DRUGPRISM: TARGET-CONDITIONED RIEMANNIAN METRICS ON MIXED-CURVATURE PRODUCT SPACES FOR PROTEIN–LIGAND AFFINITY RANKING

Anonymous authors
Paper under double-blind review

ABSTRACT

Recent spherical and hyperbolic affinity ranking models exhibit complementary performance across protein targets, suggesting that a fixed comparison geometry may be restrictive. We show that target specific geometry alone does not guarantee ranking adaptation, as single geometry rescaling changes intrinsic curvature but preserves rankings for fixed representations. This motivates relative scaling across multiple geometric components. We introduce DrugPrism, which maps protein and ligand features into a shared product manifold with spherical, Euclidean, and hyperbolic components, while a pocket conditioned metric head predicts target specific coefficients that rescale the component metrics. Under the resulting product metric, the squared geodesic distance is exactly a weighted sum of the component squared distances, allowing relative weighting to alter rankings when the components disagree. DrugPrism improves mean Spearman correlation over HypSeek by 13.8% on Merck and 19.7% on Hermite. Controlled ablations show that target conditioned weighting substantially outperforms both uniform product weighting and all single geometry variants. On an unseen OpenBind target, DrugPrism with only five tuning compounds already surpasses LigUnity and HypSeek with 25 compounds. The learned metric weights further reveal clear protein family level structure, with related proteins tending to share similar geometric preferences.

1 INTRODUCTION

Accurately ranking ligands by their binding affinity to a protein target is central to structure based drug discovery, since reliable prioritization of candidate molecules can reduce experimental cost and accelerate lead optimization. Physics based methods like FEP+ (Wang et al., 2015) can achieve high accuracy but remain computationally expensive at scale, while deep learning offers a more scalable alternative with recent gains driven by improved molecular representations and affinity aware supervision (Gao et al., 2023; Passaro et al., 2025).

Recent affinity ranking models suggest that representation geometry matters, yet no single geometry is uniformly preferable across targets. LigUnity (Feng et al., 2025) ranks normalized representations by cosine similarity, corresponding to angular distance on the unit hypersphere, whereas HypSeek (Wang et al., 2026) uses a Lorentz hyperbolic representation. As shown in Figure 1, their relative performance reverses across Merck targets, and although this is not a controlled comparison of geometry alone, the variation suggests that a fixed comparison geometry may be restrictive.

A natural response to this target wise variation is to make the comparison geometry target specific. Within a Riemannian formulation, the simplest realization is to rescale the metric of a single geometry for each target. This changes its intrinsic curvature scale, but for fixed representations uniformly scales all distances and therefore leaves ligand rankings unchanged. Target specific geometry therefore does not guarantee target specific ranking. We instead introduce multiple geometric components whose relative contributions vary across targets, yielding a basis expansion in which shared component manifolds serve as geometric primitives and target specific coefficients determine their combination.

Building on this formulation, we introduce DrugPrism, which maps protein and ligand features into a shared product of spherical, Euclidean, and hyperbolic manifolds through geometry specific projection heads. A pocket conditioned metric head predicts normalized target specific coefficients

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.

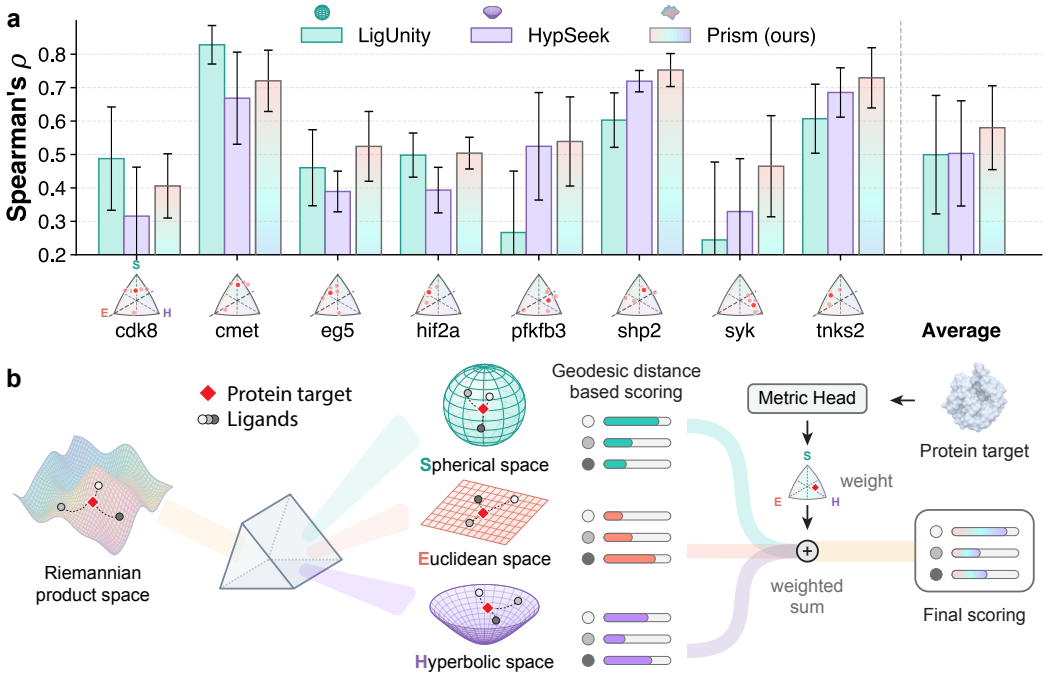


Figure 1: **Motivation and overview of DrugPrism.** (a) Spherical and hyperbolic models exhibit complementary performance across targets, revealing the limitation of a single fixed geometry. DrugPrism learns target specific mixtures whose simplex weights broadly align with these preferences and achieve better average ranking performance than either single geometry model. (b) DrugPrism represents proteins and ligands in a spherical, Euclidean, and hyperbolic product space, with a pocket conditioned metric head weighting the component geodesic distances for final scoring.

that rescale the component metrics and define a target specific product metric. Under this metric, the squared geodesic distance is exactly the weighted sum of the component squared distances, so varying the relative coefficients can alter ligand rankings when the components disagree.

We evaluate DrugPrism on JACS, Merck, and Hermite, where it improves mean Spearman correlation over HypSeek by 13.8% on Merck and 19.7% on Hermite, with over five orders of magnitude less compute than FEP+ on JACS. Controlled ablations on Merck show that target conditioned weighting outperforms both uniform weighting and all single geometry variants, confirming that the gains arise from adapting relative geometric contributions rather than simply combining multiple geometries. On an unseen OpenBind target, DrugPrism with only five tuning compounds surpasses LigUnity and HypSeek with 25 compounds. Finally, the learned metric weights show clear structure at the protein family level, with related proteins tending to share similar geometric preferences.

Our main contributions are threefold.

- We establish that target specific geometry does not guarantee ranking adaptation. Rescaling a single geometry changes intrinsic curvature but preserves rankings for fixed representations, whereas relative scaling across mixed curvature components enables ranking adaptation.
- We introduce DrugPrism, which maps protein and ligand features into a mixed curvature product manifold and predicts pocket conditioned metric coefficients that adapt the relative contributions of its components to each target.
- Extensive experiments show that DrugPrism improves affinity ranking across diverse targets, supports efficient adaptation to an unseen target with limited affinity data, and learns geometric preferences with clear protein family level structure.

2 Related Work

Riemannian representation learning. Riemannian manifolds have long provided structured non-Euclidean representation spaces in computer vision. Grassmann and symmetric positive definite manifolds have been used for video and image-set recognition, where geometry-aware distances and metric learning replace Euclidean comparisons (Huang et al., 2015; Huang & Van Gool, 2017). Riemannian representations have also been explored directly for retrieval: Li et al. match Euclidean image queries with manifold-valued video representations, while more recent hyperbolic metric learning improves image retrieval with vision transformers (Li et al., 2015; Ermolov et al., 2022). Beyond single manifolds, mixed-curvature product spaces combine spherical, Euclidean, and hyperbolic factors to model heterogeneous relational structure (Gu et al., 2019; Skopek et al., 2020).

Affinity ranking. Accurate ranking of ligand binding affinities is essential for drug discovery but remains computationally challenging. Physics-based methods such as FEP+ (Wang et al., 2015) and MM-GB/SA (Genheden & Ryde, 2015) achieve high accuracy through alchemical free-energy calculations and implicit-solvent models, respectively, but require expensive conformational sampling. Recent deep learning approaches instead improve affinity ranking through increasingly expressive protein–ligand representation learning. PBCNet (Yu et al., 2023) models pairwise ligand differences with graph neural networks, EHIGN (Yang et al., 2024) represents protein–ligand interactions using heterogeneous graphs, LigUnity (Feng et al., 2025) combines contrastive retrieval with listwise ranking for joint global screening and local prioritization, and HypSeek (Wang et al., 2026) further introduces Lorentz hyperbolic representations with geometric supervision for fine-grained affinity ranking. Despite their architectural differences, these methods all improve representation learning while relying on a shared scoring metric across protein targets.

3 Notations and Riemannian Preliminaries

DrugPrism adapts affinity ranking by changing the Riemannian metric used to compare latent protein and ligand representations. We therefore introduce the geometric notation needed to formalize metric-dependent distances, before specializing to the spherical, Euclidean, and hyperbolic components used in our product space.

Manifolds, tangent spaces, and metrics. A smooth d-manifold $\mathcal{M}$ has at each point $x \in \mathcal{M}$ a d-dimensional tangent space $T_x\mathcal{M}$, which can be viewed as the vector space of curve velocities at x. A Riemannian metric g assigns a smoothly varying positive-definite inner product g_x to each tangent space, inducing the norm $\|v\|_{g_x}^2 = g_x(v, v)$.

From local geometry to distance. For a piecewise smooth curve $\gamma : [0, 1] \to \mathcal{M}$, define

$$L_g(\gamma) = \int_0^1 \|\dot\gamma(u)\|_{g_{\gamma(u)}}\, du, \qquad E_g(\gamma) = \int_0^1 \|\dot\gamma(u)\|_{g_{\gamma(u)}}^2\, du, \qquad d_g(x, y) = \inf_{\gamma: x\to y} L_g(\gamma). \tag{1}$$

The exponential map $\mathrm{Exp}_x(v)$ follows the geodesic starting from x with initial velocity v. In a normal neighborhood, its minimizing inverse satisfies $\mathrm{Log}_x(y) \in T_x\mathcal{M}$ and $\|\mathrm{Log}_x(y)\|_{g_x} = d_g(x, y)$. On a complete connected manifold, we have $\inf_{\gamma:x\to y} E_g(\gamma) = d_g(x, y)^2$ (Lee, 2018).

Canonical Riemannian geometries. We write $\mathcal{M}_\kappa^d$ for the standard complete, simply connected geometries of constant sectional curvature $\kappa \in \{+1, 0, -1\}$, with $d \geq 2$. Curvature here describes the intrinsic geometry, not a measured property of a protein. Their concrete sphere, Euclidean, and Lorentz models are

$$\begin{aligned} &\mathbb{S}^d = \{x \in \mathbb{R}^{d+1} : x^\top x = 1\}, && T_x\mathbb{S}^d = \{v : x^\top v = 0\}, \\ &\mathbb{R}^d && T_x\mathbb{R}^d \simeq \mathbb{R}^d, \\ &\mathbb{H}^d = \{x \in \mathbb{R}^{d+1} : \langle x, x\rangle_L = -1,\ x_0 > 0\}, && T_x\mathbb{H}^d = \{v : \langle x, v\rangle_L = 0\}, \end{aligned} \tag{2}$$

where $\langle x, y\rangle_L = -x_0y_0 + \sum_{i=1}^d x_iy_i$. The spherical and Euclidean metrics restrict the ordinary inner product; the hyperbolic metric restricts the Lorentz form, which is positive definite on $T_x\mathbb{H}^d$. Their distances are respectively $\arccos(x^\top y)$, $\|x - y\|_2$, and $\mathrm{arcosh}(-\langle x, y\rangle_L)$.

4 DRUGPRISM

4.1 TARGET-SPECIFIC GEOMETRY DOES NOT GUARANTEE TARGET-SPECIFIC RANKING

The target-wise variation observed in Figure 1 suggests that a shared comparison geometry may be restrictive. A natural response is therefore to make the geometry itself target-specific. Within a single canonical geometry, the simplest way to achieve this is to rescale its Riemannian metric for each target. Let $(\mathcal{M}_\kappa^d, g_\kappa)$ denote a canonical Riemannian geometry of constant sectional curvature $\kappa \in \{+1, 0, -1\}$. A natural form of target personalization is to preserve the underlying manifold while allowing each target t to rescale its metric:

$$g_t = \lambda_t g_\kappa, \qquad \lambda_t > 0. \tag{3}$$

This operation has an intrinsic geometric interpretation. Scaling a Riemannian metric changes not only the numerical scale of distances, but also the curvature scale of the resulting geometry.

