UniPocket: Physics-Aware Geometric Graph Learning with Manifold Completeness for Ligand-Specific Binding Site Prediction

Kangxin Chen[1] , **Jieyu Zhao**[1,*] , **Jinli Hu**[1] and **Min Xie**[2]
[1]Ningbo University
[2]Chongqing University of Technology
chen_kangxin@outlook.com, zhao_jieyu@nbu.edu.cn, hu_jinli@outlook.com, cqxiemin@gmail.com

Abstract

Predicting ligand binding sites on protein surfaces requires capturing complex local geometries and satisfying physical constraints. Existing voxel-based methods suffer from high computational costs and rotation sensitivity, while standard point-cloud GNNs often lack geometric completeness—failing to distinguish chiral structures or subtle topological variations. In this work, we propose **UniPocket**, a novel E(3)-equivariant surface graph neural network. Inspired by recent advances in efficient geometric completeness, UniPocket constructs a Manifold-Aware Surface Encoder that utilizes local surface normals as virtual reference frames to capture complete geometric invariants without expensive high-order tensor products. Furthermore, we introduce a Ligand-Gated Message Passing mechanism to condition the surface features on the chemical semantics of the target ligand, and a Physics-Aware Vector Rejection Module that enforces steric constraints via orthogonal vector decomposition. Experimental results on standard benchmarks (PDBbind, COACH420, Holo4k) demonstrate that UniPocket achieves state-of-the-art performance, validating that geometric completeness and physical inductive biases are key to precise binding site detection.

1 Introduction

The precise identification of protein-ligand binding sites (pockets) is a prerequisite for structure-based drug design (SBDD) and virtual screening [Vural and Jololian, 2025]. While recent foundation models like AlphaFold 3 [Abramson *et al.*, 2024] and RoseTTAFold All-Atom [Krishna *et al.*, 2024] show unprecedented accuracy in biomolecular interaction modeling, they are computationally intensive and primarily designed for holo-structure prediction given a known ligand. They are not suited for rapidly scanning apo-structures to identify potential binding cavities. Therefore, efficient and accurate ligand-specific binding site detection on rigid protein surfaces remains a critical, standalone task.

*Corresponding author

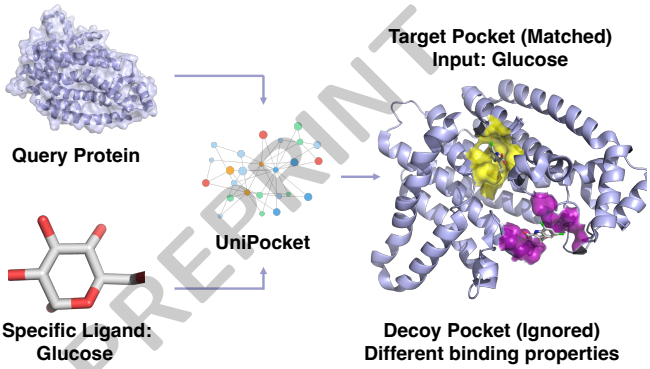


Figure 1: **Illustration of ligand-specific pocket identification.** The target pocket (yellow) specifically binds to Glucose, while the decoy pocket (purple) is rejected.

Traditional approaches for pocket detection, including geometry-based (e.g., Fpocket [Le Guilloux *et al.*, 2009]) and template-based methods [Desaphy *et al.*, 2015], along with early ML methods (e.g., P2Rank [Krivák and Hoksza, 2018], DeepSite [Jiménez *et al.*, 2017]) using handcrafted features or voxelized grids, often suffer from rotation variance and sparsity. Geometric Deep Learning (GDL) [Bronstein *et al.*, 2021; Townshend *et al.*, 2021] addressed some of these limitations by modeling proteins as geodesic surfaces [Gainza *et al.*, 2020; Mylonas *et al.*, 2021; Tubiana *et al.*, 2022] or graphs [Satorras *et al.*, 2021], enabling the capture of invariant topological patterns. However, most existing surface encoders primarily focus on protein shape complementarity, neglecting explicit physicochemical compatibility with the specific query ligand. This creates a critical expressivity-efficiency trade-off in GDL for proteins: high-order equivariant networks (e.g., TFN [Thomas *et al.*, 2018], MACE [Batatia *et al.*, 2022]) provide theoretical completeness but incur prohibitive cubic computational costs, while efficient distance-based GNNs (e.g., EGNN [Satorras *et al.*, 2021]) often lack geometric completeness crucial for subtle local topologies and precise binding site prediction. Furthermore, recent ligand-aware methods like LABind [Zhang *et al.*, 2025a] and FABind [Pei *et al.*, 2023] incorporate ligand information but often rely on complex mechanisms, increasing computational complexity or lacking the fine-grained ge-

