BEYOND STATIC POCKETS: BENCHMARKING AND MODELING MULTI-STATE TRANSPORTER DRUG DISCOVERY

Anonymous authors
Paper under double-blind review

ABSTRACT

Transporter drug discovery requires consideration of multiple conformations of the same protein, yet existing computational methods and models still struggle to rapidly and accurately identify ligand-preferred conformations. We introduce SLCState, a manually curated benchmark, and SLCHier, a model for conformational recognition. SLCState contains 762 pocket–ligand records from 622 PDB entries, covering 109 proteins, 14 structural folds, and five conformational states. Seven data splits assess generalization to unseen proteins, new ligand associations, and increasingly complex state sets. We also propose SLCHier, which uses the transporter structural hierarchy as a prior and represents it in hyperbolic space through a two-scale representation that preserves state-level pocket embeddings while organizing them with transporter-level centers. Trained separately on each split, SLCHier achieves higher top-1 accuracy than the compared zero-shot baselines in all seven settings. Controlled ablations further confirm the importance of the structural hierarchy and hyperbolic representations. Together, SLCState and SLCHier provide an evaluation framework and modeling approach for multi-state transporter drug discovery.

1 INTRODUCTION

Solute carriers (SLCs) mediate the transmembrane transport of neurotransmitters, nutrients, metabolites, and ions and are important drug targets. SLCs span diverse structural folds and transition among outward-open, occluded, and inward-open states, altering the ligand-binding pocket within the same protein (Ferrada & Superti-Furga, 2022). Figure 1b illustrates this with the VMAT2 inhibitor reserpine: its experimentally observed binding pose fits the inward-facing pocket but produces pronounced steric clashes when superposed onto occluded or outward-facing structures. These pocket changes affect recognition of the same ligand and suggest that different inhibitors may favor distinct conformations of the same transporter (Pidathala et al., 2023; Wang et al., 2024; Wu et al., 2024).

Conformation-guided drug discovery has practical pharmacological significance. Conventional selective serotonin reuptake inhibitors (SSRIs) primarily stabilize the outward-open state of the serotonin transporter (SERT), whereas ibogaine favors the inward-open state. However, ibogaine's target promiscuity and cardiotoxicity complicate mechanistic studies (Coleman et al., 2016; 2019; Rubi et al., 2017). Seeking more target-selective inhibitors, (Singh et al., 2023) screened compounds against the ibogaine-bound inward-open structure and optimized low-nanomolar inhibitors that exhibited anxiolytic- and antidepressant-like activity in mice. These findings highlight the potential of conformation-guided screening and the importance of identifying ligand-associated conformations when seeking molecules with distinct pharmacological effects.

Computational methods provide valuable tools for exploring alternative protein conformations, but limitations remain. Conventional molecular dynamics simulations struggle to adequately sample transitions across high energy barriers within practical computational budgets, while collective-variable-based enhanced sampling depends on the selection of suitable variables (Tang et al., 2026; Li & Ma, 2025). Recent structure-based protein–ligand models have provided effective tools for large-scale virtual screening. DrugCLIP enables rapid pocket–molecule retrieval through a shared representation space; LigUnity combines virtual screening with affinity ranking; and HypSeek models

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

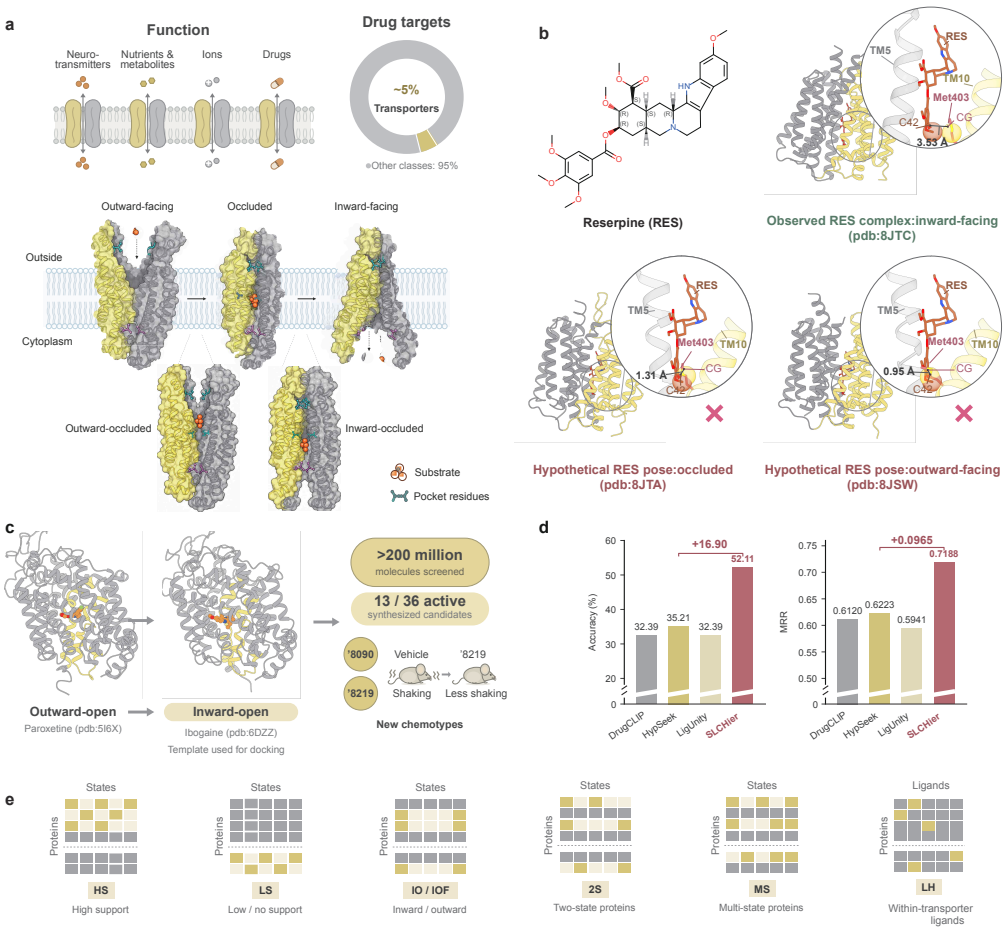


Figure 1: **Conformational selectivity in SLCs. a,** Transporter functions, alternating-access conformations, and associated binding-pocket changes. **b,** Reserpine (RES) and its experimentally observed inward-facing complex (8JTC), compared with hypothetical poses in occluded (8JTA) and outward-facing (8JSW) structures. **c,** Conformation-guided ligand discovery using serotonin transporter (SERT) structures. **d,** Retrieval accuracy and mean reciprocal rank (MRR) of DrugCLIP, HypSeek, LigUnity, and SLCHier under the low-support protein holdout (LS) split. **e,** SLCState split schematics. Yellow, test sets, grey, training sets, white, no recorded structure. Matrix sizes and proportions are illustrative.

protein–ligand relationships in hyperbolic space (Gao et al., 2023; Feng et al., 2025; Wang et al., 2026). However, current models cannot effectively distinguish ligand-binding preferences among different conformations of the same protein (Figure 1d).

To systematically investigate this problem, we formulate ligand-conditioned within-protein conformational-state ranking and establish the **SLCState** benchmark. SLCState contains 762 pocket–ligand records from 622 PDB entries, covering 109 proteins, 14 structural folds, and five conformational states (Figure 2). By holding the transporter and ligand fixed, SLCState specifically evaluates a model's ability to distinguish conformations of the same protein. Seven data splits (Figure 1e) assess generalization to unseen proteins, new ligand associations, and different combinations of conformational states.

We further propose **SLCHier**, which combines ligand-conditioned state ranking with structural hierarchy learning in hyperbolic space. The model shares information through structural relationships across proteins while preserving pocket differences among conformations of the same protein, enabling protein-level structural knowledge to support ligand-associated conformation recognition.

Under a shared evaluation setting, SLCHier trained separately on each split achieves higher top-1 accuracy than the compared zero-shot baselines across all seven evaluations. Ablation studies further

support the contributions of state supervision, structural hierarchy, and hyperbolic representations. Together, SLCState and SLCHier provide an evaluation framework and modeling approach for conformation-dependent ligand recognition, supporting conformation-guided virtual screening and the prioritization of candidates for experimental testing.

Our main contributions are twofold.

- We introduce SLCSTATE, a manually curated benchmark for ligand-conditioned conformational-state ranking, with seven complementary splits evaluating protein, ligand, and state-space generalization.
- We propose SLCHIER, a hierarchy-aware hyperbolic model with two-scale state- and transporter-level representations. Experiments and ablations show consistent gains and support the contributions of structural hierarchy, hyperbolic geometry, and state-aware supervision.

2 RELATED WORK

Protein–ligand retrieval. Representative protein–ligand modeling approaches span structure-based scoring and learned retrieval. RTMScore learns structure-based binding scores with residue–atom interaction modeling, while TANKBind predicts protein–ligand binding configurations with geometry-aware neural networks (Shen et al., 2022; Lu et al., 2022). DrugCLIP formulates virtual screening as contrastive pocket–molecule retrieval (Gao et al., 2023; Jia et al., 2026), and BindNet learns fine-grained protein–ligand interaction representations (Feng et al., 2024). More recently, LigUnity learns a shared pocket–ligand space with hierarchical affinity supervision (Feng et al., 2025), whereas HypSeek introduces hyperbolic representations for protein–ligand retrieval (Wang et al., 2026).