Proposition 4.1 (Geometry under constant metric rescaling)**.** *Let $(\mathcal{M}, g)$ be a Riemannian manifold and let $\widetilde{g} = \lambda g$ for a constant $\lambda > 0$. Then*

$$d_{\widetilde{g}}(x, y) = \sqrt{\lambda}\, d_g(x, y), \qquad K_{\widetilde{g}}(\sigma) = \frac{1}{\lambda} K_g(\sigma) \tag{4}$$

for every $x, y \in \mathcal{M}$ and every tangent 2-plane σ. In particular, if g_κ has constant sectional curvature κ, then

$$K_t = \frac{\kappa}{\lambda_t}. \tag{5}$$

The proof is given in Appendix A. Proposition 4.1 provides a useful change of viewpoint: target-specific metric rescaling can equivalently be viewed as target-specific adaptation of the intrinsic curvature scale. For spherical and hyperbolic geometry, λ_t changes the curvature magnitude while preserving its sign; Euclidean geometry remains flat, indicating that this construction indeed makes the intrinsic geometry target-specific.

However, this geometric flexibility does not translate into ranking flexibility.

Theorem 4.2 (Rank invariance under single-geometry rescaling)**.** *For a fixed target representation p_t and ligand representations $\{q_\ell\}$, let $g_t = \lambda_t g_\kappa$ with $\lambda_t > 0$. Then, for any two ligands a and b,*

$$d_{g_t}(p_t, q_a)^2 < d_{g_t}(p_t, q_b)^2 \iff d_{g_\kappa}(p_t, q_a)^2 < d_{g_\kappa}(p_t, q_b)^2. \tag{6}$$

Hence target-specific scalar rescaling of a single geometry cannot change the induced ligand ranking.

Thus, a single geometry may be personalized in its geometry, yet this personalization provides no additional degree of freedom for within-target ranking. What is missing is not target dependence itself, but *relative* geometric degrees of freedom: different geometric contributions must be adjustable against one another so that their competition can alter pairwise ligand orderings. This motivates the mixed-curvature product construction developed next.

4.2 RELATIVE GEOMETRIC SCALING ENABLES RANKING ADAPTATION

Section 4.1 shows that a single target-specific scale can change the intrinsic geometry but cannot alter within-target rankings. We therefore introduce relative geometric degrees of freedom through a product manifold and equip it with a component-wise rescaled metric:

$$\mathcal{M} = \prod_{j=1}^{m} \mathcal{M}_j, \qquad g_w = \bigoplus_{j=1}^{m} w_j g_j, \quad w_j > 0, \tag{7}$$

where $(\mathcal{M}_j, g_j)$ are canonical constant-curvature components, instantiated in DrugPrism by spherical, Euclidean, and hyperbolic factors. Within component j, the rescaling $w_j g_j$ changes its curvature scale from κ_j to κ_j / w_j, while the component-wise coefficients allow their relative geometric scales to vary independently.

Since $g_{cw} = c\, g_w$ is rank-equivalent to g_w for any $c > 0$, only relative scales matter. We therefore normalize

$$\alpha_j = \frac{w_j}{\sum_k w_k}, \qquad \alpha \in \Delta^\circ := \{\alpha > 0 : \mathbf{1}^\top \alpha = 1\}. \tag{8}$$

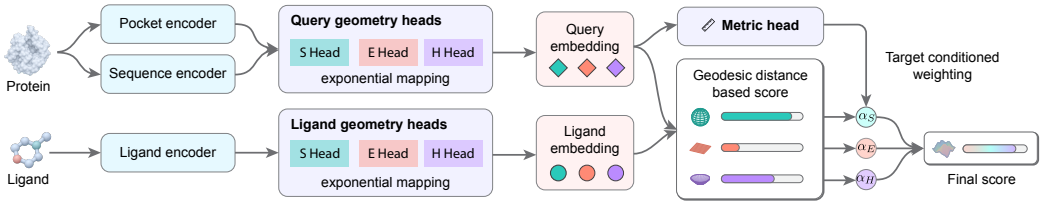


Figure 2: **DrugPrism architecture.** Protein and ligand features are mapped to spherical (S), Euclidean (E), and hyperbolic (H) embeddings. A pocket-conditioned metric head predicts weights that combine the three component geodesic distances into the final ranking score.

To connect this metric construction to an explicit comparison rule, we first characterize the geodesic distance induced by the product metric.

Theorem 4.3 (Exact geodesic realization of the product metric)**.** *Let $(\mathcal{M}_j, g_j)$, $j = 1, \ldots, m$, be complete connected Riemannian manifolds with geodesic distances d_j, and equip $\mathcal{M} = \prod_{j=1}^{m} \mathcal{M}_j$ with $g_w = \bigoplus_{j=1}^{m} w_j g_j$, $w_j > 0$. Then, for any $p, q \in \mathcal{M}$,*

$$d_{g_w}(p,q)^2 = \inf_{\gamma: p \to q} E_{g_w}(\gamma) = \sum_{j=1}^{m} w_j d_j(p_j, q_j)^2. \tag{9}$$

Thus the normalized coefficients α parameterize a genuine family of Riemannian comparison metrics. We therefore rank candidates by the negative squared geodesic distance

$$s_\alpha(p,q) := -d_{g_\alpha}(p,q)^2 = -\sum_{j=1}^{m} \alpha_j d_j(p_j, q_j)^2. \tag{10}$$

With the product-geodesic distance established, we next characterize the ranking expressivity of relative product scaling.

Theorem 4.4 (Ranking expressivity of relative product scaling)**.** *For two ligands a, b, define $c_{a,b,j} := d_j(p_j, q_{b,j})^2 - d_j(p_j, q_{a,j})^2$. Then $s_\alpha(p, q_a) - s_\alpha(p, q_b) = \alpha^\top c_{a,b}$. Over the closed simplex $\Delta = \{\alpha \geq 0 : \mathbf{1}^\top \alpha = 1\}$,*

$$\{\alpha^\top c_{a,b} : \alpha \in \Delta\} = [\min_j c_{a,b,j},\ \max_j c_{a,b,j}]. \tag{11}$$

Hence two strictly positive product metrics can induce opposite rankings of a and b if and only if

$$\min_j c_{a,b,j} < 0 < \max_j c_{a,b,j}. \tag{12}$$

Theorem 4.4 shows that relative product scaling gains ranking expressivity precisely when the component geometries disagree on a ligand pair: varying α can move the pairwise margin across zero, with $\alpha^\top c_{a,b} = 0$ defining a ranking boundary in the metric simplex. Because heterogeneous targets may impose incompatible ranking constraints, a single shared point in this simplex need not serve all targets simultaneously; a convex-geometric characterization is provided in the Appendix. DrugPrism therefore makes the metric coordinate target-dependent and learns α_t directly from each target, which we address next.

4.3 DrugPrism Architecture

DrugPrism instantiates the product construction of Section 4.2 on the mixed-curvature space

$$\mathcal{M} = \mathbb{S}^{d_S} \times \mathbb{R}^{d_E} \times \mathbb{H}^{d_H}, \tag{13}$$

where the spherical, Euclidean, and hyperbolic factors have intrinsic dimensions d_S, d_E, and d_H, with canonical metrics g_S, g_E, and g_H, respectively. For relative metric coordinates $\alpha = (\alpha_S, \alpha_E, \alpha_H) \in \Delta^\circ$, the corresponding product metric is $g_\alpha = \alpha_S g_S \oplus \alpha_E g_E \oplus \alpha_H g_H$. Thus, if component j has canonical sectional curvature κ_j, its curvature scale under g_α becomes κ_j / α_j.

Mixed-curvature product representation. Following the shared backbone of HypSeek, the pocket, protein-sequence, and ligand encoders produce Euclidean representations e_t^P, e_t^F, and e_l^L, respectively, where P, F, and L denote the pocket, sequence, and ligand branches. For each component $j \in \{S, E, H\}$, geometry-specific projection MLPs map these backbone representations to the tangent space at a fixed reference point $o_j \in \mathcal{M}_j$:

$$\begin{aligned} z_{t,j}^b &= f_j^b(e_t^b) \in \mathbb{R}^{d_j} \simeq T_{o_j}\mathcal{M}_j, \quad x_{t,j}^b = \mathrm{Exp}_{o_j}^j(z_{t,j}^b), \quad b \in \{P, F\}, \\ z_{l,j}^L &= f_j^L(e_l^L) \in \mathbb{R}^{d_j} \simeq T_{o_j}\mathcal{M}_j, \quad x_{l,j}^L = \mathrm{Exp}_{o_j}^j(z_{l,j}^L). \end{aligned} \tag{14}$$

The resulting query and ligand representations are

$$p_t^b = (x_{t,S}^b, x_{t,E}^b, x_{t,H}^b), \qquad q_l = (x_{l,S}^L, x_{l,E}^L, x_{l,H}^L) \in \mathcal{M}, \qquad b \in \{P, F\}. \tag{15}$$

Target-conditioned metric parameterization. DrugPrism predicts the relative metric coordinates for target t from the factor-specific pocket representations using a lightweight factor-aware metric head,

$$\hat{\alpha}_t = h_\theta\left(z_{t,S}^P, z_{t,E}^P, z_{t,H}^P\right) = (\hat{\alpha}_{t,S}, \hat{\alpha}_{t,E}, \hat{\alpha}_{t,H}) \in \Delta^\circ. \tag{16}$$

The predicted coefficients instantiate the target-conditioned metric $g_{\hat{\alpha}_t}$ on the shared product manifold.

The representation maps and metric predictor are trained jointly during pretraining. Consequently, for an unseen target, DrugPrism can infer its comparison metric directly from the pocket representation without target-specific affinity fitting. The resulting $\widehat{\alpha}_t$ is then used by the product-geodesic ranking score throughout training and inference, as described in Section 4.4. Architectural details, component dimensions, parameterizations, and optimization hyperparameters are provided in Appendix B.

4.4 Training Objectives and Inference

Target-conditioned affinity scoring. For query branch $b \in \{P, F\}$ and ligand l, DrugPrism scores their product-manifold representations using the target-conditioned metric predicted from the pocket:

$$s_{t,l}^b = -d_{g_{\hat{\alpha}_t}}(p_t^b, q_l)^2 = -\sum_{j \in \{S,E,H\}} \hat{\alpha}_{t,j} d_j\left(x_{t,j}^b, x_{l,j}^L\right)^2, \qquad Z_{t,l}^b = \lambda_{\mathrm{logit}} s_{t,l}^b, \tag{17}$$

where $\lambda_{\mathrm{logit}} > 0$ is a learned logit scale.

Contrastive alignment and listwise ranking. For each query branch $b \in \{P, F\}$, we retain the bidirectional assay-level contrastive loss $\mathcal{L}_{\mathrm{cont}}^b$ and the Plackett–Luce listwise ranking loss $\mathcal{L}_{\mathrm{rank}}^b$ from HypSeek (Wang et al., 2026), replacing its original similarity score with the target-conditioned logits in equation 17. Their exact formulations and masking rules are provided in Appendix C.

Geometry-aware regularization. We retain the radial–angular geometric supervision of HypSeek (Wang et al., 2026), while evaluating the corresponding norms and angles under the target-conditioned product metric $g_{\hat{\alpha}_t}$. This gives the geometry-aware loss $\mathcal{L}_{\mathrm{geom}}$; the angular-collapse regularizer R_{ang} and intra-assay heterogeneity regularizer R_{het} are retained in the same form. A coefficient-spectrum regularizer R_{spec} additionally stabilizes the predicted metric coordinates. Exact definitions are provided in Appendix C.

