MAKING THE MOST OF MOLECULAR DYNAMICS SIMULATIONS FOR BINDING AFFINITY PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Molecular dynamics (MD) simulations provide rich conformational information for protein–ligand binding affinity prediction, yet existing approaches typically treat simulated conformation samples independently or compress trajectory-level dynamics into static representations, leaving fine-grained interaction dynamics underexploited. In this work, we introduce **DrugDyn**, a unified framework that makes more effective use of molecular dynamics simulations for protein–ligand binding affinity prediction. DrugDyn integrates **Geometry-Aware Spatiotemporal Attention (GASA)** to capture geometry-dependent interactions within individual conformations and their dependencies across conformations, together with **Multi-Cardinality Contrastive Learning (MCCL)** to learn coherent and discriminative complex-level representations from different subsets of the conformational ensemble. Experiments across five backbone architectures under both standard PDBbind core-set evaluation and cross-protein LBA splits demonstrate consistent gains in both predictive accuracy and ranking consistency. On the challenging LBA-30 split, DrugDyn reduces RMSE by 13.8% and improves Kendall's τ by 9.3% on average over the baselines.

1 INTRODUCTION

Protein–ligand binding affinity quantitatively characterizes the interaction between a small-molecule ligand and its biological target protein, providing a fundamental measure for assessing molecular potency and selectivity (Gaulton et al., 2012). Accurate binding affinity estimation is therefore central to virtual screening and lead optimization (Sadybekov & Katritch, 2023; Zhou et al., 2025; Yang et al., 2026). However, experimental determination of binding affinity relies on labor-intensive and time-consuming wet-lab assays, making it impractical to exhaustively evaluate the vast chemical space of potential drug candidates (Hughes et al., 2011; Vuorinen & Schuster, 2015). To alleviate this bottleneck, deep learning (DL)-based approaches have attracted increasing attention for computational binding affinity prediction (Öztürk et al., 2018; Nguyen et al., 2021; Lim et al., 2025; Wu et al., 2026). By learning molecular interaction patterns from existing measurements, these methods enable binding affinity estimation at scale, facilitating the prioritization of ligands with favorable binding potential before costly experimental validation.

Most existing DL-based methods describe a protein–ligand complex using a single static conformation (Shin et al., 2019; Yang et al., 2023; Yin et al., 2024; Kyro et al., 2025), obtained either from experimentally determined structures (Liu et al., 2017) or computational docking (Yin et al., 2024; Huang et al., 2025). However, such a snapshot captures only a limited view of an inherently dynamic binding process. In reality, neither ligands nor proteins are rigid, but instead continuously explore diverse conformational states that may participate in binding (Boehr et al., 2009). Recent studies have begun to recognize this limitation and incorporate molecular dynamics (MD) simulations into binding affinity prediction. One common strategy, adopted by Dynaformer (Min et al., 2024), Densenucy (Libouban et al., 2025), and MDDTA (Zhao et al., 2026), uses conformations sampled from MD trajectories for data augmentation, treating them as additional training instances. However, these conformations are largely regarded as independent samples without explicitly modeling their relationships, and prediction at inference time still relies on a single conformation. Another line of work summarizes MD trajectories through motion-derived graph connectivity (Guo et al., 2025). Specifically, pairwise motion correlations are estimated across conformations, and highly correlated atom pairs are connected even when they are spatially distant. These additional edges enable

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.

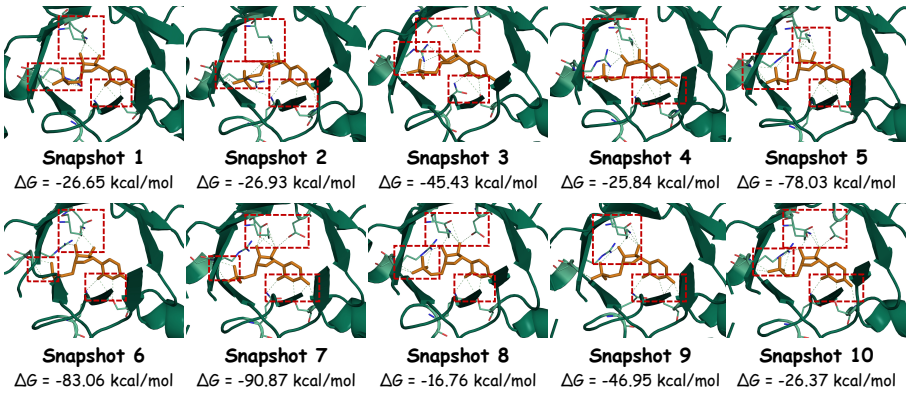


Figure 1: Conformational diversity and energetic variability of the 1LOQ complex. Representative MD conformations of orotidine monophosphate decarboxylase bound to UMP exhibit distinct local interaction geometries and substantially different MM/GBSA binding free energies.

the model to capture regions that tend to move cooperatively beyond conventional distance-based neighborhoods. However, this compresses trajectory-level dynamics into static graph edges while retaining only a single conformation as input. A more direct approach is to explicitly model the sequence of MD conformations. Timenucy and Videonucy (Libouban et al., 2025) encode successive conformations as voxelized 3D grids and employ a long-term recurrent convolutional network and a convolutional LSTM, respectively, for temporal modeling. However, these temporal architectures show no advantage over simpler MD-based data augmentation and can even yield inferior predictive performance. This suggests that simply feeding more conformations into a temporal model does not necessarily lead to more effective use of conformational dynamics. Yet, MD trajectories inherently encode rich dynamic interaction patterns that are absent from a single static snapshot and remain underexploited by existing DL-based models. Consequently, we revisit a fundamental question: how can we make better use of molecular dynamics simulations for binding affinity prediction?

To address this question, we propose **DrugDyn**, a unified framework for protein–ligand binding affinity that jointly exploits conformational dynamics through **Geometry-Aware Spatiotemporal Attention (GASA)** and **Multi-Cardinality Contrastive Learning (MCCL)**. Within each conformation, GASA performs sparse cross-molecular attention over spatially neighboring protein–ligand atom pairs. Specifically, each ligand atom attends only to nearby protein atoms, and vice versa, thereby concentrating representation refinement on the binding interface while substantially reducing computational overhead. The distances within these cross-molecular neighborhoods are further incorporated as geometric biases, enabling atom representations to capture local interaction geometry. Moreover, sparse cross-molecular attention naturally captures conformation-dependent interactions, as variations in protein–ligand geometry across conformations alter local neighborhoods and pairwise distances. GASA subsequently models each atom across conformations along the temporal dimension, capturing dependencies among these varying interaction states rather than treating each conformation independently. In contrast to Timenucy and Videonucy (Libouban et al., 2025), GASA performs atom-level temporal modeling, preserving fine-grained and heterogeneous interaction dynamics that could otherwise be obscured by early conformation-level aggregation. Meanwhile, different conformations can exhibit distinct interaction geometries and energetic profiles, as illustrated in Figure 1, suggesting that they may vary in their relevance to binding. Conformational states are not equally populated, with highly populated states occupying dominant regions of the conformational landscape (Ma et al., 2002). GASA therefore learns an atom-wise importance distribution over conformations and adaptively aggregates their representations, emphasizing informative interaction states while retaining complementary information from the overall trajectory.

