Z, et al. Exploring transition pathway and free-energy profile of large-scale protein conformational change by combining normal mode analysis and umbrella sampling molecular dynamics[J]. *J Phys Chem B*: 2014, 118: 134-143.
- [282] Zhou C Y, Jiang F, Wu Y D. Folding Thermodynamics and Mechanism of Five Trp-Cage Variants from Replica-Exchange MD Simulations with RSFF2 Force Field[J]. *J Chem Theory Comput*: 2015, 11: 5473-5480.
- [283] Wang J, Wolf R M, Caldwell J W, et al. Development and Testing of a General Amber Force Field[J]. *J Comput Chem*: 2004, 25: 1157-1174.
- [284] Ryckaert J P, Ciccotti G, Berendsen H J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes[J]. *J Comput Phys*: 1977, 23: 327-341.
- [285] Darden T, York D, Pedersen L. Particle mesh Ewald: An N method for Ewald sums in large systems[J]. *J Comput Phys*: 1993, 98: 10089.
- [286] Tribello G A, Bonomi M, Branduardi D, et al. PLUMED 2: New feathers for an old bird[J]. *Comput Phys Commun*: 2014, 185: 604-613.
- [287] Dzeja P, Terzic A. Adenylate kinase and AMP signaling networks: metabolic monitoring, signal communication and body energy sensing[J]. *Int J Mol Sci*: 2009, 10: 1729-1772.
- [288] Li D, Liu M S, Ji B. Mapping the Dynamics Landscape of Conformational Transitions in Enzyme: The Adenylate Kinase Case [J]. *Biophys J*: 2015, 109: 647-660.
- [289] Snow C, Qi G, Hayward S. Essential dynamics sampling study of adenylate kinase: comparison to citrate synthase and implication for the hinge and shear mechanisms of domain motions [J]. *Proteins*: 2007, 67: 325-337.
- [290] Schulz G E. Structure of the Complex Between Adenylate Kinase from Escherichia coli

and the Inhibitor Ap5A Refined at 1.9 Å Resolution. A model for a catalytic transition state[J]. *J Mol Biol*: 1992, 224: 159-177.

[291] Muller C W, Schlauderer G J, Reinstein J, et al. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding[J]. *Structure*: 1996, 4: 147-156.

[292] Shao Q. Enhanced conformational sampling technique provides an energy landscape view of large-scale protein conformational transitions[J]. *Phys Chem Chem Phys*: 2016, 18: 29170-29182.

[293] Matsunaga Y, Komuro Y, Kobayashi C, et al. Dimensionality of Collective Variables for Describing Conformational Changes of a Multi-Domain Protein[J]. *J Phys Chem Lett*: 2016, 7: 1446-1451.

[294] Whitford P C, Gosavi S, Onuchic J N. Conformational transitions in adenylate kinase. Allosteric communication reduces misligation [J]. *J Biol Chem*: 2008, 283: 2042-2048.

[295] Henzler-Wildman K A, Thai V, Lei M, et al. Intrinsic motions along an enzymatic reaction trajectory[J]. *Nature*: 2007, 450: 838-844.

[296] Beckstein O, Denning E J, Perilla J R, et al. Zipping and unzipping of adenylate kinase: atomistic insights into the ensemble of open<->closed transitions[J]. *J Mol Biol*: 2009, 394: 160-176.

[297] Marcos-Alcalde I, Setoain J, Mendieta-Moreno J I, et al. MEPSA: minimum energy pathway analysis for energy landscapes[J]. *Bioinformatics*: 2015, 31: 3853-3855.

[298] Arora K, Brooks 3rd C L. Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism[J]. *Proc Natl Acad Sci USA*: 2007, 104: 18496-18501.

[299] Zeller F, Zacharias M. Substrate Binding Specifically Modulates Domain Arrangements in Adenylate Kinase [J]. *Biophys J*: 2015, 109: 1978-1985.

[300] Brokaw J B, Chu J W. On the roles of substrate binding and hinge unfolding in conformational changes of adenylate kinase[J]. *Biophys J*: 2010, 99: 3420-3429.

[301] Huang Y M, Raymundo M A, Chen W, et al. Mechanism of the Association Pathways for a Pair of Fast and Slow Binding Ligands of HIV-1 Protease[J]. *Biochemistry*: 2017, 56: 1311-1323.

[302] Van der Kuyl A C, Bakker M, Jurriaans S, et al. Translational HIV-1 research: from routine diagnostics to new virology insights in Amsterdam, the Netherlands during 1983-2013[J]. *Retrovirology* 2013, 10: 93.

[303] Freedberg D I, Ishima R, Jacob J, et al. Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations[J]. *Protein Sci*: 2002, 11: 221-232.

[304] Katoh E, Louis J M, Yamazaki T, et al. A solution NMR study of the binding kinetics and the internal dynamics of an HIV-1 protease-substrate complex[J]. *Protein Sci*: 2003, 12: 1376-1385.

[305] Martin P, Vickrey J F, Proteasa G, et al. "Wide-open" 1.3 Å structure of a multidrug-resistant HIV-1 protease as a drug target [J]. *Structure*: 2005, 13: 1887-1895.

[306] King N M, Prabu-Jeyabalan M, Bandaranayake R M, et al. Extreme entropy-enthalpy

- compensation in a drug-resistant variant of HIV-1 protease[J]. ACS Chem Biol: 2012, 7: 1536-1546.
- [307] Spinelli S, Liu Q Z, Alzari P M, et al. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate B RU[J]. Biochemistry: 1991, 73: 1391-1396.
- [308] Yu Y, Wang J, Chen Z, et al. Structural insights into HIV-1 protease flap opening processes and key intermediates[J]. RSC Adv: 2017, 7: 45121-45128.
- [309] Roche J, Louis J M, Bax A. Conformation of inhibitor-free HIV-1 protease derived from NMR spectroscopy in a weakly oriented solution[J]. Chembiochem: 2015, 16: 214-218.
- [310] Deng N J, Zheng W, Gallicchio E, et al. Insights into the dynamics of HIV-1 protease: a kinetic network model constructed from atomistic simulations [J]. J Am Chem Soc: 2011, 133: 9387-9394.
- [311] Ishima R F, Wang Y X, Louis J M, et al. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function[J]. Structure: 199, 7: 1047-1055.
- [312] Butini S, Brogi S, Novellino E, et al. The structural evolution of β -secretase inhibitors: a focus on the development of small-molecule inhibitors [J]. Curr Opin Struct Biol: 2013, 13: 1787-1807.
- [313] Gorfe A A, Caflisch A. Functional plasticity in the substrate binding site of beta-secretase[J]. Structure: 2005, 13: 1487-1498.
- [314] Shimizu H, Tosaki A, Kaneko K, et al. Crystal structure of an active form of BACE1, an enzyme responsible for amyloid beta protein production [J]. Mol Cell Biol: 2008, 28: 3663-3671.
- [315] Iserloh U, Wu Y, Cumming J N, et al. Potent pyrrolidine- and piperidine-based BACE-1 inhibitors [J]. Bioorg Med Chem Lett: 2008, 18: 414-417.
- [316] Charrier N, Clarke B, Demont E, et al. Second generation of BACE-1 inhibitors part 2: Optimisation of the non-prime side substituent [J]. Bioorg Med Chem Lett: 2009, 19: 3669-3673.
- [317] Fobare W F, Solvibile W R, Robichaud A J, et al. Thiophene substituted acylguanidines as BACE1 inhibitors[J]. Bioorg Med Chem Lett: 2007, 17: 5353-5356.
- [318] Ellis C R, Shen J. pH-dependent population shift regulates BACE1 activity and inhibition[J]. J Am Chem Soc: 2015, 137: 9543-9546.
- [319] Tozzini V, Trylska J, Chang C E, et al. Flap opening dynamics in HIV-1 protease explored with a coarse-grained model[J]. J Struct Biol: 2007, 157: 606-615.
- [320] Heaslet H, Rosenfeld R, Giffin M, et al. Conformational flexibility in the flap domains of ligand-free HIV protease[J]. Acta Crystallographica Section D: 2007, 63: 866-875.
- [321] Hornak V, Okur A, Rizzo R C, et al. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations[J]. Proc Natl Acad Sci USA: 2006, 103: 915-920.
- [322] Tresadern G, Delgado F, Delgado O, et al. Rational design and synthesis of aminopiperazinones as beta-secretase (BACE) inhibitors[J]. Bioorg Med Chem Lett: 2011, 21: 7255-7260.
- [323] Efremov I V, Vajdos F F, Borzilleri K A, et al. Discovery and optimization of a novel

- spiropyrrolidine inhibitor of beta-secretase (BACE1) through fragment-based drug design[J]. *J Med Chem*: 2012, 55: 9069-9088.
- [324] Congreve M, Aharony D, Albert J, et al. Application of fragment screening by X-ray crystallography to the discovery of aminopyridines as inhibitors of β -secretase[J]. *J Med Chem*: 2007, 50: 1124-1132.
- [325] Wallace A C, Laskowski R A, Thornton J M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions[J]. *Protein Eng*: 1995, 8: 127-134.
- [326] Shan Y, Seeliger M A, Eastwood M P, et al. A conserved protonation-dependent switch controls drug binding in the Abl kinase[J]. *Proc Natl Acad Sci USA*: 2009, 106: 139-144.
- [327] Foda Z H, Shan Y, Kim E T, et al. A dynamically coupled allosteric network underlies binding cooperativity in Src kinase[J]. *Nat Commun*: 2015, 6: 5939.
- [328] Cuadrado A, Nebreda A R. Mechanisms and functions of p38 MAPK signalling[J]. *Biochem J*: 2010, 429: 403-417.
- [329] Chang C I X B, Akella R, Cobb M, et al. Crystal structures of MAP kinase p38 complexed to the docking sites on its nuclear substrate MEF2A and activator MKK3b [J]. *Mol Cell Biol*: 2002, 9: 1241-1249.
- [330] Shao Q, Xu Z, Wang J, et al. Energetics and structural characterization of the "DFG-flip" conformational transition of B-RAF kinase: a SITS molecular dynamics study[J]. *Phys Chem Chem Phys*: 2017, 19: 1257-1267.
- [331] Berteotti A, Cavalli A, Branduardi D, et al. Protein conformational transitions: the closure mechanism of a kinase explored by atomistic simulations[J]. *J Am Chem Soc*: 2008, 131: 244-250.
- [332] Gan W, Yang S, Roux B. Atomistic view of the conformational activation of Src kinase using the string method with swarms-of-trajectories[J]. *Biophys J*: 2009, 97: L8-L10.
- [333] Meng Y, Lin Y, Roux B. Computational study of the "DFG-flip" conformational transition in c-Abl and c-Src tyrosine kinases[J]. *J Phys Chem B*: 2015, 119: 1443-1456.
- [334] Xuhui H, Morten H, Byungchan K, et al. Replica Exchange with Solute Tempering: Efficiency in Large Scale Systems[J]. *J Phys Chem B*: 2007, 111: 5405-5410.
- [335] Jiang F, Han W, Wu Y-D. The Intrinsic Conformational Features of Amino Acids from a Protein Coil Library and Their Applications in Force Field Development[J]. *Phys Chem Chem Phys*: 2013, 15: 3413-3428.
- [336] Grdadolnik J, Mohacek-Grosov V, Baldwin R L, et al. Populations of the Three Major Backbone Conformations in 19 Amino Acid Dipeptides [J]. *Proc Natl Acad Sci USA*: 2010, 108: 1794-1798.
- [337] Toal S, Amidi O, Schweitzer-Stenner R. Conformational Changes of Trialanine Induced by Direct Interactions between Alanine Residues and Alcohols in Binary Mixtures of Water with Glycerol and Ethanol[J]. *J Am Chem Soc*: 2011, 133: 12728-12739.
- [338] Giordanetto F, Kihlberg J. Macrocyclic drugs and clinical candidates: what can medicinal chemists learn from their properties?[J]. *J Med Chem*: 2014, 57: 278-295.

- [339] Villar E A, Beglov D, Chennamadhavuni S, et al. How proteins bind macrocycles[J]. *Nat Chem Biol*: 2014, 10: 723-731.
- [340] Doak B C, Zheng J, Dobritsch D, et al. How Beyond Rule of 5 Drugs and Clinical Candidates Bind to Their Targets[J]. *J Med Chem*: 2016, 59: 2312-2327.
- [341] Lipinski C, Dominy B W, Feeney P J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings[J]. *Adv Drug Delivery Rev*: 2001, 46: 3-26.
- [342] Mallinson J C, Macrocycles in new drug discovery[J]. *Fut Med Chem*: 2012, 4: 1409-1438.
- [343] Driggers E M, Hale S P, Lee J, et al. The exploration of macrocycles for drug discovery-an underexploited structural class[J]. *Nat Rev Drug Discov*: 2008, 7: 608-624.
- [344] Bogdan A R, Davies N L, James K. Comparison of diffusion coefficients for matched pairs of macrocyclic and linear molecules over a drug-like molecular weight range[J]. *Org Biomol Chem*: 2011, 9: 7727-7733.
- [345] Rezai T, Yu B, Millhauser G L, et al. Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers[J]. *J Ame Chem Soc*: 2006, 128: 2510-2511.
- [346] Ganesan A. The impact of natural products upon modern drug discovery[J]. *Curr Opin Chem Biol*: 2008, 12: 306-317.
- [347] Abdelraheem E, Shaabani S, Domling A. Macrocycles: MCR synthesis and applications in drug discovery[J]. *Drug Discov Today Technol*: 2018, 29: 11-17.
- [348] Song Z J, Tellers D M, Journet M, et al. Synthesis of vaniprevir (MK-7009): lactamization to prepare a 20-membered [corrected] macrocycle[J]. *J Org Chem*: 2011, 76: 7804-7815.
- [349] Jiang Y, Andrews S W, Condroski K R, et al. Discovery of danoprevir (ITMN-191/R7227), a highly selective and potent inhibitor of hepatitis C virus (HCV) NS3/4A protease[J]. *J Med Chem*: 2014, 57: 1753-1769.
- [350] Johnson T W, Richardson P F, Bailey S, et al. Discovery of (10R)-7-amino-12-fluoro-2,10,16-trimethyl-15-oxo-10,15,16,17-tetrahydro-2H-8,4-(metheno)pyrazolo[4,3-h][2,5,11]-benzoxadiazacyclotetradecine-3-carbonitrile (PF-06463922), a macrocyclic inhibitor of anaplastic lymphoma kinase (ALK) and c-ros oncogene 1 (ROS1) with preclinical brain exposure and broad-spectrum potency against ALK-resistant mutations[J]. *J Med Chem*: 2014, 57: 4720-4744.
- [351] Marsault E, Peterson M L. Macrocycles are great cycles: applications, opportunities, and challenges of synthetic macrocycles in drug discovery[J]. *J Med Chem*: 2011, 54: 1961-2004.
- [352] Arkin M R, Tang Y, Wells J A. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality[J]. *Chem Biol*: 2014, 21: 1102-1114.
- [353] Scott D E, Bayly A R, Abell C, et al. Small molecules, big targets: drug discovery faces the protein-protein interaction challenge[J]. *Nat Rev Drug Discov*: 2016, 15: 533-550.
- [354] Estrada-Ortiz N, Neochoritis C G, Twarda-Clapa A, et al. Artificial Macrocycles as Potent

- p53-MDM2 Inhibitors[J]. ACS Med Chem Lett: 2017, 8: 1025-1030.
- [355] Ashkenazi A, Fairbrother W J, Levenson J D, et al. From basic apoptosis discoveries to advanced selective BCL-2 family inhibitors[J]. Nat Rev Drug Discov: 2017, 16: 273-284.
- [356] Wang W, Groves M R, Domling A. Artificial Macrocycles as IL-17A/IL-17RA Antagonists[J]. Medchemcomm: 2018, 9: 22-26.
- [357] Danelius E, Pettersson M, Bred M, et al. Flexibility is important for inhibition of the MDM2/p53 protein-protein interaction by cyclic beta-hairpins[J]. Org Biomol Chem: 2016, 14: 10386-10393.
- [358] Scott D E. NMR Determination of the Bioactive Conformation of Peloruside A Bound To Microtubules[J]. ACS Med Chem Lett: 2015, 10:25-30.
- [359] Roland M, Julien F, Giinter B, et al. The 3D structure of a cyclosporin analogue in water is nearly identical to the cyclophilin-bound cyclosporin conformation[J]. FEBS Lett: 1994, 340: 255-259.
- [360] Danièle A, Olivier V, Bernard R, et al. A Conformation of Cyclosporin A in Aqueous Environment Revealed by the X-ray Structure of a Cyclosporin-Fab Complex[J]. Science: 1992, 256: 92-94.
- [361] Máté E, Bernhard P, Kurt H, et al. Conformational Preferences of Natural and C3-Modified Epothilones in Aqueous Solution[J]. J Med Chem: 2008, 51: 1469-1473.
- [362] Máté E. Conformational Preferences of Natural and C3-Modified Epothilones in Aqueous Solution[J]. J Med Chem:2008, 51:1469-1473
- [363] Rossi M, Doak B C, Backlund M, et al. Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs Beyond the Rule of 5[J]. J Med Chem: 2018, 61: 4189-4202.
- [364] Jan K. Structural and conformational determinants of macrocycle cell permeability[J]. Nat Chem Biol: 2016, 12: 1065-1074.
- [365] Whitty A, Zhong M, Viarengo L, et al. Quantifying the chameleonic properties of macrocycles and other high-molecular-weight drugs[J]. Drug Discov Today: 2016, 21: 712-717.
- [366] Vistoli G, Pedretti A, Testa B. Assessing drug-likeness--what are we missing?[J]. Drug Discov Today: 2008, 13: 285-294.
- [367] Hawkins P, Skillman A, Warren G L, Ellingson B A, Stahl M T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database[J]. J Chem Inf Model: 2010, 50: 572-684.
- [368] Watts K S, Dalal P, Tebben A J, et al. Macrocycle conformational sampling with MacroModel[J]. J Chem Inf Model: 2014, 54: 2680-2696.
- [369] Coutsiar E A, Lexa K W, Wester M J, et al. Exhaustive Conformational Sampling of Complex Fused Ring Macrocycles Using Inverse Kinematics[J]. J Chem Theory Comput: 2016, 12: 4674-4687.
- [370] Wang Q, Sciabola S, Barreiro G, et al. Dihedral Angle-Based Sampling of Natural Product

- Polyketide Conformations: Application to Permeability Prediction[J]. *J Chem Inf Model*: 2016, 56: 2194-2206.
- [371] Alogheli H, Olanders G, Schaal W, et al. Docking of Macrocycles: Comparing Rigid and Flexible Docking in Glide[J]. *J Chem Inf Model*: 2017, 57: 190-202.
- [372] Sindhikara D, Spronk S A, Day T, et al. Improving Accuracy, Diversity, and Speed with Prime Macrocycle Conformational Sampling[J]. *J Chem Inf Model*: 2017, 57: 1881-1894.
- [373] Labute P. LowModeMD-Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocycles and Protein Loops[J]. *J Chem Inf Model*: 2010, 50: 792-800.
- [374] Kamenik A S, Lessel U, Fuchs J E, et al. Peptidic Macrocycles-Conformational Sampling and Thermodynamic Characterization[J]. *J Chem Inf Model*: 2018, 58: 982-992.
- [375] Shkurti A, Styliari I D, Balasubramanian V, et al. CoCo-MD: A Simple and Effective Method for the Enhanced Sampling of Conformational Space[J]. *J Chem Theory Comput*: 2019, 15: 2587-2596.
- [376] Poongavanam V, Danelius E, Peintner S, et al. Conformational Sampling of Macrocyclic Drugs in Different Environments: Can We Find the Relevant Conformations?[J]. *ACS Omega*: 2018, 3: 11742-11757.
- [377] Martin M. Molecular dynamics simulations of biomolecules[J]. *Nat Struct Biol*: 2002, 9: 646-652.
- [378] Khuntawee W, Rungrotmongkol T, Wolschann P, et al. Conformation study of varepsilon-cyclodextrin: Replica exchange molecular dynamics simulations[J]. *Carbohydr Polym*: 2016, 141: 99-105.
- [379] Geng H, Jiang F, Wu Y D. Accurate Structure Prediction and Conformational Analysis of Cyclic Peptides with Residue-Specific Force Fields[J]. *J Phys Chem Lett*: 2016, 7: 1805-1810.
- [380] Zou H Y, Li Q, Engstrom L D, et al. PF-06463922 is a potent and selective next-generation ROS1/ALK inhibitor capable of blocking crizotinib-resistant ROS1 mutations[J]. *Proc Natl Acad Sci USA*: 2015, 112: 3493-3498.
- [381] Basit S, Ashraf Z, Lee K, et al. First macrocyclic 3(rd)-generation ALK inhibitor for treatment of ALK/ROS1 cancer: Clinical and designing strategy update of lorlatinib[J]. *Eur J Med Chem*: 2017, 134: 348-356.
- [382] Elleraas J, Ewanicki J, Johnson T W, et al. Conformational Studies and Atropisomerism Kinetics of the ALK Clinical Candidate Lorlatinib (PF-06463922) and Desmethyl Congeners[J]. *Angew Chem Int Ed Engl*: 2016, 55: 3590-3595.
- [383] Shaw A T, Friboulet L, Leshchiner I, et al. Resensitization to Crizotinib by the Lorlatinib ALK Resistance Mutation L1198F[J]. *N Engl J Med*: 2016, 374: 54-61.
- [384] Erdélyi M, Danelius E, Brath U. Insight into β -Hairpin Stability: Interstrand Hydrogen Bonding[J]. *Synlett*: 2013, 24: 2407-2410.
- [385] Pahk T, Tomasso E, Daniel O. Cicero, et al. Relationship Among Ligand Conformations

- in Solution, in the Solid State, and at the Hsp90 Binding Site: Geldanamycin and Radicicol[J]. 2007, 21: 221-240..
- [386] Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019[J]. N Engl J Med: 2020, 382: 727-733.
- [387] Zhou P, Yang X L, Wang X G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin[J]. Nature: 2020, 579: 270-273.
- [388] Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China[J]. Nature: 2020, 579: 265-269.
- [389] Chan J F-W, Yuan S, Kok K-H, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster[J]. The Lancet: 2020, 395: 514-523.
- [390] Brotherton H, Usuf E, Nadjm B, et al. Dexamethasone for COVID-19: data needed from randomised clinical trials in Africa[J]. The Lancet Global Health: 2020, 8: e1125-e1126.
- [391] Selvaraj V, Dapaah-Afriyie K, Finn A, et al. Short-Term Dexamethasone in Sars-CoV-2 Patients[J]. RI Med J 2020, 103: 39-43.
- [392] Wang Y, Zhang D, Du G, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial[J]. The Lancet: 2020, 395: 1569-1578.
- [393] Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor[J]. Cell: 2020, 181: 271-280 e278.
- [394] Song W, Gui M, Wang X, et al. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2[J]. PLoS Pathog: 2018, 14: e1007236.
- [395] Daniel W, Nianshuang W, Kizzmekia S. Corbett, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation[J]. Science: 2020, 367: 1260–1263.
- [396] Henderson R, Edwards R J, Mansouri K, et al. Controlling the SARS-CoV-2 Spike Glycoprotein Conformation[J]. bioRxiv: 2020.
- [397] Barnes C O, West A P, Jr., Huey-Tubman K E, et al. Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies[J]. Cell: 2020, 367: 1260-1263.
- [398] Hsieh C L, Goldsmith J A, Schaub J M, et al. Structure-based Design of Prefusion-stabilized SARS-CoV-2 Spikes[J]. bioRxiv: 2020.
- [399] Lv Z, Deng Y Q, Ye Q. Structural basis for neutralization of SARS-CoV-2 and SARS-CoV by a potent therapeutic antibody[J]. Science: 2020, 367: 1260–1263.
- [400] Wrobel A G, Benton D J, Xu P, et al. Evolution of SARS-CoV-2 spike glycoprotein[J]. DOI: 10.21203/rs.3.rs-29398/v1: 2020.
- [401] Cao Y, Su B, Guo X, et al. Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells[J]. Cell: 2020, 182: 73-84.