Hyperbolic and hierarchical representation learning. Hyperbolic geometry is well suited to hierarchical and tree-like structure because its volume grows exponentially with radius. Lorentz embeddings enable continuous hierarchical representations (Nickel & Kiela, 2018), while constant-curvature models provide differentiable geometric operations and aggregation on non-Euclidean manifolds (Bachmann et al., 2020). Hierarchical supervision can further organize learned representations; for example, HypStructure constrains hyperbolic class centers using a predefined label hierarchy (Sinha et al., 2024).

3 PROBLEM FORMULATION

Given a ligand and a transporter, the goal is to identify the conformational state of that transporter associated with the ligand, which differs from conventional protein–ligand retrieval. We next formalize the entities, state candidates, and scoring rule for this recognition problem.

Transporters, conformational states, and structural representatives. Let $\mathcal{T} = \{1, \dots, T\}$ denote the transporter set and $\mathcal{M}$ the molecular space. Each transporter $t \in \mathcal{T}$ has a transporter-specific set of documented conformational states $\mathcal{S}_t = \{s_{t,1}, \dots, s_{t,K_t}\}, K_t = |\mathcal{S}_t|$. The state labels are transporter-specific. A state may be represented by multiple experimentally resolved pocket structures. We therefore associate each $s \in \mathcal{S}_t$ with $\mathcal{P}_{t,s} = \left\{p_{t,s}^{(1)}, \dots, p_{t,s}^{(n_{t,s})}\right\}, n_{t,s} \geq 1$. These structures are treated as alternative structural representatives of the same conformational state.

State-resolved ligand associations. Each benchmark record is represented as $q = (m_q, t_q, s_q^\star)$. Here $m_q \in \mathcal{M}, t_q \in \mathcal{T}, s_q^\star \in \mathcal{S}_{t_q}$, and $s_q^\star$ is the annotated conformational state associated with molecule m_q for transporter t_q. The collection of state-resolved associations is

$$\mathcal{D} = \{(m, t, s) : m \text{ is associated with state } s \text{ of transporter } t\}.$$

Pocket compatibility and state-level scoring. Let $f_\theta(p, m) \in \mathbb{R}$ denote the compatibility score between pocket structure p and molecule m, with larger values indicating stronger predicted compatibility. Since one state can contain multiple pocket structures, we define the state-level score as $F_\theta(m, t, s) = \max_{p \in \mathcal{P}_{t,s}} f_\theta(p, m)$. For a query (m, t), the model scores every documented state

$s \in \mathcal{S}_t$ and predicts $\widehat{s}_\theta(m,t) = \arg\max_{s\in\mathcal{S}_t} F_\theta(m,t,s)$. Thus, the complete prediction problem can be summarized as $(m,\ t,\ \{\mathcal{P}_{t,s}\}_{s\in\mathcal{S}_t}) \longmapsto \mathrm{rank}\left(\{F_\theta(m,t,s)\}_{s\in\mathcal{S}_t}\right)$.

Evaluation. For a test record $q = (m_q, t_q, s_q^\star)$, let r_q denote the rank of the annotated state $s_q^\star$ among all states in $\mathcal{S}_{t_q}$ according to F_θ. We evaluate state recognition using top-1 accuracy $\mathrm{Acc} = \frac{1}{|\mathcal{D}_{\mathrm{test}}|}\sum_{q\in\mathcal{D}_{\mathrm{test}}} \mathbb{I}\left[\widehat{s}_\theta(m_q,t_q) = s_q^\star\right]$, and mean reciprocal rank $\mathrm{MRR} = \frac{1}{|\mathcal{D}_{\mathrm{test}}|}\sum_{q\in\mathcal{D}_{\mathrm{test}}} \frac{1}{r_q}$. Accuracy measures exact recovery of the annotated state, whereas MRR also rewards models that place the correct state near the top of the ranking.

4 Preliminaries: Lorentz Hyperbolic Geometry

Hyperbolic space is a Riemannian manifold with constant negative curvature, providing a natural geometry for hierarchical representations. We use the Lorentz model for its closed-form geodesic distance and differentiable geometric operations.

Lorentz model, geodesic distance, and exponential map. For curvature $-c < 0$ with $c > 0$, the d-dimensional Lorentz model is

$$\mathbb{H}_c^d = \left\{ x \in \mathbb{R}^{d+1} : \langle x, x\rangle_L = -\frac{1}{c},\ x_0 > 0 \right\}, \qquad \langle x, y\rangle_L = -x_0 y_0 + \sum_{j=1}^{d} x_j y_j. \tag{1}$$

Its tangent space is $T_x\mathbb{H}_c^d = \{v : \langle x, v\rangle_L = 0\}$ with $\|v\|_L = \sqrt{\langle v, v\rangle_L}$. For $x, y \in \mathbb{H}_c^d$, the geodesic distance is $d_c(x,y) = c^{-1/2}\,\mathrm{arcosh}(-c\langle x, y\rangle_L)$. For $v \in T_x\mathbb{H}_c^d$, the exponential map is $\mathrm{Exp}_x^c(v) = \cosh(\sqrt{c}\,\|v\|_L)x + \frac{\sinh(\sqrt{c}\,\|v\|_L)}{\sqrt{c}\,\|v\|_L}v$. At the Lorentz origin $o = (c^{-1/2}, 0, \ldots, 0)$, $T_o\mathbb{H}_c^d \simeq \mathbb{R}^d$ via $u \mapsto (0, u)$, allowing Euclidean features to be lifted to the manifold through Exp_o^c.

Lorentz gyromidpoint. Given $z_1, \ldots, z_n \in \mathbb{H}_c^d$ and positive weights α_i with $\sum_i \alpha_i = 1$, define

$$v = \sum_{i=1}^{n} \alpha_i z_i, \qquad \zeta = \sqrt{-c\langle v, v\rangle_L}, \qquad G_c(z_1, \ldots, z_n; \alpha) = \frac{v}{\zeta}. \tag{2}$$

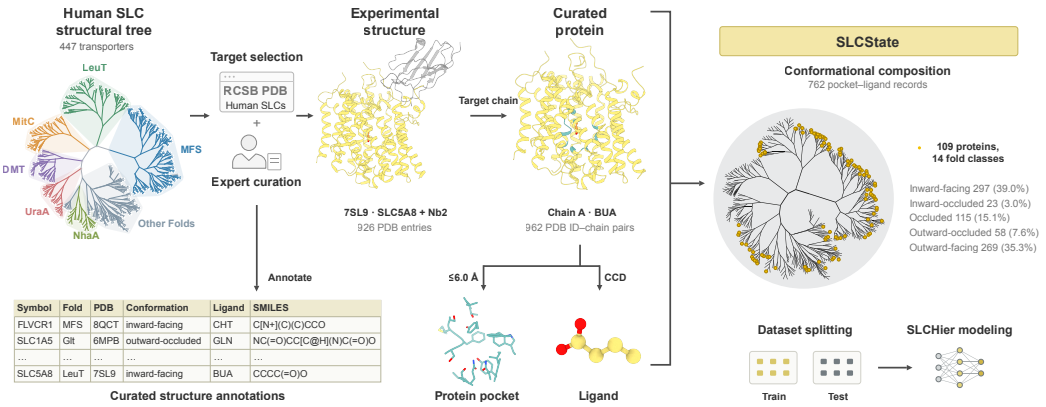
The weighted sum v is future-timelike, so its normalization again lies on $\mathbb{H}_c^d$. The gyromidpoint is a closed-form, differentiable, and permutation-invariant aggregate.

5 SLCState: A Benchmark for Multi-State Transporter Drug Discovery

5.1 Dataset construction and coverage

Information on transporter structures, conformations, and ligand functions is scattered, with inconsistent terminology and incomplete annotations. Different subunits within the same structure may adopt distinct conformations, and neither ligand presence nor protein-level activity alone establishes a ligand's functional role in a specific conformation. To address these challenges, we manually curated SLCState, integrating experimental structures, chain-level conformational annotations, and pocket–ligand associations (Figure 2). The SLC structural tree organizes structural and conformational information across different levels. Through literature review and structural visualization, we recorded ligand identity and functional role separately and assigned each protein chain to one of five standardized conformational states. Following target-chain extraction and removal of non-target components, pocket residues were defined as protein residues with at least one heavy atom within 6.0Å of any ligand heavy atom. Overlapping pockets were merged using a Jaccard similarity threshold.

The structure inventory contains 926 PDB entries and 962 PDB ID–chain pairs, with ligand-free records excluded from pocket construction. The final pocket benchmark comprises 762 pocket–ligand records from 622 PDB entries, covering 109 protein identifiers and 14 fold classes. Beyond this study, SLCState provides a curated resource for comparing transporter conformations, analyzing state-dependent ligand binding, and developing other conformation-aware modeling approaches. Full curation procedures, state counts, and structural tree reconstruction methods are provided in Appendix A.