Different protein–ligand complexes should remain distinguishable, particularly when their binding affinities differ substantially. However, as shown in Figure 2, different complexes can still be mapped to nearby regions of the latent space despite large differences in binding affinity, which motivates us to introduce a contrastive objective to enhance discriminability. Directly treating individual conformations as different views may be suboptimal, since even conformations from the same complex can correspond to substantially different interaction states, and each individual conformation can be strongly affected by high-frequency thermal noise that obscures functionally relevant

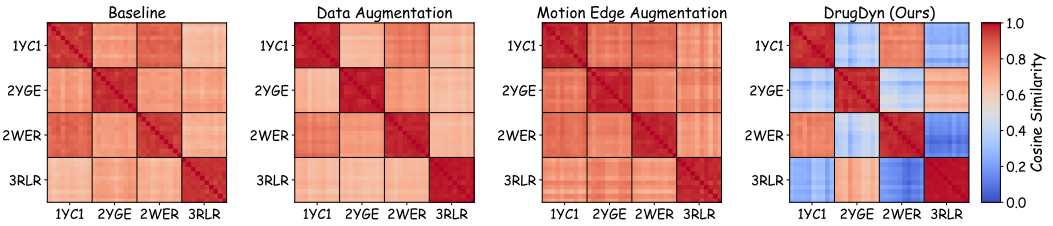


Figure 2: Pairwise similarity matrices of conformational representations. Existing MD strategies yield highly similar representations across complexes (1YC1, 2YGE, 2WER, and 3RLR) with distinct binding affinities of 6.17, 5.06, 7.05, and 7.52, whereas DrugDyn improves separability.

motions (Amadei et al., 1993; Babbitt et al., 2024). Importantly, binding affinity is an ensemble-level property: under equilibrium conditions, a protein–ligand complex explores a distribution of conformational states that collectively constitutes a thermodynamic ensemble (Boehr et al., 2009; Seo et al., 2014). We therefore consider subsets containing multiple conformations as different samples from the same conformational ensemble. Although such subsets can differ in both the number and selection of sampled conformations, they should provide compatible estimates of the shared binding-relevant information of the same complex. Based on this insight, we introduce MCCL, which constructs conformational subsets of different cardinalities as distinct views of the same complex and aligns samples from the same underlying conformational ensemble while separating those from different complexes, thereby promoting robust and discriminative representations.

In summary, our contributions can be listed as follows:

- We propose **GASA** to explicitly model conformational dynamics at the atomic level by capturing geometry-dependent intermolecular interactions within conformations, cross-conformation dependencies, and the varying importance of different conformations.
- We introduce **MCCL** to learn robust and discriminative complex-level representations by aligning conformational subsets of different cardinalities from the same underlying ensemble while separating those from different complexes.
- Extensive experiments across multiple backbone architectures and binding affinity benchmarks demonstrate that DrugDyn consistently outperforms both static baselines and existing MD-based strategies. Notably, on the most challenging LBA-30 split, DrugDyn achieves average relative improvements of 13.8% in RMSE and 9.3% in Kendall's τ over the corresponding static baselines across the five backbones. Compared with the strongest competing MD-based results, it further improves RMSE by 5.2% and Kendall's τ by 5.8%.

2 PRELIMINARIES

Given a protein–ligand complex, our goal is to predict its experimentally measured binding affinity from a set of conformations sampled from an MD trajectory. The MD trajectories are obtained from MISATO (Siebenmorgen et al., 2024), where each bound protein–ligand complex is refined using semi-empirical quantum mechanics and subsequently simulated for 10 ns with Amber20. Following the graph construction scheme of DTIGN (Yin et al., 2024), we represent the complex at the t-th conformation as a graph

$$\mathcal{G}^{(t)} = \left(\mathcal{V}, \mathcal{E}^{(t)}, \mathcal{R}^{(t)}\right) = \left(\mathcal{V}_l \cup \mathcal{V}_p, \mathcal{E}_l \cup \mathcal{E}_p \cup \mathcal{E}_{lp}^{(t)}, \mathcal{R}_l^{(t)} \cup \mathcal{R}_p^{(t)}\right), \tag{1}$$

where $\mathcal{V}_l$ and $\mathcal{V}_p$ denote the ligand and protein atom sets, respectively. Each atom $v_i \in \mathcal{V}$ is characterized by an initial feature vector $\mathbf{x}_i \in \mathbb{R}^{d_x}$ encoding its atomic properties, and a 3D coordinate $\mathbf{r}_i^{(t)} \in \mathbb{R}^3$ specifying its spatial position at conformation t. $\mathcal{E}_l$ and $\mathcal{E}_p$ describe the intramolecular connectivity within the ligand and protein, respectively, whereas $\mathcal{E}_{lp}^{(t)}$ contains intermolecular edges between spatially neighboring ligand and protein atoms. Specifically, an intermolecular edge is established when the corresponding atom pair lies within a distance cutoff δ:

$$\mathcal{E}_{lp}^{(t)} = \left\{(i,j) \mid i \in \mathcal{V}_l, j \in \mathcal{V}_p, \left\|\mathbf{r}_i^{(t)} - \mathbf{r}_j^{(t)}\right\|_2 < \delta\right\}, \tag{2}$$

where δ is set to 5 Å. While the atom identities and intramolecular topology remain fixed throughout the trajectory, the atomic coordinates $\mathcal{R}^{(t)}$ vary across conformations, leading to corresponding changes in the intermolecular edge set $\mathcal{E}_{lp}^{(t)}$.

Following the union-pocket strategy of LigoSpace (Zhou et al., 2025), we construct a shared protein pocket for each complex by taking the union of all pocket atoms identified across the sampled conformations. The resulting protein atom set $\mathcal{V}_p$ remains fixed throughout the trajectory. Together with the invariant ligand atom set $\mathcal{V}_l$, this ensures a consistent atom set across all conformations:

$$\mathcal{V}^{(1)} = \mathcal{V}^{(2)} = \cdots = \mathcal{V}^{(T)} = \mathcal{V}, \tag{3}$$

while the coordinates $\{\mathcal{R}^{(t)}\}_{t=1}^{T}$ remain conformation-specific. This construction establishes one-to-one atom correspondence across MD frames, enabling consistent atom-wise tracking across conformations. Given T sampled conformations $\{\mathcal{G}^{(t)}\}_{t=1}^{T}$, the corresponding binding affinity is denoted by $y \in \mathbb{R}$. Our objective is to learn a predictive function

$$\hat{y} = f_\theta\left(\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(T)}\right) \tag{4}$$

that integrates multiple simulated conformation samples to accurately estimate y.

3 METHOD

DrugDyn integrates **Geometry-Aware Spatiotemporal Attention (GASA)** for atom-level spatiotemporal modeling and **Multi-Cardinality Contrastive Learning (MCCL)** for robust and discriminative complex-level representation learning. The overall architecture is illustrated in Figure 3.

3.1 GEOMETRY-AWARE SPATIOTEMPORAL ATTENTION

Backbone Encoding. GASA is designed to be backbone-agnostic and can be integrated with different structure-based encoders. For each conformation $\mathcal{G}^{(t)}$, a backbone encoder f_{enc} independently extracts atom-level representations:

$$\mathbf{H}^{(t)} = f_{\text{enc}}\left(\mathcal{G}^{(t)}\right) = \left[\mathbf{h}_1^{(t)}, \ldots, \mathbf{h}_N^{(t)}\right] \in \mathbb{R}^{N \times d}, \tag{5}$$

where $N = |\mathcal{V}|$ denotes the number of atoms, which remains consistent across conformations by construction, and d is the hidden dimension. These atom-level representations serve as the initial inputs to GASA for subsequent spatial and temporal interaction modeling.

