

# Quantifying Protein-Protein Interaction with a Spatial Attention Kinetic Graph Neural Network

Yuzhi Xu<sup>1,2#</sup>, Wei Xia<sup>1,2#</sup>, Chao Zhang<sup>3</sup>, Xinxin Liu<sup>4</sup>, Chengwei Ju<sup>5</sup>, Xuhang Dai<sup>2</sup>, Pujun Xie<sup>2</sup>,  
Yuanqing Wang<sup>2\*</sup>, Guangyong Chen<sup>4\*</sup>, and John Z.H. Zhang<sup>1,3,6 \*</sup>

<sup>1</sup>Shanghai Frontiers Science Center of Artificial Intelligence and Deep Learning and NYU-ECNU  
Center for Computational Chemistry, NYU Shanghai, Shanghai 200062, China

<sup>2</sup>Simons Center for Computational Physical Chemistry, Department of Chemistry, New York  
University, New York, New York 10003, United States

<sup>3</sup>Faculty of Synthetic Biology and Institute of Synthetic Biology, Shenzhen Institute of  
Advanced Technology, Shenzhen 518055, China

<sup>4</sup>Hangzhou Institute of Medicine, Chinese Academy of Sciences, 150 Dongfang Street, Xiasha,  
Qiantang District, Hangzhou, Zhejiang 310018, China

<sup>5</sup>Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60615,  
United States

<sup>6</sup>Shanghai Engineering Research Center of Molecular Therapeutics and New Drug Development,  
School of Chemistry and Molecular Engineering, East China Normal University, 200062, China

#These authors contributed equally to this work

Corresponding author and address: John Zhang john.zhang@nyu.edu, Guangyong Chen:  
chen.guangyong2@gmail.com, Yuanqing Wang: wangyq@umich.edu

## Abstract

Accurate prioritisation of near-native protein–protein interaction (PPI) models remains a major bottleneck in structural biology. Here we present SAKE-PP, a physics-inspired, spatial-attention equivariant graph neural network that directly regresses interface RMSD (iRMSD) without any native references. Trained on docking decoys generated through our novel hierarchical sampling strategy applied to PDBBind dataset, SAKE-PP combines force-field-like attention with Laplacian-eigenvector orientation to couple local interaction forces with global topology. On the 2024PDB benchmark comprising 176 heterodimers, SAKE-PP demonstrates effective optimization and selection of AF3 decoys, achieving improvements of 13.75% based on iRMSD statistics and 12.5% based on DockQ scores. It consistently outperforms the AF3 ranking score in multiple metrics, including overlap, hit rate, and correlation. In zero-shot evaluation of 139 antibody–antigen complexes, SAKE-PP improves the score–iRMSD correlation by 0.4. By unifying geometric deep learning with physics-based realism, SAKE-PP provides a robust, plug-and-play scoring function that streamlines reliable PPI evaluation and accelerates downstream structure-guided drug-design workflows.

近天然蛋白质-蛋白质相互作用 (PPI) 模型的准确优先级排序仍然是结构生物学领域的主要瓶颈。本文我们提出 SAKE-PP，这是一个受物理启发的空间注意力等变图神经网络，它无需任何天然参考即可直接回归界面均方根差 (iRMSD)。SAKE-PP 使用我们应用于 PDBBind 数据集的新型分层采样策略生成的对接诱饵进行训练，将力场式注意力机制与拉普拉斯特征向量方向相结合，将局部相互作用力与全局拓扑结构耦合。在包含 176 个异二聚体的 2024PDB 基准数据集上，SAKE-PP 展现出对 AF3 诱饵的有效优化和选择，基于 iRMSD 统计数据实现了 13.75% 的提升，基于 DockQ 得分实现了 12.5% 的提升。它在多个指标（包括重叠度、命中率和相关性）上始终优于 AF3 排名得分。在对 139 个抗体-抗原复合物进行零样本评估中，SAKE-PP 将评分-iRMSD 相关性提高了 0.4。通过将几何深度学习与基于物理的现实主义相结合，SAKE-PP 提供了强大、即插即用的评分函数，从而简化了可靠的 PPI 评估，并加速了下游结构引导的药物设计工作流程。

## Introduction

Protein-protein interactions (PPIs) are fundamental to virtually all cellular processes, orchestrating complex biological functions from signal transduction and metabolic pathways to gene regulation.<sup>1–5</sup> Understanding the intricate structural details and energetic landscapes governing these interactions is crucial for deciphering the molecular mechanisms of life and for developing targeted therapeutic interventions.<sup>6</sup> While experimental techniques such as X-ray crystallography and cryo-electron microscopy provide invaluable atomic-level insights into the structures of protein complexes, their applicability can be limited by factors such as protein stability, sample heterogeneity, and the challenges associated with crystallizing or resolving large, flexible, or transient assemblies.<sup>7–9</sup> Consequently, computational methods, including protein-protein docking and molecular dynamics (MD) simulations, have emerged as indispensable tools to theoretically predict PPIs and explore their dynamic behavior.<sup>10,11</sup> A critical metric for evaluating the quality of computationally predicted protein complex structures is the interface Root Mean Square Deviation (iRMSD), which quantifies the structural similarity between a predicted pose and the experimentally determined (native) conformation of the binding interface.<sup>12–14</sup> Beyond its role in assessing prediction accuracy, iRMSD is also paramount in validating and monitoring the fidelity of MD simulations, enabling researchers to track interface stability and confirm simulated conformational dynamics against known experimen-

tal structures.<sup>15,16</sup> While conventional methods strive to minimize iRMSD for high-fidelity predictions and ensure robust simulations, consistently achieving low iRMSD values for diverse and challenging protein systems remains a formidable hurdle, underscoring the need for more efficient and accurate prediction strategies.

蛋白质-蛋白质相互作用（PPIs）是几乎所有细胞过程的基础，从信号转导和代谢通路到基因调控，协同调控复杂的生物功能<sup>1-5</sup>。理解这些相互作用所涉及的复杂结构细节和能量景观，对于阐明生命的分子机制以及开发有针对性的治疗干预手段至关重要

A paradigm shift in protein structure prediction has recently been catalyzed by deep learning methods, most notably AlphaFold and its subsequent iterations like AlphaFold-Multimer.<sup>17,18</sup> The latest version, AlphaFold3(AF3), has significantly expanded these capabilities, enabling accurate modeling of not only individual protein structures but also diverse biomolecular complexes, including protein-protein, protein-ligand, and protein-nucleic acid assemblies.<sup>19</sup> To provide users with crucial information on prediction reliability, AlphaFold models output intrinsic confidence metrics such as predicted Local Distance Difference Test (pLDDT) for local accuracy, Predicted Aligned Error (PAE) for relative positioning, and interface Predicted Template Modeling (ipTM) score, specifically for assessing the quality of predicted interfaces in complexes. Training of these confidence metrics requires comparing predicted structures to known experimental data, an obvious question arises concerning the output of powerful prediction methods like AlphaFold: why haven't they prioritized the direct prediction of these rigorous evaluation metrics, most notably iRMSD? Accurate estimation of iRMSD would be invaluable, as it would allow for the selection of predicted complex structures that are likely to be closer to the native state and thus serve as more suitable starting points for detailed downstream MD simulations. Deep-learning models including DeepBSP, RmsdXNA and DeepAccNet have demonstrated reference-free error regression for protein-ligand complexes or single-chain proteins, but the high-dimensional, interface-sensitive nature of PPI assemblies remains to be tackled.<sup>20-22</sup> The technical challenge of directly predicting such a metric likely stems from its inherent nature. Unlike the bounded nature of metrics such as pLDDT, PAE, and ipTM, iRMSD represents a precise, unbounded physical distance.<sup>23</sup> Accurately predicting such a continuous and potentially large value is significantly more challenging for current deep learning methods than predicting normalized confidence scores, posing a primary technical hurdle for directly modeling iRMSD.

最近，蛋白质结构预测领域正经历一场由深度学习方法引发的范式转变，最著名的例子是 AlphaFold 及其后续版本 AlphaFold-Multimer<sup>17,18</sup>。最新版本 AlphaFold3 (AF3) 显著扩展了其能力，能够精准建模不仅限于单个蛋白质结构，还包括多种生物分子复合物，如蛋白-蛋白、蛋白-配体和蛋白-核酸等组装体<sup>19</sup>。为了向用户提供关于预测可靠性的关键信息，AlphaFold 模型输出内在置信度指标，如用于局部准确性的预测局部距离差测试 (pLDDT)、用于相对定位的预测对齐误差 (PAE)，以及专用于评估复合物界面预测质量的界面预测模板建模得分 (ipTM)。这些置信度指标的训练需要将预测结构与已知实验数据进行比较，于是自然会提出一个问题：像 AlphaFold 这样强大的预测方法为何未优先考虑直接预测这些严格的评估指标，尤其是 iRMSD？若能准确估算 iRMSD，将可选出更接近天然态的预测复合体结构，从而成为更合适的下游 MD 模拟起点。已有深度学习模型如 DeepBSP、RmsdXNA 和 DeepAccNet 展示了对蛋白-配体复合物或单链蛋白的无参考误差回归能力，但高维度、界面敏感的 PPI 结构仍是一项未被攻克的挑战<sup>20-22</sup>。直接预测 iRMSD 这一指标的技术难点，可能源于其本质属性——iRMSD 是一个精确、无界的物理距离，而非如 pLDDT、PAE 和 ipTM 等有界分数。与预测归一化置信度得分相

比，当前深度学习方法更难准确预测这种连续且可能较大的值，这构成了建模 iRMSD 的核心技术障碍。

In this paper, we propose a reference-free iRMSD predictor tailored to PPIs to bridge this gap by enabling Ångström-level scoring scoring of large decoy sets and seamless integration with physics-based refinement. This predictor, named SAKE-PP, is a novel physics-inspired spatial attention kinetics graph neural network(GNN) designed specifically for protein–protein evaluation. Unlike methods that rely on or are integrated within existing structure prediction architectures like AlphaFold’s PairFormer modules, SAKE-PP is designed as an independent model framework. This independent design is crucial to avoid any potential bias towards the specific outputs or feature representations of particular structure prediction models like AF3, ensuring the predicted iRMSD is a generalizable metric of geometric accuracy. SAKE-PP integrates two complementary modules to fully capture both structural and functional relationships at protein complex interfaces: the Spatial Attention Kinetics (SAKE) component simultaneously updates node features and their 3D coordinates, using a physics-inspired attention mechanism to account for spatial relationships among residues and simulate inter-residue forces under a force field; and the Laplacian Eigenvector Orientation component, which calculates the graph Laplacian eigenvectors to generate global embeddings reflecting each residue’s position within the overall fold, the steering message passing by the gradients and magnitudes of these eigenvectors to distinguish structurally similar but functionally distinct regions: finally, multiscale feature fusion unifies local interaction patterns with the global structural context to yield a holistic representation of the protein–protein interface. Training a model to accurately predict the precise, potentially unbounded values of iRMSD is a significant technical challenge. To construct a reliable training dataset for protein–protein docking evaluation, we curated 736 complexes from PDDBind and generated 1,472,000 candidate conformations using ZDock under a 15Å distance restraint. A hierarchical sampling strategy was employed to mitigate the imbalanced iRMSD distribution of docking poses, with special emphasis on the critical 0–10 Å range. In particular, we selected representative and high-quality conformations through a combination of adaptive and uniform sampling. This procedure yielded a final dataset of 15,456 conformations, ensuring both diversity and enrichment of near-native structures to support robust regression-based model training.

在本文中，我们提出了一个针对 PPI 的无参考 iRMSD 预测器，以填补现有预测体系中的空白，实现对大规模诱饵集合的 Å 级评分以及与基于物理的精修过程的无缝整合。该预测器命名为 SAKE-PP，是一个全新的、受物理启发的空间注意力动力学图神经网络（GNN），专为蛋白质-蛋白质评估设计。与依赖或集成于 AlphaFold 的 PairFormer 模块等现有结构预测架构的方法不同，SAKE-PP 被构建为一个独立的模型框架。这种独立设计对于避免对特定结构预测模型（如 AF3）输出或特征表示的潜在偏倚至关重要，从而确保预测的 iRMSD 成为具有广泛适应性的几何精度指标。SAKE-PP 融合两个互补模块以全面捕捉蛋白质复合物界面的结构与功能关系：其一是空间注意力动力学（SAKE）模块，利用受物理启发的注意力机制，同时更新节点特征及其三维坐标，模拟残基间的关系与力场下的相互作用；其二是拉普拉斯特征向量方向模块，通过计算图拉普拉斯特征向量来生成反映各残基在整体折叠中位置的全局嵌入，并基于其梯度和幅值引导信息传递，以区分结构相似但功能有异的区域；最后，多尺度特征融合将局部相互作用模式与全局结构上下文统一起来，形成对蛋白质界面的整体表达。训练一个能够精确预测可能无界的 iRMSD 数值的模型是一项重大技术挑战。为了构建一个可靠的蛋白-蛋白对接训练数据集，我们从 PDDBind 中筛选了 736 个复合物，并利用 ZDock 在 15Å 距离约束下生成了 1,472,000 个候选构象。采用分层采样策略以缓解对接构象中 iRMSD 分布的不平衡，特别加强了对关键的 0–10 Å 范围的覆盖。我们

结合自适应与均匀采样策略，筛选出代表性和高质量构象，最终得到 15,456 个构象，确保了数据集的多样性和近天然构象的富集，以支撑鲁棒的回归模型训练。

Using this novel framework and training strategy, we demonstrate SAKE-PP’s superior performance and utility. Comprehensive benchmarking against state-of-the-art GNN-variants shows that our model achieves substantially lower RMSE and MSE, indicating superior accuracy in protein–protein interface prediction. Further, our model demonstrates effective optimization and selection of AF3 decoys on the newly curated 2024PDB dataset, achieving improvements of 13.75% based on iRMSD statistics and 12.5% based on DockQ scores. It consistently outperforms the AF3 ranking score in multiple metrics, including overlap, hit rate, and correlation. Moreover, as candidates progress along the drug-discovery pipeline, evaluation priorities shift towards developability attributes, notably energetic and dynamical properties such as binding affinity ( $\Delta G_{\text{bind}}$ ) and kinetic stability (residence time and energetic barriers).<sup>24,25</sup> In this regime, our experiments favor the previous observation that RMSD-based geometric errors correlate well with total system energies obtained from MM/GBSA, free-energy perturbation (FEP) and related alchemical workflows (Fig. 1a).<sup>26,27</sup> A finely resolved iRMSD thus supplies a continuous, differentiable signal for selecting high-affinity, kinetically stable conformers and meshes naturally with downstream free-energy and molecular-dynamics analyses. Our model consistently avoids high-energy conformation selection errors frequently made by the AF3 ranking score and identifies a class of conformationally stable heterocomplexes. Furthermore, we performed antigen-antibody tests, which showed that despite no training on antigen–antibody complexes, SAKE-PP achieved a correlation score of 0.4 higher than that of the AF3 ranking score, indicating a strong generalizability. Ultimately, our development of an accurate, reference-free iRMSD predictor provides a crucial tool for evaluating and refining the structure of predicted protein complexes, thereby accelerating the mechanistic understanding and facilitating structure-guided design and discovery efforts in biology and medicine.