Overall objective. Combining the two query-branch affinity objectives with the geometry-aware and metric regularization terms, the complete training objective is

$$\begin{aligned} \mathcal{L}_{\mathrm{total}} = \Big[w_P\big(\mathcal{L}_{\mathrm{cont}}^P + \lambda_{\mathrm{rank}}\mathcal{L}_{\mathrm{rank}}^P\big) + w_F\big(\mathcal{L}_{\mathrm{cont}}^F + \lambda_{\mathrm{rank}}\mathcal{L}_{\mathrm{rank}}^F\big) \Big] \\ + \gamma_{\mathrm{geom}}\mathcal{L}_{\mathrm{geom}} + \lambda_{\mathrm{ang}}R_{\mathrm{ang}} + \lambda_{\mathrm{het}}R_{\mathrm{het}} + \lambda_{\mathrm{spec}}R_{\mathrm{spec}}. \end{aligned} \tag{18}$$

5 Experiments

5.1 Experimental Setup

Implementation Details. We adopt the same curated assay-level training dataset as HypSeek (Wang et al., 2026) and LigUnity (Feng et al., 2025), which is constructed from ChEMBL (Mendez et al.,

Table 1: Affinity ranking results on the JACS, MERCK, and Hermite benchmark datasets.

Type	Method	JACS		Merck		Hermite	
		Pearson r	Spearman ρ	Pearson r	Spearman ρ	Pearson r	Spearman ρ
Physics	FEP+	*0.7811*	*0.7595*	*0.6960*	*0.6798*	*0.8804*	*0.8720*
	MM-GB/SA	0.1489	0.2011	0.1299	0.1299	0.2577	0.2673
DL	PBCNet	0.3939	0.3799	0.4058	0.4075	0.4486	0.4899
	EHIGN	0.5787	0.5814	0.4246	0.3830	-0.0359	0.0535
	GET	0.4034	0.3753	0.4203	0.4214	0.2241	0.3044
	BindNet	0.5481	0.5368	0.4037	0.3477	-0.2069	-0.1896
	Boltz-2	0.5231	0.5285	0.4298	0.4013	0.4903	0.4000
	LigUnity$_{\text{poc}}$ (ensemble)	0.6454	0.6460	0.5997	0.5554	0.6797	0.6543
	LigUnity$_{\text{poc}}$ (mean)	0.5705	0.5774	0.5323	0.4994	0.6047	0.6008
		$\pm$ 0.1955	$\pm$ 0.2097	$\pm$ 0.1865	$\pm$ 0.1773	$\pm$ 0.0554	$\pm$ 0.0321
	HypSeek (ensemble)	**0.7742**	**0.7819**	0.6120	0.5447	0.6612	0.5929
	HypSeek (mean)	**0.7186**	**0.7239**	0.5606	0.5034	0.5978	0.5338
		$\pm$ **0.1157**	$\pm$ **0.1321**	$\pm$ 0.1738	$\pm$ 0.1739	$\pm$ 0.0559	$\pm$ 0.0441
Ours	**DrugPrism (ensemble)**	0.7712	0.7654	**0.6592**	**0.6113**	**0.7806**	**0.6906**
	DrugPrism (mean)	0.7171	0.7131	**0.6074**	**0.5731**	**0.7082**	**0.6392**
		$\pm$ 0.0271	$\pm$ 0.0266	$\pm$ **0.0257**	$\pm$ **0.0441**	$\pm$ **0.0132**	$\pm$ **0.0358**

2019), BindingDB (Gilson et al., 2016), and PDBBind (Liu et al., 2017). DrugPrism shares the same three-tower backbone and pretrained initialization as HypSeek, consisting of Uni-Mol pocket and ligand encoders together with an ESM-2 protein sequence encoder.

Baselines and Evaluation Protocol. We compare DrugPrism with representative physics-based and learning-based baselines, including FEP+ (Wang et al., 2015), MM-GB/SA (Genheden & Ryde, 2015), PBCNet (Yu et al., 2023), EHIGN (Yang et al., 2024), GET (Kong et al., 2023), BindNet (Feng et al., 2023), Boltz-2 (Passaro et al., 2025), LigUnity (Feng et al., 2025), and HypSeek (Wang et al., 2026). Affinity-ranking performance is evaluated using Pearson's r and Spearman's ρ, which are computed independently for each target and macro-averaged across targets within each test dataset. DrugPrism, LigUnity, and HypSeek are trained with five independent random seeds under the same evaluation protocol. For these methods, we report both the mean and standard deviation across individual runs and an ensemble obtained by averaging predictions from the five models. Results for the remaining baselines are taken from their corresponding benchmark evaluations.

5.2 Affinity Ranking Across Diverse Targets

Benchmarks. We evaluate DrugPrism on three complementary affinity-ranking benchmarks. JACS (Wang et al., 2015) focuses on fine-grained ranking within congeneric lead-optimization series, Merck (Schindler et al., 2020) evaluates larger and more chemically diverse industrial lead-optimization tasks, and Hermite (Zou et al., 2025) extends the evaluation to a broader range of heterogeneous proteins.

Overall performance. Across the three benchmarks in Table 1, DrugPrism remains competitive with HypSeek on JACS while showing substantially larger gains on the broader Merck and Hermite benchmarks, improving the mean Spearman correlation over HypSeek by 13.8% and 19.7%, respectively. Beyond predictive performance, DrugPrism offers a substantial practical advantage. On JACS, it reaches FEP+-level ranking accuracy while requiring more than five orders of magnitude less compute. It also substantially outperforms the cofolding-based Boltz-2 across all three benchmarks while being over three orders of magnitude faster under the measured JACS protocol.

Target-wise analysis. Figure 1a shows that affinity-ranking performance exhibits strong target-wise reversals: LigUnity performs better on some Merck targets, whereas HypSeek is better on others, despite both being trained from the same assay-level data source. This suggests that target-specific variation is not fully explained by generic data generalization, but is closely tied to the comparison geometry used for ranking. DrugPrism addresses this issue by learning target-conditioned mixtures over spherical, Euclidean, and hyperbolic components. As a result, it substantially raises the per-target

Table 2: Ablations of geometric weighting and training objectives on Merck.

Geometric weighting			Training objective		
Weighting	**Pearson r**	**Spearman ρ**	**Variant**	**Pearson r**	**Spearman ρ**
Target-conditioned	**0.6379**	**0.6326**	Full model	**0.6379**	**0.6326**
Uniform $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	0.5341	0.4542	w/o $\mathcal{L}_{\text{geom}}$	0.5155	0.4919
Spherical $(1, 0, 0)$	0.5397	0.5039	w/o $\mathcal{R}_{\text{ang}}$	0.5614	0.5396
Euclidean $(0, 1, 0)$	0.4966	0.4655	w/o $\mathcal{R}_{\text{het}}$	0.5798	0.5523
Hyperbolic $(0, 0, 1)$	0.5425	0.5033	w/o $\mathcal{R}_{\text{spec}}$	0.5551	0.5287

floor, consistently achieving either first- or second-best performance across the Merck targets while also improving the overall average. Importantly, the learned metric weights qualitatively agree with the target preferences exposed by the fixed-geometry baselines, such as a more spherical weighting for CDK8 and a more hyperbolic weighting for PFKFB3, providing empirical support for target-specific geometric preference.

5.3 Few-Shot Adaptation

Adaptation protocol. We evaluate few-shot adaptation on the OpenBind EV-A71 2A protease affinity benchmark (OpenBind, 2026), which contains 494 compounds after quality control. We reserve 25 compounds for adaptation and the remaining 469 for evaluation. For all three methods, the Uni-Mol encoders are kept frozen and only the post-encoder MLP projection layers are fine-tuned. For each tuning size $n \in \{2, 3, 5, 8, 10, 15, 20, 25\}$, we sample 100 subsets of size n from the adaptation pool, fine-tune on each subset, and report the median evaluation-set Spearman correlation.

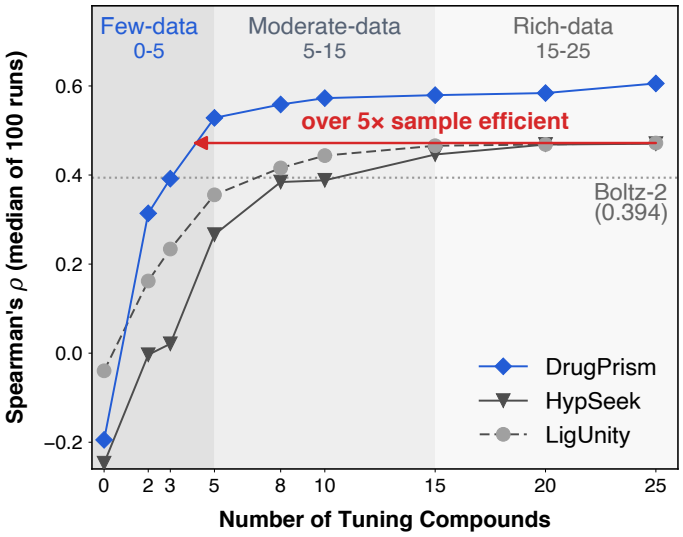


Figure 3: Few-shot adaptation on the OpenBind target.

Results. DrugPrism adapts rapidly with only a few target specific affinity labels. As shown in Figure 3, its Spearman correlation reaches the Boltz 2 reference with three tuning compounds and increases to approximately 0.53 with five. More importantly, DrugPrism shows substantially higher sample efficiency than LigUnity and HypSeek. With only five tuning compounds, it already exceeds the performance achieved by either baseline with 25 compounds. Since all three methods use frozen Uni Mol encoders and fine tune only the post encoder MLP projection layers, this result supports the effectiveness of target conditioned geometry for adaptation under limited supervision.

5.4 Ablation Studies.

Table 2 shows that the main gain comes from target conditioned geometric weighting rather than simply combining multiple geometries. Replacing the learned target specific weights with uniform weights reduces Spearman correlation from 0.6326 to 0.4542, while all three single geometry variants remain below 0.504. Notably, the uniformly weighted product is even weaker than the best single geometry variants, showing that a mixed curvature product space alone is insufficient and that adapting the relative geometric contributions to each target is essential. The training objective ablations further show that all geometric supervision and regularization terms contribute to performance. Removing L_{geom} causes the largest degradation to 0.4919, while removing R_{ang}, R_{het}, or R_{spec} also results in substantial performance drops.

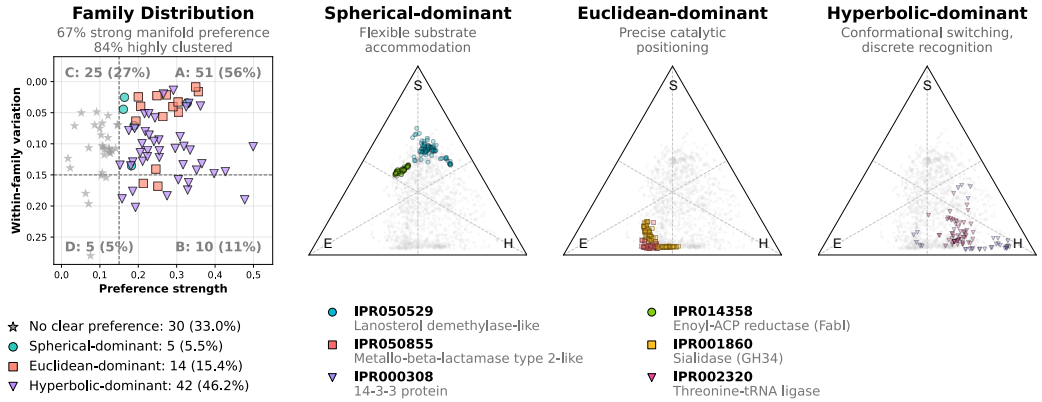


Figure 4: **Family-level geometry preferences learned by DrugPrism.** Left, each point represents an InterPro family, with preference strength indicating how strongly one geometry dominates the family-averaged weights and within-family variation indicating how consistent the weights are across proteins in that family. Right, ternary plots show protein-level spherical, Euclidean, and hyperbolic weights. Gray points denote all proteins, while colored points highlight representative families with spherical-, Euclidean-, or hyperbolic-dominant preferences.

5.5 Protein Families Exhibit Distinct Geometry Preferences

Setting. We investigate whether the geometry weights learned by DrugPrism exhibit systematic patterns at the protein family level. We identify the InterPro family of each protein target in the training set from its UniProt linked InterPro annotations and retain families with sufficient coverage in the model outputs, requiring at least 20 records, 10 structures, and 5 distinct proteins. This results in 91 families. We aggregate the spherical, Euclidean, and hyperbolic weights at the protein and family levels and analyze their preference strength and within family variation.

