arXiv:2305.05025v1 [physics.comp-ph] 8 May 2023

溶剂中蛋白质的多尺度混合建模：以 SARS-CoV2 刺突蛋白为例，研究格子玻尔兹曼与全原子分子动力学的耦合

M. Lauricella[1],L. Chiodo*[2]、F. Bonaccorso[3],M. Durve[4],A. Montessori[5],A. Tiribocchi[1],A. Loppini[2],S. Filippi[2], 和 S. Succi[4]

[1]意大利国家研究委员会计算应用研究所 IAC-CNR，邮编 00185，意大利罗马。[2]生物医学大学校园工程系，邮编 00128，意大利罗马。[3]罗马第二大学物理系及意大利国家核物理研究院，邮编 00133，意大利罗马。[4]意大利技术研究院萨皮恩扎生命纳米与神经科学中心 – IIT，邮编 00161，意大利罗马。[5]罗马第三大学工程系，维托·沃尔泰拉街62号，邮编 00146，意大利罗马。

2025年4月4日

摘要

生理溶剂流围绕生物结构流动，并由此触发其中的集体运动。典型例子包括病毒与宿主细胞相互作用以及溶剂介导的变构调节。本研究介绍一种多尺度方法，将格子玻尔兹曼流体动力学（用于模拟溶剂流）与全原子分子动力学（用于模拟蛋白质）相结合，以模拟流动与分子之间的功能性相互作用。作为应用实例，我们研究了 SARS-CoV-2 病毒刺突糖蛋白与周围溶剂的相互作用，并将周围溶剂建模为介观流体。通过适当的观测量描述野生型刺突蛋白和 Alpha变体在隐式溶剂中的平衡性质。我们将介观溶剂描述与全原子溶剂模型进行批判性比较，以量化介观流体描述的优势与局限。

1 引言

过去几十年来，方法与技术不断进步，为计算工具成功助力抗击传染病铺平了道路，使人们能够对生物分子进行计算机模拟并开展药物设计。此外，图形处理器（GPU）带来了显著的技术进步，大幅提升了处理大型系统的计算能力。在此背景下，我们采用一种多尺度建模方法，将介观溶剂表示与分子动力学相结合。最终目标是提出一种高效的生物物理计算策略，提升模拟能力并在原子级理解生物机制，同时避免对溶剂进行原子级描述所需的计算开销。

如今，分子动力学（MD）方法已展现出巨大潜力，能够在原子级表征多种蛋白质活动背后的生物学机制。计算生物物理学中的突出实例包括：近期对整个细胞器——一种来自紫色细菌的光合色素体囊泡——进行的模拟 [1] ，以及 DE Shaw研究组对 N-甲基-D-天冬氨酸（NMDA）神经受体的研究 [2]。截至目前，科研人员已投入大量精力，通过计算机模拟研究 SARS-CoV-2蛋白质的分子行为，既用于药物再利用，也用于抗体设计 [3]。例如，标准分子动力学模拟已被用于估算刺突蛋白与人血管紧张素转换酶2（ACE2）受体相互作用时的结合自由能 [4, 5, 6, 7] 以及相互作用评分 [8]。针对 SARS-CoV-2蛋白质的研究，尤其是对其刺突蛋白的高度关注，在某种程度上凸显了生物信息学和生物物理学计算工具在医学和药物发现领域的优势与局限[9, 10] ，也使这些工具受到公众的关注。

*电子邮箱： m.lauricella@iac.cnr.it；通讯作者

这里以广为人知且受到深入研究的刺突蛋白为测试案例，阐明混合多尺度描述方案的优势与不足。事实上，尽管全原子分子动力学描述是恰当刻画蛋白质功能生化本质的首选方法，但其应用的主要问题之一在于生物学机制涉及的时间尺度很长，而且从统计力学角度看，这些机制具有罕见事件的特征。通常，与生物学机制的变构响应和功能响应相关的四级结构运动发生在数微秒的时间尺度上，超出了当前标准高性能计算能力所能达到的范围，难以获得具有统计学意义的描述 [11, 12], ，即便采用针对图形处理器集群优化的代码也是如此 [13, 14]。一种可能的解决方案是采用增强采样技术 [15], ，但大多数真实机制涉及的庞大模拟规模难以轻易纳入考量。因此，过去三十年来，粗粒化模型的发展展现出克服这些限制的巨大潜力[16, 17]。粗粒化策略通常旨在减少蛋白质结构及其水性溶剂的细节。进行这种简化时必须格外谨慎，以便在适当描述蛋白质结构和功能所必需的地方保留详细信息[18, 19, 20, 21]。

另一种方法基于不同工具的耦合，从多尺度视角描述不同时间尺度和长度尺度的共存。这种多尺度方法一旦得到验证，便可成为强大的生物物理计算工具，并有望应用于药物设计和生物医学领域 [22, 23, 24]。

本研究采用基于格子玻尔兹曼（LB）方法的多尺度描述来模拟水溶液中的溶剂流体，并结合全原子分子动力学（MD）描述蛋白质结构。迄今为止，已有多种耦合 LB/MD 的多尺度方法 [25, 26, 27, 28, 29]。在本研究中，通过在两种描述的重叠区域引入斯托克斯摩擦项，实现 LB 速度场与 MD 对象的耦合，从而构建多尺度描述 [25]。

我们的测试案例是广为人知且特征明确的 SARS-CoV-2 刺突蛋白 S，这是一种高度糖基化、锚定于病毒膜上的蛋白质。它由三条链构成，每条链均具有相同的一级序列，包含 1200 多个氨基酸，其中 1146 个构成胞外结构域。三聚体的每条链由两个片段组成：与 ACE2 相互作用的受体结合片段 S1（含有受体结合域（RBD）），以及融合片段 S2 [30]。S 蛋白在第 686 位残基处被一种类弗林蛋白酶切割为 S1 和 S2 片段 [31], ，从而启动膜融合过程。我们还研究了关注变异株 202012/01（谱系 B.1.1.7），该变异株也通常称为 Alpha变体（α-刺突蛋白）。

平衡态性质（如特定位点的刚性和弹性）对于病毒蛋白质 [30, 32], 的结合及其他功能活动至关重要，因此必须在多尺度模型中得到恰当描述。我们将通过全原子分子动力学获得的 LB/MD 数据与平衡态性质进行比较，以评估多尺度描述的质量和效率。水分子对蛋白质结构和功能的关键作用也得到了强调 [33]。展望未来，我们的多尺度方法有望克服当前显式溶剂原子级描述所受的统计采样局限，尤其可提供一种独特工具，用于研究大幅构象变化以及溶剂流动对动力学的影响。然而，尽管混合方法再现了蛋白质结构和运动的主要特征，但要恰当地再现其柔性和涨落，仍需进行大量改进；这些性质关系到正确的构象采样，而构象采样又与蛋白质的正常功能密切相关。

2 方法学

2.1 模拟系统

我们使用冷冻电镜数据 [34, 35] 构建了野生型和突变型蛋白质的两个模型，以下分别称为刺突蛋白（S）和 α-刺突蛋白（αS）。我们以融合前状态刺突蛋白的 6vsb.pdb 结构 [12] 为起点，并根据 6m17.pdb 结构 [35] 补全受体结合域（RBD）中缺失的环区。在这些模拟中，我们保留了 6vsb.pdb 结构中的聚糖。 α-刺突蛋白以野生型为模板，使用 I-Tasser [36], 建模，并在序列 [37] 中引入三处缺失突变：ΔH69/ΔV70 和 ΔY144，以及六处突变：N501Y、A570D、P681H、T716I、S982A 和 D1118H（见图1）。质子化状态通过 Playmolecule 网络服务器 [38], 计算：基于 PROPKA 3.1 [39] 确定蛋白质 pKa 值，并基于 PDBTOPQR 2.1 [40] 优化蛋白质以形成有利的氢键。LB/MD 系统的模拟盒为边长 19.2 nm 的立方体，包围着全原子糖基化蛋白质（53k 个原子）。全原子模拟采用相同的盒子尺寸（676k 个原子）。通过可视化分子动力学（VMD）[41] 中的 Solvate 和 Ionize 插件构建模拟盒并使其电中性（100 nM 溶液）。

2.2 耦合格子玻尔兹曼与分子动力学（LB/MD）

LB/MD模拟使用LAMMPS（2020年3月3日稳定版） [42] 在集群上进行，该集群基于 Intel(R) Xeon(R) Platinum 8160 处理器，每个计算节点配备 24 个核心和两颗 CPU。

