# Learning Protein–Ligand Binding in Hyperbolic Space

**Jianhui Wang**[1,2*], **Wenyu Zhu**[1*], **Bowen Gao**[1,3*], **Xin Hong**[1], **Ya-Qin Zhang**[1], **Wei-Ying Ma**[1], **Yanyan Lan**[1,4†]

[1]Institute for AI Industry Research (AIR), Tsinghua University, Beijing, China
[2]University of Electronic Science and Technology of China, Chengdu, China
[3]Department of Computer Science and Technology, Tsinghua University, Beijing, China
[4]Beijing Academy of Artificial Intelligence, Beijing, China
lanyanyan@air.tsinghua.edu.cn

## Abstract

Protein-ligand binding prediction is central to virtual screening and affinity ranking, two fundamental tasks in drug discovery. While recent retrieval-based methods embed ligands and protein pockets into Euclidean space for similarity-based search, the geometry of Euclidean embeddings often fails to capture the hierarchical structure and fine-grained affinity variations intrinsic to molecular interactions. In this work, we propose HypSeek, a hyperbolic representation learning framework that embeds ligands, protein pockets, and sequences into Lorentz-model hyperbolic space. By leveraging the exponential geometry and negative curvature of hyperbolic space, HypSeek enables expressive, affinity-sensitive embeddings that can effectively model both global activity and subtle functional differences–particularly in challenging cases such as activity cliffs, where structurally similar ligands exhibit large affinity gaps. Our model unifies virtual screening and affinity ranking in a single framework, introducing a protein-guided three-tower architecture to enhance representational structure. HypSeek improves early enrichment in virtual screening on DUD-E from 42.63 to 51.44 (+20.7%) and affinity ranking correlation on JACS from 0.5774 to 0.7239 (+25.4%), demonstrating the benefits of hyperbolic geometry across both tasks and highlighting its potential as a powerful inductive bias for protein-ligand modeling.

**Code** — https://github.com/jianhuiwemi/HypSeek
**Extended version** — https://arxiv.org/pdf/2508.15480

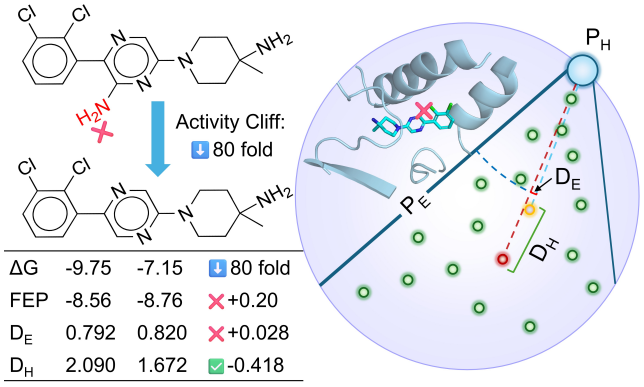


Figure 1: Illustration of how hyperbolic geometry distinguishes activity cliffs (PDB ID: 5EHR). **Left**: Two structurally similar ligands (Ligand ID: 5OD vs. its amino-substituent-removed derivative) show an ∼80-fold affinity difference. **Right**: The yellow and red points denote the two ligands; the blue point is the pocket. Dashed lines show distances in hyperbolic (red/light blue) and Euclidean (dark blue) space. Euclidean embeddings preserve structural similarity but fail to reflect affinity gaps, while hyperbolic embeddings separate such pairs via both radial and angular dimensions ($D_H$, green), enabling affinity-sensitive representations.

## 1 Introduction

Modeling protein–ligand interactions is critical for drug discovery, where accurate binding affinity prediction underpins both large-scale virtual screening and fine-grained ligand prioritization. Virtual screening seeks to identify molecules likely to bind a given protein target from large compound libraries, often containing millions or even billions of candidates. Approaches such as molecular docking (Friesner et al. 2004; Trott and Olson 2010) estimate binding compatibility by sampling ligand poses and scoring them with physics-based functions. While effective in small-scale settings, these methods are computationally intensive and scale poorly to modern library sizes. Unlike virtual screening, which emphasizes identifying likely binders from vast libraries, affinity ranking focuses on ordering a smaller set of candidate ligands by predicted binding strength, with physics-based techniques like free energy perturbation (FEP+) (Wang et al. 2015) offering high accuracy at the cost of extensive molecular dynamics simulations. These limitations restrict the practicality of traditional methods in early-stage drug discovery pipelines.

A notable shift in virtual screening came with DrugCLIP (Gao et al. 2023), which reframed the task as a dense retrieval problem. Rather than predicting binding affinity or docking poses, DrugCLIP learns contrastive embeddings of ligands and protein pockets such that interacting pairs are close in a shared Euclidean space. This design enables

*Equal contribution.
†Corresponding author.

efficient similarity-based retrieval and allows for scalable screening across billion-scale compound libraries. Despite its promising performance and efficiency, DrugCLIP struggles to capture fine-grained interaction patterns which are essential for downstream affinity ranking. Recently, LigUnity (Feng et al. 2025) extends the retrieval-based framework by unifying virtual screening and affinity ranking into a single training objective. It combines contrastive learning for global interaction patterns with listwise ranking to model pocket-specific ligand preferences, aiming to jointly learn both binding likelihood and relative affinity within a unified embedding space.

While retrieval-based methods have shown strong potential, they typically embed ligands and protein pockets into Euclidean space, where distances grow linearly and the geometry does not explicitly encourage separation based on functional or activity-related differences. As a result, standard Euclidean training objectives may fail to emphasize fine-grained distinctions in binding strength, especially when molecular structures are similar.

To enrich the embedding geometry and better capture complex protein–ligand interactions, we propose HypSeek, a retrieval-based model that embeds ligands, pockets, and protein sequences into hyperbolic space. Unlike previous dual-tower designs, HypSeek adopts a protein-guided three-tower architecture during training to promote more structured representations. The curvature of hyperbolic space enables affinity-sensitive encoding through both angular direction and radial depth, providing greater expressivity than linear Euclidean geometry. This design not only enhances fine-grained affinity discrimination, but also offers a natural mechanism to address activity cliffs—cases where structurally similar ligands exhibit large differences in binding strength. While Euclidean embeddings often enforce functional similarity among structurally similar ligands, hyperbolic geometry allows such ligands to diverge meaningfully in the embedding space, reflecting differences in interaction modes or physicochemical properties. During inference, we retain efficient similarity computation via Euclidean inner products over hyperbolically shaped representations, preserving scalability without sacrificing expressiveness.

We evaluate HypSeek across both large-scale virtual screening and fine-grained affinity ranking tasks. On the DUD-E (Mysinger et al. 2012) benchmark, HypSeek improves $EF_{1\%}$ from **42.63 to 51.44 (+20.7%)**, demonstrating strong retrieval performance across targets. For affinity ranking, it increases Spearman correlation on the JACS (Wang et al. 2015) dataset from **0.5774 to 0.7239 (+25.4%)**, consistently outperforming Euclidean baselines. These results highlight the benefits of hyperbolic geometry in capturing both global activity and nuanced affinity variation within a unified embedding space.