Geometry-Aware Cross-Molecular Attention. As shown in the lower-left panel of Figure 3, for each conformation, we refine atom representations through sparse cross-molecular attention over intermolecular edges. Since direct protein–ligand interactions are primarily localized at the binding interface, we explicitly restrict information exchange to spatially neighboring protein–ligand atom pairs. Specifically, for atom i, we define

$$\mathcal{N}_i^{(t)} = \left\{ j \mid (i,j) \in \mathcal{E}_{lp}^{(t)} \right\}, \tag{6}$$

such that each ligand atom attends only to spatially neighboring protein atoms, and vice versa. This design focuses representation refinement on binding-relevant intermolecular interactions while avoiding unnecessary computations over spatially distant atom pairs. For attention head h, the query, key, and value representations are computed as

$$\mathbf{q}_{i,h}^{(t)} = \mathbf{W}_h^Q \mathbf{h}_i^{(t)}, \quad \mathbf{k}_{j,h}^{(t)} = \mathbf{W}_h^K \mathbf{h}_j^{(t)}, \quad \mathbf{v}_{j,h}^{(t)} = \mathbf{W}_h^V \mathbf{h}_j^{(t)}. \tag{7}$$

Meanwhile, characterizing protein–ligand interactions requires accounting for the geometry of interacting atom pairs, as the formation and strength of noncovalent interactions are highly dependent on interatomic separation (Liu et al., 2008). We therefore explicitly incorporate the pairwise distance

$$d_{ij}^{(t)} = \left\| \mathbf{r}_i^{(t)} - \mathbf{r}_j^{(t)} \right\|_2 \tag{8}$$

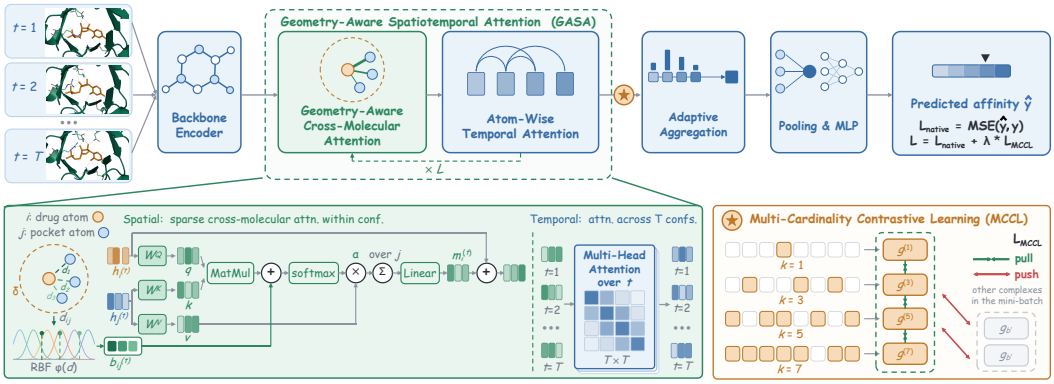


Figure 3: Overview of DrugDyn.

into the attention mechanism. Rather than using the raw scalar distance directly, we expand it using M radial basis functions (RBFs) to obtain a smooth and expressive encoding of distance:

$$\phi_m\left(d_{ij}^{(t)}\right) = \exp\left[-\left(\frac{d_{ij}^{(t)} - \mu_m}{\sigma_m}\right)^2\right], \qquad m = 1, \ldots, M, \tag{9}$$

where μ_m and σ_m denote the center and width of the m-th basis function, respectively, and are fixed according to a predefined distance range. The resulting distance embedding $\boldsymbol{\phi}(d_{ij}^{(t)}) \in \mathbb{R}^M$ is projected into a head-specific geometric bias:

$$b_{ij,h}^{(t)} = \mathrm{Linear}_{\mathrm{geo}}\left(\boldsymbol{\phi}\left(d_{ij}^{(t)}\right)\right). \tag{10}$$

The geometric bias is then incorporated into the scaled dot-product attention:

$$e_{ij,h}^{(t)} = \frac{\left(\mathbf{q}_{i,h}^{(t)}\right)^\top \mathbf{k}_{j,h}^{(t)}}{\sqrt{d_h}} + b_{ij,h}^{(t)}, \qquad \alpha_{ij,h}^{(t)} = \mathrm{softmax}_{j \in \mathcal{N}_i^{(t)}}\left(e_{ij,h}^{(t)}\right), \tag{11}$$

where d_h denotes the dimensionality of each attention head. The geometric bias adjusts the attention weights according to the spatial configuration of each protein–ligand atom pair, allowing atom pairs with similar feature compatibility but different distances to contribute differently. The cross-molecular message for atom i is then aggregated across attention heads as

$$\mathbf{m}_i^{(t)} = \mathrm{Linear}\left(\underset{h}{\mathrm{Concat}}\left[\sum_{j \in \mathcal{N}_i^{(t)}} \alpha_{ij,h}^{(t)} \mathbf{v}_{j,h}^{(t)}\right]\right). \tag{12}$$

where the linear projection maps the concatenated multi-head representation back to the hidden dimension d. Finally, we update the atom representation through a residual connection He et al. (2016) and layer normalization (LN) (Ba et al., 2016):

$$\widetilde{\mathbf{h}}_i^{(t)} = \mathrm{LN}\left(\mathbf{h}_i^{(t)} + \mathrm{Dropout}\left(\mathbf{m}_i^{(t)}\right)\right). \tag{13}$$

As protein–ligand relative positions vary across conformations, both cross-molecular neighborhoods and pairwise distances change accordingly. Consequently, sparse cross-molecular attention produces conformation-dependent atom representations that reflect distinct local intermolecular interaction states. The resulting representations $\{\widetilde{\mathbf{h}}_i^{(t)}\}_{t=1}^T$ are subsequently modeled across conformations along the temporal dimension.

Atom-Wise Temporal Attention. As illustrated in Figure 1, different conformations of the same protein–ligand complex can exhibit substantially different local interaction geometries and energetic profiles. Independently encoding each conformation therefore captures only its instantaneous interaction state, without explicitly modeling how the interaction environment of the same atom

varies across conformations. To address this limitation, GASA tracks each atom across conformations and performs temporal modeling over its conformation-dependent representations. In contrast to Timenucy and Videonucy (Libouban et al., 2025), GASA performs temporal modeling at the atom level, preserving fine-grained and heterogeneous cross-conformation interaction dynamics that could otherwise be obscured by early conformation-level aggregation.

