DYNAMICS-DRIVEN PROTEIN–LIGAND BINDING AFFINITY PREDICTION VIA CONFORMATIONAL FLEXIBILITY MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Protein-ligand binding affinity (PLBA) prediction has benefited substantially from structural information, yet available structures typically provide only static snapshots of the binding state, limiting the characterization of conformational and energetic evolution during molecular recognition. This limitation becomes particularly pronounced when binding spans multiple conformational states that reshape local interactions and energetics, leading to state-dependent variations in binding strength. To address this, we propose D-PLBA, a dynamics-driven framework that integrates dynamic pocket-ligand trajectory prediction into PLBA prediction. D-PLBA predicts motion trajectories by modeling pocket motion, ligand rigid-body displacement, and torsional deformation, while estimating the corresponding binding interaction energies. The resulting conformations and energies are then incorporated into the affinity predictor to jointly characterize the spatiotemporal evolution of binding structures and interactions. Extensive experiments demonstrate that D-PLBA substantially outperforms state-of-the-art methods across most evaluated metrics on all benchmarks, while reducing dynamic multi-conformation inference time to approximately 1% of that required by existing approaches. Our code and datasets can be found at: https://anonymous.4open.science/r/anonymous-703E.

1 INTRODUCTION

Computational prediction of protein-ligand binding affinity (PLBA) plays a critical role in accelerating the drug discovery process (Bai et al., 2023; Shah et al., 2025; Lu et al., 2025). Among existing approaches for PLBA prediction, structure-based methods (Jiang et al., 2021; Li et al., 2021) represent proteins and ligands using homogeneous structural features and explicitly model their binding conformations, enabling them to capture intermolecular interactions within a three-dimensional spatial context. The rapid advances in protein structure prediction, exemplified by AlphaFold (Jumper et al., 2021; Abramson et al., 2024) and ESMFold (Lin et al., 2023), have greatly expanded access to high-quality protein structures, further supporting the development of structure-based PLBA prediction. By incorporating explicit structural information, these methods better reflect the physical basis of molecular recognition, leading to improved predictive accuracy and interpretability (Yan et al., 2024). However, most existing structure-based PLBA methods (Zheng et al., 2019; Jones et al., 2021) rely solely on static protein-ligand structures and commonly model proteins as rigid bodies, overlooking the conformational flexibility involved in binding processes (Wu & Li, 2023). Ignoring protein conformational dynamics limits the ability of these models to capture biologically relevant structural transitions, resulting in substantial deviations between predicted binding affinities and experimentally observed values (Liu et al., 2025; Wang et al., 2026).

Specifically, most available protein structures, whether experimentally determined or computationally predicted, provide only a static view of molecular conformations. Experimental techniques (Wüthrich et al., 1983; Fernandez-Leiro & Scheres, 2016), including X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM), typically characterize proteins under specific conditions and conformational states, while structure prediction models such as AlphaFold and ESMFold primarily estimate a representative or most probable conformation rather than the underlying dynamic ensemble. Consequently, these structures

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

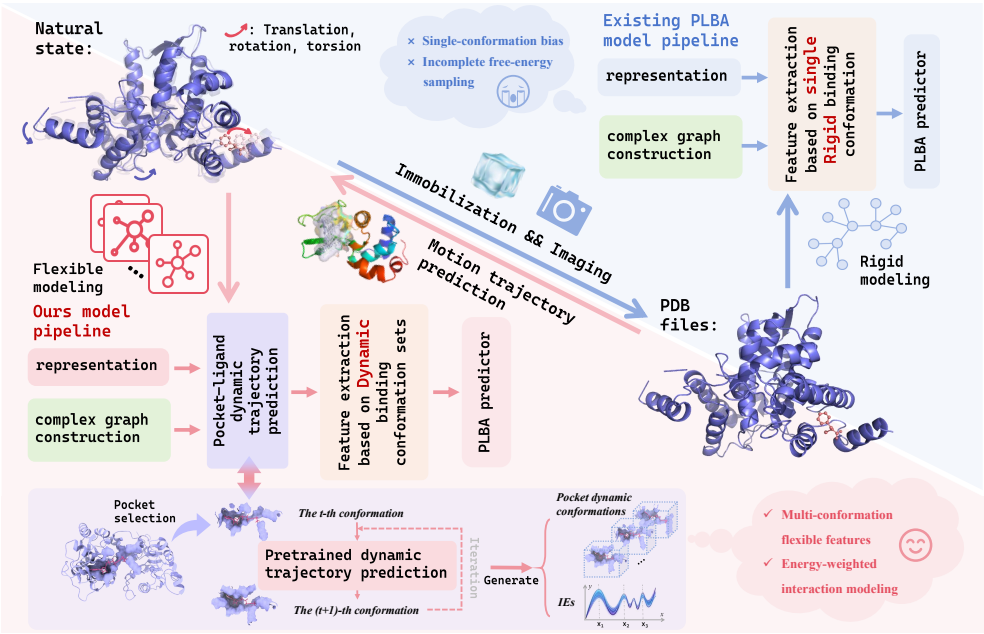

Figure 1: Blue pipeline: rigid-body modeling in conventional PLBA methods. Pink pipeline: D-PLBA, the first framework to integrate pocket-ligand dynamic trajectory prediction into PLBA modeling for flexible binding conformations.

provide limited information on binding-associated conformational motions (Lu et al., 2024; Shen et al., 2026). As illustrated by the blue pipeline in Figure 1, existing PLBA methods commonly build protein-ligand representations directly from such static structures, thereby inheriting the bias of a single conformation. This can be particularly limiting when the observed ligand pose does not fully reflect the relevant binding configurations, ultimately compromising affinity prediction (Feng et al., 2025).

More fundamentally, protein-ligand binding affinity is determined by the free-energy difference between the unbound and bound states (Wang et al., 2017; Mobley & Gilson, 2017), which comprise dynamic conformational ensembles rather than isolated structures. Accurate PLBA prediction therefore benefits from characterizing conformational flexibility and the associated energetic variations across binding states (Boehr et al., 2009). Static representations, however, provide limited information on these dynamic processes and may introduce conformational bias into PLBA models (Gkeka et al., 2020). Molecular dynamics (MD) simulations explicitly sample molecular motions and conformational trajectories (Abraham et al., 2015; Case et al., 2021), but conventional force-field-based simulations remain computationally expensive and require extensive sampling (Wang et al., 2004). AI-driven approaches (Shen et al., 2026; Cheng et al., 2026) have substantially accelerated trajectory generation, yet typically model atomic dynamics across the entire protein. This full-protein modeling paradigm incurs considerable computational costs while providing limited task-specific characterization of the coupled dynamics between binding pockets and ligands. These limitations hinder their application in large-scale PLBA prediction, highlighting the need for efficient approaches that capture molecular dynamics while preserving scalability.

To address this issue, we propose D-PLBA, a novel dynamics-driven framework for protein-ligand binding affinity prediction that explicitly accounts for conformational flexibility. As shown by the pink pipeline in Figure 1, the framework first predicts the conformational evolution of pocket-ligand complexes and subsequently integrates the resulting multi-frame conformations and interaction energies into PLBA prediction. In this way, the predicted dynamics provide a direct basis for characterizing binding interactions beyond a single static conformation. To predict pocket-ligand trajectories, we pretrain a motion prediction model to autoregressively predict the next-frame positions of pocket and ligand atoms from their current conformations. The model characterizes ligand motions through rigid displacement, SO(3)-based rotation (Celledoni & Owren, 2003; Zhou et al., 2019), and flexible torsion, while pocket trajectories are predicted under main-chain distance constraints to preserve backbone connectivity. The pretrained model further estimates the interaction energies associated with the predicted conformations, thereby characterizing both conformational and ener-

getic evolution throughout the trajectory. These multi-frame conformations and their corresponding energies are then incorporated into the PLBA prediction model, which assigns energy-aware weights to different conformational states and aggregates them into dynamic representations for affinity prediction. By jointly modeling conformational and energetic evolution, D-PLBA provides a dynamic characterization of protein–ligand interactions for PLBA prediction.

We evaluated D-PLBA on three benchmark datasets to comprehensively assess its ability to exploit conformational dynamics for protein-ligand binding affinity prediction. Extensive experiments demonstrate that D-PLBA substantially outperforms existing state-of-the-art methods across most metrics on all benchmarks, while reducing dynamic multi-conformation inference time to approximately 1% of that required by existing approaches. These gains demonstrate the effectiveness of jointly modeling pocket-ligand conformational flexibility and the temporal evolution of binding states for PLBA prediction. In addition, ablation studies validate the contributions of the proposed components, while trajectory visualizations show that the trajectory prediction module captures binding dynamics and provides informative multi-conformation inputs for PLBA prediction.

In summary, the main contributions of D-PLBA include:

- We propose D-PLBA, which advances PLBA prediction beyond static structural snapshots by explicitly learning from the temporal evolution of pocket-ligand binding conformations.
- We design a plug-and-play module that directly predicts pocket-ligand trajectories, enabling efficient and transferable modeling of localized binding dynamics.
- We incorporate the temporal evolution of interaction energies (IEs) into PLBA prediction to jointly model conformational and energetic dynamics.
- D-PLBA achieves state-of-the-art performance across all benchmarks with substantially improved dynamic multi-conformation inference efficiency.

2 Related work

Deep-learning-based PLBA prediction methods have advanced rapidly (Luo et al., 2017; Wang et al., 2023) and can be broadly categorized according to protein-ligand feature representation into heterogeneous- and homogeneous-feature models. Methods based on heterogeneous features emerged relatively early, largely benefiting from the ready availability of protein sequence and small-molecule data in public databases (Schwaller et al., 2019). With these data, Öztürk et al. (2018) established an early deep-learning framework for protein-ligand interaction prediction using sequence-based representations. Subsequent studies gradually introduced more explicit structural and semantic information into the modeling process. Nguyen et al. (2021) represented ligands as molecular graphs to better characterize their structural properties. Hua et al. (2025a) further incorporated protein representations derived from language models (LMs) (Elnaggar et al., 2021; Wang et al., 2019) to improve generalization, while Wu et al. (2024) introduced residue contact graphs to encode protein topology. More recently, Hua et al. (2025b) integrated LM-derived sequence representations with sequence-inferred structural features to enhance multimodal protein-ligand modeling. Despite these advances, heterogeneous-feature methods encode proteins and ligands in different feature spaces, hindering unified and biophysically consistent interaction modeling and potentially limiting mechanistic interpretability and confidence in drug design applications (Wang et al., 2024).