In summary, our contributions are as follows:

- We propose a hyperbolic embedding framework for protein–ligand modeling, where the geometry naturally captures hierarchical interactions and targets **the critical challenge of activity cliffs** by enabling structured separation of similar ligands with divergent affinities.
- We introduce **HypSeek**, a dense retrieval model with a protein-guided three-tower architecture that integrates structure and sequence information to learn affinity-aware representations in hyperbolic space.
- HypSeek achieves strong performance on both virtual screening and affinity ranking, capturing fine-grained binding differences more effectively than Euclidean baselines while maintaining scalable inference.

## 2 Related Work

**Virtual Screening.** Structure-based virtual screening traditionally relies on molecular docking methods such as Glide (Friesner et al. 2004) and AutoDock (Trott and Olson 2010), which predict ligand binding poses and evaluate affinities using physics-based scoring functions. Some predict binding affinity directly from protein–ligand complex structures by learning scoring functions (McNutt et al. 2021a; Jiang et al. 2021; Shen et al. 2022; Cao et al. 2024), while others infer interactions from raw structural inputs (Lu et al. 2022; Zhang et al. 2023). A major shift occurred with DrugCLIP (Gao et al. 2023), which introduced contrastive retrieval by aligning ligand and pocket embeddings in a shared Euclidean space for billion-scale similarity search. This paradigm has since inspired a range of efficient retrieval methods. For example, DrugHash (Han, Hong, and Li 2025a) employs binary hash codes for efficient retrieval with reduced memory cost, and LigUnity (Feng et al. 2025) integrates listwise ranking with contrastive screening.

**Affinity Ranking.** Accurate ranking of ligand binding affinities is essential for lead optimization but remains computationally challenging. Physics-based methods such as FEP+ (Wang et al. 2015) and MM-GB/SA (Genheden and Ryde 2015) deliver high accuracy via alchemical free-energy calculations and implicit solvent models, respectively, yet they require extensive molecular dynamics sampling. Recent deep learning approaches seek to reduce this cost: PBCNet (Yu et al. 2023) models pairwise ligand differences with graph neural networks, EHIGN (Yang et al. 2024) encodes heterogeneous protein–ligand interaction graphs, and LigUnity (Feng et al. 2025) combines contrastive screening with listwise ranking to jointly address global retrieval and local prioritization.

**Hyperbolic Representation Learning.** Hyperbolic space offers a natural geometry for data with latent hierarchies, due to its exponential volume growth and low-distortion embeddings (Nickel and Kiela 2017; Chamberlain, Clough, and Deisenroth 2017). It has been successfully applied in taxonomy modeling (Ganea, Bécigneul, and Hofmann 2018), graph learning (Bécigneul and Ganea 2018), and adapted to various neural architectures (Gulcehre et al. 2019; Bdeir, Schwethelm, and Landwehr 2024). Recent works extend its use to multimodal and vision–language tasks (Desai et al. 2023a; Poppi et al. 2025). We are the first to apply hyperbolic learning to protein–ligand retrieval, leveraging its geometric bias to better capture fine-grained affinity differences.

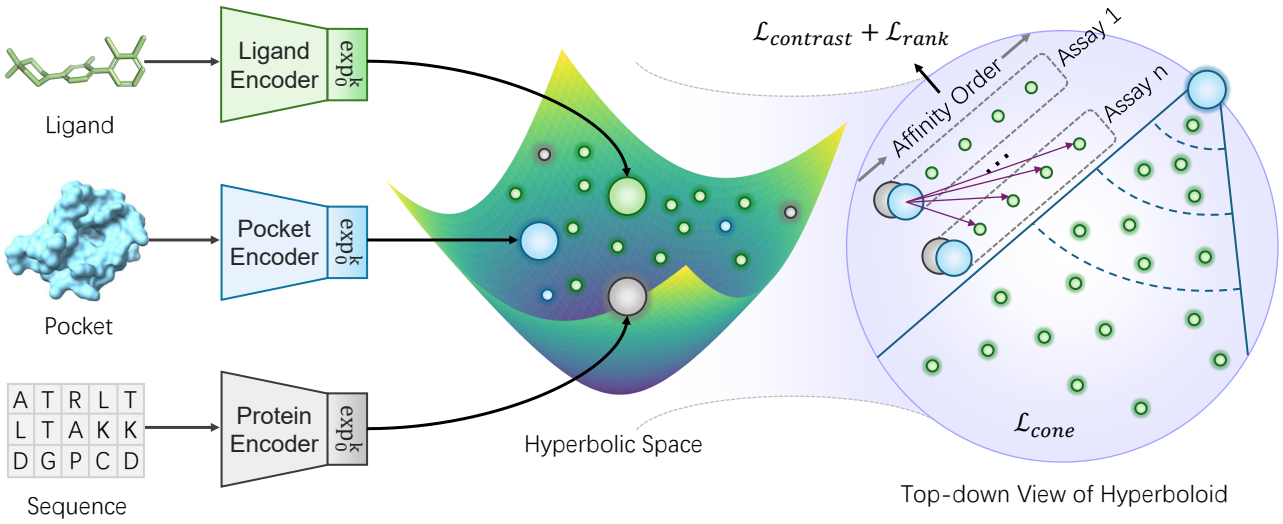


Figure 2: Overall architecture of HypSeek: three encoders lift ligands, pockets and protein sequences to a shared hyperbolic space (left); contrastive and list-wise ranking losses align pocket/sequence with ligands while the cone–hierarchy loss imposes radial–angular tiers around each pocket (right).

## 3 Preliminaries

We perform all representation learning in an $n$-dimensional hyperbolic space of constant negative curvature, using the Lorentz model (Nickel and Kiela 2018; Lin et al. 2023a; Desai et al. 2023a). This choice affords numerical stability and readily supports geodesic and exponential-map operations.

Let $\mathbb{L}^n$ denote the Lorentz (hyperboloid) model, realized as the upper sheet of a two-sheeted hyperboloid in $\mathbb{R}^{n+1}$. We first equip $\mathbb{R}^{n+1}$ with the Lorentzian inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle_{\mathbb{L}} = -\, p_0\, q_0 + \langle \tilde{\mathbf{p}}, \tilde{\mathbf{q}} \rangle_{\mathbb{E}}, \tag{1}$$

where we write $\mathbf{p} = (p_0, \tilde{\mathbf{p}}), p_0 \in \mathbb{R}, \tilde{\mathbf{p}} \in \mathbb{R}^n$ with $p_0$ the *time*-coordinate and $\tilde{\mathbf{p}}$ the *spatial*-coordinates, and $\langle \cdot, \cdot \rangle_{\mathbb{E}}$ denotes the standard Euclidean inner product.

The Lorentz model is then defined by

$$\mathbb{L}^n = \Big\{ \mathbf{p} \in \mathbb{R}^{n+1} : \langle \mathbf{p}, \mathbf{p} \rangle_{\mathbb{L}} = -\frac{1}{\kappa}, p_0 = \sqrt{\tfrac{1}{\kappa} + \|\tilde{\mathbf{p}}\|^2}, \kappa > 0 \Big\}, \tag{2}$$

where $-\kappa \in \mathbb{R}$ is the curvature of the space.