ometric completeness needed for accurate pocket detection.

To bridge these gaps, we propose UniPocket, a novel E(3)-equivariant surface graph neural network. UniPocket formulates protein pocket detection on a discretized Riemannian manifold, employing a Manifold-Aware Surface Encoder. Instead of heavy tensor products, it leverages intrinsic protein surface geometry (surface normals and curvatures) to construct local reference frames (virtual nodes), achieving efficient geometric completeness. This approach effectively captures anisotropic geometric invariants (e.g., curvature tensors) crucial for high-precision pocket recognition, without the computational overhead of high-order spherical harmonics. Combined with explicit physical constraints, UniPocket offers a robust and efficient solution for specific binding site detection.

Our main contributions are summarized as follows:

- **Manifold-Aware Geometric Completeness with Surface-Normal-Based Encoder:** We propose a Surface-Normal-Based Encoder that achieves geometric completeness on protein manifolds. By utilizing surface normals as local reference frames, our method efficiently captures anisotropic geometric invariants (e.g., curvature tensors) necessary for high-precision pocket recognition, avoiding the computational overhead of high-order spherical harmonics.
- **Conditional Ligand-Gated Message Passing for Ligand-Specificity:** We introduce a conditional graph operator that dynamically modulates the protein manifold features based on the target ligand's chemical graph. This mechanism transforms the generic pocket detection task into a ligand-specific geometric reasoning problem, effectively pruning biologically irrelevant cavities.
- **Physics-Aware Inductive Bias via Vector Decomposition:** We design a Physics-Aware Vector Rejection Module as a differentiable geometric inductive bias. By mathematically decomposing interaction vectors into orthogonal normal (clash-inducing) and tangential (surface-fitting) components, we explicitly enforce hard steric constraints within the neural architecture, ensuring physically valid predictions.

2 Related Work

2.1 Efficient Geometric Deep Learning

Geometric Deep Learning (GDL) enforces physical symmetries (e.g., SE(3)-equivariance) in molecular modeling [Bronstein *et al.*, 2021]. Early methods like TFN [Thomas *et al.*, 2018] and NequIP [Batzner *et al.*, 2022] achieved theoretical completeness but suffered from cubic computational complexity. To improve efficiency, scalarization-based approaches such as E(n)-GNN [Satorras *et al.*, 2021], SchNet [Schütt *et al.*, 2017], and ViSNet [Wang *et al.*, 2024] utilized inter-atomic distances and vector-scalar interactions. However, many distance-based GNNs lack geometric completeness, struggling with chiral structures or complex anisotropic patterns. Recent work, like Uni-EGNN [Keriven and Peyré, 2019], demonstrates that local reference frames can achieve efficient geometric completeness. Following this paradigm, UniPocket employs intrinsic surface normals as natural reference frames to construct a geometrically complete surface encoder, balancing high expressivity with linear efficiency.