- [402] Chi X, Yan R. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2[J]. *Science*: 2020, 367: 1260–1263.
- [403] He J, Tao H, Yan Y, et al. Molecular Mechanism of Evolution and Human Infection with SARS-CoV-2[J]. *Viruses*: 2020, 12.
- [404] Adhikari P, Li N, Shin M, et al. Intra- and intermolecular atomic-scale interactions in the receptor binding domain of SARS-CoV-2 spike protein: implication for ACE2 receptor binding[J]. *Phys Chem* 2020, 367: 1260–1263.
- [405] Amin M, Sorour M, Kasry A. Comparing the Binding Interactions in the Receptor Binding Domains of SARS-CoV-2 and SARS-CoV[J]. *J Phys Chem Lett*: 2020, 11: 4897-4900.
- [406] Nguyen H, Lan P, Thai N, et al. Does SARS-CoV-2 Bind to Human ACE2 More Strongly Than Does SARS-CoV?[J]. *J Phys Chem B*: 2020, 124: 7336-7347.
- [407] Spinello A, Saltalamacchia A, Magistrato A. Is the Rigidity of SARS-CoV-2 Spike Receptor-Binding Motif the Hallmark for Its Enhanced Infectivity? Insights from All-Atom Simulations[J]. *J Phys Chem Lett*: 2020, 11: 4785-4790.
- [408] Wang Y, Liu M, Gao J. Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions[J]. *Proc Natl Acad Sci U S A*: 2020, 117: 13967-13974.
- [409] Shang J, Wan Y, Luo C, et al. Cell entry mechanisms of SARS-CoV-2[J]. *Proc Natl Acad Sci U S A*: 2020, 117: 11727-11734.
- [410] Walls A C, Park Y J, Tortorici M A, et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein[J]. *Cell*: 2020, 181: 281-292 e286.
- [411] Fang L, Wenhui L, Michael F, et al. Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor[J]. *Science*: 2005, 309: 1864-1868.
- [412] Kirchdoerfer R N, Wang N, Pallesen J, et al. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis[J]. *Sci Rep*: 2018, 8: 15701.
- [413] Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor[J]. *Nature*: 2020, 581: 215-220.
- [414] Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes[J]. *Nucleic Acids Res*: 2018, 46: W296-W303.
- [415] Duan Y, Wu C, Chowdhury S, et al. A Point-charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations [J]. *J Comput Chem*: 2003, 24: 1999-2012.
- [416] Ryckaert J-P, Ciccotti G, Berendsen H. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes[J]. *J Comput Phys*: 1977, 23: 327-341.
- [417] Essmann U P, Berkowitz L.; Darden T.; Lee H.; Pedersen L. G. A smooth particle mesh Ewald method[J]. *J Comput Phys*: 1995, 103: 8577-8593.
- [418] Onufriev A, Bashford D, Case D A. Exploring protein native states and large-scale

conformational changes with a modified generalized born model[J]. *Proteins-Structure Function and Bioinformatics*: 2004, 55: 383-394.

[419] Chen F, Liu H, Sun H, et al. Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking[J]. *Phys Chem Chem Phys*: 2016, 18: 22129-22139.

[420] Wang J, Shao Q, Xu Z, et al. Exploring transition pathway and free-energy profile of large-scale protein conformational change by combining normal mode analysis and umbrella sampling molecular dynamics[J]. *J Phys Chem B*: 2014, 118: 134-143.

[421] Yuan Y, Cao D, Zhang Y, et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains[J]. *Nat Commun*: 2017, 8: 15092.

[422] Gungor O, Edward F. V, Stephen Q, et al. Adaptive Steered Molecular Dynamics of the Long-Distance Unfolding of Neuropeptide Y[J]. *J Chem Theory Comput*: 2010, 6: 3026-3038.

[423] Ghorbani M, Brooks B R, Klauda J B. Critical Sequence Hot-spots for Binding of nCoV-2019 to ACE2 as Evaluated by Molecular Simulations[J]. *bioRxiv*: 2020.

[424] Lai H T T, Nguyen L H, Kranjc A, et al. Elucidating the differences in the molecular mechanism of receptor binding between 2019-nCoV and the SARS-CoV viruses using computational tools[J]. *bioRxiv*: 2020.

[425] Ke Z, Oton J, Qu K, et al. Structures and distributions of SARS-CoV-2 spike proteins on intact virions[J]. *Nature*: 2020.

[426] Christine T, Kapil G, Sathish K, et al. Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein[J]. *science*: 2020, 370: 725-730.

[427] Chen Z, Zhang X, Peng C, et al. D3Pockets: A Method and Web Server for Systematic Analysis of Protein Pocket Dynamics[J]. *J Chem Inf Model*: 2019, 59: 3353-3358.

[428] Kao R Y, Tsui W H, Lee T S, et al. Identification of novel small-molecule inhibitors of severe acute respiratory syndrome-associated coronavirus by chemical genetics[J]. *Chem Biol*: 2004, 11: 1293-1299.

[429] Yi L, Li Z, Yuan K, et al. Small molecules blocking the entry of severe acute respiratory syndrome coronavirus into host cells[J]. *J Virol*: 2004, 78: 11334-11339.

[430] Wang X, Cao R, Zhang H, et al. The anti-influenza virus drug, arbidol is an efficient inhibitor of SARS-CoV-2 in vitro[J]. *Cell Discov*: 2020, 6: 28.

[431] Zhao G, Du L, Ma C, et al. A safe and convenient pseudovirus-based inhibition assay to detect neutralizing antibodies and screen for viral entry inhibitors against the novel human coronavirus MERS-CoV[J]. *Virol J*: 2013, 10: 266.

[432] Kim J Y, Kim Y I, Park S J, et al. Safe, high-throughput screening of natural compounds of MERS-CoV entry inhibitors using a pseudovirus expressing MERS-CoV spike protein[J]. *Int J Antimicrob Agents*: 2018, 52: 730-732.

[433] Adedeji A O, Severson W, Jonsson C, et al. Novel inhibitors of severe acute respiratory syndrome coronavirus entry that act by three distinct mechanisms[J]. *J Virol*: 2013, 87: 8017-8028.

- [434] Ghorbani M, Brooks B R, Klauda J B. Exploring dynamics and network analysis of spike glycoprotein of SARS-COV-2[J]. bioRxiv: 2020.
- [435] Casalino L, Gaieb Z, Goldsmith J A, et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein[J]. ACS Cent Sci: 2020, 6: 1722-1734.
- [436] Karplus M, McCammon J A. Molecular dynamics simulations of biomolecules[J]. Nat Struct Mol Biol: 2002, 9: 646-652.
- [437] Berendsen H J C. Collective protein dynamics in relation to function[J]. Curr Opin Struct Biol: 2000, 10: 165-169.
- [438] Venkatraman M, Alan C, Maxwell D, et al. Docking: Successes and Challenges[J]. Curr Pharm Des: 2005, 11: 323-333.
- [439] Huang S Y, Zou X. Advances and challenges in protein-ligand docking[J]. Int J Mol Sci: 2010, 11: 3016-3034.
- [440] Rousse A, Rischel C. Femtosecond x-ray crystallography[J]. Reviews of Modern Physics: 2001, 73: 17.
- [441] Bennett W S, Huber R, Engel J. Structural and Functional Aspects of Domain Motions in Proteins[J]. Critical Rev Biochem: 2008, 15: 291-384.
- [442] Skjaerven L, Hollup S M, Reuter N. Normal mode analysis for proteins[J]. J Mol Struct: 2009, 898: 42-48.
- [443] Karplus M, McCammon J A, L. P W. The internal dynamics of globular protein[J]. Critical Rev Biochem: 1981, 9: 293-349.
- [444] Careri G, Fasella P. Statistical time events in enzymes: a physical assessment[J]. Critical Rev Biochem: 1975, 3: 141-164.
- [445] Gurd F, Rothgeb T M. Motions in proteins[J]. Adv. Prot. Chem.: 1979, 33.
- [446] Andedro C. Conformational fluctuation and change in biological macromolecules[J]. Science Progress: 1980, 66: 473-497.
- [447] Konrad H, Aline T, Field M. Analysis of Domain Motions in Large Proteins[J]. Proteins:1999, 34: 369-382.
- [448] Gerstein M, Anderson B F, Norris G E, et al. Domain closure in lactoferrin[J]. J. Mol. Biol.: 1993, 234: 357-372.
- [449] Gerstein M, Lesk A M, Chothia C. Structural mechanisms for domain movements in proteins[J]. Biochemistry: 1994, 33: 6739-6749.
- [450] Nishikawa K, Ooi T, Isogai Y, et al. Representation and computation of the conformations[J]. J Phys Soc Jpn: 1972, 32: 1331-1337.
- [451] Qi G, Lee R, Hayward S. A comprehensive and non-redundant database of protein domain movements[J]. Bioinformatics: 2005, 21: 2832-2838.
- [452] Echols N. MolMovDB: analysis and visualization of conformational change and structural flexibility[J]. Nucleic Acids Res: 2003, 31: 478-482.
- [453] Piovesan D, Tabaro F, Paladin L, et al. MobiDB 3.0: more annotations for intrinsic

- disorder, conformational diversity and interactions in proteins[J]. *Nucleic acids Res*: 2018, 46: D471-D476.
- [454] Flores S, Echols N, Milburn D, et al. The Database of Macromolecular Motions: new features added at the decade mark[J]. *Nucleic acids Res*: 2006, 34: D296-301.
- [455] Gelin B R. Sidechain Torsional Potentials and Motion of Amino Acids in Proteins: Bovine Pancreatic Trypsin Inhibitor[J]. *Proc Nat Acad Sci*: 1975, 72: 2002-2006.
- [456] Gutteridge A, Thornton J. Conformational changes observed in enzyme crystal structures upon substrate binding[J]. *J Mol Biol*: 2005, 346: 21-28.
- [457] Noble M A, Miles C S, Chapman S K, et al. Roles of key active-site residues in flavocytochrome P450 BM3[J]. *Biochemical J*: 1999, 339: 371-379.
- [458] Keizers P, Lussenburg B M, de Graaf C, et al. Influence of phenylalanine 120 on cytochrome P450 2D6 catalytic selectivity and regiospecificity: crucial role in 7-methoxy-4-(aminomethyl)-coumarin metabolism[J]. *Biochem Pharmacol*: 2004, 68: 2263-2271.
- [459] Yongxia G, Zuyu Z, James J. La Clair, et al. Smoke-derived karrikin perception by the α/β -hydrolase KAI2 from *Arabidopsis*[J]. *Proc Natl Acad Sci USA*: 2013, 110: 8284-8289.
- [460] Yunjun W, Jardetzky. O. Investigation of the Neighboring Residue Effects on Protein Chemical Shifts[J]. *J Am Chem Soc*: 2002, 124: 14075-14084.
- [461] Bernstein F C, Koetzle T F, Williams G J B, et al. The Protein Data Bank: A computer - based archival file for macromolecular structures[J]. *Eur J Biochem*: 1977, 80: 319-324.
- [462] Lobanov M Y, Shoemaker B A, Garbuzynskiy S O, et al. ComSin: database of protein structures in bound (complex) and unbound (single) states in relation to their intrinsic disorder[J]. *Nucleic Acids Res*: 2010, 38: D283-287.
- [463] Chang D T, Yao T J, Fan C Y, et al. AH-DB: collecting protein structure pairs before and after binding[J]. *Nucleic Acids Res*: 2012, 40: D472-478.
- [464] Hrabe T, Li Z, Sedova M, et al. PDBFlex: exploring flexibility in protein structures[J]. *Nucleic Acids Res*: 2016, 44: D423-428.
- [465] Amemiya T, Koike R, Kidera A, et al. PSCDB: a database for protein structural change upon ligand binding[J]. *Nucleic Acids Res*: 2012, 40: D554-558.
- [466] Teague S J. Implications of protein flexibility for drug discovery[J]. *Nat Rev Drug Discov*: 2003, 2: 527-541.
- [467] Tokuriki N. Protein dynamism and evolvability[J]. *Science*: 2009, 324: 203-207.
- [468] Juritz E I, Alberti S F, Parisi G D. PCDB: a database of protein conformational diversity[J]. *Nucleic Acids Res*: 2011, 39: D475-479.
- [469] Monzon A M, Rohr C O, Fornasari M S, et al. CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state[J]. *Database (Oxford)*: 2016, 2016.
- [470] Chang C W, Chou C W, Chang D T. CCProf: exploring conformational change profile of proteins[J]. *Database (Oxford)*: 2016, 2016.
- [471] Paul N, Rognan D. ConsDock: A new program for the consensus analysis of protein-

ligand interactions[J]. *Proteins*: 2002, 47: 521-533.

[472] Trott O, Olson A J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading[J]. *J Comput Chem*: 2010, 31: 455-461.

[473] Emrick M A, Lee T, Starkey P J, et al. The gatekeeper residue controls autoactivation of ERK2 via a pathway of intramolecular connectivity[J]. *Proc Natl Acad Sci USA*: 2006, 103: 18101-18106.

[474] de Leon-Boenig G, Bowman K K, Feng J A, et al. The crystal structure of the catalytic domain of the NF-kappaB inducing kinase reveals a narrow but flexible active site[J]. *Structure*: 2012, 20: 1704-1714.

[475] Lefurgy S T, Malashkevich V N, Aguilan J T, et al. Analysis of the Structure and Function of FOX-4 Cephamycinase[J]. *Antimicrob Agents Chemother*: 2016, 60: 717-728.

[476] Wishart D S, Feunang Y D, Guo A C, et al. DrugBank 5.0: a major update to the DrugBank database for 2018[J]. *Nucleic Acids Res*: 2018, 46: D1074-D1082.

[477] Borrel A, Regad L, Xhaard H, et al. PockDrug: A Model for Predicting Pocket Druggability That Overcomes Pocket Estimation Uncertainties[J]. *J Chem Inf Model*: 2015, 55: 882-895.

[478] Herrera N G, Morano N C, Celikgil A, et al. Characterization of the SARS-CoV-2 S Protein: Biophysical, Biochemical, Structural, and Antigenic Analysis[J]. *bioRxiv*: 2020.

[479] Pinto D, Park Y J, Beltramello M, et al. Structural and functional analysis of a potent sarbecovirus neutralizing antibody[J]. *bioRxiv*: 2020.

[480] Liu L, Wang P, Nair M S, et al. Potent Neutralizing Monoclonal Antibodies Directed to Multiple Epitopes on the SARS-CoV-2 Spike[J]. *bioRxiv*: 2020.

[481] Zhou T, Tsybovsky Y, Olia A S, et al. A pH-dependent switch mediates conformational masking of SARS-CoV-2 spike[J]. *bioRxiv*: 2020.

[482] Cai Y F, Zhang J, Xiao T S, et al. Distinct conformational states of SARS-CoV-2 spike protein[J]. *Science*: 2020, 11: 233-256.

[483] Yurkovetskiy L, Pascal K E, Tompkins-Tinch C, et al. SARS-CoV-2 Spike protein variant D614G increases infectivity and retains sensitivity to antibodies that target the receptor binding domain[J]. *bioRxiv*: 2020.

[484] Huo J, Zhao Y, Ren J, et al. Neutralization of SARS-CoV-2 by Destruction of the Prefusion Spike[J]. *Cell Host Microbe*: 2020, 57:1211-1224.

[485] Melero R, Sorzano C, Foster B, et al. Continuous flexibility analysis of SARS-CoV-2 Spike prefusion structures[J]. *bioRxiv*: 2020.

[486] Xiong X, Qu K, Ciazynska K A, et al. A thermostable, closed SARS-CoV-2 spike protein trimer[J]. *Nat Struct Mol Biol*: 2020,43:23-44.

[487] Gui M, Song W, Zhou H, et al. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding[J]. *Cell Res*: 2017,

27: 119-129.

[488] Walls A C, Xiong X, Park Y J, et al. Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion[J]. Cell: 2019, 176: 1026-1039 e1015.

[489] Yasunori W, Joel D, Daniel W, et al. Site-specific glycan analysis of the SARS-CoV-2 spike[J]. Science: 2020, 369: 330-333.

[490] Watanabe Y, Berndsen Z T, Raghvani J, et al. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation[J]. Nat Commun: 2020, 11: 2688-2702.

附 录

表 S3.1 模拟副本温度

Table S3.1 Temperatures of simulation replicas

Method	Protein	Temperatures (K)							
ossPTMetaD	Adk	300.0	305.3	310.8	316.3	321.9	327.6	333.5	339.4
		345.4	351.6	357.8	364.2	370.7	377.3	384.0	390.8
		397.8	404.8	412	419.4	426.8	434.4	442.1	450.0
T-REMD	Adk	300.0	301.5	303.1	304.6	306.2	307.8	309.4	311.0
		312.6	314.2	315.8	317.4	319.1	320.7	322.4	324.0
		325.7	327.4	329.0	330.7	332.4	334.1	335.9	337.6
		339.3	341.1	342.8	344.6	346.4	348.2	349.9	351.7
		353.6	355.4	357.2	359.0	360.9	362.7	364.6	366.5
		368.4	370.3	372.2	374.1	376.0	377.9	379.9	381.8
		383.8	385.8	387.8	389.8	391.8	393.8	395.8	397.8
		399.9	402.0	404.0	406.1	408.2	410.3	412.4	414.5
		416.7	418.8	421.0	423.1	425.3	427.5	429.7	431.9
ossPTMetaD	HIV-1 PR	434.1	436.4	438.6	440.9	443.1	445.4	447.7	450.0
		300.0	308.2	316.7	325.3	334.3	343.4	352.8	362.5
ossPTMetaD	BACE1	372.4	382.6	393.1	403.9	415.0	426.3	438.0	450.0
		300.0	304.2	308.5	312.8	317.2	321.7	326.2	330.8
		335.5	340.2	345.0	349.9	354.8	359.8	364.9	370.0
		375.2	380.5	385.8	391.3	396.8	402.4	408.0	413.8
ossPTMetaD	p38 α	419.6	425.5	431.5	437.6	443.8	450.0		
		300.0	304.5	309.2	313.8	318.6	323.4	328.3	333.2
		338.3	343.4	348.6	353.9	359.2	364.7	370.2	375.8
		381.5	387.2	393.1	399.1	405.1	411.2	417.4	423.8
ossPTMetaD	c-Abl	430.2	436.7	443.3	450.0				
		300.0	304.5	309.2	313.8	318.6	323.4	328.3	333.2
		338.3	343.4	348.6	353.9	359.2	364.7	370.2	375.8
		381.5	387.2	393.1	399.1	405.1	411.2	417.4	423.8
T-REMD	Ala	430.2	436.7	443.3	450.0				
		300.0	308.2	316.7	325.3	334.3	343.4	352.8	362.5
ossPTMetaD	Ala	372.4	382.6	393.1	403.9	415.0	426.3	438.0	450.0
		300.0	317.9	336.8	356.9	378.2	400.8	424.7	450.0

表 S3.2 P38 α 晶体结构的 I141-F169 (Å)和 V38-F169 (Å)距离。Table S3.2 The I141-F169 (Å) and V38-F169 (Å) in crystal structures of P38 α .

PDB ID	I141-F169	V38-F169	PDB ID	I141-F169	V38-F169	PDB ID	I141-F169	V38-F169
2ONL	7.44	16.50	1W84	8.07	16.99	3HVC	7.30	16.06
2LGC	6.70	16.51	2ZB0	8.07	16.20	3NEW	7.45	16.10
3HV5	16.07	9.11	1WBO	8.09	17.10	4E6C	7.49	17.10
3HV4	16.02	9.16	3FC1	8.13	16.09	3FML	7.63	16.62
3GCU	15.90	6.86	2ZAZ	8.14	16.22	4E6A	7.63	17.46
2OKR	8.00	16.50	1A9U	8.16	17.08	4R3C	7.65	16.38
3DS6	6.53	16.70	2Y8O	8.16	16.76	4E5B	7.67	17.02
5ETA	7.90	15.78	1DI9	8.19	15.89	3MGY	7.69	16.06
3MVM	6.48	16.49	1YQJ	8.19	17.01	3ODZ	7.70	15.27
4KIQ	6.35	15.90	1W7H	8.21	17.23	4E5A	7.70	17.63
4KIN	6.36	15.81	5ETF	8.21	16.72	3ZYA	7.73	16.55
4KIP	6.60	15.77	1M7Q	8.23	16.09	1BMK	7.76	17.00
3MVL	6.73	15.81	5ETC	8.23	17.23	4F9W	7.77	16.90
5TBE	6.10	16.08	5XYX	8.23	15.97	3C5U	7.79	15.30
3OCG	6.38	16.13	1OUY	8.27	15.84	2ZB1	7.80	15.83
2RG6	6.45	16.41	1OUK	8.29	15.92	3FMJ	7.80	16.26
3UVQ	6.45	16.35	1WFC	8.31	17.00	3ODY	7.81	14.89
3S4Q	6.49	15.61	3MW1	8.35	16.48	5N68	7.81	16.42
3FKL	6.50	15.98	5XYX	8.35	15.48	1BL6	7.82	16.99
2RG5	6.52	16.03	3FLQ	8.37	15.80	3FSK	7.84	16.37
2YIX	6.54	16.49	1OZ1	8.38	16.22	3HA8	7.86	17.23
3MH0	6.62	17.45	3OD6	8.33	14.33	1BL7	7.89	16.46
3IPH	6.63	16.16	1IAN	10.44	16.05	5N67	7.90	16.50
5MTX	6.64	15.96	4FA2	11.85	13.01	2GFS	7.91	16.22
2QD9	6.65	16.11	4ZTH	11.93	13.21	1WBW	7.96	16.32
3D7Z	6.67	16.01	3OEF	12.47	11.43	1R3C	7.97	16.60
3E92	6.67	16.49	2BAQ	13.53	11.05	3MPA	7.98	15.77
3L8X	6.67	15.25	4EH7	13.54	10.67	3FSF	7.99	16.90
3NWW	6.67	16.02	3IW6	13.59	11.79	1R39	8.07	17.20
5TCO	6.68	15.80	3HP2	13.69	11.28	1WBT	17.85	8.02
3DT1	6.71	16.50	3IW5	13.72	12.93	1WBS	17.86	7.84
4AA4	6.71	16.37	3IW7	13.76	11.74	4L8M	16.07	7.40
3FLN	6.72	15.97	3QUD	13.82	11.27	3IW8	16.09	10.68
3RIN	6.72	16.15	3ZSI	13.83	11.34	3OBJ	16.11	9.36
4AAC	6.73	16.62	3QUE	13.88	11.35	1KV1	16.12	8.97
1ZYJ	6.75	16.06	5OMG	13.91	11.61	3NNX	16.16	7.08
3S3I	6.76	16.39	4DLI	13.97	11.37	3NNV	16.19	7.56
5MTY	6.76	15.88	3MH1	13.99	11.80	1W83	16.23	8.07
1ZZ2	6.77	15.76	4EH4	14.08	11.25	3LFC	16.23	8.39
3HUB	6.78	16.95	6ANL	14.08	11.27	3K3I	16.34	8.68
3KF7	6.78	16.43	3O8P	14.18	11.39	3MH3	16.35	7.70
3ROC	6.78	16.14	3ZSG	14.27	10.56	3HV3	16.37	9.77
3U8W	6.78	16.73	4EH3	14.32	10.52	3GCQ	16.39	6.97

4AA5	6.78	16.36	5WJJ	14.33	10.33	3K3J	16.39	9.15
5ETI	6.78	18.40	3ZS5	14.43	10.61	1WBV	16.40	9.48
3FLY	6.80	16.23	5ML5	14.44	10.36	3KQ7	16.40	9.16
3UVP	6.84	15.74	3GCP	14.49	10.86	3D83	16.45	9.28
3FLZ	6.85	16.47	3HUC	14.57	10.41	3GI3	16.45	8.47
4EWQ	6.85	16.35	3BV2	15.07	10.69	2BAK	16.58	7.32
1ZZL	6.86	16.38	5N64	15.10	10.29	3HV7	16.58	7.87
2I0H	6.86	17.22	2FSM	15.15	10.77	4A9Y	16.58	7.32
3ITZ	6.86	16.60	4EH8	15.15	11.01	3HV6	16.65	8.56
4F9Y	6.86	16.73	2NPQ	15.25	10.98	1KV2	16.70	8.17
3FMH	6.87	16.19	3BV3	15.34	10.63	3NNW	16.71	7.44
3HLL	6.87	16.07	3OC1	15.54	10.99	3HEC	16.77	8.05
3LHJ	6.87	16.24	3O8T	15.57	9.70	3UVR	16.79	6.73
3FMK	6.88	16.14	4GEO	15.60	10.34	4AA0	16.94	7.23
3OBG	6.89	16.99	3GCS	15.62	10.14	2YIW	16.96	7.33
3FLW	6.90	16.33	4EH9	15.63	9.10	5N66	16.96	8.31
3FLS	6.91	16.18	4EHV	15.64	8.93	3NNU	17.11	7.46
3FMM	6.93	16.14	5MZ3	15.66	8.68	2YIS	17.34	7.66
3MH2	6.93	16.93	3LFD	15.69	11.65	3HRB	7.07	16.43
2BAL	6.94	16.08	4EH6	15.76	9.27	3MPT	7.07	16.47
3FKO	6.97	16.37	3PG3	15.78	10.59	1OVE	7.11	16.40
3FMN	6.97	16.01	3LFE	15.83	9.64	3FKN	7.12	15.98
3GFE	6.99	16.24	3HEG	15.87	10.01	3HP5	7.12	16.39
4EH2	7.00	16.56	3LFF	15.87	7.09	3GC7	7.19	17.09
3FL4	7.02	16.17	3LFB	15.89	7.28	3BX5	7.26	15.60
3FI4	7.06	15.97	3E93	15.91	8.16	2BAJ	16.03	6.33
1WBN	16.01	9.40	3CTQ	15.93	8.07	1W82	16.04	7.61
3O8U	16.02	7.35	3L8S	15.98	10.81	3GCV	16.04	

表 S3.3 c-Abl 晶体结构的 L354-F382 (Å)和 V256-F382 (Å)距离。

Table S3.3 The L354-F382 (Å) and V256-F382 (Å) in crystal structures of c-Abl.