通过这一新颖的框架与训练策略，我们展示了 SAKE-PP 的卓越性能与实用价值。与多种先进 GNN 变体的综合基准测试表明，我们的模型在蛋白–蛋白界面预测中实现了显著更低的 RMSE 和 MSE，表现出更高的预测精度。此外，在新整理的 2024PDB 数据集上，我们的模型在对 AF3 生成的诱饵构象进行优化与筛选方面同样表现出色，基于 iRMSD 统计改善了 13.75%，基于 DockQ 得分改善了 12.5%。在包括重叠度、命中率和相关性等多项评估指标中，SAKE-PP 一贯优于 AF3 的排名评分体系。此外，随着候选分子在药物研发流程中推进，评估重点逐步转向可开发性特征，特别是结合自由能 ( $\Delta G_{\text{bind}}$ ) 及动力学稳定性（驻留时间和能垒）等能量与动力学性质<sup>24,25</sup>。在这一阶段，我们的实验证明了早前的发现，即 RMSD 几何误差与通过 MM/GBSA、自由能微扰 (FEP) 及其他炼金方法获得的系统总能之间存在良好的相关性 (见图 1a)<sup>26,27</sup>。因此，高分辨率的 iRMSD 提供了一个连续可微的信号，用于筛选高亲和力、动力学稳定的构象，并与下游自由能及分子动力学分析自然融合。我们的模型在避免 AF3 排名评分常见的高能构象选择错误方面表现稳定，能够识别一类构象稳定的异源复合物。此外，我们还进行了抗原-抗体测试，结果显示尽管未在抗原–抗体复合物上进行训练，SAKE-PP 的相关性评分仍比 AF3 排名得分高出 0.4，展现出强大的泛化能力。总之，我们开发的这一高精度、无参考 iRMSD 预测器为评估和优化预测蛋白质复合物结构提供了关键工具，从而加速对其机制的理解，并促进结构导向的生物医学设计与发现工作。

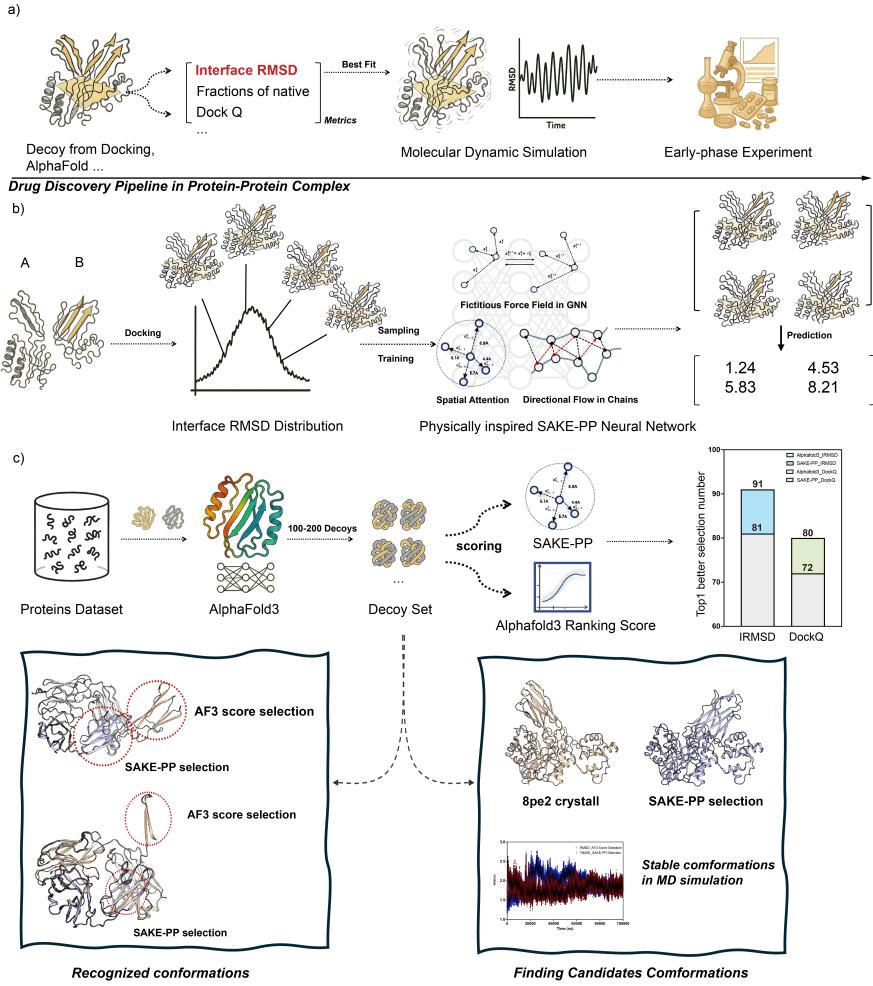


Figure 1: (a) In a typical protein–protein complex drug discovery pipeline, candidate structures are first generated via docking or AI-based prediction methods such as AlphaFold3. Each decoy is then evaluated using metrics such as interface RMSD, fraction of native contacts, or the composite DockQ score. Top-scoring conformations are subsequently subjected to all-atom molecular dynamics (MD) simulations and early-phase experimental validation.

(a) 在典型的蛋白质-蛋白质复合物药物发现流程中，候选结构首先通过对接或诸如 AlphaFold3 等 AI 预测方法生成。每一个欺骗构象随后会通过界面 RMSD、天然接触比例或复合评分 DockQ 等指标进行评估。得分最高的构象将进一步进入全原子分子动力学 (MD) 模拟和早期实验验证阶段。

(b) SAKE-PP 训练流程：给定一组由对接生成的蛋白质 A 与 B 的复合物，首先获得其界面 RMSD 分布，然后通过分层采样策略从中选取具有代表性的欺骗构象用于训练 SAKE-PP。SAKE-PP 是一个受物理启发、几何等变的图神经网络，能够学习预测未见构象的 iRMSD 值，从而指导近天然结合模式的识别。

(b) SAKE-PP 训练流程：给定一组由对接生成的蛋白质 A 与 B 的复合物，首先获得其界面 RMSD 分布，然后通过分层采样策略从中选取具有代表性的欺骗构象用于训练 SAKE-PP。SAKE-PP 是一个受物理启发、几何等变的图神经网络，能够学习预测未见构象的 iRMSD 值，从而指导近天然结合模式的识别。

(c) Cross-dataset benchmarking and representative examples: For a diverse set of protein targets, AlphaFold3 is used to generate 100 decoys per target, which are then independently scored by SAKE-PP and AF3’s native ranking function. The bar chart shows the number of targets for which SAKE-PP (colored bars) or AF3 (gray bars) selects the top-1 decoy under iRMSD or DockQ. Bottom left: Comparison of SAKE-PP-selected low-energy conformations versus AF3-selected high-energy, unstable alternatives for the same target. Bottom right: Molecular dynamics simulations of PDB 8PE2 over 1 s show that structurally distinct conformations selected by different scoring functions remain stable, suggesting the existence of multiple coexisting low-energy states in the protein–protein interaction landscape.

(c) 跨数据集的基准评估与代表性示例：对于一组多样化的蛋白质靶标，AlphaFold3 为每个靶标生成 100 个欺骗构象，这些构象随后分别由 SAKE-PP 和 AF3 原生评分函数进行独立评分。柱状图显示了在 iRMSD 或 DockQ 指标下，SAKE-PP (彩色柱) 或 AF3 (灰色柱) 能够选择出 Top-1 欺骗构象的靶标数量。左下角：展示了 SAKE-PP 选出的低能构象与 AF3 选出的高能、不稳定替代构象在同一靶标下的对比。右下角：对 PDB 8PE2 进行 1 微秒的分子动力学模拟显示，由不同评分函数选出的结构不同的构象均保持稳定，表明在蛋白质-蛋白质相互作用能量景观中可能存在多个共存的低能态。

## Result

### Comparative Analysis of SAKE and AF3 Ranking Score in Protein-Protein Complex Structure Selection

Although our model was initially trained using rigid-body decoys generated by ZDOCK, we hypothesize that SAKE-PP effectively captured relevant structural features, allowing successful transfer to the more challenging task of flexible decoy discrimination, particularly in distinguishing AlphaFold-generated decoys.

尽管我们的模型最初是使用由 ZDOCK 生成的刚体欺骗构象进行训练的，但我们假设 SAKE-PP 有效地捕捉到了相关的结构特征，从而能够成功迁移至更具挑战性的任务——即柔性欺骗构象的判别，尤其是在区分 AlphaFold 生成的构象方面表现出良好能力。

For the protein docking tasks using AlphaFold3, we conducted a systematic docking analysis. The protein data primarily consisted of structures released in 2024, with a particular focus on those published after June 2024. Based on the FASTA sequence identifiers provided by the RCSB database, we performed a stringent selection of binary complex structures. To be more specific, a selection protocol was implemented to curate binary complex structures: (1) exclusion of polypeptide chains containing fewer than 20 amino acid residues; (2) structural refinement using PDB4Fixer to remove crystallographic artifacts and resolve missing atoms; (3) systematic correction of noncanonical amino acids to their standard counterparts. As a result, we obtained a final dataset of 176 high-quality protein structures.

对于使用 AlphaFold3 进行的蛋白质对接任务，我们进行了系统性对接分析。蛋白质数据主要来源于 2024 年发布的结构，尤其聚焦于 2024 年 6 月之后发布的条目。基于 RCSB 数据库提供的 FASTA 序列标识符，我们进行了严格的二元复合物结构筛选。具体来说，我们采用如下筛选流程来整理二元复合物结构：(1) 排除包含少于 20 个氨基酸残基的多肽链；(2) 使用 PDB4Fixer 进行结构精修，以去除晶体伪影并补全缺失原子；(3) 将所有非常规氨基酸系统性地修正为其标准对应物。最终，我们获得了一个包含 176 个高质量蛋白质结构的数据集。

Based on the curated dataset of 176 high-quality protein structures, we constructed an end-to-end docking database using the AlphaFold3. Specifically, we used 20 random seeds of AlphaFold3 and generated 100 decoy conformations for each protein complex, thus creating a decoy database that captures a diverse conformational space. Furthermore, to evaluate the structural variability among these decoys, we computed and compared the interface root-mean-square deviation (iRMSD) for the decoy structures of each protein complex. To enhance the robustness of the test set, we randomly selected 20% of the samples with receptor homologs as the test set, while the remaining 80% consisted of heterologous receptors. This partitioning strategy not only ensures that the test set covers samples with high receptor homology, but also evaluates the generalizability of the model in handling heterologous receptors.

基于整理后的 176 个高质量蛋白质结构数据集，我们利用 AlphaFold3 构建了一个端到端的对接数据库。具体而言，我们使用 AlphaFold3 的 20 个随机种子，并为每个蛋白质复合物生成 100 个欺骗构象，从而建立了一个能够覆盖多样构象空间的欺骗体数据库。进一步地，为了评估这些欺骗构象之间的结构差异性，我们计算并比较了每个蛋白质复合物的界面均方根偏差 (iRMSD)。为增强测试集的稳健性，我们随机选取了其中 20% 包含受体同源体的样本作为测试集，其余 80% 则由异源受体构成。该划分策略不仅保证测试集中覆盖了具有高度受体同源性的

样本，同时也评估了模型在处理异源受体时的泛化能力。

## Overall Analysis

Figure 2 compares the effectiveness of selecting top-1 decoys based on the interface RMSD (iRMSD) and DockQ metrics between SAKE-PP and AlphaFold3’s ranking score. Among the 176 cases tested, SAKE-PP successfully identified better decoys based on iRMSD in 91 instances, whereas AlphaFold3 performed better in 81 instances, representing a 13.75% improvement for SAKE-PP. Notably, both methods selected identical top-1 decoys in four cases. Regarding DockQ, SAKE-PP achieved superior performance in selecting top-1 decoys in 80 instances compared to AlphaFold3’s 72 instances, indicating a 12.5% improvement. These results underscore the superior capability of SAKE-PP in effectively prioritizing high-quality structural decoys, highlighting its advantage over AlphaFold3’s inherent scoring function when evaluating AlphaFold-generated protein complexes.

图 2 展示了 SAKE-PP 与 AlphaFold3 评分函数在选择 Top-1 欺骗构象时，基于界面 RMSD (iRMSD) 和 DockQ 指标的有效性比较。在所测试的 176 个样本中，SAKE-PP 在 91 个样本中基于 iRMSD 选择了更优的欺骗构象，而 AlphaFold3 在 81 个样本中表现更佳，这意味着 SAKE-PP 实现了 13.75% 的提升。值得注意的是，在 4 个案例中，两种方法选择了相同的 Top-1 构象。在 DockQ 指标方面，SAKE-PP 在 80 个样本中选择的 Top-1 构象优于 AF3 的选择，而 AF3 在 72 个样本中占优，表明 SAKE-PP 在该指标上也实现了 12.5% 的改进。这些结果强调了 SAKE-PP 在优先筛选高质量结构欺骗体方面的卓越能力，凸显其在评估 AlphaFold 生成的蛋白质复合物时相较于 AlphaFold3 原生评分函数的明显优势。

In addition to evaluating top-1, beyond the top-1 assessment, we quantified the overlap between the top- $k$  decoy lists produced by the two ranking schemes (Figure 2c, d). Here, the overlap rate at a given  $k$  is simply the percentage of the 176 targets for which the top- $k$  lists from SAKE-PP or AlphaFold3, when compared with the ground-truth ranking, share at least one common decoy. This metric gauges how consistently each method enriches structurally relevant candidates suitable for downstream molecular-dynamics (MD) refinement.

除了 Top-1 评估之外，我们还对两种排序方法在 Top- $k$  构象列表中的重叠程度进行了量化分析（见图 2c, d）。其中，在给定  $k$  值下的重叠率（overlap rate）定义为：在 176 个目标样本中，SAKE-PP 或 AlphaFold3 所生成的 Top- $k$  列表与真实排名中共享至少一个构象的样本所占的百分比。该指标衡量了每种方法在为后续分子动力学 (MD) 精修流程筛选结构相关候选构象方面的一致性。

Across the entire top- $k$  spectrum, SAKE-PP consistently outperforms AF3. In the early stage ( $k = 1\text{--}3$ ), it delivers a 67% relative gain at Top-1 and maintains a clear edge up to  $k = 5$ , ensuring high precision among the highest-ranked decoys. From  $k = 6$  onward the advantage widens, with absolute gains exceeding 5 percentage points and relative improvements peaking at 18% ( $k = 6$ ) and remaining  $\geq 14\%$  through  $k = 12$ , evidencing SAKE-PP’s ability to recover high-quality decoys overlooked by AF3. At the upper bound ( $k = 15$ ) SAKE-PP achieves an 86.93% overlap versus 77.84% for AF3—a 9.09 pp improvement—and the two curves never intersect, indicating globally superior ordering. Collectively, these results demonstrate SAKE-PP’s strong early enrichment and sustained superiority across broader decoy pools, making it highly suitable for ensemble-based free-energy or MD workflows that require multiple accurate starting structures.

在整个 Top- $k$  范围内，SAKE-PP 始终优于 AF3。在早期阶段 ( $k = 1\text{--}3$ )，SAKE-PP 在 Top-1 上实现了 67% 的

相对提升，并在  $k = 5$  以内持续保持明显优势，确保了高排名构象的高精度。从  $k = 6$  开始，SAKE-PP 的优势进一步扩大，其绝对提升超过 5 个百分点，相对提升在  $k = 6$  时达到峰值（18%），并在  $k = 12$  以内保持在 14% 以上，显示了其识别 AF3 忽略的高质量构象的能力。在  $k = 15$  时，SAKE-PP 的重叠率达到 86.93%，而 AF3 为 77.84%，实现了 9.09 个百分点的提升，且两条曲线从未交叉，表明 SAKE-PP 在全局排序上具有持续优势。这些结果表明，SAKE-PP 在早期富集和广泛候选筛选中的表现非常出色，非常适合用于需要多个准确起始结构的基于集成的自由能或 MD 流程。

Besides, following the CAPRI convention, we designate the ten decoys with the lowest iRMSD for each complex as reference “near-native” poses—an operational choice that circumvents the difficulty of unambiguously identifying true native structures within a decoy set. Accordingly, the success-hit-rate is the probability of retrieving at least one of these references within the top- $k$  ranks. As illustrated in Figure 2e, SAKE-PP outperforms the AF3 ranking score from  $k = 2$  onward. Averaged over  $k = 1\text{--}5$ , the success-hit rate is 29.7% versus 27.2% (+9.2% relative). In the  $k = 6\text{--}20$  range the gap widens to 71.1% versus 61.2% (+16.2%). Across the full interval ( $k = 1\text{--}20$ ), SAKE-PP attains a mean success-hit rate of 60.7%, exceeding AF3 by 8.0 percentage points (+15.3%). The practical screening depth of  $k = 20$  benefits most, showing a 14.8 pp lift (87.5% vs 72.7%, +20%). Statistical analysis confirms the robustness of this improvement (paired  $t = 7.83$ ,  $p \approx 2.3 \times 10^{-7}$ ; Cohen’s  $d = 1.75$ , large effect). These results highlight the superior discriminative power of SAKE-PP in prioritising near-native decoys—especially when workflows allow more than a single decoy to be forwarded for downstream refinement.

此外，按照 CAPRI 的惯例，我们将每个复合物中 iRMSD 最低的十个欺骗构象视为“近天然”参考构象——这是一种操作上的选择，用以规避在欺骗集内明确识别真实天然结构的困难。因此，成功命中率定义为在 Top- $k$  排名中至少命中一个这些参考构象的概率。如图 2e 所示，SAKE-PP 从  $k = 2$  起在成功命中率上优于 AF3。在  $k = 1\text{--}5$  的区间内，SAKE-PP 的平均命中率为 29.7%，而 AF3 为 27.2%，相对提升 9.2%。在  $k = 6\text{--}20$  的区间内，这一差距扩大至 71.1% 对比 61.2%，提升 16.2%。在整个区间 ( $k = 1\text{--}20$ ) 中，SAKE-PP 的平均成功命中率为 60.7%，比 AF3 高出 8.0 个百分点（相对提升 15.3%）。在实际筛选深度  $k = 20$  时，SAKE-PP 的成功命中率为 87.5%，相比 AF3 的 72.7%，提升了 14.8 个百分点（提升 20%）。统计分析证实了这一提升的稳健性（配对  $t = 7.83$ ,  $p \approx 2.3 \times 10^{-7}$ ; Cohen’s  $d = 1.75$ , 为大效应量）。这些结果进一步凸显了 SAKE-PP 在优先筛选近天然构象方面的卓越判别能力，尤其适用于可将多个构象送入后续精修流程的任务场景。

To complement the early-ranking evaluation, we examined how well each scoring function tracks overall decoy quality. For every target we computed the Pearson correlation coefficient ( $r$ ) between its raw score and iRMSD. Head-to-head tallies (Figure 2g) indicate that SAKE-PP yields a higher  $r$  in 105 of 176 cases (60.23%), whereas the AF3 ranking score prevails in 71 cases (39.77%), corresponding to a 51.4% relative improvement in win-loss ratio. To focus on meaningful differences, we further required  $|\Delta r| > 0.10$  (“advanced” threshold). Under this stricter criterion (Figure 2h) SAKE-PP still surpasses AF3 in 87 complexes, while lagging in 57, reaffirming the robustness of the advantage. A two-sided Wilcoxon signed-rank test confirms statistical significance ( $Z = 7.35$ ,  $p \approx 1.9 \times 10^{-13}$ ).