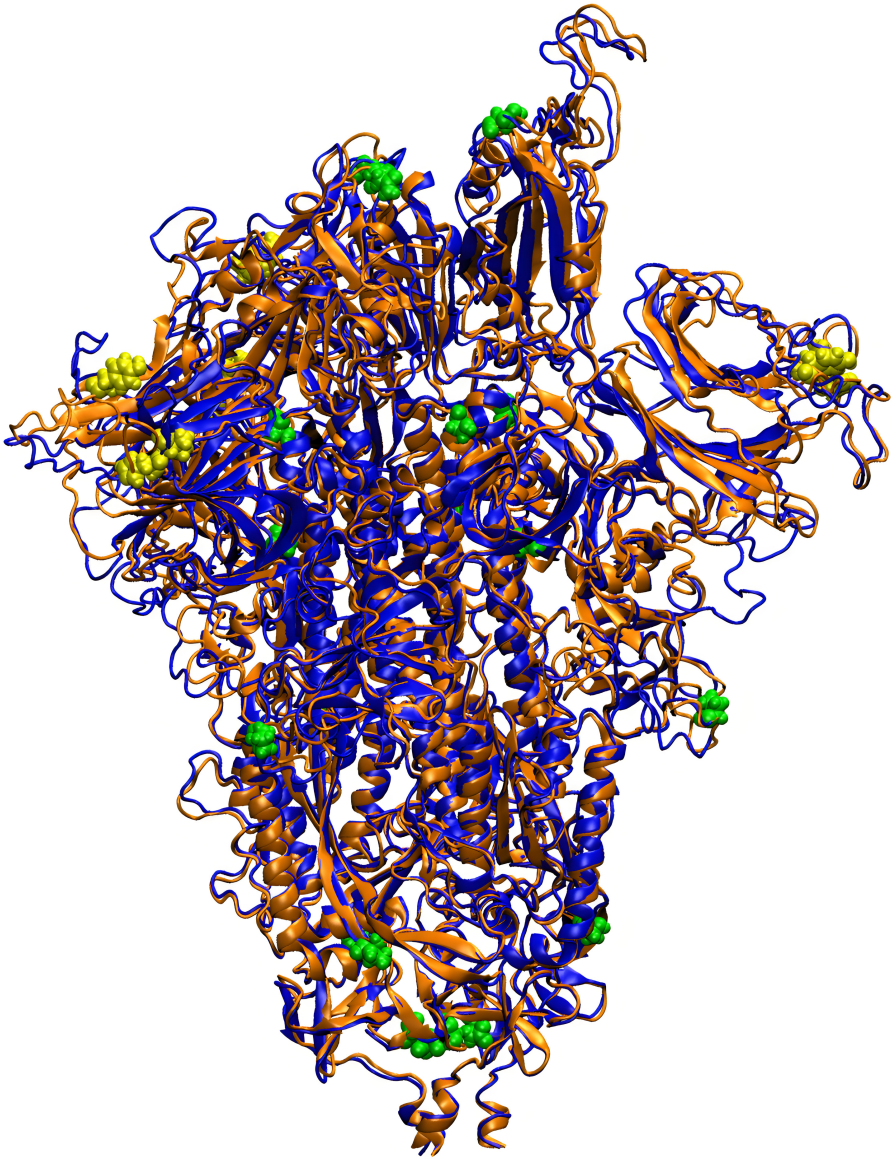

图1：野生型（蓝色）和 Alpha变体（橙色）刺突蛋白的二级结构，分别由实验PDB（6vsb.pdb）和同源建模获得。黄色和绿色分别标出了使野生型形成 Alpha变体的缺失突变和突变。

两个系统（野生型和突变型）首先进行了 10000 步共轭梯度最小化，随后在 NVT系综中进行平衡，并在 2.0 纳秒的分子动力学模拟过程中将温度升至 310 K。因此，两个系统均在 310 K 的 NVT条件下进行了 1 微秒的演化。采用直接求和法评估库仑相互作用。具体而言，通过对隐式溶剂相互作用施加库仑修正来模拟溶剂的额外屏蔽作用；该修正采用距离依赖介电常数，并在库仑公式中加入一个额外的 1/r 项进行缩放。分子间相互作用的截断距离设为 7 Å，对应于水中的 Bjerrum长度 [43]。采用 CHARMM36力场 [44] 描述分子间和分子内相互作用，包括聚糖及 N-连接聚糖键的描述。

水性溶剂通过一种特定的介观技术来描述，即格子玻尔兹曼（LB）方法，它是玻尔兹曼方程的最简晶格版本。格子玻尔兹曼方法无需直接求解连续介质流体力学中的纳维-斯托克斯方程，即可对流体流动的动态行为进行建模。在这一框架中，溶剂通过一个虚拟的粒子系综来处理，这些粒子的运动和相互作用被限制在规则的时空晶格上。与速度空间相关的自由度大幅减少，这是格子玻尔兹曼方法的主要优势。因此，溶剂被描述为：在位置 $\vec{r}$、时间 t，以速度 $\vec{c}_i$运动时，找到一定数量溶剂粒子的概率。

沿着一个可能的网格方向。在格子玻尔兹曼方法中，粒子碰撞通过向局部平衡弛豫来表示。这里，我们采用最简单的碰撞算子形式，即著名的 Bhatnagar-Gross-Krook 算子；该算子是一个简单的单时间弛豫项 [45]。

单弛豫时间（BGK）形式的标准格子玻尔兹曼方案如下：

$$f_i\left(\vec{r}+\vec{c}_i\Delta t,\, t+\Delta t\right)=f_i\left(\vec{r},\, t\right)+\omega\left[f_i\left(\vec{r},\, t\right)-f_i^{eq}\left(\vec{r},\, t\right)\right]+S_i\left(\vec{r},\, t\right), \tag{1}$$

其中， f_i 是与离散速度 $\vec{c}_i$对应的离散玻尔兹曼分布， $i=0,\ b$ 遍历离散格子；在本例中，采用的是通常记作 D3Q19 的 19速格子。在 1 中，弛豫频率ω 通过关系式 $\omega=2/(6\nu+1)$ 用于设定流体的运动黏度 ν，而 f_i^{eq} 是格子局部平衡，基本上是截断至马赫数二阶项的局部麦克斯韦-玻尔兹曼分布。质量密度和质量流量由分布函数的矩求得：

$$\rho\left(\vec{r},\, t\right)=\sum_i f_i\left(\vec{r},\, t\right) \tag{2}$$

$$\rho\left(\vec{r},\, t\right)\vec{u}\left(\vec{r},\, t\right)=\sum_i f_i\left(\vec{r},\, t\right)\vec{c}_i \tag{3}$$

其中， m 表示一个质量缩放因子，用于获得正确的水密度：在 1 atm 和 310 K 条件下为 993.4 kg/m^3 。

值得强调的是，尽管 T表示流体的有限温度，标准格子玻尔兹曼算法并未包含热涨落；根据涨落耗散关系，这些热涨落会在流体中产生自发的局部应力： $<s_{\alpha\beta}(\vec{r},\ t)s_{\gamma\nu}(\vec{r}',\ t')>=2k_BT\eta_{\alpha\beta\gamma\nu}\delta(\vec{r}-\vec{r}')\delta(t-t')$，其中 $s_{\alpha\beta}$是涨落应力张量，$\eta_{\alpha\beta\gamma\nu}$ 是黏度张量， k_B 表示玻尔兹曼常数。遵循 Adhikari 等人 [46], 的开创性工作，本文将方程 1 中的强迫项推广为 $S_i(\vec{r},\ t)=p_i+\xi_i$，其中 p_i 计入外力和耦合力（见下文），而 ξ_i 表示一组随机噪声项，用于模拟涨落应力张量，并由耗散补偿，同时保持系统中的质量守恒和动量守恒。关于实际涨落格子玻尔兹曼实现的更多细节，见参考文献 [47, 48]。因此，作为引入热噪声的结果，流体会作为分子动力学粒子的热浴，这一点已在文献中有所观察 [20, 26, 49]。

流体和粒子的耦合方式如下。粒子对周围流体的影响通过摩擦力项 $\vec{F}_{nj}=\gamma(\vec{v}_n-\vec{u}_j)$ 建模，其中 $\vec{v}_n$ 表示粒子速度， $\vec{u}_j$ 表示通过对最近的八个格点进行线性插值得到的粒子位置处的流体速度。随后，通过额外的力项 $S(\vec{r},\ t)$ 将耦合力加入式 1。因此，对粒子施加大小相等、方向相反的力，以模拟耦合项（从流体作用于粒子）的对应作用。根据文献[20], ，摩擦系数 γ 取为 0.1 fs^{-1}，而运动黏度设为 0.07 Å^2/fs，对应于 310 K 时水的运动黏度。格子玻尔兹曼方案与分子动力学积分方案同步逐步演化，时间步长为 2.0 飞秒。

2.2.1 全原子分子动力学（AA-MD）

AA-MD模拟使用GROMACS-2020.6 [50]在基于IBM Power9架构和Volta NVIDIA GPU的集群上进行。刺突蛋白和 α-刺突蛋白均先进行了10000步共轭梯度最小化。随后，按照LB/MD模拟中的做法，将系统在2.0纳秒内平衡至310 K。分子动力学模拟持续1微秒（NPT系综，310 K，1 atm）。模拟时间步长为1.6飞秒。模拟采用周期性边界条件，并使用粒子网格埃瓦尔德方法处理长程静电相互作用，网格间距为1.5 Å，结合四阶B样条电荷插值方案。库仑相互作用和伦纳德-琼斯相互作用的截断距离均为12 Å，并在10 Å处开始使用力切换函数 [51]。使用CHARMM36力场 [44]模拟分子间和分子内相互作用，包括聚糖和N-连接聚糖键的描述。水采用 CHARMM [52] 中实现的TIP3P水模型描述；该模型将水分子设定为由3个原子组成的刚性分子，并为每个原子指定伦纳德-琼斯参数和电荷。