2.2 Evolution of Binding Site Prediction

Binding site prediction has evolved from traditional geometric/ML methods (e.g., Fpocket [Le Guilloux *et al.*, 2009], P2Rank [Krivák and Hoksza, 2018]) to deep learning approaches. Early deep learning utilized 3D-CNNs (e.g., DeepSite [Jiménez *et al.*, 2017], Kalasanty [Stepniewska-Dziubinska *et al.*, 2020]) operating on voxel grids, which suffered from high memory costs, discretization errors, and rotation sensitivity despite improvements like PuResNet [Kandel *et al.*, 2021]. Subsequently, methods shifted to manifold learning (e.g., MaSIF [Gainza *et al.*, 2020], DeepSurf [Mylonas *et al.*, 2021], ScanNet [Tubiana *et al.*, 2022]) or point-cloud processing (e.g., PointSite [Yan *et al.*, 2022], EquiPocket [Zhang *et al.*, 2024]) to overcome grid limitations. While these capture static geometry, they often operate "blindly," neglecting ligand-specific chemical constraints and lacking explicit mechanisms to handle physical clashes.

2.3 Ligand-Aware Prediction and Docking

The trend in 2024-2025 is moving towards ligand-specific prediction. Regression-based frameworks such as TankBind [Lu *et al.*, 2022] and FABind [Pei *et al.*, 2023] integrate pocket prediction before docking. Other notable approaches like LABind [Zhang *et al.*, 2025a] and MP-Bind [Wang *et al.*, 2025] explicitly learn ligand-protein interactions using graph transformers or protein language models. Generative diffusion models (e.g., DiffDock [Corso *et al.*, 2023], SurfDock [Cao *et al.*, 2025]) implicitly identify binding sites during pose generation, but are often computationally expensive and may produce physically invalid structures lacking hard geometric constraints [Buttenschoen *et al.*, 2024]. UniPocket differentiates itself by offering a lightweight, geometrically complete, and physics-driven filtering mechanism, specifically designed for rapid and accurate screening of large datasets like PDBbind [Wang *et al.*, 2004] and scPDB [Desaphy *et al.*, 2015].

3 Methodology

3.1 Problem Formulation

The objective of rigid protein-ligand binding site prediction is to identify the subset of surface points on a protein that are most likely to interact with a specific query ligand. Let a protein structure be represented by its surface point cloud $\mathcal{P} = \{\mathbf{x}_i, \mathbf{h}_i\}_{i=1}^{N}$, where $\mathbf{x}_i \in \mathbb{R}^3$ denotes the spatial coordinate and $\mathbf{h}_i \in \mathbb{R}^{d_P}$ represents the physicochemical features of the i-th point. Similarly, a query ligand is defined as a molecular graph $\mathcal{G}_L = (\mathcal{V}_L, \mathcal{E}_L)$, where nodes $\mathcal{V}_L$ represent atoms and edges $\mathcal{E}_L$ represent chemical bonds. Our goal is to learn a mapping function $\mathcal{F}_\theta(\mathcal{P}, \mathcal{G}_L) \to \mathbf{Y}$, where $\mathbf{Y} \in [0, 1]^N$ represents the probability map indicating the binding likelihood of each surface point.