Symbol	Fold	PDB	Conformation	Ligand	SMILES
FLVCR1	MFS	8QCT	inward-facing	CHT	C[N+](C)(C)CCO
SLC1A5	Glt	6MPB	outward-occluded	GLN	NC(=O)CC[C@H](N)C(=O)O
…	…	…	…	…	…
SLC5A8	LeuT	7SL9	inward-facing	BUA	CCCC(=O)O

Figure 2: **SLCState construction and coverage.** A manually adjusted structural tree comprises 447 human SLC transporters. PDB retrieval and literature-guided curation yield chain-level structures and annotations; SLC5A8 (PDB: 7SL9) illustrates target-chain isolation, ligand-defined pocket extraction, and ligand representation. Yellow markers on the right-hand tree indicate benchmark coverage, and the accompanying counts summarize five conformational states. Curated records support train/test partitioning and downstream SLCHier modeling.

5.2 Generalization Protocols

We designed seven train/test splitting protocols to assess two types of generalization: protein and ligand generalization holds out entire proteins or ligand–transporter pairs to evaluate cross-protein transfer and prediction of new ligand associations on known proteins; generalization across conformational state spaces evaluates conformational discrimination on unseen proteins. Full definitions are provided in Appendix A.7.

Protein and ligand generalization. High-support protein holdout (HS) holds out entire proteins with well-represented folds in training, testing unseen-protein transfer with substantial same-fold support. Low-support protein holdout (LS) tests the same transfer with limited same-fold support. Ligand holdout within a transporter (LH) withholds all records of a ligand–transporter pair across structures and states, retaining other ligands for that transporter.

Generalization across conformational state spaces. All four protocols withhold test proteins from training. Two-state protein holdout (2S) tests all records of proteins with exactly two documented states of any combination, assessing two-state discrimination. Inward-/outward-facing two-state protein holdout (IO) selects proteins documented exclusively in these two facing states, testing directional discrimination under a fixed state composition. Protein holdout with inward-/outward-facing filtering (IOF) also admits proteins with additional states but tests only inward-/outward-facing records, excluding other states from the split. Multistate protein holdout (MS) tests all records of proteins with at least three documented states, assessing recognition among multiple conformational alternatives.

6 SLCHier: Hierarchy-Aware Hyperbolic Representation Learning

6.1 Transporter Hierarchy as a Structural Prior

Transporters exhibit substantial cross-protein structural organization. The structural resource used in this work provides pairwise distances among 447 human SLC transporters, whose hierarchical clustering yields a multi-scale organization broadly consistent with established folds such as MFS, LeuT, MitC, DMT, UraA, and NhaA. We therefore use this hierarchy as a structural prior for sharing information across transporters.

The branching nature of this organization further motivates a hyperbolic representation. Because hyperbolic space expands exponentially with radius, it provides a natural inductive bias for embedding

tree-like relationships with low distortion (Nickel & Kiela, 2017; 2018). The hierarchy is used as an architectural prior and its contribution is evaluated against matched Euclidean controls in Section 7.3.

6.2 Two-Scale Hyperbolic Representations

However, directly imposing transporter-level structure on individual conformations creates a tension: states of the same transporter should share global structural information, yet they must remain distinguishable for ligand-conditioned recognition. Treating pockets independently discards cross-transporter structure, whereas collapsing all conformations into one representation removes the state distinctions. SLCHier therefore separates the representation into two scales:

$$\underbrace{\{z_{t,s}^{(r)}\}_{s,r}}_{\text{state/pocket level}} \longrightarrow \underbrace{\mu_t}_{\text{transporter level}}, \tag{3}$$

where individual pocket representations support ligand recognition and the transporter center interfaces with the global structural hierarchy.

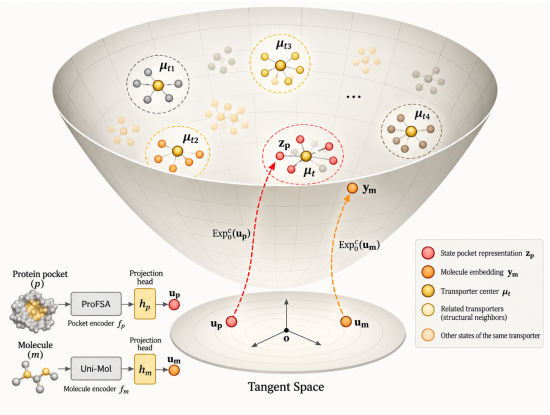


Figure 3: Overview of SLCHier. State-level pocket representations support ligand recognition, while their transporter-level centers capture structural organization in hyperbolic space.

State-level pocket–molecule representation. Let f_p and f_m denote the pocket and molecular encoders, initialized from ProFSA and Uni-Mol, and let h_p and h_m be independent projection heads into the tangent space at the Lorentz origin. The projected features are lifted to $\mathbb{H}_c^d$ through Exp_o^c, producing pocket and molecule embeddings z_p and y_m. Their compatibility is defined directly by geodesic distance, $f_\theta(p, m) = -d_c(z_p, y_m)^2$.

Transporter-level representation. For a training batch, let $\mathcal{B}_t$ denote the represented pocket files of transporter t, with $n_t = |\mathcal{B}_t|$. We summarize them using the equal-weight Lorentz gyromidpoint defined in Section 4: $\mu_t = G_c\left((z_p)_{p\in\mathcal{B}_t}; \alpha_p = \frac{1}{n_t}\right)$. Structural supervision acts on μ_t, while ligand recognition remains attached to the individual z_p; moreover, n_t counts represented pocket files and need not equal the number of canonical states K_t. This separation is theoretically compatible with state discrimination and we formalize this fixed-center expressivity result in Appendix B.

6.3 Training Objectives and Supervision

The two representation scales described in Section 6.2 are trained with complementary supervision: pocket–ligand objectives organize state-specific molecular recognition, whereas structural objectives organize transporter-level representations. Detailed formulations are provided in Appendix D.

Positive and negative supervision. Training uses both state-resolved pocket–ligand positives and coarser transporter-level positive associations. To provide broader molecular contrast, each batch is augmented with presumed negatives drawn from an FDA-approved drug library after removing all known positives. All non-positive pocket–molecule pairs within the batch act as contrastive negatives, while the other canonical states of the same transporter serve as state-aware competing negatives for state-resolved examples. Negative-construction details are provided in Appendix C.

Multi-positive pocket–molecule alignment. Let $\mathcal{B}_M$ denote the molecules in a training batch, $\mathcal{M}_p^+ \subseteq \mathcal{B}_M$ the positive molecules associated with pocket p, and $\mathcal{A}_P$ the pockets having at least one positive. We use a bidirectional multi-positive contrastive objective:

$$\begin{aligned}
\mathcal{L}_{p\to m} &= -\frac{1}{|\mathcal{A}_P|}\sum_{p\in\mathcal{A}_P}\frac{1}{|\mathcal{M}_p^+|}\sum_{m\in\mathcal{M}_p^+}\log\frac{\exp(f_\theta(p,m)/\tau_a)}{\sum_{m'\in\mathcal{B}_M}\exp(f_\theta(p,m')/\tau_a)}, \\
\mathcal{L}_{\text{align}} &= \frac{1}{2}\left(\mathcal{L}_{p\to m} + \mathcal{L}_{m\to p}\right),
\end{aligned} \tag{4}$$

where $\tau_a > 0$ is the contrastive temperature and the molecule-to-pocket direction is defined symmetrically. This provides the general pocket–molecule retrieval signal.

Within-transporter state ranking. To explicitly separate conformations of the same transporter, we aggregate the pocket scores within each canonical state using a smooth maximum. Let $A_{t,s}(m)$ denote the resulting score for state s, and let $H_{t,\neg s^\star}(m)$ denote the smooth maximum over all competing states. For a state-resolved positive $(m, t, s^\star)$, we use

$$\mathcal{L}_{\text{state}} = \mathbb{E}_{(m,t,s^\star)} \left[\tau_s \log \left(1 + \exp \frac{\gamma_s - A_{t,s^\star}(m) + H_{t,\neg s^\star}(m)}{\tau_s} \right) \right], \tag{5}$$

where $\gamma_s > 0$ is the ranking margin and $\tau_s > 0$ controls the smoothness of the ranking surrogate. This encourages the annotated state to outrank the alternative states of the same transporter. The exact smooth-max definition is given in the Appendix D.

Transporter-level structural supervision. At the coarser representation scale, we use two related but distinct structural signals. Let D_{tu} denote the pairwise structural dissimilarity between transporters t and u, and let T_{tu} denote their tree-path distance in the hierarchy obtained by hierarchical clustering of the structural distance matrix. Structural supervision acts only on transporter centers μ_t. We first match pairwise structural distance and relative ordering:

$$\begin{aligned} \mathcal{L}_{\text{struct}} &= \mathbb{E}_{(t,u)} \rho_\delta \left(d_c(\mu_t, \mu_u) - \beta D_{tu} \right), \\ \mathcal{L}_{\text{order}} &= \mathbb{E}_{(a,n,f)} \tau_o \log \left[1 + \exp \frac{d_c(\mu_a, \mu_n) - d_c(\mu_a, \mu_f) + \gamma_o}{\tau_o} \right], \end{aligned} \tag{6}$$

where ρ_δ is the Huber penalty with threshold δ, $\beta > 0$ rescales the external structural dissimilarity, $\gamma_o > 0$ is the ordering margin, and $\tau_o > 0$ controls the smoothness of the ordering surrogate. Here each triplet (a, n, f) is chosen such that n is structurally closer to anchor a than f.