Results. For each family, preference strength measures how strongly one geometry dominates the family averaged weights, while within family variation measures how consistently this preference is shared across member proteins. Using these two statistics, Figure 4 shows that the geometry weights learned by DrugPrism exhibit clear structure at the protein family level. Among the 91 families, 67% exhibit a strong geometric preference and 84% show low within family variation, indicating that related proteins often share similar geometry preferences. Notably, 69% of families with a clear preference favor the hyperbolic component, consistent with the gains obtained by introducing hyperbolic geometry into affinity ranking in HypSeek. Representative families further suggest geometry specific biological patterns. Hyperbolic dominant families such as 14-3-3 and threonyl-tRNA ligase involve selective recognition among related motifs or substrates, consistent with the hierarchical and expanding structure of hyperbolic space. Spherical dominant CYP51-like and FabI families are instead associated with flexible substrate accommodation, while Euclidean dominant metallo-beta-lactamase and sialidase families emphasize precise local positioning.

6 Conclusion

We introduced DrugPrism, a target conditioned Riemannian metric framework for protein ligand affinity ranking on a mixed curvature product manifold. Our analysis shows that target specific geometry alone does not guarantee ranking adaptation, whereas relative scaling across mixed curvature components can alter ligand rankings. DrugPrism realizes this principle through a pocket conditioned metric that adapts the relative contributions of spherical, Euclidean, and hyperbolic components, leading to improved ranking performance on Merck and Hermite, efficient few shot adaptation on an unseen OpenBind target, and FEP+ level accuracy on JACS at substantially lower computational cost. Beyond predictive gains, the learned metric weights show clear structure at the protein family level, with related proteins tending to share similar geometric preferences. Together, these results establish target conditioned comparison geometry as an effective framework for affinity ranking and for characterizing geometric variation across protein targets.

AI USE STATEMENT

We used generative AI tools to assist with language polishing and with literature retrieval and discovery. All AI-assisted text and retrieved references were reviewed and verified by the authors before inclusion in the manuscript. The authors take full responsibility for the scientific content, claims, and conclusions of this work.

REFERENCES

Colin Compas, Greg Zynda, John Sanson, Nima Pour Nejatian, Robert Sohigian, Rory Kelleher, Venkatesh Mysore, and Will Vick. NVIDIA DGX Infrastructure Solution for Drug Discovery. White Paper WP-11016-001, NVIDIA Corporation, July 2022. URL https://developer.download.nvidia.com/whitepapers/2022/NVIDIA-DGX-Infrastructure-Solution-for-Drug-Discovery.pdf.

Susan J Crennell, Elspeth F Garman, W Graeme Laver, Eric R Vimr, and Garry L Taylor. Crystal structure of a bacterial sialidase (from salmonella typhimurium lt2) shows the same fold as an influenza virus neuraminidase. *Proceedings of the National Academy of Sciences*, 90(21): 9852–9856, 1993.

Anne-Catherine Dock-Bregeon, Bernard Rees, Alfredo Torres-Larios, Gilbert Bey, Joel Caillet, and Dino Moras. Achieving error-free translation: the mechanism of proofreading of threonyl-trna synthetase at atomic resolution. *Molecular cell*, 16(3):375–386, 2004.

Aleksandr Ermolov, Leyla Mirvakhabova, Valentin Khrulkov, Nicu Sebe, and Ivan Oseledets. Hyperbolic vision transformers: Combining improvements in metric learning. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7399–7409. IEEE, 2022.

Bin Feng, Zijing Liu, Hao Li, Mingjun Yang, Junjie Zou, He Cao, Yu Li, Lei Zhang, and Sheng Wang. Hierarchical affinity landscape navigation through learning a shared pocket-ligand space. *Patterns*, 6(10), October 2025. ISSN 2666-3899. doi: 10.1016/j.patter.2025.101371. URL https://www.cell.com/patterns/abstract/S2666-3899(25)00219-3.

Shikun Feng, Minghao Li, Yinjun Jia, Weiying Ma, and Yanyan Lan. Protein-ligand binding representation learning from fine-grained interactions. *ArXiv*, abs/2311.16160, 2023. URL https://api.semanticscholar.org/CorpusID:265466810.

Bowen Gao, Bo Qiang, Haichuan Tan, Yinjun Jia, Minsi Ren, Minsi Lu, Jingjing Liu, Wei-Ying Ma, and Yanyan Lan. Drugclip: Contrastive protein-molecule representation learning for virtual screening. In *Advances in Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=lAbCgNcxm7.

Samuel Genheden and Ulf Ryde. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, 10(5):449–461, May 2015. ISSN 1746-0441. doi: 10.1517/17460441.2015.1032936. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4487606/.

Michael K Gilson, Tiqing Liu, Michael Baitaluk, George Nicola, Linda Hwang, and Jenny Chong. Bindingdb in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic acids research*, 44(D1):D1045–D1053, 2016.

Albert Gu, Frederic Sala, Beliz Gunel, and Christopher Ré. Learning mixed-curvature representations in product spaces. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=HJxeWnCcF7.

Tatiana Y Hargrove, Zdzislaw Wawrzak, Jialin Liu, W David Nes, Michael R Waterman, and Galina I Lepesheva. Substrate preferences and catalytic parameters determined by structural characteristics of sterol 14α-demethylase (cyp51) from leishmania infantum. *Journal of Biological Chemistry*, 286(30):26838–26848, 2011.

Zhiwu Huang and Luc Van Gool. A riemannian network for spd matrix learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

Zhiwu Huang, Ruiping Wang, Shiguang Shan, and Xilin Chen. Projection metric learning on grassmann manifold with application to video based face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 140–149, 2015.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. Generalist equivariant transformer towards 3d molecular interaction learning. *arXiv preprint arXiv:2306.01474*, 2023.

John M. Lee. *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, 2 edition, 2018. doi: 10.1007/978-3-319-91755-9.

Yan Li, Ruiping Wang, Zhiwu Huang, Shiguang Shan, and Xilin Chen. Face video retrieval with image query via hashing across euclidean space and riemannian manifold. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4758–4767, 2015.

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*, 50(2): 302–309, 2017.

Shuo Lu, Miranda Montoya, Liya Hu, Neetu Neetu, Banumathi Sankaran, BV Venkataram Prasad, and Timothy Palzkill. Mutagenesis and structural analysis reveal the ctx-m β-lactamase active site is optimized for cephalosporin catalysis and drug resistance. *Journal of Biological Chemistry*, 299 (5), 2023.

David Mendez, Anna Gaulton, A Patrícia Bento, Jon Chambers, Marleen De Veij, Eloy Félix, María Paula Magariños, Juan F Mosquera, Prudence Mutowo, Michał Nowotka, et al. Chembl: towards direct deposition of bioassay data. *Nucleic acids research*, 47(D1):D930–D940, 2019.

OpenBind. OpenBind EV-A71 2A dataset and benchmarks: A structure–affinity dataset for structure-based AI, 2026. URL https://doi.org/10.5281/zenodo.20026661. Dataset and reference benchmarks for docking, cofolding, and affinity prediction.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction, June 2025. URL https://www.biorxiv.org/content/10.1101/2025.06.14.659707v1.

Katrin Rittinger, Joe Budman, Jian Xu, Stefano Volinia, Lewis C Cantley, Stephen J Smerdon, Steven J Gamblin, and Michael B Yaffe. Structural analysis of 14-3-3 phosphopeptide complexes identifies a dual role for the nuclear export signal of 14-3-3 in ligand binding. *Molecular cell*, 4(2): 153–166, 1999.

Christina EM Schindler, Hannah Baumann, Andreas Blum, Dietrich Bose, Hans-Peter Buchstaller, Lars Burgdorf, Daniel Cappel, Eugene Chekler, Paul Czodrowski, Dieter Dorsch, et al. Large-scale assessment of binding free energy calculations in active drug discovery projects. *Journal of Chemical Information and Modeling*, 60(11):5457–5474, 2020.

Ondrej Skopek, Octavian-Eugen Ganea, and Gary Bécigneul. Mixed-curvature variational autoencoders. In *International Conference on Learning Representations*, 2020. URL https://arxiv.org/abs/1911.08411.

Jianhui Wang, Wenyu Zhu, Bowen Gao, Xin Hong, Ya-Qin Zhang, Wei-Ying Ma, and Yanyan Lan. Learning Protein–Ligand Binding in Hyperbolic Space. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(2):1150–1158, March 2026. ISSN 2374-3468. doi: 10.1609/aaai.v40i2.37086. URL https://ojs.aaai.org/index.php/AAAI/article/view/37086.

Lingle Wang, Yujie Wu, Yuqing Deng, Byungchan Kim, Levi Pierce, Goran Krilov, Dmitry Lupyan, Shaughnessy Robinson, Markus K. Dahlgren, Jeremy Greenwood, Donna L. Romero, Craig Masse, Jennifer L. Knight, Thomas Steinbrecher, Thijs Beuming, Wolfgang Damm, Ed Harder, Woody Sherman, Mark Brewer, Ron Wester, Mark Murcko, Leah Frye, Ramy Farid, Teng Lin, David L. Mobley, William L. Jorgensen, Bruce J. Berne, Richard A. Friesner, and Robert Abel. 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, February 2015. ISSN 0002-7863. doi: 10.1021/ja512751q. URL https://doi.org/10.1021/ja512751q.

Ziduo Yang, Weihe Zhong, Qiujie Lv, Tiejun Dong, Guanxing Chen, and Calvin Yu-Chian Chen. 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*, 46(12):8191–8208, December 2024. ISSN 1939-3539. doi: 10.1109/TPAMI.2024.3400515. URL https://ieeexplore.ieee.org/document/10530021. EHIGH.

Jie Yu, Zhaojun Li, Geng Chen, Xiangtai Kong, Jie Hu, Dingyan Wang, Duanhua Cao, Yanbei Li, Ruifeng Huo, Gang Wang, Xiaohong Liu, Hualiang Jiang, Xutong Li, Xiaomin Luo, and Mingyue Zheng. Computing the relative binding affinity of ligands based on a pairwise binding comparison network. *Nature Computational Science*, 3(10):860–872, October 2023. ISSN 2662-8457. doi: 10.1038/s43588-023-00529-9. URL https://www.nature.com/articles/s43588-023-00529-9. Number: 10.

Jiashen Zhou, Lin Zhang, Yiran Wang, Wenyan Song, Yuzhou Huang, Yajuan Mu, Werner Schmitz, Shu-Yu Zhang, Houwen Lin, Hong-Zhuan Chen, et al. The molecular basis of catalysis by sdr family members ketoacyl-acp reductase fabg and enoyl-acp reductase fabi in type-ii fatty acid biosynthesis. *Angewandte Chemie International Edition*, 62(46):e202313109, 2023.

Rongfeng Zou, Liguo Wang, Xinyan Wang, Ye Ding, and Hang Zheng. Breaking barriers in fep benchmarking: A large-scale dataset reflecting real-world drug discovery challenges. 2025.

A Proofs of Theoretical Results

This section provides proofs for the theoretical results in Section 4 and a convex-geometric characterization of the feasibility of a shared product metric across heterogeneous ranking constraints.

A.1 Proof of Proposition 4.1

Proof. Let $\widetilde{g} = \lambda g$ with $\lambda > 0$ constant.