PDB ID	L354-F382	V256-F382
2G1T	6.36	15.26
2V7A	6.41	14.94
2GQG	6.43	14.92
2F4J	6.46	14.53
2G2I	6.46	14.29
2HZ4	6.48	13.70
4TWP	6.50	13.73
2FO0	12.12	10.48
1OPL	12.33	10.64
2HZI	12.68	10.08
4YC8	12.94	11.00
3UE4	12.39	11.60
6B18	12.54	10.51
4ZOG	12.64	10.05
2G2F	12.85	10.65

4WA9	12.91	10.96
2G2H	13.57	10.08
2E2B	15.09	7.72
2HYY	15.00	7.54
2HIW	15.01	7.48
3QRJ	15.21	6.86
3QRI	15.48	6.89
3QRK	15.43	6.45
6NPE	15.57	7.12
6NPU	15.57	7.10
6NPV	15.54	7.08
5MO4	15.60	6.89
3PYY	15.70	7.24
5HU9	15.77	6.64
2HZ0	16.28	5.82

表 S4.1 CoV-2-S 与 CoV-S 在不同 ϵ_{in} 下的计算 ACE2 结合自由能Table S4.1 The predicted ACE2-binding free energy (kcal/mol) for CoV-S and CoV-2-S trimers based on three different ϵ_{in} .^a

PDB	CoV-S			CoV-2-S		
	1	2	4	1	2	4
6ACG	-10.59±0.62	-22.70±0.88	-31.49±0.97	-18.00±0.84	-24.47±0.86	-33.95±0.99
6ACJ	-12.39±0.68	-23.65±0.95	-30.28±0.82	-25.10±0.68	-35.21±0.82	-43.22±1.13
6CS2	-11.54±0.72	-22.37±1.01	-32.04±0.90	-32.40±0.86	-39.57±0.58	-45.10±0.75
6ACK	-20.63±0.60	-37.15±0.95	-45.48±0.97	-39.52±0.61	-54.01±1.02	-57.77±1.43

^a: 统计误差根据 50-100 ns MD 模拟轨迹的计算结果获得。在 MM/GBSA 计算中, 从 50-100 ns MD 模拟轨迹中均一提取 1000 个构象计算结合自由能。

表 S4.2 PDB 数据库中的 CoV-2-S 与 CoV-S 总结

Table S4.2 Summary of the CoV-S and CoV-2-S trimers in PDB.

PDB ID	Resolution (Å)	Organism	Chain	ACE2	RBD state ^a	RBD-angle (°) ^b
6X6P ^[478]	3.2	CoV-2-S	A	-	down	30.9
			B	-	down	30.2
			C	-	down	31.2
6VSB ^[395]	3.5	CoV-2-S	A	-	up	-
			B	-	down	-
			C	-	-down	-
6VXX ^[410]	2.8	CoV-2-S	A	-	down	-
			B	-	down	-
			C	-	down	-
6VYB ^[410]	3.2	CoV-2-S	A	-	down	-
			B	-	up	-

6WPS ^[479]	3.1	CoV-2-S	C	-	down	-
			A	-	down	-
			B	-	down	-
6WPT ^[479]	3.7	CoV-2-S	C	-	down	-
			A	-	down	-
			B	-	up	-
6X29 ^[396]	2.7	CoV-2-S	C	-	-	-
			A	-	down	-
			B	-	down	-
6X2A ^[396]	3.3	CoV-2-S	C	-	down	-
			A	-	down	-
			B	-	up	-
6X2B ^[396]	3.6	CoV-2-S	C	-	down	-
			A	-	up	-
			B	-	up	-
6X2C ^[396]	3.2	CoV-2-S	C	-	up	-
			A	-	down	-
			B	-	down	-
6XCM ^[397]	3.4	CoV-2-S	C	-	down	-
			A	-	down	-
			B	-	up	-
6XCN ^[397]	3.7	CoV-2-S	C	-	up	-
			A	-	up	-
			B	-	up	-
6XEY ^[480]	3.2	CoV-2-S	C	-	up	-
			A	-	down	-
			B	-	down	-
6XKL ^[398]	3.2	CoV-2-S	C	-	down	-
			A	-	up	-
			B	-	down	-
6XM5 ^[481]	3.1	CoV-2-S	C	-	down	-
			A	-	down	31.1
			B	-	down	30.7
6XR8 ^[482]	2.9	CoV-2-S	C	-	down	31.0
			A	-	down	-
			B	-	down	-
6XS6 ^[483]	3.7	CoV-2-S	C	-	down	-
			A	-	-	-
			B	-	-	-
6Z43 ^[399]	3.3	CoV-2-S	C	-	-	-
			A	-	up	-
			B	-	up	-
6Z97 ^[484]	3.4	CoV-2-S	C	-	up	-
			A	-	down	-
			B	-	up	-
6ZDH ^[399]	3.7	CoV-2-S	C	-	down	-
			A	-	up	-
			B	-	up	-
			C	-	up	-

6ZGE ^[400]	2.6	CoV-2-S	A	-	down	-
			B	-	down	-
			C	-	down	-
6ZGG ^[400]	3.8	CoV-2-S	A	-	down	-
			B	-	up	-
			C	-	down	-
6ZGH ^[400]	6.8	CoV-2-S	A	-	-	-
			B	-	down	-
			C	-	down	-
6ZGI ^[400]	2.9	CoV-2-S	A	-	down	-
			B	-	down	-
			C	-	down	-
6ZHD ^[400]	3.7	CoV-2-S	A	-	up	-
			B	-	up	-
			C	-	up	-
6ZOW ^[485]	3.0	CoV-2-S	A	-	-	-
			B	-	-	-
			C	-	-	-
6ZOX ^[485]	3.0	CoV-2-S	A	-	down	-
			B	-	down	-
			C	-	down	-
6ZOY ^[485]	3.1	CoV-2-S	A	-	down	-
			B	-	down	-
			C	-	down	-
6ZOZ ^[485]	3.5	CoV-2-S	A	-	down	-
			B	-	down	-
			C	-	down	-
6ZP0 ^[486]	3.0	CoV-2-S	A	-	down	-
			B	-	down	-
			C	-	down	-
6ZP1 ^[486]	3.3	CoV-2-S	A	-	down	-
			B	-	down	-
			C	-	down	-
6ZP2 ^[486]	3.1	CoV-2-S	A	-	down	-
			B	-	down	-
			C	-	down	-
6ZP5 ^[486]	3.1	CoV-2-S	A	-	up	-
			B	-	down	-
			C	-	down	-
6ZP7 ^[486]	3.3	CoV-2-S	A	-	up	-
			B	-	down	-
			C	-	down	-
7BYR ^[401]	3.8	CoV-2-S	A	-	up	-
			B	-	up	-
			C	-	up	-
7C2L ^[402]	3.1	CoV-2-S	A	-	up	-
			B	-	-	-
			C	-	-	-
5WRG ^[487]	4.3	CoV-S	A	-	down	30.1

			B	-	down	30.1
			C	-	down	30.1
5X58 ^[421]	3.2	CoV-S	A	-	down	31.6
			B	-	down	31.6
			C	-	down	30.7
5X5B ^[421]	3.7	CoV-S	A	-	up	84.8
			B	-	down	30.9
			C	-	down	30.9
5XLR ^[487]	3.8	CoV-S	A	-	down	32.1
			B	-	down	32.1
			C	-	down	32.1
6ACC ^[394]	3.6	CoV-S	A	-	down	33.4
			B	-	down	33.4
			C	-	down	33.4
6ACD ^[394]	3.9	CoV-S	A	-	down	32.8
		-S	B	-	down	32.8
			C	-	down	32.9
6ACG ^[394]	5.4	CoV-S	A	-	down	32.6
			B	-	down	32.7
			C	ACE2	up	54.8
6ACJ ^[394]	4.2	CoV-S	A	-	down	33.0
			B	-	down	33.3
			C	ACE2	up	68.3
6ACK ^[394]	4.5	CoV-S	A	-	down	33.1
			B	-	down	33.8
			C	ACE2	up	84.6
6CRV ^[412]	3.2	CoV-S	A	-	-	-
			B	-	-	-
			C	-	-	-
6CRW ^[412]	3.9	CoV-S	A	-	down	34.3
			B	-	up	68.8
			C	-	down	34.2
6CRX ^[412]	3.9	CoV-S	A	-	up	71.6
			B	-	up	70.6
			C	-	down	38.1
6CRZ ^[412]	3.3	CoV-S	A	-	down	34.1
			B	-	up	68.8
			C	-	down	34.1
6CS0 ^[412]	3.8	CoV-S	A	-	down	34.2
			B	-	up	68.8
			C	-	down	34.1
6CS1 ^[412]	4.6	CoV-S	A	-	up	71.6
			B	-	up	70.7
			C	-	down	38.1
6CS2 ^[412]	4.4	CoV-S	A	-	-	-
			B	ACE2	up	74.0
			C	-	-	-
6NB6 ^[488]	4.2	CoV-S	A	-	down	30.7
			B	-	up	77.9

6NB7 ^[488]	4.5	CoV-S	C	-	up	55.2
			A	-	up	75.3
			B	-	up	70.6
			C	-	up	78.5

^a: RBD 开放/闭合构象根据对应文献的描述判定。

^b: CoV-S 体系 RBD 角度由残基 D392-T608-V973 计算, CoV-2-S 由残基 D405-V622-V991 计算。

表 S4.3 CoV-2-S 糖基化实验数据^[435, 489]。

Table S4.3 The glycosylation of CoV-2-S according to experimental data.

Residues	Type	Glycan
N61A	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N61
N74A	A3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6)[bDGlcNAc(1→2)]aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N74
N122A	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N122
N149A	FA2 G2S 1	aDNeu5Ac(2→6)bDGlcNAc(1→2)aDMan(1→6) [bDGlcNAc(1→2)]aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N149
N165A	FA2 G2S 2	xaDNeu5Ac(2→6)bDGlcNAc(1→2)aDMan(1→6) [aDNeu5Ac(2→6)bDGlcNAc(1→2)]aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N165
N234A	M8	aDMan(1→2)aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) bDGlcNAc(1→)N234
N282A	FA3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)]aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)N282
N331A	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)]aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N331
N343A	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)]aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N343
N603A	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)]aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N603
N616A	A2	bDGlcNAc(1→2)aDMan(1→6) [bDGlcNAc(1→2)]aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N616
N657A	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N657
N709A	M6	aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N709
N717A	Hydrid G1	bDGlcNAc(1→2)aDMan(1→3)[aDMan(1→6) [aDMan(1→3)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N717
N801A	M6	aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N801

N1074 A	FA2 G2S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3) [bDGal(1→4)bDGlcNAc(1→2)aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N1074
N1098 A	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N1098
N61B	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N61
N74B	FA3 G3S 2	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→6)[bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6) [aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N74
N122B	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N122
N149B	FA3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)N149
N165B	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N165
N234B	M9	aDMan(1→2)aDMan(1→6)[aDMan(1→2)aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N234
N282B	FA3 G3S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4)bDGlcNAc(1→6) [bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N282
N331B	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N331
N343B	FA1	bDGlcNAc(1→2)aDMan(1→3)[aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N343
N603B	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N603
N616B	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N616
N657B	Hybrid G1	bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[aDMan(1→6) [aDMan(1→3)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N657
N709B	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N709
N717B	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N717
N801B	M7	aDMan(1→2)aDMan(1→2)aDMan(1→3)[aDMan(1→6) [aDMan(1→3)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N801
N1074 B	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N1074
N1098 B	A2	bDGlcNAc(1→2)aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N1098
N61C	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N61
N74C	A2	bDGlcNAc(1→2)aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N74
N122C	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N122
N149C	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N149

		→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N149
N165C	FA2 G2S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→6) [bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N165
N234C	M9	aDMan(1→2)aDMan(1→6)[aDMan(1→2)aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) bDGlcNAc(1→)N234
N282C	A2	bDGlcNAc(1→2)aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc (1→)N282
N331C	FA3 G3S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4) bDGlcNAc(1→6) [bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1 →4) [aLFuc(1→6)]bDGlcNAc(1→)N331
N343C	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1 →4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N343
N603C	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bD GlcNAc(1→4)bDGlcNAc(1→)N603
N616C	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1 →4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N616
N657C	Hybr id G1	bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[aDMan(1→6) [aDMan(1→3)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1 →)N657
N709C	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bD GlcNAc(1→4)bDGlcNAc(1→)N709
N717C	M6	aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1 →)N717
N801C	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bD GlcNAc(1→4)bDGlcNAc(1→)N801
N1074 C	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bD GlcNAc(1→4)bDGlcNAc(1→)N1074
N1098 C	Hybr id G1S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[aDMan(1→6)) [aDMan(1→3)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1 →)N1098

表 S4.4 CoV-S 糖基化实验数据^[490]。

Table S4.4 The glycosylation of CoV-S according to experimental data.

Residues	Type	Glycan
N29A	FA3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)N29
N65A	M6	aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N65
N73A	FA3 G3S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4) bDGlcNAc(1→6) [bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N73
N109A	FA3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)N109
N118A	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N118
N158A	M7	aDMan(1→2)aDMan(1→2)aDMan(1→3)[aDMan(1→6) [aDMan(1→3)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N158
N227A	M9	aDMan(1→2)aDMan(1→6)[aDMan(1→2)aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) bDGlcNAc(1→)N227
N269A	FA3 G3S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4) bDGlcNAc(1→6) [bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N269
N318A	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N318
N330A	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N330
N357A	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N357
N589A	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N589
N602A	A2	bDGlcNAc(1→2)aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N602
N691A	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N691
N699A	FA3 G3S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4) bDGlcNAc(1→6) [bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N699
N783A	M7	aDMan(1→2)aDMan(1→2)aDMan(1→3)[aDMan(1→6) [aDMan(1→3)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N783
N1056A	FA3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)N1056
N1080A	Hydrid G1	bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[aDMan(1→6) [aDMan(1→3)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N1080

		→)N1080
N29B	FA3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)N29
N65B	M6	aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N65
N73B	FA2 G2S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3) [bDGal(1→4)bDGlcNAc(1→2)aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N73
N109B	FA3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)N109
N118B	FA3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)N118
N158B	M6	aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N158
N227B	M9	aDMan(1→2)aDMan(1→6)[aDMan(1→2)aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N227
N269B	FA3 G3S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4)bDGlcNAc(1→6) [bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N269
N318B	FA3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)N318
N330B	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N330
N357B	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N357
N589B	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N589
N602B	A2	bDGlcNAc(1→2)aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N602
N691B	FA3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)N691
N699B	A3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N699
N783B	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N783
N1056B	FA3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)N1056
N1080B	Hydr id G1	bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[aDMan(1→6) [aDMan(1→3)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N1080
N29C	FA3 G3S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4)bDGlcNAc(1→6) [bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N29
N65C	M6	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)

		[aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N65
N73C	FA3 G3S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4)bDGlcNAc(1→6) [bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N73
N109C	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N109
N118C	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N118
N158C	M6	aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N158
N227C	M9	aDMan(1→2)aDMan(1→6)[aDMan(1→2)aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N227
N269C	FA3 G3S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4)bDGlcNAc(1→6) [bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N269
N318C	FA3 G3S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4)bDGlcNAc(1→6) [bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N318
N330C	FA3 G3S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4)bDGlcNAc(1→6) [bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N330
N357C	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N357
N589C	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N589
N602C	A3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N602
N691C	FA2	bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N691
N699C	A3	bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N699
N783C	M5	aDMan(1→6)[aDMan(1→3)]aDMan(1→6)[aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N783
N1056C	FA3 G3S 1	aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4)bDGlcNAc(1→6) [bDGal(1→4)bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)N1056
N1080C	A2	bDGlcNAc(1→2)aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)N1080

表 S4.5 劳拉替尼在 D₂O/DMSO-*d*₆ (6:4)和 CDCl₃ 溶剂中的 ¹H-NMR 谱推导。Table S4.5 ¹H-NMR assignment (ppm) of Lorlatinib in D₂O/DMSO-*d*₆ (6:4) and CDCl₃

^a Proton	δ, D ₂ O/DMSO- <i>d</i> ₆	δ, CDCl ₃
5	4.17	4.07
9'	4.47	4.36
9''	4.59	4.45
11	3.19	3.13
16	7.66	7.29
19	7.31	7.00
20	7.55	7.21
21	5.75	5.73
23	1.87	1.79
27		4.94
29	7.74	7.82
31	6.97	6.86

^a根据图 4.15 编号。表 S4.6 D₂O/DMSO-*d*₆ (6:4) (δ in ppm)中 NOE 增强测量劳拉替尼自旋间距Table S4.6 Interproton distances (Å) for Lorlatinib derived from NOE build-up measurements in D₂O/DMSO-*d*₆ (6:4) (δ in ppm).

Dis. No.	Atom type	Proton <i>i</i>	Proton <i>j</i>	δ(¹ H) <i>i</i>	δ(¹ H) <i>j</i>	σ	R ²	Dis. <i>rij</i> [Å]
1	CHCH	16	21	7.66	5.75	7.26254E-06	0.97	4.18
2	CHCH	11	20	3.19	7.55	4.98398E-06	0.99	4.45
3	CH ₂ CH	9'	20	4.47	7.55	5.30124E-05	0.99	3.00
4	CH ₂ CH	9''	20	4.59	7.55	3.83095E-05	0.99	3.17
5	CH ₂ CH	9''	31	4.59	6.97	7.20326E-06	0.98	4.19
6	CH ₂ CH	9'	31	4.47	6.97	3.23764E-05	0.99	3.26
7	CHCH ₃	31	23	6.97	1.87	3.82869E-06	0.97	4.65
8	CHCH ₃	31	11	6.97	3.19	9.93347E-06	0.99	3.97
9	CHCH ₂	31	21	6.97	5.75	0.000280931	0.99	2.27
10	CH CH ₃	21	11	5.75	3.19	5.59412E-06	0.99	4.37
11	CHCH ₂	21	9'	5.75	4.47	8.73313E-06	0.97	4.06
12	CH ₂ CH ₃	9'	11	4.47	3.19	8.72449E-06	0.99	4.06
Ref.	CHCH	19	20	7.31	7.55	0.0001564	0.96	2.51

表 S4.7 CDCl₃ (δ in ppm)中 NOE 增强测量劳拉替尼自旋间距Table S4.7 Interproton distances ($\text{\AA}$) for Lorlatinib derived from NOE build-up measurements in CDCl₃ (δ in ppm).

No.	Type	Proton <i>i</i>	Proton <i>j</i>	$\delta(^1\text{H})$ <i>i</i>	$\delta(^1\text{H})$ <i>j</i>	σ	R ²	Dis. r_{ij} [$\text{\AA}$]
1	CHCH ₃	16	23	7.29	1.79	7.17984E-05	0.99	2.47
2	CH CH ₂	20	9'	7.21	4.36	3.22055E-05	0.96	2.83
3	CHCH ₂	20	9''	7.21	4.45	4.90079E-05	0.98	2.64
4	CHCH ₃	20	11	7.21	3.13	1.11725E-06	0.97	4.95
5	CHCH ₂	31	9''	6.86	4.45	3.97831E-05	0.99	2.73
6	CHCH ₃	31	11	6.86	3.13	4.74948E-06	0.97	3.89
7	CHCH	31	21	6.86	5.73	0.000145559	0.99	2.20
8	CHCH ₃	21	23	5.73	1.79	7.97246E-05	0.98	2.43
9	CHCH ₃	9'	11	4.36	3.13	1.69377E-05	0.98	3.15
Ref	CHCH ₃	21	23	5.73	1.79	7.43545E-05	0.98	2.43

表 S4.8 蒙特-卡罗分子力学构象分析结果。

Table S4.8 Results of the Monte Carlo molecular mechanics conformational analysis.

Solvation Model	Sampling Method	Force field	Number of conformations		
			Total ^a	Within 12.6 kJ/mol ^b	Conformations added to final ensemble
CHCl₃	MCMM	OPLS-2005	5		5
	MCMM	Amber*	5		5
	MCMM	OPLS3	3		3
	MCMM	MMFF	6		6
	MCMM	OPLS	5	2	5
H₂O	MCMM	OPLS-2005	5		5
	MCMM	Amber*	5	2	5
	MCMM	OPLS3	3		3
	MCMM	MMFF	7	3	7
	MCMM	OPLS	5	2	5
Total			49		49
RCE					5 ^c
			5		
Final input ensemble for NAMFIS					5

^a 所有不同的构象数目。^b 与能量全局最低构象相差 12.55 kJ/mol (3 kcal/mol) 的构象数目。^c 根据重原子 RMSD 截断值 1 $\text{\AA}$ 消除冗余构象后得到的构象数目。

表 S4.9 劳拉替尼在 D₂O/DMSO-*d*₆ (6:4)和 CDCl₃ 溶剂中的 NAMFIS 分析。Table S4.9 Conformational populations derived by NAMFIS-analysis of Lorlatinib in D₂O/DMSO-*d*₆ (6:4) and CDCl₃ solutions.

D ₂ O		CDCl ₃	
Conformation Number	% ^a	Conformation Number	% ^a
1*	100	1*	76
		2	24

^a NAMFIS 分析的构象比例。

*在两种溶剂中发现的共同构象。

表 S4.10 NMR 构象与晶体构象的重原子 RMSD。

Table S4.10 RMSD comparison of the final conformations with the crystal structure based on the heavy atoms

RMSD _Heavy atoms		
Conf. No	Crystal structure	1
Crystal structure	0	
1	0.349	0
2	1.022	1.235

表 S5.1 结晶辅助物列表。

Table S5.1 The list of crystallographic additives.

Molecule	Structure_SMILES
GOL	<chem>OCC(O)CO</chem>
EDO	<chem>OCCO</chem>
PEG	<chem>OCCOCCO</chem>
MAN	<chem>OC[C@H]1O[C@H](O)[C@@H](O)[C@@H](O)[C@@H]1O</chem>
BMA	<chem>OC[C@H]1O[C@@H](O)[C@@H](O)[C@@H](O)[C@@H]1O</chem>
DMS	<chem>CS(C)=O</chem>
ACE	<chem>CC=O</chem>
MPD	<chem>C[C@H](O)CC(C)(C)O</chem>
MES	<chem>[O-]S(=O)(=O)CC[NH+]1CCOCC1</chem>
TRS	<chem>[NH3+]C(CO)(CO)CO</chem>
PG4	<chem>OCCOCCOCCOCCO</chem>
FUC	<chem>C[C@@H]1O[C@@H](O)[C@@H](O)[C@H](O)[C@@H]1O</chem>
GAL	<chem>OC[C@H]1O[C@@H](O)[C@H](O)[C@@H](O)[C@H]1O</chem>
EPE	<chem>OCCN1CCN(CCS(O)(=O)=O)CC1</chem>
BME	<chem>OCCS</chem>
BGC	<chem>OC[C@H]1O[C@@H](O)[C@H](O)[C@@H](O)[C@@H]1O</chem>
IMD	<chem>c1c[nH+]c[nH]1</chem>
GLC	<chem>OC[C@H]1O[C@H](O)[C@H](O)[C@@H](O)[C@@H]1O</chem>
1PE	<chem>OCCOCCOCCOCCOCCO</chem>
IPA	<chem>CC(C)O</chem>
FLC	<chem>OC(CC([O-])=O)(CC([O-])=O)C([O-])=O</chem>
MRD	<chem>C[C@@H](O)CC(C)(C)O</chem>
BOG	<chem>CCCCCCCCO[C@@H]1O[C@H](CO)[C@@H](O)[C@H](O)[C@H]1O</chem>
TLA	<chem>O[C@H]([C@@H](O)C(O)=O)C(O)=O</chem>
P6G	<chem>OCCOCCOCCOCCOCCOCCO</chem>
SCN	<chem>[S-]C#N</chem>
CMO	<chem>[C-]#[O+]</chem>
CO3	<chem>[O-]C([O-])=O</chem>
CAC	<chem>C[As](C)([O-])=O</chem>
XYP	<chem>O[C@@H]1CO[C@@H](O)[C@H](O)[C@H]1O</chem>
OLC	<chem>CCCCCCCCC=C/C(C)(C)C(=O)OC[C@H](O)CO</chem>
EOH	<chem>CCO</chem>
LMT	<chem>CCCCCCCCCCCCCO[C@@H]1O[C@H](CO)[C@@H](O[C@H]2O[C@H](CO)</chem>

	<chem>[C@@H](O)[C@H](O)[C@H]2O[C@H](O)[C@H]1O</chem>
BEN	<chem>NC(=N)c1ccccc1</chem>
LDA	<chem>CCCCCCCCCCCC[N+](C)(C)[O-]</chem>
HED	<chem>OCCSSCCO</chem>
C8E	<chem>CCCCCCCCOCCOCCOCCOCCO</chem>
CDL	<chem>CCCCCCCCCCCCCCCCCCCC(=O)OC[C@@H](CO[P@@]([O-])(=O)OC[C@H](O)CO[P@]([O-])(=O)OC[C@@H](COC(=O)CCCCCCCCCCCCCCCCC)OC(=O)CCCCCCCCCCCCCCCCC)OC(=O)CCCCCCCCCCCCCCCCC</chem>
2PE	<chem>OCCOCCOCCOCCOCCOCCOCCOCCOCCO</chem>
HEZ	<chem>OCCCCCCO</chem>
DMU	<chem>CCCCCCCCCCO[C@@H]1O[C@H](CO)[C@@H](O[C@H]2O[C@H](CO)[C@@H](O)[C@H](O)[C@H]2O)[C@H](O)[C@H]1O</chem>

表 S5.2 交叉对接打分。

Table S5.2 Docking scores of corss-docking.