3 结果

本节介绍一些通常用于评估蛋白质结构的量，并对所使用的两种方法以及所考虑的两种变体进行分析。我们还将 AA-MD 野生型的结果与文献中的数据进行比较。可以先说明的是，我们的模型能够恰当地再现蛋白质的主要特征，这些特征与采用显式溶剂进行类似的全原子或粗粒化计算所得结果一致。随后，我们着重介绍野生型与突变型刺突蛋白在结构和动力学方面的主要差异，主要涉及刚性、涨落和相关性，并在可能的情况下将这些差异与氨基酸突变联系起来。最后，本研究的主要目标，

我们分析了混合 LB-MD 描述能在多大程度上准确再现表征该蛋白质的各种特征，并探讨了可能存在差异的原因。

AA-MD 和 LB-MD 对刺突蛋白及 α-刺突蛋白的均方根偏差（RMSD）和均方根波动（RMSFs）见图 2)-3)。图 2 展示了完整轨迹的 RMSD，轨迹时长为 1 μs。

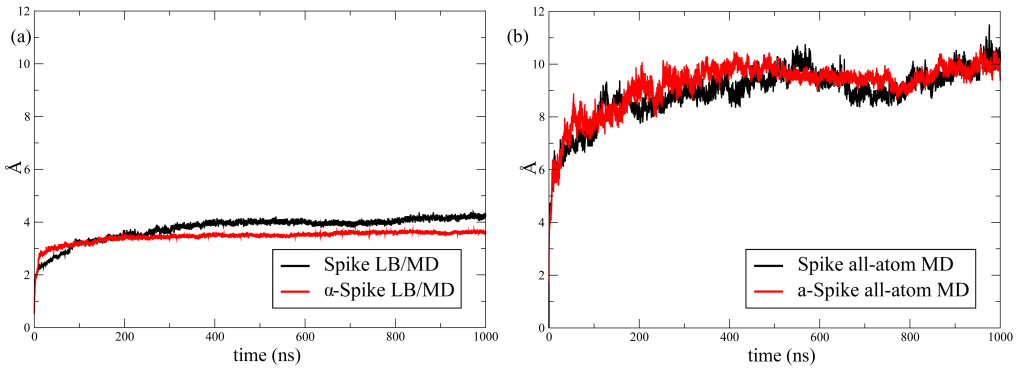

图 2：LB-MD（左图）和 AA-MD（右图）模拟中 Cα 原子的均方根偏差；轨迹时长为 1 μs。野生型与突变型之间存在轻微差异。LB-MD 情况下偏差幅度降低，与蛋白质在介观溶剂中的迁移率低于在显式溶剂中的迁移率有关。

在 AA-MD 情况下，两种结构的 RMSD 相近；相对于冷冻电镜结构，初始偏差较大，而在 200 ns 至 1000 ns 期间表现出较高的稳定性，波动范围为 8Å-10Å。按链分析的 RMSD（未展示）显示，比较天然型与突变型时，A 链和 B 链的行为相似。突变型的 C 链比天然型的 C 链更快达到松弛状态。再来看 LB-MD 结果，首先可以明确的是，与冷冻电镜参考结构相比，该结构具有稳定性。整体变化为 3.5Å-4Å，突变型比天然结构更快达到稳定平台。突变型的 A 链（未展示）RMSD 值高于天然型，差值接近 1Å；而 B 链则相反，天然型 RMSD 的平台为 5Å，远高于突变型的 2Å 平台。两种结构的 C 链表现相近。已有研究报告了不同链之间 RMSD 的差异[53]„ 与我们的发现一致。野生型的全原子描述 RMSD 值与采用类似方法得到的结果相近 [53], , 但高于其他参考数据[54, 55]。不过，这些差异可能取决于论文中所用的参考构象[54, 55, 53]。就我们的研究而言，考虑到蛋白质整体尺寸较大，包含 3 条链，每条链均含有 1000 多个残基，且我们的参考结构是由冷冻电镜和同源建模获得的模型，因此 RMSD 值是合理的。与参考文献[54], 中所考察的会导致相对于天然结构偏差更大的突变不同，本研究所考察的 α-刺突蛋白在整个结构上的 RMSD 与天然蛋白质相当或更低。

无论如何，与其关注 RMSD 的绝对值，不如考察其变化趋势，以确认结构没有发生漂移且已达到平衡。

总体而言，AA-MD 的构象平衡过程相当缓慢，与根据冷冻电镜数据和同源建模获得的初始模型相比，结构发生了较大变化。LB-MD 中的弛豫速度快得多，相对于起始结构的变化也较小。此外，AA-MD 中的 RMSD 涨落幅度大于 LB-MD。观察到的差异部分可归因于 LB-MD 描述本身具有的刚性。此外，我们无法排除起始冷冻电镜结构的影响；该结构可能代表蛋白质构象空间中一个远离平衡态的微观状态。刺突蛋白和 α-刺突蛋白的 AA-MD 模拟中，RMSD 曲线长期漂移，这表明由冷冻电镜获得的起始蛋白质结构与开放态的平衡结构并不接近。

为确保采样质量，我们评估了四条轨迹最后 500 ns 内的以下各项，忽略每条轨迹的前半部分。

我们使用 STRIDE 程序 [56],分析了刺突蛋白和 α-刺突蛋白的二级结构，分析时考虑了轨迹最后 500 ns 内每隔 1 纳秒选取的 500 个构象。在 AA-MD 和 LB-MD 系统中，我们都没有观察到刺突蛋白与 α-刺突蛋白之间存在显著差异。不过，正如文献 [57, 58] 中已有报道，我们在 LB-MD 分子动力学模拟中观察到 α-螺旋发生了轻微变形。具体而言，AA-MD 分子动力学模拟中的二级结构约有 20% 的残基处于 α-螺旋构象，而 STRIDE 程序将 LB-MD 轨迹中几乎所有的 α-螺旋结构都归类为转角单元；这些单元在主链的多个片段中连续重复，形成变形的螺旋。相比之下， β-折叠构象的分类在两种 AA-MD

和 LB-MD 系统中似乎都得以保留，平均值约为 20%。表示二级结构无序程度的卷曲比例在 LB-MD 系统中并未增加，这意味着即使存在介观溶剂，整体结构仍得以保留。

LB-MD 描述中的蛋白质刚性得到了均方根波动的证实（见图 3，结果显示的是 A 链，对应开放构象下的受体结合域）。天然态 AA-MD 中各条链的均方根波动（此处仅显示 A 链）与全原子或粗粒化模拟得出的类似数据相当[55, 59]。我们重点标出了 N 端结构域区域（至 300）内的三个峰值[55, 59], , 以及受体结合域 476-486 区域内的主要峰值[55, 60], ；该区域包含 S477，而 S477 在与人类 ACE2 的相互作用中起着关键作用。对于开放构象下受体结合域的 A 链，位于 682-689 的切割位点并不灵活；而在另外两条被研究的链中，该位点具有较强的移动性（数据未显示）。刺突蛋白与 α-刺突蛋白之间最显著的差异，出现在 A 链和 B 链的受体结合域 476-486 区域（后者未显示）。在 A 链中，突变区域的灵活性明显更高，而序列的其余部分移动性较低。在 A 链中，天然态的 476-486 区域灵活性更高；总体而言，突变结构的灵活性较低，但约在 850 处例外。