Distance scaling. For any piecewise smooth curve $\gamma : [0,1] \to \mathcal{M}$,

$$\|\dot\gamma(u)\|_{\widetilde{g}}^2 = \widetilde{g}_{\gamma(u)}\big(\dot\gamma(u), \dot\gamma(u)\big) = \lambda g_{\gamma(u)}\big(\dot\gamma(u), \dot\gamma(u)\big).$$

Hence

$$\|\dot\gamma(u)\|_{\widetilde{g}} = \sqrt{\lambda}\,\|\dot\gamma(u)\|_g,$$

and therefore

$$L_{\widetilde{g}}(\gamma) = \sqrt{\lambda}\,L_g(\gamma).$$

Taking the infimum over all curves joining x and y gives

$$d_{\widetilde{g}}(x,y) = \sqrt{\lambda}\,d_g(x,y).$$

Sectional-curvature scaling. Because λ is constant, the Levi–Civita connections of g and $\widetilde{g}$ coincide. Equivalently, their Christoffel symbols are identical, and hence their curvature operators satisfy

$$\widetilde{R}(X,Y)Z = R(X,Y)Z.$$

Let X, Y be linearly independent tangent vectors spanning a two-plane σ. The sectional curvature under $\widetilde{g}$ is

$$K_{\widetilde{g}}(\sigma) = \frac{\widetilde{g}(\widetilde{R}(X,Y)Y,X)}{\widetilde{g}(X,X)\widetilde{g}(Y,Y) - \widetilde{g}(X,Y)^2}.$$

Using $\widetilde{g} = \lambda g$ and $\widetilde{R} = R$ yields

$$\begin{aligned} K_{\widetilde{g}}(\sigma) &= \frac{\lambda\, g(R(X,Y)Y,X)}{\lambda^2\big(g(X,X)g(Y,Y) - g(X,Y)^2\big)} \\ &= \frac{1}{\lambda}K_g(\sigma). \end{aligned}$$

Thus

$$d_{\widetilde{g}}(x,y) = \sqrt{\lambda}\,d_g(x,y), \qquad K_{\widetilde{g}}(\sigma) = \lambda^{-1}K_g(\sigma).$$

In particular, if g_κ has constant sectional curvature κ, then $g_t = \lambda_t g_\kappa$ has constant sectional curvature

$$K_t = \frac{\kappa}{\lambda_t}.$$

□

A.2 Proof of Theorem 4.2

Proof. By Proposition 4.1, for $g_t = \lambda_t g_\kappa$ with $\lambda_t > 0$,

$$d_{g_t}(p_t,q)^2 = \lambda_t\, d_{g_\kappa}(p_t,q)^2.$$

Therefore, for any two ligands a and b,

$$\begin{aligned} d_{g_t}(p_t,q_a)^2 < d_{g_t}(p_t,q_b)^2 &\iff \lambda_t d_{g_\kappa}(p_t,q_a)^2 < \lambda_t d_{g_\kappa}(p_t,q_b)^2 \\ &\iff d_{g_\kappa}(p_t,q_a)^2 < d_{g_\kappa}(p_t,q_b)^2, \end{aligned}$$

where the final equivalence follows from $\lambda_t > 0$. Hence multiplication of the metric by a target-specific positive scalar rescales every ligand distance by the same factor and cannot alter the induced within-target ranking. □

A.3 Proof of Theorem 4.3

Proof. Write a curve on the product manifold $\mathcal{M} = \prod_{j=1}^{m} \mathcal{M}_j$ as

$$\gamma(u) = \big(\gamma_1(u), \dots, \gamma_m(u)\big).$$

Under the product metric

$$g_w = \bigoplus_{j=1}^{m} w_j g_j, \qquad w_j > 0,$$

the squared speed decomposes component-wise:

$$\|\dot\gamma(u)\|_{g_w}^2 = \sum_{j=1}^{m} w_j \|\dot\gamma_j(u)\|_{g_j}^2.$$

Consequently,

$$\begin{aligned} E_{g_w}(\gamma) &= \int_0^1 \|\dot\gamma(u)\|_{g_w}^2 \, du \\ &= \sum_{j=1}^{m} w_j \int_0^1 \|\dot\gamma_j(u)\|_{g_j}^2 \, du \\ &= \sum_{j=1}^{m} w_j E_{g_j}(\gamma_j). \end{aligned} \tag{19}$$

For any curve γ joining $p = (p_1, \dots, p_m)$ and $q = (q_1, \dots, q_m)$, each component γ_j joins p_j and q_j. Hence

$$E_{g_j}(\gamma_j) \geq d_j(p_j, q_j)^2,$$

and therefore

$$E_{g_w}(\gamma) \geq \sum_{j=1}^{m} w_j d_j(p_j, q_j)^2.$$

Taking the infimum over all such γ gives

$$\inf_{\gamma: p \to q} E_{g_w}(\gamma) \geq \sum_{j=1}^{m} w_j d_j(p_j, q_j)^2. \tag{20}$$

It remains to show that the bound is attainable. Since every $(\mathcal{M}_j, g_j)$ is complete and connected, the Hopf–Rinow theorem guarantees a minimizing geodesic $\eta_j : [0,1] \to \mathcal{M}_j$ from p_j to q_j. Parameterize η_j with constant speed

$$\|\dot\eta_j(u)\|_{g_j} = d_j(p_j, q_j).$$

Then

$$E_{g_j}(\eta_j) = d_j(p_j, q_j)^2.$$

The product curve

$$\eta(u) = \big(\eta_1(u), \dots, \eta_m(u)\big)$$

therefore satisfies

$$E_{g_w}(\eta) = \sum_{j=1}^{m} w_j d_j(p_j, q_j)^2.$$

Combining this construction with equation 20 yields

$$\inf_{\gamma: p \to q} E_{g_w}(\gamma) = \sum_{j=1}^{m} w_j d_j(p_j, q_j)^2.$$

Using the standard identity $d_{g_w}(p, q)^2 = \inf_{\gamma: p \to q} E_{g_w}(\gamma)$, we obtain

$$d_{g_w}(p, q)^2 = \sum_{j=1}^{m} w_j d_j(p_j, q_j)^2.$$

The minimizing curve is precisely the product of the constant-speed minimizing component geodesics. □

A.4 Proof of Theorem 4.4

Proof. For two ligands a and b, define

$$c_{a,b,j} = d_j(p_j, q_{b,j})^2 - d_j(p_j, q_{a,j})^2.$$

Since

$$s_\alpha(p, q) = -\sum_{j=1}^{m} \alpha_j d_j(p_j, q_j)^2,$$

their score difference is

$$\begin{aligned} s_\alpha(p, q_a) - s_\alpha(p, q_b) &= \sum_{j=1}^{m} \alpha_j \left[d_j(p_j, q_{b,j})^2 - d_j(p_j, q_{a,j})^2 \right] \\ &= \alpha^\top c_{a,b}. \end{aligned}$$

Now consider the closed simplex

$$\Delta = \{\alpha \geq 0 : \mathbf{1}^\top \alpha = 1\}.$$

For any $\alpha \in \Delta$, $\alpha^\top c_{a,b}$ is a convex combination of the scalars $c_{a,b,1}, \ldots, c_{a,b,m}$. Therefore

$$\min_j c_{a,b,j} \leq \alpha^\top c_{a,b} \leq \max_j c_{a,b,j}.$$

Both endpoints are attained at simplex vertices corresponding to an index achieving the minimum or maximum. Moreover, every value between the two endpoints is obtained by a convex combination of those two vertices. Hence

$$\{\alpha^\top c_{a,b} : \alpha \in \Delta\} = \left[\min_j c_{a,b,j}, \max_j c_{a,b,j} \right].$$

It remains to characterize when two *strictly positive* product metrics, corresponding to $\alpha \in \Delta^\circ$, can induce opposite rankings.

If such metrics exist, then there are $\alpha^+, \alpha^- \in \Delta^\circ$ satisfying

$$(\alpha^+)^\top c_{a,b} > 0, \qquad (\alpha^-)^\top c_{a,b} < 0.$$

Since every convex combination lies between the smallest and largest coordinates of $c_{a,b}$, necessarily

$$\min_j c_{a,b,j} < 0 < \max_j c_{a,b,j}.$$

Conversely, suppose

$$\min_j c_{a,b,j} < 0 < \max_j c_{a,b,j}.$$

Choose indices j_- and j_+ such that

$$c_{a,b,j_-} < 0, \qquad c_{a,b,j_+} > 0.$$

Let $u = \frac{1}{m}\mathbf{1} \in \Delta^\circ$ and define

$$\alpha_\varepsilon^- = (1-\varepsilon)e_{j_-} + \varepsilon u, \qquad \alpha_\varepsilon^+ = (1-\varepsilon)e_{j_+} + \varepsilon u,$$

where e_j is the j-th simplex vertex. For every $\varepsilon \in (0, 1)$, both vectors belong to Δ°. By continuity, for sufficiently small $\varepsilon > 0$,

$$(\alpha_\varepsilon^-)^\top c_{a,b} < 0, \qquad (\alpha_\varepsilon^+)^\top c_{a,b} > 0.$$

Thus two strictly positive product metrics induce opposite rankings if and only if

$$\min_j c_{a,b,j} < 0 < \max_j c_{a,b,j}.$$

□

A.5 Shared-Metric Feasibility Across Heterogeneous Targets

Theorem 4.4 characterizes a single ligand pair. We next give a convex characterization for a collection of ranking constraints pooled across multiple targets.

Suppose there are K ordered ligand-pair constraints, and let $c_1, \dots, c_K \in \mathbb{R}^m$ denote their component-wise difference vectors as in Theorem 4.4. For each constraint, correct ordering requires

$$\alpha^\top c_k > 0.$$

Define the optimal shared worst-case margin

$$\gamma^\star := \max_{\alpha \in \Delta} \min_{1 \le k \le K} \alpha^\top c_k. \tag{21}$$

Proposition A.1 (Convex characterization of shared-metric feasibility)**.** *Let Δ_K denote the K-dimensional probability simplex. Then*

$$\gamma^\star = \min_{\lambda \in \Delta_K} \max_{1 \le j \le m} \left[\sum_{k=1}^{K} \lambda_k c_k \right]_j. \tag{22}$$

In particular, there exists a strictly positive metric coordinate $\alpha \in \Delta^\circ$ satisfying all pooled ranking constraints if and only if $\gamma^\star > 0$.

Proof. For a fixed $\alpha \in \Delta$,

$$\min_k \alpha^\top c_k = \min_{\lambda \in \Delta_K} \sum_{k=1}^{K} \lambda_k \alpha^\top c_k.$$

The objective is bilinear in α and λ, and both simplices are compact and convex. Therefore the finite-dimensional minimax theorem gives

$$\begin{aligned} \gamma^\star &= \max_{\alpha \in \Delta} \min_{\lambda \in \Delta_K} \alpha^\top \left(\sum_{k=1}^{K} \lambda_k c_k \right) \\ &= \min_{\lambda \in \Delta_K} \max_{\alpha \in \Delta} \alpha^\top \left(\sum_{k=1}^{K} \lambda_k c_k \right). \end{aligned}$$

For any vector $v \in \mathbb{R}^m$,

$$\max_{\alpha \in \Delta} \alpha^\top v = \max_j v_j,$$

which proves equation 22.

If some $\alpha \in \Delta^\circ$ satisfies $\alpha^\top c_k > 0$ for every k, then

$$\min_k \alpha^\top c_k > 0,$$

and hence $\gamma^\star > 0$.

Conversely, if $\gamma^\star > 0$, let $\alpha^\star \in \Delta$ attain the maximum in equation 21. Then

$$(\alpha^\star)^\top c_k \ge \gamma^\star > 0 \qquad \text{for every } k.$$

If $\alpha^\star$ lies on the boundary of the simplex, let $u = \frac{1}{m}\mathbf{1}$ and set

$$\alpha_\varepsilon = (1 - \varepsilon)\alpha^\star + \varepsilon u.$$

For every $\varepsilon > 0$, $\alpha_\varepsilon \in \Delta^\circ$, and by continuity all constraints remain strictly positive for sufficiently small ε. Thus a strictly positive shared product metric exists.

Therefore heterogeneous target-wise constraints are jointly compatible with a shared metric exactly when $\gamma^\star > 0$; otherwise no single point of the metric simplex can satisfy all constraints with a positive margin. □

B TARGET-CONDITIONED METRIC HEAD

The target-conditioned metric predictor is factor-aware. Rather than predicting the metric coordinates directly from the raw pocket descriptor, the metric head operates on the factor-specific pocket tangent coordinates

$$(z^P_{t,S}, z^P_{t,E}, z^P_{t,H}).$$

For each factor $j \in \{S, E, H\}$, a separate residual-logit network a_j maps the corresponding tangent block to a scalar,

$$r_{t,j} = a_j(z^P_{t,j}). \tag{23}$$

Each a_j has architecture

$$\mathrm{LayerNorm}(d_j) \to \mathrm{Linear}(d_j, 128) \to \mathrm{GELU} \to \mathrm{Dropout}(0.1) \to \mathrm{Linear}(128, 1).$$

The model additionally maintains a learned global logit vector

$$b = (b_S, b_E, b_H) \in \mathbb{R}^3.$$

The final metric logits are

$$\ell_t = b + (r_{t,S}, r_{t,E}, r_{t,H}). \tag{24}$$

The factor-specific output layers and the global logits are initialized to zero, so the metric prediction is initialized at the uniform point.

The logits are normalized to obtain positive metric coordinates

$$\widehat{\alpha}_t = \mathrm{softmax}(\ell_t) = (\widehat{\alpha}_{t,S}, \widehat{\alpha}_{t,E}, \widehat{\alpha}_{t,H}) \in \Delta^\circ. \tag{25}$$

These coefficients instantiate the target-conditioned product metric

$$g_{\widehat{\alpha}_t} = \widehat{\alpha}_{t,S} g_S \oplus \widehat{\alpha}_{t,E} g_E \oplus \widehat{\alpha}_{t,H} g_H,$$

and are used to weight the factor-wise squared geodesic distances in the affinity score.