为了补充对早期排序性能的评估，我们进一步分析了每个评分函数对整体构象质量的跟踪能力。对于每个目标，我们计算其原始评分与 iRMSD 之间的皮尔逊相关系数 ( $r$ )。从逐一对比结果（见图 2g）来看，SAKE-PP 在 176 个样本中有 105 个表现出更高的相关性（60.23%），而 AF3 在 71 个样本中占优（39.77%），相对提升为 51.4%。为聚

焦更有意义的差异，我们进一步引入  $|\Delta r| > 0.10$  的“高差异”判定标准。在该严格标准下（见图 2h），SAKE-PP 仍在 87 个复合物上优于 AF3，而在 57 个样本中逊色，进一步确认了其优势的稳健性。双侧 Wilcoxon 符号秩检验结果显示该差异具有显著统计学意义 ( $Z = 7.35, p \approx 1.9 \times 10^{-13}$ )。

### Successful Cases

To further demonstrate the practical utility of SAKE-PP, we turned to three challenging yet representative complexes—8S4K, 8SOZ, and 8VGG, the conformations directly selected by AlphaFold3’s scoring function yielded iRMSD values of 17.886 Å, 16.352 Å, and 4.719 Å, respectively, all significantly exceeding the acceptable threshold of 4.0 Å. Notably, this discrepancy does not stem from structural inaccuracies of AlphaFold3-generated decoys themselves; rather, SAKE-PP successfully identified near-native conformations from these AlphaFold3 decoys, achieving markedly improved iRMSD values of 1.521 Å, 1.885 Å, and 0.925 Å, respectively. This highlights the effectiveness of SAKE-PP in accurately distinguishing conformations closely resembling native structures.

为进一步展示 SAKE-PP 在实际中的应用价值，我们选择了三个具有挑战性且具有代表性的复合物——8S4K、8SOZ 和 8VGG。AlphaFold3 评分函数直接选出的构象分别具有 iRMSD 为 17.886 Å、16.352 Å 和 4.719 Å，均显著高于可接受阈值 4.0 Å。值得注意的是，这种差异并非源自 AlphaFold3 所生成的构象本身结构错误，而是由于 SAKE-PP 能够从这些构象中成功识别出接近天然态的构象，分别将 iRMSD 显著降低至 1.521 Å、1.885 Å 和 0.925 Å。这突出显示了 SAKE-PP 在准确区分与天然结构相近构象方面的强大能力。

Specifically, among the decoys generated by AlphaFold3 (AF3), the interface RMSD (iRMSD) values exhibit a distinct bimodal distribution. Approximately 87% of the predictions cluster around a high average iRMSD of 17.26 Å (designated as the “deviated cases” group), while only about 13% cluster around a low average iRMSD of 1.526 Å (defined as the “reasonable cases” group). The overall mean iRMSD across all predictions is 15.22 Å. 具体而言，在 AlphaFold3 (AF3) 生成的欺骗构象中，其 iRMSD 值呈现出明显的双峰分布。约 87% 的预测集中在平均 iRMSD 为 17.26 Å 的高值区（定义为“偏离案例”组），而仅约 13% 的预测集中在平均 iRMSD 为 1.526 Å 的低值区（定义为“合理案例”组）。所有预测构象的整体平均 iRMSD 为 15.22 Å。

As shown in Figure 2(d), the decoy selected by AF3 belongs to the deviated group, whereas SAKE successfully identifies a decoy from the reasonable group. A notable positional displacement at a key structural region differentiates these two conformations. Further analysis of the AF3 ranking scores reveals that the decoy selected by AF3 (PDB: 8S4K) possesses the highest ranking score (0.8709), while the SAKE-selected decoy receives a lower AF3 ranking score of 0.8385.

如图 2(d) 所示，AF3 选择的构象属于偏离组，而 SAKE 成功识别出一个来自合理组的构象。两者在关键结构区域存在明显的空间位移差异。对 AF3 评分进一步分析显示，AF3 所选构象（PDB: 8S4K）具有最高评分（0.8709），而 SAKE 选出的构象在 AF3 评分中得分较低，仅为 0.8385。

To investigate whether AF3’s selection of the deviated conformation was incidental or indicative of a general preference, we randomly sampled ten decoys from each group for a detailed analysis (SIXXXX). For the deviated group, we found that the high iRMSD values primarily stem from substantial displacements of predicted structures relative to the binding site. AF3 consistently assigned these displaced decoys high ranking scores, averaging 0.8572 (standard deviation: 0.005). Conversely, the randomly sampled decoys from the reasonable group exhibited

significantly lower AF3 ranking scores, averaging 0.8399 (standard deviation: 0.004). This suggests that the AF3 ranking score function intrinsically favors the deviated cases.

为了验证 AF3 选择偏离构象是否属于偶然行为或反映了其评分函数的系统偏好，我们从每组中随机抽取十个构象进行详细分析（见 SIXXXX）。在偏离组中，我们发现高 iRMSD 主要来自预测结构相对于结合位点的明显偏移。AF3 对这些偏移构象持续给予较高评分，平均为 0.8572（标准差为 0.005）。相比之下，合理组中随机抽取的构象 AF3 评分明显较低，平均为 0.8399（标准差为 0.004）。这表明 AF3 评分函数在本质上更偏好偏离构象。

Interestingly, we also observed one exceptional case within the reasonable group: a structure with a low iRMSD of 1.495 Å obtained a relatively high AF3 ranking score of 0.8501. Although this score remains below the majority of deviated cases, it is still higher than the average score of the reasonable group. This result indicates that the AF3 ranking function does not effectively differentiate subtle but critical structural deviations at the binding site. 有趣的是，我们在合理组中也观察到一个例外案例：一个 iRMSD 仅为 1.495Å 的构象获得了相对较高的 AF3 评分 (0.8501)。虽然该评分低于大多数偏离构象，但仍高于合理组的平均评分。这一结果表明，AF3 的评分函数未能有效区分结合位点中细微却关键的结构差异。

In contrast, the top seven conformations ranked by SAKE all fall within the reasonable structural range, indicating the robustness of its predictive performance.

相比之下，SAKE 排名前七的构象全部属于合理结构范围，体现出其预测性能的稳健性。

Similar to the case of 8S4K, most decoys for 8VGG fall within the near-native range, with  $\approx 77\%$  exhibiting  $iRMSD < 2 \text{ \AA}$ . Nevertheless, about 23% of the structures show  $iRMSD$  values at or above 4 Å. These high-error outliers arise from the same issue observed in 8S4K—namely, the presence of a solvent-exposed extended loop/ $\beta$ -strand whose position is difficult to model accurately. Interestingly, even though 77% of the ensemble represents near-native conformations, the AlphaFold3 ranking score still assigns disproportionately high weight to the erroneous decoys.

与 8S4K 类似，8VGG 的大部分欺骗构象属于近天然范围，其中约 77% 的构象  $iRMSD$  低于 2Å。然而，仍有约 23% 的构象  $iRMSD$  达到或超过 4Å。这些高误差离群点与 8S4K 中观察到的问题一致——即存在一个暴露在溶剂中的延伸环或  $\beta$  链，其位置难以准确建模。有趣的是，尽管 77% 的构象为近天然状态，AlphaFold3 的评分函数仍对那些错误构象赋予了过高权重。

The case of 8SOZ is more extreme: the best conformation ranks only 7<sup>th</sup> out of 100 based on the AF3 rank score, despite having a near-native RMSD of 1.885 Å and representing just 1% of the ensemble. In contrast, the remaining 99% of conformations show substantial deviations, with an average RMSD of 16.5 Å. Structural comparison suggests that this discrepancy is similarly caused by an anomalous domain-level deviation.

8SOZ 的情况更为极端：该构象集中 RMSD 最优的构象在 AF3 评分中仅排名第 7，尽管其具有接近天然的 RMSD 值 1.885Å，且仅占总构象的 1%。相比之下，其余 99% 的构象均存在显著偏差，平均 RMSD 为 16.5Å。结构比较表明，这种偏差同样源于一个异常的结构域级偏移。

Notably, the RMSDs of the six decoys ranked in AF3 before and after the best one are all around 16.5 Å. As shown clearly in Figure SI XXX, SAKE is able to accurately identify this extreme near-native structure from the 100 conformations. This demonstrates SAKE's potential as a more effective scoring function that can complement the limitations of AF3. In cases with large structural heterogeneity, the AF3 rank score fails to discriminate between

high- and low-quality models, and the scores it assigns do not reflect the actual differences in structural quality. In contrast, SAKE assigns a score of 1.268 Å to the best conformation, while its predictions for all other conformations with higher iRMSD values are consistently above 4 Å.

值得注意的是，在 AF3 评分中排名前后紧邻该最佳构象的六个构象，其 RMSD 值均在 16.5。如图 SI XXX 所示，SAKE 能够在 100 个构象中准确识别出这一极端的近天然结构。这表明 SAKE 作为一种更有效的评分函数，有潜力弥补 AF3 的局限性。在结构异质性较大的情况下，AF3 评分函数无法有效区分高质量与低质量模型，其评分结果并未真实反映结构质量的差异。相比之下，SAKE 为最佳构象赋予 1.268，而对其余高 iRMSD 构象的预测值均稳定保持在 4。

As shown in Figures 2e and f, this is similarly a case where scoring bias arises from an outlier conformation with abnormal structural deviations. Although the proportion of such deviated structures is 23%, AF3's rank score still selected one of the deviated conformations as the top model. In contrast, SAKE once again demonstrated robustness in handling such cases by correctly identifying the near-native structure.

如图 2e 和 f 所示，该案例同样体现了由于离群构象的结构异常偏差而导致的评分偏误。尽管这些偏离构象仅占整体的 23%，AF3 评分函数仍将其中一个偏离构象选为 Top 模型。相较之下，SAKE 再次展现了其在处理此类情况时的稳健性，成功识别出近天然结构。

#### Failure Cases

Although SAKE-PP generally performs well, we identified several high-ranking cases where its predictions are still suboptimal. Closer inspection shows that these errors all arise from pose-selection issues.

尽管 SAKE-PP 整体表现良好，但我们仍识别出若干高分排序构象，其预测结果存在一定的次优情况。进一步检查发现，这些错误均源于构象选择上的偏差问题。

Taking 8PE2 as a representative example, the decoy selected by SAKE-PP ( $i\text{RMSD} = 11.536 \text{ \AA}$ ) shares a largely overlapping backbone with both the crystal structure and the best AF3 decoy ( $i\text{RMSD} = 0.987 \text{ \AA}$ ). The key difference is that the ligand chain is inserted almost 90° away from the native orientation, indicating that the interaction-determining residues were mis-positioned. We ran 1  $\mu\text{s}$  of molecular-dynamics simulations on both conformations; strikingly, both states remained stable throughout the trajectory. This finding suggests that, for this particular complex, the rotated insertion is not merely an artefact but may represent an alternative conformer of the same system.

以 8PE2 为代表性示例，SAKE-PP 所选构象 ( $i\text{RMSD} = 11.536 \text{ \AA}$ ) 与晶体结构及 AF3 评分下的最佳构象 ( $i\text{RMSD} = 0.987 \text{ \AA}$ ) 在主链骨架上高度重叠。其主要区别在于配体链偏离天然方向近 90° 插入，表明决定相互作用的残基位置发生了偏移。我们对这两种构象分别进行了 1 微秒的分子动力学模拟；令人惊讶的是，两种状态在整个轨迹中均保持稳定。该发现提示，对于该特定复合物，该旋转插入并非构象异常，而可能代表同一体系下的可共存替代构象。

#### Antibody:SAKE-PP benchmark beyond the training distribution

To probe how well SAKE-PP performs outside the distribution on which it was trained, we designed an antibody–antigen benchmark. This setting is deliberately challenging: the model was optimized only on a generic set of

dimeric protein–protein interactions and has never seen multi-chain antibody–antigen complexes during training. Its performance here therefore provides a direct measure of its ability to generalize to entirely new protein classes. 为评估 SAKE-PP 在超出训练分布情境下的表现，我们设计了一个抗体–抗原基准测试。该测试具有高度挑战性：模型仅在常规的二聚体蛋白–蛋白相互作用上进行训练，从未接触过多链抗体–抗原复合物。因此，该任务可以直接衡量模型在全新蛋白类别上的泛化能力。

For the benchmark we began with the SAbDab dataset curated by Fang et al. Redundancy was removed, chains shorter than 20 residues were discarded, and only entries containing complete antibody–antigen pairs were retained. After these filters, 139 unique antibody–antigen complexes remained and were used for evaluation. For each complex, AlphaFold3 (AF3) first generated 200 decoys. And then this decoy set was re-ranked by SAKE-PP (Fig. 4a).

在基准测试中，我们采用了 Fang 等人整理的 SAbDab 数据集。经过冗余去除、去除长度小于 20 残基的链，并仅保留包含完整抗体–抗原对的条目后，最终筛选得到 139 个独特抗体–抗原复合物用于评估。对于每个复合物，AlphaFold3 (AF3) 首先生成 200 个欺骗构象，然后由 SAKE-PP 对该构象集进行重排序（图 4a）。

Globally, the Pearson correlation between score and iRMSD is  $-0.2886$  for AF3—indicating a pronounced negative trend—whereas SAKE-PP improves the correlation to  $+0.1116$ , a gain of almost  $+0.40$ . On a per-complex basis, SAKE-PP outperforms AF3 in 106 of 139 cases (76.3%), while AF3 is superior in only 33 cases (23.7%). Strikingly, 70 complexes ( $>50\%$  of the test set) flip from negative to positive correlation when rescored with SAKE-PP, underscoring the method’s stronger interface-aware ranking capability.

从整体上看，AF3 评分与 iRMSD 之间的皮尔逊相关系数为  $-0.2886$ ，表现出明显的负相关趋势；而 SAKE-PP 将该相关性提高至  $+0.1116$ ，实现了近  $+0.40$  的提升。在每个复合物层面上，SAKE-PP 在 139 个案例中优于 AF3 的有 106 个，占比 76.3%，而 AF3 仅在 33 个案例中更优，占比 23.7%。更引人注目的是，有 70 个复合物（超过 50% 的测试集）在 SAKE-PP 重排序后由负相关变为正相关，突显出 SAKE-PP 在识别界面结构信息方面的评分优势。

Although SAKE-PP trails the native AF3 ranker by a single complex in the “pick-one” headline metric (AF3 selects the lower-iRMSD top-1 decoy in 70 cases, SAKE-PP in 69), a closer look at the difficult subset reveals a very different picture. We isolated the 18 antibody–antigen complexes for which AF3 assigns a high confidence score (ranking score  $> 0.75$ ) yet the corresponding decoy is clearly wrong ( $iRMSD > 4 \text{ \AA}$ ). In other words, these are the instances where the AF3 score is most misleading. Within this challenging cohort, SAKE-PP outperforms AF3 in 14 out of 18 cases, correctly demoting the erroneous high-score decoys and surfacing nearer-native alternatives.

虽然在“单点挑选”指标上，SAKE-PP 略逊于 AF3 一个案例（AF3 在 70 个案例中选出 iRMSD 更低的 Top-1 构象，SAKE-PP 为 69 个），但深入分析困难样本子集后却揭示出截然不同的情形。我们筛选出 18 个抗体–抗原复合物，其中 AF3 评分较高 ( $>0.75$ )，但对应构象的 iRMSD 明显偏大 ( $>4\text{\AA}$ )，即 AF3 评分严重误导的典型案例。在这一挑战性子集中，SAKE-PP 在 18 个案例中有 14 次优于 AF3，成功将评分高但错误的构象降级，并提升了更接近天然的替代构象。

Abramson et al. have argued that attaining peak accuracy with AlphaFold3 (AF3) in antibody–antigen system requires generating and re-ranking a very large number of seed decoys, which incurs substantial computational overhead. To test this claim, we expanded each antibody–antigen decoy pool from 50 to 200 structures and retained

only the top-ranked model according to either SAKE-PP or the native AF3 ranking score, while also recording the theoretical optimum (the lowest iRMSD in the pool).