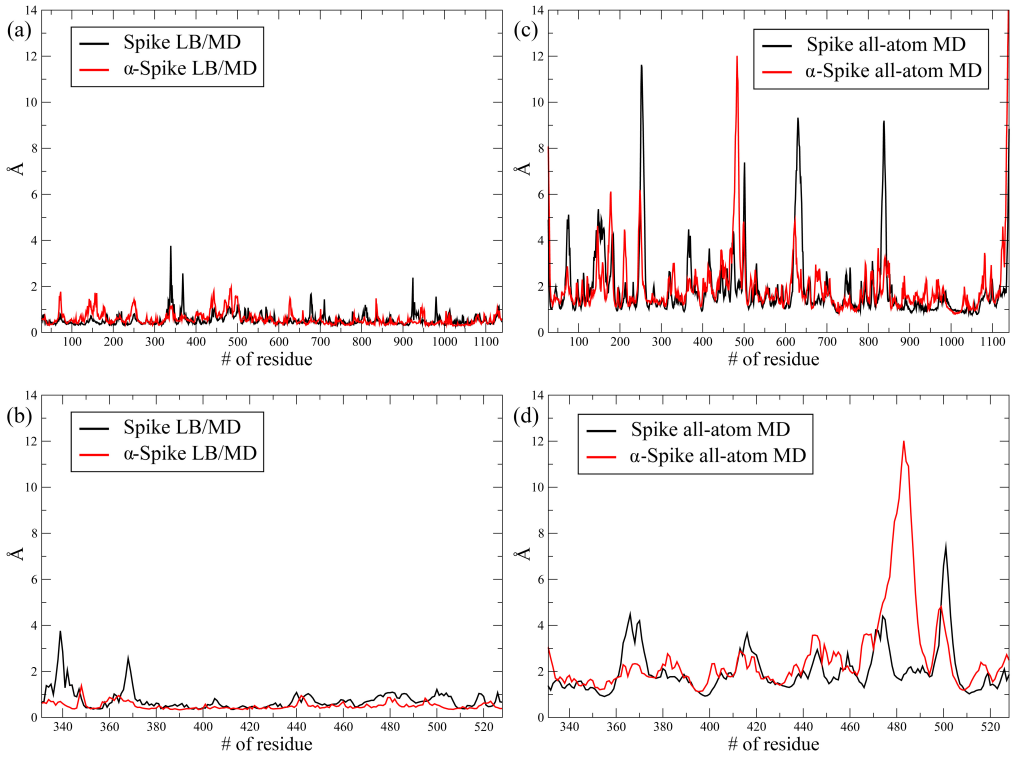

图 3：LB-MD（左图）和 AA-MD（右图）A 链模拟的均方根波动（Cα 原子，受体结合域）。上图为完整的 A 链，下图为 A 链的受体结合域。模拟开始的 500 纳秒数据被舍弃。野生型与突变型之间存在轻微差异。LB/MD 中强度的显著降低是由于蛋白质运动性较低所致。根据定性分析，这一结果是可以预期的；但据我们所知，此前尚未从定量角度对其进行考察。

LB-MD 中的涨落不如 AA-MD 中剧烈。值得注意的是，在受体结合域区域，LB-MD 重现了主要特征，尤其是识别出 L3 环（475 至 487 位残基）。这一点值得关注，因为该环与 ACE2 [32, 60, 61], 的稳定相互作用有关，而且该区域观察到的突变频率很高。总体而言，LB-MD 能够与 AA-MD 达到定性一致，两者在相同区域观察到涨落。不过，整个序列中偶尔会出现 LB-MD 涨落峰值（T470、Q920、S940、I980，来自不同链，且主要出现在天然结构中）；这些峰值与 AA-MD 描述的情况相当，值得进一步研究，尤其是在受体结合域区域。我们考察了对应于 G339 和 L368 残基的均方根波动峰值，这些峰值在野生型刺突蛋白中相当显著。图 4 展示了 LB-MD 和 AA-MD 模拟中 G339 和 L368 残基涨落的代表性快照。G339 和 L368 残基位于两条并排排列且可能相互作用的 α 螺旋中。在两种模拟中，涨落幅度相当，但 LB-MD 的表现并不合理。在整个 LB-MD 轨迹的最初 600 纳秒内，这两个残基彼此靠近（图 4(a)）；随后，在剩余模拟时间内，它们逐渐分开（图 4(b)），这可能是由于存在较高的能垒，阻碍了它们在两个位置间自由涨落。另一方面，在 AA-MD 模拟中，G339 和

L368 之间的距离在相同范围内波动得更加频繁（图 4 的 (c) 和 (d) 面板）。这一行为鲜明地表明，显式溶剂会影响局部自由能面，进而调节蛋白质结构的涨落。相比之下，介观溶剂可能会使自由能景观更加粗糙，从而限制结构的自由涨落，并促使残基位置出现受阻涨落（双峰分布）；在此处观察到的 LB-MD 模拟中 G339 和 L368 残基的情况即是如此。

关于隐式或显式不同溶剂模型对自由能采样的影响，已有大量研究，涉及蛋白质折叠过程的研究以及天然结构稳定性的分析 [62,63, 64, 57, 65]。此外，值得强调的是，已有多篇文献提出了混合溶剂方法 [66, 67, 68, 69]。具体而言，这些方法利用显式原子级表示来描述第一溶剂化层中的水分子，而将体相溶剂在连续介质层面进行建模。然而，本研究所使用的 LAMMPS 代码目前尚未包含混合溶剂表示方法。

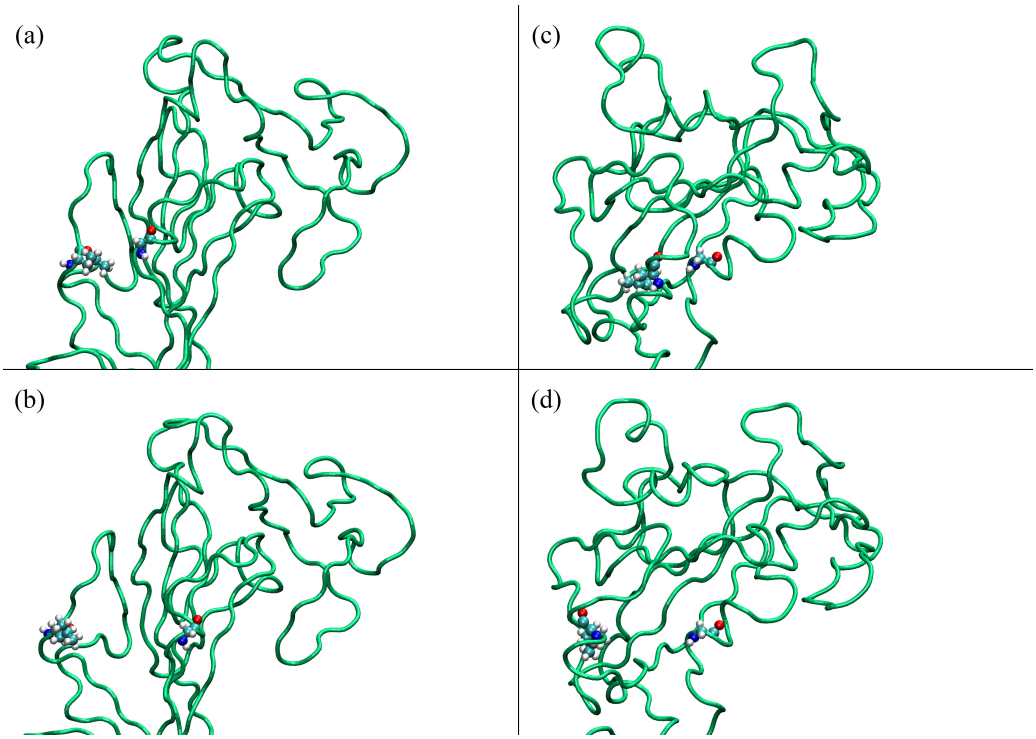

图 4：野生型刺突蛋白 A 链的受体结合域（RBD），展示了在 LB/MD 模拟（图 a、b）和 AA/MD 模拟（图 c、d）中以原子级表示突出显示的 G339 和 L368 残基的涨落。

为了深入了解三聚体的四级运动，我们对涉及相互连接的 S1 和 S2 片段的内部运动进行了主成分分析（PCA），并评估了 Cα 各对之间的交叉相关矩阵（CCM）。

主成分分析（见图5）对于比较各种方法很有参考价值。聚焦于刺突蛋白/α-刺突蛋白的比较可见，在相同方法的框架下，两者的描述相当相似。显然，由于LB-MD模型中的蛋白质柔性较低，AA-MD中的模态幅度几乎大3至4倍。此外，各模态的权重也有所不同，这意味着当溶剂采用全原子尺度或介观尺度描述时，蛋白质不同区域的柔性变化各不相同。例如，在AA-MD描述中，第一模态主要涉及NTD；而在LB-MD模型中则涉及受体结合域。

在全原子模拟中，主成分分析被用于证明，CoV-2 刺突蛋白的活性形式比 CoV-1 刺突蛋白的活性形式更稳定 [70], ，其中第一正则模与受体结合域的运动相关，正如这里由 LB 正确描述的那样。观察到 NTD 和受体结合域的协同运动构成第二正则模。在另一项模拟中 [53], ，也观察到 NTD 和受体结合域类似地参与第一正则模。

可以从相关性（CC）图（图6）中获得互补信息。乍看之下，两种方法和两种蛋白质变体的图呈现出相似的宏观行为，也与文献中的数据相当一致 [70]。与 AA-MD 相比，LB-MD 对残基间相关性的描述较为合理。这是因为在 LB-MD 方法中，未由溶剂介导的分子内局部相互作用得以保留，并得到了良好描述。值得注意的是，在两种方法中， α-刺突蛋白的相关性均高于