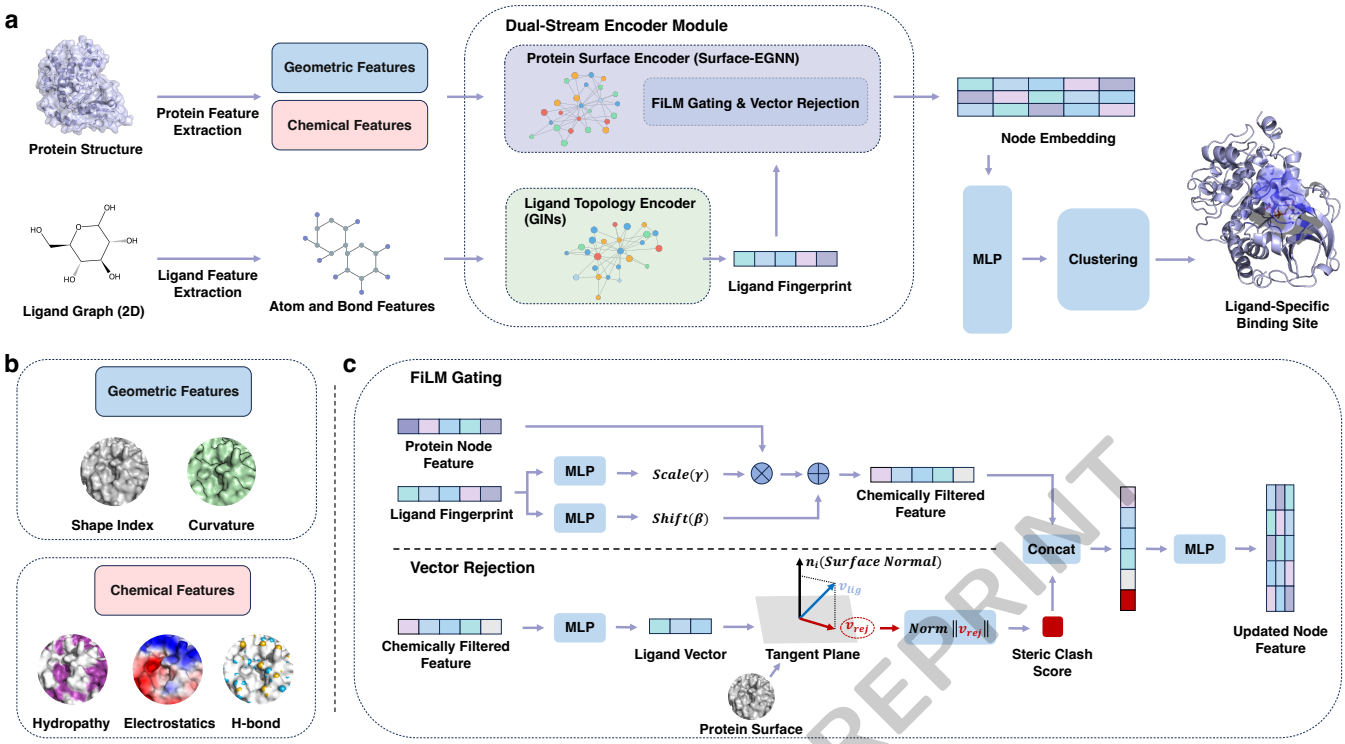


Figure 2: **The overall framework of UniPocket.** **a**, The model adopts a dual-stream architecture comprising a Ligand Topology Encoder (using GINs) to extract the ligand fingerprint and a Protein Surface Encoder (using Surface-EGNN) to process the protein surface graph. The protein node embeddings are dynamically modulated by the ligand fingerprint, followed by an MLP and clustering module to identify ligand-specific binding sites. **b**, Visualization of the input features extracted from the protein surface, including Geometric Features (Shape Index, Curvature) and Chemical Features (Hydropathy, Electrostatics, H-bond potential). **c**, Detailed illustration of the proposed interaction mechanisms within the message passing layer: (1) **FiLM Gating**, which applies ligand-conditioned affine transformations (Scale γ and Shift β) to filter chemically incompatible features; (2) **Vector Rejection**, which computes a steric clash score by projecting the predicted ligand vector onto the surface normal $\mathbf{n}_i$ relative to the tangent plane, explicitly enforcing geometric constraints.

3.2 Input Representation and Graph Construction

Protein Surface Modeling

Unlike voxel-based methods that suffer from sparsity issues, we model the protein surface as a geometric graph. Initially, the Solvent Excluded Surface (SES) [Lee and Richards, 1971] is generated using the MSMS tool [Sanner *et al.*, 1996] with a probe radius of 1.5 Å. To balance resolution and computational efficiency, we perform geodesic-based uniform sampling to obtain a fixed number of points N (default $N = 2000$).

We construct a sparse k-nearest neighbor (k-NN) graph $\mathcal{G}_P = (\mathcal{V}_P, \mathcal{E}_P)$ on these points. For each node i, the initial feature vector $\mathbf{h}_i^{(0)}$ is a concatenation of geometric and chemical descriptors:

- **Geometric Features**: As shown in Figure 2b, we compute the local curvature (mean and Gaussian), shape index, and the unit normal vector $\mathbf{n}_i \in \mathbb{R}^3$ pointing outward from the surface. The normal vector is critical for our subsequent vector rejection module.
- **Chemical Features**: We project properties from the nearest protein atoms to the surface points. These include hydropathy scores (Kyte-Doolittle scale), partial charges, and hydrogen bond donor/acceptor boolean flags.