C DETAILED TRAINING OBJECTIVES AND REGULARIZATION

This section provides the detailed training objectives summarized in Section 4.4. DrugPrism retains the assay-organized contrastive, listwise-ranking, and geometry-aware supervision of HypSeek, while replacing the original shared similarity readout with the target-conditioned product-geodesic score defined in equation 17.

C.1 TARGET-CONDITIONED LOGITS AND MASKING

For target t, query branch $b \in \{P, F\}$, and ligand l, the target-conditioned score is

$$s^b_{t,l} = -d_{g_{\widehat{\alpha}_t}}(p^b_t, q_l)^2 = -\sum_{j \in \{S,E,H\}} \widehat{\alpha}_{t,j} d_j\big(x^b_{t,j}, x^L_{l,j}\big)^2, \tag{26}$$

with logits

$$Z^b_{t,l} = \lambda_{\mathrm{logit}} s^b_{t,l}, \qquad \lambda_{\mathrm{logit}} > 0. \tag{27}$$

Training data are organized by assay. For assay i, let $\mathcal{P}_i$ denote its positive ligands and $\mathcal{B}_i$ the set of valid ligand candidates after removing invalid or conflicting assay relations. Conversely, for ligand l, let $\mathcal{Q}_l$ denote the set of valid protein queries against which it participates in the reverse retrieval objective. All masking rules follow the assay-level construction used in HypSeek.

C.2 BIDIRECTIONAL CONTRASTIVE ALIGNMENT

For each query branch $b \in \{P, F\}$, DrugPrism retains the bidirectional assay-level contrastive objective. The query-to-ligand loss is

$$\mathcal{L}^b_{q\to l} = -\sum_i \frac{1}{|\mathcal{P}_i|} \sum_{l\in\mathcal{P}_i} \log \frac{\exp(Z^b_{i,l})}{\sum_{r\in\mathcal{B}_i} \exp(Z^b_{i,r})}. \tag{28}$$

The reverse ligand-to-query objective is

$$\mathcal{L}^b_{l\to q} = -\sum_i \frac{1}{|\mathcal{P}_i|} \sum_{l\in\mathcal{P}_i} \log \frac{\exp(Z^b_{i,l})}{\sum_{i'\in\mathcal{Q}_l} \exp(Z^b_{i',l})}. \tag{29}$$

The two directions are combined as

$$\mathcal{L}^b_{\text{cont}} = \frac{1}{2}\left(\mathcal{L}^b_{q\to l} + \mathcal{L}^b_{l\to q}\right). \tag{30}$$

C.3 WITHIN-ASSAY LISTWISE RANKING

Contrastive alignment captures global protein–ligand compatibility, whereas affinity ranking requires resolving the relative ordering of ligands measured within the same assay. For each assay i, we sort its ligands by decreasing measured affinity. At ranking position k, let $\mathcal{C}_{i,k}$ contain the ligand occupying position k together with all valid weaker competitors remaining at that step. The corresponding Plackett–Luce probability is

$$P^b_{i,k} = \frac{\exp(Z^b_{i,k})}{\sum_{r\in\mathcal{C}_{i,k}} \exp(Z^b_{i,r})}. \tag{31}$$

The listwise ranking loss is

$$\mathcal{L}^b_{\text{rank}} = -\sum_i \sum_k \mu_{i,k} \log P^b_{i,k}, \tag{32}$$

where $\mu_{i,k}$ is the rank-dependent weighting used in HypSeek, placing greater emphasis on the top portion of each affinity-ranked list.

The contrastive and ranking losses are applied to both protein query branches:

$$\mathcal{L}_{\text{affinity}} = w_P\left(\mathcal{L}^P_{\text{cont}} + \lambda_{\text{rank}}\mathcal{L}^P_{\text{rank}}\right) + w_F\left(\mathcal{L}^F_{\text{cont}} + \lambda_{\text{rank}}\mathcal{L}^F_{\text{rank}}\right). \tag{33}$$

C.4 GEOMETRY-AWARE POCKET–LIGAND REGULARIZATION

The geometry-aware terms are applied to the pocket branch. Recall that

$$p^P_t = \left(x^P_{t,S}, x^P_{t,E}, x^P_{t,H}\right), \qquad q_l = \left(x^L_{l,S}, x^L_{l,E}, x^L_{l,H}\right)$$

are points on the shared product manifold $\mathcal{M} = \mathbb{S}^{d_S} \times \mathbb{R}^{d_E} \times \mathbb{H}^{d_H}$.

For tangent vectors $u = (u_S, u_E, u_H)$ and $v = (v_S, v_E, v_H)$ at p^P_t, the target-conditioned product metric induces the inner product

$$g_{\widehat{\alpha}_t}(u, v) = \sum_{j\in\{S,E,H\}} \widehat{\alpha}_{t,j} g_j(u_j, v_j), \tag{34}$$

and hence

$$\|v\|^2_{g_{\widehat{\alpha}_t}} = \sum_{j\in\{S,E,H\}} \widehat{\alpha}_{t,j} \|v_j\|^2_{g_j}. \tag{35}$$

Let

$$v_{t,l} = \operatorname{Log}_{p^P_t}(q_l), \qquad a_t = -\operatorname{Log}_{p^P_t}(o), \tag{36}$$

where $o = (o_S, o_E, o_H)$ denotes the reference point of the product manifold. The product logarithmic map is evaluated component-wise:

$$\operatorname{Log}_{p^P_t}(q_l) = \left(\operatorname{Log}^S_{x^P_{t,S}}(x^L_{l,S}), \operatorname{Log}^E_{x^P_{t,E}}(x^L_{l,E}), \operatorname{Log}^H_{x^P_{t,H}}(x^L_{l,H})\right).$$

The angular displacement between the ligand direction and the outward pocket direction is measured under the target-conditioned metric:

$$\phi_{t,l} = \arccos \frac{g_{\widehat{\alpha}_t}(v_{t,l}, a_t)}{\|v_{t,l}\|_{g_{\widehat{\alpha}_t}} \|a_t\|_{g_{\widehat{\alpha}_t}}}. \tag{37}$$

Let $r_{t,l}$ and $\psi_{t,l}$ denote the affinity-dependent radial and angular limits inherited from the geometric supervision of HypSeek. Using $[x]_+ = \max(x, 0)$, the radial penalty is

$$\mathcal{L}_{\text{rad}} = \frac{1}{N_{\text{geom}}} \sum_{(t,l)\in\mathcal{S}_{\text{geom}}} \left[\|v_{t,l}\|_{g_{\widehat{\alpha}_t}} - r_{t,l}\right]_+, \tag{38}$$

and the angular penalty is

$$\mathcal{L}_{\text{ang}} = \frac{1}{N_{\text{geom}}} \sum_{(t,l)\in\mathcal{S}_{\text{geom}}} \left[\phi_{t,l} - \psi_{t,l}\right]_+, \tag{39}$$

where $\mathcal{S}_{\text{geom}}$ is the set of pocket–ligand pairs receiving geometric supervision and $N_{\text{geom}} = |\mathcal{S}_{\text{geom}}|$. The complete geometry-aware objective is

$$\mathcal{L}_{\text{geom}} = \lambda_{\text{rad}}\mathcal{L}_{\text{rad}} + \lambda_{\text{geom-ang}}\mathcal{L}_{\text{ang}}. \tag{40}$$

We distinguish $\lambda_{\text{geom-ang}}$ from the coefficient of the additional angular-margin regularizer below to avoid notational ambiguity.

C.5 Additional Regularization

Angular-margin regularization. Following HypSeek, we discourage angular collapse by imposing an additional positive margin m beyond the affinity-dependent angular limit:

$$\mathcal{R}_{\text{ang}} = \frac{1}{N_{\text{geom}}} \sum_{(t,l)\in\mathcal{S}_{\text{geom}}} \left[\phi_{t,l} - \psi_{t,l} + m\right]_+. \tag{41}$$

Intra-assay heterogeneity regularization. We also retain the affinity-weighted enrichment term used in HypSeek. For assay i, let $\mathcal{H}_i$ denote the selected high-affinity ligands, $\rho_{i,l}$ their rank-dependent weights, and $\mathcal{G}_i$ the valid ligand set used to form the pocket-branch softmax distribution

$$\pi^P_{i,l} = \frac{\exp(Z^P_{i,l})}{\sum_{r\in\mathcal{G}_i} \exp(Z^P_{i,r})}. \tag{42}$$

The heterogeneity regularizer is

$$\mathcal{R}_{\text{het}} = -\frac{1}{C} \sum_i \sum_{l\in\mathcal{H}_i} \rho_{i,l} \log \pi^P_{i,l}, \tag{43}$$

where C normalizes over assays containing valid high-affinity ligands.

Coefficient-spectrum regularization. DrugPrism additionally regularizes the target-conditioned metric spectrum. Let

$$\mathbf{r}_t = (r_{t,S}, r_{t,E}, r_{t,H})$$

denote the target-specific residual logits produced by the factor-aware metric head, and let

$$\bar{\alpha} = \frac{1}{|\mathcal{B}|} \sum_{t\in\mathcal{B}} \widehat{\alpha}_t, \qquad u = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

be the mean metric coordinate over a mini-batch $\mathcal{B}$ and the uniform spectrum, respectively. We define

$$\mathcal{R}_{\text{spec}} = \frac{1}{|\mathcal{B}|} \sum_{t\in\mathcal{B}} \|\mathbf{r}_t\|_2^2 + \lambda_{\text{bal}} \,\mathrm{KL}\left(\bar{\alpha} \,\|\, u\right). \tag{44}$$

The first term discourages excessively large target-specific residual logits, while the second weakly prevents the batch-level metric spectrum from collapsing onto a single geometric component. Importantly, the balance term acts only on the batch mean $\bar{\alpha}$ rather than on each individual $\widehat{\alpha}_t$, thereby preserving target-specific geometric variation.

Table 3: Target-wise composition of the JACS affinity-ranking benchmark. JACS contains 8 protein targets and 199 ligands in total.

Target	No. of ligands
BACE	36
CDK2	16
JNK1	21
MCL1	42
P38	34
PTP1B	23
Thrombin	11
TYK2	16
Total	**199**

C.6 Overall Objective

Combining the terms above, the complete training objective is

$$\begin{aligned}\mathcal{L}_{\text{total}} = &\Big[w_P\big(\mathcal{L}^P_{\text{cont}} + \lambda_{\text{rank}}\mathcal{L}^P_{\text{rank}}\big) + w_F\big(\mathcal{L}^F_{\text{cont}} + \lambda_{\text{rank}}\mathcal{L}^F_{\text{rank}}\big)\Big] \\ &+ \gamma_{\text{geom}}\mathcal{L}_{\text{geom}} + \lambda_{\text{marg}}\mathcal{R}_{\text{ang}} + \lambda_{\text{het}}\mathcal{R}_{\text{het}} + \lambda_{\text{spec}}\mathcal{R}_{\text{spec}}.\end{aligned} \tag{45}$$

The representation encoders, geometry-specific projection heads, and target-conditioned metric predictor are optimized jointly. At inference time, the target pocket is encoded once to obtain $\widehat{\alpha}_t$, and candidate ligands are ranked using the same target-conditioned product-geodesic score in equation 17.

D Details of Affinity-Ranking Benchmarks

D.1 Data Deduplication

To prevent data leakage between model training and evaluation, we performed strict protein-sequence-level deduplication against all targets in the JACS, MERCK, and Hermite benchmarks. Specifically, before training, we collected the protein sequences of all evaluation targets and removed from the training corpus every record associated with a duplicated benchmark-target sequence. Consequently, none of the evaluation targets from these three benchmarks was present in the training data. This filtering procedure was applied before model optimization and was kept fixed across all reported experiments.

D.2 Benchmark Composition

We evaluate affinity-ranking performance on three complementary benchmarks: JACS, MERCK, and Hermite. JACS contains 8 protein targets and 199 ligands, MERCK contains 8 protein targets and 262 ligands, and Hermite contains 14 protein targets and 107 ligands. The detailed target-wise composition of the three benchmarks is reported in Tables 3, 4, and 5. Target names follow the nomenclature used by the corresponding benchmarks.

D.3 OpenBind EV-A71 2A Affinity Benchmark

We use the official OpenBind affinity-prediction benchmark for the EV-A71 2A protease. OpenBind reports affinity measurements for 601 compounds. During dataset generation, more than 2,000 sensorgrams from over 600 compounds, generally measured in triplicate, were manually inspected, fitted, and analyzed by trained experimentalists. After applying quality-control filters to these biophysical measurements, 494 compounds were retained in the official affinity benchmark (OpenBind, 2026).