Abramson 等人指出，在抗体-抗原系统中，若希望 AF3 达到峰值准确性，通常需要生成并重排序大量种子构象，这将带来显著的计算开销。为验证这一观点，我们将每个抗体-抗原的欺骗构象池从 50 个扩展至 200 个，仅保留按 SAKE-PP 或 AF3 评分函数排序的 Top-1 模型，同时记录理论最优构象（即构象池中 iRMSD 最低者）。

As shown in Fig. 4e, the mean iRMSD of all three traces decreases as the number of seeds doubles, confirming that deeper sampling indeed increases the chance of encountering near-native poses. AF3's average remains slightly better than that of SAKE-PP, a difference driven mainly by a few outlier complexes (e.g., 7XJ6, 7XJ8) where SAKE-PP makes large orientation errors—highlighting that the method can still be misled by extreme local heterogeneity. 如图 4e 所示，随着种子数量翻倍，三条曲线的平均 iRMSD 均出现下降，验证了更深入的采样确实提高了找到近天然构象的几率。AF3 的整体平均略优于 SAKE-PP，主要由少数离群复合物（如 7XJ6、7XJ8）推动，在这些案例中 SAKE-PP 发生了明显的方向性误判，表明该方法在面对极端局部异质性时仍可能受干扰。

From a downstream perspective, however, what matters most is the quality of the top- $k$  candidates that will proceed to molecular-dynamics refinement or affinity maturation. We therefore compared the best iRMSD found within the top 1, top 5, and top 10 decoys out of 200 (Fig. 4f) and benchmarked them against randomly chosen sets of the same size. Across all thresholds, SAKE-PP consistently outperforms the random baseline and converges to an average iRMSD of  $\sim 6 \text{ \AA}$  for the top 10, whereas AF3 lags behind random choice in the top-5 and top-10 scenarios, indicating that its native score struggles to enrich high-quality structures in large decoy pools. Taken together, while more extensive sampling benefits both methods in principle, SAKE-PP is markedly better at concentrating near-native decoys toward the very top of the ranking, delivering a smaller yet higher-quality candidate set that keeps subsequent MD simulations both reliable and computationally manageable.

然而从下游应用角度来看，最重要的是进入分子动力学精修或亲和力优化流程的 Top- $k$  候选构象的质量。因此我们在 200 个构象中比较了 Top-1、Top-5 和 Top-10 范围内的最佳 iRMSD 值（图 4f），并与相同大小的随机构象集进行了基准对比。结果表明，在所有阈值下，SAKE-PP 均明显优于随机基线，并在 Top-10 内收敛至平均约 6；而 AF3 在 Top-5 和 Top-10 情形中表现甚至低于随机选取，说明其评分函数在大构象池中难以有效富集高质量构象。综上所述，虽然更深度的采样原则上对两种方法都有所帮助，但 SAKE-PP 显著优于 AF3，能更有效地将近天然构象集中在排名前列，从而为后续的 MD 模拟提供更可靠且计算可控的高质量候选结构。

It is important to emphasize that SAKE-PP was never trained on antibody–antigen complexes; it was optimized solely on a general protein–protein dataset. The strong performance observed here therefore underscores the model's ability to generalize beyond its training distribution.

需要强调的是，SAKE-PP 从未在抗体-抗原复合物上进行训练；其仅在通用蛋白-蛋白数据集上进行了优化。因此，该方法在此类任务中的强劲表现进一步印证了其超越训练分布的泛化能力。

## Ablation Study

The comparative performance of SAKE-PP and baseline models under 5-fold and 10-fold cross-validation is summarized in Table 1 and visually depicted in Figure 5. Our proposed model demonstrates consistently superior performance across all evaluation metrics. In the 5-fold cross-validation setting, SAKE-PP achieves a mean MAE

Table 1: Model Comparison under 5-Fold and 10-Fold Cross-Validation. The results are presented as mean  $\pm$  standard deviation (Std). **Bold** indicates the best performance and underline indicates the second-best.

5-Fold	MAE	RMSE	R
Our Model	$0.9451 \pm 0.0157$	<u><math>1.2912 \pm 0.0275</math></u>	$0.6164 \pm 0.0118$
EGNN	<u><math>0.9922 \pm 0.0203</math></u>	$1.2986 \pm 0.0555$	$0.5983 \pm 0.0071$
DGN	$0.9966 \pm 0.0796$	$1.2886 \pm 0.2507$	<u><math>0.6035 \pm 0.0810</math></u>
GCN	$1.0983 \pm 0.0186$	$1.423 \pm 0.0490$	$0.4933 \pm 0.0129$
GAT	$2.0448 \pm 0.2599$	$1.9126 \pm 0.3420$	$0.0021 \pm 0.0177$
HGT	$1.2081 \pm 0.0179$	$1.575 \pm 0.0066$	$0.0059 \pm 0.0240$
10-Fold	MAE	RMSE	R
Our Model	$0.9466 \pm 0.0336$	$1.2639 \pm 0.0357$	$0.6240 \pm 0.0206$
EGNN	<u><math>0.9892 \pm 0.0292</math></u>	<u><math>1.2705 \pm 0.1119</math></u>	$0.6034 \pm 0.0244$
DGN	$0.9860 \pm 0.0932$	$1.2787 \pm 0.3093$	<u><math>0.6094 \pm 0.0831</math></u>

of  $0.9451 (\pm 0.0157)$ , significantly outperforming EGNN ( $0.9922 \pm 0.0203$ ), DGN ( $0.9966 \pm 0.0796$ ), and traditional GNN architectures such as GCN and GAT (Figure 5a). Similar performance advantages are observed for RMSE ( $1.2912 \pm 0.0275$  vs.  $1.2986$ – $1.9126$ , Figure 5b) and Pearson’s R ( $0.6164 \pm 0.0118$  vs.  $\leq 0.6035$ , Figure 5e), confirming both higher predictive accuracy and stronger correlation with ground truth values.

表 1 和图 5 总结了 SAKE-PP 与基线模型在 5 折与 10 折交叉验证中的比较表现。我们提出的模型在所有评估指标上始终展现出更优性能。在 5 折 CV 中, SAKE-PP 的平均绝对误差 (MAE) 为  $0.9451 (\pm 0.0157)$ , 显著优于 EGNN ( $0.9922 \pm 0.0203$ )、DGN ( $0.9966 \pm 0.0796$ ) 以及传统 GNN 如 GCN 和 GAT (图 5a)。在均方根误差 (RMSE) 方面也表现类似优势 ( $1.2912 \pm 0.0275$  对比  $1.2986$ – $1.9126$ , 图 5b); 在 Pearson’s R 也表现出更强相关性 ( $0.6164 \pm 0.0118$  对比最高为  $0.6035$ , 图 5e), 证实了更高的预测精度与与真实值的一致性。

Notably, SAKE-PP exhibits greater stability, evidenced by consistently lower standard deviations across all evaluated metrics. For instance, under 5-fold validation, SAKE-PP’s MAE standard deviation (0.0157) is approximately 22% smaller compared to EGNN (0.0203) and remarkably 80% smaller relative to DGN (0.0796). Such robustness is equally apparent in the 10-fold validation scenario, where SAKE-PP maintains a narrower MAE spread of 0.0336 versus DGN’s 0.0932 (Figure 5c), indicating superior generalization and less sensitivity to data splits. The RMSE metric further supports these findings, with SAKE-PP showing reduced variance and consistently lower mean errors (Figure 5d).

值得注意的是, SAKE-PP 在稳定性方面表现出更大优势, 其在各项评估指标上的标准差均较低。例如, 在 5 折验证中, SAKE-PP 的 MAE 标准差为 0.0157, 比 EGNN (0.0203) 低约 22%, 比 DGN (0.0796) 更是低了 80%。这一稳健性在 10 折验证中同样明显: SAKE-PP 的 MAE 波动范围仅为 0.0336, 而 DGN 则为 0.0932 (图 5c), 说明其具有更强的泛化能力并对数据划分不敏感。RMSE 指标进一步支持该结论, SAKE-PP 表现出更低的方差与均值误差 (图 5d)。

Overall, these analyses underscore SAKE-PP’s superior predictive performance and generalization capabilities compared to existing baseline models, highlighting its potential applicability to broader and more varied datasets. 综上, SAKE-PP 在预测性能和泛化能力方面均优于现有 baseline, 显示其在更广泛、更复杂数据集中的应用潜力。

## Discussion

In this study, we introduce SAKE-PP, a graph neural network (GNN) specifically designed for evaluating protein–protein interfaces. The model consistently outperforms the native ranking function of AlphaFold3 (AF3) across multiple evaluation metrics, particularly excelling in its ability to discriminate between structural conformations. Our findings reveal a critical bottleneck in current structure prediction workflows: although deep learning models such as AF3 are capable of generating diverse and high-quality decoy ensembles, their internal confidence scores often fail to identify conformations with the most biologically meaningful minimum-energy states.

本研究中，我们提出了 SAKE-PP——一个专为蛋白–蛋白界面评估设计的图神经网络 (GNN)。该模型在多项评估指标上持续优于 AlphaFold3 (AF3) 的原生评分函数，尤其在区分不同构象结构的能力上表现突出。我们的研究揭示了当前结构预测流程中的关键瓶颈：尽管 AF3 等深度学习模型可以生成多样且高质量的构象集合，其内部置信度评分却常常无法识别出那些具有生物学意义的最低能量构象。

We further observe that AF3 frequently favors structurally deviated conformations over near-native alternatives—even when such alternatives exist within the generated ensemble—a tendency largely driven by inflated confidence scores. This systematic bias not only hampers accurate assessment of interfacial energetics but also undermines structure-based drug design efforts that rely on high-affinity binding predictions.

我们进一步观察到，AF3 在生成的构象中常常偏好结构偏离的构象而非近天然构象，即使后者已包含在预测集合中。这种趋势主要由其夸大的置信评分驱动。这种系统性偏差不仅阻碍了界面能量的准确评估，也削弱了基于结构的药物设计对高亲和力结合预测的依赖。

SAKE-PP addresses these limitations through a physics-inspired architecture that integrates spatial attention mechanisms with Laplacian eigenvector orientation, enabling more accurate representation of interfacial features. In cross-dataset evaluations, SAKE-PP achieves a 13.75% improvement in iRMSD-based conformation selection accuracy, indicating its superior capability in recovering experimentally determined structures relative to AF3.

SAKE-PP 通过其受物理启发的架构解决了上述问题，该架构结合了空间注意机制与拉普拉斯特征向量方向信息，从而实现更准确的界面特征表征。在跨数据集评估中，SAKE-PP 在基于 iRMSD 的构象选择准确率上提升了 13.75%，显示出相较于 AF3 更强的恢复实验结构能力。

During training, we encountered several challenges. Although hierarchical sampling strategies were employed to capture diverse structural distributions, the model occasionally exhibits deviations between predicted and actual iRMSD values when encountering inputs far outside the training domain. This highlights the inherent complexity of modeling protein–protein interactions. Nevertheless, SAKE-PP maintains robust discriminative ability in fine-grained structural details, as particularly evident in its zero-shot generalization to antibody–antigen complexes: in this task, SAKE-PP outperformed AF3’s ranking score by a correlation margin of 0.4, despite AF3 having been trained on a large quantity of antibody–antigen data. This result suggests that SAKE-PP captures fundamental physical principles governing protein interfaces rather than simply learning statistical patterns.

在训练过程中，我们遇到了一些挑战。尽管我们采用了分层采样策略以覆盖多样结构分布，但当模型面对远离训练域的输入时，预测的 iRMSD 值与真实值之间仍会出现偏差。这突显了蛋白–蛋白相互作用建模的固有复杂性。尽管如此，SAKE-PP 在精细结构判别方面依然表现出强大能力，尤其是在对抗体–抗原复合物的零样本泛化任务中尤为明显：在该任务中，SAKE-PP 的相关性比分数超过 AF3 多达 0.4，尽管 AF3 接受了大量抗体–抗原数

据的训练。这一结果表明，SAKE-PP 捕捉到了控制蛋白界面的基本物理规律，而不仅仅是学习统计模式。

However, we also found that models relying solely on geometric features may sometimes diverge from biologically preferred conformations. For instance, in crystal structures derived from specific biological contexts such as signaling pathways, SAKE-PP selected a more tightly bound conformation, while AF3 favored a near-native pose. This discrepancy may arise from the absence of contextual biological information during training. Future models should consider incorporating biological priors or functional annotations to improve context-aware prediction.

但我们也发现，依赖纯几何特征的模型有时可能会偏离生物学偏好的构象。例如，在源自特定生物背景（如信号通路）的晶体结构中，SAKE-PP 倾向于选择结合更紧密的构象，而 AF3 则更倾向于选择近天然构象。这一差异可能源于训练过程中缺乏上下文生物学信息。未来的模型应考虑引入生物先验知识或功能注释，以提升其上下文感知预测能力。

To further validate these observations, we performed molecular dynamics simulations on the protein complex 8PE2, comparing competing conformations selected by different scoring functions. Surprisingly, despite significant structural differences at the outset, all conformations remained stable throughout the simulation. This suggests the existence of multiple coexisting, energetically viable states in the protein–protein interaction landscape, offering a plausible explanation for why structurally distinct yet biophysically reasonable conformations may be selected by different models.

为进一步验证上述发现，我们对蛋白复合物 8PE2 进行了分子动力学模拟，比较了由不同评分函数选择的竞争构象。令人惊讶的是，尽管初始结构差异明显，所有构象在整个模拟过程中均保持稳定。这表明在蛋白–蛋白相互作用能量景观中可能存在多个共存且能量可行的状态，合理解释了为何不同模型会选择结构上不同但在生物物理上合理的构象。

Taken together, these findings highlight the necessity of ensemble-based strategies that incorporate multiple starting conformations to fully characterize protein complex interfaces. Looking forward, the SAKE-PP framework can be extended to more complex biomolecular assemblies, including protein–nucleic acid complexes and higher-order oligomers. Moreover, its integration with free energy calculations and adaptive sampling techniques holds promise for enhancing both accuracy and efficiency in structure prediction and drug discovery workflows.

综上所述，这些发现强调了采用基于集成的策略、引入多个起始构象以全面刻画蛋白复合物界面的必要性。展望未来，SAKE-PP 框架可扩展至更复杂的生物大分子组装体，包括蛋白–核酸复合物及更高阶寡聚体。此外，将其与自由能计算和自适应采样技术相结合，有望在结构预测与药物发现流程中同步提升准确性与效率。

## Methods

### Data preparation and preprocessing

Protein–protein complexes containing two distinct polypeptide chains were identified in the PDBBind dataset. Structures were then refined using pdb4fixer by removing water molecules and other hetero atoms (e.g., small molecules), retaining only the main protein atoms. In cases where one chain was substantially shorter than the other, the shorter chain was designated as the “ligand,” while the longer chain was labeled the “receptor.” This procedure yielded 736 protein–protein complex structures in total.

我们在 PDBBind 数据库中筛选出包含两条不同多肽链的蛋白质-蛋白质复合物。随后使用 pdb4fixer 对结构进行预处理，去除水分子及其他杂原子（例如小分子），仅保留主链蛋白原子。若一条链显著短于另一条，则将短链标记为“配体”，长链标记为“受体”。该流程共获得 736 个蛋白-蛋白复合物结构。

Each protein–protein complex was docked using ZDock (version 3.0.2), generating 2000 conformations per complex. To introduce prior knowledge and ensure a reasonable docking space, a distance restraint of 15 Å was imposed between the “ligand” and “receptor,” based on their original closest contact. This restraint mimics the practical insight often employed in biochemical experiments to confine the search region around the putative binding site. Under this setting, a total of 1,472,000 candidate docking conformations were produced across all complexes.

我们使用 ZDock (版本 3.0.2) 对每个蛋白-蛋白复合物进行对接，每个复合物生成 2000 个构象。为了引入先验知识并确保合理的对接空间，我们在“配体”与“受体”之间施加了 15 Å 的距离约束，基于其原始最近接触点。该约束模拟了生化实验中常用的限定假设结合位点搜索范围的策略。在此设置下，总共生成了 1,472,000 个候选对接构象。

In the evaluation of protein-protein docking conformations, we employed a hierarchical sampling strategy centered on the interface RMSD (iRMSD). This approach was guided by prior findings: DiffDock reports an average conformational distribution iRMSD of 4.85 Å, while CAPPI considers 4 Å as the threshold distinguishing near-native conformations. Based on these observations, we designated the 4–5 Å range as a critical reference for assessing conformation quality.

在对蛋白-蛋白对接构象的评估中，我们采用了一种基于界面 RMSD (iRMSD) 的分层采样策略。该策略受到先前研究的启发：DiffDock 报告的平均构象分布 iRMSD 为 4.85 Å，而 CAPPI 将 4 Å 作为区分近天然构象的阈值。基于这些观察，我们将 4–5 Å 范围设定为评估构象质量的重要参考区间。

Although a regression-based method was utilized during training, an adaptive sampling scheme was devised to ensure a balanced set of training samples, with a particular focus on the 4–5 Å range. This scheme addresses the typically uneven and non-normal distribution of docking conformations generated by ZDock across different protein complexes. Specifically, we calculated multiple metrics, including iRMSD, the fraction of native contacts ( $f_{\text{nat}}$ ), and DockQ, to comprehensively evaluate the quality of all candidate conformations.

尽管训练中采用的是基于回归的方法，我们仍设计了一种自适应采样方案，以确保训练样本集的分布平衡，特别关注 4–5 Å 区间。该方案旨在应对 ZDock 在不同蛋白复合物中生成构象分布不均、非正态的普遍问题。具体而言，我们计算了多个指标，包括 iRMSD、天然接触比例 ( $f_{\text{nat}}$ ) 和 DockQ，以全面评估所有候选构象的质量。

The hierarchical sampling strategy was implemented as follows: If sufficient high-quality conformations were available within the 0–5 Å range, incremental sampling was performed at 10% intervals, complemented by uniform sampling in the 5–8 Å range to ensure representativeness. When the 0–5 Å range contained an insufficient number of conformations, the 10 highest-quality conformations from the 0–8 Å range were selected first, followed by uniform sampling of the remaining conformations until a total of 20 conformations was reached. In extreme cases, where the total number of conformations within the 0–8 Å range was fewer than 20, all available high-quality conformations were selected. Including the native conformations, we obtained a total of 15,456 conformations as our complete dataset.