Homogeneous-feature-based methods directly model protein-ligand complex structures, providing a unified representation that more closely reflects the physical basis of molecular binding (Mastropietro et al., 2023). Early approaches employed 3D CNNs to extract spatial features from protein-ligand complexes (Stepniewska-Dziubinska et al., 2018), but their limited modeling of atomic relationships constrained predictive performance. Jiang et al. (2021) improved upon this by explicitly capturing molecular topology and interatomic dependencies. Building on more detailed interaction modeling, Yan et al. (2024) refined binding-region representations and exploited multiple affinity labels through pretraining. More recent methods, such as EM-PLA (Xie et al., 2025) and PLXF-Pred (Li et al., 2026), further integrate biochemical and sequence information to improve affinity prediction. Despite these advances, most structure-based PLBA methods still rely on experimentally determined or computationally generated static conformations, providing only a snapshot of the binding process. In reality, proteins and ligands continuously interconvert among conformational states, and these dynamics influence molecular recognition, binding affinity, and selectivity through

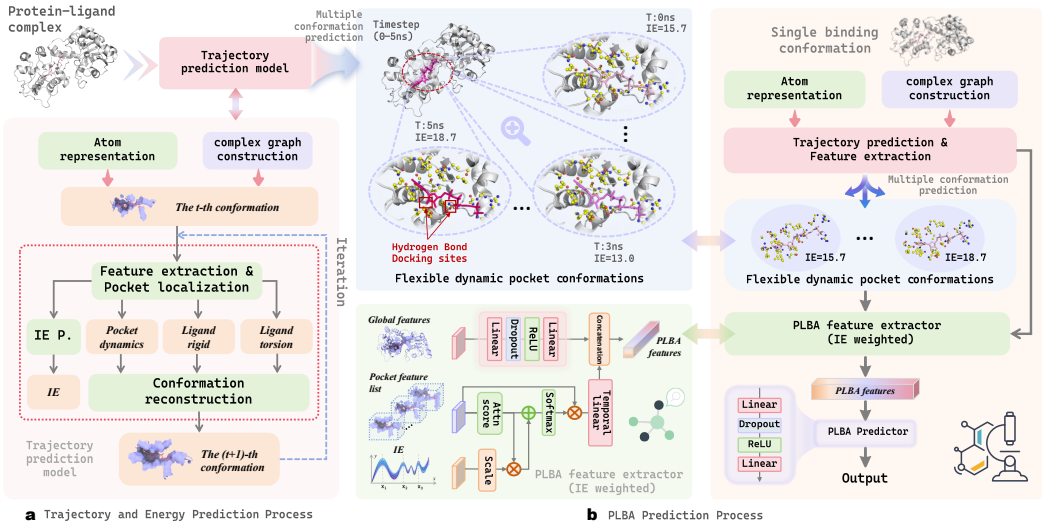

Figure 2: The D-PLBA framework. (a) Prediction process of pocket-ligand motion trajectories and interaction energies. (b) PLBA prediction process.

mechanisms such as conformational selection and induced fit (Karplus & McCammon, 2002). This highlights the importance of incorporating conformational dynamics into PLBA prediction.

3 METHOD

Accordingly, as shown in Figure 2, D-PLBA comprises two modules: dynamic pocket-ligand trajectory and interaction energy prediction, followed by PLBA prediction. The former is pretrained independently and provides dynamic information to the latter. Details are provided below.

3.1 PREDICTION OF THE POCKET-LIGAND TRAJECTORY AND INTERACTION ENERGY

As shown in Figure 2(a), the trajectory prediction model jointly predicts the interaction energy at each frame and the pocket-ligand conformation at the next frame. The protein-ligand complex is represented as an atomic graph, where each atom is encoded by its element type and molecular identity. Protein atoms are additionally associated with their corresponding residues, while ligand atoms are assigned to a shared **MOL** identifier. Each atom is treated as a node and connected to its K nearest neighbors to construct the complex graph, which is formulated as follows.

$$\begin{aligned} &\mathcal{G}_{\mathrm{PL}} = \big(\mathbf{V}, \mathbf{P}, \mathbf{E}\big), \quad \mathbf{V} = \{v_i = \mathcal{C}(t(a_i), m(a_i))\}, \\ &\mathbf{P} = \{p_i\}_{i=1}^{N}, \quad \mathbf{E} = \{(a_i, a_j) \,|\, a_j \in \mathcal{N}_K(a_i)\}, \end{aligned} \tag{1}$$

where $\mathbf{V} \in \mathbb{R}^{N \times d_v}$ denotes the representation of the atoms encoded by the element-type embedding $t(a_i) \in \mathbb{R}^{d_v/2}$ and the affiliation embedding $m(a_i) \in \mathbb{R}^{d_v/2}$, using the concatenation operator $\mathcal{C}(\cdot)$. N is the number of atoms in the complex and d_v is the feature dimension of each atom. $p_i \in \mathbb{R}^3$ is the 3D coordinates of the i-th atom and $\mathcal{N}_K(a_i)$ denotes the set of K nearest neighbors of the i-th atom in the PL complex.

Next, we input the complex representation into the EGNN-based feature extraction block. This block updates atomic features and their corresponding displacement information, each layer of which can be expressed by the following formula:

$$\begin{aligned} \mathbf{v}_i' &= \phi_h\Big(\mathbf{v}_i, \frac{1}{|\mathcal{N}_K(i)|} \sum_{j \in \mathcal{N}_K(i)} \mathbf{m}_{ij}\Big) \\ \mathbf{p}_{res,i} &= s \cdot \sum_{j \in \mathcal{N}_K(i)} (\mathbf{p}_i - \mathbf{p}_j)\, \tanh(\phi_x(\mathbf{m}_{ij})), \\ \mathbf{m}_{ij} &= \phi_e(\mathcal{C}(\mathbf{v}_i, \mathbf{v}_j, \mathbf{e}_{ij}, \|\mathbf{p}_i - \mathbf{p}_j\|)), \end{aligned} \tag{2}$$

where $\mathbf{v}_i, \mathbf{v}'_i \in \mathbb{R}^{d_v}$ denote the feature vector of the i-th atom before and after update, $\mathbf{p}_{res,i}$ denotes the residual displacement of the i-th atom coordinate. $\mathcal{N}_K(i)$ denotes the set of K nearest neighbors of atom i, $\mathbf{e}_{ij}$ denotes the edge vector between atoms i and j, and $\mathbf{m}_{ij}$ denotes the intermediate edge feature. $\phi_h(\cdot), \phi_x(\cdot), \phi_e(\cdot)$ are learnable Multilayer Perceptron (MLP) networks for updating the node feature and coordinate, as well as edge feature, respectively. s is a learnable scaling factor. In this stage, we also dynamically localize the pocket region using the formula below.

$$\mathbf{M}_p(i) = \mathbb{I}\left(a_i^{(p)} \in \bigcup_{a_j^{(l)} \in \mathcal{A}_L} \mathrm{KNN}_M(a_j^{(l)}) \right), \tag{3}$$

where $\mathbb{I}(\cdot)$ is the indicator function, $\mathcal{A}_L = \{a_j^{(l)}\}_{j=1}^{N_l}$ denotes the set of ligand atoms, $\mathrm{KNN}_M(a_j^{(l)})$ represents the set of M nearest protein atoms to the ligand atom $a_j^{(l)}$. $\mathbf{M}_p \in \{0,1\}^{N_p}$ denotes a binary pocket mask over the N_p protein atoms, with entries indicating whether each atom belongs to the binding pocket.

Using the resulting pocket and ligand features, the model predicts pocket-ligand interaction energies (IEs), pocket atomic displacements, ligand rigid-body motion, and ligand torsional changes. The IE predictor is implemented with a GNN and formulated as follows.

$$\mathrm{IE} = \phi(\mathrm{pool}(\mathbf{G})), \quad \mathbf{G} = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{V}'\mathbf{W}), \tag{4}$$

where $\mathbf{V}' = [\,\mathbf{v}'_i\,]_{v_i \in V_L \cup V_{Po}} \in \mathbb{R}^{(N_{Po}+N_L)\times d_v}$ represents the atomic feature matrix that includes all atoms of the ligand and the pocket. $\tilde{\mathbf{A}}$ is the adjacency matrix generated by KNN, $\tilde{\mathbf{D}}$ is the degree matrix, $\mathbf{W}$ is the learnable weights, $\sigma(\cdot)$ is the activation, $\mathrm{pool}(\cdot)$ denotes the global pooling, and $\phi(\cdot)$ denotes an MLP. For each pocket atom, an additional displacement term is predicted from its updated feature through an MLP, $\phi_d(\cdot)$. This term is combined with the residual displacement $\mathbf{p}_{res,i}$ obtained during feature updating to determine the final atomic displacement: $d_i = \phi_d(\mathbf{v}'_i) + \mathbf{p}_{res,i}$. To preserve topology and backbone connectivity, distance constraints are incorporated into the training supervision, as detailed in Appendix A.

