arXiv:2611.27924v1 [cs.AI] 20 Nov 2026

BioEmu2：可扩展的生物分子动力学模拟

Michael Gastegger†, Adam E. Foster†, José Jiménez-Luna†, Tim Hempel†, Sarah Lewis†, Yu Xie†, Lixin Sun†, Iryna Zaporozhets†, J. Thorben Frank†, Yaoyi Chen†, Michael Plainer†, Ludwig Winkler†, Pedro B. P. S. Reis†, Sven H. Giese†, Roberto Sordillo, Victor García Satorras, Eray Inanc, Kacper Kapuśniak, Soojung Yang, Rishwanth Raghu, Rachel Kim, Gabriel Dernbach, Deniz Gunceler, HannesSchulz, Bonnie Kruft, Marwin Segler, Christopher M. Bishop, Cecilia Clementi, Ryota Tomioka, Frank Noé*

† 贡献相同

*联系邮箱： franknoe@microsoft.com

摘要

尽管结构预测取得了重大进展，但要大规模模拟驱动生物分子功能的运动和相互作用，仍是一项核心挑战。我们推出 BioEmu2，这是一种生成模型，可生成原子级分辨率的生物分子结构，并结合静态结构预测、平衡采样和时间演化，其采样速度比分子动力学（MD）快多个数量级。该模型基于实验结构、合成结构和大量分子动力学数据进行训练，支持包含蛋白质、核酸、修饰残基、配体和离子的多组分系统。其结构预测准确度可媲美顶尖的开放权重生物分子折叠模型，上下文并行模式还可实现兆道尔顿复合物的预测。在经过测试的蛋白质和复合物中，其平衡分布与分子动力学参考相近，自由能误差约为 1 千卡/摩尔。时间传播可预测长时间尺度的动力学轨迹，以及结合和构象变化的动力学过程。通过以残基接触为条件信息，或使用用户定义的偏置势进行引导，可以控制生成的结构。增强采样可实现涉及稀有态的自由能计算。BioEmu2 使生物分子系综及其动力学研究能够实现规模化。

1 引言

蛋白质、DNA 和 RNA 是生命的基本组成部分，承担着从氧气运输、肌肉收缩到遗传信息存储与处理等多种功能。它们的活动涉及三个反复出现的过程：(i) 构象变化，即改变分子的三维排列；(ii) 结合与解离，即组装复合物并捕获或释放配体；以及 (iii) 化学反应，即转化底物或充当分子开关。这些过程往往相互耦联。例如，在血红蛋白中，氧结合会使
构象平衡从低亲和力的紧张态 (T) 转向高亲和力的松弛态 (R)。这种耦联支持血红蛋白在肺部协同装载氧气，并在组织中氧分压较低时释放氧气 [1]。在腺苷酸激酶中，催化循环将核苷酸结合、结构域闭合、磷酰基转移以及为产物释放而重新开放联系起来 [2]。这些分子活动构成相互关联的通路，支撑着细胞功能。若能以可扩展的方式理解分子功能及其在疾病中的紊乱，将有望改变药物靶点选择，阐明作用机制，并提升对药物疗效和安全性的预测能力。

新一代测序使分子序列得以广泛获取 [3], ，而 AlphaFold、RoseTTAFold 及其全原子扩展则改变了分子结构和复合物的预测方式 [4–7]。然而，单凭预测结构无法说明哪些构象会出现、它们相互转变的速度有多快，或配体结合如何改变这些性质。因此，要理解功能，就需要了解结构集合及其动力学。

生物物理实验能够揭示生物分子功能的许多方面，但通常存在通量低或分辨率低的问题。例如，冷冻电子显微镜可以分辨不同的构象状态，核磁共振波谱则可测量结构涨落和交换过程 [8, 9], ，但这两种技术都成本高昂，且需要耗时的样品制备、数据采集和数据分析。多种生物物理

技术可用于探测生物分子的形状、动力学和相互作用，包括荧光和红外光谱学、小角X射线散射以及氢氘交换质谱 [10–13]。无细胞表达、微型化和cDNA展示扩大了功能表征的规模，使生化活性和蛋白质稳定性得以进行高通量测量 [14, 15]。即便扩大规模，这些方法测量的仍是低维可观测量，而非直接解析完整的原子级构象系综。

原子级分子动力学（MD）描述在选定力场作用下，生物分子结构随时间发生的细致演变。其预测能力受相互作用势函数准确性的限制，但更主要受充分采样所需计算成本的制约。折叠、构象重排和缔合可能涉及罕见的势垒跨越，因此需要极长的模拟时间尺度，蛋白质折叠和蛋白质–蛋白质缔合模拟便是例证 [16, 17]。专用硬件和增强采样算法拓展了可触及的时间尺度，但探索复杂构象景观的计算负担仍然很重 [16–18]。机器学习粗粒化模型通过为较少的自由度学习有效相互作用，降低了这一负担 [19, 20]。近期的可迁移模型能够复现原子级模拟的结构和热力学性质，同时在基准蛋白质上将计算成本降低数个数量级 [20]。然而，要将这一准确性推广到小分子、恢复原子级细节并校准动力学，仍然面临挑战。

生成模型直接学习对目标分布进行采样，而不是求解其间的每一个模拟步骤。玻尔兹曼生成器对由已知能量函数定义的分布进行采样 [21] ，但扩展到大型系统的能力有限。DiG、AlphaFlow 和 BioEmu 从数据中学习分子系综 [22–24], ，但它们都是粗粒化的单蛋白质模型。Timewarp、MDGen 和隐式转移算子将这一方法拓展到时间上分隔的构型或轨迹片段 [25–27], ，但它们都不是能够在多种复杂结构范围内泛化的生物分子基础模型。

在此，我们介绍 BioEmu2，它在一个原子级框架中统一了三种模式：(1) 结构预测，在 FoldBench 和 Runs N' Poses [32, 33] 上，其准确度可与当前开放权重的生物分子折叠模型 [28–31] 相媲美；(2) 近似室温平衡采样，可再现大规模分子动力学模拟的自由能面，自由能误差在 1 千卡/摩尔以内，并覆盖主要极小值；(3) 时间传播，可将结构传播至未来任意时间滞后 τ，从而模拟长时间尺度的分子动力学轨迹，并预测构象变化与结合的动力学。BioEmu2 支持包含蛋白质、DNA、RNA、修饰残基、配体和离子的多组分系统，并达到重原子分辨率。上下文并行可支持对含有超过 5000 个残基的系统进行推理。接触条件化和推理时引导可控制结构生成。增强采样可提高稀有态的状态覆盖率，而有效的重加权则可恢复所学习基础系综的统计特性。若干能力仍处于实验阶段，尤其是慢动力学方面，仍需更大规模、高质量的实验数据集和进一步验证。BioEmu2 将许多分子动力学式采样任务的速度提升了数个数量级，从而推动人们快速迭代，逐步实现对生物分子动力学与功能的可扩展模拟。

2 BioEmu2 模型

BioEmu2 是一种生成模型，能够以原子分辨率预测生物分子结构、系综和动力学。与 BioEmu 的第一个版本相比， [24], BioEmu2 可以对由任意数量的蛋白质、RNA/DNA、翻译后修饰、脂质、小分子和离子组成的分子系统进行建模（图 1）。

从整体上看，BioEmu2 模型将学习问题分为两个互补部分：序列主干中的生物学推理，以及结构模块中的坐标生成 [4, 5]。在主干中，通过将蛋白质链表示为氨基酸组成、将核酸表示为核苷酸，并将其他物种直接表示为各自的组成原子，获得生物分子系统的标记化描述 $\mathcal{S}$。此外，还会通过多序列比对（MSA） [34, 35], 为蛋白质链补充进化信息。可选的主干条件输入 $\mathcal{C}_{\text{trunk}}$ 可用于实现对结构生成的推理时控制。随后，序列主干会处理这三种输入

$$(\mathcal{S}, \text{MSA}, \mathcal{C}_{\text{trunk}}) \xrightarrow{\text{sequence trunk}} \mathcal{E}_{\text{trunk}}. \tag{2.1}$$

BioEmu2 支持两种主干条件输入：一种是方法标志，用于定义采样分布，以实现精确的结构预测或近似的平衡态采样；另一种是一组（可以为空）标记对之间的物理接触，在生成结构时作为模型的软约束。作为一种扩散

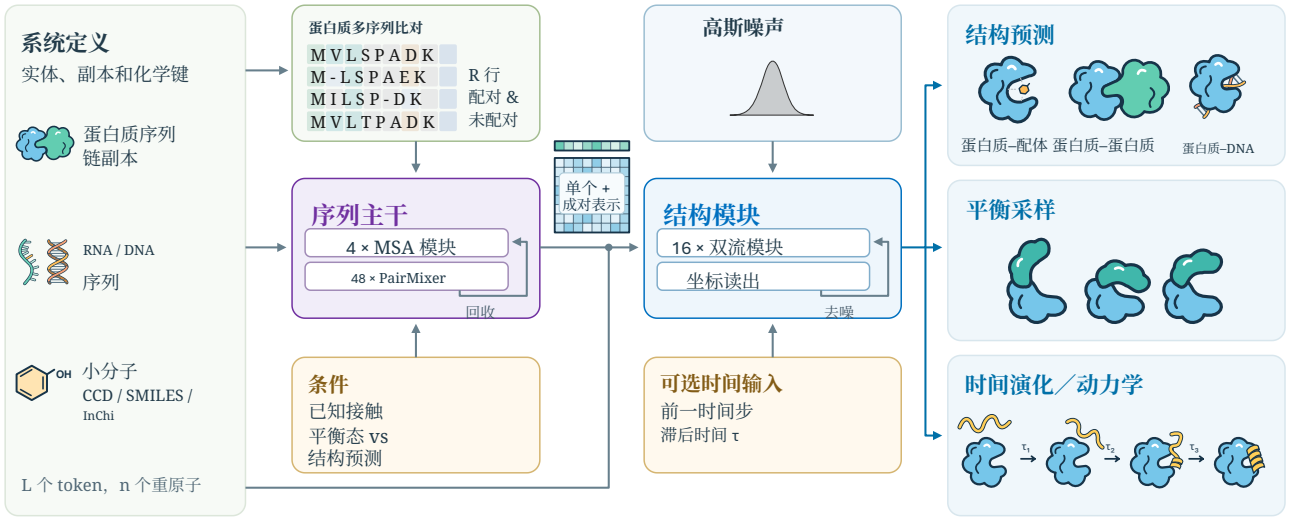


图 1：**BioEmu2 概览**：BioEmu2 的高层架构——序列模块和结构模块的详细架构请参见第 4 节。BioEmu2 主要有三种使用模式：静态结构预测、平衡结构采样，以及通过结构的时间传播来模拟分子动力学。

生成模型 [36–38], BioEmu2 的结构模块不会在单次前向传播中生成结构，而是会在不同噪声水平下逐步对结构进行去噪 σ。因此，结构模块的一次运行会将缓存的主干嵌入，以及描述部分加噪结构的坐标 x_{noisy} 和噪声水平 σ 作为输入

$$(\mathcal{S}, \mathcal{E}_{\text{trunk}}, \mathcal{C}_{\text{dyn}}; x_{\text{noisy}}, \sigma) \xrightarrow{\text{structure module}} \hat{x}, \tag{2.2}$$

从而得到去噪预测 $\hat{x}$。

与该领域早期模型不同， [5, 39], BioEmu2 还能够模拟动力学。实际上，这是通过向结构模块提供可选的条件输入来实现的 $\mathcal{C}_{\text{dyn}} = (x_t, \tau)$。该输入包含起始结构 x_t 和时间滞后 $\tau > 0$。随后，模型经过训练，能够从分布 $p_\tau(x_{t+\tau}|x_t)$ 中采样，即从 x_t 开始并在时间间隔 τ 内向前进行分子动力学模拟所得到的结构分布。

推理模式。 序列主干和结构模块可以使用不同的条件输入，因此产生了三种主要的推理模式组合，概述如下：

- **结构预测** （方法 = 结构预测，$\mathcal{C}_{\text{动力学}} = \varnothing$）：结构模块生成预测结构 $\hat{x}$，模拟通过X射线晶体学或冷冻电镜实验获得的静态结构。
- **平衡采样** （方法 = 平衡态，$\mathcal{C}_{\text{动力学}} = \varnothing$）：从对平衡分布的学习近似中生成构象样本。

- **时间传播／动力学**（方法 = 平衡态，$\mathcal{C}_{\text{dyn}} = (x_t, \tau)$）：模拟初始结构 x_t 在时间滞后 τ 内的时间演化，生成时间传播后的结构 $x \sim p_\tau(x_{t+\tau}|x_t)$。

需要注意的是，可以为模型设置同时结合动力学条件化与结构预测推理模式的输入，但不建议这样做，因为训练数据中没有使用这种方式的示例。MSA 计算和主干嵌入通常比单次结构模块评估的计算量更大。因此，BioEmu2 会计算一次主干嵌入 $\mathcal{E}_{\text{trunk}}$，并在多个样本或传播步骤中重复使用。

3 结果

本节首先通过成熟的协同折叠基准验证 BioEmu2 在静态结构预测方面的竞争力，并探究推理扩展至大型结构的能力。随后，我们分析

BioEmu2 在平衡采样方面的表现，再转向其对生物分子动力学的建模。最后，我们重点说明引导和增强采样技术如何在推理时控制所获得的系综及生成结构的质量。

3.1 结构预测

FoldBench 与 Runs N' Poses。我们首先评估 BioEmu2 准确预测单个生物分子结构的能力。在结构预测模式下，模型以实验结构数据所特有的结构为目标，例如 X射线晶体学和冷冻电镜结构，而非平衡态系综。在此设定下，我们在广受欢迎的 FoldBench 基准测试套件 [32] 以及用于蛋白质–配体结合结构预测的 Runs N' Poses基准测试 [33] 上，将 BioEmu2 与一系列协同折叠基线模型进行比较。我们选择具有合适商业许可（不包括 AlphaFold3 [5]）且训练数据截止时间相当（不包括 Boltz-2 [39]）的模型进行比较，最终得到以下基线方法：Boltz-1 [28], ESMFold2 [31], OpenFold3 [40], Protenix v1 和 v2 [29]以及 OpenDDE [30]。在对多个样本进行聚合时，我们不使用置信度头。这两个基准测试的我们所用版本的完整细节见第 4.8 节。

在这两项基准测试中，BioEmu2 的准确率均具备与所评估协同折叠基线模型相当的竞争力，确立了结构预测作为我们模型的核心能力（图 2a、b）。具体而言，从 FoldBench 结果可以看到，BioEmu2 在蛋白质和核酸单体上取得了具有竞争力的 lDDT [41] 得分，并在链间结合方面实现了较高的基于 DockQ 的 [42] 成功率。蛋白质–蛋白质结合得分更高，可能是由训练数据的构成所驱动的，因为训练数据侧重于蛋白质–蛋白质相互作用。Runs N' Poses 中的蛋白质–配体评估也突出了 BioEmu2 在建模配体定位以及其周围蛋白质结构方面的能力 [33]，其性能水平可与其他优秀模型相媲美。在第 3.4 节所示的比较中，物理性引导采样基本保持了结构预测准确率，并在某些类别中带来了小幅提升（图 6b）。