Mathematically, $\mathbf{h}_i^{(0)} = [\text{Curv}_i \,\|\, \text{Chem}_i \,\|\, \mathbf{n}_i] \in \mathbb{R}^{d_{in}}$.

Ligand Topological Graph

To ensure the model handles flexible ligands robustly without relying on potentially inaccurate 3D conformer generation, we represent the ligand strictly as a 2D topological graph. The node features $\mathbf{z}_u$ ($u \in \mathcal{V}_L$) consist of one-hot encodings of atomic number (C, N, O, F, P, S, Cl, Br, I), degree of hybridization, chirality, and aromaticity. Edge features $\mathbf{e}_{uv}$ encode bond types (single, double, triple, aromatic) and conjugated system flags. This pure 2D representation ensures that our encoder learns a conformer-invariant fingerprint of the ligand.

3.3 Dual-Stream Encoder Architecture

As illustrated in Figure 2a, our model consists of two parallel encoders.

Protein Encoder: Surface-EGNN

To capture the intrinsic geometry of the protein surface while maintaining symmetry requirements, we employ a Surface-Equivariant Graph Neural Network (Surface-EGNN). Standard GNNs are invariant to rotation, which loses directional

Method	Type	Params (M)	PDBbind 2020 (Test)			COACH420 (Generalization)			Holo4k (Generalization)		
			DCC ↓	DCA ↓	Succ. ↑	DCC ↓	DCA ↓	Succ. ↑	DCC ↓	DCA ↓	Succ. ↑
Fpocket	Geometric	-	3.12	3.45	40.1	2.85	3.10	45.2	3.50	4.10	38.5
P2Rank	ML-based	1.0	1.65	2.10	62.3	1.50	1.80	68.5	1.95	2.30	55.4
DeepSurf	3D-CNN	33.1	1.35	1.60	68.9	1.20	1.45	72.1	1.65	2.10	60.2
Kalasanty	3D-CNN	70.6	1.48	1.82	61.2	1.35	1.55	65.4	1.80	2.25	58.1
ViSNet	GNN	1.2	1.05	1.25	75.4	0.95	1.10	78.2	1.25	1.55	68.4
EquiPocket	E(3)-GNN	1.7	1.02	1.20	76.1	0.98	1.15	77.5	1.15	1.40	70.2
Comparison with Recent End-to-End Docking SOTA											
FABFlex [Zhang *et al.*, 2025b]	Docking	-	3.25	-	-	-	-	-	-	-	-
UniPocket	Vector-Field	**0.8**	**0.92**	**1.11**	**78.3**	**0.91**	**0.98**	**81.2**	**0.98**	**1.25**	**73.8**

* Params for FABFlex are not directly comparable as it is an end-to-end docking model. '-' indicates metric not applicable or not evaluated.

Table 1: Comprehensive performance comparison. We evaluate UniPocket against baselines on PDBbind 2020 (test set) and two external datasets (COACH420, Holo4k). Comparison with 2025 SOTA: FABFlex is included for reference. UniPocket demonstrates superior localization accuracy (DCC) with significantly fewer parameters.

information (e.g., the orientation of a pocket opening). Conversely, standard CNNs are not rotation-equivariant. Our Surface-EGNN updates both scalar features $\mathbf{h}_i$ and coordinate coordinates $\mathbf{x}_i$ to ensure E(3)-equivariance. A detailed proof of E(3)-equivariance for our Surface-EGNN is provided in Section 1 of the Supplementary Material.