We can measure distances by integrating the metric along geodesics. The Riemannian metric induced by the Lorentzian inner product gives the length of geodesics on $\mathbb{L}^n$, which in turn defines the hyperbolic distance.

$$d_{\mathbb{L}}(\mathbf{p}, \mathbf{q}) = \frac{1}{\sqrt{\kappa}} \cosh^{-1}\big(-\kappa \langle \mathbf{p}, \mathbf{q} \rangle_{\mathbb{L}}\big), \quad \mathbf{p}, \mathbf{q} \in \mathbb{L}^n. \tag{3}$$

At each point $\mathbf{p} \in \mathbb{L}^n$, the tangent space $T_{\mathbf{p}}\mathbb{L}^n$ provides a linear approximation of the manifold. Concretely, any tangent vector $\mathbf{v} \in T_{\mathbf{p}}\mathbb{L}^n \subset \mathbb{R}^{n+1}$ satisfies $\langle \mathbf{p}, \mathbf{v} \rangle_{\mathbb{L}} = 0$, so that

$$T_{\mathbf{p}}\mathbb{L}^n = \big\{ \mathbf{v} \in \mathbb{R}^{n+1} : \langle \mathbf{p}, \mathbf{v} \rangle_{\mathbb{L}} = 0 \big\}. \tag{4}$$

To transfer Euclidean encoder outputs into hyperbolic space, we apply the exponential map at a base point. For any $\mathbf{p} \in \mathbb{L}^n$ and $\mathbf{v} \in T_{\mathbf{p}}\mathbb{L}^n$, the exponential map is

$$\exp_{\mathbf{p}}^{\kappa}(\mathbf{v}) = \cosh\big(\sqrt{\kappa}\, \|\mathbf{v}\|_{\mathbb{L}}\big)\, \mathbf{p} + \frac{\sinh\big(\sqrt{\kappa}\, \|\mathbf{v}\|_{\mathbb{L}}\big)}{\sqrt{\kappa}\, \|\mathbf{v}\|_{\mathbb{L}}}\, \mathbf{v}, \tag{5}$$

where $\|\mathbf{v}\|_{\mathbb{L}} = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle_{\mathbb{L}}}$. In practice, we interpret the output of a Euclidean encoder as a vector in the tangent space at the point $\mathbf{0} = (\frac{1}{\sqrt{\kappa}}, 0, \ldots, 0)^\top$ on the hyperboloid, and then apply the exponential map $\exp_{\mathbf{0}}^{\kappa}$ to lift it onto $\mathbb{L}^n$ (Khrulkov et al. 2020).

## 4 Method

### 4.1 Problem Setting

Our goal is to predict the binding affinity between protein pockets and candidate ligands. The training data are organized by assay, where each assay is an experimental setup designed to evaluate ligand binding against a specific protein target. Each assay includes one protein and a subset of ligands from the full compound library that have been experimentally screened, yielding binary activity labels and optionally affinity values. Crucially, affinity values are only comparable within the same assay due to differences in experimental conditions (e.g., pH, temperature, cofactors), assay protocols (e.g., cell-based or target-based), and measurement types (e.g., $\mathrm{IC}_{50}$, $K_d$, $K_i$).

Therefore, the task is formulated as learning relative binding strength rankings within each assay rather than predicting absolute affinities across assays. Let $\mathcal{A}$ denote the set of assays. For each assay $A_i \in \mathcal{A}$, let $\mathcal{L}_i$ be the set of tested ligands, and $v_i(\ell)$ be the affinity value of ligand $\ell \in \mathcal{L}_i$. Each assay corresponds to a target protein, represented by both its amino acid sequence and a set of candidate pocket structures $\mathcal{P}_i$. During training, one pocket from $\mathcal{P}_i$ is sampled to represent the structure, and combined with the sequence information to encode the full target. The model is trained

to embed both targets and ligands into a shared hyperbolic space, enabling retrieval of active ligands and ranking them by relative binding strengths within each assay.

### 4.2 Multimodal Encoding and Lorentz Mapping

Let $x^p$ and $x^m$ denote the atom-based inputs (coordinates and types) for a protein pocket and ligand, respectively, and let $S = (s_1, \ldots, s_L)$ denote the amino acid sequence of a target protein. We define three encoder functions: $g_\phi$ and $f_\theta$ as SE(3)-equivariant 3D graph transformers for pockets and ligands (following DrugCLIP (Gao et al. 2023)), and $h_\psi$ as a protein sequence encoder based on ESM-2 (Lin et al. 2023b). As illustrated in Figure 2, each encoder maps its input to a vector in $\mathbb{R}^{d_{\text{euc}}}$:

$$E_{\text{poc}} = g_\phi(x^p), E_{\text{mol}} = f_\theta(x^m), E_{\text{seq}} = h_\psi(S). \quad (6)$$

We then lift these Euclidean embeddings to hyperbolic space via the exponential map defined in Eq. (5):

$$\mathbf{h}_{\text{poc}} = \exp_{\mathbf{0}}^{\kappa}\big(E_{\text{poc}}\big), \mathbf{h}_{\text{mol}} = \exp_{\mathbf{0}}^{\kappa}\big(E_{\text{mol}}\big), \mathbf{h}_{\text{seq}} = \exp_{\mathbf{0}}^{\kappa}\big(E_{\text{seq}}\big). \quad (7)$$

The resulting hyperbolic embeddings $\mathbf{h}_{\text{mol}}, \mathbf{h}_{\text{poc}}, \mathbf{h}_{\text{seq}} \in \mathbb{L}^n$ are subsequently employed in both the training and the inference stage.

### 4.3 Contrastive and Ranking as the Foundation.

We retain the in-batch contrastive retrieval losses of DrugCLIP (Gao et al. 2023) and LigUnity's listwise ranking term (Feng et al. 2025), applied to the hyperbolic embeddings $\tilde{\mathbf{h}}_u$. For each assay $A_i$ with query modality $u \in \{\text{poc}, \text{seq}\}$ and its $B$ candidate ligands $\{v_j\}$, we compute similarity logits $s_{i,j} = \frac{1}{\tau}\langle \tilde{\mathbf{h}}_{u_i}, \tilde{\mathbf{h}}_{v_j} \rangle$.