将推理扩展至大型复合物。 BioEmu2 的结构预测能力可扩展至更大型的分子组装体。为了在推理时支持更大型的生物分子系统，BioEmu2 实现了上下文并行（CP），将成对表示分布到多个 GPU 上 [43]。在 NVIDIA B200 GPU 上进行的基准测试表明，该方法可从单个 GPU 扩展至 4 个 GPU（2×2）和 16 个 GPU（4×4）配置，支持包含约 16,000 个残基的系统。图 2c 显示，我们的 CP 代码显著扩展了可访问系统大小的上限，同时将通信开销控制在可管理范围内。在这一规模下，预测的物理性比计算可行性更容易成为瓶颈。

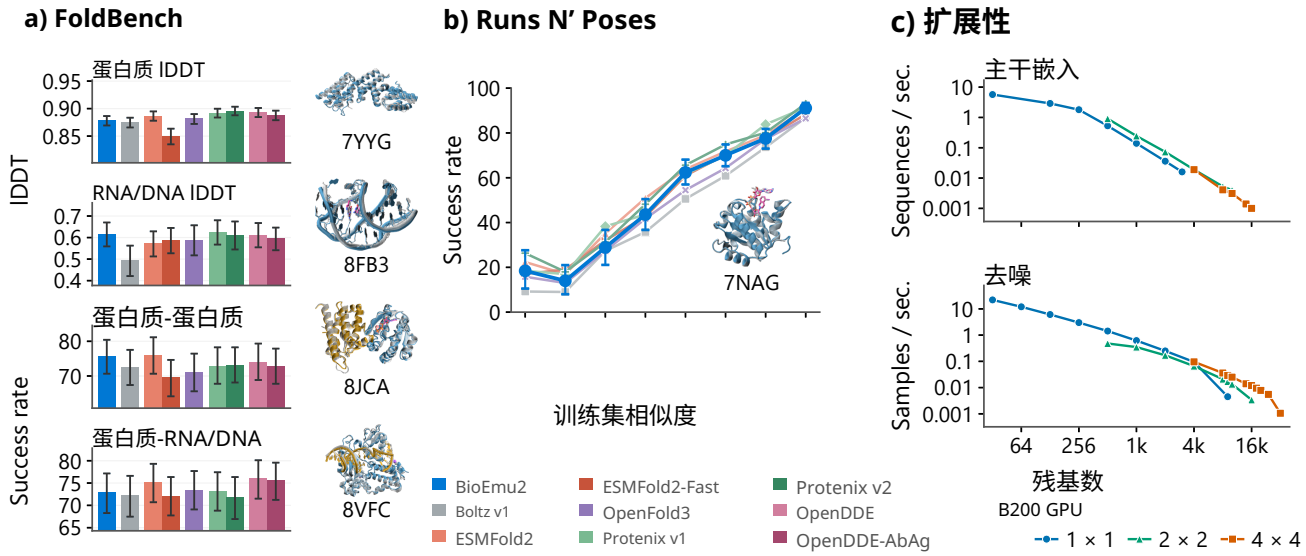
对于由数千个残基组成的各种大型蛋白质复合物，BioEmu2 均能保持局部结构准确性和整体组装构型。测试集中的四种多聚体复合物便是其中的例子：9ISQ（3,508 个残基）、8DOL（3,558 个残基）、9Q3J（4,040 个残基）和 8WW8（5,832 个残基），其主链 lDDT 得分分别为 0.81、0.87、0.94 和 0.80，全局均方根偏差分别为 5.5Å、5.7 Å、2.0Å 和 6.6 Å（图 2d）。

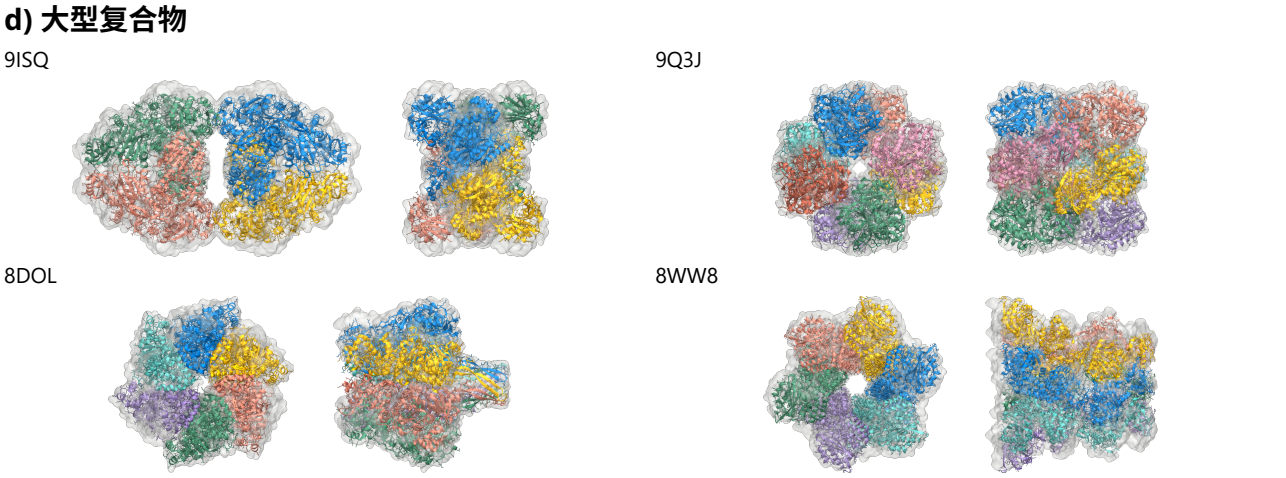
随着复合物规模不断增大，直接获得成功结构预测的概率往往会降低——这既是因为结构预测问题更加困难，也因为大型复合物往往是在细胞中组装而成，而非在溶液中形成，因此其结构通常并未得到明确定义。在这种情况下，只需提供少量额外提示，例如根据实验知识获得的残基间接触，就足以显著提升 BioEmu2 的预测效果。例如，对于含有 4,020 个残基的复合物 8WY8，只指定一个链间接触，就能大幅提高预测准确性，将均方根偏差从 71.2 Å 降至 5.0 Å，并将 DockQ 从 0.10 提高到 0.46（图 2e）。

3.2 平衡态采样

在平衡态模式下，BioEmu2 会针对指定系统，从学习到的结构分布中生成相互独立的样本。此模式并非旨在预测单一的尖锐状态，而是力求再现环境条件（室温和大气压）下，结构在水溶剂中的分布。与结构预测模式相比，这通常会带来更高的灵活性和更浅的能量极小值（图 3a）。

BioEmu2 是一种玻尔兹曼模拟器：与玻尔兹曼生成器不同， [21], 后者从已知力场指定的分布 $p_{eq}(x)$中采样；BioEmu2 则从数据中学习对 $p_\theta(x)$的近似采样方法，并能推广到不同蛋白质和生物分子系统。这延续了早期数据驱动系综模型的研究，包括 AlphaFlow [23] 和 BioEmu





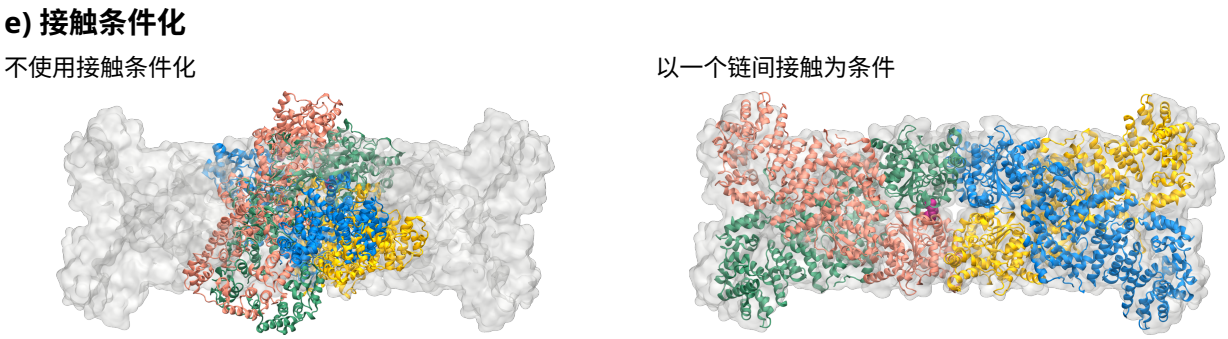


图2：BioEmu2的结构预测模式评估，以及与采用相近截止日期训练的顶尖开放权重模型的比较。彩色为示例性预测结构，灰色为PDB真实结构。a)FoldBench上的结果，按蛋白质、RNA/DNA、蛋白质复合物和蛋白质–RNA/DNA复合物分组。图例同b）。b) 高分辨率蛋白质–配体复合物的Runs N' Poses基准测试。c) 计算序列嵌入以及使用给定嵌入进行结构采样的计算成本。d) 使用上下文并行化在多个GPU上计算的大型复合物示例。e) 示例说明：对单个残基（C链和D链上的Asn202）进行接触条件化，如何修正8WY8中失败的蛋白质复合物结构预测。

[24]。形式化地，

$$x \sim p_\theta(x|\mathcal{S}) \qquad p_\theta(x|\mathcal{S}) \approx p_{eq}(x|\mathcal{S}) \tag{3.1}$$

其中，$\mathcal{S}$ 是分子系统，x 是所有重原子的三维坐标，p_θ 是 BioEmu2 的平衡分布，旨在近似真实平衡分布 p_{eq}。p_θ 是根据在 300K 温度和 1 atm 压力下、采用 Amber 力场 [44]在等温等压（NpT）系综中进行的显式溶剂全原子分子动力学模拟所生成的结构分布学习得到的。溶剂和模拟体积被边缘化，因此模型能够隐式地学习溶剂效应。

自由能面。为了将 BioEmu2 的 p_θ 分布与分子动力学数据的真实分布p_{eq}进行比较，我们首先将来自这两个分布的样本投影到控制每个测试系统中最慢跃迁过程的两个集体变量上，这些变量通过时间滞后独立成分分析（TICA）[45, 46]计算得到。该投影能够解析亚稳态（长寿命）状态，这些状态（i）可能具有生物学相关性，例如作为结合伙伴所选择的构象；（ii）与快速或局部涨落指标 [47]相比，在比较 p_θ 和 p_{eq} 时构成更具挑战性且信息量更高的测试。我们计算并比较自由能面 $U(x) = -k_B T \ln p_Y(x)$，其中 $p_Y(x)$ 是来自分子动力学或 BioEmu2 的二维 TICA 密度（图 3）。

自由能误差 用于衡量模型再现亚稳态相对占比的准确程度。令 q_n^{eq}和 q_n^θ 分别表示参考分布和模型分布中亚稳态 n 的相对占比，它们通过对 p_{eq} 和 p_θ 在构型空间相应区域内进行积分并归一化得到。随后，我们报告 TICA 投影中发现的 M 个亚稳态的平均绝对误差：

$$\mathrm{MAE}_F = \frac{k_B T}{M} \sum_n \left| \ln q_n^\theta - \ln q_n^{eq} \right|. \tag{3.2}$$

状态覆盖度 则用于衡量模型采样到的参考构象空间比例，并针对样本数量进行归一化。我们将二维 TICA 空间精细离散为多个状态，并报告

$$\mathrm{coverage} = \frac{|C_\theta \cap C_{\mathrm{ref}}|}{\widetilde{c}_{\mathrm{ref}}(N)}, \tag{3.3}$$

其中，C_θ 和 C_{ref} 分别是模型访问过的状态集合和参考分布访问过的状态集合，$\widetilde{c}_{\mathrm{ref}}(N)$表示从参考分布中独立抽取一个大小相同的样本时，预期访问的不同状态数量。当模型达到预期的参考覆盖度时，覆盖度等于一。

综合来看，这些指标能够体现 BioEmu2 是否既找到了相关构象状态，又再现了它们的相对占比。

采样模式与功能动力学。BioEmu2 可在标准结构预测和平衡采样之间进行选择。我们以腺苷酸激酶为例，探究这两种模式的影响；该酶在底物结合后会发生显著的构象变化。使用 BioEmu2 的结构预测和平衡采样模式，我们为无配体态（PDB 结构 4AKE）、ATP 结合态以及 AMP+ATP 结合态（PDB 结构 1AKE）生成构象（图 3a）。在两种模式下，BioEmu2 都会采样到接近无配体态和全酶态的结构；当向系统中加入 ATP 和 AMP+ATP 后，所采样构象之间的相对权重会发生显著变化，这与实验测定的 PDB 结构（星号）一致。通过在两种模式之间切换，BioEmu2 从分布较窄的结构预测转变为连续分布的状态系综，从而捕捉生物分子系统中热涨落的本质。后续实验均在平衡采样模式下进行，并对 BioEmu2 生成的系综质量进行定量评估。

分子动力学模拟。我们现在评估 BioEmu2 在一组测试系统上的表现，这些系统代表了几种重要的生物分子模态：单体蛋白质、蛋白质二聚体和蛋白质–小分子系统。对于蛋白质和蛋白质复合物，BioEmu2 采样到约 50% 的分子动力学所发现的状态；对于蛋白质–配体系统，平均采样到超过 100% 的状态，也就是说，采样到的状态数量超过了等量分子动力学帧所采样到的状态数量（图3b）。在采样到的状态中，自由能误差范围为 0.7 至 1.2 千卡/摩尔（图3b）。BioEmu2