We decompose the motion trajectory of the ligand into two components: global rigid-body motion and internal torsional deformation. The rigid-body motion comprises translation and rotation, which are predicted as follows.

$$\begin{aligned} &\mathbf{P}'^{(l)} = \mathbf{R}\mathbf{P}^{(l)} + \boldsymbol{\tau}, \\ &\mathbf{R} = \exp([\boldsymbol{\omega}]_\times) = \mathbf{I} + \frac{\sin\theta}{\theta}[\boldsymbol{\omega}]_\times + \frac{1-\cos\theta}{\theta^2}[\boldsymbol{\omega}]_\times^2, \\ &[\boldsymbol{\omega}, \boldsymbol{\tau}] = \Psi_{\boldsymbol{\phi}}(\mathbf{V}'^{(l)}), \end{aligned} \tag{5}$$

where $\mathbf{P}^{(l)}$ and $\mathbf{P}'^{(l)}$ denote the ligand atomic coordinates in the current and next frames, respectively. $\boldsymbol{\tau} \in \mathbb{R}^3$ is the translational vector and $\boldsymbol{\omega} \in \mathbb{R}^3$ is the rotation vector in the Lie algebra $\mathfrak{so}(3)$ with $\theta = \|\boldsymbol{\omega}\|$ (Zhou et al., 2019). The operator $[\boldsymbol{\omega}]_\times = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}$ is the skew-symmetric matrix representation of $\boldsymbol{\omega}$, and $\mathbf{R} = \exp([\boldsymbol{\omega}]_\times)$ yields a valid rotation matrix on $SO(3)$ through the exponential map. The motion parameters $[\boldsymbol{\omega}, \boldsymbol{\tau}]$ are predicted by an MLP-based predictor $\Psi_{\boldsymbol{\phi}}$, which takes the ligand atom feature $\mathbf{V}'^{(l)}$ as input. To model ligand internal deformation, we consider torsional motion around rotatable bonds. Movable atoms are first identified from the rotatable-bond topology and partitioned into connected torsion groups $\{\mathcal{S}_q\}_{q=1}^{Q}$ using depth-first search (DFS). For each group, predicted atomic displacements are corrected to enforce collective torsional motion around the corresponding rotatable bond.

$$\begin{aligned} &\mathbf{P}'_q = \mathcal{T}_q\left(\mathbf{P}_q, \Delta\mathbf{P}_q\right), \quad \Delta\mathbf{P}_q = \Psi_\phi\left(\mathbf{V}'^{(l)}, \mathcal{S}_q\right), \\ &d_{\min}^{(b)} \le \left\|\mathbf{p}'_i - \mathbf{p}'_j\right\| \le d_{\max}^{(b)}, \qquad \forall (i,j) \in \mathcal{E}_L, \\ &\left\|\mathbf{p}'_i - \mathbf{p}'_j\right\| \ge d_{\min}^{(lp)}, \qquad \forall i \in \mathcal{V}_L,\ \forall j \in \mathcal{V}_P. \end{aligned} \tag{6}$$

where $\mathbf{P}_q$ and $\mathbf{P}'_q$ denote the coordinate matrices of torsion group $\mathcal{S}_q$ before and after the torsional transformation, respectively. $\Delta\mathbf{P}_q$ denotes the predicted atomic displacements for this group, and $\mathcal{T}_q$ applies an inference-time correction to enforce collective torsional motion around the corresponding

rotatable bond while preserving the fragment geometry. $\mathcal{E}_L$, $\mathcal{V}_L$, and $\mathcal{V}_P$ denote the ligand bond-edge, ligand atom, and pocket atom sets, respectively; $d_{\min}^{(b)}$, $d_{\max}^{(b)}$, and $d_{\min}^{(\mathrm{lp})}$ define the bond-length bounds and minimum ligand-pocket distance. After each prediction step, the structural updates are used to reconstruct the subsequent conformation, which is then propagated for interaction energy and motion prediction at the next frame. Iterative repetition of this process yields the dynamic binding trajectory and associated interaction energies for downstream PLBA prediction.

3.2 Prediction of protein-ligand binding affinity

As shown in Figure 2(b), our PLBA prediction model explicitly incorporates dynamic protein-ligand information for more reliable affinity estimation across diverse binding conformations. Specifically, following the same atomic representation used in trajectory prediction, the protein-ligand complex is represented as $\mathcal{G} = \big(\mathbf{V}, \mathbf{P}, \mathbf{E}\big)$ according to atom types and molecular affiliations. The pretrained trajectory and energy prediction model generates pocket conformations and corresponding interaction energies over T frames. For each frame, pocket features are extracted using an EGNN-based encoder following Equation 2. In parallel, global protein and ligand features are also encoded, with the protein encoder adopting the same EGNN formulation as the pocket encoder. The ligand feature extraction is defined as follows.

$$\begin{aligned} \mathbf{h}_v^{(\ell+1)} &= \mathrm{LN}\left(\mathbf{h}_v^{(\ell)} \; + \sum_{u \in \mathcal{N}(v)} \left(\mathbf{m}_{(u,v)}^{(\ell)} + \mathbf{e}_t\right)\right), \\ \mathbf{m}_e^{(\ell)} &= g_e \cdot \phi_{lig}\left(\mathcal{C}(\mathbf{h}_u^{(\ell)}, \mathbf{e}_t)\right), \quad g_e = \sigma(\gamma[t_e]), \quad \mathbf{e}_t = \varphi(t_e), \end{aligned} \tag{7}$$

where $\mathbf{h}_v^{(\ell)} \in \mathbb{R}^{d_h}$ is the atom representation at layer ℓ, $t_e \in \{1, \ldots, N_b\}$ is the bond-type index of the edge $e = (u, v)$. $\mathcal{N}(v)$ denotes the set of bonded neighbors of atom v in molecular graph $\mathcal{G}$, and $\mathrm{LN}(\cdot)$ denotes layer normalization. $\varphi(\cdot)$ and γ are learnable parameters mapping each bond type to a continuous embedding $\mathbf{e}_t$ and a scalar gate $g_e = \sigma(\gamma[t_e])$, respectively. $\phi_{lig}(\cdot)$ is a two-layer MLP network. The protein and ligand features are concatenated into a global complex representation, which is then jointly fed with the multi-frame pocket features into the PLBA feature extractor.

Within the PLBA feature extractor, the predicted IEs serve as attention-like weights to modulate pocket features across conformations according to their energetic contributions, thereby producing more discriminative features for affinity prediction. This process is formulated as follows.

$$\begin{aligned} \mathbf{F} &= \mathcal{C}\left(\phi_{global}(\mathbf{F}_{global}), \phi_{pocket}(\mathbf{A} \cdot \mathcal{F}_{pocket})\right), \\ \mathbf{A} &= \mathrm{softmax}\left(\phi_{atten}(\mathcal{F}_{pocket}) + \zeta \odot \mathbf{E}_{\mathrm{IE}}\right), \end{aligned} \tag{8}$$

Where $\mathbf{F} \in \mathbb{R}^{B \times C}$ denotes the extracted PLBA feature, $\mathbf{F}_{global} \in \mathbb{R}^{B \times C}$ represents the protein-ligand global structure feature, and $\mathcal{F}_{pocket} \in \mathbb{R}^{B \times T \times C}$ corresponds to the pocket features, which are weighted by the attention vector $\mathbf{A} \in \mathbb{R}^{B \times T}$. $\phi_{global}(\cdot)$, $\phi_{pocket}(\cdot)$ and $\phi_{atten}(\cdot)$ are MLP-based networks. In particular, $\phi_{pocket(\cdot)}$ incorporates a linear projection to aggregate the attention-weighted multi-frame features into a unified representation, while $\phi_{atten}(\cdot)$ maps the pocket features to frame-wise attention scores aligned with the predicted energy vector $\mathbf{E}_{\mathrm{IE}} \in \mathbb{R}^{B \times T}$. $\zeta \in \mathbb{R}^T$ is a learnable temporal energy-scaling vector and $\odot$ denotes broadcast element-wise multiplication. Finally, the extracted PLBA feature $\mathbf{F}$ is fed into an MLP-based predictor to predict the protein-ligand binding affinity. Training and more implementation details are provided in Appendix A.

4 Results and Discussion

In this section, we conduct comprehensive experiments to evaluate the effectiveness of D-PLBA by comparing it with state-of-the-art methods and analyzing the results to understand its advantages.

4.1 Comparison with state-of-the-art methods for PLBA prediction

In this part, we evaluate D-PLBA on two tasks: PLBA prediction and pocket-ligand trajectory prediction. For PLBA prediction, we benchmark D-PLBA on PDBbind v2020 (Liu et al., 2015), CSAR-HiQ (Dunbar Jr et al., 2013), and OpenBind (OpenBind Consortium, 2026) against state-of-the-art

Table 1: Comparison of test performance on the PDBbind v2020 Core set and CSAR-HiQ dataset. The best results are highlighted in **bold**, and the second-best results are underlined.

Method	PDBbind				CSAR-HiQ			
	RMSE↓	MAE↓	**R**↑	**SD**↓	RMSE↓	MAE↓	**R**↑	**SD**↓
OnionNet	$1.437_{(0.046)}$	$1.168_{(0.040)}$	$0.753_{(0.021)}$	$1.401_{(0.049)}$	$1.939_{(0.053)}$	$1.533_{(0.030)}$	$0.672_{(0.031)}$	$1.917_{(0.059)}$
GraphDTA	$1.612_{(0.043)}$	$1.218_{(0.046)}$	$0.717_{(0.027)}$	$1.608_{(0.053)}$	$2.130_{(0.051)}$	$1.543_{(0.037)}$	$0.633_{(0.041)}$	$2.019_{(0.054)}$
IGN	$1.372_{(0.045)}$	$1.123_{(0.040)}$	$0.803_{(0.023)}$	$1.417_{(0.047)}$	$1.759_{(0.063)}$	$1.363_{(0.040)}$	$0.732_{(0.043)}$	$1.761_{(0.061)}$
SIGN	$1.306_{(0.039)}$	$1.127_{(0.032)}$	$0.809_{(0.025)}$	$1.382_{(0.046)}$	$1.743_{(0.055)}$	$1.359_{(0.034)}$	$0.727_{(0.039)}$	$1.713_{(0.069)}$
MBP	$1.231_{(0.036)}$	$0.965_{(0.034)}$	$0.815_{(0.018)}$	$1.244_{(0.040)}$	$\underline{1.657}_{(0.060)}$	$1.265_{(0.038)}$	$\underline{0.747}_{(0.026)}$	$\underline{1.641}_{(0.064)}$
MISATO	$1.223_{(0.036)}$	$\underline{0.926}_{(0.024)}$	$0.819_{(0.016)}$	$1.193_{(0.037)}$	$1.683_{(0.062)}$	$1.304_{(0.038)}$	$0.727_{(0.026)}$	$1.686_{(0.065)}$
EM-PLA	$1.230_{(0.031)}$	$0.931_{(0.027)}$	$0.810_{(0.017)}$	$1.229_{(0.039)}$	$1.668_{(0.055)}$	$1.271_{(0.041)}$	$0.739_{(0.033)}$	$1.674_{(0.070)}$
PLXFPred	$\underline{1.218}_{(0.035)}$	$0.940_{(0.031)}$	$\mathbf{0.825}_{(0.020)}$	$\underline{1.187}_{(0.042)}$	$1.662_{(0.053)}$	$\underline{1.263}_{(0.040)}$	$0.743_{(0.027)}$	$1.652_{(0.068)}$
D-PLBA	$\mathbf{1.151}_{(0.026)}$	$\mathbf{0.879}_{(0.012)}$	$\underline{0.824}_{(0.010)}$	$\mathbf{1.148}_{(0.025)}$	$\mathbf{1.601}_{(0.056)}$	$\mathbf{1.204}_{(0.039)}$	$\mathbf{0.766}_{(0.021)}$	$\mathbf{1.573}_{(0.063)}$