For atom i, we collect its geometry-aware representations over T conformations as

$$\widetilde{\mathbf{H}}_i = \left[\widetilde{\mathbf{h}}_i^{(1)}, \widetilde{\mathbf{h}}_i^{(2)}, \ldots, \widetilde{\mathbf{h}}_i^{(T)}\right] \in \mathbb{R}^{T\times d}. \tag{14}$$

We then apply multi-head self-attention (MHA) along the conformational dimension for each atom independently, followed by a residual connection and layer normalization:

$$\mathbf{A}_i = \mathrm{MHA}\left(\widetilde{\mathbf{H}}_i, \widetilde{\mathbf{H}}_i, \widetilde{\mathbf{H}}_i\right) \in \mathbb{R}^{T\times d}, \qquad \widehat{\mathbf{H}}_i = \mathrm{LN}\left(\widetilde{\mathbf{H}}_i + \mathbf{A}_i\right) \in \mathbb{R}^{T\times d}. \tag{15}$$

This atom-wise attention allows each conformational state to be contextualized by the other states of the same atom, thereby capturing dependencies among the varying local intermolecular environments encountered across the MD trajectory. Specifically, we employ multiple GASA blocks that alternate spatial cross-molecular attention and atom-wise temporal attention to jointly capture local interactions and cross-conformation dependencies.

Adaptive Aggregation. Furthermore, we perform adaptive aggregation by learning an atom-specific conformational importance distribution over the temporally contextualized representations. Let $\widehat{\mathbf{h}}_i^{(t)}$ denote the representation of atom i at conformation t after spatiotemporal modeling. We compute its importance score through an MLP and normalize it over the T conformations as

$$s_i^{(t)} = \mathrm{MLP}\left(\widehat{\mathbf{h}}_i^{(t)}\right), \qquad \beta_i^{(t)} = \frac{\exp\left(s_i^{(t)}\right)}{\sum_{\tau=1}^{T}\exp\left(s_i^{(\tau)}\right)}. \tag{16}$$

The representation of atom i is then aggregated according to the learned importance weights, after which they are pooled into a complex-level embedding for binding affinity prediction:

$$\mathbf{z}_i = \sum_{t=1}^{T}\beta_i^{(t)}\widehat{\mathbf{h}}_i^{(t)}, \qquad \mathbf{z}_{\mathcal{G}} = \sum_{i=1}^{N}\mathbf{z}_i, \qquad \hat{y} = \mathrm{MLP}\left(\mathbf{z}_{\mathcal{G}}\right). \tag{17}$$

The corresponding binding affinity regression objective is defined using mean squared error (MSE):

$$\mathcal{L}_{\mathrm{native}} = \mathrm{MSE}(\hat{y}, y), \tag{18}$$

where $\hat{y}$ and y denote the predicted and experimental binding affinities, respectively.

3.2 Multi-Cardinality Contrastive Learning

Ideally, conformations sampled from complexes with different binding affinities should be distinguishable. However, as shown in Figure 2, existing MD-based strategies can map conformations from complexes with substantially different binding affinities to highly similar latent representations. This motivates us to impose a contrastive objective on the representation space to enhance separation between different protein–ligand complexes. A further question is how to define the views. Directly treating individual conformations as separate views may be suboptimal, since they can exhibit substantially different interaction states while also being affected by high-frequency thermal noise. Under equilibrium conditions, a protein–ligand complex explores a distribution of conformational states, and any subset of MD conformations provides only a partial observation of this underlying ensemble. Therefore, the representation of the complex should remain coherent across different numbers of sampled conformations. In light of this, we introduce **Multi-Cardinality Contrastive Learning (MCCL)**, which constructs multiple views of each complex from conformational subsets of different cardinalities, as illustrated in the lower-right panel of Figure 3.

For the t-th conformation, we first pool the atom representations after spatiotemporal modeling to obtain a conformation-level representation:

$$\mathbf{c}^{(t)} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{h}}_i^{(t)} \in \mathbb{R}^d. \tag{19}$$

For each complex, we then sample conformational subsets with cardinalities $\mathcal{K} = \{1, 3, 5, 7\}$. Given a subset $\mathcal{S}^{(k)} \subseteq \{1, \ldots, T\}$ containing k conformations, we obtain its k-cardinality representation by averaging the corresponding conformation embeddings:

$$\mathbf{g}^{(k)} = \frac{1}{k} \sum_{t \in \mathcal{S}^{(k)}} \mathbf{c}^{(t)}, \qquad \left|\mathcal{S}^{(k)}\right| = k. \tag{20}$$

Different cardinalities thus provide different observations of the same conformational ensemble at varying sampling scales. We treat representations constructed from different cardinalities of the same complex as positives, while those from different complexes serve as negatives. For each complex b, $\mathbf{g}_b^{(k)}$ serves as an anchor, while the representations $\{\mathbf{g}_b^{(k')} \mid k' \in \mathcal{K}, k' \neq k\}$ constitute its positive set. Accordingly, we optimize the following contrastive objective:

$$\mathcal{L}_{\mathrm{MCCL}} = -\frac{1}{B|\mathcal{K}|} \sum_{b=1}^{B} \sum_{k \in \mathcal{K}} \frac{1}{|\mathcal{K}| - 1} \sum_{k' \in \mathcal{K},\, k' \neq k} \log \frac{\exp\left(\mathrm{sim}\left(\mathbf{g}_b^{(k)}, \mathbf{g}_b^{(k')}\right)/\gamma\right)}{\sum_{b' \neq b,\, k'' \in \mathcal{K}} \exp\left(\mathrm{sim}\left(\mathbf{g}_b^{(k)}, \mathbf{g}_{b'}^{(k'')}\right)/\gamma\right)}. \tag{21}$$

where B denotes the batch size, γ is the temperature parameter, and $\mathrm{sim}(\cdot, \cdot)$ denotes cosine similarity. By aligning different-cardinality observations of the same complex while separating representations across complexes, MCCL encourages the representation of a protein–ligand complex to remain coherent across different observations of its conformational ensemble while preserving complex-level discriminability.

Training Objective. We train the entire framework end-to-end by jointly optimizing the binding affinity regression loss and the MCCL objective:

$$\mathcal{L} = \mathcal{L}_{\mathrm{native}} + \lambda \mathcal{L}_{\mathrm{MCCL}}, \tag{22}$$

where λ balances the contrastive regularization.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate DrugDyn on the widely used PDBbind dataset (Liu et al., 2017; Li et al., 2014), with MD conformations obtained from MISATO (Siebenmorgen et al., 2024). Following MDDTA (Zhao et al., 2026), we select 10 representative conformations from the 100 conformations provided for each complex to reduce computational cost.

Evaluation Protocols. We consider two evaluation settings. First, following the protocol of existing methods (Li et al., 2021; Yang et al., 2023), we train models on the PDBbind v2019 general set and evaluate them on the PDBbind v2013 and v2016 core sets, providing a standard benchmark for predictive performance. We further assess model generalization to diverse proteins using the ligand-binding affinity (LBA) splits from Atom3D (Townshend et al., 2020). Specifically, the PDBbind v2019 refined set is partitioned according to protein sequence-identity thresholds of 30% and 60%, referred to as LBA-30 and LBA-60, respectively. By reducing sequence similarity between training and test proteins, these splits provide a more challenging evaluation of cross-protein generalization.

Baselines. We adopt five representative structure-based models, EGNN (Satorras et al., 2021), GIGN (Yang et al., 2023), DTIGN (Yin et al., 2024), DeepRLI (Lin et al., 2025), and BA-Pred (Sim & Lee, 2026) as backbones. For each backbone, we compare DrugDyn against two representative MD-based strategies: MD-based data augmentation (DA) and motion-derived edge augmentation (MDEA). This setup enables a fair comparison of different strategies for exploiting conformational dynamics while assessing the general effectiveness of DrugDyn across backbone architectures.