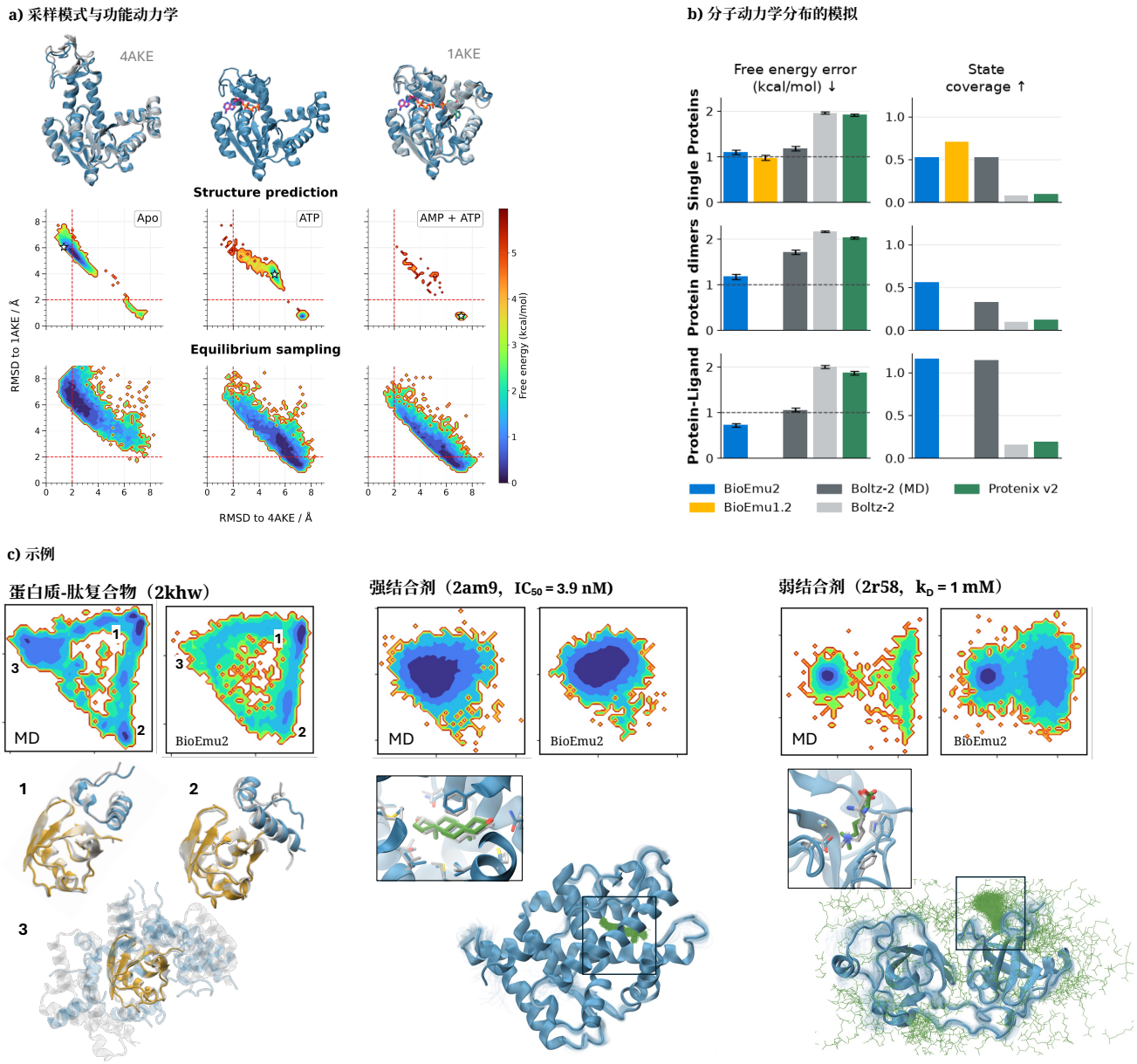


图 3：**功能动力学与分子动力学模拟。a**）在无配体、结合 ATP 和结合 AMP+ATP 条件下，腺苷酸激酶在开放态（4AKE）和闭合态（1AKE）之间的构象采样。1AKE 与 4AKE 的均方根偏差投影展示了标准结构预测（中间）与平衡采样（底部）模式之间的差异。**b**）分子动力学分布模拟在单体蛋白质、蛋白质二聚体和蛋白质–配体复合物上的表现。评估方法（BioEmu2、BioEmu1.2、Boltz-2 (MD)、Boltz-2 和 Protenix v2）的指标包括自由能误差（kcal mol^{-1}，越低越好）和构象状态覆盖度（越高越好）。**c**）代表性示例：比较分子动力学与 BioEmu2 的自由能景观及其对应的结构集合，包括多状态蛋白质–肽复合物（PDB 2KHW，状态 1–3）、强结合剂（PDB 2AM9，$\text{IC}_{50} = 3.9\text{nM}$）和表现出构象柔性的弱结合剂（PDB 2R58，$K_{\text{d}} = 1\text{mM}$）。

在状态覆盖率和自由能误差两方面，BioEmu2 的表现都显著优于未经分子动力学数据训练的模型（此处以 Proteinx v2 和 Boltz-2 默认模式为例）。在单蛋白质动力学方面，BioEmu1 仍然表现最佳，这可能是因为它经过了实验测定的蛋白质稳定性数据微调，而当前发布的 BioEmu2 尚未经过此类微调。不过，BioEmu1 是单蛋白质模型，因此无法在蛋白质–蛋白质和蛋白质–配体类别中与其他模型竞争。Boltz-2 具有分子动力学模式，在该模式下，其状态覆盖度接近 BioEmu2，但 BioEmu2 的自由能误差显著更低。

图 3c 展示了三个具体示例系统的 TICA 能量景观和结构，并将其与分子动力学参考进行比较。BioEmu2 始终能再现 TICA 能量景观，找到所有模式，并较为准确地估计它们的相对占比，不过亚稳态之间的能垒往往略低。以一个蛋白质–肽系统为例，我们发现了两种肽结合模式以及一种未结合模式，这三种模式在分子动力学参考中也都有体现。此外，对于随机选取的强结合剂和弱结合剂示例，BioEmu2 对结构集合的表现定性上符合预期：对于强结合剂，所有样本都处于结合态，且与晶体结构（2AM9）相符；对于弱结合剂，采样到了结合结构和未结合结构，在未结合态下，配体会在蛋白质表面扩散。

3.3 动力学与动力学过程

分子动力学可以用跃迁密度 $p_\tau(y \mid x)$ 描述，即在构型 x 之后经过物理时间 τ 观察到构型 y 的概率密度。它定义了一个转移算子，可将构象集合向前传播。近期的可迁移生成模型，如 Timewarp [25], MDGen [26] 和隐式转移算子 (ITO) [27, 48]，会隐式学习这种密度，并可直接采样未来构型。BioEmu2 采用了这种方法并进行采样

$$x_{t+\tau} \sim p_{\theta,\tau}(\cdot \mid x_t), \tag{3.4}$$

其中 $p_{\theta,\ \tau}$ 是通过学习分子动力学轨迹中相隔一个时间滞后 τ 的构型对（x_t，$x_{t+\tau}$）获得的。模型生成端点 $x_{t+\tau}$，而不解析中间的运动；迭代这一过程即可得到时间步长比分子动力学的飞秒积分步长长许多个数量级的轨迹。

对于遍历性物理动力学，跃迁密度 $p_\tau(y \mid x)$ 在滞后时间足够长时会失去对起始构型 x 的记忆，并趋近平衡分布 $p_{eq}(y)$。通过合适的参数化（第 4.6 节），BioEmu2 可以复现这一极限，从而在同一个模型中统一其转移算子和平衡分布。随着 τ 的时间尺度增大，时间传播因此会在同一个网络中连续转变为平衡采样；经过 $\tau \gg 1$ ms 后，模型将完全失去对起始构型的记忆。尽管存在这一渐近极限，但在当前版本的 BioEmu2 中，对于较慢的结合、解离、折叠、解折叠及其他重排过程，其预测尚未被证实具有定量准确的动力学。

分子动力学时间尺度上的自相关。 平衡分布匹配并不一定意味着动力学具有真实性，因为不同的动力学可能产生相同的平衡态。检验动力学的一种直接方法，是考察 BioEmu2轨迹沿 MD 所揭示的最慢运动是否能像 MD 一样快速去相关。为此，我们将 78 个系统的 BioEmu2 ITO轨迹与 MD 进行比较：包括 CATH 结构域、PDBbind 蛋白质-配体复合物和蛋白质异源二聚体（图 4）。对于每个系统，我们将 MD 轨迹和 100 条 BioEmu2轨迹投影到 MD 的两个最慢 TICA坐标 IC1 和 IC2 上。自相关 $C(\tau)$ 会随着轨迹逐渐失去对较早构象的记忆而衰减至零，而衰减越快，表示沿该坐标的动力学越快。对于大多数系统，在 $\tau = 100$纳秒处与 MD 的差异处于 MD 数据随机半组数据之间差异的范围内；对于差异较大的情况，BioEmu2 大多衰减得更快（图 4）。无论是通过 $\tau = 100$ 纳秒的一步，还是通过 $\tau = 1$纳秒的 100 步覆盖 100 纳秒，差异都相近，没有明确迹象表明误差会在多步过程中累积。总体而言，我们观察到 BioEmu2 在许多系统中都接近 MD，但有时弛豫得过快。

腺苷酸激酶的配体耦合结构域闭合。 我们通过生成连续的滚动推演（x_0、x_{τ_1}、$x_{\tau_1+\tau_2}$、…、x_T），使用 BioEmu2 模拟分子动力学模拟。在这里，可以选择相等的滞后时间，即 $\tau_1 = \tau_2 = \cdots$，也可以选择不同的滞后时间，例如精细解析初始弛豫过程，并在之后采用较长的时间步长，以覆盖较大的时间尺度T。我们首先介绍腺苷酸激酶（ADK），这是一个配体结合与大幅结构域运动耦合的经典例子：其 LID 和 NMP 结构域在 ATP 和 AMP 上方闭合，形成催化活性状态 [2, 49]。平衡态

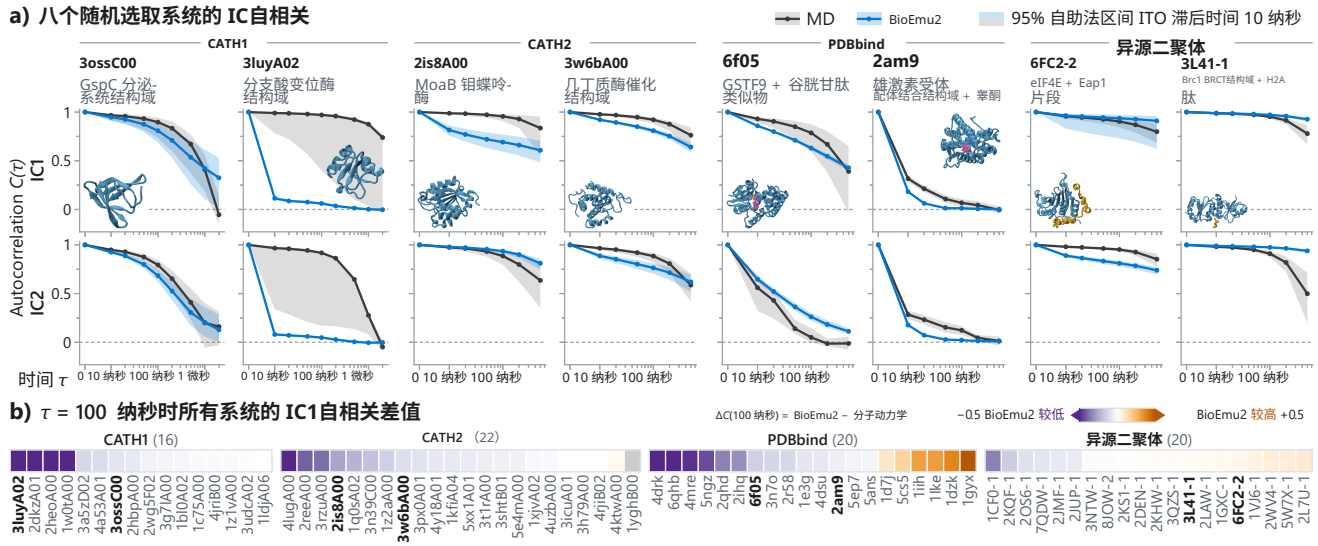


图4：**比较连续 BioEmu2 ITO 展开与分子动力学中最慢 TICA 模态的自相关** $C(\tau)$。**a)** 随机选取的八个系统中，分子动力学与 BioEmu2（95% 区间）的 $C(\tau)$，展示了来自 CATH 结构域、PDBbind 蛋白质-配体复合物和异源二聚体的内坐标 IC1（上）和 IC2（下）。b) ΔC，即在 $\tau = 100$ ns 时，78 个系统中 IC1 的 BioEmu2 自相关减去分子动力学自相关：$\Delta C < 0$ 表示 BioEmu2 中衰减快于分子动力学，$\Delta C > 0$ 表示衰减较慢。

采样表明，BioEmu2 能够捕捉 ADK 的配体依赖状态（图 3a），但无法确定配体结合和结构域闭合的先后顺序。在所示轨迹中，在 50 纳秒的滞后时间下（图 5a–c），我们将 ATP·Mg^{2+} 和 AMP 置于开放的无配体酶旁边，且各自更靠近另一个配体的结合位点。在 3 μ秒内，两个配体都找到各自的结合位点，随后 ATP 大部分时间保持结合状态，而 NMP 结构域几乎经过 30 μ秒才在 AMP 上方闭合，随后不到 1 μ秒后 LID 结构域也闭合。在 43 μ秒时，该酶与闭合晶体结构 1ANK [49], 的均方根偏差处于 0.1 纳米以内，且两个配体均处于其晶体学构象中（图 5c）。在 50 条轨迹中，当配体存在时，约 40% 的轨迹在 50 μ秒内发生闭合；而在配体不存在时，没有一条轨迹发生闭合（图 5d）。因此，BioEmu2 使酶与其配体和 Mg^{2+} 共同经历配体结合与结构域闭合，并为配体诱导的构象转变机制提供了可检验的预测。

PUMA 与 MCL-1 的偶联折叠与结合。许多内在无序蛋白质只有在与伙伴结合时才会折叠，而它们达到结合态结构的路径很难直接观察。PUMA 的 BH3 结构域单独存在时处于无序状态，并在与蛋白质 MCL-1 的沟槽结合时折叠成螺旋，因此它是研究偶联折叠与结合的模型系统 [20, 50, 51]。我们探究 BioEmu2 能否在 PUMA 从沟槽外部起始时找到天然复合物，以及它会沿着怎样的路径到达。我们从一个起始结构运行 500 条轨迹，每条轨迹总长为 $T = 6.5$ ms，滞后时间从 1 ns 增至 5 μs，模拟时间总计超过 3 s。在所示轨迹中（图 5e–g），PUMA 在 MCL-1 表面探索约100 μs，随后在沟槽中折叠成螺旋，但其寄存器偏移了七个残基（约两个螺旋圈）。这种寄存器错位的螺旋持续存在超过半毫秒，直到在单个 1 μs 步内突然转入天然寄存器，且 PUMA 的 D146 形成一个盐桥。之后，PUMA 一直保持在天然寄存器，最终停留在接近核磁共振结构 2ROC 的位置 [50]。在 500 条轨迹中，78% 发生结合（图 5h），其中约四分之三首先经过一个非天然螺旋中间态。