methods to assess its predictive performance and generalization. Details of the datasets and evaluation metrics are provided in Appendix B.

In particular, we compare D-PLBA with representative state-of-the-art PLBA prediction methods, including OnionNet (Zheng et al., 2019), GraphDTA (Nguyen et al., 2021), IGN (Jiang et al., 2021), SIGN (Li et al., 2021), MBP (Yan et al., 2024), MISATO (Siebenmorgen et al., 2024), EM-PLA (Xie et al., 2025), and PLXFPred (Li et al., 2026), as shown in Table 1. On the PDBbind v2020 benchmark, D-PLBA achieves the best overall performance across the majority of the evaluated metrics. Compared with the best-performing baseline, PLXFPred, it reduces RMSE from 1.218 to 1.151, a **5.5%** relative reduction, while further reducing MAE by **5.1%** compared with MISATO. The improvement in standard deviation (**SD**), together with a Pearson correlation coefficient (**R**) comparable to the best-performing baseline, further indicates reliable agreement with experimental affinities. These results highlight the value of explicitly modeling dynamic protein-ligand binding conformations for PLBA prediction.

To further assess out-of-distribution generalization, we train D-PLBA exclusively on PDBbind and evaluate it on the structurally diverse CSAR-HiQ benchmark, comprising proteins and ligands that are entirely unseen during training. D-PLBA achieves the best overall performance across all evaluation metrics. Compared with MBP, a strong generalization baseline, D-PLBA reduces RMSE, MAE, and **SD** by **3.4%**, **4.7%**, and **4.1%**, respectively, while improving **R** by **2.5%**, demonstrating its superior accuracy and robustness under out-of-distribution conditions. Notably, several methods that perform competitively on PDBbind v2020 show substantial degradation on CSAR-HiQ, suggesting greater dependence on dataset-specific structural patterns. In contrast, D-PLBA retains strong predictive performance by modeling pocket-ligand conformational dynamics and evolving intermolecular interactions across binding states. Such dynamic information characterizes binding behavior beyond a single static conformation, providing a more transferable physical basis for affinity prediction and contributing to the stronger generalization of D-PLBA.

To evaluate D-PLBA in a realistic virtual screening scenario, we perform PLBA prediction on OpenBind for the enterovirus A71 (EV-A71) 2A protease, where multiple ligands are screened against the same target, and compare D-PLBA with PLXFPred and MBP. As shown in Figure 3(a–c), the experimental affinities are mainly concentrated between 4 and 6. Although the baseline methods achieve comparable RMSE values, their predicted distributions deviate markedly from the experimental distribution. In contrast, D-PLBA leverages pocket conformational changes and interaction-energy evolution across binding states to better distinguish ligands within this narrow affinity range, yielding predictions that more faithfully reproduce the experimental distribution and demonstrating its advantage in single-target, multi-ligand screening. To further verify whether this performance is supported by reliable modeling of binding dynamics, we compare the predicted interaction energies with those obtained from independent MD simulations on unseen OpenBind samples. Their consistent variation trends provide further evidence that D-PLBA captures physically meaningful dynamic information underlying affinity prediction. Detailed results are provided in Appendix C.

Following the evaluation setting adopted in prior work, we also report results for models trained on the general set in Figure 3(d,e). As training data increases, prediction performance generally improves, while D-PLBA consistently achieves the best result among all compared methods.

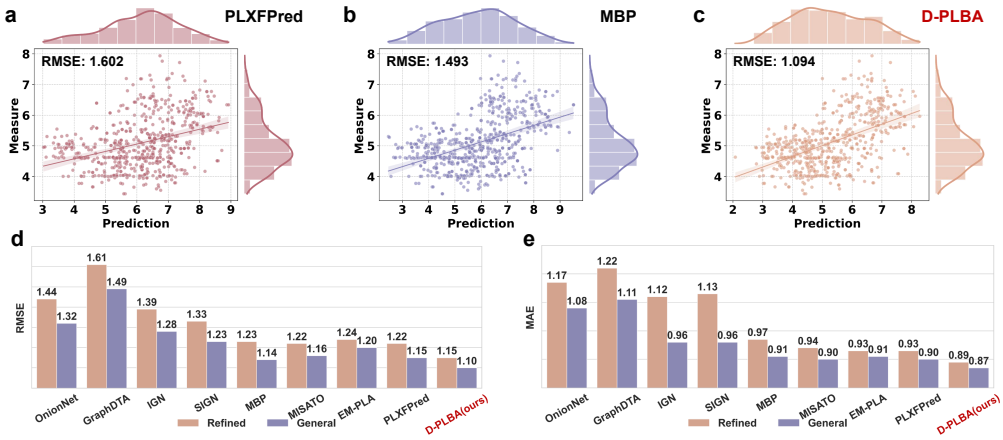

Figure 3: (a–c) Comparison between our method and state-of-the-art approaches on the OpenBind dataset. (d,e) Comparison of PLBA prediction results trained on the general and refined sets.

Table 2: **Evaluation on 1200 test samples of MISATO**. The best results are highlighted in **bold** and the second-best results are underlined.

	Method	RMSD(t)↓	RMSF↑	Clashes↓	RMSD_{IE}↓	Infer T.(h)↓
L. P.	GNNMD	5.212	0.197	5.639	–	–
	NeuralMD-SDE	6.109	0.283	0.087	–	–
	VerletMD	14.677	0.201	0.426	–	–
	DynamicBind	–	0.327	0.072	–	–
	ConfRover	4.142	0.654	0.035	–	–
	ProAR	3.651	0.571	0.023	–	–
	D-PLBA	**2.804**	**0.680**	**0.011**	–	–
PL. P.	MDSimulation	–	–	–	–	>200
	DynamicBind	–	0.298	2.965	–	–
	ConfRover	4.274	0.437	0.431	–	9.30
	ProAR	3.419	**0.467**	0.270	–	6.76
	D-PLBA	**2.814**	0.453	**0.081**	1.986	**0.06**

4.2 Performance comparison and analysis of trajectory prediction

Beyond PLBA prediction, we further benchmark the pretrained model against state-of-the-art methods on dynamic trajectory prediction using the MISATO test set (1200 samples). The baselines include representative learning-based molecular dynamics approaches, such as GNNMD (Siebenmorgen et al., 2024), NeuralMD-SDE (Liu et al., 2025), VerletMD (Liu et al., 2025), DynamicBind (Lu et al., 2024), ConfRover (Shen et al., 2026), ProAR (Cheng et al., 2026), and MDSimulation (Abraham et al., 2015) for predicting molecular dynamics and conformational flexibility. As reported in Table 2, D-PLBA achieves strong performance on both ligand flexibility prediction (L.P.) and pocket-ligand flexibility prediction (PL.P.). In terms of trajectory accuracy and physical plausibility, quantified by RMSD(t) and Clashes, D-PLBA consistently outperforms ConfRover (Shen et al., 2026), ProAR (Cheng et al., 2026), and other competing approaches across both tasks. For RMSF, D-PLBA exhibits the highest RMSF on L.P. and the second-highest RMSF on PL.P., while maintaining low clash scores, suggesting substantial conformational flexibility without severe structural distortions. By focusing on local pocket-ligand dynamics rather than global full-atom motion, D-PLBA maintains high trajectory accuracy while preserving physically meaningful flexibility. Among the methods evaluated under the same hardware and experimental settings, D-PLBA completes dynamic structure inference for 1200 samples in only 0.06 hours. This substantial acceleration facilitates its integration with downstream affinity prediction and enables practical inference. Further analysis of the dynamic structure inference process and all ablation studies on the contribution of individual components to PLBA prediction are provided in Appendix D.