Uniptot_ID	PDB_ID	Ligand	Score	Protein_another	Score	Difference
A0A024B7W1	5MRK	SFG	-8.76	5KQS	-8.57	0.19
A0A024B7W1	5KQS	SAM	-8.47	5MRK	-8.99	0.52
A0A059WZ16	5F49	MLC	-9.42	5F48	-9.37	0.05
A0A059WZ16	5F48	COA	-8.77	5F49	-9.18	0.41
A0A0H2UPS5	2GZ3	NAP	-10.24	2GZ2	-10.77	0.53
A0A0H2UPS5	2GZ2	A2P	-8.75	2GZ3	-9.25	0.50
A0A0H2WY27	5TY2	NFF	-7.81	5TW4	-7.95	0.14
A0A0H2WY27	5TW4	AI8	-9.66	5TY2	-8.43	1.23
A0A0H2Z9F5	5U8W	NAI	-10.94	5U8V	-10.93	0.01
A0A0H2Z9F5	5U8V	NAD	-9.17	5U8W	-9.47	0.3
A0A0H3LT39	3OZM	DXL	-6.23	3OP2	-6.18	0.05
A0A0H3LT39	3OP2	AKG	-6.91	3OZM	-6.25	0.66
A0A0J1HJU0	4MYX	2F0	-8.52	4MYA	-8.84	0.32
A0A0J1HJU0	4MYA	2EY	-9.13	4MYX	-8.77	0.36
A0A0J9X1Y0	4D44	JA3	-8.47	4D43	-8.05	0.42
A0A0J9X1Y0	4D43	9W7	-9.24	4D44	-9.08	0.16
A0A0R4I995	5CB4	TIV	-10.78	5XI7	-7.88	2.90
A0A0R4I995	5XI7	PO7	-8.22	5CB4	-8.63	0.41
A0A0S7ASE9	5LCD	AMP	-8.52	5LDB	-8.02	0.5
A0A0S7ASE9	5LDB	ADP	-9.06	5LCD	-9.37	0.31
A0A0X1KGP2	3X1M	COA	-9.08	3X1J	-9.38	0.30
A0A0X1KGP2	3X1J	ACO	-9.06	3X1M	-9.23	0.17
A0A0X1KHF9	5FBT	5WQ	-9.74	5FBU	-10.50	0.76
A0A0X1KHF9	5FBU	5WP	-11.30	5FBT	-9.77	1.53
A0A117IMA6	5K5C	TRE	-7.37	5L3K	-8.20	0.83
A0A117IMA6	5L3K	ADP	-8.73	5K5C	-8.37	0.36
A0A1B0UHJ4	5KCQ	GST	-7.21	5KCL	-6.67	0.54
A0A1B0UHJ4	5KCL	6C7	-6.50	5KCQ	-5.99	0.51
A0R2B1	3ZHU	TD8	-8.76	3ZHS	-7.62	1.14

A0R2B1	3ZHS	TD6	-7.57	3ZHU	-8.31	0.74
A2QHE5	6EVD	BYN	-13.39	4ZA9	-13.66	0.27
A2QHE5	4ZA9	4MJ	-14.77	6EVD	-14.12	0.65
A3FQ16	3IGO	ANP	-8.78	4Y5Q	-8.10	0.68
A3FQ16	4Y5Q	AMP	-8.03	3IGO	-7.71	0.32
A5H660	4BZ6	SHH	-6.91	4CQF	-6.61	0.30
A5H660	4CQF	9Z8	-5.25	4BZ6	-5.03	0.22
A5K1A2	2YND	646	-11.69	4CAF	-10.41	1.28
A5K1A2	4CAF	370	-10.04	2YND	-8.88	1.16
A5K1A2	5V0W	KFK	-11.38	2YND	-8.62	2.76
A5K1A2	2YND	646	-10.19	5V0W	-10.37	0.18
A5K4U6	5HN8	HXK	-7.37	3RBM	-8.12	0.75
A5K4U6	3RBM	B73	-8.42	5HN8	-7.58	0.84
A5KE01	3EWD	MCF	-8.65	2PGR	-8.60	0.05
A5KE01	2PGR	DCF	-8.91	3EWD	-7.74	1.17
A6L2E5	5L7U	NGA	-6.59	5L7V	-6.12	0.47
A6L2E5	5L7V	GNL	-7.40	5L7U	-6.94	0.46
A6ZSR0	5ESG	1YN	-11.73	4ZDY	-8.95	2.78
A6ZSR0	5HS1	VOR	-8.19	5ESG	-7.75	0.44
A6ZSR0	4ZDY	1YN	-11.87	5EAG	-10.88	0.99
A6ZSR0	5EAG	5LU	-7.00	5HS1	-7.12	0.12
A7Z2A9	5DM0	5D8	-8.97	5DLY	-8.86	0.11
A7Z2A9	5DLY	5D7	-7.39	5DM0	-7.13	0.26
A8C927	3P2K	SAM	-9.57	3P2E	-9.36	0.21
A8C927	3P2E	SAH	-9.54	3P2K	-9.56	0.02
A8DG50	4NWL	2R9	-9.25	4NWK	-9.90	0.65
A8DG50	4NWK	2R8	-10.04	4NWL	-8.64	1.40
A8DG50	3SU6	SU3	-12.28	3SU3	-7.21	5.07
A8DG50	3SV8	SV6	-6.57	3SU6	-6.98	0.41
A8DG50	3SU3	SU3	-12.57	3SUE	-7.57	5.00
A8DG50	3SUE	SUE	-12.42	3SV8	-11.04	1.38
A8PWE4	2XGT	NSS	-12.16	2XTI	-10.74	1.42
A8PWE4	2XTI	ATP	-11.05	2XGT	-11.36	0.31
A9JQL9	2ZCQ	B65	-9.69	3ACX	-9.76	0.07
A9JQL9	3ACX	673	-9.77	2ZCQ	-8.88	0.89
A9XG43	5C1U	GHV	-7.54	5DP7	-7.50	0.04
A9XG43	5DP7	5EC	-7.08	5C1U	-6.64	0.44
B2HIZ6	4B55	P18	-6.05	2VFC	-5.54	0.51
B2HIZ6	2VFC	COA	-7.13	4B55	-7.98	0.85
B2HVG8	3O83	IXN	-10.19	3U16	-9.60	0.59
B2HVG8	3U16	H89	-10.10	3O83	-9.51	0.59
B3Q7C0	4GW9	BLA	-11.65	4XTQ	-10.62	1.03
B3Q7C0	4XTQ	BL8	-10.71	4GW9	-11.24	0.53
B4Y381	4JEM	C5P	-8.86	4OHB	-8.54	0.32
B4Y381	4OHB	5HM	-8.66	4JEM	-9.54	0.88
B5XYG3	4F7W	PN4	-6.66	4GI7	-6.62	0.04
B5XYG3	4GI7	0JR	-7.81	4F7W	-7.16	0.65
B6ZK72	4RZM	LSD	-9.55	3RGA	-9.87	0.32

B6ZK72	3RGA	ILD	-10.87	4RZM	-8.42	2.45
B8J605	4KAE	TGG	-8.74	4KDY	-8.35	0.39
B8J605	4KDY	GSH	-6.60	4KAE	-7.35	0.75
C0STY1	3W79	ORI	-7.47	3W78	-7.26	0.21
C0STY1	3W78	CBD	-9.23	3W79	-10.02	0.79
C1KIK8	5ESB	SU3	-12.05	5ETX	-11.39	0.66
C1KIK8	5ETX	5RS	-9.28	5ESB	-8.74	0.54
C3W5S0	4AWK	CI1	-8.76	5WCS	-7.84	0.92
C3W5S0	5WCS	9KU	-9.62	4AWK	-9.51	0.11
C4M4W4	4DVG	GSP	-10.88	3REF	-10.74	0.14
C4M4W4	3REF	GDP	-9.97	4DVG	-6.26	3.71
C7Q5P8	5BW4	SAM	-8.73	5D1N	-9.42	0.69
C7Q5P8	5D1N	SAH	-9.32	5BW4	-8.77	0.55
C8W8H7	4R1G	CXU	-8.17	4R3J	-8.53	0.36
C8W8H7	4R3J	CFU	-7.11	4R1G	-7.03	0.08
C9DRU9	4FWH	CIX	-8.22	4FWD	-8.15	0.07
C9DRU9	4FWD	BO2	-7.69	4FWH	-7.82	0.13
D8KFM5	4N4D	XX8	-9.90	4MHW	-7.65	2.25
D8KFM5	4MHW	0YN	-10.04	4N4D	-9.64	0.40
D9J2T9	3RL9	ADE	-5.59	4FXA	-5.78	0.19
D9J2T9	4FXA	AAG	-6.12	3RL9	-5.70	0.42
E1BQ43	5KX5	ADP	-9.99	5JCB	-9.08	0.91
E1BQ43	5JCB	ACP	-9.23	5KX5	-10.37	1.14
E9QY26	4UYM	VOR	-8.14	4UYL	-7.89	0.25
E9QY26	4UYL	VNI	-10.54	4UYM	-10.43	0.11
G1C794	3UDX	IM2	-6.35	3UE0	-6.49	0.14
G1C794	3UE0	AZR	-7.42	3UDX	-7.42	0.00
G1C794	3UE1	UE1	-9.71	3UDX	-9.52	0.19
G1C794	3UDX	IM2	-6.10	3UE1	-6.43	0.33
G4V7G8	3UQ9	TBN	-9.28	3UQ6	-9.24	0.04
G4V7G8	3UQ6	ADN	-8.87	3UQ9	-8.86	0.01
G7CGT2	5K44	T6P	-8.03	5K41	-8.07	0.04
G7CGT2	5K41	ADQ	-10.31	5K44	-9.36	0.95
G8G134	6C0M	8LM	-8.95	5KV8	-9.23	0.28
G8G134	5KV8	6Y6	-7.73	6C0M	-7.02	0.71
G8NVB5	4XYB	NDP	-13.01	4XYE	-11.80	1.21
G8NVB5	4XYE	NAD	-12.02	4XYB	-12.32	0.30
I6L8L2	4B5D	SW4	-8.00	4AFH	-7.21	0.79
I6L8L2	4AFH	L0B	-8.78	4B5D	-8.09	0.69
I6L8L8	4EPP	APR	-11.80	4EPQ	-9.93	1.87
I6L8L8	4EPQ	0RR	-10.20	4EPP	-6.62	3.58
I6YFP0	4XU0	44N	-13.63	4XTU	-13.98	0.35
I6YFP0	4XTU	44J	-13.49	4XU0	-12.72	0.77
J9VFT1	5K8F	ATP	-11.15	5K8F	-9.70	1.45
J9VFT1	5K8F	6R9	-10.31	5K8F	-10.26	0.05
O00214	4HAN	NAD	-8.77	5T7I	-8.22	0.55
O00214	5T7I	LAT	-6.30	4HAN	-5.65	0.65
O00329	5M6U	7KA	-8.25	5DXU	-7.96	0.29

O00329	5DXU	5H5	-8.85	5M6U	-8.30	0.55
O06543	2GD0	MRS	-6.88	2YIM	-6.50	0.38
O06543	2YIM	MC4	-8.12	2GD0	-8.36	0.24
O06644	1T4C	COA	-10.08	1VGQ	-9.25	0.83
O06644	1VGQ	CAO	-9.27	1T4C	-7.69	1.58
O07347	2NG1	GDP	-9.51	2C04	-10.00	0.49
O07347	2C04	GCP	-10.56	2NG1	-9.62	0.94
O13833	4FHW	GTP	-8.75	4NKT	-9.33	0.58
O13833	4NKT	2KH	-8.26	4FHW	-9.35	1.09
O14744	5EMM	5QL	-11.08	5EMK	-10.81	0.27
O14744	5EMK	5QH	-11.61	5EMM	-10.90	0.71
O14965	3UP2	0C8	-9.79	3UO5	-8.81	0.98
O14965	4DHF	0K6	-9.47	3UP2	-9.63	0.16
O14965	3UO5	0BX	-9.15	3VAP	-8.98	0.17
O14965	3VAP	0FY	-10.48	4DHF	-9.28	1.20
O14965	5DN3	5DN	-8.59	4ZTR	-8.56	0.03
O14965	6HJJ	G7T	-11.38	5DN3	-10.58	0.8
O14965	5DR6	5E1	-8.61	5DR6	-7.63	0.98
O14965	4ZTR	4RJ	-11.87	6HJJ	-8.92	2.95
O15264	4EYJ	N61	-10.57	5EKN	-10.07	0.50
O15264	5EKN	N58	-10.00	4EYJ	-10.29	0.29
O15826	2YB0	DUR	-7.34	2CJE	-6.81	0.53
O15826	2CJE	DUN	-9.60	2YB0	-9.33	0.27
O15826	2YAZ	UMP	-8.84	2YB0	-8.96	0.12
O15826	2YB0	DUR	-7.65	2YAZ	-7.63	0.02
O22476	4OH4	ANP	-9.45	5LPB	-8.67	0.78
O22476	5LPB	ADP	-7.28	4OH4	-7.74	0.46
O25511	2GN9	UPG	-10.70	2GNA	-9.93	0.77
O25511	2GNA	GDU	-10.07	2GN9	-10.45	0.38
O26232	3WK3	O7E	-7.80	3NQE	-8.15	0.35
O26232	3NQE	BMP	-9.40	3WK3	-8.89	0.51
O28344	4GVQ	N4M	-10.54	4GVS	-9.19	1.35
O28344	4GVS	F4M	-8.37	4GVQ	-9.50	1.13
O30297	1Z0U	NAP	-8.93	1Z0Z	-8.55	0.38
O30297	1Z0Z	NAD	-9.27	1Z0U	-8.95	0.32
O31801	4B0H	DUR	-6.69	2XCE	-6.12	0.57
O31801	2XCE	DUP	-6.63	4B0H	-6.40	0.23
O32080	2HMT	NAI	-10.09	4J91	-9.60	0.49
O32080	4J91	ADP	-10.37	2HMT	-8.36	2.01
O32727	1TID	ATP	-9.35	1TH8	-9.51	0.16
O32727	1TH8	ADP	-9.24	1TID	-9.49	0.25
O33877	4B0I	KBP	-4.75	4B0B	-4.73	0.02
O33877	4B0B	54F	-5.36	4B0I	-6.02	0.66
O34453	4D3K	12S	-11.02	4D3K	-10.50	0.52
O34453	4UGR	6E5	-10.83	4D3K	-10.28	0.55
O34453	4UGF	Q16	-9.83	5G6F	-10.27	0.44
O34453	4D3M	IVU	-11.28	4UGR	-11.36	0.08
O34453	4D3K	12S	-10.75	4D3M	-10.99	0.24

O34453	5G6F	2SN	-9.69	4UGF	-9.68	0.01
O34757	3EHH	ADP	-10.34	5IUN	-10.33	0.01
O34757	5IUN	ACP	-10.04	3EHH	-8.54	1.50
O34919	4AO5	UMP	-5.46	2Y1T	-6.25	0.79
O34919	2Y1T	DUD	-5.90	4AO5	-5.79	0.11
O35904	5IS5	6CY	-10.87	5I6U	-10.00	0.87
O35904	6FTN	E78	-10.34	5IS5	-8.17	2.17
O35904	5I6U	68R	-10.76	6FTN	-8.85	1.91
O35904	6FTN	E78	-10.34	6FTN	-7.78	2.56
O35963	1Z06	GNP	-12.82	2G77	-12.37	0.45
O35963	2G77	GDP	-10.73	1Z06	-10.84	0.11
O43173	5BO6	CDP	-10.61	5CXY	-10.17	0.44
O43173	5CXY	55T	-11.66	5BO6	-11.15	0.51
O49686	5UR6	8KM	-8.69	3WG8	-8.79	0.10
O49686	5YGV	8V6	-9.23	5UR6	-8.77	0.46
O49686	3WG8	6AS	-8.05	5YGV	-8.27	0.22
O49686	5YGV	8V6	-9.23	5YGV	-10.32	1.09
O54438	4BO3	U98	-7.01	4BNX	-6.56	0.45
O54438	4BNX	O74	-7.50	4BO3	-7.44	0.06
O57883	2DTH	ADP	-7.08	2ZGW	-6.87	0.21
O57883	2ZGW	ADN	-6.95	2DTH	-7.22	0.27
O59521	5H7K	GDP	-10.27	5H7J	-9.21	1.06
O59521	5H7J	GCP	-8.62	5H7K	-8.64	0.02
O60218	4I5X	FLF	-9.89	5LIU	-9.33	0.56
O60218	5LIU	388	-8.88	4I5X	-9.33	0.45
O60674	4JI9	1M3	-9.78	4GFM	-9.65	0.13
O60674	5L3A	6DP	-6.46	4JI9	-6.77	0.31
O60674	5CF8	50V	-9.27	5CF8	-8.63	0.64
O60674	4GFM	0X2	-7.95	5L3A	-7.42	0.53
O60674	5WIN	SKE	-9.61	6M9H	-10.71	1.1
O60674	6M9H	J9D	-8.76	5WIN	-8.29	0.47
O60885	2YEM	WSH	-10.44	5UEU	-8.67	1.77
O60885	5UEU	0S6	-8.16	2YEM	-8.60	0.44
O67135	1C3R	TSN	-8.57	1C3S	-8.55	0.02
O67135	1C3S	SHH	-9.31	1C3R	-8.33	0.98
O67575	5TV8	APC	-9.88	5TVA	-9.48	0.4
O67575	5TVA	AMP	-9.86	5TV8	-9.14	0.72
O67648	5DRP	5EP	-9.62	3P3C	-10.14	0.52
O67648	3P3C	3P3	-10.49	5DRP	-10.14	0.35
O70370	4BSQ	QQV	-7.64	4MZS	-7.94	0.3
O70370	4BQV	8PW	-7.89	4BQV	-7.33	0.56
O70370	4BS6	JG7	-7.79	4BS6	-7.40	0.39
O70370	4MZS	2EV	-8.11	4BSQ	-7.85	0.26
O73948	3NTU	ANP	-8.51	2FPK	-8.72	0.21
O73948	2FPK	ADP	-8.42	3NTU	-8.28	0.14
O74036	4D6P	ANP	-8.29	4UQO	-8.63	0.34
O74036	4UQO	ADP	-7.31	4D6P	-8.12	0.81
O75460	4YZC	STU	-13.22	4Z7H	-6.54	6.68

O75460	4Z7H	4L5	-9.03	4YZC	-8.17	0.86
O75469	6BNS	XGH	-10.78	4NY9	-10.67	0.11
O75469	4NY9	2Q4	-9.60	6BNS	-8.96	0.64
O76290	4CMK	FQW	-9.42	3BMO	-9.13	0.29
O76290	3BMO	AX4	-6.97	4CMK	-7.04	0.07
O80992	3NJ1	PYV	-8.91	3KB3	-9.02	0.11
O80992	3KB3	A8S	-8.68	3NJ1	-7.84	0.84
O88703	5KHJ	6SY	-11.39	5KHH	-10.21	1.18
O88703	5KHH	6SW	-11.14	5KHJ	-10.55	0.59
O92972	4JU1	1NZ	-11.15	4JTZ	-11.86	0.71
O92972	4JTZ	1NW	-11.61	4JU1	-10.92	0.69
O95271	4N4V	2GY	-12.69	4MSK	-8.67	4.02
O95271	4MSK	2C8	-14.76	4N4V	-12.82	1.94
O95819	4U42	3C8	-10.57	4OBQ	-8.88	1.69
O95819	4OBQ	2QT	-10.63	4U42	-10.49	0.14
O96013	5VEF	M77	-8.25	5VEE	-7.04	1.21
O96013	5VEE	981	-8.52	5VEF	-8.62	0.1
O96935	4K5P	1OS	-8.79	4K5L	-8.01	0.78
O96935	4K5L	19N	-7.05	4K5P	-7.19	0.14
P00157	2FYU	FDN	-11.99	1SQB	-8.70	3.29
P00157	1SQB	AZO	-8.74	2FYU	-8.42	0.32
P00178	4H1N	CGE	-7.93	3UAS	-7.61	0.32
P00178	3UAS	0BV	-8.99	4H1N	-8.49	0.50
P00179	1N6B	DMZ	-7.65	1NR6	-8.32	0.67
P00179	1NR6	DIF	-7.46	1N6B	-7.27	0.19
P00349	1PGO	NDP	-8.75	1PGN	-9.07	0.32
P00349	1PGN	NBP	-9.33	1PGO	-9.69	0.36
P00374	3S7A	684	-10.53	4KAK	-10.45	0.08
P00374	4KAK	06U	-10.51	3S7A	-9.76	0.75
P00374	3FS6	DH1	-8.38	3NXT	-7.69	0.69
P00374	3NXT	D2E	-8.84	3FS6	-9.47	0.63
P00438	1PDH	FAS	-13.17	1PBF	-12.14	1.03
P00438	1PBF	FAD	-13.42	1PDH	-12.71	0.71
P00439	4PAH	LNR	-7.15	4ANP	-7.33	0.18
P00439	4ANP	3QI	-7.66	4PAH	-8.19	0.53
P00519	2HYY	STI	-12.91	2HZ0	-9.92	2.99
P00519	2HZ0	GIN	-11.82	2HYY	-10.80	1.02
P00520	3KFA	B91	-14.15	3IK3	-12.98	1.17
P00520	3IK3	0LI	-13.81	3KFA	-10.58	3.23
P00523	5D10	563	-9.84	4O2P	-9.06	0.78
P00523	4O2P	11V	-9.29	5D10	-7.66	1.63
P00636	1FBF	AHM	-7.16	1FBE	-7.72	0.56
P00636	1FBE	AHG	-7.41	1FBF	-7.21	0.2
P00653	1RDS	GPC	-10.17	1RMS	-8.06	2.11
P00653	1RMS	3GP	-8.21	1RDS	-7.69	0.52
P00720	5CGD	51E	-11.78	4OO9	-7.89	3.89
P00720	4OO9	2U8	-12.10	5CGD	-6.40	5.7
P00734	4BAQ	M4Z	-10.32	3TU7	-8.83	1.49

P00734	1T4U	81A	-9.32	1T4U	-8.89	0.43
P00734	1YPK	CCR	-10.02	1YPK	-8.69	1.33
P00734	3TU7	0BM	-9.81	4BAQ	-9.29	0.52
P00742	3HPT	YET	-10.01	2RA0	-9.39	0.62
P00742	2RA0	JNJ	-11.51	3HPT	-9.89	1.62
P00760	1Y5B	TL3	-8.53	1V2N	-8.69	0.16
P00760	1V2N	BBA	-10.30	1Y5B	-8.61	1.69
P00772	1ELE	0QN	-7.42	1BMA	-7.38	0.04
P00772	1BMA	0QH	-7.99	1ELE	-7.54	0.45
P00772	1E34	TPX	-5.83	2EST	-5.44	0.39
P00772	2EST	2Z5	-7.59	1E34	-7.71	0.12
P00797	1BIL	0IU	-10.36	1HRN	-9.90	0.46
P00797	1HRN	03D	-8.89	1BIL	-9.53	0.64
P00953	1I6K	TYM	-12.44	5DK4	-12.28	0.16
P00953	5DK4	5BX	-8.15	1I6K	-7.92	0.23
P01837	2AJX	TGN	-5.68	2AJY	-5.12	0.56
P01837	2AJY	ECG	-4.21	2AJX	-4.45	0.24
P02638	4FQO	AZ3	-5.14	5DKR	-5.45	0.31
P02638	5DKR	5CZ	-7.48	4FQO	-7.02	0.46
P02701	5MYQ	HBF	-9.54	5IRU	-9.04	0.5
P02701	5IRU	B9P	-10.72	5MYQ	-9.75	0.97
P02753	5NU7	RTL	-9.90	3FMZ	-9.03	0.87
P02753	3FMZ	2T1	-11.59	5NU7	-9.35	2.24
P02768	4L9Q	9TP	-12.15	2XW1	-10.42	1.73
P02768	3LU8	IQX	-9.27	4L9Q	-9.61	0.34
P02768	2XW1	9NV	-8.40	5YOQ	-6.18	2.22
P02768	5YOQ	CLT	-6.85	3LU8	-6.49	0.36
P02879	4HV3	19L	-9.23	4HV7	-10.07	0.84
P02879	4HV7	19J	-8.09	4HV3	-7.59	0.5
P03195	2WE1	UMP	-8.50	2WE3	-7.73	0.77
P03195	2WE3	DUT	-8.19	2WE1	-7.95	0.24
P03303	1VRH	SD8	-10.19	1R09	-9.09	1.10
P03303	1R09	JEN	-9.56	1VRH	-9.04	0.52
P03366	4RW9	3X6	-11.15	4O4G	-9.59	1.56
P03366	4O4G	2RT	-10.59	4RW9	-10.74	0.15
P03366	1NPV	L27	-8.15	1HVK	-7.32	0.83
P03366	1HVK	A79	-6.80	1NPV	-7.44	0.64
P03372	4IWC	1GV	-10.64	4IWF	-6.50	4.14
P03372	4IWF	15Q	-9.15	4IWC	-9.14	0.01
P03819	3EYW	NAD	-10.41	3L9W	-9.81	0.60
P03819	3L9W	AMP	-9.98	3EYW	-8.16	1.82
P03951	4CRC	OTM	-10.22	1ZPB	-5.85	4.37
P03951	1ZPB	995	-7.81	4CRC	-7.11	0.70
P04035	1DQ9	HMG	-6.63	3CD5	-7.35	0.72
P04035	3CD5	7HI	-8.50	1DQ9	-8.41	0.09
P04058	2CEK	N8T	-12.54	1ZGB	-12.52	0.02
P04058	1ZGB	A1E	-11.24	2CEK	-11.61	0.37
P04150	6EL7	B9T	-9.70	5UC3	-5.97	3.73