3.4 引导与增强采样

推理时，可以通过引导操控 BioEmu2 的样本生成。这既可用于提升生成样本的物理质量，也可结合增强采样技术提高稀有状态覆盖率 [52]。设 $p_\theta(x)$表示未引导的模型分布，势函数 $v(x)$定义了一个期望的倾斜分布，

$$p_v(x) \propto p_\theta(x) \exp[-v(x)]. \tag{3.5}$$

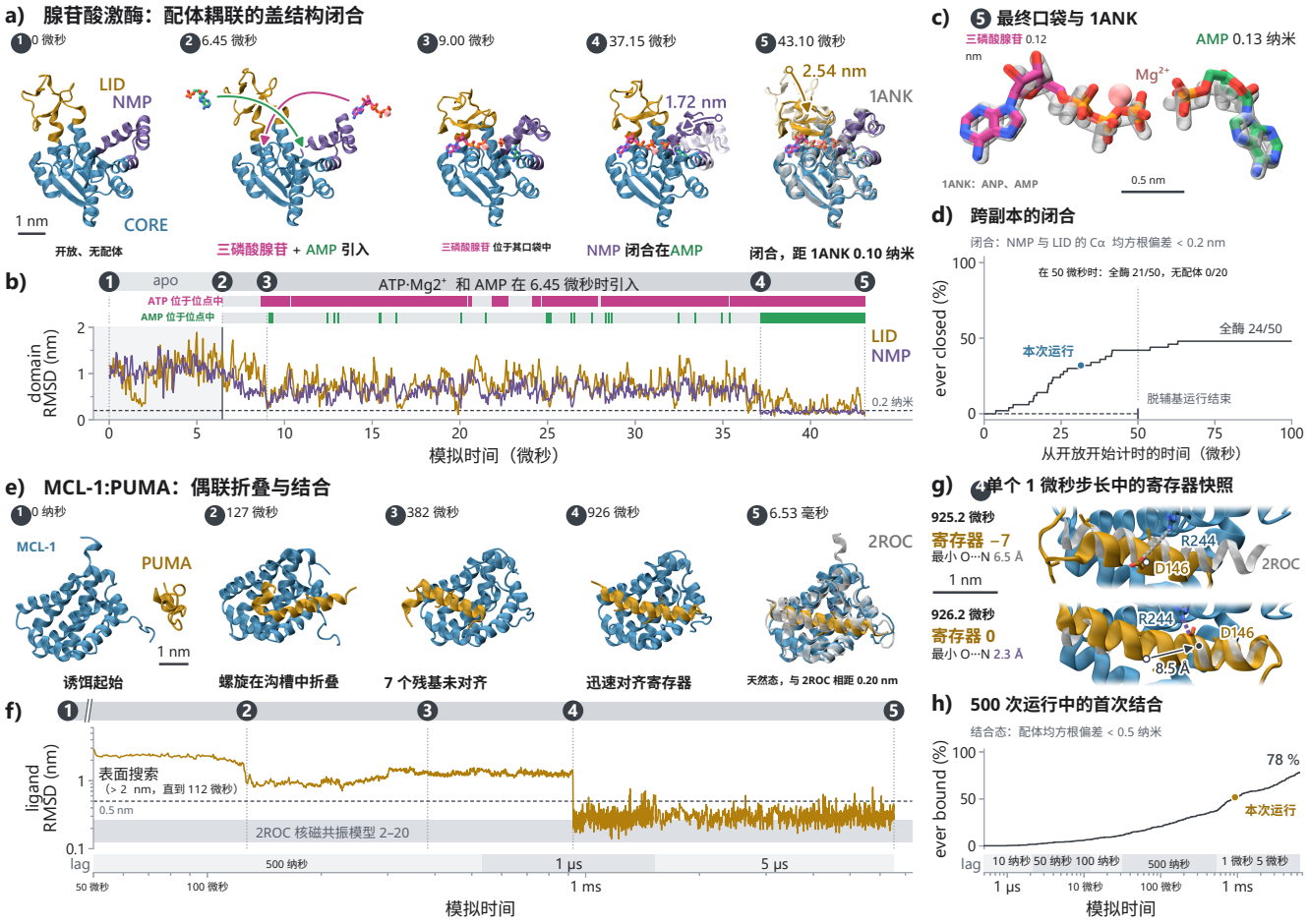


图 5：**BioEmu2 ITO模拟捕捉到腺苷酸激酶中配体耦合的结构域闭合，以及 PUMA 与 MCL-1 的偶联折叠与结合。**快照是通过序贯 ITO 模拟生成的单条轨迹样本；编号事件（ 1–5 ）标记模拟时间上的样本。

腺苷酸激酶 (ADK)。**a)** 腺苷酸激酶的一条轨迹在三磷酸腺苷·Mg^{2+} 之前和之后的快照，以及添加 AMP（滞后 50 纳秒）。最终构象与晶体结构 1ANK 的 Cα 均方根偏差为 0.10 纳米。**b)** 经 CORE 比对后，LID 和 NMP 相对于 1ANK 的 Cα 均方根偏差；条带标示与对应 1ANK 构象均方根偏差在 0.2 纳米以内的配体构象。**c)**最终结合口袋与 1ANK（ANP 和 AMP）叠合。**d)** 已闭合的独立轨迹累计比例。

MCL-1:PUMA。**e)** 一条 PUMA–MCL-1 轨迹中的快照（滞后时间从 1 ns 增至 5 μs）；端点均方根偏差是相对于最接近的 2ROC 核磁共振模型计算的。**f)**在 MCL-1 结合位点上对齐后，PUMA 相对于 2ROC（模型 1）的 Cα 均方根偏差；同时显示核磁共振模型 2–20 的分布范围（色带）、0.5 nm 的结合阈值（虚线）和 ITO 滞后时间（色条）。**g)** 寄存器移动（4）前后的 PUMA D146 和 MCL-1 R244，与 2ROC 叠合（快照间隔 1 μs）。**h)** 已发生结合的独立轨迹累积比例（PUMA Cα 均方根偏差低于 0.5 nm）：500 条轨迹中，262 条在 1 ms 内结合，391 条在 6.53 ms 内结合。

势函数可以惩罚几何缺陷、促进接触或参考构象，也可以将集体变量限制在一定范围内。对于物理合理性导向，我们在分离去噪坐标上使用引导，并可选择进行费曼-卡茨重采样 [28, 53]。对于增强采样，我们通过去噪器对势函数求导，并采用一种改编自扭曲扩散采样器（TDS）的离散转移比加权方案 [54]。

物理合理性导向。对于结构预测，我们使用几何约束来抑制键畸变、空间位阻重叠和立体化学缺陷，同时不强加特定的目标折叠。我们的实现结合了源自 Boltz [28, 39] 的势函数，并加入了显式的聚合物几何和碰撞项，其中包括针对核酸的约束（详情见第 4.7 节）。图 6b 中的结果比较了 FoldBench 基准测试中无引导采样与在去噪最后约 20% 阶段应用物理合理性导向的效果。对于大多数类别，结构准确性基准不受引导影响；不过，DNA单体（0.571 升至 0.604）和 RNA单体（0.660 升至 0.679）的最佳五次平均 lDDT 有显著提升。对于单个 DNA/RNA 系统，提升幅度可能相当可观（图 6b，右侧）。在各个类别中，出现聚合物冲突、键长异常值或键角异常值的引导样本显著减少（图 6b，下方）。对于 DNA 和 RNA 单体，不符合物理规律的样本比例分别从 70.0% 降至 1.4%，以及从 61.3% 降至 4.0%。

稀有态的增强扩散采样。增强采样是一类用于分子动力学模拟的方法，可在自由能差较大时进行自由能计算 [55, 56]。其原理是利用偏置势，人为提高平衡态中稀有态出现的频率，然后对有偏样本重新加权，从而恢复无偏自由能差。我们将同一原理应用于扩散模型采样[52]：TDS [54] 或费曼-卡茨校正器（FKC） [57] 等推理时引导方法会生成有偏系综，多状态Bennett接受比（MBAR） [58–60]则将这些系综结合起来，以估算零偏置下的模型分布。

我们以磷酸甘油酸激酶（PGK）为例，说明在已知另一种参考构象、但无引导样本很少覆盖该构象且模型下的相对布居未明确的情况下，BioEmu2 如何进行增强采样。对于 PGK，我们根据两个结构域相对于已知闭合构象 2X13 的相对位姿来定义偏置势。更多细节见第 4.7 节。在图6c中，我们比较了总样本量相同的无引导采样与增强采样。增强采样在无引导时采样充分的区域保留了定量上相近的自由能景观，并沿着类闭合参考构象分支进一步延伸。代表性构象展示了占主导地位的上臂区域（A）、类开放参考构象（B）和类闭合参考构象（C）。

4 方法

4.1 分子系统定义

我们将分子系统描述为

$$\mathcal{S} = (\mathcal{E}, \mathcal{I}, \mathcal{B}), \tag{4.1}$$

其中，$\mathcal{E}$ 表示分子实体，$\mathcal{I}$ 表示其链副本，$\mathcal{B}$ 表示其共价连接关系。这种划分将分子身份与化学计量区分开来：两个相同的蛋白质链共享一个实体定义，但在生成的复合物中仍是不同的链。该表示遵循 PDBx/mmCIF 的实体与链组织方式，以及 AlphaFold 3 [5] 的全原子建模方法。

化学组成与连接关系。系统定义支持单体、多链组装体，以及包含蛋白质、RNA、DNA、修饰残基、小分子和支化组分的复合物。蛋白质由氨基酸序列指定，RNA 和 DNA 则由碱基序列指定。化学组分可通过化学组分词典（CCD）代码 [61], 指定，该代码描述原子身份、化学键和立体化学；也可通过受支持的基于文本的化学编码指定。包括脂质和离子在内的配体可使用 CCD 代码、SMILES 或 InChI；SMILES 和 InChI 直接描述化学结构。

修饰的和非天然聚合物残基通过其序列位置以及 CCD代码或 InChI 指定。这会替换指定位置上的化学身份。例如，可以在苏氨酸位置使用 CCD代码 TPO 指定磷酸苏氨酸。

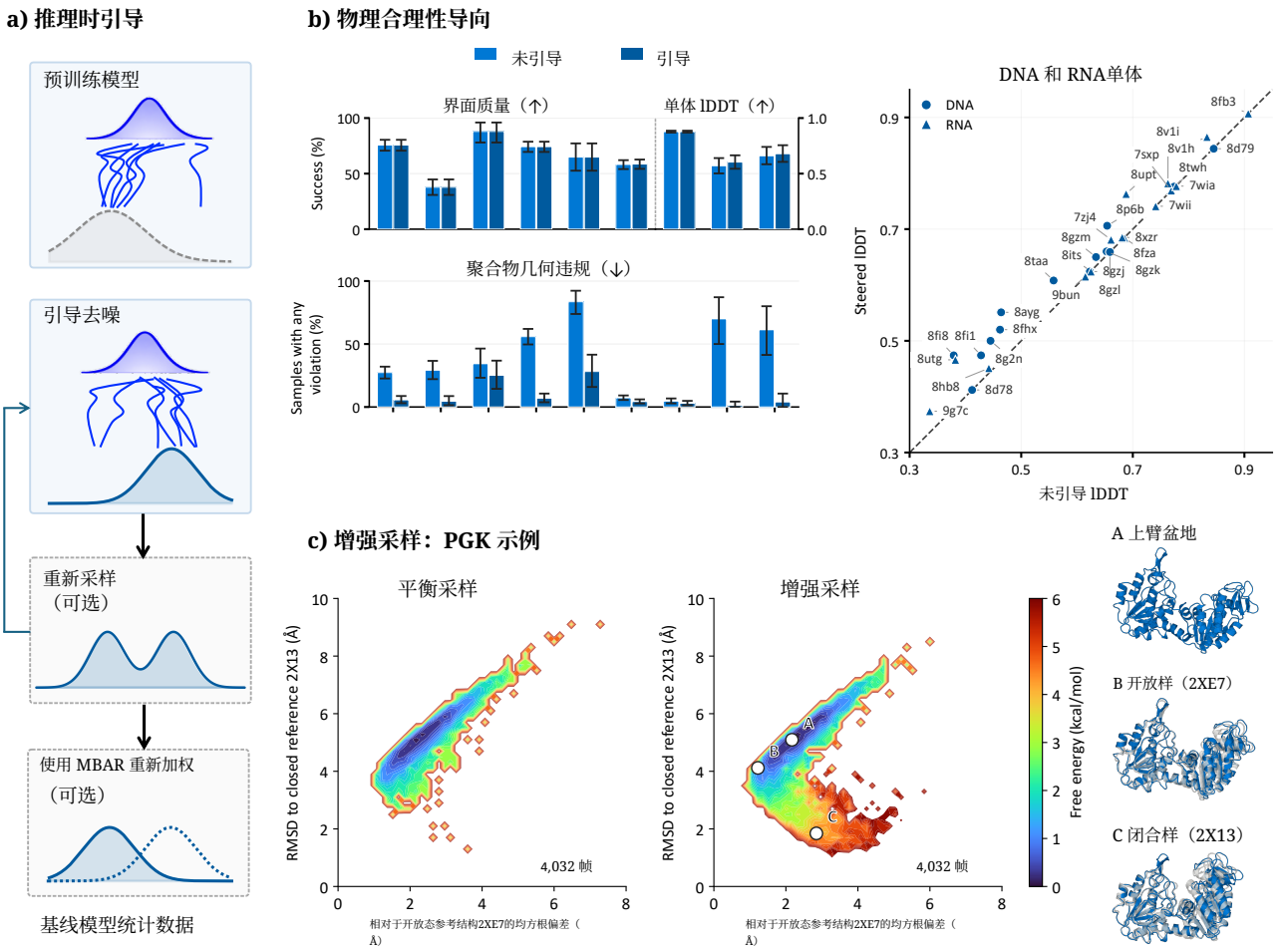


图 6：**针对聚合物几何的推理时引导与增强构象采样。a)**推理时引导示意图。包括引导、可选的粒子重采样，以及可选的采样后重加权，以估计基线模型统计数据。b) FoldBench 上未引导（浅蓝色）与引导（深蓝色）的结果。上方柱状图显示五次采样中最佳的界面成功率和平均预言单体 lDDT；下方柱状图显示所有样本中出现任意 OST聚合物冲突、键长或键角违规的比例。误差线：95% 自助法区间。散点图比较了 14 个 DNA（圆形）和 15 个 RNA（三角形）单体各自独立选出的预言 lDDT。c) 脱辅基PGK：无引导采样，以及结合 TDS 偏置样本进行 MBAR 重加权的零偏置估计。坐标为相对于开放态2XE7和闭合样 2X13 的 Cα 均方根偏差。颜色表示 300 K 下以千卡/摩尔为单位的自由能。A–C 标示上臂、开放样和闭合样构象；预测结果为蓝色，PDB 参考结构为灰色。采样、对齐和重加权的详细信息见第 4.7 节。

标准聚合物连接根据序列顺序构建。可以添加额外的共价键，以指定链顺序无法确定的连接关系，例如二硫键或链间共价连接。链定义、链副本和共价键共同决定分子拓扑。支化组分可以表示，但当前训练数据中不包含这类组分。

原子级和词元级表示。 系统被展开为原子级图和词元序列，遵循全原子结构预测中使用的混合分辨率表示 [5]。每个常规聚合物残基对应一个词元；配体、修饰残基和支化单体则每个重原子对应一个词元。因此，词元是计算单元，并非对生成结构进行统一的粗粒化描述。

4.2 多序列比对

蛋白质多序列比对（MSAs）提供了与分子组成互补的进化信息。序列保守性和相关替换有助于模型识别蛋白质内部及相互作用链之间的结构约束，这与基于多序列比对的结构预测方法 [4, 35] 类似。BioEmu2 将这些输入用于结构预测和平衡态模拟。

链比对与配对。 每种不同的蛋白质序列都会分别进行独立比对。默认情况下，通过 ColabFold [35] 搜索多序列比对。每个比对都包含查询序列，以及与其残基位置对齐的同源序列。非蛋白质组分无需进行进化比对。

对于异源多聚体系统，配对比对还会关联不同蛋白质链之间的同源序列，例如，在同一生物体的基因组中匹配这两种序列的出现位置。配对行提供的对应关系有助于推断分子间相互作用；未配对比对则提供链内进化背景，但不会表明两条链各自比对之间存在特殊关系。BioEmu2 会结合这两类行，并记录哪些位置来自配对比对。

示意而言，比对准备过程会生成

$$\{\text{chain alignments}\} \longrightarrow (A, d, q) \longrightarrow \text{MSA features}, \tag{4.2}$$