We visualize dynamic trajectories and interaction energy landscapes to further examine the mechanistic advantages of introducing pocket flexibility. As shown in Figure 4, the predicted conformations within the first 5 ns closely follow the corresponding reference trajectories for four randomly selected complexes, consistent with the quantitative evaluation in Figure 6(f). The temporal evolu-

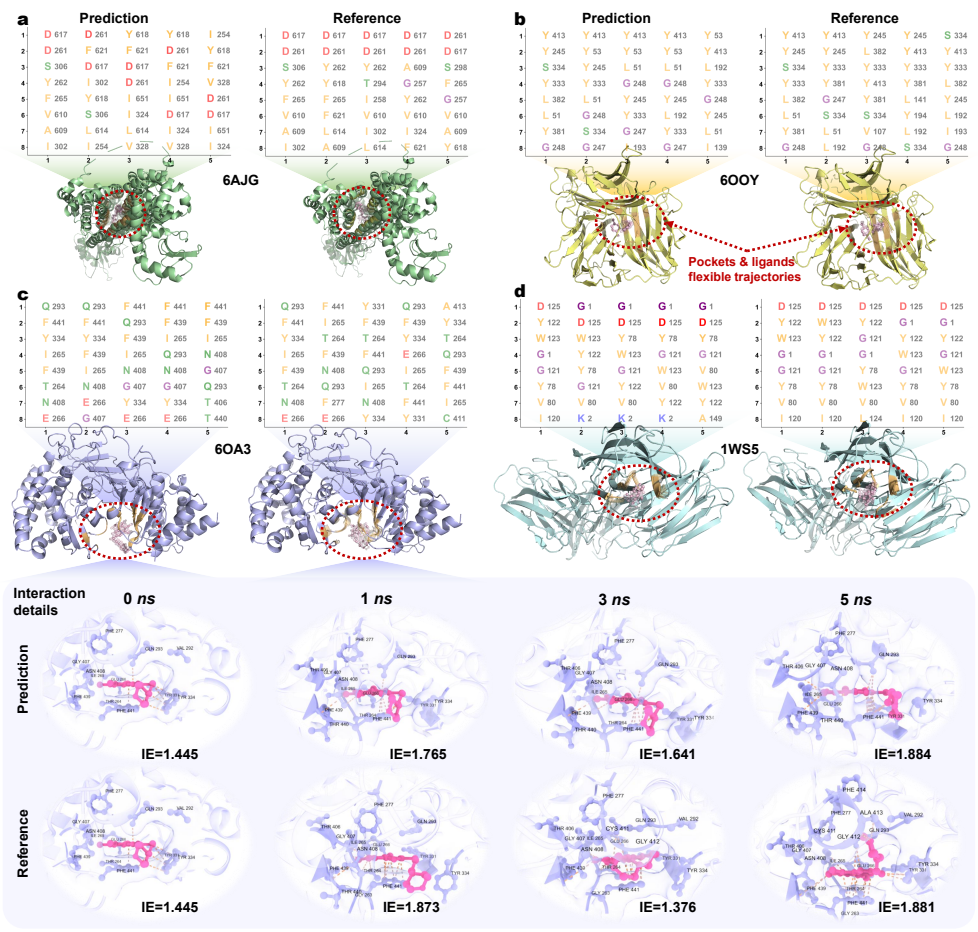

Figure 4: Four randomly selected samples comparing predicted and reference pocket–ligand dynamic trajectories. The enlarged 6OA3 view highlights pocket evolution and interaction energies.

tion of ligand-interacting residues also shows high agreement with the reference trajectories, indicating that the model captures dynamic changes in the local binding environment rather than merely reproducing global ligand motion. For the 6OA3 complex, we further examine the pocket-ligand conformations at a finer structural scale. The predicted structures remain highly consistent with the reference trajectory, with only minor local deviations in ligand geometry, such as slight bond-angle variations associated with thermal fluctuations. Importantly, the overall binding mode and interaction energy evolution are well preserved. These observations provide structural evidence that the predicted conformations within the first 5 ns retain the key interaction patterns and local dynamic changes relevant to PLBA prediction, supporting the performance gains achieved by incorporating them into the dynamic representation.

5 Conclusion

We introduce D-PLBA, a dynamics-aware framework for protein-ligand binding affinity prediction that improves both accuracy and interpretability in structure-based drug discovery. Unlike conventional static-structure approaches, D-PLBA explicitly models the dynamic evolution of pocket-ligand binding conformations through a pretrained trajectory prediction module, capturing protein pocket conformational rearrangements together with ligand translation, rotation, and torsional flexibility across multiple binding states. Spatiotemporal positional states are then combined with interaction energy terms into a unified feature representation, providing a more complete and physically grounded description of protein-ligand interactions. Across multiple benchmark datasets, D-PLBA consistently outperforms state-of-the-art methods in prediction accuracy and reliability. However, prediction errors gradually accumulate over long trajectories, leading to increased RMSD and steric clashes. Future work will focus on improving long-horizon stability to address this limitation.

AI USE STATEMENT

In this work, we used generative AI tools for language-related assistance, including proofreading, grammatical correction, and refinement of English expression to improve the clarity and readability of the manuscript.

We have not used generative AI tools for research idea generation, study design, methodology development, experiment design or execution, code generation, data generation or analysis, figure or table creation, result interpretation, or scientific conclusion development, and the remaining disclosure categories are not applicable to this work.

Additionally, we used generative AI tools to assist with stylistic polishing and language consistency throughout the manuscript. We have reviewed all AI-assisted work. In particular, all AI-assisted revisions were carefully and repeatedly checked by the authors against the original manuscript, experimental results, and intended scientific meaning to ensure that no factual information, technical interpretation, or scientific claim was altered or distorted. The scientific content, experiments, analyses, and substantive writing of the manuscript were completed by the authors, and any AI-assisted wording was further revised by the authors where necessary.

We take responsibility for the final content of this work, including all text, claims, and other materials produced with the aid of generative AI.

ETHICS STATEMENT

This work involves only computational experiments and does not involve human participants, animal experiments, sensitive personal data, or other ethical concerns. We are not aware of any ethical issues associated with this study.

REPRODUCIBILITY STATEMENT

We have made substantial efforts to ensure the reproducibility of all experimental results. Detailed model configurations, training and evaluation procedures, and data-processing settings are provided in the main text and appendices. All source code required to reproduce the reported results is provided through an anonymous repository, together with the necessary instructions and configurations.

AUTHOR CONTRIBUTIONS

Detailed author contributions will be provided after the anonymous review stage.

ACKNOWLEDGMENTS

Acknowledgments, including funding information, will be provided after the anonymous review stage.

REFERENCES

Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

Peizhen Bai, Filip Miljković, Bino John, and Haiping Lu. Interpretable bilinear attention network with domain adaptation improves drug–target prediction. *Nature Machine Intelligence*, 5(2):126–136, 2023.

David D Boehr, Ruth Nussinov, and Peter E Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology*, 5(11):789–796, 2009.

David A Case, H Metin Aktulga, Kellon Belfon, Ido Ben-Shalom, Scott R Brozell, David S Cerutti, Thomas E Cheatham III, Vinícius Wilian D Cruzeiro, Tom A Darden, Robert E Duke, et al. *Amber 2021*. University of California, San Francisco, 2021.

Elena Celledoni and Brynjulf Owren. Lie group methods for rigid body dynamics and time integration on manifolds. *Computer Methods in Applied Mechanics and Engineering*, 192(3-4):421–438, 2003.

Kaiwen Cheng, Yutian Liu, Zhiwei Nie, Mujie Lin, Yanzhen Hou, Yiheng Tao, Chang Liu, Jie Chen, Youdong Mao, and Yonghong Tian. Proar: Probabilistic autoregressive modeling for molecular dynamics. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(1):147–155, 2026.

James B Dunbar Jr, Richard D Smith, Kelly L Damm-Ganamet, Aqeel Ahmed, Emilio Xavier Esposito, James Delproposto, Krishnapriya Chinnaswamy, You-Na Kang, Ginger Kubish, Jason E Gestwicki, et al. Csar data set release 2012: ligands, affinities, complexes, and docking decoys. *Journal of Chemical Information and Modeling*, 53(8):1842–1852, 2013.

Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rehawi, Yu Wang, Llion Jones, Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, et al. Prottrans: toward understanding the language of life through self-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7112–7127, 2021.

Bin Feng, Jiying Zhang, Xinni Zhang, Zijing Liu, and Yu Li. Biomd: All-atom generative model for biomolecular dynamics simulation. *International Conference on Learning Representations (ICLR)*, 2025. URL https://openreview.net/forum?id=LQDeJk6NOr.

Rafael Fernandez-Leiro and Sjors HW Scheres. Unravelling biological macromolecules with cryo-electron microscopy. *Nature*, 537(7620):339–346, 2016.

Ross Girshick. Fast r-cnn. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, 2015.

Paraskevi Gkeka, Gabriel Stoltz, Amir Barati Farimani, Zineb Belkacemi, Michele Ceriotti, John D Chodera, Aaron R Dinner, Andrew L Ferguson, Jean-Bernard Maillet, Hervé Minoux, et al. Machine learning force fields and coarse-grained variables in molecular dynamics: application to materials and biological systems. *Journal of Chemical Theory and Computation*, 16(8):4757–4775, 2020.

Yang Hua, Zhenhua Feng, Xiaoning Song, Xiao-Jun Wu, and Josef Kittler. Mmdg-dti: drug–target interaction prediction via multimodal feature fusion and domain generalization. *Pattern Recognition*, 157:110887, 2025a.

Yang Hua, Tianyang Xu, Xiaoning Song, Zhenhua Feng, Rui Wang, Wenjie Zhang, and Xiaojun Wu. R-dti: Drug target interaction prediction based on second-order relevance exploration. *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 17368–17376, 2025b.

Peter J Huber. Robust estimation of a location parameter. *Breakthroughs in Statistics: Methodology and Distribution*, pp. 492–518, 1992.

Dejun Jiang, Chang-Yu Hsieh, Zhenxing Wu, Yu Kang, Jike Wang, Ercheng Wang, Ben Liao, Chao Shen, Lei Xu, Jian Wu, et al. Interactiongraphnet: A novel and efficient deep graph representation learning framework for accurate protein–ligand interaction predictions. *Journal of Medicinal Chemistry*, 64(24):18209–18232, 2021.

Derek Jones, Hyojin Kim, Xiaohua Zhang, Adam Zemla, Garrett Stevenson, WF Drew Bennett, Daniel Kirshner, Sergio E Wong, Felice C Lightstone, and Jonathan E Allen. Improved protein–ligand binding affinity prediction with structure-based deep fusion inference. *Journal of Chemical Information and Modeling*, 61(4):1583–1592, 2021.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

Martin Karplus and J Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9):646–652, 2002.

Jixiang Li, Ruilin Cai, Ziteng Wang, Ye Sun, Wenge Yang, and Yonghong Hu. Plxfpred: interpretable cross-attention networks with hierarchical fusion of multi-modal features for predicting protein–ligand interactions and affinities. *Bioinformatics*, 42(1):btaf662, 2026.