Metrics. We evaluate each model in terms of both scoring and ranking performance. Scoring performance is assessed using root mean square error (RMSE), mean absolute error (MAE), and Pearson correlation coefficient (R), which quantify predictive accuracy and linear agreement between predicted and experimental binding affinities, while ranking performance is measured by Spearman's rank correlation coefficient (ρ) and Kendall's rank correlation coefficient (τ), which assess the consistency between predicted and experimental binding affinity rankings.

Table 1: Performance comparison on the standard PDBbind benchmarks.

Backbone	Method	PDBbind v2013 core set					PDBbind v2016 core set				
		RMSE↓	MAE↓	R↑	ρ↑	τ↑	RMSE↓	MAE↓	R↑	ρ↑	τ↑
EGNN	Baseline	1.963	1.513	0.580	0.627	0.443	2.054	1.587	0.550	0.585	0.414
	DA	1.713	1.357	0.663	0.681	0.492	1.780	1.392	0.615	0.645	0.465
	MDEA	1.766	1.367	0.662	0.675	0.486	1.913	1.487	0.607	0.613	0.438
	DrugDyn	**1.585**	**1.266**	**0.724**	**0.730**	**0.526**	**1.702**	**1.364**	**0.668**	**0.662**	**0.472**
GIGN	Baseline	1.760	1.376	0.646	0.647	0.464	1.572	1.209	0.699	0.686	0.506
	DA	1.665	1.306	0.687	0.698	0.505	1.415	1.098	0.760	0.747	0.555
	MDEA	1.601	1.325	0.719	0.711	0.519	1.447	1.161	0.749	0.730	0.540
	DrugDyn	**1.458**	**1.133**	**0.777**	**0.805**	**0.605**	**1.330**	**1.040**	**0.798**	**0.803**	**0.604**
DTIGN	Baseline	1.626	1.348	0.706	0.693	0.507	1.523	1.231	0.723	0.705	0.515
	DA	1.603	1.294	0.713	0.723	0.524	1.430	1.105	0.755	0.753	0.560
	MDEA	1.596	1.269	0.716	0.707	0.515	1.481	1.170	0.735	0.716	0.527
	DrugDyn	**1.535**	**1.223**	**0.751**	**0.753**	**0.559**	**1.330**	**1.070**	**0.802**	**0.790**	**0.597**
DeepRLI	Baseline	1.431	1.134	0.772	0.762	0.568	1.467	1.127	0.756	0.780	0.581
	DA	1.345	1.059	0.791	0.782	0.589	1.260	0.898	0.821	0.825	0.646
	MDEA	1.477	1.190	0.751	0.739	0.547	1.526	1.202	0.734	0.757	0.561
	DrugDyn	**1.287**	**1.032**	**0.812**	**0.799**	**0.606**	**1.163**	**0.894**	**0.859**	**0.864**	**0.675**
BA-Pred	Baseline	1.430	1.126	0.775	0.756	0.572	1.551	1.228	0.719	0.733	0.551
	DA	1.429	1.173	0.773	0.754	0.564	1.394	1.102	0.787	0.775	0.582
	MDEA	1.461	1.192	0.784	0.765	0.572	1.565	1.247	0.720	0.714	0.521
	DrugDyn	**1.327**	**1.070**	**0.812**	**0.804**	**0.608**	**1.375**	**1.076**	**0.797**	**0.801**	**0.604**

4.2 Main Results

Standard PDBbind Benchmark. DrugDyn consistently achieves the best performance across all five backbones on both the v2013 and v2016 core sets, as shown in Table 1. It outperforms the corresponding static baseline, DA, and MDEA on both scoring and ranking metrics, demonstrating improvements in predictive accuracy and ranking consistency. Specifically, DrugDyn achieves an average relative RMSE reduction of 13.7% over the corresponding static baselines and 5.8% over the strongest competing MD-based strategies. For example, with GIGN on the v2013 core set, it reduces RMSE from 1.601 to 1.458 relative to the best MD-based baseline, corresponding to an 8.9% improvement, while increasing R from 0.719 to 0.777. Similarly, with DTIGN on the v2016 core set, DrugDyn lowers RMSE from 1.430 to 1.330 and improves R from 0.755 to 0.802. The improvements are also evident in ranking performance, with ρ increasing from 0.765 to 0.804 on BA-Pred with the v2013 core set. These consistent gains demonstrate the effectiveness of DrugDyn across diverse model architectures.

Diverse Protein Evaluation. We further evaluate generalization under the more challenging LBA-30 and LBA-60 settings, with results summarized in Table 2. DrugDyn again delivers stronger overall performance than the corresponding static baselines and alternative MD-based strategies across all five backbones. Across the five backbones and two LBA splits, DrugDyn reduces RMSE by an average of 15.1% relative to the corresponding static baselines. On LBA-30, for example, DrugDyn reduces RMSE from 1.603 to 1.498 with GIGN and from 1.762 to 1.612 with BA-Pred, corresponding to improvements of 6.6% and 8.5% over the strongest competing MD-based strategies, respectively. The gains become particularly pronounced with DTIGN on LBA-60, where RMSE decreases from 1.521 to 1.395 and R increases from 0.682 to 0.734. Likewise, with GIGN on LBA-60, DrugDyn achieves an RMSE of 1.335 compared with 1.405 for DA, together with improvements in R, ρ, and τ. These results demonstrate that DrugDyn better exploits MD trajectories and generalizes more effectively across diverse protein targets.

4.3 Ablation Studies

To investigate the contribution of each component, we conduct ablation studies across multiple backbone architectures on the PDBbind v2016 core set. Starting from each backbone baseline, we independently incorporate Geometry-Aware Spatiotemporal Attention (GASA) and Multi-Cardinality Contrastive Learning (MCCL). We further compare with Single-Conformation Contrastive Learning

Table 2: Performance comparison on the challenging LBA-30 and LBA-60 splits.