分层采样策略具体如下：若在 0–5 Å 范围内存在足够的高质量构象，则以 10% 间隔进行递增采样，并在 5–8 Å 范围内均匀采样以确保代表性；若 0–5 Å 范围内构象数量不足，则首先从 0–8 Å 中选出 10 个最高质量构象，再从剩余构象中均匀采样至总数达到 20 个；若 0–8 Å 范围内总构象数不足 20 个，则选取所有可用的高质量构象。加上天然构象，共获得 15,456 个构象，构成完整数据集。

Fig b illustrates the conformational distribution obtained after the application of a strategic sampling approach. According to the CAPRI classification criteria (with a threshold of 4 Å), the proportion of acceptable conformations is 42.87%, while the proportion of unacceptable conformations is 57.13%. Besides, among the acceptable conformations, high-quality conformations (RMSD 0–2 Å) account for 34.01%, whereas borderline acceptable conformations (RMSD 2–4 Å) constitute 65.99%.

图 b 展示了应用策略采样方法后获得的构象分布。根据 CAPRI 分类标准 (4 Å 阈值)，可接受构象占比为 42.87%，不可接受构象占比为 57.13%。在可接受构象中，高质量构象 (RMSD 0–2 Å) 占 34.01%，而边缘可接受构象 (RMSD 2–4 Å) 占 65.99%。

Figure c illustrates the distribution of 1,472,000 candidate docking conformations within the range of 0–10 Å. The distribution exhibits a sandglass-like pattern, with a relatively higher number of conformations at high RMSD values. After sampling, the distribution is transformed into the pattern shown in Figure d.

图 c 展示了 1,472,000 个候选对接构象在 0–10 Å 范围内的分布。该分布呈现出“沙漏”形态，在高 RMSD 区间具有相对较多的构象。经过采样后，分布被调整为图 d 所示的模式。

Our sampling strategy ensures that the dataset achieves broad coverage of conformational distributions while maintaining a high level of sample representativeness in the critical range (RMSD 0–4 Å).

我们的采样策略保证了数据集在构象分布上的广泛覆盖，同时在关键区间 (RMSD 0–4 Å) 中维持了较高的样本代表性。

## Protein Structure Representation

We represent a protein structure as a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where each node  $v \in \mathcal{V}$  corresponds to one residue in the protein backbone, and edges  $\mathcal{E}$  capture their spatial proximity and biochemical relationships. Let  $|\mathcal{V}| = N$  denote the number of residues (nodes), and let  $\mathcal{E} \subseteq \{(u, v) \mid u, v \in \mathcal{V}\}$  be the edge set.

我们将蛋白质结构表示为图  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ ，其中每个节点  $v \in \mathcal{V}$  对应于蛋白质主链中的一个残基，边集  $\mathcal{E}$  用于捕捉它们之间的空间接近性和生化关系。设  $|\mathcal{V}| = N$  表示残基（节点）数量， $\mathcal{E} \subseteq \{(u, v) \mid u, v \in \mathcal{V}\}$  为边集合。

## Node Representation

### 节点表示

For each residue  $v \in \mathcal{V}$ , we collect both spatial and chemical features: 对于每个残基  $v \in \mathcal{V}$ ，我们提取其空间特征和化学特征：

Spatial Features.

空间特征：

- $C\alpha$  coordinates: We treat the entire protein-ligand complex (or multi-chain complex) as a single structure and extract all  $C\alpha$  atoms from the backbone. Each  $C\alpha$  atom is indexed by a node  $v$ , whose 3D position is given by  $x_v \in \mathbb{R}^3$  (i.e., the  $(x, y, z)$  coordinates). These coordinates allow the model to capture the overall geometry of the protein complex.

将整个蛋白-配体复合物（或多链复合物）视为一个整体结构，从其主链中提取所有  $C\alpha$  原子。每个  $C\alpha$  原子对应一个节点  $v$ ，其三维坐标为  $x_v \in \mathbb{R}^3$ （即  $(x, y, z)$  坐标）。这些坐标用于捕捉蛋白复合物的整体几何结构。

- Backbone dihedral angles: For each residue  $v$ , we calculate the main-chain dihedral angles  $(\phi_v, \psi_v)$  following the standard protein geometry definition. In rare cases where numerical issues arise (e.g., a tiny subset of angles failing to compute properly), we assign a small default value (e.g.,  $10^{-5}$ ) to avoid undefined behaviors in subsequent computations.

对于每个残基  $v$ ，根据标准蛋白质几何定义计算其主链二面角  $(\phi_v, \psi_v)$ 。在极少数数值问题（如部分角度计算失败）情况下，使用默认小值（如  $10^{-5}$ ）替代，以避免后续计算中的未定义行为。

- Angle-derived embedding: We adopt the angle-derived embedding  $a_v \in \mathbb{R}^3$  to convert dihedral angles into a 3D direction vector. Formally, 我们采用角度派生的嵌入  $a_v \in \mathbb{R}^3$ ，将二面角转换为三维方向向量。其形式定义为：

$$a_v = \text{angle2vec}(\phi_v, \psi_v) = [\sin(\phi_v) \cos(\psi_v), \sin(\phi_v) \sin(\psi_v), \cos(\phi_v)]. \quad (1)$$

#### Chemical Features. 化学特征：

- Residue type encoding: We represent each canonical amino acid with a one-hot vector or learned embedding,  $r_v \in \mathbb{R}^{20}$ .

将每种标准氨基酸用一个独热编码或可学习的嵌入向量表示，记为  $r_v \in \mathbb{R}^{20}$ 。

- Chain identifier: Each residue  $v$  is associated with a chain identifier  $c_v \in \mathbb{N}$ , which indicates which protein chain the residue belongs to (e.g., chain A, chain B, etc.).

每个残基  $v$  关联一个链标识符  $c_v \in \mathbb{N}$ ，指明其属于哪一条蛋白链（例如链 A、链 B 等）。

- $C\alpha$  flag: We include a binary flag  $f_v \in \{0, 1\}$  to specify whether the entry corresponds to a  $C\alpha$  atom (main-chain  $\alpha$  carbon).

我们引入一个二值标记  $f_v \in \{0, 1\}$ ，指示该条目是否对应于  $C\alpha$  原子（主链 碳）。

By concatenating the chemical features above with the spatial features described previously, we arrive at the complete node feature vector. Specifically, we form

将上述化学特征与前述空间特征连接后，即可构建出完整的节点特征向量。具体如下：

$$h_v = [r_v \oplus a_v \oplus x_v] \in \mathbb{R}^{26}, \quad (2)$$

where  $\oplus$  denotes vector concatenation. In this construction,

- $r_v \in \mathbb{R}^{20}$  encodes the residue identity,

- $\mathbf{a}_v \in \mathbb{R}^3$  (angle-derived embedding) represents dihedral angles,
- $\mathbf{x}_v \in \mathbb{R}^3$  stores the C $\alpha$  coordinates.

Hence, each residue  $v$  is ultimately represented by a 26-dimensional feature vector that integrates both chemical and spatial information. This design aligns with our code's practice of combining residue type, angle-based features, and 3D coordinates into a single node embedding.

因此，每个残基  $v$  最终被表示为一个 26 维的特征向量，融合了化学信息与空间信息。这一设计与我们代码实现中将残基类型、角度特征和三维坐标整合为单一节点嵌入的做法保持一致。

### Edge Representation 边表示

To fully capture inter-residue interactions, we define both the edge set  $\mathcal{E}$  and the edge features  $\mathbf{f}_{e_{uv}}$  in one integrated procedure. Specifically, we first construct edges based on complementary criteria, then compute a rich set of geometric and biochemical descriptors for each edge.

为全面刻画残基间相互作用，我们在一个统一的流程中定义边集合  $\mathcal{E}$  与边特征  $\mathbf{f}_{e_{uv}}$ 。具体而言，我们首先基于互补准则构建边结构，随后为每条边计算一组丰富的几何和生化描述符。

**Edge Construction.** We adopt two strategies to determine whether an edge should exist between residues  $u$  and  $v$ :

1. Contact-based rule:

$$\mathcal{E}_d = \{(u, v) \mid \|\mathbf{x}_u - \mathbf{x}_v\| < d_0, u \neq v\}, \quad (3)$$

where  $d_0 = 10 \text{ \AA}$  is a typical cutoff for identifying residue-residue contacts.

2. k-Nearest Neighbors (kNN): For each residue  $v$ , we pick  $k = 30$  closest neighbors by Euclidean distance on C $\alpha$  coordinates:

$$\mathcal{N}_k(v) = \text{argsort}_u(\|\mathbf{x}_u - \mathbf{x}_v\|)_-1:k, \quad \mathcal{E}_k = \bigcup_{v \in \mathcal{V}} \{(v, u) \mid u \in \mathcal{N}_k(v)\}. \quad (4)$$

We then merge these sets to form the final edge set:

$$\mathcal{E} = \mathcal{E}_d \cup \mathcal{E}_k, \quad (5)$$

ensuring that both long-range contacts (within  $d_0$ ) and local neighbors ( $k = 30$ ) are included.

**Edge Features.** Once the edge set  $\mathcal{E}$  is established, we compute an edge feature vector  $\mathbf{f}_{e_{uv}}$  for each  $(u, v) \in \mathcal{E}$ . This vector encodes geometric and biochemical information essential for downstream message passing. Concretely, we combine:

- Displacement vector and Distance:

$$\vec{\mathbf{e}}_{uv} = \mathbf{x}_v - \mathbf{x}_u, \quad l_{uv} = \|\vec{\mathbf{e}}_{uv}\|. \quad (6)$$

- Direction:

$$\hat{e}_{uv} = \frac{\vec{e}_{uv}}{l_{uv} + \varepsilon}, \quad (7)$$

where  $\varepsilon$  is a small positive constant (e.g.,  $10^{-6}$ ) to avoid division by zero.

- Radial Basis Expansion: In our implementation, we expand the raw distance  $l_{uv}$  via a set of radial basis functions:

$$RBF(l_{uv}) = \left[ \exp\left(-\frac{(l_{uv} - \mu_i)^2}{2\sigma^2}\right) \right]_{i=1}^{N_{rbf}}, \quad (8)$$

where  $\{\mu_i\}_{i=1}^{N_{rbf}}$  are centers evenly spaced in  $[0, d_0]$ ,  $\sigma$  is a bandwidth, and we fix  $N_{rbf} = 16$ . This 16-channel representation provides a smooth and more expressive encoding of inter-residue distances.

- Source and Target Residue Types: For each edge  $(u, v)$ , we can include the amino acid encodings  $r_u, r_v \in \mathbb{R}^{20}$ , thereby allowing the model to learn pairwise interaction patterns (e.g., hydrophobic vs. hydrophobic).
- Chain Interaction Flag:

$$\delta_{uv} = \mathbb{I}[c_u = c_v], \quad (9)$$

which indicates whether residues  $u$  and  $v$  belong to the same chain. In interface-focused tasks, one might specifically highlight the inter-chain edges (i.e.,  $c_u \neq c_v$ ).

Hence, the final edge feature vector (concatenating all the above) is:

$$f_{e_{uv}} = [l_{uv}, \hat{e}_{uv}, RBF(l_{uv}), r_u, r_v, \delta_{uv}, \dots]. \quad (10)$$

## SAKE Interaction Layer: Spatial Attention and Kinetic Updates

### SAKE 交互层：空间注意与动力学更新

In the previous work, we present a novel neural network named SAKE (spatial attention kinetic network with E(n)-Equivariance) layer aims to simultaneously update the features of (i) nodes (residue embeddings) and (ii) their 3D positions with E(n)-equivariance. In the following, we outline the essential components of SAKE and explain how they integrate into a single framework.

在前述工作中，我们提出了一种新型神经网络结构，称为 SAKE（具有 E(n) 等变性的空间注意动力网络）层，旨在同时更新 (i) 节点特征（即残基嵌入）与 (ii) 节点的三维空间位置，且保持 E(n) 等变性。下文将概述 SAKE 的关键组成部分，并解释其如何在统一框架中协同工作。

**Spatial Attention Definition. 空间注意机制定义：** Given a node  $v$  with embedding  $h_v \in \mathbb{R}^C$  and position  $x_v \in \mathbb{R}^n$  in graph  $\mathcal{G}$ , the spatial attention  $\phi^{SA}$  is designed to capture geometric relationships among neighbors  $u \in \mathcal{N}(v)$ : 设图  $\mathcal{G}$  中节点  $v$  的嵌入特征为  $h_v \in \mathbb{R}^C$ , 空间位置为  $x_v \in \mathbb{R}^n$ , 空间注意机制  $\phi^{SA}$  旨在捕捉节点  $v$  与其邻居  $u \in \mathcal{N}(v)$  之间的几何关系:

$$\phi^{SA}(v) = \mu \left( \bigoplus_{i=1}^{N_\lambda} \left\| \sum_{u \in \mathcal{N}(v)} \lambda_i(h_{e_{uv}}) f(\vec{e}_{uv}) \right\| \right), \quad (11)$$

where  $h_{e_{uv}}$  is the edge embedding,  $\lambda_i$  are attention weight functions (i.e., small MLPs) mapping edge embeddings to scalars, and  $f$  is an  $E(n)$ -equivariant transformation on the displacement vector  $\vec{e}_{uv} = \mathbf{x}_v - \mathbf{x}_u$ . 其中  $h_{e_{uv}}$  为边的嵌入表示,  $\lambda_i$  为注意力权重函数 (即小型多层感知机, MLP), 将边嵌入映射为标量;  $f$  是作用于位移向量  $\vec{e}_{uv} = \mathbf{x}_v - \mathbf{x}_u$  上的  $E(n)$  等变变换函数。

**Edge Embedding Update. 边嵌入更新:** Prior to applying attention weights, each edge  $(u, v)$  is assigned an embedding  $h_{e_{uv}}$  that combines both chemical and geometric features: 在应用注意力权重之前, 每条边  $(u, v)$  都会被赋予一个边嵌入  $h_{e_{uv}}$ , 该嵌入整合了化学特征与几何特征:

$$h_{e_{uv}}^{(k)} = \phi^e \left( h_u^{(k)} \oplus h_v^{(k)} \oplus \|\vec{e}_{uv}\| \oplus \text{RBF}(\|\vec{e}_{uv}\|) \odot f_r(h_u^{(k)} \oplus h_v^{(k)}) \right), \quad (12)$$

where RBF encodes distance into radial basis channels, and  $f_r$  is a small network that modulates the combined embeddings.

**Mixed Attention.** SAKE uses a combination of:

- A distance-based cutoff  $\alpha_{uv}^X$  to smoothly suppress edges beyond a certain threshold  $d_0$
- A semantic attention  $\alpha_{uv}^H$  akin to GAT-style softmax weights that highlight edges with higher feature similarity

These are combined as:

$$\alpha_{uv}^{X \times H} = \frac{\alpha_{uv}^X \alpha_{uv}^H}{\sum_{w \in \mathcal{N}(v)} \alpha_{wv}^X \alpha_{wv}^H}. \quad (13)$$

**Velocity and Position Updates.** The position updates are coupled with velocity vectors:

$$\mathbf{v}_v^{(k+1)} = \phi^{v \rightarrow V}(h_v^{(k)}) \mathbf{v}_v^{(k)} + \mathbf{W}_v \sum_i \sum_{u \in \mathcal{N}(v)} \lambda_i(h_{e_{uv}}^{(k)}) f(\vec{e}_{uv}^k), \quad \mathbf{x}_v^{(k+1)} = \mathbf{x}_v^{(k)} + \mathbf{v}_v^{(k)}. \quad (14)$$

## Laplacian Eigenvector Directional Graph Neural Network 拉普拉斯特征向量方向图神经网络

**Laplacian Matrix and Eigenvectors. 拉普拉斯矩阵与特征向量:** For an undirected graph  $\mathcal{G} = (V, E)$ , we define the graph Laplacian matrix  $\mathbf{L} \in \mathbb{R}^{N \times N}$  either in the combinatorial form  $\mathbf{L} = \mathbf{D} - \mathbf{A}$  or in the normalized form  $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$ , where  $\mathbf{A}$  is the adjacency matrix (with  $A_{ij} = 1$  if  $(i, j) \in E$  and 0 otherwise) and  $\mathbf{D}$  is the degree matrix ( $D_{ii} = \sum_j A_{ij}$ ).

对于一个无向图  $\mathcal{G} = (V, E)$ , 我们定义其图拉普拉斯矩阵  $\mathbf{L} \in \mathbb{R}^{N \times N}$ , 可以采用组合形式  $\mathbf{L} = \mathbf{D} - \mathbf{A}$ , 或标准化形式  $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$ 。其中,  $\mathbf{A}$  为邻接矩阵 (若  $(i, j) \in E$  则  $A_{ij} = 1$ , 否则为 0),  $\mathbf{D}$  为度矩阵 ( $D_{ii} = \sum_j A_{ij}$ )。

The Laplacian is symmetric positive semidefinite, so it admits an eigendecomposition  
拉普拉斯矩阵是对称正半定的, 因此它可以进行特征分解:

$$\mathbf{L} \mathbf{U} = \mathbf{U} \mathbf{\Lambda} \quad \text{or} \quad \tilde{\mathbf{L}} \mathbf{U} = \mathbf{U} \mathbf{\Lambda},$$

where  $\mathbf{U} = [\phi_0 \phi_1 \dots \phi_{N-1}]$  is an orthonormal basis of eigenvectors, and  $\mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{N-1})$  are the corresponding eigenvalues (sorted as  $0 \leq \lambda_0 \leq \lambda_1 \leq \dots$ ).