To preserve local tree topology, we further define two neighborhood distributions over $u \neq t$,

$$P_t^{\text{tree}}(u) \propto \exp(-T_{tu}/\tau_T), \qquad P_t^{\text{hyp}}(u) \propto \exp(-d_c(\mu_t, \mu_u)/\tau_H),$$

and minimize $\mathcal{L}_{\text{nbr}} = \mathbb{E}_t D_{\text{KL}} \left(P_t^{\text{tree}} \,\|\, P_t^{\text{hyp}} \right)$. Here τ_T and τ_H are the temperatures of the tree-derived and hyperbolic neighborhood distributions. Thus, $\mathcal{L}_{\text{struct}}$ preserves structural scale, $\mathcal{L}_{\text{order}}$ preserves near–far ordering, and $\mathcal{L}_{\text{nbr}}$ preserves local hierarchical neighborhoods. Full definitions of the structural objectives and their training details are provided in Appendix D.

Bounded local dispersion and joint objective. To avoid excessive separation of pockets from their transporter summary without forcing collapse, we use $\mathcal{L}_{\text{local}} = \mathbb{E}_{t,p\in\mathcal{B}_t} \left[d_c(z_p, \mu_t) - r \right]_+^2$, where $r > 0$ is the local radius margin. The complete training objective is

$$\mathcal{L} = \mathcal{L}_{\text{align}} + \lambda_s \mathcal{L}_{\text{state}} + \lambda_d \mathcal{L}_{\text{struct}} + \lambda_o \mathcal{L}_{\text{order}} + \lambda_n \mathcal{L}_{\text{nbr}} + \lambda_l \mathcal{L}_{\text{local}}. \tag{7}$$

Here the λ coefficients control the relative strengths of the corresponding objectives. At inference, transporter centers and hierarchy supervision are discarded; candidate states are ranked solely from pocket–molecule scores.

7 Experiments

7.1 Experimental Setup

Training and model selection. SLCHier is trained on the training partition of each SLCState split. Hyperparameters and checkpoints are selected using a validation subset drawn exclusively from the corresponding training data, after which the selected checkpoint is frozen for test evaluation.

Baselines and evaluation protocol. We compare SLCHier with representative learning-based baselines, including RTMScore (Shen et al., 2022), TANKBind (Lu et al., 2022), DrugCLIP (Gao et al., 2023; Jia et al., 2026), BindNet (Feng et al., 2024), LigUnity (Feng et al., 2025), and HypSeek (Wang et al., 2026). All baselines are evaluated zero-shot. For each test record, the ligand and transporter identity are fixed and each method ranks the candidate conformational states using the maximum score over the structural representatives of each state. We report record-level top-1 accuracy (Acc, reported in percent) and mean reciprocal rank (MRR) as the primary evaluation metrics.

Table 1: Protein and ligand generalization on SLCState.

	HS		LS		LH	
Method	Acc ↑	MRR ↑	Acc ↑	MRR ↑	Acc ↑	MRR ↑
RTMScore (Shen et al., 2022)	34.62	0.6026	18.31	0.4826	35.90	0.6229
TANKBind (Lu et al., 2022)	65.38	0.8098	33.80	0.5965	28.21	0.6026
DrugCLIP (Gao et al., 2023; Jia et al., 2026)	61.54	0.7970	32.39	0.6120	39.74	0.6699
BindNet (Feng et al., 2024)	35.90	0.6560	49.30	0.6812	39.74	0.6549
LigUnity (Feng et al., 2025)	69.23	0.8376	32.39	0.5941	38.46	0.6581
HypSeek (Wang et al., 2026)	67.95	0.8098	35.21	0.6223	38.46	0.6688
SLCHier	**71.79**	**0.8419**	**52.11**	**0.7188**	**51.28**	**0.7254**

Table 2: State-space generalization on SLCState.

	2S		IO		IOF		MS	
Method	Acc ↑	MRR ↑	Acc ↑	MRR ↑	Acc ↑	MRR ↑	Acc ↑	MRR ↑
RTMScore (Shen et al., 2022)	42.68	0.7134	53.66	0.7683	60.76	0.7627	21.05	0.4825
TANKBind (Lu et al., 2022)	59.76	0.7988	51.22	0.7561	54.43	0.7500	47.37	0.6831
DrugCLIP (Gao et al., 2023; Jia et al., 2026)	75.61	0.8792	67.07	0.8394	59.49	0.7932	48.68	**0.7248**
BindNet (Feng et al., 2024)	71.95	0.8598	64.63	0.8232	43.04	0.6551	25.00	0.4912
LigUnity (Feng et al., 2025)	71.95	0.8598	63.41	0.8171	46.84	0.6951	51.32	0.6930
HypSeek (Wang et al., 2026)	69.51	0.8476	62.20	0.8110	50.63	0.7352	48.68	0.7072
SLCHier	**87.65**	**0.9383**	**75.61**	**0.8780**	**69.62**	**0.8418**	**53.95**	0.7226

7.2 Evaluation across SLCState

We evaluate SLCHier across all SLCState splits, grouping them into protein/ligand generalization (HS, LS, and LH) and conformational state-space generalization (2S, IO, IOF, and MS).

Protein and ligand generalization. As shown in Table 1, SLCHier consistently outperforms all baselines on HS, LS, and LH. On HS, it reaches 71.79% Acc and 0.8419 MRR, exceeding all baselines. Its advantage becomes clearer on LS, where it obtains 52.11% Acc and 0.7188 MRR. Particularly large gain appears on LH, where SLCHier reaches 51.28% Acc and 0.7254 MRR, improving over the strongest baseline by 11.54 percentage points in Acc (+29.0%). These results show consistent transfer to both unseen proteins and unseen ligand associations.

Conformational state-space generalization. As shown in Table 2, SLCHier achieves the highest Acc on all four state-space protocols, outperforming the strongest baseline by 12.04, 8.54, 8.86, and 2.63 percentage points on 2S, IO, IOF, and MS, respectively. It also obtains the best MRR on 2S, IO, and IOF; on MS, DrugCLIP is marginally higher in MRR, while SLCHier still achieves the best Acc (53.95%). Overall, SLCHier generalizes consistently from simple two-state discrimination to more complex multi-state recognition.

7.3 Ablation Studies

We conduct ablations on the low-support protein holdout (LS). As shown in Table 3, removing either structural distance/order supervision or tree-neighborhood matching reduces Acc from 52.11% to 33.80%, while shuffling the hierarchy further lowers it to 21.13%. Replacing Lorentz geometry with a matched Euclidean counterpart also substantially degrades performance to 29.58% Acc, supporting the importance of both meaningful hierarchy and hyperbolic representation.

Table 3: Ablation studies on LS.

Variant	Acc ↑	MRR ↑
Full SLCHier	**52.11**	**0.7188**
w/o state ranking	46.48	0.6894
w/o structural distance/order	33.80	0.6329
w/o tree neighborhood	33.80	0.6026
w/o FDA negatives	42.25	0.6425
w/o state-aware negatives	49.29	0.7034
Euclidean geometry	29.58	0.5756
Shuffled hierarchy	21.13	0.5289

The negative construction also contributes materially. Removing the FDA-derived background negatives decreases performance to 42.25% Acc and 0.6425 MRR, showing that broad molecular contrast provides useful supervision beyond the molecules already present in a training batch. Removing state-aware negatives or the explicit state-ranking objective also consistently degrades performance. These ablations show that the performance of SLCHier depends jointly on meaningful tree-structured supervision, hyperbolic representation learning, and the proposed negative-sampling and training design.

7.4 Cross-Protein Generalization and Multi-State Recognition

Beyond split-level results, we further examine two complementary capabilities: cross-protein generalization under varying same-fold training support, and conformational recognition as the number of candidate states increases.

Cross-protein generalization under limited support. Combining HS and LS, performance improves steadily with same-fold training support (Figure 4a): Acc rises from 36.36% with no support to 75.64% with more than five supporting transporters. Notably, SLCHier retains substantial recognition ability even without same-fold training examples.

Multi-state recognition with increasing candidate sets. Across 2S and MS, SLCHier remains highly accurate for two- and three-state systems and reaches 51.61% Acc and 70.16% MRR with four candidate states (Figure 4b), demonstrating recognition beyond binary conformational discrimination.