We adopt a symmetric InfoNCE objective over each assay $A_i$. Let $L_i \subseteq \{1, \ldots, B\}$ denote the indices of true binders for $u_i$. We compute:

$$\mathcal{L}^{(i)}_{\text{p}\to\text{l}} = -\frac{1}{|L_i|}\sum_{k\in L_i} \log \frac{\exp(s_{i,k})}{\sum_{j=1}^{B}\exp(s_{i,j})}, \quad (8)$$

$$\mathcal{L}^{(i)}_{\text{l}\to\text{p}} = -\frac{1}{|L_i|}\sum_{k\in L_i} \log \frac{\exp(s_{i,k})}{\sum_{n=1}^{B}\exp(s_{i,n})}, \quad (9)$$

The total contrastive loss is then

$$\mathcal{L}_{\text{contrast}} = \frac{1}{2}\sum_i \Big(\mathcal{L}^{(i)}_{\text{p}\to\text{l}} + \mathcal{L}^{(i)}_{\text{l}\to\text{p}}\Big). \quad (10)$$

For each assay $A_i$ the screened ligands are sorted by measured affinity, yielding an ordered list $(v_{i,1}, \ldots, v_{i,B})$. Following the Plackett–Luce model (Cao et al. 2007), the probability of selecting ligand $v_{i,k}$ at step $k$ (from the remaining set $\mathcal{R}_{i,k} = \{k, k+1, \ldots, B\}$) is

$$p_{i,k}(v_{i,k}) = \frac{\exp\big(s_{i,k}\big)}{\sum_{j\in\mathcal{R}_{i,k}}\exp\big(s_{i,j}\big)}, \quad (11)$$

where $s_{i,k} = \langle \tilde{\mathbf{h}}_{u_i}, \tilde{\mathbf{h}}_{v_{i,k}} \rangle/\tau$. We use the decay $\mu_k = \frac{1}{\sqrt{B}\,\log(k+1)}$. The listwise loss for assay $A_i$ is therefore

$$\mathcal{L}^{(i)}_{\text{rank}} = -\sum_{k=1}^{B}\mu_k \log p_{i,k}\big(v_{i,k}\big). \quad (12)$$

### 4.4 Hyperbolic Geometry as a Structural Prior

Beyond simply embedding pockets and ligands into a shared hyperbolic space, we aim to further leverage the geometric structure of $\mathbb{L}^n$ to encode fine-grained inductive biases about binding affinity. The exponential capacity of hyperbolic space allows for natural modeling of hierarchical relationships, while the Lorentz model enables cone-based entailment mechanisms. We therefore introduce a cone–hierarchy learning process that exploits both the radial and angular dimensions of hyperbolic space to reflect the graded nature of ligand binding strength.

Within an assay $A_i$, the protein pocket is represented by a Lorentz-model vector $\mathbf{h}_{\text{poc},i} \in \mathbb{L}^n$, and every screened ligand $j \in \mathcal{L}_i$ has its own embedding $\mathbf{h}_{\text{mol},ij} \in \mathbb{L}^n$. Each hyperbolic vector splits into a time-like coordinate and an $n$-dimensional spatial part: $\mathbf{h}_{\text{poc},i} = \big(p_{0,i},\ \tilde{\mathbf{p}}_i\big), \mathbf{h}_{\text{mol},ij} = \big(m_{0,ij},\ \tilde{\mathbf{m}}_{ij}\big)$, with $p_{0,i}, m_{0,ij} \in \mathbb{R}$ and $\tilde{\mathbf{p}}_i, \tilde{\mathbf{m}}_{ij} \in \mathbb{R}^n$. These components satisfy the hyperboloid constraint $p_{0,i}^2 - \|\tilde{\mathbf{p}}_i\|^2 = m_{0,ij}^2 - \|\tilde{\mathbf{m}}_{ij}\|^2 = 1/\kappa$.

The geodesic distance $d_{i,j} = d_{\mathbb{L}}(\mathbf{h}_{\text{poc},i}, \mathbf{h}_{\text{mol},i,j})$ is computed via Eq. (3). The exterior angle at the pocket,

$$\phi_{i,j} = \arccos\Big(\frac{m_{0,i,j} + \kappa\big(\langle\tilde{\mathbf{p}}_i, \tilde{\mathbf{m}}_{i,j}\rangle - p_{0,i}m_{0,i,j}\big)p_{0,i}}{\|\tilde{\mathbf{p}}_i\|\sqrt{\big[\kappa(\langle\tilde{\mathbf{p}}_i, \tilde{\mathbf{m}}_{i,j}\rangle - p_{0,i}m_{0,i,j})\big]^2 - 1}}\Big), \quad (13)$$

follows from the hyperbolic law of cosines and measures how far the ligand "leans" away from the pocket direction.

Each pocket defines a surface of admissible directions. Its half-aperture angle is formulated by Le et al. (2019); Desai et al. (2023b) as

$$\omega_i = \arcsin\Big(\tfrac{2r_0}{\sqrt{\kappa}\,\|\tilde{\mathbf{p}}_i\|}\Big), \quad (14)$$

with a small constant $r_0 > 0$ to keep the expression bounded near the origin; larger $\|\tilde{\mathbf{p}}_i\|$ (a pocket already pushed towards the boundary) therefore yields a narrower cone.

Given the assay–specific affinity values $\{v_{i,j}\}_{j=1}^{|\mathcal{L}_i|}$, we draw $K$ thresholds $t_0 < t_1 < \cdots < t_K$ and assign each ligand a bucket index

$$b_{i,j} = \big\{\, k \in \{0, \ldots, K\} \,:\, v_{i,j} \in [t_k, t_{k+1})\,\big\}. \quad (15)$$

Bucket 0 therefore collects the weakest binders and bucket $K$ the strongest. For every ligand we derive a bucket–specific radial limit $r_{i,j}$ and angular–scaling factor $\eta_{i,j}$

$$r_{i,j} = r_0 + b_{i,j}\,\Delta r, \qquad \eta_{i,j} = \eta_0 - b_{i,j}\,\Delta\eta, \quad (16)$$

where $r_0$ and $\eta_0$ are the base radius/angle for the weakest tier, and $\Delta r, \Delta\eta > 0$ are the per-tier increments. Smaller $b_{i,j}$ thus yields a smaller radius cap and a larger cone. We penalise violations in radius and angle:

$$L_{\text{rad}} = \frac{1}{\sqrt{N}}\sum_{i,j}\max\big(d_{i,j} - r_{i,j},\, 0\big), \quad (17)$$

$$L_{\text{ang}} = \frac{1}{\sqrt{N}}\sum_{i,j}\max\big(\phi_{i,j} - \eta_{i,j}\,\omega_i,\, 0\big), \quad (18)$$

and combine them as

$$\mathcal{L}_{\text{cone}} = \lambda_{\text{rad}}\, L_{\text{rad}} + \lambda_{\text{ang}}\, L_{\text{ang}}. \tag{19}$$

We furthur introduce two regularization terms that operate on angular structure and intra-assay heterogeneity, respectively. To prevent trivial angular collapse, we introduce a fixed angular margin $m > 0$ beyond the cone boundary:

$$R_{\text{ang}} = \frac{1}{\sqrt{N}} \sum_{i,j} \max\big(\phi_{i,j} - \eta_{i,j}\omega_i + m,\ 0\big), \tag{20}$$

We also re-weight active ligands within each assay using rank-based weights $w_{i,j}$ and intra-assay softmax scores $p_{i,j}$:

$$R_{\text{het}} = \frac{1}{\max(C,1)} \sum_{i} \sum_{\substack{j \\ v_{i,j} < v_{\text{th}}}} -w_{i,j} \log p_{i,j}, \tag{21}$$

where $C$ is the number of assays with at least one active ligand, and $v_{\text{th}}$ is a predefined affinity threshold.

### 4.5 Addressing Activity Cliffs with Hyperbolic Geometry

While structurally similar ligands often cluster in Euclidean space, such geometry can underrepresent functional differences—especially in activity cliffs, where minor structural changes lead to large affinity shifts. As formalized in Proposition 1, hyperbolic space provides exponentially greater separation via angular variation, offering a principled mechanism for distinguishing such cases. The theoretical derivation is provided in Appendix.