其中， A_{ri} 是第 r 行在标记位置 i处的比对符号， d_{ri} 记录相对于查询序列的插入信息，而 q_{ri} 是表示是否配对的二元标记。空位表示缺失的比对位置，以及在给定行中没有序列的链。将这些特征拼接后即可得到多序列比对特征向量 u_{ri}。此外，还会通过对多序列比对各行进行简单的均值池化，将多序列比对信息的粗略摘要纳入主干。

4.3 序列主干与条件信息

序列主干将化学和进化输入转换为主干嵌入 $E_{\text{trunk}} = (E_{\text{single}}, E_{\text{pair}})$。对于$L$个令牌，单令牌嵌入和令牌对嵌入的形状为

$$E_{\text{single}} \in \mathbb{R}^{L \times d_s}, \qquad E_{\text{pair}} \in \mathbb{R}^{L \times L \times d_z}, \tag{4.3}$$

其中， d_s 和 d_z 分别是单令牌和令牌对的特征维度。在 BioEmu2 模型中，我们设定 $d_s = 384$， $d_z = 256$。主干的输出取决于分子系统、比对和条件信息。条件输入$\mathcal{C}_{\text{trunk}}$ 由模式标签 $m \in${平衡态、结构预测} 以及可选的接触集合$\mathcal{H}$组成。随后对主干进行如下计算：

$$(E_{\text{single}}, E_{\text{pair}}) = F_\theta(\mathcal{S}, \text{MSA}, m, \mathcal{H}). \tag{4.4}$$

这种分离遵循 AlphaFold 3 和 Boltz [5, 28] 的表征与扩散架构。它使一次表征计算能够在条件信息不变的情况下支持多个坐标样本。

我们采用了一种 PairMixer 序列主干 [62] ，该主干省略了 PairFormer [5] 中的三角注意力和单表征更新。包括 Minifold [63]，在内的早期研究也表明，更简单的主干架构能够达到与 PairFormer 基本相同的性能。

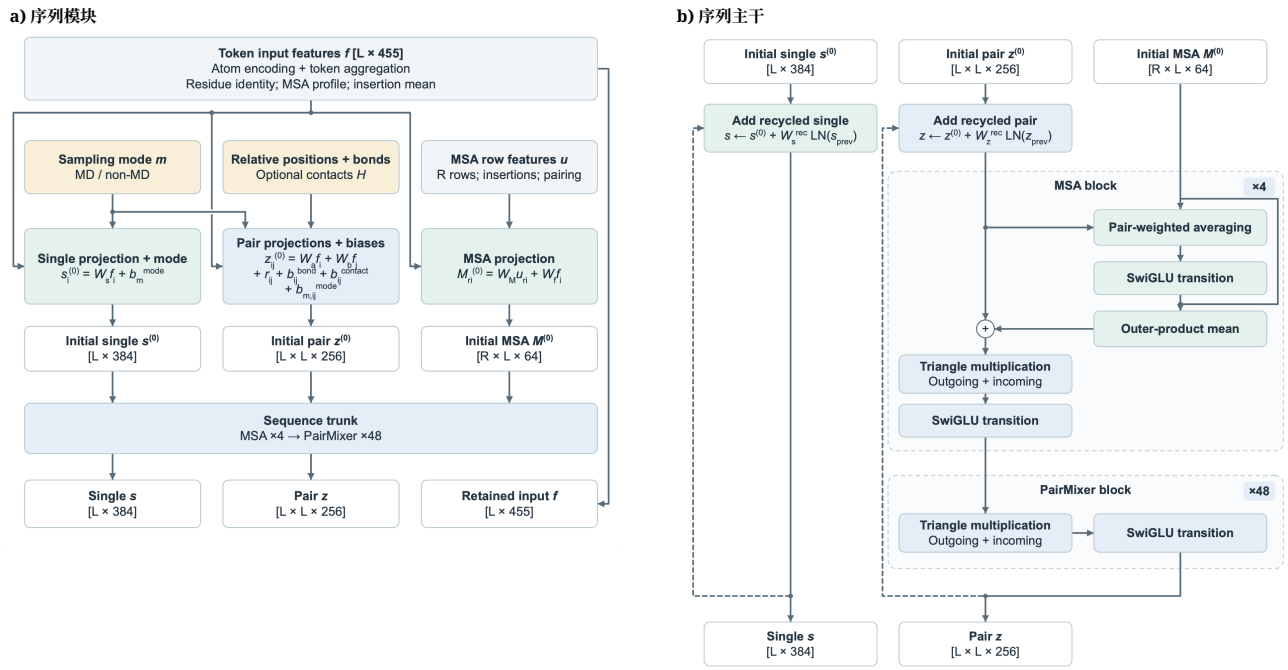


图7：**a) 序列模块**。 共享的词元输入 f 被投影为初始单表示和成对表示$s^{(0)}$ 和 $z^{(0)}$，并为初始多序列比对表示 $M^{(0)}$提供信息。相同的变量也会输入下方的主干示意图。**b) 序列主干**。 多序列比对处理和成对混合会对成对表示进行优化，并可选择回收单表示和成对状态。单表示的维度为 384，成对表示的维度为 256，词元输入的维度为 455。

初始特征。 主干的初始原子特征对电荷、元素、名称和构象参考位置进行编码（从而给出孤立状态下单个残基、核苷酸和小分子的规范几何构型）。随后，这些特征会经过处理并汇聚到词元级别。为表示残基或核苷酸身份（用于蛋白质链和核酸链），以及多序列比对特征的粗略摘要（用于蛋白质链），系统还会添加额外的初始令牌特征。455 的词元输入特征宽度由以下部分拼接而成：384 维的聚合原子编码、33 维的残基类型编码、34 维的粗略多序列比对特征向量，以及四个未使用的通道。

给定此特征向量 f_i，单表示和成对表示初始化如下：

$$\begin{aligned} s_i^{(0)} &= W_s f_i, \\ z_{ij}^{(0)} &= W_a f_i + W_b f_j + r_{ij} + b_{ij}^{\text{bond}}, \end{aligned} \tag{4.5}$$

其中， r_{ij} 编码了关于 i、j 相对位置的信息，例如它们是否属于同一条链或同一实体，以及它们的序列间隔；b_{ij}^{bond}则携带共价键信息。

模式条件化。 采样模式 m 用于区分分子动力学条件化平衡态模拟和非分子动力学结构预测。我们通过一个学习得到的偏置向量实现这一点，该向量会加到每个词元的初始单表示中；另一个偏置则加到初始成对表示中。

接触条件化。沿用 Boltz2 [39], ，接触输入是由词元对和距离阈值构成的列表，

$$\mathcal{H} = \{(i, j, d_{ij}^{\max})\}. \tag{4.6}$$

每个所提供的配对都应落在其阈值范围内。该列表被转换为成对特征：一个类别用于区分选定接触、未选定配对以及接触条件化的缺失。学习得到的接触偏置

添加到成对表示中的部分记为 b_{ij}^{contact}。对于选定的配对，

$$b_{ij}^{\text{contact}} = \text{Embed}_{\text{contact}}(\text{category}_{ij}, d_{ij}^{\max}), \tag{4.7}$$

配对 (j, i) 被对称更新。主干中的接触条件化不同于引导，后者是在推理时影响采样过程的一种手段。

MSA 输入。 行 r 在令牌 i 处的初始 MSA 特征包含残基（或空位）的独热编码，以及插入、缺失和配对的标志。给定这些 MSA 特征 u_{ri}，我们将其与初始令牌特征f_i 相结合，经由

$$M_{ri}^{(0)} = W_M u_{ri} + W_f f_i. \tag{4.8}$$

主干网络层。 主干先使用 MSA 模块，随后接一个 PairMixer 堆栈 [62]。MSA 模块通过配对条件平均耦合序列行，并通过外积聚合将比对信息传递到配对。经过 4 个 MSA 层后，随后的 48 个 PairMixer 层仅对成对表示进行运算。

MSA 模块与 PairMixer 均使用三角乘法，合并共享同一中间令牌的各配对信息。在略去学习得到的投影、门控和归一化后，两种三角运算具有如下形式

$$\Delta z_{ij}^{\text{out}} \sim \sum_k A(z_{ik}) \odot B(z_{jk}), \qquad \Delta z_{ij}^{\text{in}} \sim \sum_k A'(z_{ki}) \odot B'(z_{kj}), \tag{4.9}$$

其中 $\odot$ 表示逐元素乘法。这些运算通过令牌三元组进行通信，使局部化学与进化证据能够为整个体系中的关系提供信息。PairMixer 将向外与向内的三角更新与成对转换网络相结合，而不使用三角注意力层 [62]。主堆栈细化成对特征；单令牌特征则保留其依赖于输入和模式的信息。

回收。 我们重复 MSA 和成对处理阶段，同时将经过归一化和投影的单表示与成对表示馈送回其初始状态。推理时默认使用 3 个回收步骤。训练时，我们采用确定性的回收步骤调度，使所有秩保持同步。

4.4 结构模块与采样

双流结构架构。 我们的结构模块深受 NeuralPLexer3 [64] 提出的 Flow Blocks 启发。其 16 个模块均交错执行原子级和令牌级更新，而非像 AlphaFold 3 [5] 那样将原子编码、令牌变换器和原子解码分开处理。

原子流采用窗口注意力和转换多层感知机；令牌流采用全令牌注意力和转换多层感知机。跨流更新会在指定的令牌中心原子处汇集特征，并将令牌特征广播至其关联原子。两个流均以主干表示和扩散噪声水平为条件信息。对于时间传播，滞后时间特征和提供的起始坐标会作为额外条件信息。堆栈之后接有坐标读出，而 EDM 预条件化将此更新与噪声输入结合，以获得去噪坐标估计。

条件去噪。 生物分子系统经序列主干嵌入后，BioEmu2 的结构模型使用条件扩散模型生成结构。条件定义如下：

$$\mathcal{C} = (\mathcal{S}, E_{\text{single}}, E_{\text{pair}}; [x_t, \tau]), \tag{4.10}$$

- $\mathcal{S}$：系统嵌入部分所述的静态系统特征。
- $E_{\text{单}}$ 和 $E_{\text{对}}$：由序列主干生成的单令牌嵌入和令牌对嵌入。
- $[x_t, \tau]$：初始物理时间下的可选构型，以及用于时间传播的正滞后时间 τ。在提供平衡模态条件化时，这些信息指定了在x_t条件下采样未来端点 $x_{t+\tau}$。

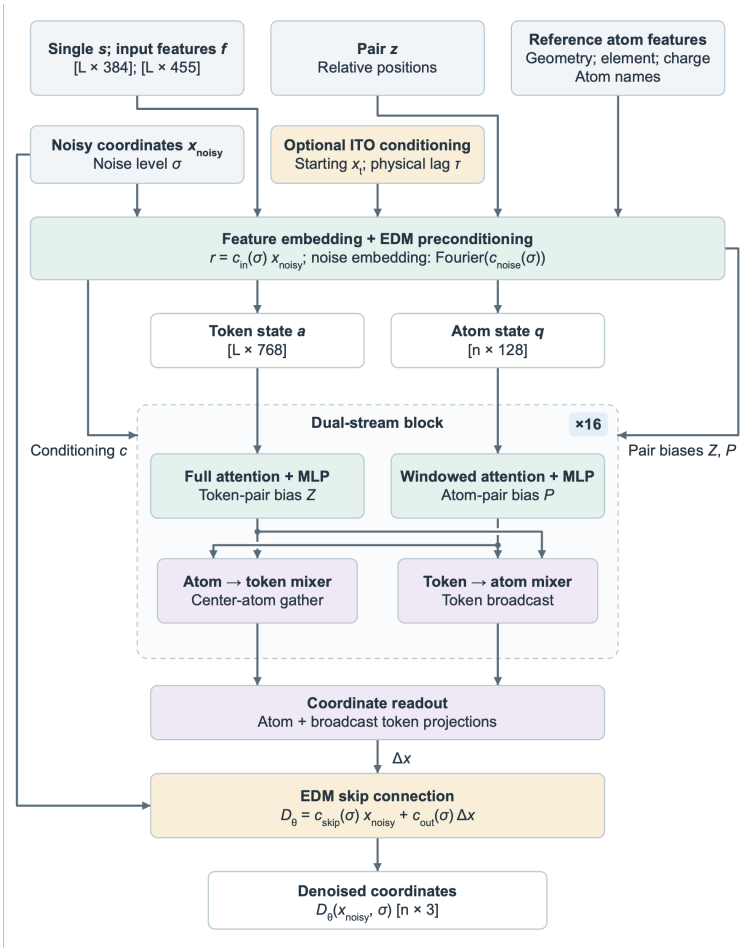


图8：**结构模块**。 输出是评估后的去噪器， $\hat{x} = D_\theta(x_{\text{noisy}}, \sigma \mid \mathcal{C})$。重复采样更新使用该预测来生成坐标。

该扩散模型是一个针对全原子笛卡尔坐标的条件得分模型。在每个去噪步骤中，除了 $\mathcal{C}$ 外，它还会接收当前的含噪坐标 x_{noisy} 及其噪声水平 σ。静态系统特征和嵌入会在各步骤间缓存。得分用于估计经噪声扰动的条件密度的对数梯度，

$$s_\theta(x_{\text{noisy}}, \sigma \mid \mathcal{C}) \approx \nabla_{x_{\text{noisy}}} \log p_\sigma(x_{\text{noisy}} \mid \mathcal{C}), \tag{4.11}$$

其中 p_σ 是目标结构分布与协方差为 $\sigma^2 I$ 的高斯噪声卷积后的分布。BioEmu2 使用去噪器参数化 [65]：网络重建干净坐标 $D_\theta(x_{\text{noisy}}, \sigma \mid \mathcal{C})$，并据此定义得分：

$$s_\theta(x_{\text{noisy}}, \sigma \mid \mathcal{C}) = \frac{D_\theta(x_{\text{noisy}}, \sigma \mid \mathcal{C}) - x_{\text{noisy}}}{\sigma^2}. \tag{4.12}$$

从模型中采样意味着逐步对结构进行去噪。采样从高斯坐标$x_{\text{noisy},\ \sigma_{\max}} \sim \mathcal{N}(0,\ \sigma_{\max}^2 I)$开始，并沿着递减的噪声调度 [65]逐步对其去噪。不进行 ITO 条件化时，结果是从所选结构分布中抽取的样本。使用 x_t 和 τ时，结果是一个未来构型 $x_{t+\tau}$。（噪声水平 σ和采样器的去噪步骤不同于物理滞后时间 τ。）

ITO 输入的处理。时间条件输入 $[x_t, \tau]$仅通过结构模块传入。在训练时，我们对结构 x_t 施加噪声 $N(0, \varsigma^2)$，以提高模型对原始结构误差（$\varsigma = 0.1$Å）的鲁棒性。我们通过 $e_k(\tau) = e^{-\tau/s_k}\{1, \sin(2\pi q\tau/s_k), \cos(2\pi q\tau/s_k)\}$对 τ进行特征化，并将其在不同衰减率 s_k 和不同 q上拼接。这些特征与扩散噪声尺度嵌入一同进入单一条件流。时间条件坐标 x_t 与含噪坐标 x_{noisy}一同传入，并乘以门控 $g(\tau)$；该门控的构造方式确保了当 $\tau \to \infty$时，$g \to 0$。