P04150	5UC3	486	-4.46	6EL7	-0.07	4.39
P04181	2OAT	PFM	-7.63	5VWO	-6.78	0.85
P04181	5VWO	9QJ	-7.23	2OAT	-8.07	0.84
P04585	1JLB	NVP	-10.18	3DLG	-10.20	0.02
P04585	3DLG	GWE	-10.24	1JLB	-10.21	0.03
P04695	1TND	GSP	-14.13	1TAG	-12.72	1.41
P04695	1TAG	GDP	-13.12	1TND	-13.18	0.06
P04825	2DQM	BES	-9.91	5MFR	-8.41	1.50
P04825	5MFR	7MK	-8.87	2DQM	-8.31	0.56
P04925	4MA8	Z80	-5.79	4MA7	-5.94	0.15
P04925	4MA7	P2Z	-5.86	4MA8	-5.70	0.16
P05326	3ZKU	HCV	-7.32	2JB4	-7.23	0.09
P05326	2JB4	A14	-7.96	3ZKU	-7.38	0.58
P05798	2SAR	3GP	-7.23	1GMP	-7.71	0.48
P05798	1GMP	2GP	-8.41	2SAR	-7.31	1.10
P06401	3ZRA	ORB	-9.41	2W8Y	-9.19	0.22
P06401	2W8Y	486	-6.54	3ZRA	8.53	15.07
P06401	3KBA	WOW	-10.25	3ZRA	-8.52	1.73
P06401	3ZRA	ORB	-9.19	3KBA	-8.85	0.34
P06730	2V8W	MGO	-9.36	1IPB	-9.46	0.10
P06730	1IPB	GTA	-8.97	2V8W	-9.19	0.22
P06737	3CEM	AVD	-10.51	3DDS	-8.35	2.16
P06737	3DDS	26B	-9.14	3CEM	-8.05	1.09
P06745	1U0F	G6P	-8.91	1U0G	-8.91	0.00
P06745	1U0G	E4P	-6.46	1U0F	-6.20	0.26
P07097	1QFL	COA	-7.98	1OU6	-10.27	2.29
P07097	1OU6	168	-7.05	1QFL	-6.13	0.92
P07097	2VU2	PN5	-5.59	2WKV	-6.06	0.47
P07097	2WKV	COA	-7.41	2VU2	-7.84	0.43
P07251	3ZIA	ATP	-9.95	3ZIA	-9.37	0.58
P07251	3ZIA	ADP	-9.17	3ZIA	-8.19	0.98
P07332	3BKB	STU	-13.60	4E93	-9.12	4.48
P07332	4E93	GUI	-9.67	3BKB	-9.60	0.07
P07333	3KRL	KRL	-8.80	2I1M	-9.78	0.98
P07333	2I1M	5CN	-9.35	3KRL	-8.27	1.08
P07711	4AXM	V65	-7.87	3BC3	-7.93	0.06
P07711	3BC3	OPT	-7.13	4AXM	-6.91	0.22
P07788	4Q8B	SXX	-5.22	1OF0	-4.98	0.24
P07788	1OF0	EBS	-7.66	4Q8B	-8.40	0.74
P07814	4HVC	ANP	-11.46	5V58	-11.17	0.29
P07814	5V58	8X1	-12.43	4HVC	-11.52	0.91
P07858	6AY2	C1G	-6.88	1GMY	-7.30	0.42
P07858	1GMY	APD	-5.78	6AY2	-5.71	0.07
P07900	5J20	6FJ	-10.56	4AWQ	-10.16	0.40
P07900	4CWF	H05	-9.27	5J20	-8.41	0.86
P07900	4AWQ	592	-11.88	5J6L	-7.60	4.28
P07900	5J6L	6GC	-10.64	4CWF	-9.79	0.85
P07949	2IVU	ZD6	-9.36	4CKJ	-8.09	1.27

P07949	4CKJ	ADN	-8.19	2IVU	-7.68	0.51
P07986	3CUI	X4S	-9.72	3CUF	-9.85	0.13
P07986	3CUF	9MR	-7.40	3CUI	-7.69	0.29
P08037	1TW5	UDH	-9.84	1YRO	-9.89	0.05
P08037	1YRO	GDU	-12.14	1TW5	-11.14	1.00
P08049	5V48	TIO	-7.10	4ZR5	-6.87	0.23
P08049	4ZR5	RDF	-10.82	5V48	-7.80	3.02
P08191	5L4W	6KH	-7.49	4X50	-6.76	0.73
P08191	4X50	3X8	-9.28	5L4W	-9.21	0.07
P08246	3Q77	2HY	-8.77	1H1B	-7.14	1.63
P08246	1H1B	151	-6.92	3Q77	-6.56	0.36
P08473	2YB9	HA0	-8.66	1R1H	-9.19	0.53
P08473	1R1H	BIR	-9.20	2YB9	-9.21	0.01
P08473	4CTH	RDF	-10.36	1R1J	-9.01	1.35
P08473	1R1J	OIR	-8.31	4CTH	-8.27	0.04
P08519	4BV7	BV7	-4.85	4BV5	-4.78	0.07
P08519	4BV5	5C3	-6.53	4BV7	-7.43	0.90
P08581	3C1X	CKK	-12.51	5DG5	-10.55	1.96
P08581	5DG5	5B4	-12.62	3C1X	-9.92	2.70
P08659	3IES	M24	-13.23	5GYZ	-12.78	0.45
P08659	5GYZ	7BV	-10.78	3IES	-10.68	0.10
P08684	4NY4	2QH	-11.73	3UA1	-12.01	0.28
P08684	3UA1	08Y	-13.87	4NY4	-12.05	1.82
P08709	4ZXY	4T1	-11.02	4Z6A	-11.36	0.34
P08709	4Z6A	0Z6	-7.77	4ZXY	-7.73	0.04
P09147	1XEL	UPG	-11.16	1NAI	-10.77	0.39
P09147	1NAI	UDP	-8.90	1XEL	-9.64	0.74
P09148	1HXP	UDP	-8.17	1HXP	-8.16	0.01
P09148	1HXP	U5P	-7.32	1HXP	-7.30	0.02
P09237	2Y6D	TQJ	-8.72	2Y6C	-7.46	1.26
P09237	2Y6C	TQI	-8.39	2Y6D	-9.52	1.13
P09467	3KBZ	2T4	-7.78	4MJO	-7.20	0.58
P09467	4MJO	2C1	-7.99	3KBZ	-7.44	0.55
P09838	4I2I	AP5	-9.53	4IQT	-8.48	1.05
P09838	4IQT	1FO	-7.34	4I2I	-7.82	0.48
P09874	1WOK	CNQ	-9.24	3L3M	-8.89	0.35
P09874	1UK1	FRQ	-11.22	1WOK	-9.57	1.65
P09874	3L3M	A92	-10.33	1UK0	-10.34	0.01
P09874	1UK0	FRM	-11.27	1UK1	-11.11	0.16
P0A0N3	3BT9	DEQ	-9.77	3BTI	-9.68	0.09
P0A0N3	3BTI	BER	-10.30	3BT9	-10.68	0.38
P0A2K1	1TTQ	PLP	-9.34	2J9X	-9.19	0.15
P0A2K1	2J9X	P1T	-10.98	1TTQ	-7.96	3.02
P0A3R9	2CBQ	TH2	-8.81	1NCO	-8.73	0.08
P0A3R9	1NCO	CHR	-11.75	2CBQ	-8.52	3.23
P0A3Y5	3TM0	ANP	-9.22	1J7L	-9.81	0.59
P0A3Y5	1J7L	ADP	-10.42	3TM0	-9.30	1.12
P0A6I6	6CCQ	EX7	-9.05	1B6T	-8.39	0.66

P0A6I6	1B6T	COD	-9.90	6CCQ	-7.66	2.24
P0A6K3	1BSJ	MLN	-7.66	3K6L	-8.48	0.82
P0A6K3	3K6L	2BB	-8.17	1BSJ	-7.55	0.62
P0A6R0	5BNR	4VL	-9.51	5BNM	-8.64	0.87
P0A6R0	1HNJ	MLC	-7.41	5BNR	-8.17	0.76
P0A6R0	5BNM	4VK	-9.07	5BNR	-9.72	0.65
P0A6R0	5BNR	4VL	-9.51	1HNJ	-8.91	0.60
P0A786	2FZK	EOZ	-9.16	2FZG	-9.65	0.49
P0A786	2FZG	EOB	-8.33	2FZK	-8.07	0.26
P0A794	1HO4	PXP	-8.05	1IXO	-6.93	1.12
P0A794	1IXO	G3P	-5.80	1HO4	-6.31	0.51
P0A7F3	4FYY	CTP	-7.30	4AT1	-7.42	0.12
P0A7F3	4AT1	ATP	-7.65	4FYY	-7.54	0.11
P0A7G6	1XMS	ANP	-9.91	1XMV	-10.21	0.30
P0A7G6	1XMV	ADP	-9.19	1XMS	-9.33	0.14
P0A884	1JG0	DDT	-10.43	1DDU	-10.47	0.04
P0A884	1DDU	CB3	-9.65	1JG0	-7.85	1.80
P0A8M3	1EVL	TSB	-11.87	4HWR	-10.79	1.08
P0A8M3	4HWR	1B2	-10.94	1EVL	-11.52	0.58
P0A988	4N95	2HQ	-5.52	4MJR	-5.33	0.19
P0A988	4MJR	0LA	-6.73	4N95	-6.50	0.23
P0A9J6	1RK2	ADP	-9.86	1GQT	-9.35	0.51
P0A9J6	1GQT	ACP	-9.66	1RK2	-10.27	0.61
P0A9M2	5KNV	6W9	-8.14	5KNS	-8.23	0.09
P0A9M2	5KNS	3L7	-7.71	5KNV	-7.26	0.45
P0ABE7	5N2R	8JN	-8.67	5IUB	-9.05	0.38
P0ABE7	5IUB	6DV	-9.51	5N2R	-10.27	0.76
P0AC14	5U10	PT1	-9.76	5U0Z	-9.11	0.65
P0AC14	5U0Z	7PM	-8.71	5U10	-8.98	0.27
P0ACT4	4ABZ	T1C	-10.04	2O7O	-8.62	1.42
P0ACT4	2O7O	DXT	-10.98	4ABZ	-10.31	0.67
P0AEB2	3MZD	CXV	-7.62	1Z6F	-8.24	0.62
P0AEB2	1Z6F	BO9	-7.66	3MZD	-8.28	0.62
P0AEB2	5J8X	OK3	-7.13	3MZF	-7.19	0.06
P0AEB2	3MZF	IM2	-6.35	5J8X	-6.12	0.23
P0AES6	1AJ6	NOV	-9.83	5L3J	-8.44	1.39
P0AES6	5L3J	6G9	-6.96	1AJ6	-7.11	0.15
P0AEX9	4WMW	3R6	-5.36	4WMU	-5.34	0.02
P0AEX9	4WMU	19H	-9.59	4WMW	-7.08	2.51
P0AGJ9	1VBM	YSA	-12.30	2YXN	-10.61	1.69
P0AGJ9	2YXN	AZY	-8.22	1VBM	-7.79	0.43
P0C6U8	4TWW	3A7	-7.34	4MDS	-7.43	0.09
P0C6U8	4MDS	23H	-7.70	4TWW	-7.26	0.44
P0C6U8	2ALV	CY6	-7.28	2ALV	-6.73	0.55
P0C6U8	2ZU4	ZU3	-8.98	2ALV	-8.02	0.96
P0C6U8	2ALV	CY6	-7.28	3SZN	-6.84	0.44
P0C6U8	3SZN	G75	-6.81	2ZU4	-7.50	0.69
P0C6W3	4YOJ	RFM	-9.28	4YOI	-8.60	0.68

P0C6W3	4YOI	4F4	-8.23	4YQJ	-8.67	0.44
P0C6X7	2A5K	AZP	-7.26	4TWY	-7.28	0.02
P0C6X7	4TWY	3BL	-8.28	4TWY	-7.48	0.80
P0C6X7	3D62	959	-5.70	5N5O	-5.50	0.20
P0C6X7	4TWY	3BL	-8.28	3D62	-7.29	0.99
P0C6X7	5N5O	8O5	-8.42	3D62	-6.63	1.79
P0C6X7	3D62	959	-5.70	2A5K	-5.37	0.33
P0DMV8	5MKR	TI8	-8.71	5BN9	-8.73	0.02
P0DMV8	5BN9	ADP	-10.59	5MKR	-9.80	0.79
P0DPI1	4HEV	AXM	-8.94	5V8U	-7.33	1.61
P0DPI1	5V8U	90M	-7.82	4HEV	-7.48	0.34
P10153	2C02	ADP	-7.87	1HI3	-7.66	0.21
P10153	1HI3	A2P	-7.91	2C02	-7.55	0.36
P10408	2FSH	ANP	-8.26	2FSI	-6.70	1.56
P10408	2FSI	ADP	-9.36	2FSH	-8.84	0.52
P10584	2CVQ	NDP	-8.48	1BDM	-8.48	0.00
P10584	1BDM	NAX	-9.39	2CVQ	-9.80	0.41
P10613	5TZ1	VT1	-10.27	5V5Z	-10.47	0.20
P10613	5V5Z	1YN	-11.27	5TZ1	-11.31	0.04
P10632	2VN0	TDZ	-10.94	2NNI	-10.85	0.09
P10632	2NNI	MTK	-12.00	2VN0	-10.98	1.02
P10760	1XWF	ADN	-10.89	2H5L	-10.79	0.10
P10760	2H5L	3DD	-9.40	1XWF	-8.04	1.36
P10824	5KDL	GSP	-14.09	1GDD	-12.74	1.35
P10824	1GDD	GDP	-12.80	5KDL	-13.40	0.60
P10828	2J4A	OEF	-11.20	1R6G	-6.18	5.02
P10828	1R6G	442	-6.38	2J4A	-6.12	0.26
P11021	5F2R	ACP	-11.49	3LDP	-11.06	0.43
P11021	3LDP	3P1	-10.01	5F2R	-10.91	0.90
P11086	2G72	F21	-10.21	3KR0	-8.20	2.01
P11086	3KR0	172	-7.30	2G72	-7.34	0.04
P11142	6BII	ADP	-10.88	3FZM	-9.87	1.01
P11142	3FZM	3GO	-11.53	6BII	-10.98	0.55
P11309	5TEX	7AU	-10.11	4MBI	-10.31	0.20
P11309	4MBI	26K	-6.95	5TEX	-7.13	0.18
P11712	5A5I	XI1	-7.72	5K7K	-8.24	0.52
P11712	5K7K	6RJ	-8.68	5K7K	-8.33	0.35
P11712	5K7K	6RJ	-8.68	5A5J	-9.29	0.61
P11712	5A5J	6YF	-8.89	5A5I	-7.69	1.20
P12499	5AH7	C7J	-7.17	4K4P	-7.35	0.18
P12499	4K4P	3TL	-6.70	5AH7	-6.13	0.57
P12694	2BEU	THV	-8.54	1V1M	-7.28	1.26
P12694	1V1M	TDP	-7.38	2BEU	-7.19	0.19
P12821	4BZS	K26	-10.03	6EN5	-9.47	0.56
P12821	6EN5	BJ2	-8.36	4BZS	-8.15	0.21
P12823	1R6A	RVP	-6.96	2P1D	-7.06	0.1
P12823	2P1D	5GP	-7.74	1R6A	-7.61	0.13
P13051	3FCI	3FI	-7.90	2HXM	-7.53	0.37

P13051	2HXM	302	-8.15	3FCI	-8.46	0.31
P13551	4M1K	GDP	-8.41	2J7K	-8.86	0.45
P13551	2J7K	GCP	-9.17	4M1K	-8.65	0.52
P14210	5CT2	CXS	-5.32	5CT3	-5.45	0.13
P14210	5CT3	54O	-4.14	5CT2	-3.99	0.15
P14555	3U8D	U8D	-9.18	1POE	-9.03	0.15
P14555	1POE	GEL	-6.97	3U8D	-6.65	0.32
P14618	4G1N	NZT	-8.63	3U2Z	-9.17	0.54
P14618	3U2Z	07T	-8.00	4G1N	-8.01	0.01
P14677	1QMF	CES	-7.99	2ZC3	-7.87	0.12
P14677	2ZC3	BMG	-6.90	2ZC3	-7.41	0.51
P14677	2ZC3	BMG	-6.90	2Z2M	-7.49	0.59
P14677	2Z2M	CDS	-8.95	1QMF	-8.25	0.7
P14780	2OVZ	5MR	-10.41	2OVX	-9.55	0.86
P14780	2OVX	4MR	-11.51	2OVZ	-10.62	0.89
P14920	3ZNO	SE5	-8.32	3ZNP	-8.22	0.10
P14920	3ZNP	SE2	-7.66	3ZNO	-7.44	0.22
P15291	4EEM	UDP	-10.10	3EE5	-9.68	0.42
P15291	3EE5	UDH	-9.50	4EEM	-9.57	0.07
P15468	5ARK	UMP	-6.84	5ARL	-6.81	0.03
P15468	5ARL	DCM	-7.00	5ARK	-7.64	0.64
P15659	5IFD	6AQ	-5.02	5IEQ	-5.72	0.70
P15659	5IEQ	1CI	-5.58	5IFD	-5.01	0.57
P15790	4JR7	GNP	-10.31	4FI1	-10.45	0.14
P15790	4FI1	ATP	-9.99	4JR7	-8.78	1.21
P15917	1PWU	GM6	-7.70	5D1T	-8.27	0.57
P15917	5D1T	56R	-9.43	1PWU	-7.92	1.51
P16094	4YP2	NCA	-6.49	1AHB	-5.79	0.70
P16094	1AHB	FMP	-8.02	4YP2	-6.93	1.09
P16330	3ZBR	NAP	-10.53	2YDD	-9.99	0.54
P16330	2YDD	2AM	-7.82	3ZBR	-8.05	0.23
P17252	3IW4	LW4	-11.23	4RA4	-9.42	1.81
P17252	4RA4	3KZ	-10.30	3IW4	-10.48	0.18
P17612	4UJA	4L7	-10.30	3OXT	-9.98	0.32
P17612	3OXT	3SB	-9.16	4UJA	-8.63	0.53
P18031	1Q1M	234	-8.67	2FJN	-8.55	0.12
P18031	1NZ7	901	-6.78	1Q1M	-6.27	0.51
P18031	2FJN	73	-10.95	3CWE	-10.05	0.90
P18031	3EB1	LZQ	-7.45	1NZ7	-6.90	0.55
P18031	3CWE	825	-10.35	2CNE	-8.28	2.07
P18031	2CNE	DFJ	-10.11	3EB1	2.39	12.5
P18946	3L74	FMX	-12.21	3L71	-10.42	1.79
P18946	3L71	AZO	-8.69	3L74	-8.56	0.13
P19483	1NBM	ATP	-9.14	1E79	-9.15	0.01
P19483	1W0K	ADP	-8.95	1W0K	-8.56	0.39
P19483	1E79	ADP	-8.63	2JIZ	-8.53	0.10
P19483	2JIZ	ANP	-9.25	1NBM	-9.77	0.52
P19491	3RTF	CWD	-8.78	5FHO	-8.29	0.49

P19491	5FHO	5XN	-10.66	3RTF	-7.30	3.36
P19866	1RM3	NDP	-9.44	1NBO	-9.59	0.15
P19866	1NBO	NAD	-10.33	1RM3	-8.93	1.40
P20231	2BM2	PM2	-8.29	2FPZ	-8.07	0.22
P20231	2FPZ	270	-6.48	2BM2	-5.82	0.66
P20339	3MJH	GTP	-12.89	1N6I	-12.99	0.10
P20339	1N6I	GDP	-12.00	3MJH	-12.06	0.06
P20701	3BQM	BQM	-10.49	4IXD	-8.30	2.19
P20701	2ICA	2IC	-10.20	2ICA	-8.62	1.58
P20701	1XDG	AB8	-8.10	1XDG	-8.09	0.01
P20701	4IXD	1HV	-9.82	3BQM	-7.15	2.67
P20906	4JUC	TPP	-10.09	3FSJ	-8.94	1.15
P20906	3FSJ	D7K	-9.89	4JUC	-10.56	0.67
P21524	2EUD	GCQ	-9.14	2CVX	-8.46	0.68
P21524	2CVX	ADP	-10.27	2EUD	-9.61	0.66
P21673	2B58	COA	-9.34	2FXF	-8.96	0.38
P21673	2FXF	ACO	-10.07	2B58	-8.52	1.55
P21860	4OTW	DB8	-7.93	3LMG	-8.36	0.43
P21860	3LMG	ANP	-9.24	4OTW	-9.11	0.13
P21873	3DVA	TPW	-8.34	3DUF	-7.85	0.49
P21873	3DUF	R1T	-7.71	3DVA	-7.76	0.05
P22102	4ZZ3	4DW	-8.95	4ZYT	-9.00	0.05
P22102	4ZYT	3Y9	-9.46	4ZZ3	-8.83	0.63
P22102	1NJS	KEU	-11.15	1ZLY	-10.32	0.83
P22102	1ZLY	DQB	-9.23	1NJS	-9.48	0.25
P22629	1SRJ	NAB	-10.94	1N43	-7.94	3.00
P22629	1N43	BTN	-7.19	1SRJ	-6.28	0.91
P22734	3HVJ	705	-12.42	3NW9	-8.66	3.76
P22734	4PYL	TCW	-7.88	3HVJ	-7.59	0.29
P22734	3NW9	637	-13.64	5P8Y	-13.09	0.55
P22734	5P8Y	763	-10.53	4PYL	-9.93	0.60
P22894	5H8X	5XT	-10.23	1ZP5	-10.74	0.51
P22894	1ZP5	2NI	-9.77	5H8X	-8.67	1.10
P23458	6AAH	9T6	-10.52	5E1E	-9.82	0.70
P23458	5E1E	5JG	-8.90	6AAH	-8.89	0.01
P23467	2H04	4UN	-9.38	2H03	-8.76	0.62
P23467	2H03	3UN	-7.71	2H04	-7.43	0.28
P23467	2I5X	UA5	-8.01	2I4G	-7.43	0.58
P23467	2I4G	UA1	-7.94	2I5X	-8.64	0.70
P23919	1E2G	TYD	-8.87	1NN3	-9.38	0.51
P23919	1NN3	2DT	-8.44	1E2G	-8.33	0.11
P23946	1T31	OHH	-11.45	2HVX	-9.51	1.94
P23946	2HVX	DRX	-8.61	1T31	-8.89	0.28
P24941	3TNW	F18	-7.61	2UZL	-7.74	0.13
P24941	2UZL	C94	-9.26	3TNW	-9.48	0.22
P25440	4QEV	3IO	-8.26	4MR5	-7.95	0.31
P25440	4MR5	1K0	-8.26	4QEV	-8.42	0.16
P26281	6AN4	J1F	-11.33	1EX8	-10.68	0.65