其中,  $\mathbf{U} = [\phi_0 \phi_1 \dots \phi_{N-1}]$  是由特征向量组成的正交归一基,  $\mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{N-1})$  为对应的特征值对角矩阵, 特征值按顺序排列:  $0 \leq \lambda_0 \leq \lambda_1 \leq \dots$ 。

The first eigenvector  $\phi_0$  associated with  $\lambda_0 = 0$  is the constant vector; The trivial eigenvector  $\phi_0$  associated with the constant vector is discarded, and the subsequent  $k$  nontrivial eigenvectors  $\{\phi_1, \dots, \phi_k\}$  are retained to capture more meaningful structural properties of  $\mathcal{G}$ .

第一个特征向量  $\phi_0$  对应特征值  $\lambda_0 = 0$ , 为常数向量; 我们舍弃这一平凡特征向量  $\phi_0$ , 保留其后  $k$  个非平凡特征向量  $\{\phi_1, \dots, \phi_k\}$ , 用于捕捉图  $\mathcal{G}$  中更具结构意义的属性信息。

Laplacian Eigenvector and directional aggregators in Laplacian Eigenvector Directional Graph Neural Network  
Let  $\mathcal{G} = (V, E)$  be the input graph with  $|V| = N$  nodes and  $|E| = M$  edges. Denote the node features by  $\mathbf{X} \in \mathbb{R}^{N \times d_{\text{in}}}$  and the edge features by  $\mathbf{E} \in \mathbb{R}^{M \times d_e}$ . We first map  $\mathbf{X}$  to the hidden dimension  $d_{\text{hid}}$  via

$$\mathbf{H}^{(0)} = \text{Dropout}\left(\sigma(\mathbf{X} \mathbf{W}_{\text{in}})\right),$$

where  $\mathbf{W}_{\text{in}} \in \mathbb{R}^{d_{\text{in}} \times d_{\text{hid}}}$  is a learnable matrix, and  $\sigma$  is the LeakyReLU activation function.

We compute the first  $k = 3$  Laplacian eigenvectors  $\in \mathbb{R}^{N \times 3}$ , serving as position embeddings (PE). That is, each row  $i,:$  stores the eigenvector values  $(\phi_1(i), \phi_2(i), \phi_3(i))$  for node  $i$ .

我们计算前  $k = 3$  个拉普拉斯特征向量, 形成位置嵌入 (PE), 记为  $\in \mathbb{R}^{N \times 3}$ 。其中, 每一行  $i,:$  存储节点  $i$  在三个特征向量维度上的值, 即  $(\phi_1(i), \phi_2(i), \phi_3(i))$ 。

Specifically, for an edge  $(i, j)$ , let  $\phi_k$  denote the  $k$ th Laplacian eigenvector. The directional difference  $F_{i,j} = \phi_k(j) - \phi_k(i)$  encodes how node  $j$  differs from node  $i$  in the  $k$ th eigenvector dimension.

具体而言, 对于一条边  $(i, j)$ , 设  $\phi_k$  为第  $k$  个拉普拉斯特征向量, 方向差  $F_{i,j} = \phi_k(j) - \phi_k(i)$  描述了节点  $j$  相对于节点  $i$  在第  $k$  个特征向量维度上的差异。

Two specialized aggregators, directional average and directional derivative, utilize  $|F_{i,j}|$  or  $F_{i,j}$  respectively to weigh neighbor features  $\mathbf{h}_j$ .

我们设计了两种方向性聚合器: 方向平均 (directional average) 与方向导数 (directional derivative), 分别使用  $|F_{i,j}|$  和  $F_{i,j}$  对邻居特征  $\mathbf{h}_j$  加权。

In our work, these directional aggregators are combined with standard ones such as mean, sum, and max, along with degree-based scalers (identity, amplification, attenuation), to capture both global and local structures. 在我们的设计中, 这些方向性聚合器与标准聚合器 (如 mean、sum 和 max) 相结合, 并引入基于节点度数的缩放器 (如 identity、amplification 和 attenuation), 以同时捕捉全局与局部的结构信息。

Hence, DGNConv offers richer geometric expressiveness than purely isotropic graph convolutions.

因此, 相比纯各向同性的图卷积, DGNConv 在几何表示能力上更加丰富。

Laplacian Eigenvector Directional Graph Neural Network Layers.

**拉普拉斯特征向量方向图神经网络层** For each layer  $\ell = 1, \dots, L$ , we apply a Laplacian Eigenvector Directional Graph Neural Network module, denoted by  $\mathcal{F}^{(\ell)}$ , which takes the graph  $(\mathcal{G}, \mathbf{H}^{(\ell-1)}, \mathbf{E}, \mathbf{ )}$  as input.

对于每一层  $\ell = 1, \dots, L$ , 我们应用一个拉普拉斯特征向量方向图神经网络模块, 记为  $\mathcal{F}^{(\ell)}$ , 其输入包括图结构  $(\mathcal{G}, \mathbf{H}^{(\ell-1)}, \mathbf{E}, \mathbf{ )}$ 。

Inside  $\mathcal{F}^{(\ell)}$ , we use five aggregators: mean, max, sum, dir1-av, dir1-dx, and three scalers: identity, amplification, attenuation.

在模块  $\mathcal{F}^{(\ell)}$  内部, 我们使用五种聚合器: mean、max、sum、dir1-av、dir1-dx, 以及三种缩放器: identity、amplification、attenuation。

Combining each aggregator with each scaler yields multiple feature transforms that are concatenated and projected back to  $d_{\text{hid}}$ .

每个聚合器与每个缩放器组合后, 生成多个特征变换, 这些变换被拼接并投影回隐藏维度  $d_{\text{hid}}$ 。

We then apply LeakyReLU and a dropout with rate  $p$  (e.g.,  $p = 0.15$ ).

随后我们对输出应用 LeakyReLU 激活函数, 并设置丢弃率为  $p$  (例如  $p = 0.15$ ) 的 dropout 操作。

The output of layer  $\ell$  is

第  $\ell$  层的输出为:

$$\mathbf{H}^{(\ell)} = \text{Dropout}\left(\sigma(\mathcal{F}^{(\ell)}(\mathcal{G}, \mathbf{H}^{(\ell-1)}, \mathbf{E}, \mathbf{ )})\right).$$

After stacking  $L$  layers of DGNConv, we obtain

$$\mathbf{H}^{(L)} \in \mathbb{R}^{N \times d_{\text{hid}}},$$

which we map to the output dimension  $d_{\text{out}}$  via a linear projection:

$$\mathbf{Y} = \mathbf{H}^{(L)} \mathbf{W}_{\text{out}} \in \mathbb{R}^{N \times d_{\text{out}}}.$$

Here,  $\mathbf{W}_{\text{out}} \in \mathbb{R}^{d_{\text{hid}} \times d_{\text{out}}}$  is a learnable parameter matrix. The returned node-level output  $\mathbf{Y}$  thus has one row per node, each row being a  $d_{\text{out}}$ -dimensional feature.

其中,  $\mathbf{W}_{\text{out}} \in \mathbb{R}^{d_{\text{hid}} \times d_{\text{out}}}$  是可学习的参数矩阵。最终返回的节点级输出  $\mathbf{Y}$  为一个矩阵, 每行对应一个节点的  $d_{\text{out}}$  维特征表示。

## SAKE-PP: Integrating SAKE and Laplacian Eigenvector Directional GNN

### SAKE-PP: 融合 SAKE 与拉普拉斯特征向量方向图神经网络

SAKE layers update both the node features and their 3D coordinates/velocities.

SAKE 层同时更新节点特征与其三维坐标/速度信息。

We initialize the node features via a learnable linear map and optional activation/dropout, 我们通过一个可学习的线性映射对节点特征进行初始化, 并可选择性地添加激活函数和 dropout 操作,

$$\mathbf{h}_{\text{SAKE}}^{(0)} = \text{Dropout}\left(\sigma(\mathbf{h} \mathbf{W}_{\text{init}})\right),$$

where  $W_{\text{init}} \in \mathbb{R}^{d_h \times d_{\text{SAKE}}}$ . Then each SAKE Interaction Layer applies the spatial attention and kinetic updates (11–14), producing

$$(h_{\text{res}}^{(\ell)}, x^{(\ell)}, v^{(\ell)}) = \text{SAKE\_Layer}(h_{\text{SAKE}}^{(\ell-1)}, x^{(\ell-1)}, v^{(\ell-1)}, \text{pairlist}).$$

Each residual node embedding  $h_{\text{res}}^{(\ell)}$  is then projected via an additional linear layer and concatenated with previous features, e.g.,

$$\tilde{h}_{\text{res}}^{(\ell)} = h_{\text{res}}^{(\ell)} W_{\text{add}}^{(\ell)}, \quad h_{\text{SAKE}}^{(\ell)} = \text{ConcatMLP}\left([h_{\text{SAKE}}^{(0)}, \dots, \tilde{h}_{\text{res}}^{(\ell)}]\right),$$

where ConcatMLP denotes a learned concatenation transform followed by batch normalization, activation, and dropout.

其中，ConcatMLP 表示一个可学习的拼接变换，后接批量归一化 (batch normalization)、激活函数以及 dropout 操作。

After the final SAKE Interaction Layer (say  $\ell = L_{\text{SAKE}}$ ), we obtain updated node embeddings  $h_{\text{SAKE}} \in \mathbb{R}^{N \times d_{\text{SAKE}}}$ , as well as new positions/velocities  $(x, v)$ .

在最后一层 SAKE 交互层 (记为  $\ell = L_{\text{SAKE}}$ ) 之后，我们获得了更新后的节点嵌入  $h_{\text{SAKE}} \in \mathbb{R}^{N \times d_{\text{SAKE}}}$ ，以及新的空间位置/速度信息  $(x, v)$ 。

In parallel, original node features  $h$  and the edge features  $E$  (together with the graph structure  $\mathcal{G}$ ) are feeded into the Laplacian Eigenvector Directional GNN (Sec. ).

同时，原始节点特征  $h$  与边特征  $E$  (以及图结构  $\mathcal{G}$ ) 会被送入拉普拉斯特征向量方向图神经网络模块 (见第 节)。

This module, denoted as  $dgn$ , outputs another node-level representation:

该模块 (记为  $dgn$ ) 输出另一组节点级表示：

$$h_{\text{DGN}} = dgn(\mathcal{G}, h, E).$$

Internally,  $dgn$  uses multiple DGNConv layers, each exploiting both standard and directional aggregators (e.g. dir1-av, dir1-dx) with degree-based scalers, ultimately producing

在内部， $dgn$  使用多个 DGNConv 层，每个层都利用标准和方向聚合器 (例如 dir1-av、dir1-dx) 以及基于度的缩放器，最终生成  $h_{\text{DGN}} \in \mathbb{R}^{N \times d_{\text{DGN}}}$ 。

We then concatenate the SAKE output  $h_{\text{SAKE}}$  and the DGN output  $h_{\text{DGN}}$  along the feature dimension:

$$h_{\text{combined}} = [h_{\text{SAKE}}, h_{\text{DGN}}] \in \mathbb{R}^{N \times (d_{\text{SAKE}} + d_{\text{DGN}})}.$$

An additional learnable transformation  $\text{combined\_processing}$  is applied, which may be a linear layer or a small MLP with activation:

$$h_{\text{combined}} = \text{combined\_processing}(h_{\text{combined}}).$$

To obtain a graph-level representation, we perform global pooling over the node dimension:

$$u_{\text{mean}} = \text{mean}(h_{\text{combined}}), \quad u_{\text{max}} = \text{max}(h_{\text{combined}}), \quad u_{\text{sum}} = \text{sum}(h_{\text{combined}}).$$

Here, mean, max, sum are taken node-wise (and separately within each subgraph if using a batched graph). We concatenate the three pooled vectors,

$$\mathbf{u}_{\text{graph}} = [\mathbf{u}_{\text{mean}}, \mathbf{u}_{\text{max}}, \mathbf{u}_{\text{sum}}],$$

and finally project to the desired scalar output (or any target dimension) via

$$\mathbf{y}_{\text{out}} = \mathbf{u}_{\text{graph}} \mathbf{W}_{\text{out}} \in \mathbb{R}^1 \quad (\text{or the appropriate output size}).$$

Thus, the SAKE-PP model jointly leverages the SAKE interaction mechanism (Eq. 11–14) and Laplacian Eigen-vector Directional Aggregators to capture both local geometric relations (via positions and velocities) and global structural cues from the eigenvector-based directional GNN.

因此，SAKE-PP 模型联合利用了 SAKE 交互机制（见公式 11–14）与拉普拉斯特征向量方向聚合器，从而同时捕捉局部几何关系（通过节点位置与速度）以及来自基于特征向量的方向图神经网络的全局结构信息。

## Data and Code Availability

We are currently consolidating and organizing the codebase. The source code will be released with the second version of the manuscript. All relevant data are provided in the Supplementary Information.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgement

This work was supported by National Natural Science Foundation of China (Grant nos. 21933010, 22250710136, 22333006). We sincerely thank the High-Performance Computing (HPC) resources at NYU Abu Dhabi and Greene at New York University, as well as the dedicated staff and their technical support.

## References

- [1] Nada, H.; Choi, Y.; Kim, S.; Jeong, K. S.; Meanwell, N. A.; Lee, K. Signal Transduct. Target Ther. 2024, 9, 1–32.
- [2] Westermarck, J.; Ivaska, J.; Corthals, G. L. Mol. Cell. Proteomics. 2013, 12, 1752–1763.
- [3] Lu, H.; Zhou, Q.; He, J.; Jiang, Z.; Peng, C.; Tong, R.; Shi, J. Signal transduction and targeted therapy 2020, 5, 213.
- [4] Batista, I. d. A. A.; Helguero, L. A. Signal Transduct. Target Ther. 2018, 3, 19.
- [5] Ni, D.; Lu, S.; Zhang, J. Med. Res. Rev. 2019, 39, 2314–2342.

- [6] Zinzalla, G.; Thurston, D. E. Future Med. Chem. 2009, 1, 65–93.
- [7] Jiang, L.; Zhang, K.; Zhu, K.; Zhang, H.; Shen, C.; Hou, T. Wiley Interdiscip. Rev. Comput. Mol. Sci. 2025, 15, e70016.
- [8] Carter, R.; Luchini, A.; Liotta, L.; Haymond, A. Curr. Pathobiol. Rep. 2019, 7, 61–71.
- [9] Sacquin-Mora, S.; Prévost, C. Biomolecules 2021, 11, 1529.
- [10] Rakers, C.; Bermudez, M.; Keller, B. G.; Mortier, J.; Wolber, G. Wiley Interdiscip. Rev. Comput. Mol. Sci. 2015, 5, 345–359.
- [11] Rehman, A. U.; Khurshid, B.; Ali, Y.; Rasheed, S.; Wadood, A.; Ng, H.-L.; Chen, H.-F.; Wei, Z.; Luo, R.; Zhang, J. Expert Opin. Drug Discov. 2023, 18, 315–333.
- [12] Zhao, N.; Wu, T.; Wang, W.; Zhang, L.; Gong, X. Interdiscip. Sci. Comput. Life Sci. 2024, 16, 261–288.
- [13] Wiehe, K.; Peterson, M. W.; Pierce, B.; Mintseris, J.; Weng, Z. Protein Struct. Predict. 2008, 283–314.
- [14] Gao, M.; Skolnick, J. Proteins Struct. Funct. Bioinform. 2011, 79, 1623–1634.
- [15] Kingsley, L. J.; Esquivel-Rodríguez, J.; Yang, Y.; Kihara, D.; Lill, M. A. J. Comput. Chem. 2016, 37, 1861–1865.
- [16] Shinobu, A.; Takemura, K.; Matubayasi, N.; Kitao, A. J. Chem. Phys. 2018, 149.
- [17] Evans, R.; O’Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; others biorxiv 2021, 2021–10.
- [18] Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; others Nature 2021, 596, 583–589.
- [19] Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; others Nature 2024, 630, 493–500.
- [20] Bao, J.; He, X.; Zhang, J. Z. J. Chem. Inf. Model. 2021, 61, 2231–2240.
- [21] Tan, L. H.; Kwoh, C. K.; Mu, Y. Brief. Bioinform. 2024, 25, bbae166.
- [22] Hiranuma, N.; Park, H.; Baek, M.; Anishchenko, I.; Dauparas, J.; Baker, D. Nat. Commun. 2021, 12, 1340.
- [23] Cao, Y.; Shen, Y. J. Chem. Theory Comput. 2020, 16, 5334–5347.
- [24] Nicolaou, K. Angew. Chem. 2014, 126, 9280–9292.
- [25] Parenti, M. D.; Rastelli, G. Biotechnol. Adv. 2012, 30, 244–250.
- [26] Govind Kumar, V.; Polasa, A.; Agrawal, S.; Kumar, T. K. S.; Moradi, M. Nat. Comput. Sci. 2023, 3, 59–70.
- [27] Tao, H.; Zhang, Y.; Huang, S.-Y. J. Chem. Inf. Model. 2020, 60, 2377–2387.