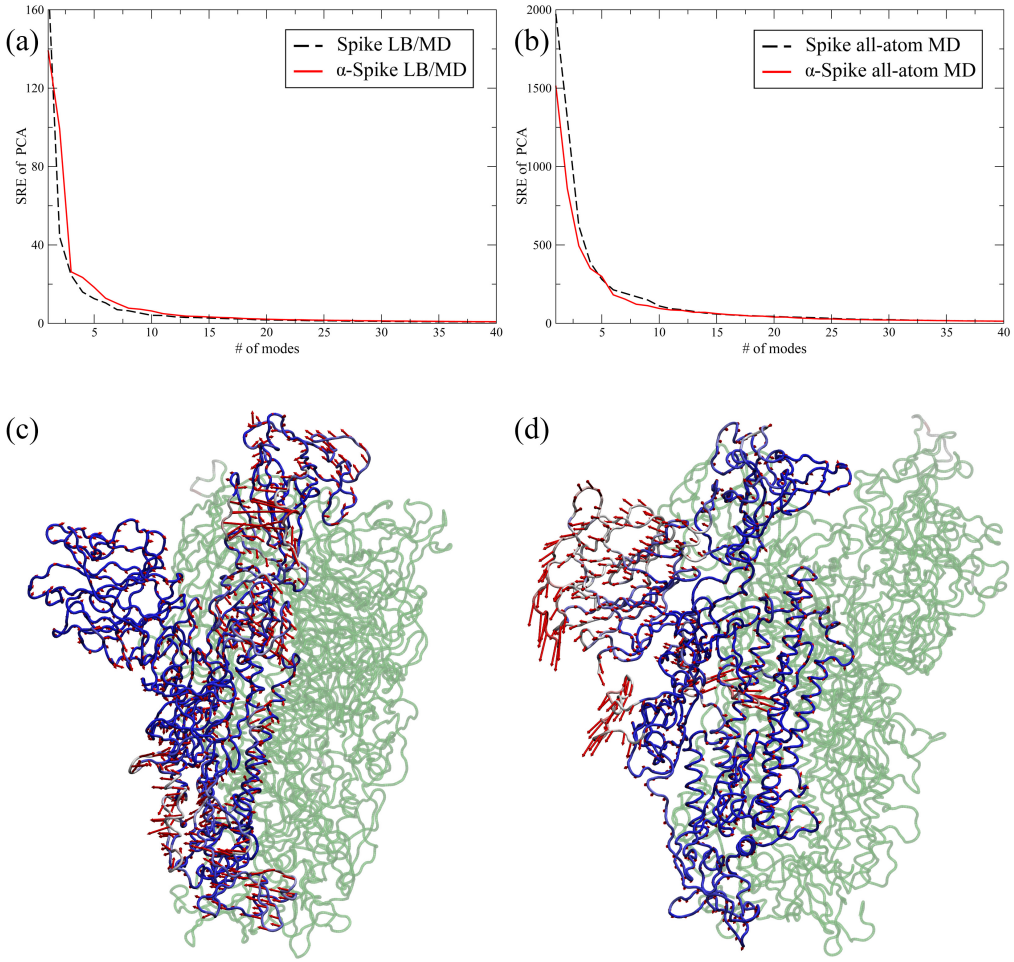

图5：上图：LB/MD（左图）和 AA/MD（右图）模拟中，PCA 得到的 A 链最大 40 个模式的特征值平方根（SRE）。野生型和突变型之间存在细微差异。LB-MD 情况下强度显著降低，这对应于蛋白质活动能力的减弱。下图：LB/MD 模拟（a）和 AA/MD 模拟（b）中，野生型刺突蛋白 A 链在主成分分析（PCA）中的主要变化模式。

野生型，这表明突变确实会影响蛋白质结构和相互作用倾向，而这些基本特征在介观描述中得以保留。高相关性区域与蛋白质结构域密切相关。

4 讨论与结论

在本研究中，我们展示了 LB-MD 方案如何准确描述蛋白质的一些基本结构和动力学特征。该方案将蛋白质的原子级描述与溶剂的介观模型耦合起来。不过，我们也明确指出并量化了若干关键问题。

首先，由于采用了隐式溶剂模型，LB-MD方案完全忽略了蛋白质内部以及蛋白质与溶剂分子之间的氢键。氢键以及与水分子的相互作用，对蛋白质稳定性、构象灵活性以及蛋白质-蛋白质相互作用都至关重要。例如，在这一具体案例中，氢键的形成是刺突蛋白-ACE2相互作用的基础[71]。此外，由于采用了介观尺度溶剂模型，蛋白质的灵活性和构象涨落受到强烈抑制。构象涨落至关重要，因为它关系到蛋白质改变结构构象并进而发挥功能的可能性。对于刺突蛋白这一具体案例，受体结合域中含有S477的L3环发生涨落，对于确保刺突蛋白锚定到ACE至关重要 [60]。因此，对融合前、未结合状态下的刺突蛋白受体结合域进行可靠的结构描述和分析，是非常关键的一步；例如，这有助于设计能够阻断人血管紧张素转换酶2与受体结合域相互作用的抑制剂或抗体 [60]。LB-MD仅能部分保证受体结合域的运动，其提供的描述本质上比AA-MD更为刚性。不过，LB-MD能够正确识别该区域具有较高的灵活性，并且

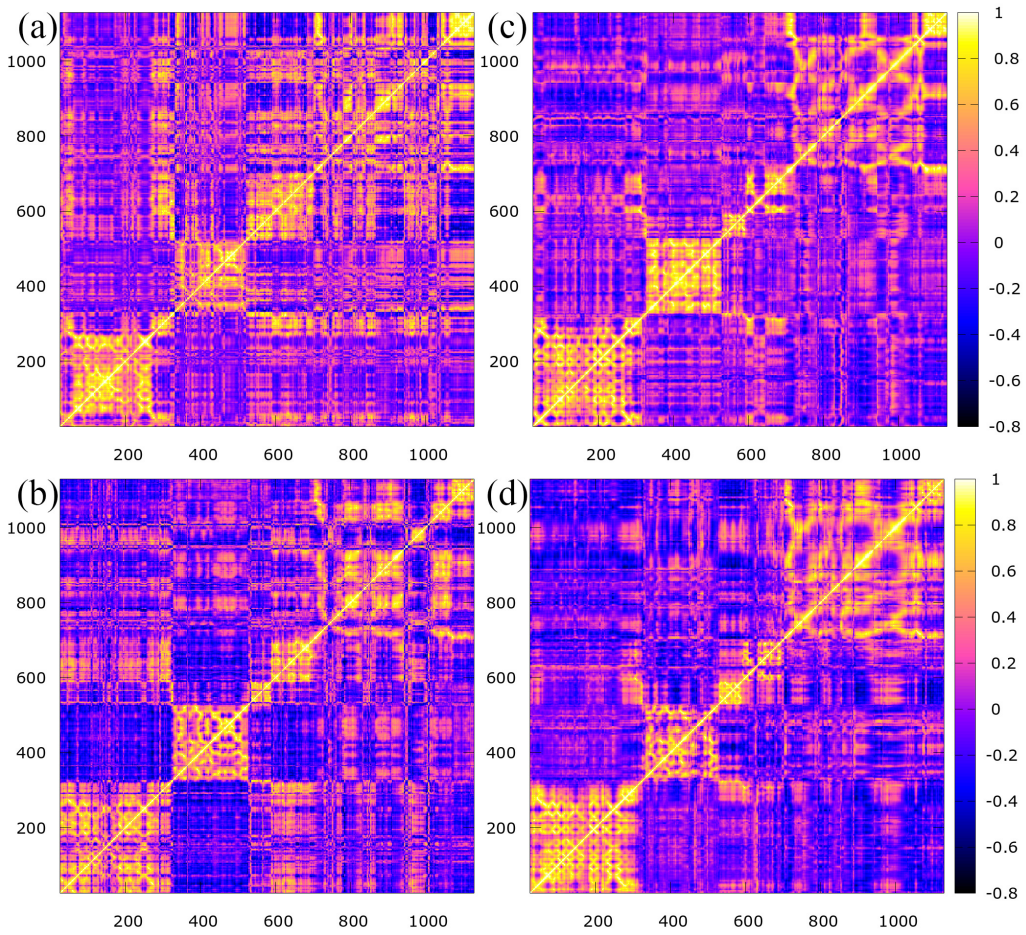

图6：LB/MD模拟（面板a、b）和AA/MD模拟（面板c、d）中，刺突蛋白及 α-刺突蛋白A 链的交叉相关矩阵交叉相关图。所有情况均显示一个显著的相关性区块，与受体结合域（残基331–528）的残基有关。全原子模拟中蛋白质较高的流动性，体现为交叉相关矩阵更丰富的结构（面板c、d）；不过，在LB/MD中，相关模式的定性结构仍得以保留。

与序列的其余部分相比，呈现出交叉相关性，例如CC图和均方根波动（RMSFs）所示。尽管两种方法得到的均方根波动在定量上存在差异，但在比较刺突蛋白和 α-刺突蛋白时，其定性变化趋势相似。特别是，与野生型刺突蛋白相比，我们观察到 α-刺突蛋白的受体结合域具有更高的刚性。亚结构域内部的主要相关性得以保留，而 α-刺突蛋白的CC图中受体结合域区块表现出的较强相关性，进一步证实了突变结构的稳定性信息。事实上，从统计学角度来看， α-刺突蛋白的受体结合域维持刚性结构的能力，可以与其显著的人际传播能力联系起来 [32], ，因为较小的涨落限制了受体结合域对可能构象盆地的采样范围，使其局限于能够更有效地与血管紧张素转化酶2结合的结构子区域。这种行为在某种程度上类似于SARS-CoV-2相较于SARS病毒所表现出的受体结合域刚性特征 [32]。此外，在LB/MD模拟中，对刺突蛋白和 α-刺突蛋白结构也观察到了相同的相关性趋势，这表明即使在对溶剂采用较低描述层次的情况下，多尺度方法仍能够保留关键信息。