**Proposition 1.** ***(Hyperbolic Separation of Activity Cliffs)*** *Let $\ell_1, \ell_2$ be structurally similar ligands with large affinity differences. Under constant radial norm and small angular deviation, hyperbolic embeddings yield significantly larger geodesic distance than their Euclidean counterparts:*

$$d_{\mathbb{H}}(h_H(\ell_1), h_H(\ell_2)) \gg d_E(h_E(\ell_1), h_E(\ell_2)).$$

*This highlights the capacity of hyperbolic geometry to distinguish functionally divergent ligands without distorting local structural similarity.*

### 4.6 Training and Inference

The core learning signal is driven by the pocket–ligand relationship. Accordingly, we apply hyperbolic regularisation only to the structure-based (pocket) branch, where geometric alignment in Lorentz space is both meaningful and effective. The sequence pathway provides complementary information to enhance generalisation, but does not participate in hyperbolic supervision.

Our full training objective is given by:

$$\begin{aligned}\mathcal{L}_{\text{total}} \;=\; & \underbrace{\alpha_{\text{poc}}\Big(\mathcal{L}_{\text{cont}}^{\text{poc}\leftrightarrow\text{lig}} + \lambda_{\text{rank}}\, \mathcal{L}_{\text{rank}}^{\text{poc}}\Big)}_{\text{pocket} \leftrightarrow \text{ligand}} \\ & + \underbrace{\alpha_{\text{seq}}\Big(\mathcal{L}_{\text{cont}}^{\text{seq}\leftrightarrow\text{lig}} + \lambda_{\text{rank}}\, \mathcal{L}_{\text{rank}}^{\text{seq}}\Big)}_{\text{sequence} \leftrightarrow \text{ligand}} \\ & + \underbrace{\gamma_{\text{cone}}\, \mathcal{L}_{\text{cone}} + \lambda_{\text{ang}}\, R_{\text{ang}} + \lambda_{\text{het}}\, R_{\text{het}}}_{\text{pocket} \leftrightarrow \text{ligand}}.\end{aligned} \tag{22}$$

At inference time, we simply embed a query pocket and each candidate ligand into hyperbolic space, extract their spatial components $\tilde{\mathbf{h}}_{\text{poc}}$ and $\tilde{\mathbf{h}}_{\text{mol},j}$, and compute similarity scores by their inner product $s_j = \tilde{\mathbf{h}}_{\text{poc}}^{\top}\, \tilde{\mathbf{h}}_{\text{mol},j}$ . We then rank all ligands in descending order of $s_j$.

## 5 Experiments

### 5.1 Experimental Settings

**Implemention Details.** We adopt the same curated assay–level training dataset as LigUnity (Feng et al. 2025), which is constructed from ChEMBL (Mendez et al. 2018), BindingDB (Gilson et al. 2015), and PDBBind (Liu et al. 2017). For virtual screening, we strictly exclude any target UniProt IDs present in the DUD-E (Mysinger et al. 2012), LIT-PCBA (Tran-Nguyen, Jacquemard, and Rognan 2020) test sets. For affinity ranking tasks, we perform ligand-level deduplication by removing redundant small molecules and non-redundant assay IDs. Training is run on four NVIDIA A100 GPUs for 50 epochs, using the Adam optimizer with an initial learning rate of $1 \times 10^{-4}$ and the curvature parameter $\kappa$ (absolute value of negative curvature) fixed to 1.

**Benchmark.** In virtual screening, evaluations are performed on DUD-E (Mysinger et al. 2012) and LIT-PCBA (Tran-Nguyen, Jacquemard, and Rognan 2020). DUD-E includes 102 protein targets, each associated with experimentally verified actives and 50 property-matched decoys, designed to test enrichment capability under artificially constructed decoy scenarios. LIT-PCBA, in contrast, contains 15 targets with over 400K experimentally confirmed inactives, offering a more realistic and challenging setting without synthetic decoy bias. For affinity ranking, the evaluation is conducted on JACS (Wang et al. 2015) and Merck (Schindler et al. 2020). JACS consists of eight high-quality congeneric series extracted from real lead optimization projects, emphasizing precise ranking within narrow chemical series, while Merck serves as a large-scale benchmark for FEP-based lead optimization with diverse chemical scaffolds and higher experimental noise.

**Evaluation Metrics.** For virtual screening, we use AUROC, $\text{BEDROC}_{80.5}$, Enrichment Factor (EF), and ROC-enrichment (RE) to assess model performance. For fine-grained affinity ranking, we evaluate using Pearson's and Spearman's rank correlation coefficients.

**Baselines.** We compare our method against a broad spectrum of existing approaches, including classical physics-based docking tools, empirical scoring functions, and modern deep learning models. These baselines reflect diverse modeling paradigms, ranging from structure-based simulations to neural networks trained on large protein–ligand datasets. For affinity ranking benchmarks, we additionally include methods based on free energy perturbation, energy decomposition, and recent representation learning techniques. All baselines are evaluated using their reported protocols or open-source implementations, ensuring consistency with prior work.

| Method | DUD-E (n = 102) | | | LIT-PCBA (n = 15) | | |
|---|---|---|---|---|---|---|
| | AUROC | $\text{BEDROC}_{80.5}$ | $\text{EF}_{1\%}$ | AUROC | $\text{BEDROC}_{80.5}$ | $\text{EF}_{1\%}$ |
| Glide-SP (Friesner et al. 2004) | 0.7670 | 0.4070 | 16.18 | 0.5315 | 0.4000 | 3.41 |
| Surflex (Spitzer and Jain 2012) | 0.7426 | 0.2387 | 13.35 | 0.5147 | — | 2.50 |
| DeepDTA (Öztürk, Özgür, and Ozkirimli 2018) | 0.5836 | 0.0513 | 2.28 | 0.5627 | 0.0253 | 1.47 |
| Gnina (McNutt et al. 2021b) | 0.7817 | 0.2994 | 17.73 | 0.6093 | 0.0540 | 4.63 |
| BigBind (Brocidiacono et al. 2022) | 0.5014 | 0.0240 | 1.18 | 0.6278 | 0.0502 | 3.79 |
| RTMScore (Shen et al. 2022) | 0.7529 | 0.4341 | 27.10 | 0.5247 | 0.0388 | 2.94 |
| Tankbind (Lu et al. 2022) | 0.7509 | 0.3300 | 13.00 | 0.5970 | 0.0389 | 2.90 |
| DrugCLIP (Gao et al. 2023) | 0.8093 | 0.5052 | 31.89 | 0.5717 | 0.0623 | 5.51 |
| GenScore (Shen et al. 2023) | 0.8160 | 0.4726 | 28.53 | 0.5957 | 0.0654 | 5.14 |
| Planet (Zhang et al. 2024) | 0.7160 | — | 8.83 | 0.5731 | — | 3.87 |
| EquiScore (Cao et al. 2024) | 0.7760 | 0.4320 | 17.68 | 0.5678 | 0.0490 | 3.51 |
| DrugHash (Han, Hong, and Li 2025b) | 0.8373 | 0.5716 | 37.18 | 0.5458 | 0.0714 | 6.14 |
| $\text{LigUnity}_{\text{poc}}$ (Feng et al. 2025) | 0.8922 | 0.6526 | 42.63 | 0.5985 | 0.1133 | 6.47 |
| **HypSeek** | **0.9435** | **0.7892** | **51.44** | **0.6210** | **0.1196** | **6.81** |