We reserve a fixed pool of 25 compounds for adaptation and use the remaining 469 compounds exclusively for evaluation. For every tuning size, the adaptation compounds are sampled only from the fixed 25-compound pool, while the 469-compound evaluation set remains unchanged.

Table 4: Target-wise composition of the MERCK affinity-ranking benchmark. MERCK contains 8 protein targets and 262 ligands in total.

Target	No. of ligands
CDK8	32
cMet	24
EG5	28
HIF2A	41
PFKFB3	40
SHP2	26
SYK	44
TNKS2	27
Total	**262**

Table 5: Target-wise composition of the Hermite affinity-ranking benchmark. Hermite contains 14 protein targets and 107 ligands in total.

Target	No. of ligands
3CLPro	5
ACK1	6
BCL6	6
CDK4	6
Eg5	7
FKBP12	6
FXR	6
Galectin-3	8
HNE	6
HSP90	11
KRAS-G12D	10
PAK4	10
Pka	8
mGluR5	12
Total	**107**

For pocket construction, we use the official protein–ligand complex structures released with the OpenBind benchmark. For each available complex, we extract a candidate binding pocket using a spatial cutoff of 6 Å (six angstroms): a protein residue is included in the pocket if at least one of its atoms lies within 6 Å of any atom of the bound ligand. Among all candidate pockets extracted from the official complex structures, we select the pocket containing the largest number of protein residues and use it as the fixed pocket input for all 494 compounds. Thus, all adaptation and evaluation compounds are scored using the same receptor-pocket definition rather than a compound-dependent pocket.

Table 6 summarizes the resulting OpenBind evaluation protocol.

E Computational Latency and Cost

This appendix evaluates the computational cost of producing complete target-specific ligand rankings on the JACS benchmark. We report direct wall-clock measurements for DrugPrism and Boltz-2 and complement them with a literature-based A100 throughput estimate for FEP+. Because these methods have different native computational units and standard inputs, we use an output-aligned campaign unit for the cross-method comparison: the total compute required to produce ligand-level scores, and hence a complete ranking, for all compounds in each target series.

Table 6: Summary of the OpenBind EV-A71 2A affinity benchmark and the pocket-construction protocol used in our experiments.

Item	Setting
Compounds with reported affinity measurements	601
Compounds retained after quality control	494
Fixed adaptation pool	25
Held-out evaluation set	469
Pocket extraction cutoff	6 Å (six angstroms)
Pocket-selection criterion	Largest number of residues

Table 7: Composition of the JACS benchmark used for latency evaluation. The ligand and perturbation counts follow the original FEP+ benchmark.

Target	No. ligands	No. FEP edges
BACE1	36	58
CDK2	16	25
JNK1	21	31
MCL1	42	71
P38	34	56
PTP1B	23	49
Thrombin	11	16
TYK2	16	24
Total	**199**	**330**

E.1 Benchmark and Timing Protocol

Benchmark composition. JACS contains eight target-specific congeneric series with 199 ligands and 330 alchemical perturbation edges in total (Wang et al., 2015). Table 7 reports the number of ligands and FEP edges for each target. DrugPrism and Boltz-2 are evaluated on the same 199 final protein–ligand pairs.

DrugPrism pipeline. DrugPrism is timed from the standard structure-based benchmark artifacts: ligand SMILES, a processed receptor structure, a reference ligand, and the target FASTA sequence. For every target, the runtime pipeline regenerates a one-conformer ligand LMDB and a pocket LMDB using a 6 Å pocket radius. It then performs molecule encoding, pocket encoding, target-conditioned metric-head inference, and computation. The run uses half-precision inference and a ligand batch size of 64.

The principal DrugPrism latency includes all raw-input processing, temporary LMDB construction, encoder forward passes, and metric scoring. Checkpoint loading and prediction-array serialization are measured separately and excluded:

$$T_{\mathrm{DrugPrism}} = \frac{\sum_{p=1}^{8} T_p^{\mathrm{raw}\to\mathrm{score}}}{199}. \tag{46}$$

The resulting target-summed runtime is 15.024 s. For completeness, subtracting only checkpoint loading from the complete script wall clock gives a more conservative value of 18.035 s; all primary comparisons below use equation 46, which is available target-by-target and excludes prediction-file writing.

Boltz-2 pipeline. For each protein–ligand pair, we generate one YAML input and invoke `boltz predict` with the MSA-server option. Each invocation performs complex-structure prediction, and affinity prediction (Passaro et al., 2025). Although the official interface can accept a directory of YAML files [1], the evaluated per-ligand implementation processes ligands through independent subprocesses. For each target, the first invocation is treated as the warm-up/load call, and the reported

[1] https://github.com/jwohlwend/boltz/blob/main/docs/prediction.md

Table 8: Per-target computational latency on JACS, excluding checkpoint loading. DrugPrism is reported in milliseconds per pair and Boltz-2 in seconds per pair because their runtimes differ by over three orders of magnitude. Slowdown is Boltz-2 divided by DrugPrism.

Target	n	DrugPrism (ms/pair)	Boltz-2 (s/pair)	Slowdown
BACE1	36	78.78	211.5	2685×
CDK2	16	102.45	216.1	2109×
JNK1	21	69.92	228.8	3272×
MCL1	42	59.24	201.4	3400×
P38	34	76.85	219.1	2851×
PTP1B	23	85.75	203.6	2374×
Thrombin	11	109.14	168.1	1540×
TYK2	16	50.44	198.5	3936×
All targets	**199**	**75.50**	**208.5**	**2762×**

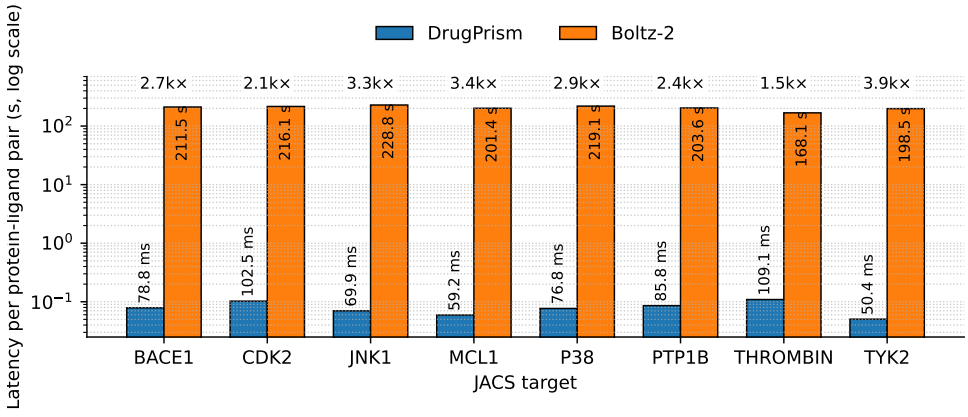


Figure 5: Per-target latency of DrugPrism and Boltz-2 on the 199 JACS protein–ligand pairs. The vertical axis is logarithmic. Values above the bars report DrugPrism latency in milliseconds and Boltz-2 latency in seconds; annotations above each target give the corresponding slowdown.

metric is

$$T_{\text{Boltz-2}} = \frac{\sum_{p=1}^{8} \left(T_p^{\text{total}} - T_p^{\text{first}}\right)}{199}. \tag{47}$$

Because T_p^{first} also contains one complete ligand prediction, this convention is conservative for Boltz-2.

Hardware and timing. The measured pipelines use single NVIDIA A100 80 GB GPUs and are timed with `time.perf_counter()`. The primary measurements exclude the load terms defined above while retaining the recurring data preparation and inference required to score each ligand series. DrugPrism and Boltz-2 use their standard deployment inputs rather than identical input modalities: DrugPrism uses a prepared receptor structure and reference ligand to define the pocket, whereas Boltz-2 predicts directly from sequence and ligand descriptions.

E.2 Measured JACS Latency

Table 8 reports the per-target latency. DrugPrism scores the complete 199-pair benchmark in 15.024 s, or 75.50 ms per pair. Boltz-2 requires 41,493 s (11.53 h), or 208.5 s per pair. Under the measured deployment protocols, Boltz-2 therefore requires approximately 2.76×10^3 times more wall-clock compute than DrugPrism.

The DrugPrism profile is dominated by raw feature-database construction: ligand and pocket LMDB generation together require 13.60 s, accounting for approximately 90.5% of its 15.02 s raw-input-to-

Table 9: Campaign-level compute required to produce complete JACS ligand rankings. DrugPrism and Boltz-2 are measured on single A100 GPUs. FEP+ is a literature-based, throughput-normalized A100 range.

Method	Total JACS cost	Average per target	Amortized per ranked ligand
DrugPrism	**0.00417 A100 GPU·h** (15.02 s)	5.22×10^{-4} **A100 GPU·h** (1.88 s)	2.10×10^{-5} **A100 GPU·h** (75.50 ms)
Boltz-2	**11.526 A100 GPU·h** (11.53 h)	**1.4407 A100 GPU·h** (86.44 min)	**0.0579 A100 GPU·h** (208.5 s)
FEP+	**495–1815 A100 GPU·h**	**61.88–226.88 A100 GPU·h**	**2.49–9.12 A100 GPU·h**

score runtime. By contrast, final target-conditioned metric scoring requires only 8.57 ms for all 199 pairs. Thus, the reported DrugPrism latency already includes its dominant data-preparation overhead rather than measuring only the lightweight scoring head.

E.3 Campaign-Level Comparison with FEP+

Seconds per protein–ligand input cannot be directly compared with seconds per alchemical perturbation edge. We therefore compare the common scientific output: producing complete ligand rankings for all eight JACS target series.

For FEP+, we use the NVIDIA DGX A100 drug-discovery white paper (Compas et al., 2022). It reports aggregate throughputs of 128.8 perturbations/day for a BACE-1 one-edge workload and 35 perturbations/day for a larger six-edge workload under high-density MIG scheduling. A DGX A100 contains eight A100 GPUs; therefore, the throughput-normalized cost is

$$C_{\text{edge}} = \frac{8 \times 24}{P_{\text{DGX}}} = 1.49\text{–}5.49 \text{ A100 GPU}\cdot\text{h/edge}, \tag{48}$$

which we round to 1.5–5.5 A100 GPU·h per perturbation.

For the 330-edge JACS network,

$$C_{\text{FEP+}}^{\text{JACS}} = 330 \times (1.5\text{–}5.5) = 495\text{–}1815 \text{ A100 GPU}\cdot\text{h}. \tag{49}$$

The corresponding measured neural-model costs are

$$C_{\text{DrugPrism}}^{\text{JACS}} = \frac{15.024}{3600} = 0.00417 \text{ A100 GPU}\cdot\text{h}, \tag{50}$$

and

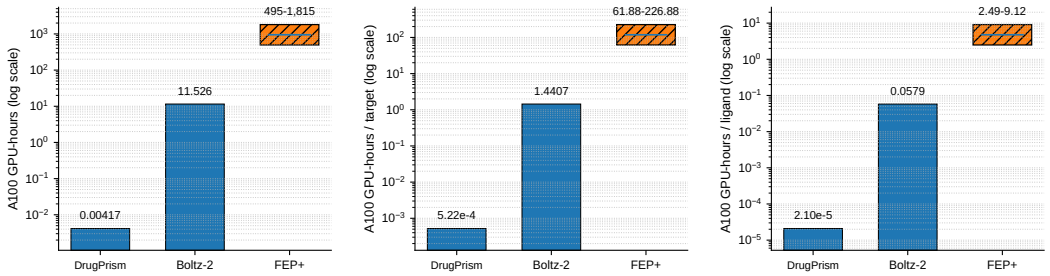
$$C_{\text{Boltz-2}}^{\text{JACS}} = \frac{41493}{3600} = 11.526 \text{ A100 GPU}\cdot\text{h}. \tag{51}$$

Table 9 reports the full campaign cost, the mean cost per target obtained by dividing by eight, and the amortized cost per ranked ligand obtained by dividing by 199. For FEP+, the last quantity is an output-normalized share of the full perturbation network, not the runtime of an independent protein–ligand prediction.

Under the common campaign-level unit, Boltz-2 consumes approximately $2762\times$ the compute of DrugPrism. The FEP+ estimate is $43.0\text{–}157.5\times$ the Boltz-2 cost and approximately 1.19×10^5–4.35×10^5 times the DrugPrism cost. On a fully utilized eight-A100 DGX system, the FEP+ estimate corresponds to approximately 2.58–9.45 days of aggregate throughput for the complete JACS perturbation network.