Shuangli Li, Jingbo Zhou, Tong Xu, Liang Huang, Fan Wang, Haoyi Xiong, Weili Huang, Dejing Dou, and Hui Xiong. Structure-aware interactive graph neural networks for the prediction of protein-ligand binding affinity. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 975–985, 2021.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

Shengchao Liu, Weitao Du, Hannan Xu, Yanjing Li, Zhuoxinran Li, Vignesh Bhethanabotla, Divin Yan, Christian Borgs, Anima Anandkumar, Hongyu Guo, et al. A multi-grained symmetric differential equation model for learning protein-ligand binding dynamics. *Nature Communications*, 17(1):1049, 2025.

Zhihai Liu, Yan Li, Li Han, Jie Li, Jie Liu, Zhixiong Zhao, Wei Nie, Yuchen Liu, and Renxiao Wang. Pdb-wide collection of binding data: current status of the pdbbind database. *Bioinformatics*, 31 (3):405–412, 2015.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *International Conference on Learning Representations (ICLR)*, 2019.

Wei Lu, Jixian Zhang, Weifeng Huang, Ziqiao Zhang, Xiangyu Jia, Zhenyu Wang, Leilei Shi, Chengtao Li, Peter G Wolynes, and Shuangjia Zheng. Dynamicbind: predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nature Communications*, 15(1):1071, 2024.

Zhangli Lu, Guoqiang Song, Huimin Zhu, Chuqi Lei, Xinliang Sun, Kaili Wang, Libo Qin, Yafei Chen, Jing Tang, and Min Li. Dtiam: a unified framework for predicting drug-target interactions, binding affinities and drug mechanisms. *Nature Communications*, 16(1):2548, 2025.

Yunan Luo, Xinbin Zhao, Jingtian Zhou, Jinglin Yang, Yanqing Zhang, Wenhua Kuang, Jian Peng, Ligong Chen, and Jianyang Zeng. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature Communications*, 8(1):573, 2017.

Andrea Mastropietro, Giuseppe Pasculli, and Jürgen Bajorath. Learning characteristics of graph neural networks predicting protein–ligand affinities. *Nature Machine Intelligence*, 5(12):1427–1436, 2023.

David L Mobley and Michael K Gilson. Predicting binding free energies: frontiers and benchmarks. *Annual Review of Biophysics*, 46:531–558, 2017.

Thin Nguyen, Hang Le, Thomas P Quinn, Tri Nguyen, Thuc Duy Le, and Svetha Venkatesh. Graphdta: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, 37(8):1140–1147, 2021.

OpenBind Consortium. Openbind structure–affinity data release: Enterovirus a71 (ev-a71) / coxsackievirus a16 (cva16) 2a protease, 2026. URL `https://doi.org/10.5281/zenodo.20026661`.

Hakime Öztürk, Arzucan Özgür, and Elif Ozkirimli. Deepdta: deep drug–target binding affinity prediction. *Bioinformatics*, 34(17):i821–i829, 2018.

Philippe Schwaller, Teodoro Laino, Théophile Gaudin, Peter Bolgar, Christopher A Hunter, Costas Bekas, and Alpha A Lee. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS central science*, 5(9):1572–1583, 2019.

Pir Masoom Shah, Huimin Zhu, Zhangli Lu, Kaili Wang, Jing Tang, and Min Li. Deepdtagen: a multitask deep learning framework for drug-target affinity prediction and target-aware drugs generation. *Nature Communications*, 16(1):5021, 2025.

Yuning Shen, Lihao Wang, Huizhuo Yuan, Yan Wang, Bangji Yang, and Quanquan Gu. Simultaneous modeling of protein conformation and dynamics via autoregression. *Advances in Neural Information Processing Systems*, 38:109782–109822, 2026.

Till Siebenmorgen, Filipe Menezes, Sabrina Benassou, Erinc Merdivan, Kieran Didi, André Santos Dias Mourão, Radosław Kitel, Pietro Liò, Stefan Kesselheim, Marie Piraud, et al. Misato: machine learning dataset of protein–ligand complexes for structure-based drug discovery. *Nature Computational Science*, 4(5):367–378, 2024.

Marta M Stepniewska-Dziubinska, Piotr Zielenkiewicz, and Pawel Siedlecki. Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. *Bioinformatics*, 34(21):3666–3674, 2018.

Baoli Wang, Chenglin Wang, Jingyang Chen, Danlin Liu, Changzhi Sun, Jie Zhang, Kai Zhang, and Honglin Li. Conditional diffusion with locality-aware modal alignment for generating diverse protein conformational ensembles. *Nature Machine Intelligence*, 8(3):415–434, 2026.

Debby D Wang, Wenhui Wu, and Ran Wang. Structure-based, deep-learning models for protein-ligand binding affinity prediction. *Journal of Cheminformatics*, 16(1):2, 2024.

Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9): 1157–1174, 2004.

Sheng Wang, Yuzhi Guo, Yuhong Wang, Hongmao Sun, and Junzhou Huang. Smiles-bert: large scale unsupervised pre-training for molecular property prediction. *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pp. 429–436, 2019.

Yong Wang, Joao Miguel Martins, and Kresten Lindorff-Larsen. Biomolecular conformational changes and ligand binding: from kinetics to thermodynamics. *Chemical Science*, 8(9):6466–6473, 2017.

Yuxuan Wang, Ying Xia, Junchi Yan, Ye Yuan, Hong-Bin Shen, and Xiaoyong Pan. Zerobind: a protein-specific zero-shot predictor with subgraph matching for drug-target interactions. *Nature Communications*, 14(1):7861, 2023.

Fang Wu and Stan Z Li. Diffmd: A geometric diffusion model for molecular dynamics simulations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(4):5321–5329, 2023.

Lirong Wu, Yufei Huang, Cheng Tan, Zhangyang Gao, Bozhen Hu, Haitao Lin, Zicheng Liu, and Stan Z Li. Psc-cpi: Multi-scale protein sequence-structure contrasting for efficient and generalizable compound-protein interaction prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 310–319, 2024.

K Wüthrich, M Billeter, and W Braun. Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. *Journal of Molecular Biology*, 169(4):949–961, 1983.

Zhiqi Xie, Peng Zhang, Zipeng Fan, Qingpeng Zhang, and Qianxi Lin. Em-pla: environment-aware heterogeneous graph-based multimodal protein–ligand binding affinity prediction. *Bioinformatics*, 41(7):btaf298, 2025.

Jiaxian Yan, Zhaofeng Ye, Ziyi Yang, Chengqiang Lu, Shengyu Zhang, Qi Liu, and Jiezhong Qiu. Multi-task bioassay pre-training for protein-ligand binding affinity prediction. *Briefings in Bioinformatics*, 25(1):bbad451, 2024.

Liangzhen Zheng, Jingrong Fan, and Yuguang Mu. Onionnet: a multiple-layer intermolecular-contact-based convolutional neural network for protein–ligand binding affinity prediction. *ACS Omega*, 4(14):15956–15965, 2019.

Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5738–5746, 2019.

A Appendix: Training and implementation details

The training of D-PLBA proceeds in two successive stages, where the dynamic trajectory prediction model is first trained independently, and its learned parameters are subsequently frozen when training the affinity prediction model. For the first stage, the total loss function is the sum of the conformation prediction losses across multiple time frames, $\mathcal{L}^{\text{total}} = \sum_{t=1}^{T} \mathcal{L}^t$, where the t-th conformation prediction loss $\mathcal{L}^t$ is defined as follows:

$$\begin{aligned}
\mathcal{L}^t &= \underbrace{\mathcal{L}_{\text{poc}} + \lambda_p \cdot \mathcal{P}}_{\text{pocket}} + \underbrace{\mathcal{L}_{\text{lig}}}_{\text{ligand}} + \mathcal{L}_{\text{IE}}, \\
\mathcal{L}_{\text{poc}} + \lambda_p \cdot \mathcal{P} &= \ell\Big(\hat{\mathbf{P}}_p^t, \mathbf{P}_p^{t+1}\Big) + \lambda_p \cdot \mathcal{P}\Big(\hat{\mathbf{P}}_p^t, \mathbf{P}_p^{ini}\Big) \\
\mathcal{L}_{\text{lig}} &= \ell\big(\mathbf{R}\,\mathbf{P}_l^t + \boldsymbol{\tau},\ \mathbf{P}_l^{t+1}\big) + \alpha\,\ell\Big(\boldsymbol{\delta},\ \mathbf{P}_l^{t+1} - \hat{\mathbf{P}}_l^t\Big) \\
\mathcal{L}_{\text{IE}} &= \ell(\hat{a},\ \mathcal{E})
\end{aligned} \tag{9}$$

where the loss function $\mathcal{L}^t$ consists of three main components. The pocket loss ($\mathcal{L}_{\text{poc}} + \lambda_p \cdot \mathcal{P}$)) aligns the predicted protein pocket with the ground-truth structure. Specifically, $\mathcal{L}_{\text{poc}}$ measures pocket geometry accuracy, while $\lambda_p \cdot \mathcal{P}$ penalizes internal residue clashes or chain breaks. $\hat{\mathbf{P}}_p^t$ denotes the predicted pocket coordinates for frame $t+1$ based on frame t, while $\mathbf{P}_p^{t+1}$ denotes the corresponding reference coordinates. The latter enforces distance constraints only on the protein backbone using its initial coordinates $\mathbf{P}_p^{ini}$ as reference, preserving backbone connectivity while allowing conformational flexibility. The ligand loss ($\mathcal{L}_{\text{lig}}$) consists of two terms, the rigid-body alignment loss $\ell\big(\mathbf{R}\,\mathbf{P}_l^t + \boldsymbol{\tau},\ \mathbf{P}_l^{t+1}\big)$ and the torsion loss $\alpha\,\ell\Big(\boldsymbol{\delta},\ \mathbf{P}_l^{t+1} - \hat{\mathbf{P}}_l^t\Big)$, which together ensure the accurate prediction of the ligand's trajectory and internal geometry. $\mathbf{R} = \exp([\boldsymbol{\omega}]_\times)$ represents the rotation matrix derived from the rotation vector $\boldsymbol{\omega}$, as defined in Equation 5. α is a weighting factor and $\boldsymbol{\delta}$ represents the per-atom torsion residual. The interaction energy loss ($\mathcal{L}_{\text{IE}}$) minimizes the Huber loss $\ell(\hat{a}, \mathcal{E})$ (Huber, 1992; Girshick, 2015) between the predicted result $\hat{a}$ and the reference interaction energy $\mathcal{E}$. The loss function $\mathcal{L}^t$ is minimized to optimize the model parameters using the AdamW optimizer (Loshchilov & Hutter, 2019).