P26281	1EX8	A4P	-11.21	6AN4	-10.40	0.81
P26662	4KTC	1X3	-12.04	4I33	-11.74	0.30
P26662	4I33	1BV	-12.39	4KTC	-11.66	0.73
P26664	2GVF	NHN	-8.44	1RTL	-7.69	0.75
P26664	1RTL	CPX	-7.80	2GVF	-6.47	1.33
P27001	1B7Y	FYA	-11.83	2AKW	-7.96	3.87
P27001	2AKW	200	-7.70	1B7Y	-6.30	1.40
P27487	4N8E	2KV	-8.85	4N8D	-8.76	0.09
P27487	4N8D	2KS	-8.55	4N8E	-9.32	0.77
P27487	2G63	AAF	-8.09	5T4H	-8.61	0.52
P27487	5T4H	75J	-9.64	2G63	-8.39	1.25
P27601	3CX8	GSP	-13.53	1ZCB	-9.69	3.84
P27601	1ZCB	GDP	-12.95	3CX8	-12.90	0.05
P27694	5E7N	5KR	-6.69	4LUZ	-6.49	0.20
P27694	4LUZ	1XT	-8.79	5E7N	-8.06	0.73
P27915	5E9Q	SAM	-9.00	3P8Z	-8.52	0.48
P27915	3P8Z	36A	-10.19	5E9Q	-9.79	0.40
P27958	6BQK	Z1E	-9.74	2F9V	-7.53	2.21
P27958	2F9V	BN6	-7.26	6BQK	-7.50	0.24
P28012	1JEP	DFL	-8.87	1FM8	-9.66	0.79
P28012	1FM8	DDC	-9.52	1JEP	-9.36	0.16
P28161	1HNC	GDN	-6.56	3GUR	-5.83	0.73
P28161	3GUR	BYG	-6.74	1HNC	-6.66	0.08
P28248	2V9X	DUT	-9.23	1XS4	-8.49	0.74
P28248	1XS4	DCP	-8.40	2V9X	-7.27	1.13
P28327	3C4Z	ADP	-8.98	4L9I	-9.11	0.13
P28327	4L9I	8PR	-8.58	3C4Z	-8.14	0.44
P28523	3PWD	CZ0	-9.79	1DAW	-10.49	0.70
P28523	1DAW	ANP	-10.08	3PWD	-8.98	1.10
P28845	4K1L	SFF	-8.61	4HFR	-7.83	0.78
P28845	4HFR	14M	-8.92	4K1L	-9.28	0.36
P29317	5IA1	ZZL	-9.89	5NKH	-9.20	0.69
P29317	5NKH	8ZQ	-10.52	5IA1	-9.73	0.79
P29375	5IVE	6E8	-9.04	5IVC	-6.67	2.37
P29375	5IVC	6E7	-9.62	5IVE	-10.20	0.58
P29474	5VVB	9P7	-9.73	5UOC	-10.12	0.39
P29474	5UOC	8FD	-10.56	5VVB	-10.62	0.06
P29476	4IMW	1EV	-10.82	3TYM	-11.17	0.35
P29476	3TYM	08R	-9.09	4IMW	-8.75	0.34
P29477	1QW4	3AR	-6.08	3EBF	-5.20	0.88
P29477	4UX6	YWO	-10.33	1QW4	-10.39	0.06
P29477	2Y37	A54	-8.36	2Y37	-8.87	0.51
P29477	3EBF	332	-9.93	4UX6	-9.70	0.23
P29597	4GJ3	0XP	-8.70	4GIH	-8.02	0.68
P29597	6AAM	9T6	-9.43	4GJ3	-9.80	0.37
P29597	6DBM	G4J	-10.29	4WOV	-8.82	1.47
P29597	4GIH	0X5	-8.05	6AAM	-7.76	0.29
P29597	5C03	AGS	-9.55	5C03	-9.87	0.32

P29597	4WOV	3SM	-8.94	6DBM	-8.73	0.21
P30405	5CCR	4ZT	-6.92	5CBW	-6.61	0.31
P30405	5CBW	4ZO	-6.89	5CCR	-6.85	0.04
P30419	4C2Z	646	-10.49	3IU2	-8.78	1.71
P30419	3IU2	96	-9.09	4C2Z	-8.88	0.21
P31013	6ECG	PM9	-8.95	2YCN	-8.28	0.67
P31013	2YCN	P61	-9.93	6ECG	-9.75	0.18
P31749	3MVH	WFE	-10.14	4EKK	-9.38	0.76
P31749	3QKM	SM9	-10.72	4EKK	-9.43	1.29
P31749	4EKK	ANP	-9.28	3CQW	-8.93	0.35
P31749	3CQW	CQW	-8.94	3QKM	-8.28	0.66
P31749	4EKK	ANP	-8.78	3QKM	-9.06	0.28
P31749	3QKM	SM9	-10.72	3MVH	-9.85	0.87
P31751	2UW9	GVP	-9.87	3E87	-8.88	0.99
P31751	3E87	G95	-8.90	2UW9	-9.42	0.52
P33590	3MZ9	BHN	-7.64	5ON4	-7.15	0.49
P33590	5ON4	9YK	-6.88	3MZ9	-7.45	0.57
P33981	4O6L	2QK	-10.18	4JT3	-9.90	0.28
P33981	5NTT	SVE	-10.15	4O6L	-9.61	0.54
P33981	5N9S	8QW	-12.28	5N9S	-9.28	3.00
P33981	4JT3	1PH	-10.06	5NTT	-9.08	0.98
P34913	3I28	34N	-11.59	3KOO	-10.82	0.77
P34913	3KOO	24D	-8.96	3I28	-9.71	0.75
P35557	3ID8	MRK	-7.80	4IWV	-8.29	0.49
P35557	4IWV	1J9	-10.77	3ID8	-8.14	2.63
P35747	5DQF	CZE	-7.19	5ID9	-7.87	0.68
P35747	5ID9	6A4	-6.84	5DQF	-6.32	0.52
P35968	2QU6	857	-10.80	3VNT	-10.84	0.04
P35968	3VNT	0JA	-12.59	2QU6	-8.75	3.84
P36897	5USQ	8LY	-10.85	3TzM	-10.25	0.60
P36897	3TzM	85	-11.27	5USQ	-10.21	1.06
P37231	4HEE	14R	-12.58	3TY0	-10.74	1.84
P37231	2VV1	4HD	-9.97	4HEE	-8.08	1.89
P37231	4R6S	3K2	-12.34	4HEE	-11.00	1.34
P37231	4HEE	14R	-12.58	4R6S	-9.72	2.86
P37231	6-Aug	BXG	-8.96	4R6S	-8.68	0.28
P37231	5Z6S	RTF	-10.01	2VV1	-9.83	0.18
P37231	6AVI	GW9	-7.97	5DV3	-7.73	0.24
P37231	3TY0	82	-11.17	6-Aug	-10.48	0.69
P37231	5DV3	B05	-9.42	6AVI	-8.56	0.86
P37231	4R6S	3K2	-11.41	5Z6S	-9.75	1.66
P37268	3WCI	BH5	-7.57	3WCH	-7.57	0.00
P37268	3WCH	8PH	-7.39	3WCI	-6.94	0.45
P37344	2C99	ANP	-9.79	4QOS	-9.48	0.31
P37344	4QOS	ADP	-8.86	2C99	-8.98	0.12
P39058	2VAV	CSC	-8.59	2VAT	-7.35	1.24
P39058	2VAT	COA	-8.43	2VAV	-8.43	0.00
P40233	2CSN	CKI	-6.80	1CSN	-7.63	0.83

P40233	1CSN	ATP	-9.62	2CSN	-8.41	1.21
P41222	3O2Y	PLM	-6.30	4IMN	-6.16	0.14
P41222	4IMN	1PG	-4.40	3O2Y	-4.41	0.01
P41279	4Y83	49B	-10.52	4Y85	-9.27	1.25
P41279	4Y85	499	-11.18	4Y83	-9.23	1.95
P41743	5LI1	ANP	-10.70	5LI9	-7.90	2.8
P41743	5LI9	ACP	-10.68	5LI1	-10.57	0.11
P42216	1GQC	CMK	-10.17	1GQC	-9.53	0.64
P42216	1GQC	C5P	-8.19	1GQC	-8.25	0.06
P42330	4FAL	0T0	-10.89	4FA3	-10.17	0.72
P42330	6F78	CVN	-10.10	4FAL	-10.02	0.08
P42330	4FA3	0SL	-10.56	5JM5	-9.96	0.6
P42330	4WDW	WDW	-8.56	6F78	-8.50	0.06
P42330	5JM5	6LG	-9.51	1RY8	-9.92	0.41
P42330	1RY8	RUT	-12.86	4WDW	-12.67	0.19
P42336	5UL1	8DY	-8.29	4WAF	-8.44	0.15
P42336	4WAF	3K6	-9.84	5UL1	-8.94	0.90
P42574	1NMS	161	-7.59	1NMQ	-6.09	1.50
P42574	1NMQ	160	-6.81	1NMS	-6.77	0.04
P43405	4YJV	4DT	-9.30	3FQH	-9.77	0.47
P43405	5Y5U	8OU	-9.63	4YJV	-9.07	0.56
P43405	3FQH	057	-8.82	5CXH	-8.72	0.10
P43405	5CXH	55M	-8.80	5Y5U	-7.95	0.85
P43889	2V0J	H2U	-8.53	4E1K	-7.29	1.24
P43889	4E1K	0N5	-10.22	2V0J	-8.69	1.53
P45452	3ZXH	E41	-7.90	2YIG	-7.97	0.07
P45452	2YIG	5EL	-11.38	3ZXH	-10.22	1.16
P47811	2GHL	LIB	-9.58	4TYH	-9.68	0.10
P47811	5O90	SB4	-9.81	2GHL	-8.18	1.63
P47811	2EWA	SB2	-10.89	2EWA	-8.21	2.68
P47811	4TYH	39G	-8.93	5O90	-8.72	0.21
P47934	1NDI	COA	-8.09	2H3P	-8.24	0.15
P47934	2H3P	ACO	-8.46	1NDI	-7.82	0.64
P48730	5OKT	9XK	-8.66	4TWC	-8.43	0.23
P48730	5IH6	AUG	-8.74	5OKT	-9.36	0.62
P48730	4TWC	37J	-9.68	5IH5	-9.20	0.48
P48730	5IH5	AUE	-9.69	5IH6	-8.95	0.74
P48735	6ADI	NDP	-10.39	5SVO	-8.63	1.76
P48735	5SVO	NAP	-9.44	6ADI	-9.50	0.06
P49336	4F6W	0SS	-11.01	4F6S	-10.18	0.83
P49336	4F6S	0SQ	-9.50	4F6W	-8.33	1.17
P49336	4F7J	0SU	-10.11	4F70	-8.54	1.57
P49336	4F70	0ST	-9.62	4F7J	-9.53	0.09
P49610	4AZG	OAN	-7.33	4AZH	-7.90	0.57
P49610	4AZH	LOG	-6.54	4AZG	-6.21	0.33
P49841	3ZDI	UGJ	-8.34	4DIT	-7.35	0.99
P49841	3F88	3HT	-9.31	3F88	-8.39	0.92
P49841	4DIT	0KD	-8.65	4NM0	-7.63	1.02

P49841	4NM0	ADP	-8.34	3ZDI	-8.33	0.01
P49902	2XJE	ATP	-7.72	4H4B	-7.46	0.26
P49902	2XJC	5GP	-9.56	2XJC	-9.56	0.00
P49902	2XJE	ATP	-9.72	2XJE	-9.26	0.46
P49902	6DDH	IMP	-9.02	2XJE	-9.46	0.44
P49902	5OPM	DTP	-7.72	5OPM	-7.43	0.29
P49902	4H4B	11H	-8.99	6DDH	-8.40	0.59
P49789	3FIT	A	-7.82	6FIT	-6.24	1.58
P48789	6FIT	AMW	-8.23	3FIT	-6.77	1.46
P50225	2D06	EST	-10.71	3U3K	-8.36	2.35
P50225	3U3K	03V	-7.76	2D06	-7.65	0.11
P50579	2ADU	R20	-7.24	1YW7	-7.13	0.11
P50579	1YW7	A41	-8.26	2ADU	-7.62	0.64
P50750	3LQ5	SLQ	-8.14	3TN8	-8.95	0.81
P50750	3TN8	F18	-8.03	3LQ5	-6.91	1.12
P51149	1T91	GTP	-14.22	3LAW	-13.10	1.12
P51149	3LAW	GNP	-13.79	1T91	-13.69	0.10
P51449	4WPF	3SN	-10.95	4QM0	-9.48	1.47
P51449	6BN6	XGH	-11.40	4WPF	-10.84	0.56
P51449	6CN6	F7J	-14.23	6CN6	-10.39	3.84
P51449	4QM0	39K	-10.29	6BN6	-8.88	1.41
P51857	3CAS	ASD	-11.68	3UZX	-10.80	0.88
P51857	3UZX	AOM	-12.29	3CAS	-11.71	0.58
P51955	2XKF	BX1	-8.67	2W5B	-7.62	1.05
P51955	2W5B	AGS	-8.24	2XKF	-8.00	0.24
P52333	4QPS	37Q	-9.77	4HVG	-9.00	0.77
P52333	4HVG	19Q	-7.53	4QPS	-8.06	0.53
P52699	5EWA	9BZ	-5.64	5HH4	-5.56	0.08
P52699	5HH4	60M	-6.97	5EWA	-5.84	1.13
P52700	2GFJ	VI	-7.68	2FU9	-7.27	0.41
P52700	2FU9	MP2	-6.52	2GFJ	-6.47	0.05
P52732	5ZO8	4C5	-10.54	2GM1	-10.36	0.18
P52732	2GM1	2AZ	-11.27	5ZO8	-8.90	2.37
P53355	2W4K	ADP	-8.36	5AUT	-8.19	0.17
P53355	5AUT	2AN	-8.91	2W4K	-8.30	0.61
P53582	5YKP	OVA	-7.14	4HXX	-6.45	0.69
P53582	4HXX	1AY	-9.01	5YKP	-9.33	0.32
P53779	1PMN	984	-9.98	4U79	-8.41	1.57
P53779	3FV8	JK3	-8.26	1PMN	-6.88	1.38
P53779	4U79	3EL	-11.24	6EKD	-11.70	0.46
P53779	6EKD	B9K	-9.67	3FV8	-8.56	1.11
P55055	1PQ6	965	-13.02	4DK8	-10.09	2.93
P55055	4DK8	0KT	-10.34	1PQ6	-6.15	4.19
P56221	3STD	MQ0	-12.29	4STD	-9.46	2.83
P56221	4STD	BFS	-10.42	3STD	-9.53	0.89
P56273	4J7D	I31	-8.87	4LWV	-8.12	0.75
P56273	4LWV	20W	-9.38	4J7D	-8.36	1.02
P60045	1TD7	NFL	-7.73	3NJU	-7.48	0.25

P60045	3NJU	ANN	-5.97	1TD7	-6.16	0.19
P60568	1PY2	FRH	-9.28	1PW6	-7.51	1.77
P60568	1PW6	FRB	-8.80	1PY2	-8.95	0.15
P60766	3EG5	GNP	-12.60	5C2J	-11.05	1.55
P60766	5C2J	GDP	-9.87	3EG5	-10.13	0.26
P61157	2P9U	ANP	-10.30	2P9P	-10.43	0.13
P61157	2P9P	ADP	-10.35	2P9U	-10.40	0.05
P61586	4XOI	GTP	-12.30	1S1C	-12.30	0.00
P61586	1S1C	GNP	-12.86	4XOI	-12.40	0.46
P62483	3EB4	NDP	-15.16	2R9R	-9.49	5.67
P62483	2R9R	NAP	-14.90	3EB4	-15.14	0.24
P62491	5C46	GSP	-13.41	1OIX	-13.29	0.12
P62491	1OIX	GDP	-12.82	5C46	-12.64	0.18
P62508	2ZAS	1OH	-9.89	2GPP	-8.63	1.26
P62508	2GPP	1BA	-11.36	2ZAS	-7.30	4.06
P62575	2YA7	ZMR	-8.95	5KKY	-8.01	0.94
P62575	5KKY	6UD	-8.71	2YA7	-8.78	0.07
P62593	1AXB	FOS	-6.95	1JWZ	-6.83	0.12
P62593	1JWZ	105	-7.90	1AXB	-8.18	0.28
P62617	1H48	C5P	-6.40	2AMT	-5.67	0.73
P62617	2AMT	1AA	-5.66	1H48	-5.89	0.23
P62937	5T9U	7HG	-8.76	5T9W	-8.80	0.04
P62937	5T9W	78E	-9.20	5T9U	-8.96	0.24
P62993	3IN7	AYQ	-8.53	3IMJ	-7.61	0.92
P62993	3IMJ	AYI	-6.85	3IN7	-7.72	0.87
P63000	3SBE	GNP	-9.64	1I4D	-10.33	0.69
P63000	1I4D	GDP	-11.61	3SBE	-11.56	0.05
P63086	4N4S	2H1	-12.78	4N4S	-8.32	4.46
P63086	4XOY	DX4	-4.87	4N4S	-5.19	0.32
P63086	4GVA	ADP	-7.61	4GVA	-8.47	0.86
P63086	4N4S	2H1	-12.41	4XOY	-9.92	2.49
P63820	4NAU	2W3	-9.24	4NAH	-9.31	0.07
P63820	4NAH	2VJ	-8.92	4NAU	-8.20	0.72
P65248	2CCG	TMP	-8.69	4GFD	-7.17	1.52
P65248	4GFD	0YB	-9.43	2CCG	-8.96	0.47
P67091	4R2L	ATP	-10.75	4R2M	-10.04	0.71
P67091	4R2M	ANP	-10.06	4R2L	-10.04	0.02
P69834	5MRP	6BC	-11.08	4NAN	-7.66	3.42
P69834	4NAN	2JM	-7.68	5MRP	-5.64	2.04
P76143	3GND	5RP	-6.60	4P2V	-6.66	0.06
P76143	4P2V	26T	-6.41	3GND	-6.28	0.13
P78536	3LEA	Z93	-9.30	3LGP	-8.98	0.32
P78536	3LGP	50X	-12.74	3LEA	-9.06	3.68
P80561	1TF9	PHI	-7.22	1XBU	-6.79	0.43
P80561	1XBU	IOY	-6.63	1TF9	-7.11	0.48
P82597	4KE8	1QY	-5.92	4KE7	-5.50	0.42
P82597	4KE7	1QX	-5.65	4KE8	-5.99	0.34
P93330	4GY9	ZIP	-6.01	4JHI	-6.18	0.17

P93330	4JHI	EMU	-7.15	4GY9	-7.08	0.07
P93836	1TFZ	869	-5.60	1TG5	-6.46	0.86
P93836	1TG5	645	-6.73	1TFZ	-6.11	0.62
P95607	3HGI	BEZ	-6.05	3I4Y	-5.74	0.31
P95607	3I4Y	35C	-6.48	3HGI	-5.57	0.91
P96907	4FNB	3HC	-5.95	4FND	-5.14	0.81
P96907	4FND	3H9	-6.98	4FNB	-6.12	0.86
P97612	3K84	K84	-8.86	3K83	-8.42	0.44
P97612	3K83	F27	-11.61	3K84	-10.13	1.48
P98170	6EY2	C3T	-8.23	5C3K	-7.05	1.18
P98170	5C3K	4XF	-6.24	6EY2	-6.42	0.18
P9WFX1	3RV6	VAE	-9.60	3RV9	-8.39	1.21
P9WFX1	3RV9	RVD	-6.96	3RV6	-7.64	0.68
P9WGR0	5VRN	9JM	-11.12	5VRM	-12.05	0.93
P9WGR0	5VRM	9JJ	-12.42	5VRN	-10.82	1.60
P9WGR1	4D0S	9G4	-8.98	4OXY	-9.61	0.63
P9WGR1	4OXY	1TN	-7.80	4D0S	-7.88	0.08
P9WHE9	1W19	T1P	-8.14	2C9D	-7.81	0.33
P9WHE9	2C9D	PHR	-7.15	1W19	-7.58	0.43
P9WI81	3F69	XDR	-14.37	5U94	-11.41	2.96
P9WI81	5U94	G93	-10.03	3F69	-8.95	1.08
P9WIL5	2A84	ATP	-11.38	3ISJ	-10.35	1.03
P9WIL5	3ISJ	A8D	-8.16	2A84	-7.14	1.02
P9WMC0	5F0C	5TE	-11.77	5F0H	-11.03	0.74
P9WMC0	5F0H	5TC	-13.31	5F0C	-10.40	2.91
P9WMC1	5F04	5TB	-9.64	4M3D	-9.65	0.01
P9WMC1	4M3D	2H2	-12.22	5F04	-11.35	0.87
P9WMC1	5J3L	6FR	-8.57	5F04	-7.30	1.27
P9WMC1	5F04	5TB	-9.64	5J3L	-10.47	0.83
P9WNH5	2WUG	HPK	-7.59	5JZB	-7.07	0.52
P9WNH5	5JZB	6OT	-7.29	2WUG	-5.63	1.66
P9WNX1	4KM2	ATR	-9.86	2CIG	-9.06	0.80
P9WNX1	2CIG	1DG	-13.08	4KM2	-11.34	1.74
P9WNX1	5U26	MMV	-7.86	6DDP	-8.26	0.4
P9WNX1	6DDP	G6Y	-10.34	5U26	-10.66	0.32
P9WQ81	4WYD	3VR	-6.22	4MQQ	-6.13	0.09
P9WQ81	4MQQ	2B6	-9.05	4WYD	-8.67	0.38
Q00441	1JIN	KTN	-8.16	1OXA	-8.22	0.06
Q00441	1OXA	DEB	-8.73	1JIN	-9.04	0.31
Q00610	4G55	VH2	-8.62	2XZG	-8.10	0.52
Q00610	2XZG	VH1	-9.60	4G55	-9.25	0.35
Q00955	5CTB	57J	-8.88	3TVW	-9.28	0.40
Q00955	3TVW	07H	-8.58	5CTB	-7.85	0.73
Q00972	4E01	ANP	-10.13	3TZ5	-11.07	0.94
Q00972	3TZ5	ADP	-10.59	4E01	-9.93	0.66
Q00987	5LAY	6SS	-9.10	4ZFY	-9.87	0.77
Q00987	4ZFY	4T4	-10.21	5LAY	-9.57	0.64
Q01064	5B25	4QJ	-12.50	5W6E	-9.91	2.59

Q01064	5W6E	0NY	-8.61	5B25	-7.61	1.00
Q02127	2B0M	201	-11.22	5HIN	-10.67	0.55
Q02127	5HIN	1KL	-10.66	2B0M	-8.74	1.92
Q02293	4GTO	7TO	-8.10	1O1S	-8.88	0.78
Q02293	1O1S	1NH	-8.15	4GTO	-7.82	0.33
Q03181	3SP9	IL2	-10.06	5U44	-8.87	1.19
Q03181	5U44	7SV	-10.56	3SP9	-10.36	0.20
Q05097	5MIH	7NU	-6.15	4YW7	-4.95	1.20
Q05097	4YW7	4J0	-7.01	5MIH	-6.95	0.06
Q05769	1DDX	PGX	-9.48	3MDL	-8.80	0.68
Q05769	3OLU	1AG	-8.50	3OLU	-9.11	0.61
Q05769	3MDL	1AG	-7.67	5FDQ	-7.53	0.14
Q05769	5FDQ	60A	-4.50	1DDX	-4.43	0.07
Q06187	4ZLZ	4RV	-8.84	4NWM	-8.11	0.73
Q06187	4NWM	2P5	-11.50	4ZLZ	-11.19	0.31
Q06528	5JR3	4MU	-7.28	4WXH	-7.20	0.08
Q06528	4WXH	3VL	-13.10	5JR3	-10.45	2.65
Q07343	2QYL	NPV	-10.55	5K6J	-10.25	0.30
Q07343	5K6J	6QQ	-8.43	2QYL	-8.19	0.24
Q07817	3ZLO	X8U	-12.57	3ZLN	-10.85	1.72
Q07817	3ZLN	H0Y	-12.17	3ZLO	-11.75	0.42
Q07820	4OQ5	2UU	-12.01	4HW2	-11.95	0.06
Q07820	4HW2	19H	-10.26	4OQ5	-7.29	2.97
Q07869	1K7L	544	-10.66	3VI8	-10.57	0.09
Q07869	3VI8	13M	-11.45	1K7L	-10.47	0.98
Q08210	4CQA	ID6	-11.25	1TV5	-8.90	2.35
Q08210	1TV5	A26	-8.90	4CQA	-8.26	0.64
Q08499	5K1I	6PT	-8.77	4W1O	-8.33	0.44
Q08499	4W1O	3GJ	-8.49	5K1I	-8.45	0.04
Q08499	3SL8	JN7	-8.77	3IAK	-7.59	1.18
Q08499	3IAK	EV1	-7.91	3SL8	-7.88	0.03
Q08603	3C72	CX1	-9.38	3PZ1	-9.00	0.38
Q08603	3PZ1	3PZ	-8.03	3C72	-8.58	0.55
Q12051	2ZEU	B71	-8.52	2Z4W	-7.27	1.25
Q12051	2Z4W	749	-8.72	2ZEU	-8.99	0.27
Q12341	4PSX	COA	-8.45	1BOB	-7.78	0.67
Q12341	1BOB	ACO	-7.44	4PSX	-8.24	0.8
Q12852	5CEP	50E	-8.60	5CEO	-8.76	0.16
Q12852	5CEO	50D	-9.20	5CEP	-8.23	0.97
Q13231	1HKM	ALI	-6.30	5NRA	-6.56	0.26
Q13231	5NRA	95K	-8.77	1HKM	-8.92	0.15
Q13546	6C4D	EJP	-13.60	5HX6	-11.84	1.76
Q13546	5HX6	65U	-12.14	6C4D	-11.93	0.21
Q13627	4MQ2	2C4	-9.45	4MQ1	-9.41	0.04
Q13627	4MQ1	2C3	-10.76	4MQ2	-10.14	0.62
Q13882	5DA3	58V	-9.85	5H2U	-10.14	0.29
Q13882	5H2U	1N1	-10.90	5DA3	-8.07	2.83
Q13976	4QX5	CMP	-9.99	5J48	-9.55	0.44