Table 1: Virtual-screening results on the DUD-E and LIT-PCBA benchmarks.

| Type | Method | JACS | | Merck | |
|---|---|---|---|---|---|
| | | Pearson $r$ | Spearman $\rho$ | Pearson $r$ | Spearman $\rho$ |
| Physics | FEP+ (Wang et al. 2015) | 0.7811 | 0.7595 | 0.6960 | 0.6798 |
| | MM-GB/SA (Genheden and Ryde 2015) | 0.1489 | 0.2011 | 0.1299 | 0.1299 |
| DL | PBCNet (Yu et al. 2023) | 0.3939 | 0.3799 | 0.4058 | 0.4075 |
| | EHIGN (Yang et al. 2024) | 0.5787 | 0.5814 | 0.4246 | 0.3830 |
| | GET (Kong, Huang, and Liu 2024) | 0.4034 | 0.3753 | 0.4203 | 0.4214 |
| | BindNet (Feng et al. 2024) | 0.5481 | 0.5368 | 0.4037 | 0.3477 |
| | Boltz-2 (Passaro et al. 2025) | 0.5231 | 0.5285 | 0.4298 | 0.4013 |
| | $\text{LigUnity}_{\text{poc}}$ (ensemble) (Feng et al. 2025) | 0.6454 | 0.6460 | 0.5997 | 0.5554 |
| | $\text{LigUnity}_{\text{poc}}$ ($\text{mean}_{\text{std}}$) (Feng et al. 2025) | $0.5705_{0.1955}$ | $0.5774_{0.2097}$ | $0.5323_{0.1865}$ | $0.4994_{0.1773}$ |
| Ours | **HypSeek (ensemble)** | 0.7742 | 0.7819 | 0.6120 | 0.5447 |
| | **HypSeek ($\text{mean}_{\text{std}}$)** | $\mathbf{0.7186_{0.1157}}$ | $\mathbf{0.7239_{0.1321}}$ | $\mathbf{0.5606_{0.1738}}$ | $\mathbf{0.5034_{0.1739}}$ |

Table 2: Affinity ranking results on the JACS and MERCK benchmark datasets.

## 5.2 Quantitative Results

**Virtual Screening.** As shown in Table 1, HypSeek substantially outperforms all baselines across both DUD-E and LIT-PCBA. On DUD-E, HypSeek achieves an AUROC of 0.9435, improving over the next best method (LigUnity) by more than 5 points, and delivers a $\text{BEDROC}_{80.5}$ of 0.7892, nearly 0.14 higher than LigUnity. Its $\text{EF}_{1\%}$ of 51.44 is more than 20 points above the highest competing model, demonstrating exceptional early retrieval of actives. Similarly, on the more challenging LIT-PCBA benchmark, HypSeek attains the top AUROC (0.6210), the highest $\text{BEDROC}_{80.5}$ (0.1196), and an $\text{EF}_{1\%}$ of 6.81, consistently surpassing both docking-based and deep learning approaches. These results highlight HypSeek's superior ability to rank true binders early in the list, making it particularly well suited for high-throughput virtual screening applications.

**Affinity Ranking.** We evaluate HypSeek on the JACS and Merck datasets using five independent random seeds to assess both accuracy and robustness. We report two sets of our results: "ensemble," which averages the five models' predictions before computing metrics, and "$\text{mean}_{\text{std}}$," which gives the mean and standard deviation of Pearson's $r$ and Spearman's $\rho$ across the five runs. As shown in Table 2, on JACS HypSeek (ensemble) achieves Pearson $r = 0.7742$ and Spearman $\rho = 0.7819$, closely matching the physics-based FEP+ (Pearson $r = 0.7811$, Spearman $\rho = 0.7595$) and significantly outperforming all deep-learning baselines. On Merck, HypSeek (ensemble) attains Pearson $r = 0.6120$ and Spearman $\rho = 0.5447$, leading the non-physics methods. Moreover, HypSeek's standard deviations are lower than those reported for LigUnity's $\text{mean}_{\text{std}}$ results, indicating more consistent performance across random seeds.

| Setting | Module | | | | DUD-E (n = 102) | | JACS | |
|---|---|---|---|---|---|---|---|---|
| | $\mathcal{L}_{\text{cone}}$ | $R_{\text{ang}}$ | $R_{\text{het}}$ | Seq | $\text{BEDROC}_{80.5}$ | $\text{EF}_{1\%}$ | Pearson $r$ | Spearman $\rho$ |
| **Hyperbolic space** | | | | | | | | |
| Full model | ✓ | ✓ | ✓ | ✓ | 0.7892 | 51.44 | 0.7518 | 0.7580 |
| – no hyp | × | × | × | ✓ | 0.7671 | 49.14 | 0.6839 | 0.6906 |
| – no $R_{\text{ang}}$ | ✓ | × | ✓ | ✓ | 0.7856 | 50.52 | 0.7340 | 0.7529 |
| – no $R_{\text{het}}$ | ✓ | ✓ | × | ✓ | 0.7773 | 50.42 | 0.7047 | 0.7074 |
| – no Seq | ✓ | ✓ | ✓ | × | 0.7351 | 47.70 | 0.7194 | 0.7050 |
| **Euclidean space** | | | | | | | | |
| Contrastive + rank | × | × | × | × | 0.6565 | 42.87 | 0.5978 | 0.6060 |

Table 3: Ablation results on the DUD-E and JACS benchmarks.

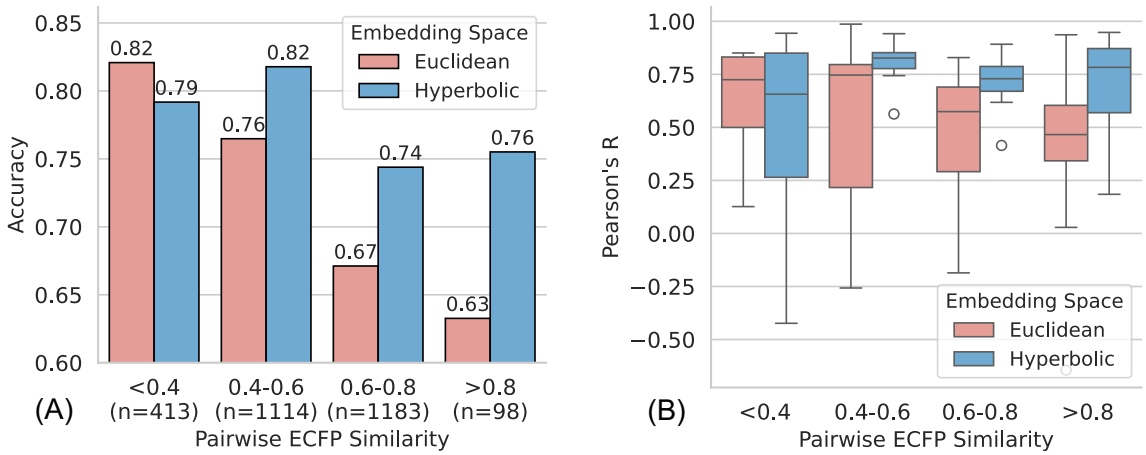


Figure 3: Pairwise analysis on the JACS benchmark. (A) Accuracy of affinity change prediction on ligand pairs with different ECFP4 similarity, comparing Euclidean and hyperbolic spaces; (B) Pearson's $R$ between predicted score difference and ground truth affinity gap.