EDM 预条件化与推理更新。我们遵循 EDM 扩散模型的设定 [65]。令 Δx_θ表示网络的坐标更新读出，则 EDM 去噪器 [65] 为

$$D_\theta(x_{\text{noisy}}, \sigma \mid \mathcal{C}) = c_{\text{skip}}(\sigma) x_{\text{noisy}} + c_{\text{out}}(\sigma) \Delta x_\theta(c_{\text{in}}(\sigma) x_{\text{noisy}}, c_{\text{noise}}(\sigma) \mid \mathcal{C}), \tag{4.13}$$

with

$$\begin{aligned} c_{\text{skip}}(\sigma) &= \frac{\sigma_{\text{data}}^2}{\sigma^2 + \sigma_{\text{data}}^2}, \qquad c_{\text{out}}(\sigma) = \frac{\sigma \sigma_{\text{data}}}{\sqrt{\sigma^2 + \sigma_{\text{data}}^2}}, \\ c_{\text{in}}(\sigma) &= \frac{1}{\sqrt{\sigma^2 + \sigma_{\text{data}}^2}}, \quad c_{\text{noise}}(\sigma) = \frac{1}{4} \log \frac{\sigma}{\sigma_{\text{data}}}. \end{aligned} \tag{4.14}$$

在这里，σ_{data} 设定用于预条件化的坐标尺度；在我们的案例中，该尺度为 16Å。

设 $\sigma_0 > \cdots > \sigma_K = 0$ 为推理噪声调度，并令 $x^{(0)} \sim \mathcal{N}(0, \sigma_0^2 I)$。在步骤 k中，噪声注入（“噪声扰动”）会暂时提高噪声水平，

$$\widehat{\sigma}_k = (1 + \gamma_k)\sigma_k, \qquad x_{\text{noisy},k} = x^{(k)} + S_{\text{noise}} \sqrt{\widehat{\sigma}_k^2 - \sigma_k^2}\, \epsilon_k, \qquad \epsilon_k \sim \mathcal{N}(0, I). \tag{4.15}$$

当噪声水平低于某个阈值时，我们关闭噪声扰动，只需设定 $x_{\text{noisy},k} = x^{(k)}$ 和 $\widehat{\sigma}_k = \sigma_k$。给定结构$x_{\text{noisy},k}$，一次去噪器评估得到 $\widehat{x}_k = D_\theta(x_{\text{noisy},k}, \widehat{\sigma}_k \mid \mathcal{C})$，随后

$$x_{\text{update},k} = \frac{x_{\text{noisy},k} - \widehat{x}_k}{\widehat{\sigma}_k}, \qquad x^{(k+1)} = x_{\text{noisy},k} + S_{\text{step}}(\sigma_{k+1} - \widehat{\sigma}_k) x_{\text{update},k}. \tag{4.16}$$

每个步骤都会应用全局刚性增强。如果使用 ITO 条件结构 x_t，则对其应用相同的刚性增强。在推理过程中，与 AF3 中一样，在依据（4.16）进行更新之前，会将每个反向步骤中的含噪坐标进行刚性对齐，使其与去噪预测对齐。

采用直接引导时，在计算 $x_{\text{update},k}$之前，会将 $\widehat{x}_k$ 替换为引导去噪估计。相应的方程和粒子校正变体见第 4.7 节。

结构预测锐化。对于结构预测，我们在 EDM采样框架 [65] 中采用经验锐化，通过相对于注入噪声放大去噪更新。在上述更新中，这相当于在方程 (4.16) 中设置 $S_{\text{step}} = 1.5$，$S_{\text{noise}} = 1.003$，而平衡采样时则设置为 $S_{\text{step}} = S_{\text{noise}} = 1$。这种结构预测修改也用于 AlphaFold 3 [5] 和 Boltz [28, 39]。

4.5 平衡分布采样

我们的分子动力学模拟采用等温等压系综，即 NpT 系综：粒子数 N、外部压力 p和温度 T 保持固定，而模拟体积 V 则会波动。BioEmu2 的目标是在环境条件下进行平衡采样： NpT 训练模拟采用室温 T 和 $p = 1$ atm 的压力。令 x表示 BioEmu2 生成的生物分子坐标，令 z 表示其余构型自由度，包括显式溶剂。对动量积分消去后，完整构象的无量纲约化能为

$$u(x, z, V) = \beta \left[U(x, z; V) + pV \right], \qquad \beta = \frac{1}{k_{\mathrm{B}} T}, \tag{4.17}$$

其中 U 是 势能， k_{B} 是玻尔兹曼常数。联合平衡密度为

$$\rho(x, z, V) = Z^{-1} \exp[-u(x, z, V)], \tag{4.18}$$

其归一化常数为 Z ，并默认采用与体积相关的笛卡尔构型域。与模型相关的平衡分布是对溶剂和体积取边缘分布后的结果，

$$p_{eq}(x) = \int_0^\infty \mathrm{d}V \int \mathrm{d}z \, \rho(x, z, V). \tag{4.19}$$

因此，BioEmu2 隐式地纳入了溶剂效应：它生成生物分子坐标而不包含显式溶剂坐标，同时学习一个已将溶剂和体积自由度边缘化的分布。因此，该结构分布并非简单的孤立蛋白质的玻尔兹曼因子。

BioEmu2 是一种玻尔兹曼模拟器，通过以下方式近似这一平衡分布

$$x \sim p_\theta, \qquad p_\theta(x) \approx p_{eq}(x), \tag{4.20}$$

其中 θ 表示共享学习参数，为便于阅读，省略了对系统的序列嵌入的依赖。平衡采样是一个近似目标，并不能保证对每个系统都能得到校准后的布居或定量自由能。

4.6 时间传播

马尔可夫状态模型（马尔可夫状态模型）将分子动力学描述为在选定滞后时间 τ [66–68]下离散构象状态之间的跃迁。当许多短轨迹共同采样相关状态和跃迁时，马尔可夫状态模型可以恢复长时间弛豫和平衡布居，而无需每条轨迹都达到平衡[67, 68]。然而，传统的马尔可夫状态模型是针对特定系统的分析模型，无法跨分子体系迁移。

近期的工作转向了能够跨分子体系泛化的生成式神经网络。例子包括深度生成式马尔可夫状态模型 [69], Timewarp [25], MDGen [26], 以及隐式转移算子（ITO）框架的可迁移扩展 [48, 70]。BioEmu2遵循原始ITO 公式 [27], , 其中以起始坐标 x_t 和物理滞后时间 τ 为条件的扩散模型进行采样

$$x_{t+\tau} \sim p_{\theta,\tau}(\cdot \mid x_t), \tag{4.21}$$

近似物理跃迁密度 $p_\tau(x_{t+\tau} \mid x_t)$ 。这种表述使 BioEmu2 能够直接生成端点 $x_{t+\tau}$ ，而无需解析中间的分子动力学轨迹。通过重复进行条件采样，可以采样出长时间运行的、有时间顺序的轨迹。由于仅保留生物分子坐标，其他动力学自由度（例如动量和溶剂）通常不会从一个步骤传递到下一个步骤。

平衡极限。 对于遍历系统，动力学会弛豫至唯一的平衡分布，物理跃迁密度也会失去对其初始构型的记忆。对于该模型，相应的一致性条件为

$$\lim_{\tau \to \infty} p_{\theta,\tau}(x_{t+\tau} \mid x_t) = p_\theta(x_{t+\tau}), \tag{4.22}$$

其中 p_θ 是公式（4.20）中所述的平衡模型。我们要求模型与分子动力学训练数据所定义的经验平衡保持一致，而不是像马尔可夫状态模型那样从跃迁中推断平衡态。

为此，我们借鉴 BoPITO [71] 的思路，将跃迁密度拆分为平稳部分和随滞后时间衰减的时变部分，前者由平衡模型给出。实际上，网络并不会分别添加平衡得分和动力学得分，而是随着滞后时间增长，逐渐减弱动力学条件信息，从而消除时变部分。滞后时间特征带有指数包络 $e_k(\tau)$，而对初始构型的条件化则由这些包络的一个经学习得到的正线性组合进行门控：

$$e_k(\tau) = \exp(-\tau/s_k), \qquad g(\tau) = \sum_k w_k e_k(\tau), \qquad w_k > 0, \quad 0 < s_k \le s_{\max}. \tag{4.23}$$

添加到单表示中的滞后时间嵌入采用阻尼谐振子的形式，并以相同的时间尺度衰减。当 $s_{\max} = 1$ ms 时，对于 $\tau \gg s_{\max}$，门控和滞后时间嵌入都会消失，因此只剩下平稳部分，网络会准确恢复平衡去噪器。这种交叉过渡的形式属于架构选择，但权重 w_k 和时间尺度 s_k 是从数据中学习得到的。因此，不应将这一交叉过渡解读为物理弛豫时间的估计：它保证了与所学习的平衡分布的一致性，而该分布不一定与精确的热力学平衡相吻合。

训练。对于一部分训练样本，我们从同一条分子动力学轨迹中抽取帧对（x_t，$x_{t+\tau}$）。我们将目标帧进行刚性对齐，使其与起始帧对齐，然后对两者施加相同的随机旋转和平移，从而保留相对于起始构象的运动。其余分子动力学帧作为独立的平衡样本使用，并屏蔽动力学条件信息，将滞后设为 $\tau = \infty$。这两项任务共享同一个网络和 EDM 去噪损失（第 4.4 节）。

BioEmu2 几乎所有用于动力学训练的数据都来自时长不超过 $1\,\mu$ 秒的分子动力学模拟。因此，对于结合、解离、折叠、解折叠及其他慢速重排等较慢过程，其预测的速率和路径需要进行独立验证。在模型针对这些时间尺度的实验数据完成训练和验证之前，此类预测应视为定性假设，而非定量动力学结果。

滚动推演。每个 ITO 步都会运行完整的扩散采样器，每次去噪器评估都以先前的构象 x_t 和选定的滞后 τ为条件；主干嵌入则针对每个系统计算一次，并在各步之间重复使用。在采样器设置固定的情况下，每步的计算成本与 τ 无关，所需计算量随模拟时间线性增长。滞后可在不同步骤之间变化，起始构型可以是实验结构或预测结构、平衡样本，也可以是经过人工修改的结构，例如在脱辅基蛋白旁放置配体。

4.7 引导与增强采样

引导可在不重新训练模型的情况下，改变推理时采样得到的结构分布。与会改变主干嵌入的模式条件化或接触条件化不同，引导作用于坐标生成。它可以偏向物理几何、特定构象，或未引导模型下出现概率较低的区域。BioEmu2 实现了势能梯度引导和基于粒子的 Feynman-Kac 方法 [53, 54, 57], ，后者也可支持增强采样工作流 [52]。

势函数与倾斜分布。令 $p_{\mathrm{base}}(x)$ 表示固定系统和条件信息下的未引导分布。无量纲势函数 $v(x)$ 用于指定所需的倾斜，

$$p_v(x) = Z_v^{-1} p_{\mathrm{base}}(x) \exp[-v(x)], \qquad Z_v = \int p_{\mathrm{base}}(x) \exp[-v(x)]\, \mathrm{d}x. \tag{4.24}$$

势函数值越低越有利。对于物理偏置能 $B(x)$，对应的无量纲势函数为$v(x) = \beta B(x)$；也可以直接以无量纲形式定义几何惩罚项。多个目标可以通过对各自的势函数求和来组合。倾斜定义了目标分布，但并不保证每种引导过程都能精确地从中采样。

坐标记法。按照第4.4节的记法，i 表示去噪步骤，而非物理时间。坐标 $x^{(i)}$ 是预定噪声水平 σ_i下的当前含噪状态；$x_{\mathrm{noisy},\,i}$ 表示经过数据增强和噪声注入后、处于重新加热后的噪声水平 $\hat{}\sigma_i$的坐标；而 $\hat{}x_i = D_\theta(x_{\mathrm{noisy},\,i},\ \hat{}\sigma_i \mid C)$ 是相应的

带有条件信息 C的干净坐标预测。我们省略粒子索引。这两种方法使用相同的坐标，但再加热调度以及梯度的计算位置不同。

物理合理性导向：几何惩罚。 物理合理性势函数会在坐标生成过程中抑制局部几何缺陷。我们改编 Boltz导向势 [28] ，用于配体距离几何、手性、键立体化学、平面性和链重叠，并通过显式的键长与键角约束以及聚合物碰撞惩罚对其进行扩展。聚合物碰撞惩罚会抑制链内和链间的原子碰撞。通用的链内和链间碰撞惩罚会排除由三条以内共价键连接的原子对。另有一项局部惩罚，用于防止不同残基中符合条件的蛋白质或核酸原子之间出现过近接触；这些原子由两条或三条化学键连接。几何约束采用列表形式的参考值，包括肽和核苷酸连接以及已声明的二硫键；若无匹配的参考值，则使用较宽的边界。

按照 Boltz导向势的做法，可通过平底惩罚来强制执行允许区间。单侧约束会省去相应的边界项。在允许区间内，违约量为零；超出区间后，违约量线性增长；其影响程度由特定势函数的缩放系数和采样器系数决定。对于具有下界和上界 $q_{\min}$ 与 $q_{\max}$的标量几何量 q ，未加权的违约量为

$$h(q;q_{\min},q_{\max})=\max(q_{\min}-q,0)+\max(q-q_{\max},0). \tag{4.25}$$

其他势函 数可以偏向指定的接触、参考构象或集体变量范围。

物理合理性导向：分离坐标引导。我们使用第4.4节中的churn调度，并在不通过去噪器求导的情况下计算$\hat{x}_i$ 。在每个引导扩散步，我们初始化一个单独的优化变量 $y_i^{(0)}=\hat{x}_i$。此处 ℓ 表示内部势函数梯度步数，而非扩散步数； j 表示势函数的索引:

$$y_i^{(\ell+1)}=y_i^{(\ell)}-\sum_j \eta_{j,i,\ell}\nabla_y U_j(y_i^{(\ell)}),\qquad \hat{x}_i^{\,g}=y_i^{(L_i)}. \tag{4.26}$$

经过 L_i 个内部步后，$\hat{x}_i^{\,g}=y_i^{(L_i)}$ 即为引导干净坐标预测；上标 g 表示引导。系数$\eta_{j,\,i,\,\ell}$ 包含每个势函数的引导强度和激活调度，并且引导应用于非终止转移。随后，我们应用式4.16中的EDM更新，以引导后的干净预测$\hat{x}_i^{g}$ 替代$\hat{x}_i$。如果启用了重采样，每个选中的粒子都会同时带上其含噪坐标和引导后的预测。如果启用了刚性对齐，我们会先将这些含噪坐标与该预测对齐。

因此，势函数会修改用于传输的去噪预测，而不是通过去噪器反向传播至其含噪输入。可选的 FK重采样使用引导前评估的加权势能之差，将其乘以全局权重强度，并结合基于所携带引导位移和当前高斯噪声的校正项。