Backbone	Method	LBA-30					LBA-60				
		RMSE↓	MAE↓	R↑	ρ↑	τ↑	RMSE↓	MAE↓	R↑	ρ↑	τ↑
EGNN	Baseline	2.084	1.622	0.440	0.446	0.307	1.945	1.482	0.501	0.534	0.373
	DA	1.754	1.405	0.449	0.448	0.307	1.794	1.381	0.560	0.592	0.419
	MDEA	1.838	1.441	0.446	0.455	0.311	1.959	1.485	0.532	0.571	0.405
	DrugDyn	**1.735**	**1.392**	**0.458**	**0.462**	**0.318**	**1.765**	**1.339**	**0.582**	**0.608**	**0.433**
GIGN	Baseline	1.997	1.617	0.486	0.466	0.319	1.769	1.406	0.662	0.658	0.474
	DA	1.603	1.291	0.478	0.458	0.319	1.405	1.091	0.744	0.739	0.548
	MDEA	1.771	1.369	0.402	0.448	0.308	1.779	1.323	0.644	0.673	0.489
	DrugDyn	**1.498**	**1.198**	**0.526**	**0.500**	**0.350**	**1.335**	**1.036**	**0.756**	**0.757**	**0.562**
DTIGN	Baseline	1.967	1.581	0.412	0.403	0.276	1.918	1.504	0.621	0.632	0.449
	DA	1.676	1.338	0.422	0.425	0.289	1.521	1.138	0.682	0.686	0.506
	MDEA	1.725	1.389	0.475	0.461	0.318	1.730	1.366	0.656	0.672	0.483
	DrugDyn	**1.566**	**1.281**	**0.504**	**0.480**	**0.331**	**1.395**	**1.080**	**0.734**	**0.726**	**0.538**
DeepRLI	Baseline	1.670	**1.301**	0.537	0.529	0.374	1.625	1.261	0.684	0.691	0.504
	DA	1.709	1.346	0.500	0.498	0.348	1.480	1.116	0.722	0.736	0.546
	MDEA	1.700	1.327	0.522	0.520	0.367	1.582	1.224	0.676	0.689	0.502
	DrugDyn	**1.641**	1.310	**0.541**	**0.549**	**0.384**	**1.428**	**1.089**	**0.728**	**0.743**	**0.552**
BA-Pred	Baseline	1.701	1.370	0.464	0.453	0.312	1.572	1.236	0.655	0.656	0.472
	DA	1.787	1.420	0.473	0.462	0.318	1.479	1.157	0.686	0.697	0.508
	MDEA	1.762	1.409	0.459	0.445	0.308	1.525	1.182	0.673	0.679	0.493
	DrugDyn	**1.612**	**1.284**	**0.503**	**0.492**	**0.345**	**1.442**	**1.129**	**0.723**	**0.719**	**0.528**

(SCCL), where each contrastive view contains only one sampled conformation, to isolate the benefit of constructing multi-cardinality conformational views.

As shown in Table 3, both GASA and MCCL consistently improve the baseline across all three evaluated backbones. GASA benefits prediction by modeling geometry-dependent interactions and cross-conformation dependencies, while MCCL enhances complex-level representations by enforcing consistency across different conformational views. Notably, MCCL also consistently outperforms SCCL. For example, with DTIGN, MCCL reduces RMSE from 1.497 to 1.400 and increases R from 0.740 to 0.770, demonstrating the benefit of constructing contrastive views with different numbers of conformations. Finally, combining GASA and MCCL yields the best performance, which confirm the complementary benefits of spatiotemporal interaction modeling and ensemble-level representation learning.

Table 3: Ablation study of the proposed components on the PDBbind v2016 core set.

Backbone	Method	RMSE↓	MAE↓	R↑	ρ↑	τ↑
EGNN	Baseline	2.054	1.587	0.550	0.585	0.414
	+ GASA w/MHA	1.995	1.595	0.578	0.603	0.423
	+ GASA	1.868	1.530	0.602	0.608	0.428
	+ SCCL	1.882	1.484	0.572	0.589	0.411
	+ MCCL	1.807	1.433	0.607	0.613	0.439
	Full	**1.702**	**1.364**	**0.668**	**0.662**	**0.472**
GIGN	Baseline	1.572	1.209	0.699	0.686	0.506
	+ GASA w/MHA	1.473	1.184	0.739	0.722	0.534
	+ GASA	1.385	1.115	0.775	0.767	0.572
	+ SCCL	1.477	1.174	0.735	0.713	0.528
	+ MCCL	1.464	1.156	0.741	0.726	0.534
	Full	**1.330**	**1.040**	**0.798**	**0.803**	**0.604**
DTIGN	Baseline	1.523	1.231	0.723	0.705	0.515
	+ GASA w/MHA	1.422	1.125	0.760	0.752	0.558
	+ GASA	1.399	1.107	0.770	0.762	0.570
	+ SCCL	1.497	1.157	0.740	0.737	0.545
	+ MCCL	1.400	1.082	0.770	0.761	0.571
	Full	**1.330**	**1.070**	**0.802**	**0.790**	**0.597**

5 Conclusion

In this work, we revisit how conformational dynamics can be effectively exploited for protein–ligand binding affinity prediction. We introduce DrugDyn, a unified framework that goes beyond treating MD conformations as independent samples or compressing trajectory information into static representations. DrugDyn combines geometry-aware spatiotemporal attention to capture fine-grained intermolecular interactions and cross-conformation dependencies with multi-cardinality contrastive learning to learn coherent and discriminative representations from different observations of the conformational ensemble. More broadly, our study highlights the importance of moving beyond the mere inclusion of multiple conformations toward explicitly modeling the interaction dynamics and ensemble-level information encoded in MD trajectories.

AI USE STATEMENT

In this work, we used generative AI tools for writing polish. We have not used generative AI tools to design or provide feedback on the research methodology or experiments, implement methods, assist with translation, support qualitative or thematic data analysis, or interpret results. The following uses of generative AI tools are not applicable to this work: generate synthetic data sets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, and clean and reformat datasets. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Andrea Amadei, Antonius BM Linssen, and Herman JC Berendsen. Essential dynamics of proteins. *Proteins: Structure, Function, and Bioinformatics*, 1993.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

Gregory A Babbitt, Madhusudan Rajendran, Miranda L Lynch, Richmond Asare-Bediako, Leora T Mouli, Cameron J Ryan, Harsh Srivastava, Patrick Rynkiewicz, Kavya Phadke, Makayla L Reed, et al. Atomdance: Kernel-based denoising and choreographic analysis for protein dynamic comparison. *Biophysical Journal*, 2024.

David D Boehr, Ruth Nussinov, and Peter E Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nature chemical biology*, 2009.

Anna Gaulton, Louisa J Bellis, A Patricia Bento, Jon Chambers, Mark Davies, Anne Hersey, Yvonne Light, Shaun McGlinchey, David Michalovich, Bissan Al-Lazikani, et al. Chembl: a large-scale bioactivity database for drug discovery. *Nucleic acids research*, 2012.

Pengkang Guo, Bruno Correia, Pierre Vandergheynst, and Daniel Probst. Boosting protein graph representations through static-dynamic fusion. In *ICML*. JMLR.org, 2025.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*. IEEE, 2016.

Yanwen Huang, Bowen Gao, Yinjun Jia, Hongbo Ma, Wei-Ying Ma, Ya-Qin Zhang, and Yanyan Lan. Redefining the task of bioactivity prediction. In *ICLR*, 2025.

James P Hughes, Stephen Rees, S Barrett Kalindjian, and Karen L Philpott. Principles of early drug discovery. *British journal of pharmacology*, 2011.

Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design*, 2016.

Gregory W Kyro, Anthony M Smaldone, Yu Shee, Chuzhi Xu, and Victor S Batista. T-alpha: A hierarchical transformer-based deep neural network for protein–ligand binding affinity prediction with uncertainty-aware self-learning for protein-specific alignment. *Journal of Chemical Information and Modeling*, 2025.

Jiashan Li and Xinqi Gong. Harnessing pre-trained models for accurate prediction of protein-ligand binding affinity. *BMC bioinformatics*, 2025.

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. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021.

Yan Li, Zhihai Liu, Jie Li, Li Han, Jie Liu, Zhixiong Zhao, and Renxiao Wang. Comparative assessment of scoring functions on an updated benchmark: 1. compilation of the test set. *Journal of chemical information and modeling*, 2014.

Pierre-Yves Libouban, Camille Parisel, Maxime Song, Samia Aci-Sèche, Jose C Gómez-Tamayo, Gary Tresadern, and Pascal Bonnet. Spatio-temporal learning from molecular dynamics simulations for protein–ligand binding affinity prediction. *Bioinformatics*, 2025.