The affinity prediction model is optimized via a composite loss combining an MSE-based regression term (ℓ_{reg}) for supervised affinity prediction (p) and constrained objectives on protein-ligand global and dynamic pocket features (p_{glo}, p_{poc}), jointly ensuring affinity-aware semantic alignment and complementary representational capacity. The loss functions are defined as follows:

$$\begin{aligned}
\mathcal{L} &= \ell_{\text{reg}}(p, y) + \lambda \cdot (\ell_{\text{reg}}(p_{glo}, y) + \ell_{\text{reg}}(p_{poc}, y) + g \cdot \ell_{\text{div}}(p_{glo}, p_{poc}, y)) \\
\ell_{\text{div}} &= \left(\frac{(p_{glo} - y)^\top (p_{poc} - y)}{\|p_{glo} - y\|_2 \, \|p_{poc} - y\|_2 + \varepsilon} \right)^2
\end{aligned} \tag{10}$$

where y and λ are the affinity label and the weighting coefficient. g is a gating factor that activates the diversity constraint for non-trivial errors, and ℓ_{div} penalizes aligned residual directions to encourage complementary predictions.

We implemented our method in Python 3.8 using PyTorch 2.4.1. All the experiments were performed on a machine with the Ubuntu 20.04 operating system, Intel(R) Xeon(R) Silver 4214 CPU. An NVIDIA A100 GPU was used to pretrain the trajectory and energy prediction model, while an NVIDIA GeForce RTX 4090 GPU was used for downstream PLBA training. The hyperparameters used for network training are reported in Table 3. For complete implementation details, including the source code, loss functions, and other technical settings, please refer to our open-source repository: https://anonymous.4open.science/r/anonymous-703E.

Table 3: The hyper-parameter setting of D-PLBA.

Trajectory Prediction		PLBA Prediction	
Hyper-parameter	Value	Hyper-parameter	Value
Atom representation d_v	128	d_h	128
Number of intramolecular neighbors K	6	K_{global}	6
Number of inter-molecular neighbors M	10	K_{pocket}	10
Number of conformational states T	-	T	5
Weight coefficients λ_p	0.5	λ	1
Learning rate lr_1	1×10^{-3}	lr_2	1×10^{-4}
Batch size B_1	8	B_2	64
Weight decay W_1	1×10^{-4}	W_2	0.1

B Appendix: Datasets and metrics

We employed distinct datasets for model training and evaluation across different tasks. Specifically, PDBbind (Liu et al., 2015), CSAR-HiQ (Dunbar Jr et al., 2013), and OpenBind (OpenBind Consortium, 2026) were used for protein-ligand binding affinity (PLBA) prediction, whereas the MISATO dataset (Siebenmorgen et al., 2024) was employed for molecular dynamics trajectory prediction. Detailed descriptions are provided below.

PDBbind v2020 (Liu et al., 2015) is a widely adopted benchmark for protein-ligand binding affinity (PLBA) prediction. After preprocessing and deduplication, 4545 complexes from the refined set are used for training, 482 complexes are used for validation, and the 285-complex core set is used for testing. We additionally conduct a larger-scale experiment using 18585 complexes from the general set for training, with the same validation and test sets.

CSAR-HiQ (Dunbar Jr et al., 2013) is a publicly available benchmark comprising experimentally determined 3D protein-ligand complexes with binding affinity annotations. We follow the data index of the established CSAR-NRC HiQ release and retrieve the corresponding complex structures from the latest database records. To ensure an independent evaluation, complexes overlapping with the PDBbind v2020 refined set are removed, resulting in 149 complexes. This filtered dataset is used as an external held-out benchmark to assess the generalization capability of the proposed model.

OpenBind (OpenBind Consortium, 2026) is a structure-affinity paired dataset focused on the Enterovirus A71 (EV-A71) 2A protease, encompassing 925 crystal structures alongside 601 experimentally determined binding affinity measurements. As its affinity measurements are derived directly from drug discovery campaigns against a single biologically relevant target, OpenBind more faithfully reflects real-world screening conditions.

MISATO (Siebenmorgen et al., 2024) is a molecular dynamics dataset containing trajectories for 16972 protein-ligand complexes. Each complex was simulated for 10 ns, with 100 coordinate snapshots retained after discarding the first 2 ns as equilibration, resulting in more than 170 μs of total simulation time. The dataset is split into 13577, 2195, and 1200 complexes for training, validation, and testing, respectively. For trajectory prediction, we sample conformations at 1-ns intervals, such that each autoregressive step corresponds to 1 ns of physical time. The predicted conformation at each step is used as the input for the subsequent step, producing a multi-frame trajectory through recursive prediction.

Since MISATO and PDBbind both contain PDB-derived protein–ligand complexes, we examine their overlap to prevent information leakage from trajectory pretraining to downstream PLBA evaluation. Before trajectory pretraining, we remove MISATO complexes that overlap with any downstream test sample at the complex, protein, or ligand level. Complex-level overlap is identified using PDB identifiers, protein-level overlap using identical protein sequences, and ligand-level overlap using identical canonical SMILES. Consequently, the proteins, ligands, and protein-ligand complexes in the PDBbind Core, CSAR-HiQ, and OpenBind test sets are not observed during trajectory pretraining.

We use different evaluation metrics to validate the model's performance for the PLBA prediction task and the trajectory prediction task. For PLBA prediction, Root Mean Square Error (RMSE),

Mean Absolute Error (MAE), Standard Deviation (SD) (Yan et al., 2024) and Pearson's correlation coefficient (R) are used to evaluate the performance of our model. The formal formulas of them are as follows:

$$
\begin{aligned}
RMSE &= \sqrt{\frac{1}{|D|}\sum_{i=1}^{|D|}(\hat{y}_i - y_i)^2}, \quad MAE = \frac{1}{|D|}\sum_{i=1}^{|D|}|\hat{y}_i - y_i|, \\
SD &= \sqrt{\frac{1}{|D|-1}\sum_{i=1}^{|D|}[y_i - (a + b\hat{y}_i)]^2}, \quad R = \frac{\sum_{i=1}^{|D|}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{|D|}(\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^{|D|}(y_i - \bar{y})^2}}
\end{aligned}
\tag{11}
$$

where a and b are the y-intercept and the slope of the regression line, respectively.

For trajectory prediction, we employ three structural metrics and one energetic metric. RMSF_i characterizes the conformational flexibility of atom i and is interpreted together with structural validity. Clash quantifies atomic overlap violations to assess structural plausibility, while $\mathrm{RMSD}(t)$ measures the coordinate deviation of all atoms between the predicted conformation and the reference MD conformation at time t. In addition, $\mathrm{RMSD}_{\mathrm{IE}}$ measures the deviation of the predicted interaction-energy trajectory from its reference values. Furthermore, we report Inference time (h) to evaluate computational efficiency, as faster inference directly enables more practical deployment in downstream PLBA prediction tasks.

$$
\begin{aligned}
RMSF_i &= \sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_i^p(t) - \bar{\mathbf{x}}_i^p\|^2}, \quad Clash = \sum_{i<j}\max(0, (R_i + R_j - \delta) - d_{ij}), \\
RMSD(t) &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i^p(t) - \mathbf{x}_i^r(t)\|^2}, \quad RMSD_{\mathrm{IE}} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\|E^p(t) - E^r(t)\|^2},
\end{aligned}
\tag{12}
$$

where T denotes the number of trajectory frames, $\mathbf{x}_i^p(t)$ is the predicted coordinate of atom i at frame t, and $\bar{\mathbf{x}}_i^p$ is its time-averaged coordinate over the predicted trajectory. For the clash metric, d_{ij} denotes the Euclidean distance between atoms i and j, R_i and R_j are their van der Waals radii, respectively, and δ is the tolerance threshold. N is the number of atoms; $\mathbf{x}_i^p(t)$ and $\mathbf{x}_i^r(t)$ are the predicted and reference coordinates of atom i at frame t after alignment, respectively; and $E^p(t)$ and $E^r(t)$ are the predicted and reference interaction energies (IEs) at frame t, respectively. $\mathrm{RMSD}_{\mathrm{IE}}$ is computed on the log-transformed interaction-energy scale. The reference interaction-energy labels are derived from protein-ligand binding interaction energies computed using AMBER (Case et al., 2021). Since these energies are predominantly negative and span a wide numerical range, we define the interaction strength as $IE = -IE_{\mathrm{AMBER}}/10$, where IE_{AMBER} denotes the AMBER-computed binding interaction energy. This transformation is intended to preserve the magnitude of interaction-energy variations as much as possible while reducing the optimization burden during model training. Nevertheless, the resulting labels still span a relatively wide range of $(-16.987, 64.835)$, with substantial difference across samples, requiring the model to capture interaction-energy changes across diverse magnitude scales.