Q13976	5J48	6FW	-11.56	4QX5	-9.54	2.02
Q14397	4PX3	2WX	-11.89	4OHK	-11.26	0.63
Q14397	4OHK	2TE	-11.47	4PX3	-10.93	0.54
Q14416	5CNJ	52Q	-9.23	4XAS	-9.19	0.04
Q14416	4XAS	40H	-9.83	5CNJ	-9.12	0.71
Q15119	2BU8	ADP	-8.62	5M4N	-6.92	1.70
Q15119	5M4N	7FV	-10.18	2BU8	-10.83	0.65
Q16620	4AT4	T6E	-12.86	4AT5	-10.99	1.87
Q16620	4AT5	MUJ	-10.70	4AT4	-9.04	1.66
Q16651	3E16	B4C	-8.21	3FVF	-8.56	0.35
Q16651	3FVF	1JZ	-6.40	3E16	-6.50	0.10
Q16769	2AFZ	NVI	-3.81	2AFX	-4.07	0.26
Q16769	2AFX	1BN	-6.50	2AFZ	-6.29	0.21
Q1R2J4	5AAP	VNY	-9.02	5AAL	-8.24	0.78
Q1R2J4	5AAL	8L8	-9.45	5AAP	-9.11	0.34
Q21A49	5JFM	COA	-7.07	5JFM	-6.93	0.14
Q21A49	5JFM	1VU	-7.33	5JFM	-7.02	0.31
Q27895	4KH6	AU1	-10.33	4KH4	-10.53	0.20
Q27895	4KH4	ANP	-11.63	4KH6	-10.83	0.80
Q2LG68	4CB7	41G	-8.55	4CB6	-8.80	0.25
Q2LG68	4CB6	29R	-7.03	4CB7	-6.62	0.41
Q2MG72	5U1E	9CS	-9.46	5U1I	-9.59	0.13
Q2MG72	5U1I	7QM	-9.73	5U1E	-9.53	0.20
Q2RSB2	2OO5	TXD	-11.63	1U2G	-11.43	0.20
Q2RSB2	1U2G	APR	-10.60	2OO5	-9.50	1.10
Q2YY41	3F0X	53T	-10.64	3F0Q	-10.81	0.17
Q2YY41	3F0Q	52V	-10.36	3F0X	-10.23	0.13
Q385E8	2X2N	X2N	-10.21	4G3J	-10.67	0.46
Q385E8	4G3J	VNT	-10.17	2X2N	-9.86	0.31
Q389T8	2WP6	WP6	-9.66	6BU7	-6.02	3.64
Q389T8	6BU7	RD0	-8.15	2WP6	-8.17	0.02
Q3F0V8	4KGK	GTP	-9.52	4KGM	-9.59	0.07
Q3F0V8	4KGM	ATP	-9.67	4KGK	-9.92	0.25
Q3JP94	3GQT	UFO	-6.61	3EON	-6.36	0.25
Q3JP94	3EON	341	-4.75	3GQT	-4.73	0.02
Q3JRA0	3MBM	CYT	-4.37	3F0G	-4.18	0.19
Q3JRA0	3F0G	C5P	-6.25	3MBM	-6.38	0.13
Q40577	5EAT	FHP	-7.57	5EAU	-7.42	0.15
Q40577	5EAU	FFF	-7.57	5EAT	-8.37	0.80
Q460N5	3SMI	QDR	-9.36	4F1L	-9.67	0.31
Q460N5	4F1L	0RY	-7.58	3SMI	-8.03	0.45
Q46822	2VNP	DED	-6.77	1Q54	-7.03	0.26
Q46822	1Q54	BHI	-7.57	2VNP	-7.75	0.18
Q46893	1I52	CTP	-9.76	1INI	-9.49	0.27
Q46893	1INI	CDM	-9.63	1I52	-9.22	0.41
Q47066	4X69	OP0	-7.42	1IYO	-7.05	0.37
Q47066	1IYO	CEF	-7.94	4X69	-7.34	0.60
Q47NQ8	4OMR	CAA	-8.01	4JVT	-6.41	1.60

Q47NQ8	4JVT	ACO	-7.26	4OMR	-6.67	0.59
Q4D3W2	3W85	W85	-8.94	3W7O	-8.87	0.07
Q4D3W2	3W7O	W7O	-9.17	3W85	-9.09	0.08
Q4DA54	4YRR	691	-5.45	4YRM	-5.59	0.14
Q4DA54	4YRM	282	-4.93	4YRR	-4.23	0.70
Q4DA73	5Y4Q	8OF	-9.52	4YUY	-8.07	1.45
Q4DA73	4YUY	1SQ	-7.22	5Y4Q	-6.94	0.28
Q4QC75	4UXJ	TTP	-10.04	4UXH	-11.27	1.23
Q4QC75	4UXH	T5A	-11.68	4UXJ	-10.62	1.06
Q4W1X2	5VWT	FAD	-14.89	5HHF	-14.55	0.34
Q4W1X2	5HHF	62F	-15.80	5VWT	-12.46	3.34
Q51504	3OCL	CB9	-8.75	4FSF	-8.30	0.45
Q51504	4FSF	0W0	-10.95	3OCL	-9.84	1.11
Q52424	5US1	COA	-8.55	5US1	-7.67	0.88
Q52424	5US1	ACO	-7.98	5US1	-8.19	0.21
Q53752	3D4A	3GP	-8.49	3DGY	-7.42	1.07
Q53752	3DGY	2GP	-7.51	3D4A	-7.46	0.05
Q55891	2D1E	BLA	-13.59	3AJH	-9.67	3.92
Q55891	3AJH	BL3	-12.68	2D1E	-12.59	0.09
Q57977	3BBH	SFG	-9.01	3BBD	-9.05	0.04
Q57977	3BBD	SAH	-8.86	3BBH	-9.04	0.18
Q58F21	5VBR	IBI	-7.99	4KCX	-8.19	0.20
Q58F21	4KCX	1QK	-7.86	5VBR	-7.71	0.15
Q5A9A4	2WFG	ZZB	-11.39	5AGI	-11.41	0.02
Q5A9A4	5AGI	ANZ	-10.98	2WFG	-11.11	0.13
Q5AU62	1TI7	NAP	-10.83	2VUT	-10.96	0.13
Q5AU62	2VUT	NAD	-10.04	1TI7	-9.56	0.48
Q5G940	3CF9	AGI	-8.85	3DOY	-7.26	1.59
Q5G940	3DOY	2BE	-9.11	3CF9	-8.28	0.83
Q5H3Z2	5CVQ	BB2	-8.18	5CWY	-6.60	1.58
Q5H3Z2	5CWY	4WL	-6.30	5CVQ	-5.60	0.70
Q5JIZ8	3WDL	ATP	-10.48	3WDM	-7.87	2.61
Q5JIZ8	3WDM	ADN	-8.94	3WDL	-9.36	0.42
Q5NFC4	3M5P	F6P	-7.32	3Q7I	-7.81	0.49
Q5NFC4	3Q7I	6PG	-8.09	3M5P	-6.66	1.43
Q5SGX2	3HFZ	MTY	-8.43	1JJC	-8.39	0.04
Q5SGX2	1JJC	FA5	-11.93	3HFZ	-11.22	0.71
Q5SHZ3	2ZDQ	ATP	-11.36	2ZDG	-8.87	2.49
Q5SHZ3	2ZDG	ADP	-12.51	2ZDQ	-11.48	1.03
Q5SLL8	4X3L	MTA	-9.22	4X3M	-8.86	0.36
Q5SLL8	4X3M	ADN	-9.78	4X3L	-8.87	0.91
Q5SQI0	4U9Z	COA	-10.64	4B5O	-10.27	0.37
Q5SQI0	4B5O	ACO	-10.32	4U9Z	-9.57	0.75
Q5TCY1	4BTM	F8E	-7.61	4BTJ	-7.36	0.25
Q5TCY1	4BTJ	ATP	-9.51	4BTM	-8.82	0.69
Q5TLG6	6D38	GYC	-6.44	4HQC	-6.26	0.18
Q5TLG6	4HQC	CR8	-6.02	6D38	-6.77	0.75
Q5UQL3	3FBE	GDP	-9.23	3FC9	-9.84	0.61

Q5UQL3	3FC9	CTP	-10.04	3FBE	-9.14	0.90
Q5ZUA2	4BRL	GMV	-14.36	4BRF	-12.43	1.93
Q5ZUA2	4BRF	AMP	-10.51	4BRL	-10.54	0.03
Q5ZUA2	4BRH	TMV	-11.80	4BRG	-11.64	0.16
Q5ZUA2	4BRG	GNP	-11.74	4BRH	-12.21	0.47
Q62230	1OD7	SUW	-7.78	1OD9	-6.24	1.54
Q62230	1OD9	BND	-6.47	1OD7	-6.45	0.02
Q63T71	3IEW	CTP	-6.77	3IEW	-6.85	0.08
Q63T71	3IEW	CDP	-6.47	3IEW	-6.64	0.17
Q64610	5M0E	5JK	-11.76	5DLW	-12.06	0.30
Q64610	5DLW	5D5	-11.21	5M0E	-11.22	0.01
Q6B856	5LP6	71P	-7.44	5JVD	-6.80	0.64
Q6B856	5JVD	6NL	-8.42	5LP6	-8.34	0.08
Q6DE08	4C2V	YJA	-9.03	4B8L	-8.29	0.74
Q6DE08	4B8L	A0P	-11.32	4C2V	-10.59	0.73
Q6FPH0	3EEM	53V	-10.82	3EEK	-10.99	0.17
Q6FPH0	3EEK	53S	-11.46	3EEM	-11.28	0.18
Q6N089	3QCU	NKN	-6.16	3QCV	-5.86	0.30
Q6N089	3QCV	18L	-6.03	3QCU	-5.99	0.04
Q6N5P6	2I4O	ATP	-12.08	2I4N	-11.31	0.77
Q6N5P6	2I4N	5CA	-11.00	2I4O	-10.46	0.54
Q6P179	5K1V	6PX	-9.64	5J6S	-10.00	0.36
Q6P179	5J6S	6GA	-10.58	5K1V	-10.28	0.30
Q6PL18	4TYL	39O	-5.49	4TU4	-5.92	0.43
Q6PL18	4TU4	37N	-8.15	4TYL	-7.69	0.46
Q6VT83	6AWR	2AN	-6.20	4MA6	-6.69	0.49
Q6VT83	4MA6	28E	-8.75	6AWR	-6.78	1.97
Q6YMS4	5JJS	6L2	-9.18	5F3T	-8.38	0.80
Q6YMS4	5F3T	5UH	-7.49	5JJS	-7.71	0.22
Q74FS9	4GVL	AMP	-9.42	4GX1	-9.84	0.42
Q74FS9	4GX1	ADP	-9.00	4GVL	-9.07	0.07
Q79SH7	3HTH	PRL	-10.82	3HTJ	-10.54	0.28
Q79SH7	3HTJ	ET	-13.84	3HTH	-7.09	6.75
Q7BG50	2X0E	TYD	-8.38	2X0F	-8.83	0.45
Q7BG50	2X0F	TRH	-11.24	2X0E	-10.31	0.93
Q7K4Y6	4XPH	42J	-7.80	4XP9	-6.51	1.29
Q7K4Y6	4XP9	1WE	-6.27	4XPH	-7.13	0.86
Q7M537	4B8S	AMP	-8.62	1GC5	-8.02	0.60
Q7M537	1GC5	ADP	-9.22	4B8S	-9.33	0.11
Q7Z1V1	4CK8	LFD	-9.79	4C27	-9.66	0.13
Q7Z1V1	4C27	26N	-11.31	4CK8	-11.48	0.17
Q81R22	3FL8	RAR	-11.84	3JWF	-9.14	2.70
Q81R22	3JWF	5WA	-8.43	3FL8	-7.60	0.83
Q831W7	2J3L	P5A	-13.03	2J3M	-12.52	0.51
Q831W7	2J3M	ATP	-11.70	2J3L	-11.12	0.58
Q83XK4	2XLS	NAP	-9.97	2XLU	-9.58	0.39
Q83XK4	2XLU	NA7	-9.17	2XLS	-9.40	0.23
Q86W56	4BII	A8P	-10.45	4BIJ	-9.37	1.08

Q86W56	4B1J	A1R	-11.77	4B1I	-10.82	0.95
Q86X55	2Y1W	849	-10.16	5U4X	-9.58	0.58
Q86X55	5U4X	7VM	-9.80	2Y1W	-9.34	0.46
Q873X9	2IUZ	D1H	-8.12	2A3E	-8.38	0.26
Q873X9	2A3E	AMI	-6.58	2IUZ	-6.17	0.41
Q89VT8	4H2W	AMP	-10.19	4H2W	-9.48	0.71
Q89VT8	4H2W	5GP	-9.52	4H2W	-10.18	0.66
Q89ZI2	2WCA	NP6	-8.86	2J47	-8.85	0.01
Q89ZI2	2J47	GDV	-9.07	2WCA	-8.02	1.05
Q8GIQ0	2IXL	TRH	-7.86	1NYW	-8.40	0.54
Q8GIQ0	1NYW	DAU	-8.68	2IXL	-7.90	0.78
Q8II73	3B7P	SPM	-5.84	4BP3	-5.65	0.19
Q8II73	4BP3	4MN	-6.50	3B7P	-5.80	0.7
Q8IXJ6	5G4C	CNA	-14.62	4RMJ	-13.67	0.95
Q8IXJ6	4RMJ	AR6	-13.81	5G4C	-12.81	1.00
Q8KI25	3EPT	FDA	-14.19	2R0P	-11.84	2.35
Q8KI25	2R0P	FAD	-13.79	3EPT	-13.47	0.32
Q8N8S7	5NEG	8VK	-8.98	4MY6	-8.30	0.68
Q8N8S7	4MY6	3VH	-8.91	5NEG	-9.15	0.24
Q8NKB0	3WZM	ZER	-11.27	5IE4	-11.45	0.18
Q8NKB0	5IE4	36J	-11.87	3WZM	-10.76	1.11
Q8PC69	2JLB	UDM	-11.18	2XGS	-9.83	1.35
Q8PC69	2XGS	44P	-9.04	2JLB	-9.44	0.40
Q8TEK3	5MW4	5JU	-11.43	5DTQ	-11.03	0.40
Q8TEK3	5DTQ	5F6	-11.29	5MW4	-7.88	3.41
Q8WWQ0	5ENC	5QD	-5.55	5ENE	-5.83	0.28
Q8WWQ0	5ENE	5Q8	-6.64	5ENC	-6.66	0.02
Q8YUQ7	3L6J	Z90	-11.38	4JQH	-10.61	0.77
Q8YUQ7	4JQH	1MF	-12.98	3L6J	-12.21	0.77
Q92769	4LXZ	SHH	-7.34	6G3O	-6.95	0.39
Q92769	6G3O	EL8	-7.44	4LXZ	-7.39	0.05
Q92830	5TRL	SCA	-8.13	1Z4R	-7.56	0.57
Q92830	1Z4R	ACO	-7.64	5TRL	-8.48	0.84
Q93IG4	2JAS	DTP	-10.45	2JAQ	-10.80	0.35
Q93IG4	2JAQ	DCP	-11.40	2JAS	-9.91	1.49
Q93PQ0	3BFF	SFR	-7.41	3BFC	-7.35	0.06
Q93PQ0	3BFC	IM2	-7.09	3BFF	-6.65	0.44
Q965D7	3AZA	KM0	-6.49	3AZ9	-6.70	0.21
Q965D7	3AZ9	K91	-6.71	3AZA	-7.07	0.36
Q96C86	3BL7	DD1	-9.88	5OSY	-9.42	0.46
Q96C86	5OSY	AJQ	-10.98	3BL7	-11.08	0.10
Q96PN6	5IV3	LRI	-8.74	4OYA	-6.90	1.84
Q96PN6	4OYA	1VE	-11.03	5IV3	-9.96	1.07
Q96RI1	5Q0J	9KV	-13.70	5Q0X	-12.30	1.40
Q96RI1	5Q17	9MV	-11.44	5Q0J	-8.27	3.17
Q96RI1	5Q14	9MM	-13.07	5Q14	-12.72	0.35
Q96RI1	5Q0X	643	-12.72	5Q17	-10.01	2.71
Q97ZE2	2J4J	ACP	-10.84	2J4J	-10.77	0.07

Q97ZE2	2J4J	4TC	-14.03	2J4J	-12.54	1.49
Q980Q4	1VST	PRP	-8.42	3G6W	-8.44	0.02
Q980Q4	3G6W	HSX	-7.28	1VST	-7.32	0.04
Q98SW5	5KZV	HCD	-10.51	5KZY	-9.46	1.05
Q98SW5	5KZY	CY8	-12.30	5KZV	-12.13	0.17
Q99418	4JMO	JAF	-7.07	4L5M	-7.20	0.13
Q99418	4L5M	HRC	-7.06	4JMO	-6.68	0.38
Q9AIU0	2UXI	G50	-7.82	2UXP	-7.90	0.08
Q9AIU0	2UXP	CLM	-6.85	2UXI	-6.75	0.10
Q9BH77	2OP1	8PC	-10.24	2OP0	-9.43	0.81
Q9BH77	1NHW	TCC	-7.68	2OP1	-8.48	0.80
Q9BH77	2OP0	7PC	-9.71	2OP1	-10.06	0.35
Q9BH77	2OP1	8PC	-10.24	1NHW	-8.95	1.29
Q9BQ65	5V1M	U5P	-7.60	6D31	-7.64	0.04
Q9BQ65	6D31	A	-7.79	5V1M	-8.06	0.27
Q9BY41	4RN0	L6G	-6.32	3SFH	-6.46	0.14
Q9BY41	3SFH	1DI	-10.37	4RN0	-7.46	2.91
Q9BYW2	5JLB	SAH	-9.91	5LSX	-10.35	0.44
Q9BYW2	5LSX	76O	-12.29	5JLB	-11.95	0.34
Q9BZP6	3RM8	RM8	-10.62	3RM4	-9.22	1.40
Q9BZP6	3RM4	3RM	-8.85	3RM8	-7.91	0.94
Q9F663	3RXX	NPB	-6.20	6B1J	-7.11	0.91
Q9F663	6B1J	C8V	-8.88	3RXX	-8.05	0.83
Q9GPQ4	2FF2	IMH	-10.20	1HP0	-10.62	0.42
Q9GPQ4	1HP0	AD3	-10.17	2FF2	-9.69	0.48
Q9GZ28	1XMZ	CRK	-5.89	1XQM	-5.88	0.01
Q9GZ28	1XQM	CH6	-6.58	1XMZ	-6.26	0.32
Q9H0U4	5SZH	GNP	-14.10	4I1O	-12.29	1.81
Q9H0U4	4I1O	GDP	-11.65	5SZH	-11.76	0.11
Q9H2K2	5FPG	Q28	-15.92	4W5I	-11.95	3.97
Q9H2K2	4W5I	3GX	-11.49	5FPG	-10.34	1.15
Q9H4A3	5WE8	A7Y	-11.77	5TF9	-10.41	1.36
Q9H4A3	5TF9	7AV	-10.48	5WE8	-10.04	0.44
Q9H8M2	4Z6H	4L2	-8.85	4XY8	-8.27	0.58
Q9H8M2	4XY8	43U	-7.92	4Z6H	-7.52	0.40
Q9H8M2	5TWX	7P7	-8.58	5E9V	-7.87	0.71
Q9H8M2	5E9V	5L0	-9.17	5TWX	-8.97	0.20
Q9H9B1	3MO2	E67	-8.11	3MO0	-7.39	0.72
Q9H9B1	3MO0	E11	-6.69	3MO2	-7.48	0.79
Q9HU22	4B42	942	-6.73	3ZLL	-7.08	0.35
Q9HU22	3ZLL	4WF	-7.90	4B42	-7.96	0.06
Q9K2N0	5LCH	6TU	-9.45	5LCF	-8.29	1.16
Q9K2N0	5LCF	6TJ	-9.68	5LCH	-8.69	0.99
Q9KTW3	4QNE	NAJ	-10.69	4X3Z	-11.11	0.42
Q9KTW3	4X3Z	NAD	-11.45	4QNE	-10.78	0.67
Q9KU37	2OXN	OAN	-7.11	3GS6	-6.82	0.29
Q9KU37	3GS6	NP6	-7.11	2OXN	-7.03	0.08
Q9M9P3	2ICX	UTP	-9.77	2ICY	-9.33	0.44

Q9M9P3	2ICY	UPG	-10.50	2ICX	-10.88	0.38
Q9NAV7	5LK9	LZ1	-5.60	5LLZ	-5.38	0.22
Q9NAV7	5LLZ	0CT	-5.27	5LK9	-5.29	0.02
Q9NAV8	4HSX	BML	-5.68	3LB3	-5.12	0.56
Q9NAV8	3LB3	4CH	-4.98	4HSX	-5.48	0.50
Q9NPB1	4MWO	2E02	-12.12	4MWO	-11.46	0.66
Q9NPB1	4YIK	2O2	-11.40	4MWO	-10.53	0.87
Q9NPB1	4MWO	2E02	-12.12	4NFL	-11.12	1.00
Q9NPB1	4NFL	2JW	-11.10	4YIK	-10.46	0.64
Q9NR97	4R07	UCG	-10.15	4R09	-9.45	0.70
Q9NR97	4R09	06S	-9.37	4R07	-9.02	0.35
Q9NRW1	2FFQ	GSP	-13.44	2FE4	-12.00	1.44
Q9NRW1	2FE4	GDP	-11.84	2FFQ	-11.99	0.15
Q9NWT6	4AI8	DZA	-5.16	5OP8	-4.32	0.84
Q9NWT6	5OP8	A1H	-8.12	4AI8	-8.38	0.26
Q9NZJ5	4X7K	3Z3	-15.48	4X7J	-14.45	1.03
Q9NZJ5	4X7J	3Z1	-13.07	4X7K	-12.45	0.62
Q9QLL6	5EFA	MGT	-7.48	5EFC	-7.22	0.26
Q9QLL6	5EFC	GTP	-8.65	5EFA	-8.60	0.05
Q9R194	4I6G	FAD	-12.71	4MLP	-11.24	1.47
Q9R194	4MLP	2CX	-10.24	4I6G	-9.70	0.54
Q9R1E6	5OLB	6ZO	-10.63	5LIA	-10.85	0.22
Q9R1E6	5LIA	6XN	-12.39	5OLB	-10.75	1.64
Q9RA63	4LJ7	MNT	-9.62	4FD2	-10.32	0.70
Q9RA63	4FD2	ADP	-10.41	4LJ7	-9.85	0.56
Q9RC92	2AHG	UCD	-8.91	2FV1	-8.55	0.36
Q9RC92	2FV1	GAD	-6.63	2AHG	-7.48	0.85
Q9UGL1	5FZL	UOI	-8.02	5A3T	-7.60	0.42
Q9UGL1	5A3T	MMK	-8.29	5FZL	-8.20	0.09
Q9UHI6	3B7G	ANP	-8.64	2OXC	-8.60	0.04
Q9UHI6	2OXC	ADP	-8.27	3B7G	-7.42	0.85
Q9UJ71	5G6U	YJM	-6.04	4N32	-5.83	0.21
Q9UJ71	4N32	2F8	-4.99	5G6U	-4.93	0.06
Q9UKV8	3QX9	ATP	-6.98	3LUD	-6.82	0.16
Q9UKV8	3LUD	AMP	-7.12	3QX9	-7.17	0.05
Q9UM73	4FNZ	NZF	-11.16	4FOB	-8.68	2.48
Q9UM73	4FOB	0US	-9.62	4FNZ	-8.39	1.23
Q9UVX3	4CAX	646	-9.82	4CAV	-9.59	0.23
Q9UVX3	4CAV	2XQ	-10.84	4CAX	-9.06	1.78
Q9UVX3	4CAW	P3U	-11.10	5T5U	-10.22	0.88
Q9UVX3	5T5U	75T	-11.44	4CAW	-10.88	0.56
Q9UYR9	1YR7	GSP	-9.64	1YRB	-9.99	0.35
Q9UYR9	1YRB	GDP	-10.01	1YR7	-9.28	0.73
Q9V2G1	5HJK	SAH	-9.86	5HJM	-10.34	0.48
Q9V2G1	5HJM	MTA	-9.46	5HJK	-9.00	0.46
Q9WUL6	5T8Q	76Y	-6.49	4G3E	-7.04	0.55
Q9WUL6	4G3E	0WC	-11.48	5T8Q	-8.75	2.73
Q9X1X0	4O5G	2PV	-7.84	4O4K	-6.28	1.56

Q9X1X0	4O4K	2PK	-8.39	4O5G	-7.87	0.52
Q9Y3R4	2F11	IEM	-6.65	1VCU	-7.29	0.64
Q9Y3R4	1VCU	DAN	-7.88	2F11	-6.79	1.09
Q9Y3Z3	6DWK	HFD	-10.10	4TNY	-9.86	0.24
Q9Y3Z3	4TNY	DGT	-10.32	6DWK	-10.32	0.00
Q9Y6E0	4QO9	627	-9.45	4QMS	-8.57	0.88
Q9Y6E0	4QMS	1N1	-9.14	4QO9	-7.75	1.39
Q9ZAG3	5YQT	3ZS	-3.95	5JPU	-3.98	0.03
Q9ZAG3	5JPU	3ZQ	-4.89	5YQT	-5.45	0.56
U6AVY0	4OOM	2U3	-9.50	4OOL	-9.83	0.33
U6AVY0	4OOL	2U2	-8.91	4OOM	-9.03	0.12

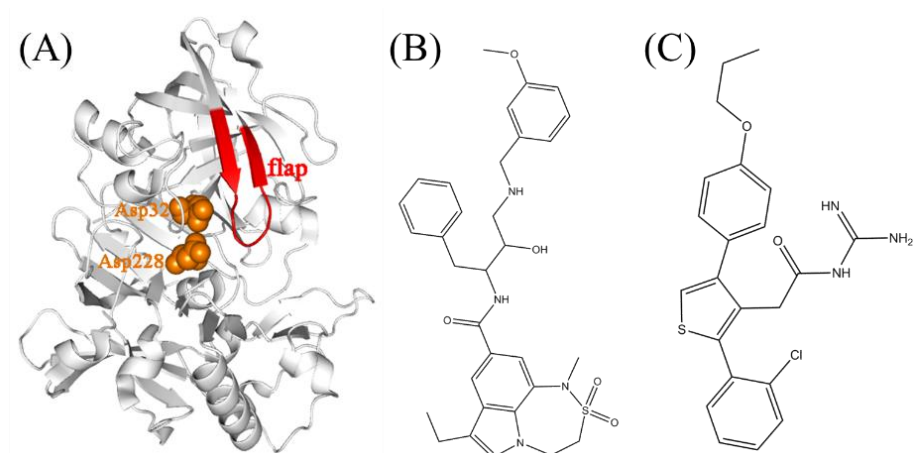


图 S3.1 BACE1 蛋白的晶体结构 (2QP8, A) 以及抑制剂复合结构 (2WF1, B 和 2QU3, C)

Figure S3.1 The crystal structures of BACE1 (2QP8, A) and inhibitors in structure 2WF1 (B) and 2QU3 (C).