Hyukjun Lim, Sun Kim, and Sangseon Lee. Cheapnet: Cross-attention on hierarchical representations for efficient protein-ligand binding affinity prediction. In *ICLR*, 2025.

Haoyu Lin, Jintao Zhu, Shiwei Wang, Yibo Li, Jianfeng Pei, and Luhua Lai. Deeprli: a multi-objective framework for universal protein–ligand interaction prediction. *Digital Discovery*, 4(8): 2083–2103, 2025.

Zhiguo Liu, Guitao Wang, Zhanting Li, and Renxiao Wang. Geometrical preferences of the hydrogen bonds on protein- ligand binding interface derived from statistical surveys and quantum mechanics calculations. *Journal of Chemical Theory and Computation*, 2008.

Zhihai Liu, Minyi Su, Li Han, Jie Liu, Qifan Yang, Yan Li, and Renxiao Wang. Forging the basis for developing protein–ligand interaction scoring functions. *Accounts of chemical research*, 2017.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Buyong Ma, Maxim Shatsky, Haim J Wolfson, and Ruth Nussinov. Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein science*, 2002.

Yaosen Min, Ye Wei, Peizhuo Wang, Xiaoting Wang, Han Li, Nian Wu, Stefan Bauer, Shuxin Zheng, Yu Shi, Yingheng Wang, et al. From static to dynamic structures: Improving binding affinity prediction with graph-based deep learning. *Advanced Science*, 2024.

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*, 2021.

Hakime Öztürk, Arzucan Özgür, and Elif Özkirimli Ölmez. Deepdta: deep drug-target binding affinity prediction. *Bioinformatics*, 2018.

Anastasiia V Sadybekov and Vsevolod Katritch. Computational approaches streamlining drug discovery. *Nature*, 2023.

Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *ICML*. PMLR, 2021.

Min-Hee Seo, Jongmin Park, Eunyoung Kim, Sungchul Hohng, and Hyun-Soo Kim. Protein conformational dynamics dictate the binding affinity for a ligand. *Nature Communications*, 2014.

Bonggun Shin, Sungsoo Park, Keunsoo Kang, and Joyce C Ho. Self-attention based molecule representation for predicting drug-target interaction. In *Machine learning for healthcare conference*. PMLR, 2019.

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*, 2024.

Jaemin Sim and Juyong Lee. Ba-pred and rmsd-pred: Integrated graph neural network models for accurate protein–ligand binding affinity and binding pose prediction. *Journal of Chemical Information and Modeling*, 66(7):3480–3495, 2026.

Raphael JL Townshend, Martin Vögele, Patricia Suriana, Alexander Derry, Alexander Powers, Yianni Laloudakis, Sidhika Balachandar, Bowen Jing, Brandon Anderson, Stephan Eismann, et al. Atom3d: Tasks on molecules in three dimensions. *arXiv preprint arXiv:2012.04035*, 2020.

Anna Vuorinen and Daniela Schuster. Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods*, 2015.

Guishen Wang, Hangchen Zhang, Mengting Shao, Shisen Sun, and Chen Cao. Mmpd-dta: Integrating multi-modal deep learning with pocket-drug graphs for drug-target binding affinity prediction. *Journal of Chemical Information and Modeling*, 2025.

Qianyang Wu, Jingwei Lv, Zilong Zhang, and Feifei Cui. Generalizable drug–target interaction prediction via esm-2 representations and progressive contrastive curriculum learning. In *AAAI*, 2026.

Yiming Yang, Zhiyuan Zhou, Yueming Yin, Hoi-Yeung Li, and Adams Wai-Kin Kong. Test-time adaptation without source data for out-of-domain bioactivity prediction. In *ICLR*, 2026.

Ziduo Yang, Weihe Zhong, Qiujie Lv, Tiejun Dong, and Calvin Yu-Chian Chen. Geometric interaction graph neural network for predicting protein–ligand binding affinities from 3d structures (gign). *The journal of physical chemistry letters*, 2023.

Yueming Yin, Hilbert Yuen In Lam, Yuguang Mu, Hoi-Yeung Li, and Adams Wai-Kin Kong. Advancing bioactivity prediction through molecular docking and self-attention. *IEEE J. Biomed. Health Informatics*, 2024.

Long Zhao, Hongmei Wang, Ximin Zeng, and Shaoping Shi. Mddta: A drug target binding affinity prediction method based on molecular dynamics simulation data enhancement. *IEEE Journal of Biomedical and Health Informatics*, 2026.

Zhiyuan Zhou, Yueming Yin, Yiming Yang, Yuguang Mu, Hoi-Yeung Li, and Adams Wai Kin Kong. Enhancing bioactivity prediction via spatial emptiness representation of protein-ligand complex and union of multiple pockets. In *NeurIPS*, 2025.

A Related Work

Binding Affinity Prediction. Early DL-based methods mainly rely on ligand-centric representations, typically modeling small molecules as 2D graphs (Kearnes et al., 2016). Later approaches incorporate protein information to model protein–ligand interactions. DeepDTA (Öztürk et al., 2018) encodes protein sequences and ligand SMILES with CNNs, while MT-DTI (Shin et al., 2019) further introduces attention-based sequence modeling. With increasing access to structural data, structure-based methods directly model protein–ligand complexes. GraphDTA (Nguyen et al., 2021) introduces 3D molecular graphs for drug representation, while GIGN (Yang et al., 2023) and DTIGN (Yin et al., 2024) further construct 3D drug–pocket interaction graphs to capture binding geometry. More recent studies have advanced binding affinity prediction through multimodal fusion and representation learning. MMPD-DTA (Wang et al., 2025) integrates graph- and sequence-level representations of proteins and drugs, while T-ALPHA (Kyro et al., 2025) models pocket surface geometry using oriented point clouds. Meanwhile, SableBind (Li & Gong, 2025) improves molecular representations through self-supervised structural perturbations, and ESP-DTI (Wu et al., 2026) adopts cross-modal contrastive alignment to enhance generalization under distribution shifts.

Molecular Dynamics for Binding Affinity Prediction. Recent studies have begun to incorporate molecular dynamics (MD) simulations into binding affinity prediction. A common strategy leverages MD-sampled conformations as additional training data, as in Dynaformer (Min et al., 2024), DenseNucy (Libouban et al., 2025), and MDDTA (Zhao et al., 2026). A second line of work adopts motion-informed graph augmentation. Guo et al. (2025) derives additional edges from correlated motions across MD frames, thereby incorporating trajectory-level motion information into a static graph. More explicit temporal modeling is explored by TimeNucy and VideoNucy (Libouban et al., 2025), which process successive conformations using recurrent architectures. However, these approaches do not outperform simpler MD-based data augmentation. Overall, existing methods mainly exploit conformational dynamics through independent-state augmentation, motion-informed edge construction, or conventional temporal sequence modeling, leaving fine-grained interaction evolution and ensemble-level representation learning insufficiently addressed.