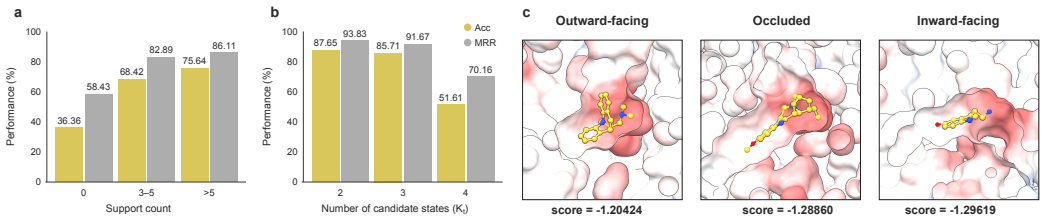


Figure 4: **Cross-protein generalization and multi-state recognition.** (a) Analysis on the number of same-fold training transporters (HS and LS combined). (b) Analysis on the number of candidate states (2S and MS combined). (c) SLC6A4/SERT pockets in three conformational states with co-crystallized ligands and SLCHier scores for a sertraline-like SSRI query.

7.5 Case Study: Conformational Recognition in SERT

To illustrate ligand-conditioned conformational recognition in a well-characterized system, we consider SLC6A4/SERT, for which SSRI inhibitors are known to stabilize an outward-open conformation (Coleman et al., 2016; Coleman & Gouaux, 2018). As shown in Figure 4c, for a sertraline-like SSRI inhibitor, SLCHier assigns the highest score to the annotated outward-facing state (-1.20424), ahead of the occluded (-1.28860) and inward-facing (-1.29619) alternatives. This case illustrates that SLCHier can distinguish ligand-associated conformations within the same transporter in a manner consistent with established SERT pharmacology.

8 Conclusion

We introduce ligand-conditioned conformational-state ranking as a distinct problem in transporter drug discovery and construct **SLCState**, a manually curated benchmark with seven complementary generalization protocols. We further propose **SLCHier**, a hierarchy-aware hyperbolic model that combines state-specific pocket–ligand recognition with transporter-level structural organization through a two-scale representation. Across SLCState, SLCHier consistently improves conformational-state recognition under diverse generalization and multi-state settings. Together, SLCState and SLCHier provide a unified framework for conformation-aware transporter drug discovery.

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

Gregor Bachmann, Gary Bécigneul, and Octavian Ganea. Constant curvature graph convolutional networks. In *International conference on machine learning*, pp. 486–496. PMLR, 2020.

Jonathan A Coleman and Eric Gouaux. Structural basis for recognition of diverse antidepressants by the human serotonin transporter. *Nature structural & molecular biology*, 25(2):170–175, 2018.

Jonathan A. Coleman, Evan M. Green, and Eric Gouaux. X-ray structures and mechanism of the human serotonin transporter. *Nature*, 532:334–339, 2016. doi: 10.1038/nature17629.

Jonathan A. Coleman, Dongxue Yang, Zhiyu Zhao, Po-Chao Wen, Craig Yoshioka, Emad Tajkhorshid, and Eric Gouaux. Serotonin transporter–ibogaine complexes illuminate mechanisms of inhibition and transport. *Nature*, 569:141–145, 2019. doi: 10.1038/s41586-019-1135-1.

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), 2025.

Shikun Feng, Minghao Li, Yinjun Jia, Wei-Ying Ma, and Yanyan Lan. Protein-ligand binding representation learning from fine-grained interactions. In *International Conference on Learning Representations*, volume 2024, pp. 10625–10640, 2024.

Evandro Ferrada and Giulio Superti-Furga. A structure and evolutionary-based classification of solute carriers. *iScience*, 25(10):105096, 2022. doi: 10.1016/j.isci.2022.105096.

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*, volume 36, 2023. doi: 10.52202/075280-1932.

Ulrich Goldmann, Tabea Wiedmer, Andrea Garofoli, Vitaly Sedlyarov, Manuel Bichler, Ben Haladik, Gernot Wolf, Eirini Christodoulaki, Alvaro Ingles-Prieto, Evandro Ferrada, Fabian Frommelt, Shao Thing Teoh, Philipp Leippe, Gabriel Onea, Martin Pfeifer, Mariah Kohlbrenner, Lena Chang, Paul Selzer, Jürgen Reinhardt, Daniela Digles, Gerhard F. Ecker, Tanja Osthushenrich, Aidan MacNamara, Anders Malarstig, David Hepworth, and Giulio Superti-Furga. Data- and knowledge-derived functional landscape of human solute carriers. *Molecular Systems Biology*, 21(6):599–631, 2025. doi: 10.1038/s44320-025-00108-2.

Yinjun Jia, Bowen Gao, Jiaxin Tan, Jiqing Zheng, Xin Hong, Wenyu Zhu, Haichuan Tan, Yuan Xiao, Liping Tan, Hongyi Cai, et al. Deep contrastive learning enables genome-wide virtual screening. *Science*, 391(6781):eads9530, 2026.

Huiyu Li and Ao Ma. Enhanced sampling of protein conformational changes via true reaction coordinates from energy relaxation. *Nature Communications*, 16:786, 2025. doi: 10.1038/s41467-025-55983-y.

Wei Lu, Qifeng Wu, Jixian Zhang, Jiahua Rao, Chengtao Li, and Shuangjia Zheng. Tankbind: Trigonometry-aware neural networks for drug-protein binding structure prediction. *Advances in neural information processing systems*, 35:7236–7249, 2022.

Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30, 2017.

Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *International conference on machine learning*, pp. 3779–3788. PMLR, 2018.

Shabareesh Pidathala, Shuyun Liao, Yaxin Dai, Xiao Li, Changkun Long, Chi-Lun Chang, Zhe Zhang, and Chia-Hsueh Lee. Mechanisms of neurotransmitter transport and drug inhibition in human VMAT2. *Nature*, 623:1086–1092, 2023. doi: 10.1038/s41586-023-06727-9.

Lena Rubi, Daniel Eckert, Stefan Boehm, Karlheinz Hilber, and Xaver Koenig. Anti-addiction drug ibogaine prolongs the action potential in human induced pluripotent stem cell-derived cardiomyocytes. *Cardiovascular Toxicology*, 17(2):215–218, 2017. doi: 10.1007/s12012-016-9366-y.

Chao Shen, Xujun Zhang, Yafeng Deng, Junbo Gao, Dong Wang, Lei Xu, Peichen Pan, Tingjun Hou, and Yu Kang. 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, 2022.

Isha Singh, Anubha Seth, Christian B. Billesbølle, Joao Braz, Ramona M. Rodriguiz, Kasturi Roy, Bethlehem Bekele, Veronica Craik, Xi-Ping Huang, Danila Boytsov, Vladimir M. Pogorelov, Parnian Lak, Henry O'Donnell, Walter Sandtner, John J. Irwin, Bryan L. Roth, Allan I. Basbaum, William C. Wetsel, Aashish Manglik, Brian K. Shoichet, and Gary Rudnick. Structure-based discovery of conformationally selective inhibitors of the serotonin transporter. *Cell*, 186(10): 2160–2175.e17, 2023. doi: 10.1016/j.cell.2023.04.010.

Aditya Sinha, Siqi Zeng, Makoto Yamada, and Han Zhao. Learning structured representations with hyperbolic embeddings. *Advances in Neural Information Processing Systems*, 37:91220–91259, 2024.

Chenyu Tang, Mayank Prakash Pandey, Cheng Giuseppe Chen, Alberto Megías, François Dehez, and Christophe Chipot. Breaking timescales with generative sampling of conformational transitions. *Nature*, 657:944–949, 2026. doi: 10.1038/s41586-026-11025-1.

Jianhui Wang, Wenyu Zhu, Bowen Gao, Xin Hong, Ya-Qin Zhang, Wei-Ying Ma, and Yanyan Lan. Learning protein–ligand binding in hyperbolic space. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 1150–1158, 2026.

Yuwei Wang, Pei Zhang, Yulin Chao, Zhini Zhu, Chuanhui Yang, Zixuan Zhou, Yaohui Li, Yonghui Long, Yuehua Liu, Dianfan Li, Sheng Wang, and Qianhui Qu. Transport and inhibition mechanism for VMAT2-mediated synaptic vesicle loading of monoamines. *Cell Research*, 34(1):47–57, 2024. doi: 10.1038/s41422-023-00906-z.

Di Wu, Qihao Chen, Zhuoya Yu, Bo Huang, Jun Zhao, Yuhang Wang, Jiawei Su, Feng Zhou, Rui Yan, Na Li, Yan Zhao, and Daohua Jiang. Transport and inhibition mechanisms of human VMAT2. *Nature*, 626:427–434, 2024. doi: 10.1038/s41586-023-06926-4.

A SLCSTATE: DETAILED CONSTRUCTION AND EVALUATION METHODS

A.1 DATASET RATIONALE AND CONSTRUCTION WORKFLOW

Public databases, including the Protein Data Bank (PDB), UniProt, the Transporter Classification Database (TCDB), and ChEMBL, provide complementary information on transporter structures, sequences, functions, classifications, and molecular activities. Training and evaluating models for conformation-dependent ligand recognition requires standardized samples grounded in experimentally determined structures, with consistent conformational annotations, clearly defined ligand identities and functional roles, and accurate pocket–ligand associations. Although existing resources provide essential components of this information, their records require integration and verification at the level of individual experimental structures before they can serve as training and evaluation data for this purpose. In particular, the presence of a ligand in a structure does not establish its functional role as a substrate or inhibitor, and protein-level activity measurements cannot be directly assigned to a specific experimentally observed conformation.