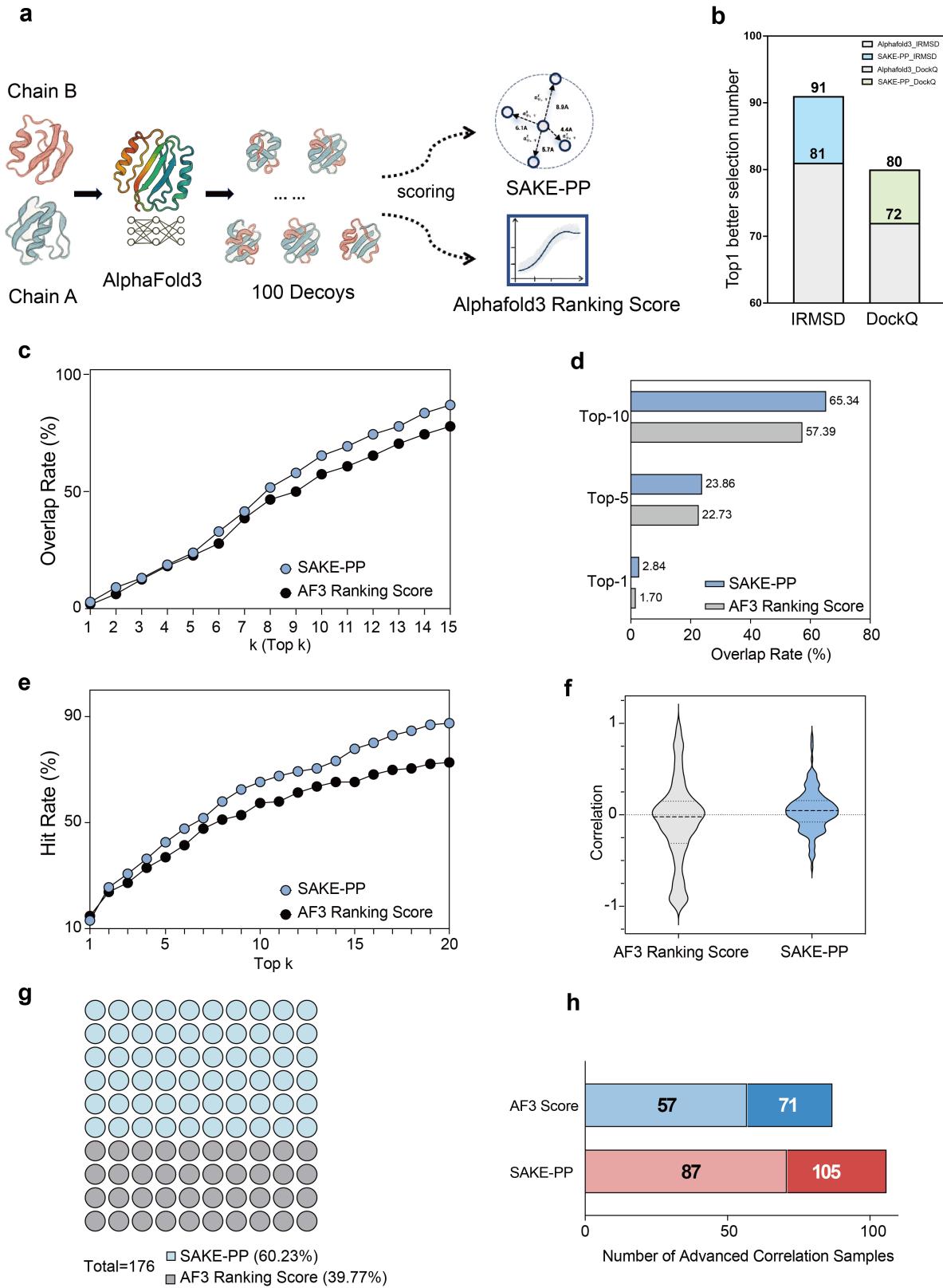


Figure 2: Benchmarking SAKE-PP against the AlphaFold3 (AF3) ranking score on a 176-complex docking set.  
在一个包含 176 个复合物的对接数据集中，对 SAKE-PP 与 AlphaFold3 (AF3) 评分函数的性能进行基准评估。  
(a) Workflow schematic. For each heterodimeric target, AF3 generates 100 decoys that are rescored independently by SAKE-PP and AF3 ranking score.

- (a) 工作流程示意图：对于每一个异源二聚体靶标，AF3 生成 100 个欺骗构象，并分别由 SAKE-PP 与 AF3 的原生评分函数独立重新评分。
- (b) Top-1 selection quality: 柱状图表示在各方法中，所选 Top-1 欺骗构象具有更低 iRMSD (左图) 或更高 DockQ 评分 (右图) 的复合物数量。
- (c) Overlap-rate summarised at Top-1, Top-5, and Top-10.
- (d) Success-hit-rate curves: probability of finding at least one of the ten lowest-iRMSD decoys within the top- $k$  ranks ( $1 \leq k \leq 20$ )。
- (e) Distribution of Pearson correlations ( $r$ ) between predicted score and iRMSD for each complex.
- (f) Distribution of Pearson correlations ( $r$ ) between predicted score and iRMSD for each complex.
- (g) Win–loss matrix (1 square = 1 complex). Blue squares denote complexes where SAKE-PP achieves a higher correlation; grey squares denote AF3

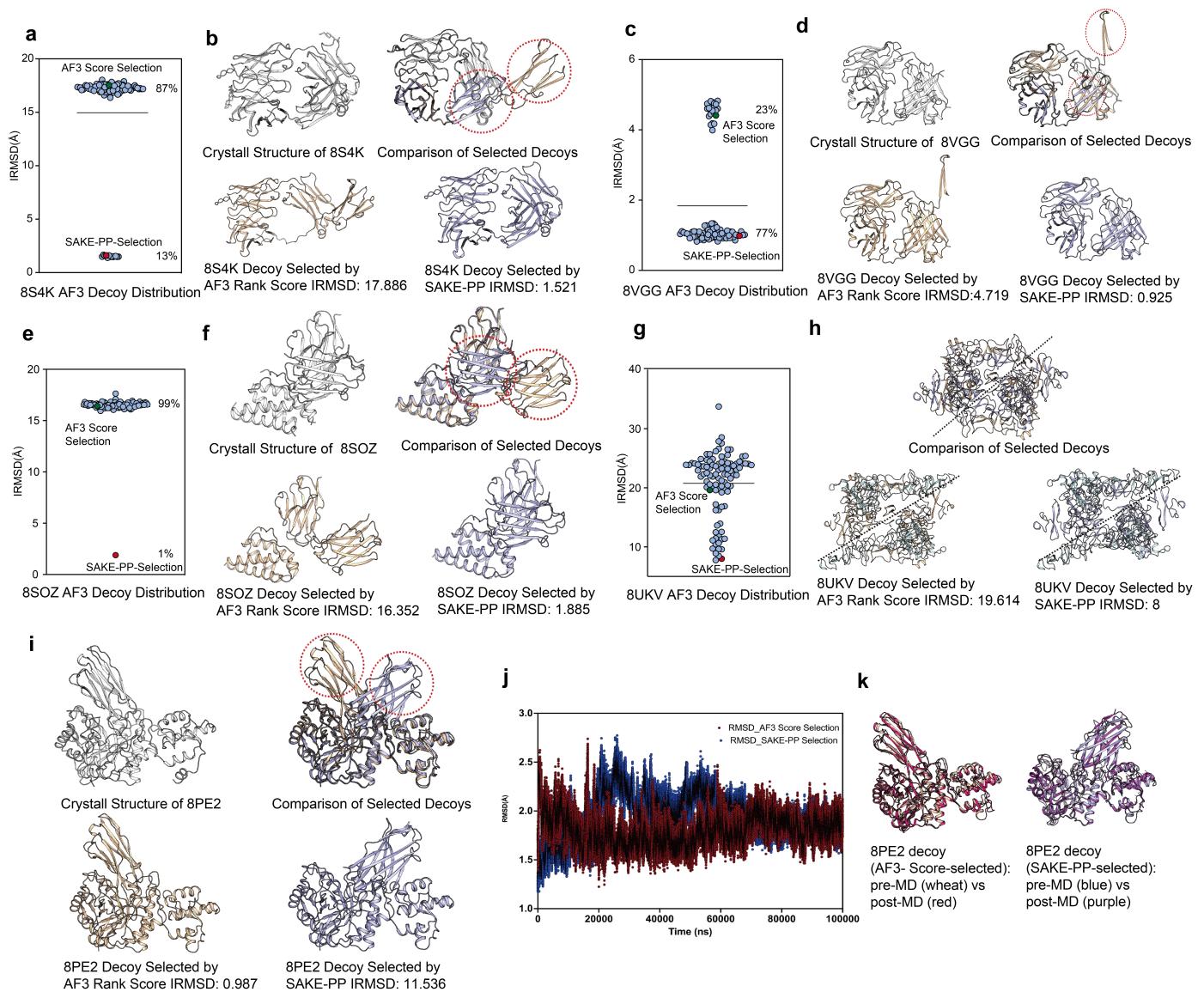


Figure 3: (a, c, e, g) Scatterplots of iRMSD for all 100 AF3-generated decoys of PDB IDs 8S4K, 8VGG, 8SOZ, and 8UKV. The symbol highlighted in red marks the decoy chosen by SAKE-PP; the blue symbol marks that chosen by the AF3 ranking score. Percentages indicate the fraction of decoys with equal or lower iRMSD than the AF3-selected model.

(a, c, e, g) PDB ID 为 8S4K、8VGG、8SOZ 和 8UKV 的蛋白复合物中，所有由 AF3 生成的 100 个欺骗构象的 iRMSD 散点图。红色标记表示 SAKE-PP 所选构象，蓝色标记表示 AF3 评分函数所选构象。图中百分比表示具有等于或低于 AF3 所选构象的 iRMSD 的构象所占比例。(b, d, f, h) Structural overlays. For each target the crystal structure is shown in grey. Decoys selected by AF3 (tan) and SAKE-PP (slate-blue) are superposed; red dashed circles highlight regions where the two selections differ markedly. Corresponding iRMSD values are listed beneath each panel.

(b, d, f, h) 结构叠加图：每个目标的晶体结构以灰色显示，AF3 选择的构象以浅棕色表示，SAKE-PP 选择的构象以青蓝色表示。红色虚线圈强调了两种选择在结构上显著差异的区域。每个子图下方标注了对应构象的 iRMSD 值。(i) 同样是 8PE2 的结构叠加图，但 SAKE-PP 有意选择了更不相似的起始构象 (iRMSD = 11.536 Å) 来测试采样鲁棒性。

(i) PDB ID 为 8PE2 的同类结构叠加图，其中 SAKE-PP 有意选择了一个更不相似的起始构象 (iRMSD = 11.536 Å) 以测试采样鲁棒性。(j) All-atom MD refinement of the AF3-selected (wheat) and SAKE-PP-selected (blue) 8PE2 decoys: backbone RMSD vs simulation time over 1  $\mu$ s. Post-MD conformations are coloured red (AF3) and purple (SAKE-PP).

(j) 对 8PE2 中由 AF3 选择（浅黄色）与 SAKE-PP 选择（蓝色）的欺骗构象进行全原子分子动力学精修：主链 RMSD 随 1 微秒模拟时间的变化。MD 模拟后构象分别用红色（AF3）和紫色（SAKE-PP）表示。(k) Final MD snapshots of the two 8PE2 decoys, showing that the SAKE-PP start converges towards the experimental structure while the AF3 start drifts away.

(k) 两个 8PE2 欺骗构象的 MD 最终快照，显示 SAKE-PP 选择的起始构象趋近于实验结构，而 AF3 选择的起始构象则发生偏移远离实验结构。

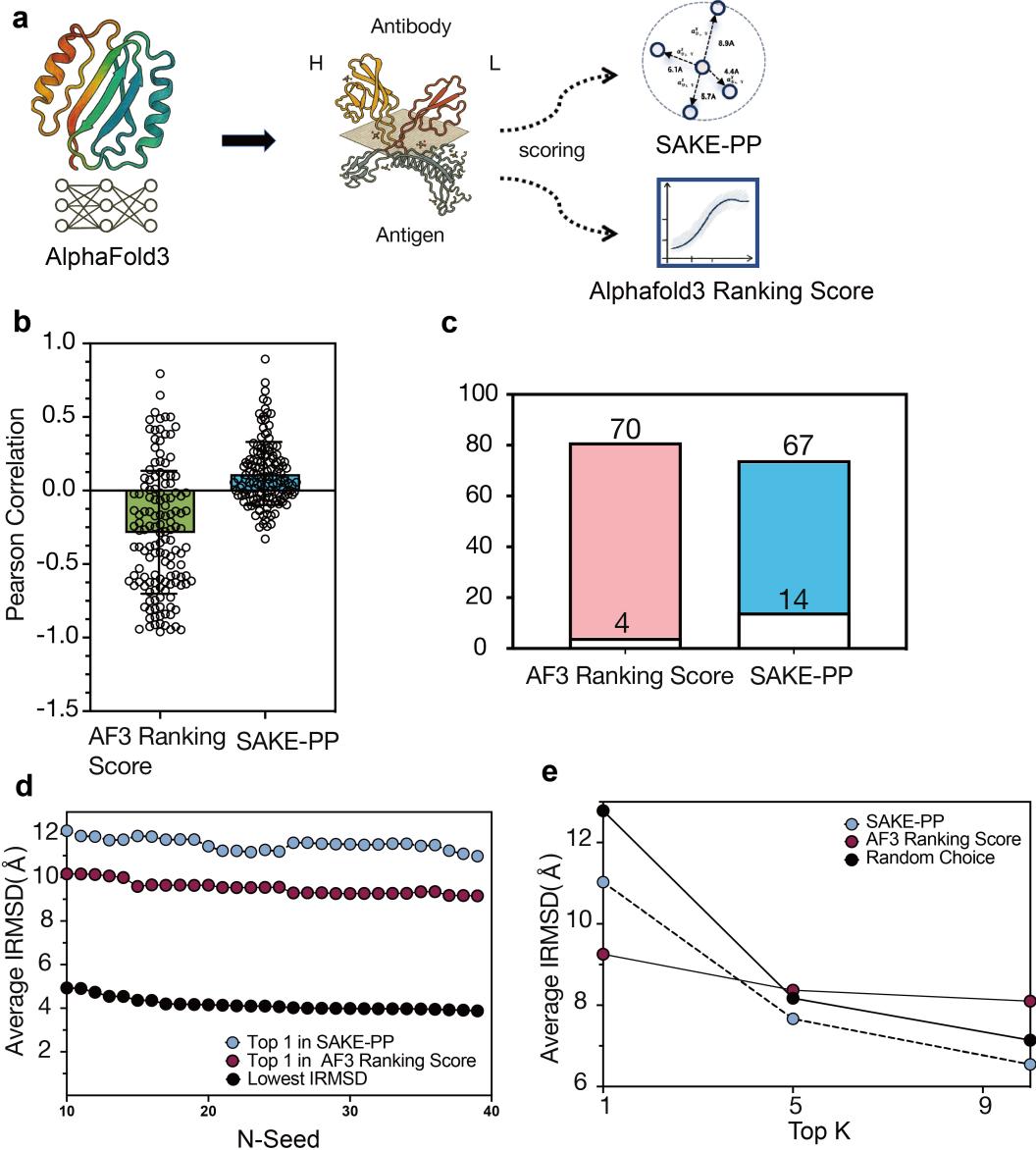


Figure 4: Antigen benchmark: SAKE-PP versus the AlphaFold3 (AF3) ranking score.

抗原基准测试：SAKE-PP 与 AlphaFold3 (AF3) 评分函数的比较。

(a) Workflow. AF3 first generates 100 decoys for each heavy-/light-chain antibody bound to its antigen; the decoys are then rescored by SAKE-PP (blue) or by the native AF3 ranking score (grey).

(a) 工作流程：AF3 首先为每个重链/轻链抗体与其抗原生成 100 个欺骗构象；这些构象随后由 SAKE-PP（蓝色）或 AF3 原生评分函数（灰色）进行重评分。(b) Pearson correlations ( $r$ ) between predicted score and iRMSD for every complex ( $n = 74$ ). Each point is a single complex; boxes indicate the inter-quartile range. AF3 correlations cluster around negative values (green), whereas SAKE-PP correlations are predominantly positive (blue).

(b) 每个复合物 ( $n = 74$ ) 中预测评分与 iRMSD 之间的皮尔逊相关系数 ( $r$ )。每个点表示一个复合物；箱线图显示四分位区间。AF3 的相关性集中在负值（绿色），而 SAKE-PP 的相关性则主要为正值（蓝色）。(c) Count of complexes with positive versus negative correlation. AF3 yields only 4 positive and 70 negative cases, while SAKE-PP reverses the trend with 67 positive and 14 negative cases.