### 5.3 Ablation and Analysis of HypSeek

**Impact of Key Components.** As summarised in Table 3, switching off hyperbolic–specific terms (no hyp) already degrades virtual–screening performance on DUD-E ($\text{BEDROC}_{80.5}$ drops from 0.7892 to 0.7671; $\text{EF}_{1\%}$ from 51.44 to 49.14), while the Euclidean baseline is markedly worse. The advantage becomes even more pronounced for affinity ranking on JACS, where Pearson $r$ falls from 0.7518 to 0.6839 without hyperbolic supervision and to 0.5978 in purely Euclidean space. In the affinity ranking task, due to limited computational resources, we conducted each ablation with a single random seed. Ablating either the angular or heterogeneity regulariser alone (no $R_{\text{ang}}$, no $R_{\text{het}}$) yields intermediate losses, confirming that both angle control and intra-assay weighting contribute complementary signals beyond the core cone loss. Finally, removing the protein sequence pathway (no Seq) also degrades performance, indicating that protein-sequence features serve mainly as an auxiliary signal that further shapes the embeddings.

**Pairwise Affinity Prediction.** Figure 3 (A)-(B) highlight how Euclidean and hyperbolic models behave across increasing ECFP4 (Rogers and Hahn 2010) similarity. Performance is comparable on dissimilar ligand pairs, but once the two molecules become structurally close, Euclidean accuracy and correlation deteriorate, whereas the hyperbolic model remains consistently strong. This suggests that the richer angular separation in hyperbolic space better captures subtle affinity shifts characteristic of activity cliffs, which are often compressed in Euclidean geometry.

## 6 Conclusion

We introduced HypSeek, a hyperbolic protein–ligand binding prediction model that embeds ligands, protein pockets, and sequences into a shared hyperbolic space using a three-tower architecture. By leveraging the negative curvature and exponential geometry of hyperbolic space, HypSeek captures both global interaction patterns and fine-grained affinity differences—especially in challenging cases like activity cliffs, where Euclidean embeddings often fail. Meanwhile, it retains efficient retrieval through inner product similarity, enabling large-scale virtual screening. Extensive experiments show that HypSeek consistently outperforms existing baselines across both screening and ranking tasks. HypSeek provides a geometry-aware solution for binding prediction.

## Acknowledgments

This work was supported by the National Key R&D Program of China No. 2025ZD1802501, the National Natural Science Foundation of China (NSFC) No. 62506193, and the Beijing Academy of Artificial Intelligence.

## References

Bdeir, A.; Schwethelm, K.; and Landwehr, N. 2024. Fully Hyperbolic Convolutional Neural Networks for Computer Vision. In *ICLR*.

Bécigneul, G.; and Ganea, O.-E. 2018. Riemannian adaptive optimization methods. *arXiv preprint arXiv:1810.00760*.

Brocidiacono, M.; Francoeur, P.; Aggarwal, R.; Popov, K.; Koes, D.; and Tropsha, A. 2022. BigBind: Learning from Nonstructural Data for Structure-Based Virtual Screening.

Cao, D.; Chen, G.; Jiang, J.; Yu, J.; Zhang, R.; Chen, M.; Zhang, W.; Chen, L.; Zhong, F.; Zhang, Y.; Lu, C.; Li, X.; Luo, X.; Zhang, S.; and Zheng, M. 2024. Generic protein–ligand interaction scoring by integrating physical prior knowledge and data augmentation modelling. *Nature Machine Intelligence*, 6(6): 688–700. Publisher: Nature Publishing Group.

Cao, Z.; Qin, T.; Liu, T.-Y.; Tsai, M.-F.; and Li, H. 2007. Learning to rank: from pairwise approach to listwise approach. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, 129–136. New York, NY, USA: Association for Computing Machinery. ISBN 9781595937933.

Chamberlain, B. P.; Clough, J.; and Deisenroth, M. P. 2017. Neural embeddings of graphs in hyperbolic space. *arXiv preprint arXiv:1705.10359*.

Desai, K.; Nickel, M.; Rajpurohit, T.; Johnson, J.; and Vedantam, S. R. 2023a. Hyperbolic image-text representations. 7694–7731. PMLR.

Desai, K.; Nickel, M.; Rajpurohit, T.; Johnson, J.; and Vedantam, S. R. 2023b. Hyperbolic Image-text Representations. In Krause, A.; Brunskill, E.; Cho, K.; Engelhardt, B.; Sabato, S.; and Scarlett, J., eds., *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, 7694–7731. PMLR.

Feng, B.; Liu, Z.; Yang, M.; Zou, J.; Cao, H.; Li, Y.; Zhang, L.; and Wang, S. 2025. A foundation model for protein-ligand affinity prediction through jointly optimizing virtual screening and hit-to-lead optimization. *bioRxiv*, 2025–02.

Feng, S.; Li, M.; Jia, Y.; Ma, W.-Y.; and Lan, Y. 2024. Protein-ligand binding representation learning from fine-grained interactions. In *The Twelfth International Conference on Learning Representations*.

Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; and Shenkin, P. S. 2004. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, 47(7): 1739–1749.

Ganea, O.; Bécigneul, G.; and Hofmann, T. 2018. Hyperbolic neural networks. In *NeurIPS*.

Gao, B.; Qiang, B.; Tan, H.; Jia, Y.; Ren, M.; Lu, M.; Liu, J.; Ma, W.-Y.; and Lan, Y. 2023. Drugclip: Contrastive protein-molecule representation learning for virtual screening. *Advances in Neural Information Processing Systems*, 36: 44595–44614.

Genheden, S.; and Ryde, U. 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery*, 10(5): 449–461.

Gilson, M. K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; and Chong, J. 2015. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Research*, 44(D1): D1045–D1053.

Gulcehre, C.; Denil, M.; Malinowski, M.; Razavi, A.; Pascanu, R.; Hermann, K. M.; Battaglia, P.; Bapst, V.; Raposo, D.; Santoro, A.; and de Freitas, N. 2019. Hyperbolic Attention Networks. In *ICLR*.

Han, J.; Hong, Y.; and Li, W.-J. 2025a. DrugHash: Hashing Based Contrastive Learning for Virtual Screening. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 17041–17049.

Han, J.; Hong, Y.; and Li, W.-J. 2025b. DrugHash: Hashing Based Contrastive Learning for Virtual Screening. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(16): 17041–17049.