此外，我们进行了初步模拟（未展示），结果表明，采用LB-MD方法也能成功实现ACE2与受体结合域的相互作用，克服了环区柔性不足和缺乏氢键这两个问题。

总体而言，与另一种还原论方法——粗粒化描述 [17] 的比较表明，LB-MD也能得到类似结果。我们在此仅研究了刺突蛋白的开放态。针对不同构象状态的最新研究 [59]表明，锁定态和闭合态比开放态更为刚性。因此，与柔性更强的开放态相比，LB-MD模型可能更适合描述这些功能状态。尽管存在局限性，LB-MD仍能追踪与突变相关的柔性结构变化，因此有望保留与不同谱系病毒行为相关的生物学信息。

总体而言，根据我们的研究，得出了两个主要结果：a）多尺度方法在萌芽阶段 展现出的潜力

可用于描述大型系统，并有可能将流动效应纳入蛋白质动力学；b）蛋白质/溶剂界面需要采用更精细的耦合方案，例如仅在靠近蛋白质界面的区域显式描述溶剂，以可靠地描述水分子与氨基酸之间的相互作用；这种相互作用对于决定生物分子的平衡性质和动力学性质都至关重要。这一需求仍然带来诸多问题，相关研究论文不断发表便证明了这一点 [62, 72, 73, 74, 63]。此外，LB 方法也可扩展为直接在晶格表示中纳入氢键相互作用。例如，过去几十年中，人们开发了多种类水势，以在晶格表示中描述水分子的极性。在这些势中，每个立方盒可容纳约一个水分子 [75, 76]，也可容纳由多个水分子组成的簇 [77, 78]。因此，通常会定义一个有效哈密顿量，用于模拟体相水及蛋白质水合层中的氢键相互作用 [79, 80]。一个实际例子是，将三维 Ben- Naim [81] 水势纳入 LB 方法，以模拟体相水中的氢键 [82]。不过，采用专为隐式处理溶剂而开发的粗粒化势后，LB 模型便无需再扩展以纳入氢键。隐式溶剂粗粒化模型的典型例子包括 OPEP [83] 和 dry-Martini [84] 力场。特别是，OPEP 力场已成功与 LB 方法结合，用于研究高剪切速率下的蛋白质解折叠 [85]。此外，粗粒化力场还可用于模拟大型系统，使层流得以充分发展，从而超越同时存在的平面膜、受体和病毒衣壳等大型结构所造成的局部热涨落效应。

总而言之，一方面，我们的模拟展示并定量估算了显式处理水分子对于获得生物分子统计可靠表征所起的关键作用。在这方面，尽管使用 LB 溶剂在计算上更具优势，但无法提供定量准确的信息。另一方面，我们的研究表明，在 LB-MD 介观溶剂描述中，许多对生物活性至关重要的结构特征仍得以保留。

未来可能的研究方向包括开发一类新型 LB 模型，使其能够支持比目前更大的涨落。沿这一方向的一个有前景的途径是采用更高阶的离散速度集合，并扩展模型以恰当描述氢键的形成。

致谢

M. L.、F. B.、M. D.、A. M.、A. T. 和 S. S. 感谢欧洲研究理事会通过欧盟“地平线 2020”框架计划（No. FP/2014-2020）提供资助，资助协议编号为 ERC Grant Agreement No.739964（COPMAT）。L. C.、A. L. 和 S. F. 感谢国际相对论天体物理中心网络（ICRANet）的支持。M. L.、L. C.、A. L. 和 S. F. 感谢意大利国家数学物理学组（GNFM-INdAM）的支持。所有作者衷心感谢 ENEA 提供高性能计算资源，并为 LB/MD模拟中使用的 HPC CRESCO 设施提供支持；该设施由 Associazione Big Data COVID-19 Fast Track 倡议提供。我们感谢 CINECA ISCRA 倡议下获批的 ABRISP 项目，该项目为全原子分子动力学模拟提供了高性能计算资源和支持。

参考文献

[1] J.-D. Rochaix, 《在生命起源处对一个含有一亿个原子的细胞器进行动态建模》, 《细胞》, 第 179 卷, 第 5 期, 第 1012–1014 页, 2019 年。[2] X. Song、M. Ø. Jensen、V. Jogini、R. A. Stein、C.-H. Lee、H. S. Mchaourab、D. E. Shaw 和 E. Gouaux, 《NMDA 受体通道被 MK-801 和美金刚阻断的机制》, 《自然》, 第 556 卷, 第 7702 期, 第 515–519 页, 2018 年。[3] D. Gadioli、E. Vitali、F. Ficarelli、C. Latini、C. Manelfi、C. Talarico、C. Silvano、C. Cavazzoni、G. Palermo 和 A. R. Beccari, 《Exscalate：一种超大规模的计算机虚拟筛选平台，可在 81 petaflops 超级计算机上于 60 小时内评估 1 万亿种化合物》, *arXiv* 预印本 *arXiv:2110.11644*, 2021 年。[4] E. Taka、S. Z. Yilmaz、M. Golcuk、C. Kilinc、U. Aktas、A. Yildiz 和 M. Gur, 《SARS-CoV-2 刺突糖蛋白与人类 ACE2受体之间的关键相互作用》, 《物理化学杂志 *B*》, 2021 年。[5] J. He、H. Tao、Y. Yan、S.-Y. Huang 和 Y. Xiao, 《SARS-CoV-2 的演化及其感染人类的分子机制》, 《病毒》, 第 12 卷, 第 4 期, 第 428 页, 2020 年。[6] V. Armijos-Jaramillo、J. Yeager、C. Muslin 和 Y. Perez-Castillo, 《SARS-CoV-2：从进化视角看其与人类 ACE2 的相互作用，揭示了维持复合物稳定性所必需的未知氨基酸》, 《进化应用》, 第 13 卷, 第 9 期, 第 2168–2178 页, 2020 年。[7] J. Zou、J. Yin、L. Fang、M. Yang、T. Wang、W. Wu、M. A. Bellucci 和 P. Zhang, 《通过相对自由能计算预测突变对 SARS-CoV-2 结合的影响》, 《化学信息与建模杂志》, 第 60 卷, 第 12 期, 第 5794–5802 页, 2020 年。[8] E. S. Brielle、D. Schneidman-Duhovny 和 M. Linial, 《SARS-CoV-2 与 ACE2 人类受体相互作用时采用了独特策略》, 《病毒》, 第 12 卷, 第 5 期, 第 497 页, 2020 年。[9] K. Gao、R. Wang、J. Chen、L. Cheng、J. Frishcosy、Y. Huzumi、Y. Qiu、T. Schluckbier 和 G.-W. Wei, 《以方法论为中心的 SARS-CoV-2 分子建模、模拟和预测综述》, *arXiv* 预印本 *arXiv:2102.00971*, 2021 年。[10] E. N. Muratov、R. Amaro、C. H. Andrade、N. Brown、S. Ekins、D. Fourches、O. Isayev、D. Kozakov、J. L. Medina-Franco、K. M. Merz、等, 《用于 COVID-19 药物发现的计算方法综述》, 《化学学会评论》, 2021 年。[11] A. Ali 和 R. Vijayan, 《ACE2–SARS-CoV-2/SARS-CoV 刺突蛋白界面的动力学揭示了独特机制》, 《科学报告》, 第 10 卷, 第 1 期, 第 1–12 页, 2020 年。[12] W. Zheng、H. Wen、G. J. Iacobucci 和 G. K. Popescu, 《通过粗粒化建模探究 NMDA 受体激活的结构动力学》, 《生物物理学杂志》, 第 112 卷, 第 12 期, 第 2589–2601 页, 2017 年。[13] L. Casalino、A. C. Dommer、Z. Gaieb、E. P. Barros、T. Sztain、S.-H. Ahn、A. Trifan、A. Brace、A. T. Bogetti、A. Clyde、等, 《人工智能驱动的多尺度模拟揭示 SARS-CoV-2 刺突蛋白动力学机制》, 《国际高性能计算应用杂志》, 第 10943420211006452 页, 2021 年。[14] B. Lev、S. Murail、F. Poitevin、B. A. Cromer、M. Baaden、M. Delarue 和 T. W. Allen, 《用弦方法求解五聚体配体门控离子通道的门控通路》, 《美国国家科学院院刊》, 第 114 卷, 第 21 期, 第 E4158–E4167 页, 2017 年。[15]C. Abrams 和 G. Bussi, 《在分子动力学中使用元动力学、副本交换和温度加速增强采样》, 《熵》, 第 16 卷, 第 1 期, 第 163–199 页, 2014 年。[16] A. Yu、A. J. Pak、P. He、V. Monje-Galvan、L. Casalino、Z. Gaieb、A. C. Dommer、R. E. Amaro 和 G. A. Voth, 《SARS-CoV-2 病毒粒子的多尺度粗粒化模型》, 《生物物理学杂志》, 第 120 卷, 第 6 期, 第 1097–1104 页, 2021 年。[17] T. Leong、C. Voleti 和 Z. Peng, 《冠状病毒刺突蛋白和 ACE2受体的粗粒化建模》, 《物理学前沿》, 第 9 卷, 2021 年。[18] L. Monticelli、S. K. Kandasamy、X. Periole、R. G. Larson、D. P. Tieleman 和 S.-J. Marrink, 《Martini 粗粒化力场：扩展至蛋白质》, 《化学理论与计算杂志》, 第 4 卷, 第 5 期, 第 819–834 页, 2008 年。