B Implementation Details

All models are trained for 50 epochs using AdamW (Loshchilov & Hutter, 2017) with a learning rate of 1×10^{-4}. The batch size is set to 128 for EGNN, GIGN, and DTIGN, and reduced to 32 for DeepRLI and BA-Pred due to GPU memory constraints. Both the geometry-aware cross-molecular attention and atom-wise temporal attention use 8 attention heads. For MCCL, the temperature γ is set to 0.2 for EGNN, GIGN, and DTIGN, and 0.05 for DeepRLI and BA-Pred, with the loss weight λ fixed at 0.5 across all backbones. All experiments are conducted on a server equipped with 8 NVIDIA H200 NVL GPUs.

C Hyperparameter Sensitivity Analysis

We investigate the sensitivity of DrugDyn to the temperature γ in MCCL and contrastive loss weight λ. Experiments are conducted on LBA-60 using DeepRLI and BA-Pred, with $\gamma \in \{0.05, 0.1, 0.2, 0.5\}$ and $\lambda \in \{0.1, 0.2, 0.5, 1.0\}$, while keeping all other training settings unchanged. As shown in Table 4, DrugDyn consistently delivers stronger overall performance than the corresponding baseline, DA, and MDEA variants across all evaluated hyperparameter combinations on both backbones, demonstrating its robustness to hyperparameter variations.

D Evaluation Metrics

We evaluate predictive accuracy and the ability to preserve relative binding affinity relationships using five metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Pearson correlation coefficient (R), Spearman rank correlation coefficient (ρ), and Kendall's tau (τ). Let n denote the number of samples, and let y_i and $\hat{y}_i$ denote the ground-truth and predicted binding affinities of the i-th sample, respectively.

Root Mean Squared Error (RMSE). RMSE measures the square root of the average squared difference between predicted and ground-truth binding affinities:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}.$$

By assigning greater weight to larger deviations, RMSE is particularly sensitive to substantial prediction errors. It therefore reflects the model's ability to maintain accurate predictions while avoiding large deviations from experimental measurements. Lower RMSE values indicate better predictive accuracy.

Mean Absolute Error (MAE). MAE measures the average absolute difference between predicted and ground-truth binding affinities:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|.$$

Unlike RMSE, MAE does not disproportionately penalize large deviations, making it less sensitive to outliers. It provides an intuitive measure of the typical prediction error across protein–ligand complexes and complements RMSE in assessing overall predictive accuracy. Lower MAE values indicate better model performance.

Pearson Correlation Coefficient (R). Pearson's R measures the strength and direction of the linear relationship between predicted and ground-truth binding affinities:

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}},$$

Table 4: Hyperparameter sensitivity analysis of MCCL on LBA-60 using DeepRLI and BA-Pred.

Backbone	Method	γ	λ	RMSE↓	MAE↓	R↑	ρ↑	τ↑
DeepRLI	Baseline	–	–	1.625	1.261	0.684	0.691	0.504
	DA	–	–	1.480	1.116	0.722	0.736	0.546
	MDEA	–	–	1.582	1.224	0.676	0.689	0.502
	DrugDyn	0.05	0.1	1.421	1.072	0.740	0.756	0.567
			0.2	1.431	1.068	0.735	0.762	0.571
			0.5	1.428	1.089	0.728	0.743	0.552
			1.0	1.461	1.099	0.731	0.745	0.554
		0.1	0.1	1.440	1.096	0.742	0.758	0.566
			0.2	1.493	1.125	0.731	0.752	0.561
			0.5	1.452	1.105	0.738	0.744	0.556
			1.0	1.474	1.117	0.734	0.753	0.561
		0.2	0.1	1.503	1.132	0.724	0.741	0.553
			0.2	1.465	1.119	0.736	0.753	0.559
			0.5	1.463	1.093	0.732	0.749	0.560
			1.0	1.448	1.072	0.735	0.750	0.566
		0.5	0.1	1.506	1.130	0.726	0.748	0.559
			0.2	1.469	1.116	0.734	0.750	0.559
			0.5	1.458	1.093	0.726	0.744	0.555
			1.0	1.456	1.095	0.734	0.746	0.558
BA-Pred	Baseline	–	–	1.572	1.236	0.655	0.656	0.472
	DA	–	–	1.479	1.157	0.686	0.697	0.508
	MDEA	–	–	1.525	1.182	0.673	0.679	0.493
	DrugDyn	0.05	0.1	1.448	1.120	0.713	0.703	0.517
			0.2	1.439	1.116	0.717	0.709	0.522
			0.5	1.442	1.129	0.723	0.719	0.528
			1.0	1.418	1.099	0.730	0.729	0.537
		0.1	0.1	1.458	1.126	0.711	0.703	0.518
			0.2	1.431	1.115	0.728	0.722	0.533
			0.5	1.462	1.135	0.718	0.709	0.522
			1.0	1.458	1.123	0.713	0.711	0.521
		0.2	0.1	1.466	1.124	0.706	0.697	0.516
			0.2	1.433	1.107	0.720	0.714	0.527
			0.5	1.477	1.150	0.710	0.704	0.516
			1.0	1.501	1.187	0.701	0.706	0.516
		0.5	0.1	1.427	1.118	0.730	0.724	0.532
			0.2	1.441	1.131	0.725	0.718	0.525
			0.5	1.426	1.103	0.721	0.715	0.528
			1.0	1.482	1.151	0.734	0.736	0.545

where $\bar{y}$ and $\bar{\hat{y}}$ denote the means of the ground-truth and predicted binding affinities, respectively. The coefficient ranges from -1 to 1, with values approaching 1 indicating strong positive linear correlation, values approaching -1 indicating strong negative linear correlation, and values near 0 indicating weak linear association. A higher positive R indicates that the model more closely captures the linear variation in experimental binding affinities.

Spearman Rank Correlation Coefficient (ρ). Spearman's ρ measures the strength and direction of the monotonic relationship between predicted and ground-truth binding affinities based on their ranks:

$$\rho = \frac{\sum_{i=1}^{n}(r_i - \bar{r})(\hat{r}_i - \bar{\hat{r}})}{\sqrt{\sum_{i=1}^{n}(r_i - \bar{r})^2}\sqrt{\sum_{i=1}^{n}(\hat{r}_i - \bar{\hat{r}})^2}},$$

where r_i and $\hat{r}_i$ denote the ranks of y_i and $\hat{y}_i$, respectively, and $\bar{r}$ and $\bar{\hat{r}}$ are their corresponding mean ranks. Unlike Pearson correlation, Spearman correlation captures monotonic relationships without requiring linear dependence. It evaluates whether the model preserves the relative ordering of protein–ligand complexes according to their binding affinities. Higher values indicate stronger agreement between predicted and experimental affinity rankings.

Kendall's Tau (τ). Kendall's tau measures ordinal association by comparing the relative ordering of all pairs of samples. A pair is concordant if its predicted and ground-truth binding affinities follow the same ordering, and discordant if their orderings disagree. We adopt Kendall's τ, which accounts for tied values:

$$\tau = \frac{C - D}{\sqrt{(C + D + T_y)(C + D + T_{\hat{y}})}},$$

where C and D denote the numbers of concordant and discordant pairs, respectively. T_y denotes the number of pairs tied only in ground-truth binding affinities, while $T_{\hat{y}}$ denotes the number of pairs tied only in predicted binding affinities. By explicitly accounting for pairwise ordering agreement and disagreement, Kendall's tau provides an additional measure of ranking consistency across protein–ligand complexes. Higher values indicate stronger agreement between predicted and ground-truth affinity rankings.