Jiang, D.; Hsieh, C.-Y.; Wu, Z.; Kang, Y.; Wang, J.; Wang, E.; Liao, B.; Shen, C.; Xu, L.; Wu, J.; Cao, D.; and Hou, T. 2021. InteractionGraphNet: A Novel and Efficient Deep Graph Representation Learning Framework for Accurate Protein–Ligand Interaction Predictions. *Journal of Medicinal Chemistry*, 64(24): 18209–18232. Publisher: American Chemical Society.

Khrulkov, V.; Mirvakhabova, L.; Ustinova, E.; Oseledets, I.; and Lempitsky, V. 2020. Hyperbolic image embeddings. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6418–6428.

Kong, X.; Huang, W.; and Liu, Y. 2024. Generalist Equivariant Transformer Towards 3D Molecular Interaction Learning. In Salakhutdinov, R.; Kolter, Z.; Heller, K.; Weller, A.; Oliver, N.; Scarlett, J.; and Berkenkamp, F., eds., *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, 25149–25175. PMLR.

Le, M.; Roller, S.; Papaxanthos, L.; Kiela, D.; and Nickel, M. 2019. Inferring Concept Hierarchies from Text Corpora via Hyperbolic Embeddings. In Korhonen, A.; Traum, D. R.; and Màrquez, L., eds., *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, 3231–3241. Association for Computational Linguistics.

Lin, Y.-W. E.; Coifman, R. R.; Mishne, G.; and Talmon, R. 2023a. Hyperbolic diffusion embedding and distance for hierarchical representation learning. 21003–21025. PMLR.

Lin, Z.; Akin, H.; Rao, R. M.; Das, R.; Doshi, P.; Bepler, T.; et al. 2023b. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637): 1123–1130.

Liu, Z.; Su, M.; Han, L.; Liu, J.; Yang, Q.; Li, Y.; and Wang, R. 2017. Forging the basis for developing protein–ligand interaction scoring functions. *Accounts of chemical research*, 50(2): 302–309.

Lu, W.; Wu, Q.; Zhang, J.; Rao, J.; Li, C.; and Zheng, S. 2022. TANKBind: Trigonometry-Aware Neural NetworKs for Drug-Protein Binding Structure Prediction. *Advances in Neural Information Processing Systems*, 35: 7236–7249.

McNutt, A. T.; Francoeur, P.; Aggarwal, R.; Masuda, T.; Meli, R.; Ragoza, M.; Sunseri, J.; and Koes, D. R. 2021a. GNINA 1.0: molecular docking with deep learning. *Journal of Cheminformatics*, 13(1).

McNutt, A. T.; Francoeur, P.; Aggarwal, R.; Masuda, T.; Meli, R.; Ragoza, M.; Sunseri, J.; and Koes, D. R. 2021b. GNINA 1.0: molecular docking with deep learning. *Journal of cheminformatics*, 13(1): 1–20.

Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.; Segura-Cabrera, A.; Hersey, A.; and Leach, A. R. 2018. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Research*, 47(D1): D930–D940.

Mysinger, M. M.; Carchia, M.; Irwin, J. J.; and Shoichet, B. K. 2012. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *Journal of medicinal chemistry*, 55(14): 6582–6594.

Nickel, M.; and Kiela, D. 2017. Poincaré embeddings for learning hierarchical representations. 30.

Nickel, M.; and Kiela, D. 2018. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. 3779–3788. PMLR.

Öztürk, H.; Özgür, A.; and Ozkirimli, E. 2018. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics*, 34(17): i821–i829.

Passaro, S.; Corso, G.; Wohlwend, J.; Reveiz, M.; Thaler, S.; Somnath, V. R.; Getz, N.; Portnoi, T.; Roy, J.; Stark, H.; Kwabi-Addo, D.; Beaini, D.; Jaakkola, T.; and Barzilay, R. 2025. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. *bioRxiv*.

Poppi, T.; Kasarla, T.; Mettes, P.; Baraldi, L.; and Cucchiara, R. 2025. Hyperbolic Safety-Aware Vision-Language Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Rogers, D.; and Hahn, M. 2010. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*, 50(5): 742–754. PMID: 20426451.

Schindler, C. E.; Baumann, H.; Blum, A.; Bose, D.; Buchstaller, H.-P.; Burgdorf, L.; Cappel, D.; Chekler, E.; Czodrowski, P.; Dorsch, D.; et al. 2020. Large-scale assessment of binding free energy calculations in active drug discovery projects. *Journal of Chemical Information and Modeling*, 60(11): 5457–5474.

Shen, C.; Zhang, X.; Deng, Y.; Gao, J.; Wang, D.; Xu, L.; Pan, P.; Hou, T.; and Kang, Y. 2022. Boosting Protein-Ligand Binding Pose Prediction and Virtual Screening Based on Residue-Atom Distance Likelihood Potential and Graph Transformer. *Journal of Medicinal Chemistry*, 65(15): 10691–10706.

Shen, C.; Zhang, X.; Hsieh, C.-Y.; Deng, Y.; Wang, D.; Xu, L.; Wu, J.; Li, D.; Kang, Y.; Hou, T.; and Pan, P. 2023. A generalized protein-ligand scoring framework with balanced scoring, docking, ranking and screening powers. *Chemical Science*, 14(30): 8129–8146.

Spitzer, R.; and Jain, A. N. 2012. Surflex-Dock: Docking benchmarks and real-world application. *Journal of computer-aided molecular design*, 26: 687–699.

Tran-Nguyen, V.-K.; Jacquemard, C.; and Rognan, D. 2020. LIT-PCBA: An Unbiased Data Set for Machine Learning and Virtual Screening. *Journal of Chemical Information and Modeling*. Publisher: American Chemical Society.

Trott, O.; and Olson, A. J. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of computational chemistry*, 31(2): 455–461.

Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; et al. 2015. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *Journal of the American Chemical Society*, 137(7): 2695–2703.

Yang, Z.; Zhong, W.; Lv, Q.; Dong, T.; Chen, G.; and Chen, C. Y.-C. 2024. Interaction-based inductive bias in graph neural networks: enhancing protein-ligand binding affinity predictions from 3d structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Yu, J.; Li, Z.; Chen, G.; Kong, X.; Hu, J.; Wang, D.; Cao, D.; Li, Y.; Huo, R.; Wang, G.; et al. 2023. Computing the relative binding affinity of ligands based on a pairwise binding comparison network. *Nature Computational Science*, 3(10): 860–872.

Zhang, X.; Gao, H.; Wang, H.; Chen, Z.; Zhang, Z.; Chen, X.; Li, Y.; Qi, Y.; and Wang, R. 2024. PLANET: A Multi-objective Graph Neural Network Model for Protein–Ligand Binding Affinity Prediction. *Journal of Chemical Information and Modeling*, 64(7): 2205–2220. Publisher: American Chemical Society.

Zhang, X.; Zhang, O.; Shen, C.; Qu, W.; Chen, S.; Cao, H.; Kang, Y.; Wang, Z.; Wang, E.; Zhang, J.; Deng, Y.; Liu, F.; Wang, T.; Du, H.; Wang, L.; Pan, P.; Chen, G.; Hsieh, C.-Y.; and Hou, T. 2023. Efficient and accurate large library ligand docking with KarmaDock. *Nature Computational Science*, 3(9): 789–804. Publisher: Nature Publishing Group.