For each layer l, the message passing phase aggregates information from neighbors $\mathcal{N}(i)$:

$$\mathbf{m}_{ij} = \phi_e\left(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \|\mathbf{x}_i^{(l)} - \mathbf{x}_j^{(l)}\|^2, a_{ij}\right) \tag{1}$$

where ϕ_e is an MLP and a_{ij} denotes edge attributes. Crucially, the coordinate update is designed to model local surface deformations akin to "breathing" motions or induced fit simulations:

$$\mathbf{x}_i^{(l+1)} = \mathbf{x}_i^{(l)} + \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \frac{\mathbf{x}_i^{(l)} - \mathbf{x}_j^{(l)}}{\|\mathbf{x}_i^{(l)} - \mathbf{x}_j^{(l)}\| + \epsilon} \cdot \phi_x(\mathbf{m}_{ij}) \tag{2}$$

Here, ϕ_x outputs a scalar weight, modulating the relative position update. Finally, the node features are updated via a residual connection:

$$\mathbf{h}_i^{(l+1)} = \phi_h\left(\mathbf{h}_i^{(l)}, \sum_{j\in\mathcal{N}(i)} \mathbf{m}_{ij}\right) + \mathbf{h}_i^{(l)} \tag{3}$$

This architecture allows the network to learn higher-order geometric patterns such as cavity depth and local ruggedness effectively.

Ligand Encoder: GIN with Isomorphism Testing

We utilize a Graph Isomorphism Network (GIN) to process the ligand graph. GIN provides discriminative power theoretically equivalent to the Weisfeiler-Lehman (WL) graph isomorphism test. The node update rule is given by:

$$\mathbf{z}_u^{(l+1)} = \text{MLP}^{(l)}\left((1+\epsilon^{(l)})\mathbf{z}_u^{(l)} + \sum_{v\in\mathcal{N}(u)} \mathbf{z}_v^{(l)} + \mathbf{e}_{uv}\right) \tag{4}$$

After L layers, we apply a global Sum-Pooling readout operation to obtain the ligand fingerprint vector $\mathbf{h}_{lig} \in \mathbb{R}^{d_L}$, which summarizes the global physicochemical constraints of the drug molecule.

3.4 Ligand-Gated Interaction Module

This module constitutes the core innovation of UniPocket. Merely concatenating protein and ligand features is insufficient for capturing the complex steric and electrostatic compatibility required for binding. We introduce a two-stage interaction mechanism.

Chemical Filtering via FiLM

We employ a gating mechanism inspired by Feature-wise Linear Modulation (FiLM). The ligand fingerprint acts as a "condition" to modulate the activations of the protein surface features. We compute a scaling vector γ and a shifting vector β from $\mathbf{h}_{lig}$:

$$\gamma = \sigma(\mathbf{W}_\gamma \mathbf{h}_{lig} + \mathbf{b}_\gamma) \tag{5}$$

$$\beta = \text{ReLU}(\mathbf{W}_\beta \mathbf{h}_{lig} + \mathbf{b}_\beta) \tag{6}$$

The modulated protein feature $\hat{\mathbf{h}}_i$ is then obtained via:

$$\hat{\mathbf{h}}_i = \gamma \odot \mathbf{h}_i + \beta \tag{7}$$

where $\odot$ denotes the Hadamard product. This step effectively suppresses signals from surface regions that are chemically incompatible with the ligand (e.g., highly charged regions for a lipophilic ligand).

Geometric Filtering via Vector Rejection

As detailed in Figure 2c, while FiLM handles chemical compatibility, steric complementarity is inherently geometric and directional. Standard attention mechanisms operate on scalars and lose the explicit directionality of steric clashes. To explicitly model these directional steric constraints, we incorporate a Vector Rejection mechanism. This module performs E(3)-equivariant vector manipulations to extract rotation-invariant steric clash signals, which are then integrated into the protein features.