对于图 6b 中的物理性比较，我们采用了不进行粒子重采样的分离坐标梯度引导，该引导在 200 次去噪跃迁的最后约 20% 阶段启用。采样队列和评估流程见第 4.8 节。更好的局部几何结构并不保证更优的平衡布居，但有助于保持并改善结构预测指标，同时显著减少非物理性。

增强采样：经去噪器引导和 TDS 权重。 对于 PGK 示例，我们改用扭曲扩散采样器（TDS） [54]。设 $v_i(x)$ 表示第 i 步干净坐标上的偏置，其中包括任何随步变化的扩展性或参数变化。含噪坐标上对应的偏置为 $V_i(x_{\text{noisy}})=v_i(D_\theta(x_{\text{noisy}},\ \hat{\sigma}_i\mid C))$ 。这里的 V_i 是一个通过去噪器依赖于含噪输入坐标的函数，不同于针对分离的 y 所进行的物理性优化。我们使用西格玛再加热，而不是物理性扰动调度：对于 $i>0$，$\hat{\sigma}_0=\sigma_0$ 和 $\hat{\sigma}_i=\sigma_{i-1}$，注入噪声方差为 $\nu_i=\hat{\sigma}_i^2-\sigma_i^2$。在 $S_{\text{step}}=1$ 下，非终端提议为

$$x^{(i+1)}=x_{\text{noisy},i}+\frac{\hat{\sigma}_i-\sigma_{i+1}}{\hat{\sigma}_i}\left[D_\theta(x_{\text{noisy},i},\hat{\sigma}_i\mid C)-x_{\text{noisy},i}\right]-\frac{\nu_i}{2}\nabla_{x_{\text{noisy},i}}V_i(x_{\text{noisy},i}), \tag{4.27}$$

梯度通过去噪器作用于含噪坐标 $x_{\text{noisy},\ i}$ 。令 w_i 表示再加热状态 $x_{\text{noisy},\ i}$下的未归一化粒子权重。对数权重在重采样事件之间按以下方式累积:

$$\log w_i=\log w_{i-1}-V_i(x_{\text{noisy},i})+V_{i-1}(x_{\text{noisy},i-1})+\log\frac{K_i}{Q_i},\qquad i\geq 1. \tag{4.28}$$

这里， K_i 和 Q_i 分别是参考转移密度和引导提案转移密度，在 $x_{\mathrm{noisy},\,i}$ 处求值。它们分别结合了此前的无引导或有引导输运，以及当前的高斯刷新，并以相同的采样刚性增强为条件。如果它们的均值相差 $\delta_i = m_{Q,\,i} - m_{K,\,i}$，且 $r_i = x_{\mathrm{noisy},\,i} - m_{Q,\,i}$ 是采样得到的高斯残差，那么 $\log(K_i/Q_i) = -r_i^{\mathsf{T}}\delta_i/\nu_i - \|\delta_i\|^2/(2\nu_i)$。在应用共享的旋转和平移之前，均值会分别进行中心化。

初始对数权重为 $-V_0(x_{\mathrm{noisy},\,0})$，对于恒定强度引导，它不一定为零。在最后一次转移中m， $\sigma_{m+1} = 0$ 引导被禁用。将式 4.27 中的引导项设为零，并令 $a = 1$，即可得到完成采样 $X = x^{(m+1)} = D_\theta(x_{\mathrm{noisy},\,m},\ \hat{\sigma}_m \mid C)$；最终对数权重增量为 $-v(X) + V_m(x_{\mathrm{noisy},\,m})$。由此得到的目标分布为 $p_v(X) \propto p_\theta(X)e^{-v(X)}$，其中 p_θ 是有限步无引导参考采样器的端点分布。有限粒子群和权重退化会限制对该目标分布的估计。我们的 PGK 示例采用了式 4.28 [54], 中的离散 TDS 核比率，而非先前工作 [52] 中的 Feynman–Kac 校正器（FKC）权重规则 [57] 。与 FKC 不同，TDS 规则不需要 $\partial_\sigma U$，即经由去噪器计算的势函数噪声水平导数。

增强采样：粒子重采样。我们在每个粒子组内对权重进行归一化，并计算$N_{\mathrm{eff}} = 1/\sum_{n=1}^{N_p} \bar{w}_n^2$。TDS 实现支持按设定的间隔进行重采样，或在任一组的占比 N_{eff}/N_p 低于阈值时进行重采样。系统重采样会复制所有依赖祖先的状态，并重置累积对数权重；提供势函数时，还会执行最终重采样。返回的粒子权重相等，但可能拥有共同的祖先，因此不应将其视为独立观测。

增强采样：恢复无偏模型统计量。偏置势可以使未引导模型很少采样到的构象更受青睐 [52]。在倾斜系综中，结构 x 的概率与其未引导概率 $p_{\mathrm{base}}(x)$乘以 $\exp[-v(x)]$成正比。因此，只要偏置系综与未引导系综之间有足够的重叠，我们就可以通过对每个结构赋予逆因子权重，来估计可观测量 A 的未引导均值：

$$\langle A\rangle_{\mathrm{base}} = \frac{\langle A(x)\exp[v(x)]\rangle_v}{\langle \exp[v(x)]\rangle_v}. \tag{4.29}$$

当使用多个窗口施加不同偏置时，可以通过使用 MBAR 进行多状态重加权来估计它们的相对归一化常数 [60]。需要指出的是，引导和重加权方法可以用于稀有事件采样，以恢复基础模型的统计量；但它们无法消除基础模型本身的误差。

示例：PGK结构域姿态增强采样。 为说明已知的另一种构象如何引导探索而不预设其群体比例，我们在平衡模态条件化下对无配体PGK进行采样，并将其N端和C端核心的相对平移与旋转偏置到闭合参考结构2X13。结果见图6c。我们使用164个N端核心原子和186个C端核心原子，将居中后的闭合参考结构核心拟合到相应的采样Cα坐标。设 R_N， R_C为采用行向量约定拟合得到的正规旋转， c_N， c_C为采样得到的核心质心， $c_N^\star$， $c_C^\star$为其闭合参考结构质心。偏置集体变量为N端核心坐标系中的平移误差和相对旋转弦误差：

$$d(X) = \left\|(c_C - c_N)R_N^{\mathsf{T}} - (c_C^\star - c_N^\star)\right\|, \qquad c(X) = \frac{\left\|R_C R_N^{\mathsf{T}} - I\right\|_{\mathrm{F}}}{\sqrt{2}}. \tag{4.30}$$

对于强度为 k的窗口，偏置 $v(X;k)$在平底容差范围之外呈二次增长：

$$v(X;k) = \frac{k}{2}\left[\left(\frac{[d(X)-d_0]_+}{w_d}\right)^2 + \left(\frac{[c(X)-c_0]_+}{w_c}\right)^2\right], \qquad [u]_+ = \max(u,0). \tag{4.31}$$

我们使用了 $d_0 = 1$Å、 $w_d = 2$Å、 $c_0 = 2\sin(5^\circ/2)$和 $w_c = 2\sin(15^\circ/2)$。这会约束拟合得到的结构域间姿态，而非直接惩罚核心均方根偏差。我们使用了两个偏置窗口 $k = 0.5$ 和 $k = 1$，采用TDS权重，并对每个包含14个粒子的群体进行了200次扩散跃迁。当$N_{\mathrm{eff}}/14 < 0.5$ 时触发系统重采样，并在最终端点也进行了重采样。对于每个非零强度，所有返回的1,008个结构均予以保留，未进行基于均方根偏差的群体筛选。

为便于可视化，我们在共同的406原子选区上，分别计算了与开放态2XE7和闭合态2X13进行刚性对齐后的Cα均方根偏差。基线面板（图6c左）包含4,032个未引导结构。增强

面板（图6c右）通过普通MBAR [60], ，将2,016个基线结构与来自每个牵引系综的1,008个结构合并，估计无偏（即 $k=0$）分布，其中 $k=0.5$ 和 $k=1$。我们在三个偏置条件下评估了每个保存的结构，即 $k=0$，0。5，1，并将这些偏置能量和样本数提供给MBAR；MBAR会计算用于估计无偏基线统计量的独立权重。自由能景观以 $\Delta F=-RT\log(p/p_{\max})$的等高线表示，其中 $T=300$K。

代表性结构 A–C 是从该投影指定区域内保存的源池帧中选出的，并通过了蛋白质几何结构质量检查。这些选择并未对能量景观进行筛选。结构示意图采用 N端核心比对，并使用相同的视角和比例；而图中绘制的均方根偏差则是分别针对每个参考结构进行的全蛋白质独立比对。

4.8 基准测试详情

结构预测基准测试。 我们使用 FoldBench 和 Runs N’ Poses 的不同变体评估了结构预测。FoldBench 涵盖蛋白质和核酸单体，以及六种分子界面类别 [32]；Runs N’ Poses 则评估不同配体–口袋相似度分层下的蛋白质–配体协同折叠 [33]。

为确保各竞争方法之间完全统一，我们决定针对相同的任务定义和多序列比对进行局部重新采样，然后将样本输入通用的 OpenStructure (OST) v2.11.1 流程 [72]。对于每个纳入的模型，我们使用公开代码中的默认预测设置，根据适配后的任务生成了新的预测。出于许可原因，未纳入 AlphaFold3 [5] ；由于 Boltz-2 [39] 的训练截止日期较晚，也未纳入该方法。在可用时，我们采用了基于 ColabFold 的多序列比对变体 [35]；ESMFold2-Fast 则在不使用多序列比对的情况下进行评估 [31]。

FoldBench 采样任务按照作者的描述 [32]，经由 AlphaFold3 流程进行转换。Runs N’ Poses 系统定义保留了上游的蛋白质和配体实体，尽可能使用 CCD 定义，否则使用 SMILES（例如聚糖链或肽）。32 个 FoldBench 系统因常见 OST 错误（14）、常见配体建模问题（8）、有问题的原始 PDB 提交信息（5）以及 DNA/RNA 杂合体（5）而被排除。39 个 Runs N’ Poses 系统因 PLINDER 中的上游修复（21）、令牌数超过 2,560 而对多个基线造成问题（10）、常见 OST 错误（6）以及常见配体建模问题（2）而被排除。为避免训练集–测试集污染，我们进一步移除了 6 个 FoldBench 系统和 584 个 Runs N’ Poses 系统，这些系统的来源 PDB 条目在 2021-09-30 或之前提交，但在此后才发布。经过上述所有排除后的最终集合作为成功率的共同分母。因此，与这些基准的其他一些实现不同，我们确保所有模型都在同一组测试系统上进行比较。

比较中，每个系统使用五个生成的结构。我们先在 NVIDIA A100-80GB GPU 上以相同批次采样五个结构；遇到内存溢出情况时，则降低批次大小和/或改用 B200 GPU 重新运行。

为在不使用置信度排序器的情况下进行公平比较，所有模型均采用五选最优（预言机）选择进行评估。单体性能以 lDDT 五选最优的平均值概括。非配体界面成功要求 DockQ≥ 0.23；配体成功则要求单次预测中的均方根偏差 <2 Å 和 lDDT-PLI >0.8 。Runs N’ Poses 根据来源分配，对带注释的适当配体分别评分。分箱和可视化采用上游的 SuCOS 口袋相似性区间，并省略未预先分配数值相似度的配体。不确定性通过 1,000 次靶标级重采样计算 95% 百分位数自助法区间进行估计。

聚合物几何评估。对于图 6b，我们对 1,490 个符合条件的系统，在每种采样条件下比较五个结构。准确性采用上述五选最优标准；几何评估则涵盖全部五个结构，包括被已保存的过滤流程剔除的结构。两种条件均使用相同的 OpenStructure 2.11.1 Scorer/Molck 准备流程和 StereoCheck 进行评估。若报告至少一项聚合物冲突、键长异常值或键角异常值，则将该结构计为受影响。检查范围涵盖整个系统的聚合物，不包括配体内部几何和聚合物-配体冲突。

几何异常发生率是每个类别中各唯一系统的五个结构里受影响结构所占的比例，并对这些系统取平均。不同类别之间可能存在重叠。在此比较中，准确率以符合条件的界面或单体靶标实例为准，缺失或无法评分的靶标计为零；这些实例是进行1,000次重复准确率自助法的重采样单位。几何指标采用5,000次配对系统级自助重采样，保留全部五个结构，并在不同条件间匹配系统身份，以获得95%百分位区间。各采样活动所用的源代码修订版本、编译和数值设置不同，因此该比较仅作描述，并非对引导作用进行独立的因果消融实验。

参考文献

[1] M. F. Perutz. 血红蛋白协同效应的立体化学：血红素–血红素相互作用与变构问题。《自然》，228:726–734，1970年。doi: 10.1038/228726a0.[2] K. A. Henzler-Wildman 等。沿酶促反应轨迹的内在运动。《自然》，450:838–844，2007年。doi: 10.1038/nature06410.[3] M. L. Metzker. 测序技术——下一代技术。《自然·遗传学评论》，11:31–46，2010年。doi:10.1038/nrg2626.[4] J. Jumper 等。利用 AlphaFold 实现高精度蛋白质结构预测。《自然》，596:583–589，2021年。doi:10.1038/s41586-021-03819-2.[5] J. Abramson 等。利用 AlphaFold 3 精确预测生物分子相互作用结构。《自然》，630: 493–500，2024年。doi: 10.1038/s41586-024-07487-w.[6] M. Baek 等。利用三轨神经网络精确预测蛋白质结构和相互作用。《科学》，373:871–876，2021年。doi: 10.1126/science.abj8754.[7] R. Krishna 等。利用 RoseTTAFold All-Atom 进行通用生物分子建模与设计。《科学》，384: eadl2528，2024年。doi: 10.1126/science.adl2528.[8] X.-C. Bai、G. McMullan 和 S. H. W. Scheres. 冷冻电镜如何推动结构生物学变革。《生化科学趋势》，40:49–57，2015年。doi: 10.1016/j.tibs.2014.10.005.[9] I. R. Kleckner 和 M. P. Foster. 基于核磁共振的方法简介：测量蛋白质动力学。《生物化学与生物物理学学报》，1814:942–968，2011。doi: 10.1016/j.bbapap.2010.10.012.[10] E. Lerner 等。迈向动态结构生物学：单分子Förster共振能量转移的二十年发展。《科学》，359:eaan1133，2018年。doi: 10.1126/science.aan1133.[11] Z. Ganim 等。蛋白质的酰胺I二维红外光谱。《化学研究述评》，41: 432–441，2008。doi: 10.1021/ar700188n.[12] D. I. Svergun 和 M. H. J. Koch. 溶液中生物大分子的小角散射研究。《物理学进展报告》，66:1735–1782，2003。doi: 10.1088/0034-4885/66/10/r05.[13] J. R. Engen. 利用氢/氘交换质谱分析蛋白质构象与动力学。《分析化学》，81:7870–7875，2009年。doi: 10.1021/ac901154s.[14] C. J. Markin、D. A. Mokhtari、F. Sunden、M. J. Appel、E. Akiva、S. A. Longwell、C. Sabatti、D. Herschlag 和 P. M. Fordyce. 通过高通量微流控酶动力学揭示酶功能架构。《科学》，373:eabf8761，2021年。doi: 10.1126/science.abf8761.[15] K. Tsuboyama、J. Dauparas、J. Chen、E. Laine、Y. Mohseni Behbahani、J. J. Weinstein、N. M. Mangan、S. Ovchinnikov 和 G. J. Rocklin. 对生物学与设计中的蛋白质折叠稳定性进行超大规模实验分析。《自然》，620:434–444，2023。doi: 10.1038/s41586-023-06328-6.[16] K. Lindorff-Larsen、S. Piana、R. O. Dror 和 D. E. Shaw. 快速折叠蛋白质如何折叠。《科学》，334:517–520，2011。doi: 10.1126/science.1208351.[17] N. Plattner、S. Doerr、G. De Fabritiis 和 F. Noé. 通过分子动力学模拟和马尔可夫建模揭示原子细节下完整的蛋白质-蛋白质缔合动力学。《自然-化学》，9:1005–1011，2017年。doi: 10.1038/nchem.2785.[18] J. Hénin、T. Lelièvre、M. R. Shirts、O. Valsson 和 L. Delemotte. 分子动力学模拟的增强采样方法。《计算分子科学生活期刊》，4:1583，2022。doi: 10.33011/livecoms.4.1.1583.