A major challenge in dataset curation is that the relevant information is dispersed across sources and described using inconsistent terminology. The evidence supporting conformational assignments and ligand functional roles is not always explicitly annotated in structure entries. Instead, it often needs to be retrieved from the main text, figure legends, and supplementary materials of the original publications and linked to specific protein chains and ligand instances. Automated retrieval and structural parsing therefore require subsequent literature review and visual inspection of the structures. To address these challenges, we constructed SLCState through a workflow comprising experimental structure collection, conformational and ligand annotation, chain-level structural curation, pocket extraction, and manual screening.

SLCState was designed specifically for transporters, integrating chain-level conformational annotations, binding pockets, and ligand information to capture their multiple conformational states and state-dependent ligand-binding relationships. In addition to supporting model training and evaluation in this study, the dataset provides a standardized basis for comparing pockets across transporter conformations, analyzing ligand-binding modes, and developing other conformation-aware modeling approaches.

A.2 EXPERIMENTAL STRUCTURE COLLECTION AND LITERATURE-BASED ANNOTATION

We searched the PDB for candidate experimental structures of human solute carrier (SLC) transporters using target protein names or gene symbols. We then verified protein identities and excluded entries that did not correspond to the intended targets, as well as structures containing only local fragments that did not adequately represent the main transporter architecture.

For each candidate entry, we reviewed the corresponding primary publication, including the main text, structural figures, and supplementary materials, to annotate the conformational state, ligand identity, and reported ligand functional role. Ligand identity and functional role were recorded separately to distinguish the chemical entity observed in a structure from its reported function as a substrate, inhibitor, activator, or other type of modulator. Conformational descriptions were standardized into five categories: inward-facing, inward-occluded, occluded, outward-occluded, and outward-facing. Synonymous descriptions were mapped to the corresponding categories. When terminology could not be mapped directly, assignments were based on descriptions of gating states and structural comparisons in the original publication. Conformational annotations were assigned to individual protein chains to avoid imposing a single label on subunits adopting different conformations within the same PDB entry.

A.3 CHAIN-LEVEL STRUCTURAL CURATION AND MANUAL REVIEW

Target transporter chains and their associated ligands were extracted from the source structures. Homologous subunits adopting different conformations within the same entry were retained as separate chain-level records. We then manually reviewed the extracted structures against the original publications and through structural visualization. This process removed non-target accessory proteins, such as antigen-binding fragments (Fabs), non-target chemical components, and substrate molecules located far from the target binding site. Additional protein components assigned to the same chain

identifier were also removed when they were disconnected from, and did not belong to, the main transporter structure.

Structures lacking a retained target ligand after curation were kept in the protein structure inventory under a ligand-free category but were excluded from pocket construction. The remaining structures were used for pocket extraction. The curated protein structure collection comprised 926 unique PDB entries, corresponding to 962 unique PDB ID–chain combinations.

A.4 Ligand-defined pocket extraction and ligand representation

For each ligand instance i, we defined a pocket residue set P_i comprising residues in the target protein chain with at least one heavy atom within 6.0 Å of any ligand heavy atom. All heavy-atom coordinates available in the source structure were retained for each selected residue. The resulting pocket structures contained only protein atoms and excluded ligand coordinates.

For chain-level structures containing multiple ligands, we used the overlap between ligand-defined pocket residue sets to determine whether the corresponding pockets should be merged. The Jaccard similarity between two pocket residue sets was defined as

$$J(P_i, P_j) = \frac{|P_i \cap P_j|}{|P_i \cup P_j|}. \tag{8}$$

We constructed a graph in which nodes represented ligand instances and an edge connected two nodes when $J(P_i, P_j) \geq 0.35$. For each connected component, the union of the corresponding pocket residue sets was used to generate a merged pocket. Ligand instances without connections to other nodes retained separate pockets. Multiple components confirmed to form a single covalently connected ligand were processed jointly as one complete ligand. Ligand instances that did not define a valid protein neighborhood were excluded from subsequent pocket construction.

Ligands associated with each pocket were identified from the source structure. SMILES representations were obtained from the Chemical Component Dictionary (CCD), validated and standardized using RDKit, and retained with their stereochemical and charge information. Peptide ligands comprising multiple residues were represented as complete molecules, whereas distinct ligands and ions were recorded separately.

A.5 Final sample organization and quality control

Following manual screening, retained pocket structures and ligand representations were compiled into a sample table containing protein identifiers, fold classes, PDB identifiers, conformational states, ligand identifiers, SMILES strings, and pocket filenames. Chain-level provenance remained traceable through the corresponding structure records.

The current release contains 926 unique PDB entries and 962 unique PDB ID–chain combinations in the protein structure collection. The pocket collection comprises 762 pocket–ligand records from 622 PDB entries, covering 109 protein identifiers and 14 fold classes. Of these records, 297 are inward-facing, 23 inward-occluded, 115 occluded, 58 outward-occluded, and 269 outward-facing.

A.6 Reconstruction of the SLC structural similarity tree

Following the RESOLUTE structural tree workflow described by Goldmann et al. (2025), we constructed a symmetric distance matrix for 447 human SLC proteins. Structural similarities were converted to distances as $d_{ij} = 1 - s_{ij}$, where s_{ij} denotes the structural similarity between proteins i and j. Hierarchical clustering was performed in R using the `ward.D2` method. The resulting branch lengths were square-root transformed, and the tree was visualized using an equal-angle unrooted layout implemented in the `ape` package.

A.7 Data splitting and generalization evaluation protocols

We organize seven protocols into two complementary groups aligned with the evaluation questions. Protein and ligand generalization (HS, LS, and LH) assesses transfer to unseen proteins under different levels of same-fold training support or to unseen ligand associations on known proteins.

Generalization across conformational state spaces (2S, IO, IOF, and MS) assesses state recognition on held-out proteins as the number and composition of documented states vary. These state-based protocols do not require state labels to be absent from training. The groups describe evaluation emphases rather than statistically independent factors or a guaranteed ordering of difficulty.

Protein and ligand generalization.

1. **High-support protein holdout (HS).** Proteins were assigned exclusively to either the training or test set. The fold classes of test proteins were well represented by distinct proteins in the training set. This protocol evaluates cross-protein generalization when substantial training support is available within the same fold class.
2. **Low-support protein holdout (LS).** Proteins were assigned exclusively to either the training or test set, with only limited representation of the test proteins' fold classes by distinct training proteins. Fold classes represented by a single protein could be entirely absent from training. This protocol evaluates cross-protein generalization under sparse training support within the same fold class.
3. **Ligand holdout within a transporter (LH).** Ligand–transporter pairs served as the holdout units. All records for a held-out pair, across structures and conformational states, were assigned to the test set, while records of the same transporter with other ligands were retained for training. This protocol evaluates generalization to unseen ligand associations for previously seen proteins. It does not require a test ligand to be absent from training records involving other proteins.

Generalization across conformational state spaces.

4. **Two-state protein holdout (2S).** Proteins with exactly two recorded conformational states were held out in their entirety, without restricting the particular combination of states. This protocol evaluates discrimination between two documented conformational states on unseen proteins, without restricting the state pair.
5. **Inward-/outward-facing two-state protein holdout (IO).** Proteins whose recorded conformational states consisted exclusively of inward-facing and outward-facing states were held out in their entirety, with both states included in the test set. This protocol evaluates discrimination between the two canonical facing directions on unseen proteins under a fixed inward-/outward-facing state composition.
6. **Protein holdout with inward-/outward-facing filtering (IOF).** Proteins with recorded states that included both inward-facing and outward-facing conformations were withheld from training. Only records in these two states were included in the test set; records in any other conformational states were excluded from that split. This protocol evaluates inward-/outward-facing discrimination across a broader eligible protein population than IO while maintaining the same two-state evaluation setting.
7. **Multistate protein holdout (MS).** Proteins with at least three recorded standardized conformational states were held out in their entirety, and all their conformational records were included in the test set. This protocol evaluates state recognition on unseen proteins among at least three documented conformational alternatives, rather than a two-state choice.

A.8 Conformational state accuracy and mean reciprocal rank

For N test records, let r_q denote the rank of the annotated state $s_q^\star$ under the evaluation scoring rule, and let $\widehat{s}_q$ denote the predicted state. The primary metrics are averaged over records:

$$\mathrm{Acc} = \frac{1}{N}\sum_{q\in\mathcal{D}^{\mathrm{te}}} \mathbf{1}\{\widehat{s}_q = s_q^\star\}, \qquad \mathrm{MRR} = \frac{1}{N}\sum_{q\in\mathcal{D}^{\mathrm{te}}} \frac{1}{r_q}, \tag{9}$$

where $\mathcal{D}^{\mathrm{te}}$ is the test set and $\mathbf{1}\{\cdot\}$ is the indicator function. Accuracy counts a prediction as correct only when the annotated state ranks first. Mean reciprocal rank (MRR) also assigns credit to lower-ranked annotated states, with ranks two, three, and beyond contributing $1/2$, $1/3$, and so on.