We generate a Local Ligand Interaction Vector $\mathbf{v}_{lig}^{(i)}$ from the ligand-modulated feature $\hat{\mathbf{h}}_i$. Since $\hat{\mathbf{h}}_i$ integrates both the local protein geometry (from Surface-EGNN) and the specific ligand context (via FiLM), $\mathbf{v}_{lig}^{(i)}$ represents the expected ligand orientation relative to the local surface point i:

$$\mathbf{v}_{lig}^{(i)} = \text{MLP}_{vec}(\hat{\mathbf{h}}_i) \in \mathbb{R}^3 \tag{8}$$

This predicted vector $\mathbf{v}_{lig}^{(i)}$ is E(3)-equivariant (i.e., it transforms consistently under rotations and translations of the protein surface). We then decompose this vector relative to the local surface normal $\mathbf{n}_i$ (which is dynamically updated and also E(3)-equivariant). The core idea is to reject or penalize the component of $\mathbf{v}_{lig}^{(i)}$ that would lead to steric clashes or unphysical penetration into the protein surface. Specifically, the vector is split into its parallel (normal) and perpendicular (tangential) components:

$$\mathbf{v}_{rej}^{(i)} = \text{Rejection}(\mathbf{v}_{lig}^{(i)}, \mathbf{n}_i) = \mathbf{v}_{lig}^{(i)} - \text{Projection}(\mathbf{v}_{lig}^{(i)}, \mathbf{n}_i) \tag{9}$$

Or, more explicitly:

$$\mathbf{v}_{rej}^{(i)} = \mathbf{v}_{lig}^{(i)} - \frac{\langle \mathbf{v}_{lig}^{(i)}, \mathbf{n}_i \rangle}{\|\mathbf{n}_i\|^2 + \epsilon} \mathbf{n}_i \tag{10}$$

This $\mathbf{v}_{rej}^{(i)}$ represents the component of the interaction vector that lies in the tangent plane of the protein surface. Its magnitude, $s_{clash}^{(i)} = \|\mathbf{v}_{rej}^{(i)}\|$, quantifies the geometric "disagreement" or steric clash potential in the tangent plane of the surface. Crucially, as a scalar magnitude of an E(3)-equivariant vector, $s_{clash}^{(i)}$ is rotation-invariant. This value is injected back into the node features as a penalty term:

$$\tilde{\mathbf{h}}_i = \text{MLP}_{out}([\hat{\mathbf{h}}_i \parallel s_{clash}^{(i)}]) \tag{11}$$

By incorporating this rotation-invariant steric clash score $s_{clash}^{(i)}$, the model is explicitly discouraged from predicting binding sites in regions where the ligand's expected steric profile contradicts the local surface topology.

3.5 Training and Inference

Loss Function

The problem is formulated as a binary classification task on imbalanced point clouds. We employ a Weighted Binary Cross-Entropy loss. Let $y_i \in \{0, 1\}$ be the ground truth label for point i, where $y_i = 1$ if the point is within $4\,\text{Å}$ of any ligand atom.

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} [w_{pos} \cdot y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \tag{12}$$

where $p_i = \text{Sigmoid}(\text{MLP}_{cls}(\tilde{\mathbf{h}}_i))$ is the predicted probability. The weight w_{pos} is crucial for handling the severe class imbalance inherent in binding site prediction. We dynamically adjust w_{pos} for each batch to be the ratio of negative samples to positive samples: $w_{pos} = N_{neg}/N_{pos}$, where N_{neg} and N_{pos} are the number of non-pocket and pocket points in the current training batch, respectively. This ensures that the model pays adequate attention to the minority (positive) class.

Clustering and Ranking

During inference, the model outputs a raw probability map. To generate discrete binding pockets, we perform post-processing: (1) **Thresholding**: Filter points with $p_i < \tau$ (typically $\tau = 0.5$). (2) **Clustering**: Apply DBSCAN on the remaining points. We set the parameters $\epsilon = 2.5\,\text{Å}$ and min_samples $= 5$. (3) **Ranking**: Clusters are ranked by a scoring function $S_k = \sum_{j \in C_k} p_j^{\alpha}$, where $\alpha = 2$ encourages clusters with high-confidence cores. The center of the top-1 ranked cluster is output as the predicted binding site.