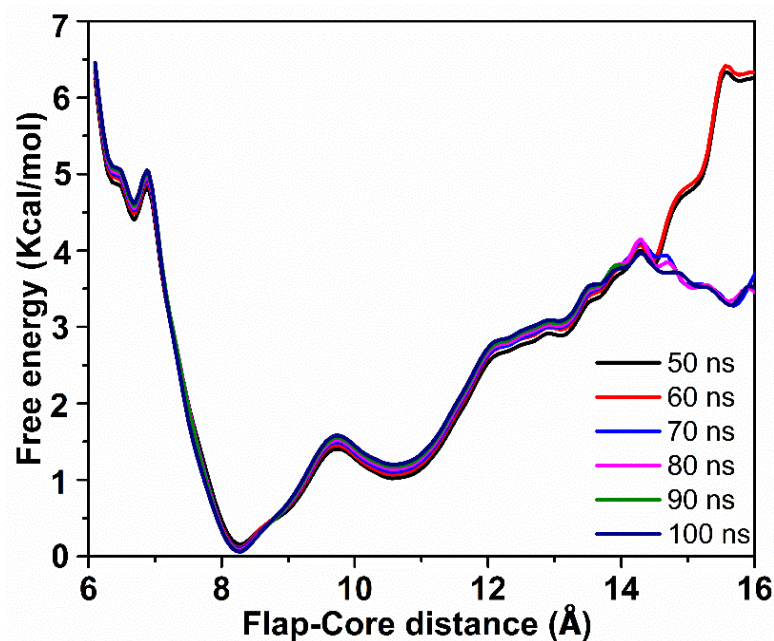


图 S3.2 BACE1 空蛋白体系的 ossPTMetaD 模拟收敛性。不同时间段以 Flap-Core 距离为反应坐标计算的一维自由能图景。

Figure S3.2 The convergence of ossPTMetaD simulation for apo BACE1. One-dimensional free energy profile as function of Flap-Core distance calculated with trajectories of different simulation time of ossPTMetaD.

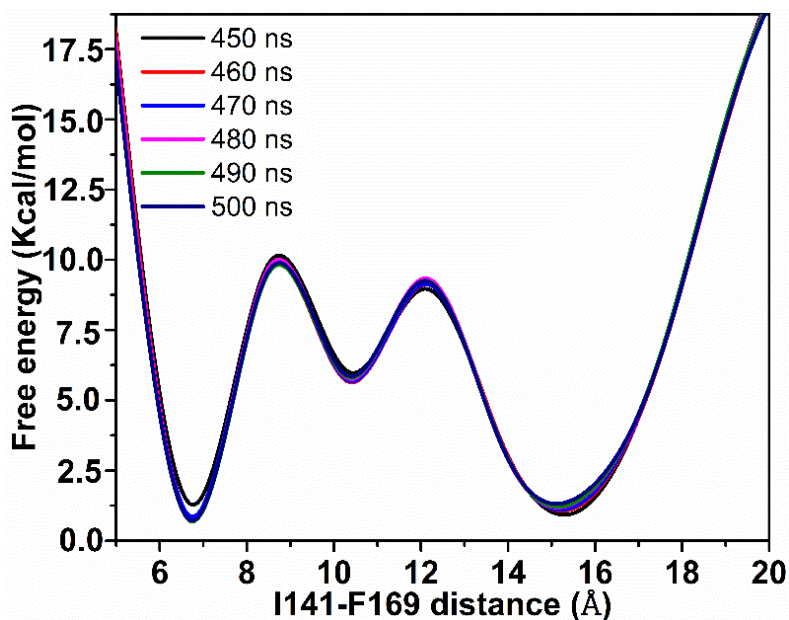


图 S3.2 p38 α 激酶体系的 ossPTMetaD 模拟收敛性。不同时间段以 F169-I141 距离为反应坐标计算的一维自由能图景。

Figure S3.3. The convergence of ossPTMetaD simulation for p38 α kinase. One-dimensional free energy profile as function of F169-I141 distance calculated with trajectories of different simulation time of ossPTMetaD.

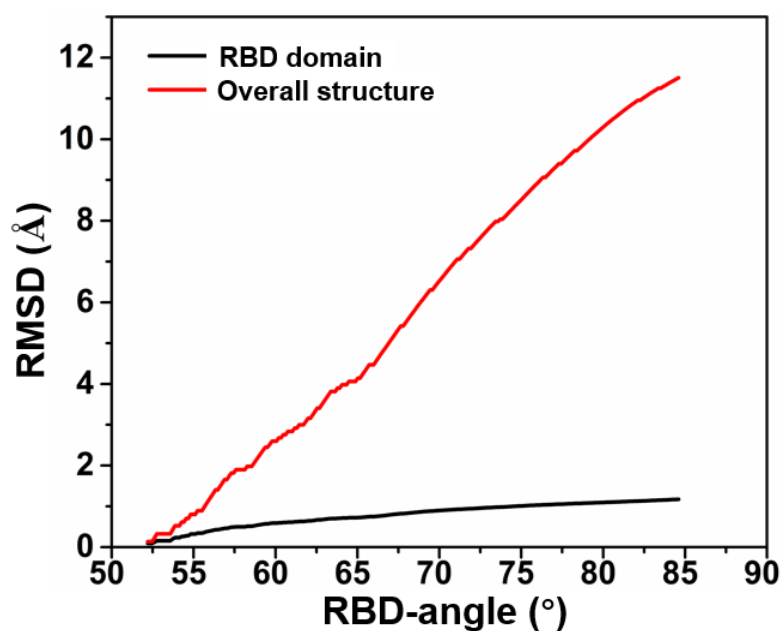


图 S4.1 RBD 和整体结构的 RMSD 值随 RBD 角度的变化。

Figure S4.1 The RMSD values of RBD domain and overall structure along with the RBD-angle.

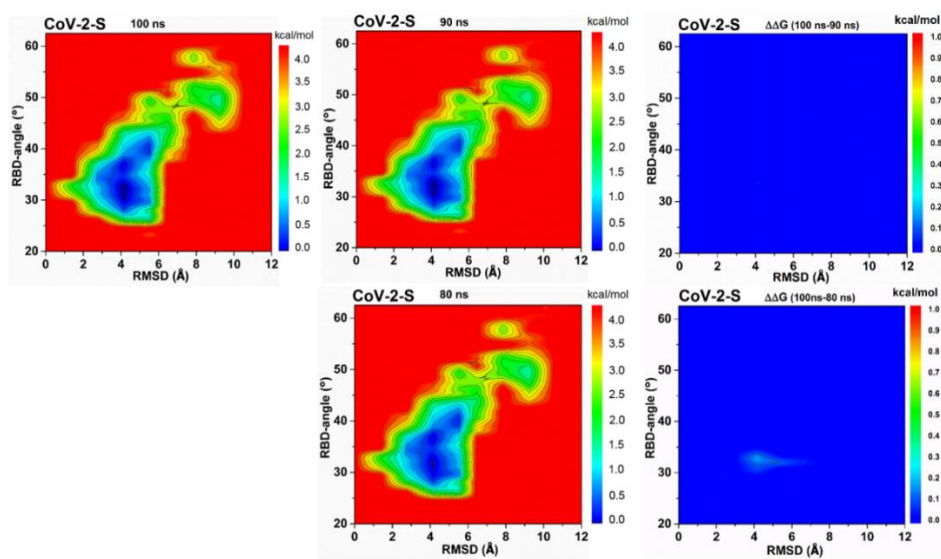


图 S4.2 CoV-2-S 在不同模拟时长下的自由能图景。80 ns 和 90 ns 与 100 ns 的自由能差异在右栏。

Figure S4.2 Free energy profiles of CoV-2-S calculated with trajectories of different simulation time (80, 90 and 100 ns). The free energy differences ($\Delta\Delta G$) between 100 ns and 90, 80 ns, respectively, were shown on the right of each column.

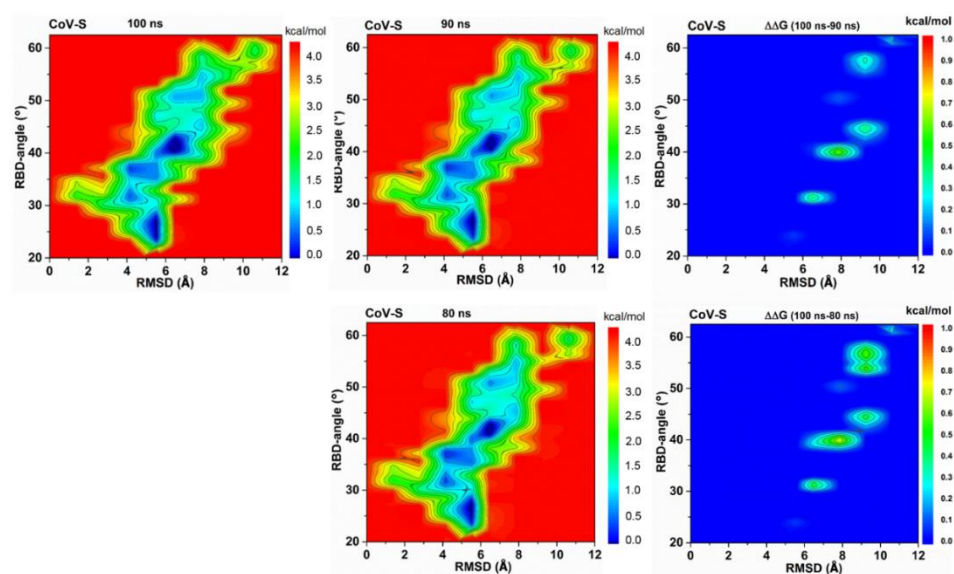


图 S4.3 CoV-2-S 在不同模拟时长下的自由能图景。80 ns 和 90 ns 与 100 ns 的自由能差异在右栏。

Figure S4.3 Free energy profiles of CoV-S calculated with trajectories of different simulation time (80, 90 and 100 ns). The free energy differences ($\Delta\Delta G$) between 100 ns and 90, 80 ns, respectively, were shown on the right of each column.

致 谢

时维五月，春芳已归，榴花未住，天气如秋，四六无暑。鹤雾凝远山之翠，烟云析日光之髓。栖高蝉而响疏桐，舞蜂蝶而度暗香。幽木萃天地之精气，潺水通盘古之经络。浮华苍宇，几处楼阁，谈笑风生，络绎鸿儒。久居静室，不觉物转星移；乍见春光，匆匆挥手言别。来时桂花香、锦鲤欢，奈何年年负花期。柳絮零落，悲喜才上心头；月影成双，聚散终是人生。谨以拙文，感师朋亲友之恩，山水一程，三生有幸。

虽资质愚钝，蒙恩师不弃，添列朱门，不吝言传身教。恩师治学之严谨，授业之尽瘁，处世之平和，学识之渊博，待人之谦逊，浸于天地，惠及及人。恩师常嘱余树正心，求正理，君子于天地间，有才无忧，有能独立。杨时少智且程门立雪，明帝九尊犹师敬桓荣，此情春晖，寸草难报。

常忆吾师进安，受恩一两载，已别三四年，虽跨洋求学，然指点可及。王师博学，室内外皆知，冥冥之志，恢恢其才。游园畅言，王师句句珠玑、指点迷津。徐师志建，温和教业，循循善诱，如沐春风。欲谢诸师，邵强、鑫贲、婷婷、李波、张勇等门中导师，何敏、林巍等学府导师，恐难具陈，不以表闻。

师门济济，皆良师益友。师兄照强，有求必应，师姐玉琪，屡屡解惑，师姐桂敏之达学，师兄正诞之严谨，师姐米娜之热忱，同窗王宇之立行，师弟海国之勤学，禹龙之研学，延青之好学，家鑫之笃学，师妹凯洁之阔达，乐云之关微，皆余钦佩所向，吾辈榜样，其余诸君，恐难具陈，亦不以表闻。室友三人，为宁、福、昊，子曰“三人行，必有我师焉”，此三人亦常为我师，师宁之明理，乐福之豁达，敬昊之志远。亦交长青、德建、尊蕴、丽萍、美娜、叙潼、小红甚欢，实乃吾幸。

今学至博士，窃以为父居启蒙。吾少时弃学，不顾惩戒，吾父虽为生计所累，然日常剥时教学，动情晓理，言传身教。终以菲薄之智，入足煌煌学府。叩谢天恩，父母康健，乌鸟私情，愿乞回报。有姊名兰，如父身教，如母顾行，如友知心。吾闻天命，知遇一女，陌上花开，与君同行。

自觉天资不聪，言语不善，气运不佳，容貌不秀，苟活至今，幸遇贵人无数。天干地支，轮回有期，求学五载，不似少年，研路漫长，自觉不可失初心。是以

于师长常存崇敬之心，故日久不忘其嘱；于课业常存敬畏之心，故相待不以狡黠；
于长技常存谦卑之心，故学以维新；于亲友常存智信之心，故行而无虑；于世人
常存同理之心，故不辩善恶。

2021 年 5 月

作者简历及攻读学位期间发表的学术论文与研究成果

作者简历:

2012 年 9 月——2016 年 6 月, 在武汉大学药学院获得学士学位。

2016 年 9 月——2021 年 6 月, 在中国科学院上海药物研究所攻读博士学位。

获奖情况:

2014-2015 年 “武汉大学优秀学生”

2019 年 中国科学院上海药物所 “优秀共青团员”

2020 年 中国科学院上海药物所 “新冠疫情科研攻关优秀奖”

2019-2020 年 “中国科学院大学三好学生”

2021 年 “上海市优秀毕业生”

参加学术会议:

2019 年于南京参加第 14 届全国计算(机)化学学术会议

2018 年于杭州参加中国化学会第 31 届学术年会

2019 年于上海参加第 15 届全国计算(机)化学学术会议

2021 年于珠海参加中国化学会第 32 届学术年会

已发表(或正式接受)的学术论文:(书写格式同参考文献)

1. Ren WM¹, Zhang JJ¹, **Peng C¹**, Xiang HJ, Chen JJ, Peng CY, Zhu WL, Huang RM, Zhang HY, Hu YH. Fluorescent Imaging of β -Amyloid Using BODIPY Based Near-Infrared Off-On Fluorescent Probe. *Bioconjugate Chem.* **2018**, 29, 3459-3466. **(Co-first author)**
2. **Peng C**, Wang JA, Yu YQ, Wang GM, Chen ZQ, Xu ZJ, Cai TT, Shao Q, Shi JY, Zhu WL. Improving the accuracy of predicting protein-ligand binding-free energy with semiempirical quantum chemistry charge. *Fut. Med. Chem.* **2019**, 11: 303-321.
3. **Peng C¹**, Atilaw Y¹, Wang JA, Xu ZJ, Poongavanam V, Shi JY, Kihlberg J, Zhu WL, Erdélyi M. Conformation of the Macrocyclic Drug Lorlatinib in Polar and Nonpolar Environments: A MD Simulation and NMR Study. *Acs omega* **2019**, 4: 22245-22250. **(Co-first author)**
4. Shi YL¹, Zhang XB¹, Mu KJ¹, **Peng C¹**, Zhu ZD, Wang XY, Yang YQ, Xu ZJ, Zhu WL. D3Targets-2019-nCoV: a webserver for predicting drug targets and for multi-

- target and multi-site based virtual screening against COVID-19. *Acta Pharmaceutica Sinica B*. **2020**, 10: 1239-1248. (Co-first author)
5. Peng C, Wang JA, Xu ZJ, Cai TT, Zhu WL. Accurate Prediction of Relative Binding Affinities of a Series of HIV-1 Protease Inhibitors Using Semi-Empirical Quantum Mechanical Charge. *J. Comput. Chem.* **2020**, 41: 1773-1780.
 6. Peng C¹, Zhu ZD¹, Shi YL, Wang XY, Mu KJ, Yang YQ, Zhang XB, Xu ZJ, Zhu WL. Computational Insights into the Conformational Accessibility and Binding Strength of SARS-CoV-2 Spike Protein to Human Angiotensin-Converting Enzyme 2. *J Phys Chem Lett.* **2020**, 11: 10482-10488 (Co-first author)
 7. Peng C, Wang JA, Shi YL, Xu ZJ, Zhu WL. Increasing the Sampling Efficiency of Protein Conformational Change by Combining a Modified Replica Exchange Molecular Dynamics and Normal Mode Analysis. *J. Chem. Theory Comput.* **2020**, 17: 13-28.
 8. Peng C¹, Zhang XB¹, Xu ZJ, Chen ZQ, Yang YQ, Cai TT, Zhu WL. D3PM: A Comprehensive Database for Protein Motions Ranging from Residue to Domain. *BMC Bioinformatics*, **2020**, In revision. (Co-first author)
 9. 彭诚, 吴乐云, 徐志建, 朱维良. 副本交换分子动力学的发展现状与展望[J]. 化学进展, **2021**, DOI: 10.7536/PC201226.
 10. Wu LY¹, Peng C¹, Xu ZJ, Zhu WL. Predicting the Potential Effect of E484K Mutation on the Binding of 28 Antibodies to the Spike Protein of SARS-CoV-2 by Molecular Dynamics Simulation and Free Energy Calculation. ChemRxiv, <https://chemrxiv.org/s/ddad6665b4cf0595bd4e>
 11. Chen JZ, Peng C, Wang JA, Zhu WL. Exploring Molecular Mechanism of Allosteric Inhibitor to Relieve Drug Resistance of Multiple Mutations in HIV-1 Protease by Enhanced Conformational Sampling. *Proteins* **2018**, 86: 1294-1305.
 12. Wang Y, Peng C, Wang GM, Xu ZJ, Luo YF, Wang JA, Zhu WL. Exploring binding mechanisms of VEGFR2 with three drugs lenvatinib, sorafenib, and sunitinib by molecular dynamics simulation and free energy calculation. *Chem. Biol. Drug Des.* **2019**, 93: 934-948.
 13. Zheng ZG, Zhang X, Liu XX, Jin XX, Dai LZ, Cheng HM, Jing D, Thu P, Li HY, Zhu J, Liu C, Xue B, Li Y, Chen LG, Peng C, Zhu WL, Wang L, Liu JL, Li HJ, Li P, Xu XJ. Inhibition of HSP90 β Improves Lipid Homeostasis Disorder by Promoting Mature SREBPs Ubiquitin-proteasome Degradation. *Theranostics*. **2019**,

- 9: 5769-5783.
14. Chen ZQ, Zhang XB, **Peng C**, Wang JA, Xu ZJ, Chen KX, Shi JY, Zhu WL. D3Pockets: A Method and Web Server for Systematic Analysis of Protein Pocket Dynamics. *J. Chem. Inf. Model.* **2019**, 59: 3353-3358.
 15. Wang JA, **Peng C**, Yu YQ, Chen ZQ, Xu ZJ, Cai TT, Shao Q, Shi JY, Zhu WL. Exploring Conformational Change of Adenylate Kinase by Replica Exchange Molecular Dynamic Simulation. *Biophys. J.* **2020**, 118: 1009-1018.
 16. Wang XY, **Peng C**, Yang YQ, Shi YL, Zhu ZD, Mu KJ, Xu ZJ, Zhu WL. Research progress of anti-pathogenic coronavirus active compounds and their potential target proteins. *Acta Pharm Sin.* **2020**.
 17. Zhang QH, Smalley A, Zhu ZD, Xu ZJ, **Peng C**, Chen ZQ, Yao GM, Shi JY, Zhu WL. Computational study of the substituent effect of halogenated fused-ring heteroaromatics on halogen bonding. *J Mol. Model.* **2020**, 26: 270.
 18. Xu ZJ, **Peng C**, Shi YL, Zhu ZD, Mu KJ, Wang XY, Zhu WL. Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. *BioRxiv.* **2020**, DOI: 10.1101/2020.01.27.921627.
 19. Wang XY, Zhu ZD, Yang YQ, Zhang XB, Mu KJ, Shi YL, **Peng C**, Xu ZJ, Zhu WL. Ligand-based approach for predicting drug targets and for virtual screening against COVID-19. *Brief Bioinform.* **2021**, DOI: 10.1093/bib/bbaa422.
 20. Abula A, Xu ZJ, Zhu ZD, **Peng C**, Chen ZQ, Zhu WL, Aisa, HA. The substitution effect of the trifluoromethyl group on the bioactivity in medicinal chemistry: statistical analysis and energy calculations. *J. Chem. Inf. Model.* **2020**, 60: 6242–6250.
 21. Wang XY; Zhang XB, **Peng C**, Shi YL, Li HY, Xu ZJ, Zhu WL. Most Distal Mutations Affect Enzyme Activity. *J. Chem. Inf. Model.* *Accept.*
 22. Mu KJ, **Peng C**, Yang YQ, Shi YL, Zhang Y, Xu ZJ, Zhu WL. Combining custom scoring function, MM/GBSA calculations, and anti-CoV ligand similarity analysis to identify a potential TMPRSS2 inhibitor. *J Proteome Res.* Under review.

申请或已获得的专利：（书写格式同参考文献）

1. 徐志建; 姚航平; 许叶春; 吴南屏; **彭诚**; 卢翔云; 苏海霞; 程林芳; 刘福民;

吴海波; 靳昌忠; 吴志刚; 沈敬山; 朱维良; 蒋华良; 李兰娟; 奈非那韦在制备防治新冠肺炎药物中的应用, 2020-2-17, 中国, 202010096827.2.

2. 朱维良; 徐志建; 石禹龙; 张鑫贲; 穆凯洁; **彭诚**; 朱正诞; 王晓宇; 杨延青; D3Docking 抗新冠肺炎药物靶标预测及虚拟筛选软件。国家版权局, 计算机软件著作权登记, 登记号: 2020SR0414683, 证书号: 软著登字第 5293379 号.

3. 朱维良; 徐志建; 杨延青; 朱正诞; 王晓宇; 张鑫贲; 穆凯洁; 石禹龙; **彭诚**; D3Similarity 抗新冠肺炎药物靶标预测及虚拟筛选软件。国家版权局, 计算机软件著作权登记, 登记号: 2021SR0218009, 证书号: 软著登字第 7399436 号.