[19] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, 和 A. H. De Vries, “Martini力场：用于生物分子模拟的粗粒化模型”, 《物理化学杂志 *B*》, 第111卷, 第27期, 第7812–7824页, 2007年。[20] F. Sterpone, P. Derreumaux, 和 S. Melchionna, “流体中的蛋白质模拟：将OPEP粗粒化力场与流体动力学耦合”, 《化学理论与计算杂志》, 第11卷, 第4期, 第1843–1853页, 2015年。[21] F. Sterpone, S. Melchionna, P. Tuffery, S. Pasquali, N. Mousseau, T. Cragnolini, Y. Chebaro, J.-F. St-Pierre, M. Kalimeri, A. Barducci, 等, “OPEP蛋白质模型：从单分子、淀粉样蛋白形成、拥挤效应和流体动力学到DNA/RNA系统”, 《化学学会评论》, 第43卷, 第13期, 第4871–4893页, 2014年。[22] P. V. Coveney, A. Hoekstra, B. Rodriguez, 和 M. Viceconti, “计算生物医学。第二部分：器官与系统”, 《界面焦点》, 第11卷, 第1期, 文章编号20200082, 2021年。[23] W. Edeling, H. Arabnejad, R. Sinclair, D. Suleimenova, K. Gopalakrishnan, B. Bosak, D. Groen, I. Mah- mood, D. Crommelin, 和 P. V. Coveney, “不确定性对COVIDSIM流行病学代码预测的影响”, 《自然·计算科学》, 第1卷, 第2期, 第128–135页, 2021年。[24] F. Di Palma, S. Succi, F. Sterpone, M. Lauricella, F. Pérot, 和 S. Melchionna, “NMDA神经受体的三阶段多尺度建模”, 《分子物理学》, 第119卷, 第19–20期, 文章编号e1928312, 2021年。[25] M. Bernaschi, S. Melchionna, 和 S. Succi, “物理学-化学-生物学交叉领域的介观模拟”, 《现代物理评论》, 第91卷, 第2期, 文章编号025004, 2019年。[26] F. Mackay 和 C. Denniston, “通过使用保守力将分子动力学粒子与格子玻尔兹曼流体耦合”, 《计算物理学杂志》, 第237卷, 第289–298页, 2013年。[27] A. Dupuis, E. Kotsalis, 和 P. Koumoutsakos, “致密流体中格子玻尔兹曼模型与分子动力学模型的耦合”, 《物理评论 *E*》, 第75卷, 第4期, 文章编号046704, 2007年。[28] M. G. Fyta, S. Melchionna, E. Kaxiras, 和 S. Succi, “分子动力学与流体动力学的多尺度耦合：应用于DNA穿过纳米孔的易位”, 《多尺度建模与模拟》, 第5卷, 第4期, 第1156–1173页, 2006年。[29] A. Chatterji 和 J. Horbach, “分子动力学与格子玻尔兹曼的结合：用于模拟（带电）胶体系统的混合方法”, 《化学物理学杂志》, 第122卷, 第18期, 文章编号184903, 2005年。[30] Y. Cai, J. Zhang, T. Xiao, H. Peng, S. M. Sterling, R. M. Walsh, S. Rawson, S. Rits-Volloch, 和 B. Chen, “SARS-CoV-2刺突蛋白的不同构象状态”, 《科学》, 第369卷, 第6511期, 第1586–1592页, 2020年。[31] B. J. Bosch, R. Van der Zee, C. A. De Haan, 和 P. J. Rottier, “冠状病毒刺突蛋白是一种I类病毒融合蛋白：融合核心复合物的结构与功能表征”, 《病毒学杂志》, 第77卷, 第16期, 第8801–8811页, 2003年。[32] A. Spinello, A. Saltalamacchia, 和 A. Magistrato, “SARS-CoV-2刺突蛋白受体结合基序的刚性是否是其感染性增强的标志？来自全原子模拟的见解”, 《物理化学快报杂志》, 第11卷, 第12期, 第4785–4790页, 2020年。[33] M.-C. Bellissent-Funel, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. Van Der Spoel, Y. Xu, 和 A. E. Garcia, “水决定蛋白质的结构与动力学”, 《化学评论》, 第116卷, 第13期, 第7673–7697页, 2016年。[34] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, 和 J. S. McLellan, “融合前构象下2019-nCoV刺突蛋白的冷冻电镜结构”, 《科学》, 第367卷, 第6483期, 第1260–1263页, 2020年。[35] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, 和 Q. Zhou, “全长人血管紧张素转化酶2识别SARS-CoV-2的结构基础”, 《科学》, 第367卷, 第6485期, 第1444–1448页, 2020年。[36] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, 和 Y. Zhang, “I-TASSER套件：蛋白质结构与功能预测”, 《自然·方法》, 第12卷, 第1期, 第7–8页, 2015年。

[37] B. Meng, S. A. Kemp, G. Papa, R. Datir, I. A. Ferreira, S. Marelli, W. T. Harvey, S. Lytras, A. Mohamed, G. Gallo, 等, “SARS-CoV-2刺突蛋白H69/V70缺失的反复出现及其在Alpha变体B.1.1.7中的作用”, 《细胞报告》, 第35卷, 第13期, 文章编号109292, 2021年。[38] G. Martínez-Rosell, T. Giorgino, 和 G. De Fabritiis, “PlayMolecule ProteinPrepare：用于分子动力学模拟蛋白质制备的网络应用程序”, 《化学信息与建模杂志》, 第57卷, 第7期, 第1511–1516页, 2017年。[39] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, 和 J. H. Jensen, “PROPKA3：对经验性pKa预测中内部和表面残基的一致处理”, 《化学信息与建模杂志》, 第7卷, 第2期, 第525–537页, 2011年。[40] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, 和 N. A. Baker, “PDB2PQR：用于设置泊松–玻尔兹曼静电计算的自动化流程”, 《核酸研究》, 第32卷, 增刊_2, 第W665–W667页, 2004年。[41] W. Humphrey, A. Dalke, 和 K. Schulten, “VMD：可视化分子动力学”, 《分子图形学杂志》, 第14卷, 第1期, 第33–38页, 1996年。[42] S. Plimpton, “短程分子动力学的快速并行算法”, 《计算物理学杂志》, 第117卷, 第1期, 第1–19页, 1995年。[43] U. Micka, C. Holm, 和 K. Kremer, “不良溶剂中的强带电柔性聚电解质：分子动力学模拟”, 《朗缪尔》, 第15卷, 第12期, 第4033–4044页, 1999年。[44] J. Huang 和 A. D. MacKerell Jr, “CHARMM36全原子加和型蛋白质力场：基于与核磁共振数据比较的验证”, 《计算化学杂志》, 第34卷, 第25期, 第2135–2145页, 2013年。[45] S. Succi 和 S. Succi, 《格子玻尔兹曼方程：用于流动物质的复杂状态》。牛津大学出版社, 2018年。[46] R. Adhikari, K. Stratford, M. Cates, 和 A. Wagner, “涨落格子玻尔兹曼”, 《EPL（欧洲物理快报）》, 第71卷, 第3期, 文章编号473, 2005年。[47] F. Mackay, S. T. Ollila, 和 C. Denniston, “通过格子玻尔兹曼流体在LAMMPS中实现流体动力学力”, 《计算物理学通讯》, 第184卷, 第8期, 第2021–2031页, 2013年。[48] S. T. Ollila, C. Denniston, M. Karttunen, 和 T. Ala-Nissila, “用于复杂流体的涨落格子玻尔兹曼模型”, 《化学物理学杂志》, 第134卷, 第6期, 文章编号064902, 2011年。[49] B. Dünweg 和 A. J. Ladd, “软物质系统的格子玻尔兹曼模拟”, 《软物质科学的高级计算机模拟方法 III》, 第89–166页, 2009年。[50] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, 和 E. Lindahl, “GROMACS：通过从笔记本电脑到超级计算机的多级并行实现高性能分子模拟”, 《SoftwareX》, 第1卷, 第19–25页, 2015年。[51] P. J. Steinbach 和 B. R. Brooks, “大分子模拟中长程力的新型球形截断方法”, 《计算化学杂志》, 第15卷, 第7期, 第667–683页, 1994年。[52] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, 等, “用于蛋白质分子建模和动力学研究的全原子经验势”, 《物理化学杂志 B》, 第102卷, 第18期, 第3586–3616页, 1998年。[53] M. Ghorbani, B. R. Brooks, 和 J. B. Klauda, “探索SARS-CoV-2刺突糖蛋白的动力学与网络分析”, 《生物物理学杂志》, 第120卷, 第14期, 第2902–2913页, 2021年。[54] A. Khan, D.-Q. Wei, K. Kousar, J. Abubaker, S. Ahmad, J. Ali, F. Al-Mulla, S. S. Ali, N. Nizam-Uddin, A. Mohammad Sayaf, 等, “初步结构数据表明, SARS-CoV-2 B.1.617变体的受体结合域与ACE2受体的结合比野生型更强, 从而增强感染性”, 《ChemBioChem》, 第22卷, 第16期, 第2641–2649页, 2021年。[55] P. V. Raghuvamsi, N. K. Tulsian, F. Samsudin, X. Qian, K. Purushotorman, G. Yue, M. M. Kozma, W. Y. Hwa, J. Lescar, P. J. Bond, 等, “SARS-CoV-2刺突蛋白与ACE2的相互作用揭示了新的变构靶点”, 《eLife》, 第10卷, 文章编号e63646, 2021年。[56] D. Frishman 和 P. Argos, “基于知识的蛋白质二级结构指定”, 《蛋白质：结构、功能与生物信息学》, 第23卷, 第4期, 第566–579页, 1995年。