(c) 正相关与负相关复合物的数量统计。AF3 仅有 4 个正相关与 70 个负相关，而 SAKE-PP 的趋势完全相反，有 67 个为正相关，仅 14 个为负相关。(d) Seed-selection experiment. As the number of seeds ( $N$ ) increases, the average iRMSD of the single best seed chosen by SAKE-PP (blue) or by AF3 (purple) is compared with the theoretical optimum—the lowest-iRMSD decoy in the pool (black).

(d) 种子选择实验：随着种子数量  $N$  的增加，SAKE-PP（蓝色）与 AF3（紫色）各自选出的最优欺骗构象的平均 iRMSD 与理论最优值（黑色）进行比较——理论最优值为构象池中 iRMSD 最低者。(e) Average iRMSD achieved when retaining the top- $k$  decoys ( $k = 1, 5, 9$ ). SAKE-PP (blue) consistently outperforms AF3 (purple) and a random baseline (black dashed).

(e) 在保留 Top- $k$  构象 ( $k = 1, 5, 9$ ) 时获得的平均 iRMSD。SAKE-PP（蓝色）在所有阈值下均优于 AF3（紫色）与随机基线（黑色虚线）。

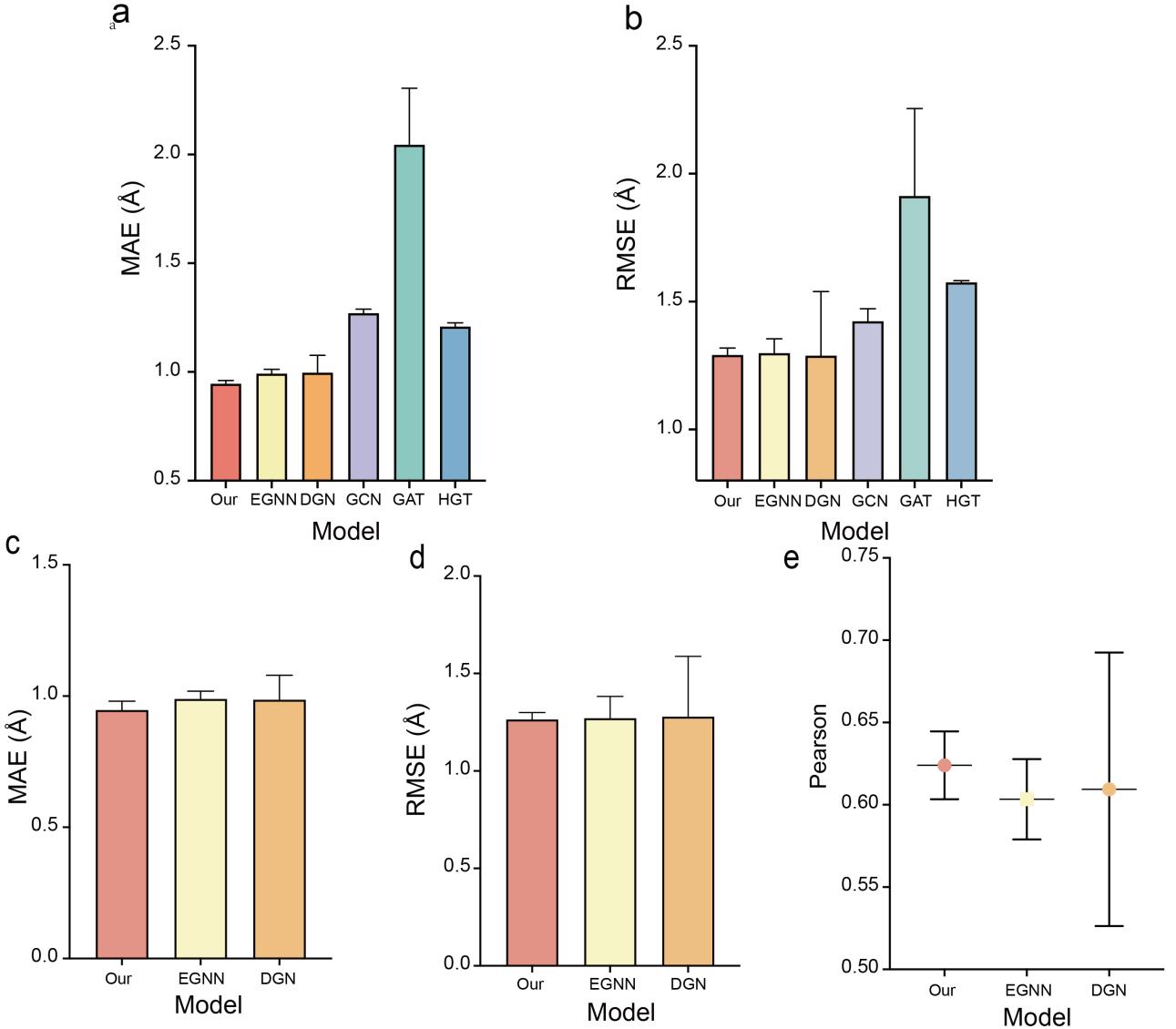
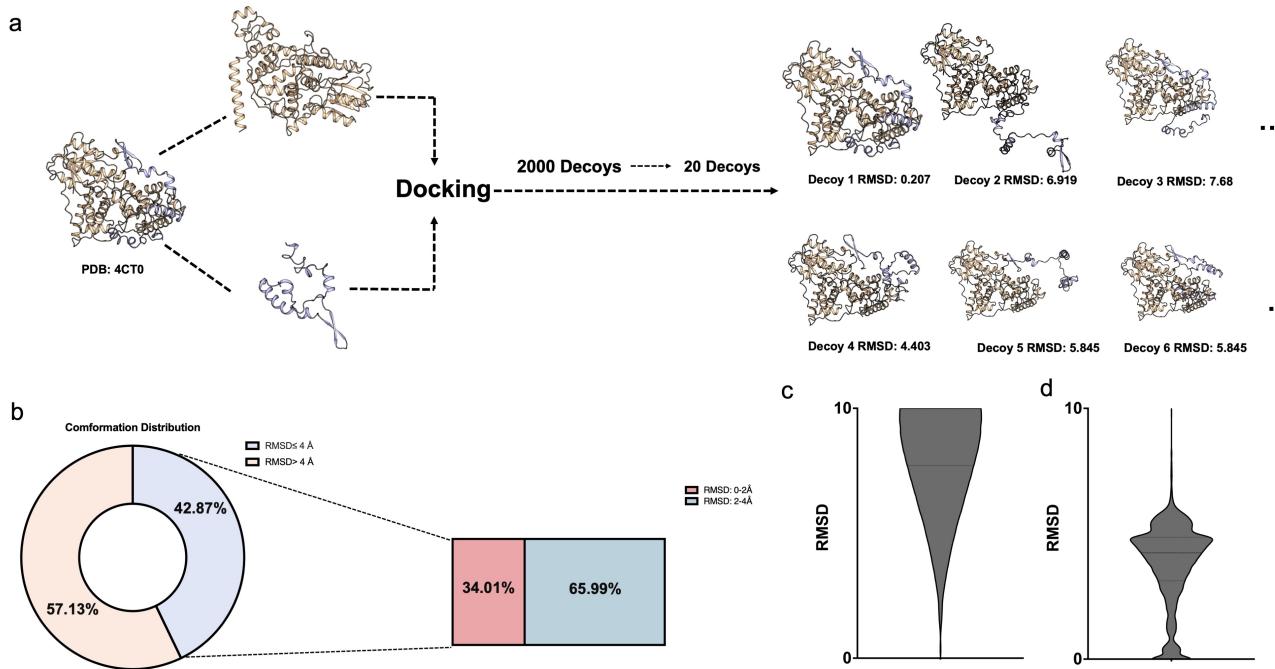


Figure 5: Comparative performance analysis of SAKE-PP and baseline graph neural network (GNN) models under 5-fold (a, b) and 10-fold (c–e) cross-validation. Panels (a) and (b) depict Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), respectively, comparing SAKE-PP against EGNN, DGN, GCN, GAT, and HGT in the 5-fold setting. Panels (c–e) illustrate the performance metrics (MAE, RMSE, and Pearson correlation coefficient, respectively) under 10-fold cross-validation.

SAKE-PP 与基线图神经网络 (GNN) 模型在 5 折 (a, b) 与 10 折 (c–e) 交叉验证下的性能对比分析。图 (a) 与 (b) 分别展示了平均绝对误差 (MAE) 和均方根误差 (RMSE)，对比对象包括 EGNN、DGN、GCN、GAT 与 HGT；图 (c)–(e) 展示了 10 折验证下的 MAE、RMSE 与皮尔逊相关系数 (Pearson R) 表现。



(b) Conformational distribution after sampling. By CAPRI criteria ( $i\text{RMSD} \leq 4 \text{ \AA}$  vs.  $> 4 \text{ \AA}$ ), 42.87% of decoys are acceptable (light blue) and 57.13% are unacceptable (peach); within the acceptable set, high-quality poses ( $0-2 \text{ \AA}$ ; 34.01%, red) and borderline poses ( $2-4 \text{ \AA}$ ; 65.99%, blue) are shown.

(c)  $i\text{RMSD}$  distribution of all 1,472,000 decoys in the 0–10  $\text{\AA}$  range, showing a “sandglass” shape with enrichment at high  $i\text{RMSD}$ .

(d)  $i\text{RMSD}$  distribution after sampling, illustrating balanced coverage and enrichment in the critical 0–4  $\text{\AA}$  region.

Figure 6: (a) From PDDBind, 736 protein–protein complexes were docked with ZDock under a 15  $\text{\AA}$  distance restraint to generate 2,000 decoys each, then hierarchically sampled down to 20 representative poses. Right: representative decoys for PDB 4CT0 and their interface RMSD ( $i\text{RMSD}$ ) values.

(b) Conformational distribution after sampling. By CAPRI criteria ( $i\text{RMSD} \leq 4 \text{ \AA}$  vs.  $> 4 \text{ \AA}$ ), 42.87% of decoys are acceptable (light blue) and 57.13% are unacceptable (peach); within the acceptable set, high-quality poses ( $0-2 \text{ \AA}$ ; 34.01%, red) and borderline poses ( $2-4 \text{ \AA}$ ; 65.99%, blue) are shown.

(c)  $i\text{RMSD}$  distribution of all 1,472,000 decoys in the 0–10  $\text{\AA}$  range, showing a “sandglass” shape with enrichment at high  $i\text{RMSD}$ .

(d)  $i\text{RMSD}$  distribution after sampling, illustrating balanced coverage and enrichment in the critical 0–4  $\text{\AA}$  region.

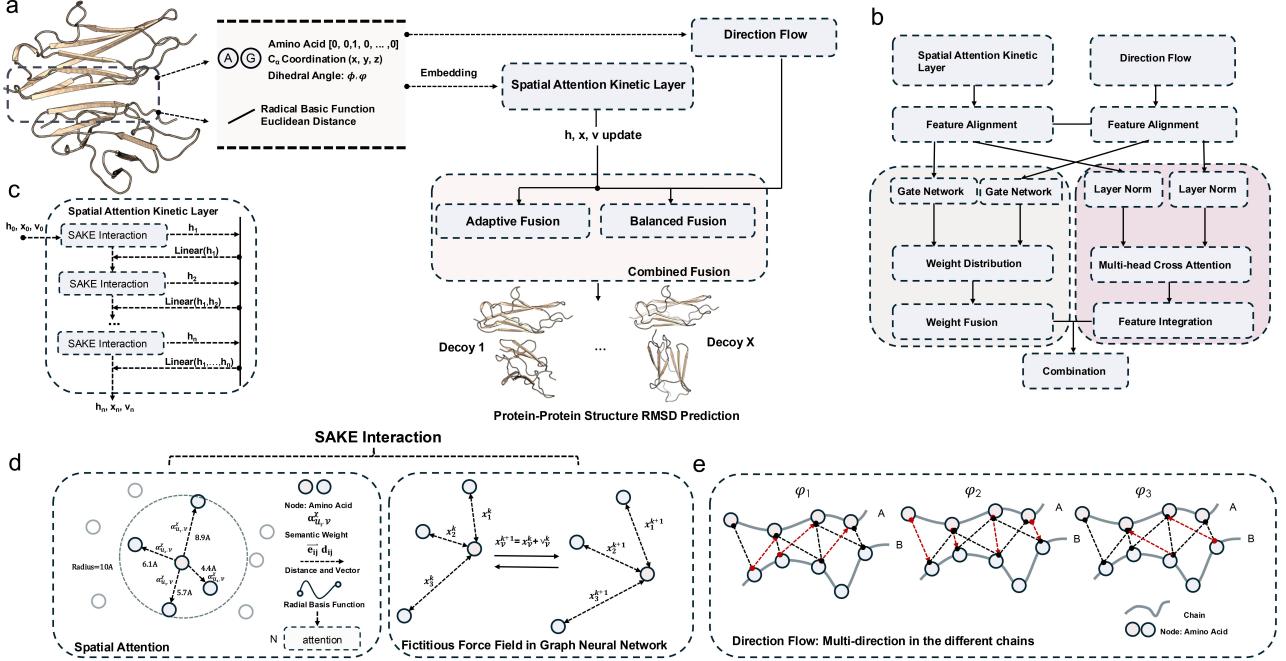


Figure 7: Workflow and architecture of SAKE-PP for protein–protein structure iRMSD prediction.

SAKE-PP 用于蛋白–蛋白结构 iRMSD 预测的工作流程与架构。(a) Workflow of SAKE-PP: 每个蛋白复合物通过氨基酸类型、二面角 ( $\phi, \psi$ )、C $\alpha$  原子坐标 ( $x, y, z$ ) 以及基于欧几里得距离的 10 Å 界面区域内的特征进行编码。所得表示被输入至 Direction Flow 模块与 Spatial Attention Kinetic Layer 模块，其输出将在融合模块中整合，用于预测标量形式的 iRMSD 值。(b) Feature fusion mechanism: 两个表示通路先经过特征对齐处理，随后通过门控网络、权重融合和多头交叉注意机制生成统一的潜在表示。(c) Spatial Attention Kinetic Layer: Composed of stacked SAKE Interactions, each connected via residual links. Each SAKE layer updates node features  $h$ , positions  $x$ , and velocities  $v$  in an  $E(n)$ -equivariant manner.

(a) SAKE-PP 工作流程：每个蛋白质复合物通过以下特征进行编码——氨基酸类型、二面角 ( $\phi, \psi$ )、C $\alpha$  原子坐标 ( $x, y, z$ ) 以及基于欧几里得距离的 10 Å 界面区域内的特征。所得表示被输入至 Direction Flow 模块与 Spatial Attention Kinetic Layer 模块，其输出将在融合模块中整合，用于预测标量形式的 iRMSD 值。(b) Feature fusion mechanism: 两个表示通路先经过特征对齐处理，随后通过门控网络、权重融合和多头交叉注意机制生成统一的潜在表示。(c) Spatial Attention Kinetic Layer: Composed of stacked SAKE Interactions, each connected via residual links. Each SAKE layer updates node features  $h$ , positions  $x$ , and velocities  $v$  in an  $E(n)$ -equivariant manner.

(b) 特征融合机制：两个表示通路先经过特征对齐处理，随后通过门控网络、权重融合和多头交叉注意机制生成统一的潜在表示。(c) Spatial Attention Kinetic Layer: Composed of stacked SAKE Interactions, each connected via residual links. Each SAKE layer updates node features  $h$ , positions  $x$ , and velocities  $v$  in an  $E(n)$ -equivariant manner.

(c) Space Attention Kinetic Layer: 由多个堆叠的 SAKE 交互模块组成，各层通过残差连接相连。每个 SAKE 层在  $E(n)$  等变框架下更新节点特征  $h$ 、节点位置  $x$  与速度向量  $v$ 。(d) SAKE Interaction Module: Integrates spatial attention and a fictitious force field for joint node and coordinate updates. The spatial attention mechanism combines radial basis encoding of pairwise distances with edge-based semantic attention, capturing local geometric dependencies. A mixed attention score  $\alpha_{uv}^{X \times H}$  modulates neighborhood aggregation. Position updates are coupled with learned velocity vectors that evolve under a physics-inspired force field: (d) SAKE 交互模块：将空间注意机制与拟力场相结合，实现节点特征与坐标的联合更新。空间注意机制通过对成对距离的径向基编码与基于边的语义注意融合，有效捕捉局部几何依赖关系。混合注意得分  $\alpha_{uv}^{X \times H}$  用于调控邻域聚合过程。位置更新由学习得到的速度向量驱动，在受物理启发的力场作用下演化： $x_v^{(k+1)} = x_v^{(k)} + v_v^{(k)}, v_v^{(k+1)} = \phi^{v \rightarrow V}(h_v^{(k)})v_v^{(k)} + \sum_{u \in \mathcal{N}(v)} \lambda_i(h_{e_{uv}}^{(k)}) f(\vec{e}_{uv}^k)$ 。

(e) Direction Flow: Captures multi-directional spatial interactions across protein chains using chain-aware angular relationships ( $\varphi_1, \varphi_2, \varphi_3$ ) to guide geometric feature propagation.

(e) Direction Flow 模块：利用链感知的角度关系 ( $\varphi_1, \varphi_2, \varphi_3$ ) 建模跨蛋白链的多方向空间交互，引导几何特征的传播与整合。