C Appendix: Comparison of interaction energy trajectories

As the samples in OpenBind were entirely unseen by the trained model and better reflect realistic conditions, we examined whether the predicted interaction energies (IEs) are consistent with trends measured in molecular dynamics (MD) simulations. To assess this consistency, we use $\mathrm{RMSD} * \mathrm{IE}$ for quantitative evaluation and further visualize nine randomly selected examples in Figure 5(a–i). The resulting $\mathrm{RMSD} * \mathrm{IE}$ indicates a reasonable consistency between the predicted and reference interaction-energy trajectories. Before evaluation, the GROMACS-derived reference energies are converted to the same unit and transformed using the same linear transformation as the training labels. As shown, the predicted IE trajectories generally follow the temporal variation trends of the MD references, providing supporting evidence for the model's ability to capture dynamic

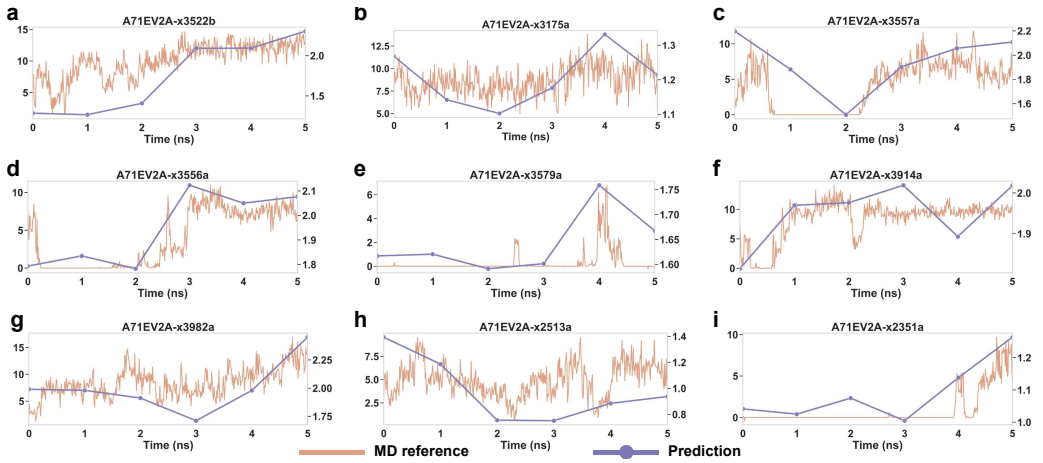

Figure 5: (a–i) Comparison of predicted and reference interaction-energy trajectories for nine randomly selected samples from the OpenBind dataset.

Table 4: Ablation study of D-PLBA on the PDBbind and CSAR-HiQ datasets.

Variant	**PDBbind**			**CSAR-HiQ**		
	RMSE↓	MAE↓	**R**↑	RMSE↓	MAE↓	**R**↑
D-PLBA	$\mathbf{1.151}_{\pm 0.026}$	$\mathbf{0.879}_{\pm 0.012}$	$\mathbf{0.824}_{\pm 0.010}$	$\mathbf{1.601}_{\pm 0.056}$	$\mathbf{1.204}_{\pm 0.039}$	$\mathbf{0.766}_{\pm 0.021}$
w/o FDPC, w/o IEs	$1.400_{\pm 0.011}$	$1.108_{\pm 0.018}$	$0.698_{\pm 0.012}$	$1.793_{\pm 0.090}$	$1.352_{\pm 0.084}$	$0.691_{\pm 0.042}$
w/ FDPC, w/o IEs	$1.225_{\pm 0.029}$	$0.920_{\pm 0.020}$	$0.797_{\pm 0.010}$	$1.628_{\pm 0.054}$	$1.266_{\pm 0.032}$	$0.754_{\pm 0.024}$
w/o FDPC, w/ IEs	$1.345_{\pm 0.028}$	$0.999_{\pm 0.007}$	$0.725_{\pm 0.011}$	$1.620_{\pm 0.034}$	$1.207_{\pm 0.055}$	$0.753_{\pm 0.014}$

interaction-energy changes. Nevertheless, since the training interaction-energy labels and the MD reference trajectories shown in these visualizations are generated using different software packages, namely AMBER (Case et al., 2021) and GROMACS (Abraham et al., 2015), respectively, their absolute IE values exhibit numerical discrepancies. These discrepancies arise from differences in force-field implementation, energy evaluation protocols, and numerical settings between the two simulation pipelines. Therefore, we focus primarily on the consistency of temporal IE variations rather than direct agreement in absolute energy values.

D Appendix: Ablation Study

We conduct ablation studies to evaluate contributions of our core informational components, including flexible dynamic pocket conformation (FDPC) and interaction energies (IEs), while justifying our representation and hyperparameter selections.

As shown in Table 4, both FDPC and IEs are indispensable to D-PLBA, and their integration yields the optimal results across both benchmarks. The full model achieves an RMSE of 1.151 and an **R** of 0.824 on PDBbind, and an RMSE of 1.601 and an **R** of 0.766 on CSAR-HiQ. The ablation of FDPC severely compromises predictive accuracy. On PDBbind, removing FDPC increases the RMSE to 1.400 without IEs and to 1.345 even when IEs are retained, while the corresponding **R** values decrease to 0.698 and 0.725, respectively. A similar trend is observed on CSAR-HiQ, demonstrating the consistent contribution of FDPC across different datasets. These results indicate that static pocket representations are insufficient to characterize ligand-induced conformational rearrangements, and that the resulting loss of dynamic structural information cannot be fully compensated for by energetic features alone. Complementarily, removing IEs while preserving FDPC weakens the model's ability to resolve fine-grained affinity variations among complexes with highly similar pocket geometries. On PDBbind, this omission increases the RMSE from 1.151 to 1.225 and decreases the **R** from 0.824 to 0.797. This performance degradation highlights the importance of incorporating subtle interaction energetics, such as electrostatic and van der Waals interactions,

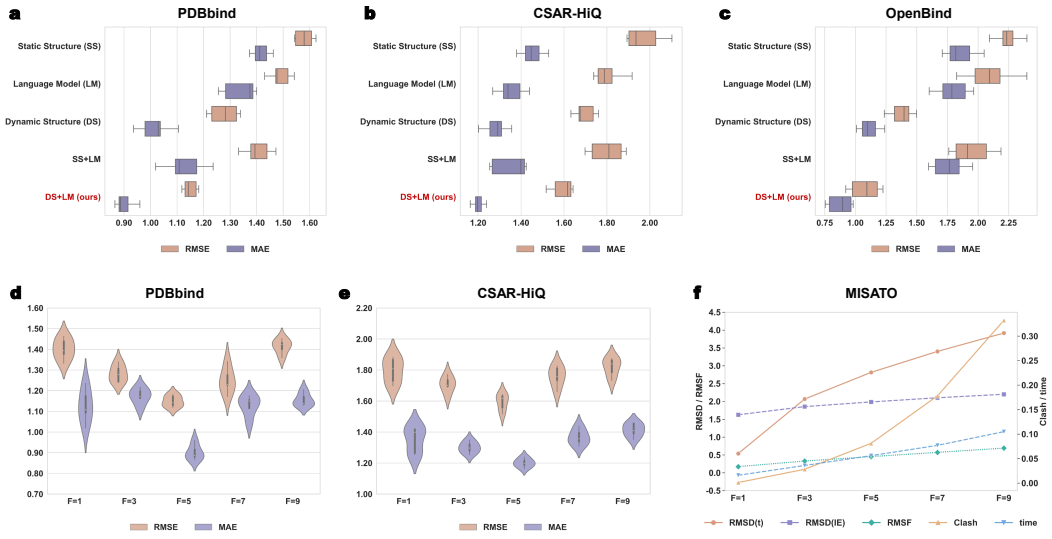

Figure 6: (a–c) Performance comparison of five different representation strategies across three datasets. (d,e) Effect of varying the number of conformations on PLBA performance. (f) Variation of conformational prediction metrics under dynamic conformational modeling as the number of time steps increases.

which are not explicitly captured by geometric features. Collectively, FDPC captures conformational dynamics, whereas IEs characterize the accompanying energy variations. Their complementary modeling underpins the robust and accurate predictions of D-PLBA.

To validate the effectiveness of the representations, we evaluate five strategies, namely static structure (SS), language model (LM), dynamic structure (DS), static structure combined with language model (SS+LM), and dynamic structure combined with language model (DS+LM). The results are presented in Figure 6(a–c). The results reveal that relying solely on static structural representation yields suboptimal performance in both source-dataset accuracy (PDBbind) and cross-dataset generalization (CSAR-HiQ and OpenBind). This underscores that static structures alone are insufficient to sustain robust and generalizable predictions, thereby constraining their practical utility. In contrast, representations generated by the language model (Elnaggar et al., 2021) pretrained on extensive protein sequence corpora are enriched with sample-level and semantic information, leading to relatively stronger performance on the PLBA task. As shown in Figure 6, combining LM-derived and static structural representations consistently improves performance across all three datasets, indicating their complementary contributions to PLBA prediction. Notably, dynamic structural representations alone outperform the combination of LM-derived and static structural representations, underscoring the greater relevance of conformational flexibility and binding interaction energetics to PLBA prediction. Building on this advantage, integrating LM-derived representations with dynamic structural representations (DS+LM) achieves the best overall performance, suggesting that sequence-derived information from language models further enriches the representation of dynamic protein-ligand binding states.

Then, we assessed the effect of using different numbers of conformations in the dynamic representation for PLBA prediction. As shown in Figure 6(d,e), the predictive performance initially improved as the number of sequentially sampled conformations increased, reached a maximum, and then gradually declined. The best performance was achieved using five conformations, corresponding to one conformation sampled per nanosecond. This trend was observed not only on the PDBbind dataset but also consistently generalized to the cross-dataset CSAR-HiQ benchmark. To further characterize the performance of the dynamic conformation prediction model, we examined the variation in multiple evaluation metrics as the number of sequentially sampled conformations increased, with one conformation sampled per nanosecond, as shown in Figure 6(f). As the model recursively predicted subsequent conformations, *RMSF* increased, indicating progressively richer conformational flexibility while the structures remained physically plausible within the first 5 ns. However, prediction

errors also accumulated over time, leading to a gradual increase in *RMSD*. Beyond 5 ns, the predicted conformations exhibited a marked increase in steric clashes, which likely offset the benefits of incorporating additional predicted conformations and explains the performance decline observed beyond the five-conformation setting. Addressing this limitation will be an important direction for future work.