[19] J. Wang 等。粗粒化分子动力学力场的机器学习。《ACS Central Science》，5: 755–767，2019。doi: 10.1021/acscentsci.8b00913.[20] N. E. Charron 等。利用机器学习的可迁移粗粒化模型探索蛋白质景观。《Nature Chemistry》，17:1284–1292，2025。doi: 10.1038/s41557-025-01874-0.[21] F. Noé、S. Olsson、J. Köhler 和 H. Wu。玻尔兹曼生成器：利用深度学习对多体系统的平衡态进行采样。《Science》，365:eaaw1147，2019。doi: 10.1126/science.aaw1147.[22] S. Zheng 等。利用深度学习预测分子系统的平衡分布。《Nature Machine Intelligence》，6:558–567，2024。doi: 10.1038/s42256-024-00837-3.[23] B. Jing、B. Berger 和 T. Jaakkola。AlphaFold 遇上流匹配：生成蛋白质系综。载于第 41 届国际机器学习会议论文集，第 235 卷《机器学习研究论文集》，第 22277–22303 页。PMLR，2024。URL https://proceedings.mlr.press/v235/jing24a.html.[24] S. Lewis 等。利用生成式深度学习实现蛋白质平衡系综的可扩展仿真。《Science》，389: eadv9817，2025。doi: 10.1126/science.adv9817.[25] L. Klein 等。Timewarp：通过学习时间粗粒化动力学加速分子动力学，并实现可迁移性。arXiv:2302.01170，2023。URL https://arxiv.org/abs/2302.01170.[26] B. Jing、H. Stärk、T. Jaakkola 和 B. Berger。分子动力学轨迹的生成式建模。载于《神经信息处理系统进展》，第 37 卷，第 40534–40564 页，2024。doi: 10.52202/079017-1282.[27] M. Schreiner、O. Winther 和 S. Olsson。隐式转移算子学习：分子动力学的多时间分辨率模型。载于《神经信息处理系统进展》，第 36 卷，第 36449–36462 页，2023。URL https://arxiv.org/abs/2305.18046.[28] J. Wohlwend 等。Boltz-1：推动生物分子相互作用建模民主化。bioRxiv，2024。doi: 10.1101/2024.11.19.624167.[29] Y. Zhang、C. Gong、J. Sun、J. Guan、M. Ren、S. Xue、H. Zhang、W. Ma、Z. Liu、X. Chen 等。Protenix-v2：拓展结构预测和生物分子设计的应用范围。bioRxiv，2026。doi: 10.64898/2026.04.10.717613.[30] Aureka AI OpenDDE 项目。利用开源药物发现引擎进行折叠、推理和扩展。arXiv:2607.03787，2026。URL https://arxiv.org/abs/2607.03787.[31] S. Candido、T. Hayes、A. Derry、R. Rao、Z. Lin、R. Verkuil、B. Z. Wu、J. S. Lee、E. S. Bruguera、J. A. Keval、M. Kopylov、J. E. Pak、W. Wu、N. Thomas、S. Mataraso、A. Hsu、A. C. Trotman-Grant、K. Fatras、A. dos Santos Costa、R. Badkundri、H. Akın、D. Oktay、J. Deaton、E. Montabana、H. Sitwala、Y. Yu、M. Wiggert、D. A. Carlin、A. W. Goering、T. Blazejewski、M. Sandora、M. Hla、T. Z. Jia、L. H. Kloker、N. J. Sofroniew、M. Uehara、J. Pannu、S. Bachas、D. S. Liu、T. Sercu 和 A. Rives。语言建模将蛋白质生物学具体化为一个世界模型。bioRxiv，2026。doi: 10.64898/2026.06.03.729735.[32] S. Xu、Q. Feng、L. Qiao、H. Wu、T. Shen、Y. Cheng、S. Zheng 和 S. Sun。使用 FoldBench 对全原子生物分子结构预测进行基准测试。《Nature Communications》，17:442，2026。doi: 10.1038/s41467-025-67127-3.[33] P. Škrinjar、J. Eberhardt、G. Studer、G. Tauriello、T. Schwede 和 J. Durairaj。评估蛋白质–配体共折叠方法的泛化能力。《Nature Structural & Molecular Biology》，33:782–794，2026。doi: 10.1038/s41594-026-01797-5.[34] M. Steinegger 和 J. Söding。MMseqs2 支持灵敏的蛋白质序列搜索，用于分析海量数据集。《Nature Biotechnology》，35:1026–1028，2017。doi: 10.1038/nbt.3988。

[35] M. Mirdita、K. Schütze、Y. Moriwaki、L. Heo、S. Ovchinnikov 和 M. Steinegger。ColabFold：让所有人都能进行蛋白质折叠。《Nature Methods》，19:679–682，2022。doi：10.1038/s41592-022-01488-1.[36] J. Sohl-Dickstein、E. Weiss、N. Maheswaranathan 和 S. Ganguli。利用非平衡热力学进行深度无监督学习。载于第 32 届国际机器学习会议论文集，第 37 卷《机器学习研究论文集》，第 2256–2265 页。PMLR，2015。URLhttps://proceedings.mlr.press/v37/sohl-dickstein15.html.[37] J. Ho、A. Jain 和 P. Abbeel。去噪扩散概率模型。载于《神经信息处理系统进展》，第 33 卷，第 6840–6851 页，2020。URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.[38] Y. Song、J. Sohl-Dickstein、D. P. Kingma、A. Kumar、S. Ermon 和 B. Poole。通过随机微分方程实现基于分数的生成建模。载于《国际学习表征会议》，2021。URLhttps://arxiv.org/abs/2011.13456.[39] S. Passaro 等。Boltz-2：迈向精准高效的结合亲和力预测。bioRxiv，2025。doi:10.1101/2025.06.14.659707.[40] OpenFold3 团队。Openfold3-preview，2025。URL https://github.com/aqlaboratory/openfold-3.[41] V. Mariani、M. Biasini、A. Barbato 和 T. Schwede。lDDT：一种用于通过距离差异测试比较蛋白质结构和模型的局部无叠合得分。《Bioinformatics》，29:2722–2728，2013。doi:10.1093/bioinformatics/btt473.[42] S. Basu 和 B. Wallner。DockQ：蛋白质-蛋白质对接模型的质量度量。《PLOS ONE》，11: e0161879，2016。doi：10.1371/journal.pone.0161879.[43] D. Lin、S. Chu、V. Iyer、Y. Lee、J. St John、K. Boyd、B. Roland、X. Ren、G. Zhou、Z. Cao 等。Fold- cp：用于生物分子建模的上下文并行框架。arXiv:2603.14806，2026。URL https://arxiv.org/abs/2603.14806.[44] J. Wang、R. M. Wolf、J. W. Caldwell、P. A. Kollman 和 D. A. Case。通用 Amber 力场的开发与测试。《Journal of Computational Chemistry》，25:1157–1174，2004。doi：10.1002/jcc.20035.[45] G. Pérez-Hernández、F. Paul、T. Giorgino、G. De Fabritiis 和 F. Noé。识别用于马尔可夫模型构建的慢分子序参量。《The Journal of Chemical Physics》，139:015102，2013。doi:10.1063/1.4811489.[46] C. R. Schwantes 和 V. S. Pande。改进马尔可夫状态模型构建揭示 NTL9 折叠过程中的许多非天然相互作用。《Journal of Chemical Theory and Computation》，9:2000–2009，2013。doi:10.1021/ct300878a.[47] S. Olsson、F. Noé、G. M. Rotskoff 和 K. Lindorff-Larsen。不要只看路灯下：生成式分子动力学中的评估。载于《神经信息处理系统进展》，2026。URLhttps://arxiv.org/abs/2610.12518。Position Track。[48] J. Viguera Diez、M. Schreiner 和 S. Olsson。可迁移生成模型连接飞秒至纳秒时间步长的分子动力学。《Science Advances》，12:eaed2333，2026。doi：10.1126/sciadv.aed2333.[49] M. B. Berry、B. Meador、T. Bilderback、P. Liang、M. Glaser 和 G. N. Phillips。高度柔性蛋白质的闭合构象：结合 AMP 和 AMPPNP 的大肠杆菌腺苷酸激酶结构。《 Proteins: Structure, Function, and Bioinformatics 》，19:183–198，1994。doi：10.1002/prot.340190304.[50] C. L. Day、C. Smits、F. C. Fan、E. F. Lee、W. D. Fairlie 和 M. G. Hinds。p53 诱导型 BH3-only 蛋白 Noxa 和 PUMA 的 BH3 结构域与 MCL-1 形成复合物的结构。《 Journal of Molecular Biology》，380: 958–971，2008。doi：10.1016/j.jmb.2008.05.071.[51] J. M. Rogers、A. Steward 和 J. Clarke。内在无序蛋白质的折叠与结合：速度很快，但并非“扩散受限”。《Journal of the American Chemical Society》，135:1415–1422，2013。doi：10.1021/ja309527h。

[52] Y. Xie 等. 增强扩散采样：使用扩散模型高效采样稀有事件并计算自由能. arXiv:2602.16634, 2026. URL https://arxiv.org/abs/2602.16634.[53] R. Singhal 等. 扩散模型推理时扩展与引导的通用框架. 收录于第42届国际机器学习会议论文集，机器学习研究论文集第267卷，第55810–55827页. PMLR, 2025. URL https://proceedings.mlr.press/v267/singhal25b.html.[54]L. Wu, B. Trippe, C. Naesseth, D. Blei, 和 J. Cunningham. 扩散模型中实用且渐近精确的条件采样. 收录于神经信息处理系统进展，第36卷，第31372–31403页，2023. URL https://papers.nips.cc/paper_files/paper/2023/hash/63e8bc7bbf1cfea36d1d1b6538aecce5-Abstract-Conference.html.[55] R. W. Zwanzig. 通过微扰方法建立高温状态方程。I. 非极性气体. 《化学物理学杂志》，22:1420–1426, 1954. doi: 10.1063/1.1740409.[56] G. M. Torrie 和 J. P. Valleau. 蒙特卡罗自由能估计中的非物理采样分布：伞式采样. 《计算物理学杂志》，23:187–199, 1977. doi: 10.1016/0021-9991(77)90121-8.[57] M. Skreta 等. 扩散中的费曼-卡茨校正项：退火、引导与专家乘积. 收录于第42届国际机器学习会议论文集，机器学习研究论文集第267卷，第55906–55949页. PMLR, 2025. URL https://proceedings.mlr.press/v267/skreta25a.html.[58] C. Bartels. 分析有偏的蒙特卡罗和分子动力学模拟. 《化学物理快报》，331: 446–454, 2000. doi: 10.1016/S0009-2614(00)01215-X.[59] A. Kong, P. McCullagh, X.-L. Meng, D. Nicolae, 和 Z. Tan. 蒙特卡罗积分统计模型理论. 《英国皇家统计学会杂志：B辑（统计方法学）》，65:585–604, 2003. doi:10.1111/1467-9868.00404.[60] M. R. Shirts 和 J. D. Chodera. 对来自多个平衡态的样本进行统计最优分析. 《化学物理学杂志》，129:124105, 2008. doi: 10.1063/1.2978177.[61] J. D. Westbrook 等. 化学组分词典：对蛋白质数据库中实验测定的三维大分子所含组成分子的完整描述. 《生物信息学》，31:1274–1278, 2015. doi: 10.1093/bioinformatics/btu789.[62] J. Ouyang-Zhang 等. 三角乘法是生物分子结构表示所需的一切. arXiv:2510.18870, 2025. URL https://arxiv.org/abs/2510.18870.[63] J. Wohlwend, M. Reveiz, M. McPartlon, A. Feldmann, W. Jin, 和 R. Barzilay. Minifold：简单、快速且准确的蛋白质结构预测. 《机器学习研究汇刊》，2025. URL https://openreview.net/forum?id=1p9hQTbjgo.[64] J. Z. Qiao, F. Ding, T. Dresselhaus, M. Rosenfeld, X. Han, O. Howell, A. Iyengar, S. Opalenski, A. Christensen, S. K. Sirumalla 等. NeuralPLexer3：使用流模型进行准确的生物分子复合物结构预测. 收录于神经信息处理系统进展，第38卷，第1–38页，2025. doi: 10.52202/085713-0001.[65] T. Karras, M. Aittala, T. Aila, 和 S. Laine. 阐明基于扩散的生成模型的设计空间. 收录于神经信息处理系统进展，第35卷，第26565–26577页，2022. URL https://arxiv.org/abs/2206.00364.[66] C. Schütte, A. Fischer, W. Huisinga, 和 P. Deuflhard. 一种基于混合蒙特卡罗的构象动力学直接研究方法. 《计算物理学杂志》，151:146–168, 1999. doi: 10.1006/jcph.1999.6231.[67] W. C. Swope, J. W. Pitera, 和 F. Suits. 通过分子动力学模拟描述蛋白质折叠动力学。1. 理论. 《物理化学杂志B辑》，108:6571–6581, 2004. doi: 10.1021/jp037421y.[68] J.-H. Prinz 等. 分子动力学的马尔可夫模型：生成与验证. 《化学物理学杂志》，134:174105, 2011. doi: 10.1063/1.3565032.

[69] H. Wu, A. Mardt, L. Pasquali, 和 F. Noé. 深度生成式马尔可夫状态模型. 收录于神经信息处理系统进展，第31卷，2018年. URL https://arxiv.org/abs/1805.07601.[70] P. Antoniadis, B. Pavesi, S. Olsson, 和 O. Winther. 蛋白质语言模型嵌入提升隐式转移算子的泛化能力. arXiv:2602.11216, 2026. URL https://arxiv.org/abs/2602.11216.[71] J. Viguera Diez, M. Schreiner, O. Engkvist, 和 S. Olsson. 隐式转移算子的玻尔兹曼先验. 收录于国际学习表征会议，2025. URL https://arxiv.org/abs/2410.10605.[72] G. Studer, X. Robin, S. Bienert, J. Durairaj, P. Škrinjar, G. Tauriello, A. M. Waterhouse, 和 T. Schwede. 用于比较大分子复合物的全自动基准测试套件. 《自然·方法》，23:387–394, 2026. doi: 10.1038/s41592-025-02973-z.