All models are evaluated using identical candidate pockets and the same within-state maximum aggregation rule. Candidate sets are fixed independently of the scoring model. Reproducible evaluation additionally requires explicit ranking conventions. Random-reference values and supplementary target-balanced reporting provide additional diagnostics. These diagnostics supplement, rather than replace, the record-averaged metrics defined above.

B Fixed-Center Expressivity of the Lorentz Gyromidpoint

We provide the theoretical justification for the two-scale representation used in SLCHier. The result below shows that constraining a transporter-level Lorentz gyromidpoint does not uniquely determine its constituent pocket representations: a high-dimensional family of local configurations remains available, and motions within this family can modify pocket–molecule scores while preserving the transporter center.

For completeness, $\mathrm{Log}_x^c(y)$ denotes the Riemannian logarithmic map on $\mathbb{H}_c^d$; its closed form is

$$\mathrm{Log}_x^c(y) = \frac{\operatorname{arcosh}(\xi)}{\sqrt{\xi^2-1}}\,(y-\xi x)\,, \qquad \xi = -c\langle x,y\rangle_L.$$

Theorem 1 (Fixed-center expressivity)**.** *Let*

$$G_c : (\mathbb{H}_c^d)^n \to \mathbb{H}_c^d$$

be the Lorentz gyromidpoint with fixed weights $\alpha_i > 0$ satisfying $\sum_{i=1}^n \alpha_i = 1$:

$$G_c(z_1,\dots,z_n) = \frac{v}{\zeta}, \qquad v = \sum_{i=1}^n \alpha_i z_i, \qquad \zeta = \sqrt{-c\langle v,v\rangle_L}.$$

Then G_c is a smooth, surjective submersion. At $Z = (z_1,\dots,z_n)$ with center $\mu = G_c(Z)$, its differential is

$$\mathrm{d}(G_c)_Z(\eta_1,\dots,\eta_n) = \zeta^{-1}\Pi_\mu\left(\sum_{i=1}^n \alpha_i\eta_i\right), \qquad \Pi_\mu w = w + c\langle\mu,w\rangle_L\mu. \tag{10}$$

Consequently, every fixed-center fiber $G_c^{-1}(\mu)$ is a smooth submanifold of dimension $(n-1)d$.

For equal weights $\alpha_i = 1/n$ and $n \geq 2$, let $h \in T_\mu\mathbb{H}_c^d$ and define

$$z_1(\varepsilon) = \mathrm{Exp}_\mu^c(\varepsilon h), \qquad z_2(\varepsilon) = \mathrm{Exp}_\mu^c(-\varepsilon h), \qquad z_i(\varepsilon) = \mu \quad (i \geq 3).$$

Then

$$G_c\big(z_1(\varepsilon),\dots,z_n(\varepsilon)\big) = \mu$$

for every ε. Moreover, for any $y \in \mathbb{H}_c^d$ with $y \neq \mu$,

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left[-d_c(z_1(\varepsilon),y)^2 + d_c(z_2(\varepsilon),y)^2\right]\Big|_{\varepsilon=0} = 4\left\langle \mathrm{Log}_\mu^c(y), h\right\rangle_L. \tag{11}$$

Proof. Define the normalization map

$$N(v) = \frac{v}{\sqrt{-c\langle v,v\rangle_L}}.$$

Since a positive weighted sum of future-directed timelike vectors remains future-directed timelike, $v = \sum_i \alpha_i z_i$ satisfies $-c\langle v,v\rangle_L > 0$. Hence N and therefore G_c are smooth.

Let

$$\delta v = \sum_{i=1}^n \alpha_i \eta_i.$$

Differentiating $\zeta = (-c\langle v,v\rangle_L)^{1/2}$ gives

$$\mathrm{d}\zeta_v(\delta v) = -\frac{c\langle v,\delta v\rangle_L}{\zeta}.$$

Using $v = \zeta\mu$, we obtain

$$\begin{aligned} \mathrm{d}N_v(\delta v) &= \frac{\delta v}{\zeta} - \frac{v}{\zeta^2}\mathrm{d}\zeta_v(\delta v) \\ &= \frac{1}{\zeta}\left(\delta v + c\langle \mu, \delta v\rangle_L \mu\right) \\ &= \zeta^{-1}\Pi_\mu(\delta v), \end{aligned} \tag{12}$$

which proves equation 10.

The map is surjective because, for any $\mu \in \mathbb{H}^d_c$,

$$G_c(\mu, \ldots, \mu) = \mu.$$

To show that G_c is a submersion, note that Π_μ is the Lorentz-orthogonal projection onto $T_\mu\mathbb{H}^d_c$, with

$$\ker \Pi_\mu = \mathrm{span}\{\mu\}.$$

For any $z_i \in \mathbb{H}^d_c$, the restriction

$$\Pi_\mu\big|_{T_{z_i}\mathbb{H}^d_c} : T_{z_i}\mathbb{H}^d_c \to T_\mu\mathbb{H}^d_c$$

is injective. Indeed, if a nonzero vector in $T_{z_i}\mathbb{H}^d_c$ belonged to $\mathrm{span}\{\mu\}$, then $\mu \in T_{z_i}\mathbb{H}^d_c$, which would imply $\langle z_i, \mu\rangle_L = 0$. This is impossible because

$$-c\langle z_i, \mu\rangle_L = \cosh\big(\sqrt{c}\, d_c(z_i, \mu)\big) \geq 1.$$

Both tangent spaces have dimension d, so this restriction is an isomorphism. Since every $\alpha_i > 0$, any tangent vector in $T_\mu\mathbb{H}^d_c$ can therefore be realized by equation 10. Thus $\mathrm{d}(G_c)_Z$ is surjective at every Z, and G_c is a submersion.

Because $(\mathbb{H}^d_c)^n$ has dimension nd and $\mathbb{H}^d_c$ has dimension d, the regular level-set theorem gives

$$\dim G_c^{-1}(\mu) = nd - d = (n-1)d.$$

We next consider the explicit equal-weight perturbation. By the Lorentz exponential-map formula,

$$\begin{aligned} \mathrm{Exp}^c_\mu(\varepsilon h) &= \cosh(a)\mu + \frac{\sinh(a)}{\sqrt{c}\,\|h\|_L} h, \\ \mathrm{Exp}^c_\mu(-\varepsilon h) &= \cosh(a)\mu - \frac{\sinh(a)}{\sqrt{c}\,\|h\|_L} h, \end{aligned}$$

where $a = \sqrt{c}\,\varepsilon\|h\|_L$. Hence

$$z_1(\varepsilon) + z_2(\varepsilon) = 2\cosh(a)\mu.$$

Together with $z_i(\varepsilon) = \mu$ for $i \geq 3$, the unnormalized equal-weight center is

$$v(\varepsilon) = \frac{2\cosh(a) + n - 2}{n}\,\mu = \kappa(\varepsilon)\mu,$$

where $\kappa(\varepsilon) > 0$. Since $\langle \mu, \mu\rangle_L = -1/c$,

$$\sqrt{-c\langle v(\varepsilon), v(\varepsilon)\rangle_L} = \kappa(\varepsilon),$$

and therefore

$$G_c\big(z_1(\varepsilon), \ldots, z_n(\varepsilon)\big) = \mu.$$

Finally, on hyperbolic space the first-variation identity for squared geodesic distance is

$$\mathrm{grad}_x \frac{1}{2} d_c(x, y)^2 = -\,\mathrm{Log}^c_x(y).$$

Because

$$\dot z_1(0) = h, \qquad \dot z_2(0) = -h,$$

we have

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left[-d_c(z_1(\varepsilon), y)^2\right]\Big|_{\varepsilon=0} &= 2\left\langle \mathrm{Log}^c_\mu(y), h\right\rangle_L, \\ \frac{\mathrm{d}}{\mathrm{d}\varepsilon} d_c(z_2(\varepsilon), y)^2\Big|_{\varepsilon=0} &= 2\left\langle \mathrm{Log}^c_\mu(y), h\right\rangle_L. \end{aligned}$$

Adding the two derivatives yields equation 11. □

The theorem establishes a capacity statement rather than a guarantee of state separation. Center-level structural supervision constrains only d collective degrees of freedom, leaving $(n-1)d$ local degrees of freedom available to the constituent pocket representations. Direct state-ranking supervision is therefore still required to make use of this capacity for ligand-conditioned conformational discrimination.

C Positive Associations and Negative Construction

We provide additional details on the molecular supervision used to train SLCHier. The training data contain two levels of positive associations. A *state-resolved positive* associates a molecule m with a specific canonical state s of transporter t, whereas a *transporter-level positive* specifies that m is associated with transporter t without assigning it to a particular canonical state. State-resolved positives contribute to both pocket–molecule alignment and within-transporter state ranking, while transporter-level positives are used only for the general alignment objective.