E.4 Limitations and Interpretation

The comparison measures deployed end-to-end cost for the same final output—a ranked ligand series—rather than the efficiency of architecturally identical tasks. DrugPrism uses a prepared receptor structure and reference ligand to define a local pocket and returns a target-conditioned ranking score. Boltz-2 additionally predicts a three-dimensional complex structure and affinity from sequence-level inputs, while FEP+ performs explicit alchemical molecular dynamics.



(a) Total JACS cost **(b)** Average per target **(c)** Amortized per ligand

Figure 6: Computational cost under three output-aligned normalizations. All vertical axes use logarithmic scales. DrugPrism and Boltz-2 are point measurements. The hatched FEP+ bars show the full range induced by the publicly reported 1.5–5.5 A100 GPU·h cost per perturbation; the horizontal segment marks the geometric midpoint of each range.

The Boltz-2 result corresponds to the measured per-ligand subprocess implementation. The official interface also supports directory inputs, and optimized deployments could reuse cached MSAs or persistent model workers (Passaro et al., 2025). Conversely, the benchmark excludes one complete first Boltz-2 prediction per target as part of the load term, so the reported Boltz-2 average is conservative.

Finally, the FEP+ results are literature-derived throughput estimates rather than a direct rerun of the JACS network in our environment. They do not include manual setup, failed-edge reruns, or additional convergence analysis and should be interpreted as aggregate A100 compute cost rather than isolated perturbation latency.

F Family-level analysis of learned geometry preferences

This appendix provides additional details for the protein family analysis in Section 5.5. Consistent with the terminology in the main text, we use *geometry preference* to describe the relative spherical, Euclidean, and hyperbolic metric weights learned by DrugPrism. The analysis asks whether these learned metric weights show clear structure at the protein family level and whether related proteins tend to share similar geometric preferences.

F.1 Family aggregation and preference measures

Each protein target is associated with an InterPro family through its UniProt-linked InterPro annotation. Following the filtering used in the main text, we retain families in the training set represented by at least 20 model-output records, 10 structures, and 5 distinct proteins, resulting in 91 families.

For a protein p, let

$$\boldsymbol{\alpha}_p = (\alpha_{p,S}, \alpha_{p,E}, \alpha_{p,H}) \tag{52}$$

denote its spherical, Euclidean, and hyperbolic metric weights. When a protein is represented by multiple model-output records, we first average records corresponding to the same structure and then average across structures. The family-level mean for family f is

$$\bar{\boldsymbol{\alpha}}_f = \frac{1}{n_f} \sum_{p \in f} \boldsymbol{\alpha}_p, \tag{53}$$

where n_f is the number of distinct proteins in the family. This aggregation gives equal weight to each protein regardless of how many structures or model-output records are available for that protein.

We summarize each family using the same two quantities reported in Figure 4. Let the components of $\bar{\boldsymbol{\alpha}}_f$ be ordered as $\bar{\alpha}_{f,(1)} \geq \bar{\alpha}_{f,(2)} \geq \bar{\alpha}_{f,(3)}$. The *preference strength* is

$$P_f = \bar{\alpha}_{f,(1)} - \frac{1}{2}\left(\bar{\alpha}_{f,(2)} + \bar{\alpha}_{f,(3)}\right) = \frac{3}{2}\left(\bar{\alpha}_{f,(1)} - \frac{1}{3}\right), \tag{54}$$

which measures how strongly the largest family-mean weight departs from the uniform mixture. We use $P_f \geq 0.15$ to denote a strong geometric preference, equivalent to $\bar{\alpha}_{f,(1)} \geq 0.4333$.

Table 10: Summary of family-level geometry preferences. Preference strength and within-family variation follow Eqs. 54 and 55.

Criterion	Families	Share
Strong geometric preference, $P_f \geq 0.15$	61	67.0%
Low within-family variation, $V_f < 0.15$	76	83.5%
Both criteria	51	56.0%

Table 11: Dominant geometry among families with strong geometric preference.

Dominant geometry	Families	Share
Spherical	5	8.2%
Euclidean	14	23.0%
Hyperbolic	42	68.9%
Total	61	100%

The *within-family variation* is

$$V_f = \max_{j \in \{S,E,H\}} \sqrt{\frac{1}{n_f} \sum_{p \in f} (\alpha_{p,j} - \bar{\alpha}_{f,j})^2}. \tag{55}$$

We use $V_f < 0.15$ to denote low within-family variation. These thresholds are descriptive summaries used to characterize the learned weights rather than per-family significance tests.

The dominant geometry of a family is defined by the largest component of $\bar{\boldsymbol{\alpha}}_f$. This label indicates which component has the largest family mean, but it does not imply that the remaining components are negligible. Families with similar values for the two largest components are therefore naturally interpreted as mixed preferences even when one component is nominally dominant.

F.2 Family-level geometry preferences

The learned metric weights show clear structure at the protein family level. Among the 91 families, 61 families, or 67.0%, exhibit a strong geometric preference, while 76 families, or 83.5%, show low within-family variation. A total of 51 families, or 56.0%, satisfy both criteria. These observations support the main-text result that related proteins often share similar geometric preferences.

Among the 61 families with a strong geometric preference, 42 are hyperbolic-dominant, 14 are Euclidean-dominant, and 5 are spherical-dominant. Thus, 68.9% of the strongly preferred families favor the hyperbolic component. This is the most frequent dominant preference in this subset, but it is not universal. The presence of clear spherical and Euclidean preferences is consistent with the target-wise variation that motivates DrugPrism and with the need to adapt relative geometric contributions rather than use one fixed geometry for every target.

Table 12 reports the representative families highlighted in Figure 4, together with two examples whose two largest family-mean weights are nearly tied. The representative families combine a clear dominant component with relatively low within-family variation. By contrast, Avidin and legume lectin illustrate why the dominant geometry should be interpreted as the largest component of a mixture rather than as an exclusive assignment.

F.3 Representative biological patterns

The representative families in Figure 4 suggest several geometry-specific biological patterns. These observations provide examples that are consistent with the family-level geometric preferences learned by DrugPrism.

- **Spherical preference.** CYP51-like and FabI families are associated with flexible substrate accommodation. CYP51 proteins perform related sterol transformations while their substrate preferences vary with active-site topology, and FabI enzymes accommodate acyl substrates

Table 12: Representative family-level geometry preferences. Mean weights are ordered as spherical, Euclidean, and hyperbolic.

Family	InterPro	n_f	Dominant	Mean (S, E, H)	P_f	V_f
CYP51-like	IPR050529	5	S	(.552,.192,.256)	.327	.034
FabI	IPR014358	5	S	(.443,.358,.199)	.164	.025
Metallo-β-lactamase type 2	IPR050855	5	E	(.060,.571,.369)	.357	.016
Sialidase GH34	IPR001860	19	E	(.082,.527,.392)	.290	.040
14-3-3	IPR000308	7	H	(.063,.271,.666)	.499	.105
Threonyl-tRNA ligase	IPR002320	6	H	(.114,.312,.575)	.362	.039
Avidin	IPR005469	5	H	(.046,.477,.478)	.216	.052
Legume lectin	IPR016363	6	E	(.071,.467,.462)	.200	.025

with different chain lengths through changes near the substrate-binding region (Hargrove et al., 2011; Zhou et al., 2023).

- **Euclidean preference.** Metallo-β-lactamase and sialidase families emphasize precise local positioning around the active site. Structural studies of β-lactamases and sialidases illustrate the importance of constrained local arrangements for catalysis and substrate recognition (Lu et al., 2023; Crennell et al., 1993).
- **Hyperbolic preference.** 14-3-3 and threonyl-tRNA ligase involve selective recognition among related motifs or substrates. 14-3-3 proteins recognize specific phosphorylated peptide motifs, while threonyl-tRNA ligase must discriminate among related amino-acid substrates during aminoacylation and editing (Rittinger et al., 1999; Dock-Bregeon et al., 2004).

These examples should be interpreted as descriptive biological context rather than a one-to-one mapping between a geometry and a biological mechanism. The current analysis shows that learned geometric preferences differ across protein families and are often consistent within a family. It does not directly measure pocket flexibility, ligand diversity, conformational dynamics, or catalytic mechanism. Establishing a general mechanistic relationship between these properties and the learned geometry weights would require dedicated pocket-level and ligand-level analyses.

F.4 Families with strong preference and low within-family variation

For completeness, Table 13 lists all 51 families that satisfy both $P_f \geq 0.15$ and $V_f < 0.15$. These families provide the clearest examples of a family-level geometric preference that is also shared consistently across member proteins.

Table 13: Families with strong geometric preference and low within-family variation. Families are grouped by the dominant component of the family-mean metric weights.

InterPro	Family name	n_f	P_f	V_f
Spherical-dominant				
IPR050529	Lanosterol 14-alpha demethylase-like	5	0.327	0.034
IPR005990	Inosine-5'-monophosphate dehydrogenase	7	0.189	0.071
IPR002397	Cytochrome P450, B-class	5	0.182	0.135
IPR014358	Enoyl-[acyl-carrier-protein] reductase (NADH)	5	0.164	0.025
IPR050494	Serine/threonine and dual-specificity kinase	5	0.161	0.045
Euclidean-dominant				
IPR050855	Metallo-beta-lactamase type 2-like	5	0.357	0.016
IPR000652	Triosephosphate isomerase	6	0.350	0.009
IPR058199	Metallo-beta-lactamase type 2/VIM/IMP-1	6	0.305	0.049
IPR058131	Beta-lactamase CTX-M	7	0.304	0.033
IPR001860	Glycoside hydrolase, family 34	19	0.290	0.040
IPR014033	Arginase	5	0.273	0.021
IPR026856	Sialidase family	8	0.264	0.056
IPR050111	C-type lectin and snaclec domain-containing protein	5	0.249	0.023
IPR000941	Enolase	6	0.246	0.141
IPR040234	Glutaminyl-peptide cyclotransferase-like	5	0.206	0.040
IPR016363	Legume lectin	6	0.200	0.025
IPR058044	Metallo-beta-lactamase VIM-1/2/7	5	0.194	0.063
Hyperbolic-dominant				
IPR000308	14-3-3 protein	7	0.499	0.105
IPR000413	Integrin alpha chain	5	0.427	0.145
IPR030678	Peptide/nickel binding protein, MppA-type	10	0.397	0.148
IPR014412	Genome polyprotein, Flavivirus	11	0.366	0.132
IPR002320	Threonine-tRNA ligase, class IIa	6	0.362	0.039
IPR050150	Immunoglobulin Variable Light Chain	11	0.352	0.142
IPR025799	Protein arginine N-methyltransferase	6	0.335	0.110
IPR050784	Inhibitor of Apoptosis	5	0.332	0.035
IPR005904	Hypoxanthine phosphoribosyl transferase	9	0.325	0.041
IPR006060	Maltose/Cyclodextrin ABC transporter, substrate-binding protein	6	0.319	0.104
IPR001461	Aspartic peptidase A1	13	0.317	0.134
IPR051313	Bacterial iron-siderophore binding	5	0.309	0.088
IPR002398	Peptidase C14 family	5	0.296	0.107
IPR044156	Galectin-like	5	0.291	0.014
IPR010049	MTA/SAH nucleosidase	5	0.265	0.020
IPR023561	Carbonic anhydrase, alpha-class	7	0.255	0.122
IPR001314	Peptidase S1A, chymotrypsin family	18	0.255	0.111
IPR001360	Glycoside hydrolase family 1	7	0.254	0.094
IPR013126	Heat shock protein 70 family	7	0.243	0.058
IPR000239	GPCR kinase	6	0.241	0.096
IPR050314	Glycosyl hydrolase family 18	9	0.228	0.086
IPR001427	Pancreatic ribonuclease	5	0.227	0.052
IPR001574	Ribosome-inactivating protein	6	0.224	0.111
IPR018094	Thymidylate kinase	10	0.224	0.121
IPR039702	Farnesyl pyrophosphate synthase-like	6	0.218	0.080
IPR005469	Avidin	5	0.216	0.052
IPR050198	Non-receptor tyrosine kinases involved in cell signaling	18	0.211	0.128
IPR008181	Deoxyuridine triphosphate nucleotidohydrolase	8	0.210	0.100
IPR003138	Picornavirus coat protein VP4	5	0.205	0.114
IPR023635	Peptide deformylase	13	0.190	0.075
IPR015433	Phosphatidylinositol 3-/4-kinase	9	0.186	0.129
IPR002166	RNA dependent RNA polymerase, hepatitis C virus	12	0.181	0.136
IPR021190	Peptidase M10A	10	0.175	0.079
IPR001211	Phospholipase A2	10	0.152	0.134