[57] R. Zhou，《水中蛋白质折叠的自由能景观：显式溶剂与隐式溶剂》，《蛋白质：结构、功能与生物信息学》，第 53 卷，第 2 期，第 148–161 页，2003 年。[58] N. Calimet、M. Schaefer 和 T. Simonson，《采用广义 Born/ACE 溶剂模型的蛋白质分子动力学》，《蛋白质：结构、功能与生物信息学》，第 45 卷，第 2 期，第 144–158 页，2001 年。[59] G. M. Verkhivker，《分子模拟与网络建模揭示 SARS-CoV-2 刺突蛋白中的变构信号传导》，《蛋白质组研究杂志》，第 19 卷，第 11 期，第 4587–4608 页，2020 年。[60] A. Singh、G. Steinkellner、K. Köchl、K. Gruber 和 C. C. Gruber，《丝氨酸 477 在 SARS-CoV-2 刺突蛋白与人类受体血管紧张素转化酶2相互作用中发挥关键作用》，《科学报告》，第 11 卷，第 1 期，第 1–11 页，2021 年。[61] V. A. Ngo 和 R. K. Jha，《确定 SARS-CoV-2/ACE2 紧密相互作用的关键决定因素及其动力学》，《PLOS ONE》，第 16 卷，第 9 期，文章编号 e0257905，2021 年。[62] Y. Chen、A. Krämer、N. E. Charron、B. E. Husic、C. Clementi 和 F. Noé，《用于分子动力学的机器学习隐式溶剂化》，《化学物理学杂志》，第 155 卷，第 8 期，文章编号 084101，2021 年。.[63] M. Kulke、N. Geist、D. Moller 和 W. Langel，《基于副本的蛋白质结构采样方法：在显式溶剂与隐式溶剂之间折中》，《物理化学杂志 B》，第 122 卷，第 29 期，第 7295–7307 页，2018 年。[64] Q. Shao 和 W. Zhu，《评估隐式溶剂中蛋白质折叠所用的 AMBER 力场》，《物理化学化学物理》，第 20 卷，第 10 期，第 7206–7216 页，2018 年。[65] A. Suenaga，《采用隐式溶剂对小型蛋白质折叠进行副本交换分子动力学模拟》，《分子结构杂志：THEOCHEM》，第 634 卷，第 1–3 期，第 235–241 页，2003 年。[66] A. V. Onufriev 和 S. Izadi，《生物分子模拟中的水模型》，《Wiley 跨学科评论：计算分子科学》，第 8 卷，第 2 期，文章编号 e1347，2018 年。[67] D. R. Roe、A. Okur、L. Wickstrom、V. Hornak 和 C. Simmerling，《广义 Born 溶剂模型中的二级结构偏差：显式与隐式溶剂化下构象系综及溶剂极化自由能的比较》，《物理化学杂志 B》，第 111 卷，第 7 期，第 1846–1857 页，2007 年。[68] A. Okur、L. Wickstrom、M. Layten、R. Geney、K. Song、V. Hornak 和 C. Simmerling，《通过采用显式/隐式混合溶剂化模型提高副本交换模拟的效率》，《化学理论与计算杂志》，第 2 卷，第 2 期，第 420–433 页，2006 年。[69] Y. Deng 和 B. Roux，《氨基酸侧链的水合：通过采用显式水分子的分阶段分子动力学自由能模拟计算非极性和静电贡献》，《物理化学杂志 B》，第 108 卷，第 42 期，第 16567–16576 页，2004 年。[70] V. G. Kumar、D. S. Ogden、U. H. Isu、A. Polasa、J. Losey 和 M. Moradi，《SARS 冠状病毒 1 型和 2 型融合前刺突蛋白的差异动态行为》，BioRxiv，第 2020–12 页，2021 年。[71] S. Piplani 、 P. K. Singh、D. A. Winkler 和 N. Petrovsky，《不同物种 SARS-CoV-2 刺突蛋白与 ACE2 结合亲和力的计算机模拟比较及其对病毒起源的启示》，《科学报告》，第 11 卷，第 1 期，第 1–13 页，2021 年。[72] P. Tao 和 Y. Xiao，《采用广义 Born 表面积模型折叠蛋白质可实现更有效的采样，同时在定性上保留折叠景观》，《物理评论 E》，第 101 卷，第 6 期，文章编号 062417，2020 年。[73] A. Gao、R. C. Remsing 和 J. D. Weeks，《用于离子相关性和疏水缔合的短程溶剂模型》，《美国国家科学院院刊》，第 117 卷，第 3 期，第 1293–1302 页，2020 年。[74] A. Chakravorty、Z. Jia、L. Li、S. Zhao 和 E. Alexov，《从单一结构再现蛋白质系综平均极性溶剂化能：用于大分子建模的高斯平滑介电函数》，《化学理论与计算杂志》，第 14 卷，第 2 期，第 1020–1032 页，2018 年。[75] K. Stokely、M. G. Mazza、H. E. Stanley 和 G. Franzese，《氢键协同性对水行为的影响》，《美国国家科学院院刊》，第 107 卷，第 4 期，第 1301–1306 页，2010 年。[76] G. Franzese 和 H. E. Stanley，《水的哈密顿模型中的液-液临界点：解析解》，《凝聚态物理杂志》，第 14 卷，第 9 期，文章编号 2201，2002 年。

[77] A. Ben-Naim, 《水和水溶液的一维模型。I. 纯液态水》, 《化学物理学杂志》, 第 128 卷, 第 2 期, 文章编号 024505, 2008 年。[78] R. A. Lovett 和 A. Ben-Naim, 《惰性气体水溶液的一维模型》, 《化学物理学杂志》, 第 51 卷, 第 7 期, 第 3108–3119 页, 1969 年。[79] V. Bianco、N. Pagès-Gelabert、I. Coluzza 和 G. Franzese, 《折叠蛋白质的稳定性如何取决于界面水性质和残基间相互作用》, 《分子液体杂志》, 第 245 卷, 第 129–139 页, 2017 年。[80] V. Bianco、G. Franzese、C. Dellago 和 I. Coluzza, 《水在常温及极端热力学条件下稳定蛋白质筛选中的作用》, 《物理评论 *X*》, 第 7 卷, 第 2 期, 文章编号 021047, 2017 年。[81] A. Ben-Naim, 《二维中“类水”粒子的统计力学。I. 物理模型及 Percus–Yevick 方程的应用》, 《化学物理学杂志》, 第 54 卷, 第 9 期, 第 3682–3695 页, 1971 年。[82] S. Succi、N. Moradi、A. Greiner 和 S. Melchionna, 《类水流体的格子玻尔兹曼建模》, 《物理学前沿》, 第 2 卷, 文章编号 22, 2014 年。[83] F. L. Barroso da Silva、F. Sterpone 和 P. Derreumaux, 《OPEP6：采用 OPEP 粗粒化力场的全新恒定 pH 分子动力学模拟方案》, 《化学理论与计算杂志》, 第 15 卷, 第 6 期, 第 3875–3888 页, 2019 年。[84] C. Arnarez、J. J. Uusitalo、M. F. Masman、H. I. Ingólfsson、D. H. De Jong、M. N. Melo、X. Periole、A. H. De Vries 和 S. J. Marrink, 《Dry Martini：用于脂质膜模拟的隐式溶剂粗粒化力场》, 《化学理论与计算杂志》, 第 11 卷, 第 1 期, 第 260–275 页, 2015 年。[85] F. Sterpone、P. Derreumaux 和 S. Melchionna, 《剪切作用下蛋白质解折叠的分子机制：一项格子玻尔兹曼分子动力学研究》, 《物理化学杂志 *B*》, 第 122 卷, 第 5 期, 第 1573–1579 页, 2018 年。