Background molecular negatives. To provide molecular contrast beyond the observed positive associations, we construct an additional background pool from an FDA-approved drug library. For each training split, let $\mathcal{M}_{\mathrm{FDA}}$ denote the molecules in this library and let

$$\mathcal{M}^{+}_{\mathrm{known}}$$

contain every molecule with either a state-resolved or transporter-level positive annotation in the corresponding training partition. The background pool is

$$\mathcal{M}^{-}_{\mathrm{bg}} = \mathcal{M}_{\mathrm{FDA}} \setminus \mathcal{M}^{+}_{\mathrm{known}}. \tag{13}$$

These molecules are treated as *presumed negatives* for contrastive training. Importantly, absence from the positive set is not interpreted as experimental evidence of non-binding.

Batch-level contrastive negatives. A molecular batch is first populated with the positive molecules required by the sampled transporters and is then supplemented with molecules drawn from $\mathcal{M}^{-}_{\mathrm{bg}}$. The resulting negative supervision is broader than the explicit FDA-derived pool. For each pocket, all molecules in the batch that are not annotated as its positives remain eligible comparisons in the contrastive denominator. Consequently, a molecule that is positive for one transporter or pocket can act as a negative comparison for another pocket for which no positive association is recorded.

State-aware competing negatives. State-resolved examples additionally induce a second form of negative supervision at the conformational level. For a positive association $(m, t, s^{\star})$, all other canonical states

$$s \in \mathcal{S}_t \setminus \{s^{\star}\}$$

are treated as competing states in the state-ranking objective. This distinguishes state-aware negatives from molecular contrastive negatives: the former compare alternative conformations of the same transporter, whereas the latter distinguish compatible from non-positive molecules in the shared pocket–molecule representation space.

Thus, training combines three complementary sources of negative signal: FDA-derived background molecules, non-positive pocket–molecule pairs within each batch, and competing canonical states of the same transporter.

D Detailed Training Objectives

We provide the complete definitions of the training objectives summarized in Section 6.3. Throughout, pocket–molecule compatibility is

$$f_{\theta}(p, m) = -d_c(z_p, y_m)^2,$$

where $z_p, y_m \in \mathbb{H}^d_c$ are the hyperbolic pocket and molecule representations.

D.1 Multi-Positive Pocket–Molecule Alignment

Let $\mathcal{B}_M$ denote the set of molecules in a training batch, and let $\mathcal{M}_p^+ \subseteq \mathcal{B}_M$ be the positive molecules associated with pocket p. Denote by

$$\mathcal{A}_P = \{p : \mathcal{M}_p^+ \neq \varnothing\}$$

the set of pockets having at least one positive molecule. The pocket-to-molecule objective is

$$\mathcal{L}_{p\to m} = -\frac{1}{|\mathcal{A}_P|} \sum_{p\in\mathcal{A}_P} \frac{1}{|\mathcal{M}_p^+|} \sum_{m\in\mathcal{M}_p^+} \log \frac{\exp(f_\theta(p,m)/\tau_a)}{\sum_{m'\in\mathcal{B}_M} \exp(f_\theta(p,m')/\tau_a)}. \tag{14}$$

The reverse direction is defined analogously. Let $\mathcal{P}_m^+$ denote the positive pockets associated with molecule m, let $\mathcal{B}_P$ denote the pockets in the batch, and define

$$\mathcal{A}_M = \{m : \mathcal{P}_m^+ \neq \varnothing\}.$$

Then

$$\mathcal{L}_{m\to p} = -\frac{1}{|\mathcal{A}_M|} \sum_{m\in\mathcal{A}_M} \frac{1}{|\mathcal{P}_m^+|} \sum_{p\in\mathcal{P}_m^+} \log \frac{\exp(f_\theta(p,m)/\tau_a)}{\sum_{p'\in\mathcal{B}_P} \exp(f_\theta(p',m)/\tau_a)}. \tag{15}$$

The bidirectional alignment loss is

$$\mathcal{L}_{\text{align}} = \frac{1}{2}\left(\mathcal{L}_{p\to m} + \mathcal{L}_{m\to p}\right). \tag{16}$$

Multiple experimentally supported associations are retained as positives rather than collapsed to a single matched pair.

D.2 Within-Transporter State Ranking

A canonical state may contain multiple structural pocket representatives. For differentiable training, we use the smooth maximum

$$\operatorname{smax}_\tau(a_1,\ldots,a_n) = \tau \log \sum_{i=1}^{n} \exp(a_i/\tau). \tag{17}$$

For transporter t, state s, and molecule m, define the aggregated state score

$$A_{t,s}(m) = \operatorname{smax}_{\tau_p}\left\{f_\theta(p,m) : p \in \mathcal{B}_t \cap \mathcal{P}_{t,s}\right\}. \tag{18}$$

For a state-resolved positive $(m,t,s^\star)$, the strongest competing state is approximated by

$$H_{t,\neg s^\star}(m) = \operatorname{smax}_{\tau_s}\left\{A_{t,s}(m) : s \in \mathcal{S}_t \setminus \{s^\star\}\right\}. \tag{19}$$

The state-ranking loss is

$$\mathcal{L}_{\text{state}} = \mathbb{E}_{(m,t,s^\star)}\left[\tau_s \log\left(1 + \exp\frac{\gamma_s - A_{t,s^\star}(m) + H_{t,\neg s^\star}(m)}{\tau_s}\right)\right]. \tag{20}$$

This objective encourages the annotated state to exceed the strongest competing state by margin γ_s. The smooth aggregation is used only during training; evaluation uses the exact max-over-pocket state score defined in Section 3.

D.3 Transporter-Level Structural Supervision

Let D_{tu} denote the pairwise structural dissimilarity between transporters t and u, and let T_{tu} denote their path distance in the hierarchy reconstructed from the structural similarity matrix. Structural supervision is applied only to the transporter-level centers μ_t.

Pairwise structural distance. For the set $\mathcal{Q}$ of transporter pairs available in a structural training batch, we use

$$\mathcal{L}_{\text{struct}} = \frac{1}{|\mathcal{Q}|} \sum_{(t,u)\in\mathcal{Q}} \rho_\delta\left(d_c(\mu_t,\mu_u) - \beta D_{tu}\right), \tag{21}$$

where ρ_δ is the Huber penalty and β controls the scale between the external structural dissimilarity and the learned hyperbolic distance.

Relative structural ordering. Absolute structural distance need not be reconstructed perfectly for its relative ordering to remain informative. Let $\mathcal{R}$ denote a set of ordered triplets (a, n, f) such that transporter n is structurally closer to anchor a than transporter f. We minimize

$$\mathcal{L}_{\text{order}} = \frac{1}{|\mathcal{R}|} \sum_{(a,n,f)\in\mathcal{R}} \tau_o \log\left[1 + \exp\frac{d_c(\mu_a, \mu_n) - d_c(\mu_a, \mu_f) + \gamma_o}{\tau_o}\right]. \tag{22}$$

This term requires structurally nearer transporters to remain closer in the learned representation than structurally more distant ones.

Tree-neighborhood matching. The structural tree provides an additional local-topology signal. For each anchor transporter t, define the tree-derived neighborhood distribution

$$P^{\text{tree}}_{tu} = \frac{\exp(-T_{tu}/\tau_T)}{\sum_{v\neq t}\exp(-T_{tv}/\tau_T)}, \qquad u \neq t, \tag{23}$$

and the corresponding distribution induced by hyperbolic center distances

$$P^{\text{hyp}}_{tu} = \frac{\exp(-d_c(\mu_t, \mu_u)/\tau_H)}{\sum_{v\neq t}\exp(-d_c(\mu_t, \mu_v)/\tau_H)}. \tag{24}$$

We then minimize

$$\mathcal{L}_{\text{nbr}} = \mathbb{E}_t D_{\text{KL}}\left(P^{\text{tree}}_t \,\|\, P^{\text{hyp}}_t\right). \tag{25}$$

Unlike $\mathcal{L}_{\text{struct}}$, this objective does not require hyperbolic distances to numerically reproduce tree-path lengths; it only requires their induced neighborhood distributions to agree.

D.4 Bounded Local Dispersion and Joint Objective

Transporter-level structural supervision should organize the center μ_t without collapsing all constituent pockets onto it. We therefore penalize only excessive pocket dispersion:

$$\mathcal{L}_{\text{local}} = \text{mean}_{t,\, p\in\mathcal{B}_t}\left[d_c(z_p, \mu_t) - r\right]^2_+, \tag{26}$$

where $[x]_+ = \max(x, 0)$. Hence pockets incur no local penalty while they remain within radius r of their transporter center.

The complete objective is

$$\mathcal{L} = \mathcal{L}_{\text{align}} + \lambda_s\mathcal{L}_{\text{state}} + \lambda_d\mathcal{L}_{\text{struct}} + \lambda_o\mathcal{L}_{\text{order}} + \lambda_n\mathcal{L}_{\text{nbr}} + \lambda_l\mathcal{L}_{\text{local}}. \tag{27}$$

All objectives are optimized jointly. Because the transporter center μ_t is computed differentiably from the current pocket embeddings, gradients from the structural objectives propagate through μ_t to the constituent pockets and the pocket encoder.

At inference, transporter centers and structural hierarchy information are not used as scoring objects. Given a ligand and transporter, SLCHier scores each candidate pocket independently using $f_\theta(p, m)$, aggregates multiple pocket representatives of the same canonical state using the exact maximum, and ranks the resulting state scores.