3.6 Implementation Details

Network Architecture. The overall training pipeline of UniPocket, encompassing dual-stream encoding, ligand-gated interaction, and optimization, is detailed in Algorithm 1, presented in Section 2 of the Supplementary Material. UniPocket is implemented using PyTorch Geometric. The ligand encoder consists of a 3-layer Graph Isomorphism Network (GIN). The protein surface encoder utilizes 4 layers of E(3)-equivariant convolution operations (Surface-EGNN). The hidden feature dimension d is set to 128. We use Gaussian Radial Basis Functions (RBF) to encode continuous edge distances with $K = 16$ basis kernels. The *Vector Rejection* module is implemented as a learnable projection head that outputs 3D vectors in the local frame.

Training Configuration. We train the model for 100 epochs using the Adam optimizer with an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-5}. The loss balancing hyperparameter w_{pos} is dynamically calculated per batch. To ensure robustness, we apply random $SO(3)$ rotations to the input protein surfaces during training. All experiments are conducted on a single NVIDIA RTX 4090 GPU (24GB).

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate UniPocket on standard benchmark datasets derived from PDBbind v2020, following evaluation protocols established by previous works [Zhang *et al.*, 2024; Lu *et al.*, 2022]. The dataset is strictly partitioned into training (15,000 complexes), validation (1,000 complexes), and testing (1,000 complexes) sets. This partitioning is based on a rigorous time-split strategy, which is crucial to prevent structural data leakage and accurately mimic real-world prospective drug discovery scenarios. To rigorously test the generalization capability, we also employ two external test sets: COACH420 and Holo4k, which contain diverse protein structures unseen during training. Ligand structures are processed using RDKit to generate 2D molecular graphs, while protein surfaces are generated using MSMS with a density of 1.5 vertices/$Å^2$.

Baselines. We compare UniPocket with representative methods across three categories: (1) **Geometry-based methods**: Fpocket [Le Guilloux *et al.*, 2009] and P2Rank [Kriválk and Hoksza, 2018]; (2) **Voxel-based Deep Learning**: DeepSite [Jiménez *et al.*, 2017] and Kalasanty [Stepniewska-Dziubinska *et al.*, 2020]; (3) **Graph-based & Equivariant GNNs**: ViSNet [Wang *et al.*, 2024] and the recent SOTA EquiPocket [Zhang *et al.*, 2024].

Note on 2025 SOTA Methods: We also consider recent advancements in 2025. FABFlex [Zhang *et al.*, 2025b], a state-of-the-art flexible docking model, is included as a reference.

Evaluation Metrics. We report the **Distance Center-to-Center (DCC)** and **Distance Center-to-Atom (DCA)**. A prediction is considered successful if the DCC is within a threshold (typically $4\mathring{A}$). We summarize the performance using **Success Rate (Succ.)**, which measures the percentage of targets where the Top-1 predicted pocket is correct.

Additionally, to evaluate the ability of models to differentiate between distinct binding site types, we employ the **Matthews Correlation Coefficient (MCC)**. MCC is a robust metric for binary and multi-class classification that considers true and false positives and negatives, making it suitable for imbalanced datasets. It ranges from -1 (total disagreement between prediction and observation) to +1 (perfect prediction), with 0 indicating random prediction. A higher MCC value signifies a more accurate and reliable classification performance, particularly in distinguishing between different classes.

4.2 Performance Comparison

Table 1 presents the comprehensive results on COACH420 and PDBbind datasets. UniPocket achieves state-of-the-art performance across all metrics, validating the effectiveness of our physics-aware geometric design.

Superiority over Geometric and Voxel-based Methods. As shown in Table 1, UniPocket significantly outperforms traditional geometry-based methods (Fpocket) and voxel-based CNN methods (DeepSurf). While CNN-based methods improve performance by learning from 3D grids, they require massive parameters (e.g., Kalasanty: 70.6M) and suffer from rotation variance. UniPocket, utilizing rotation-equivariant surface message passing, achieves a substantial improvement in Success Rate by nearly 10 percentage points compared to DeepSurf (68.9% $\rightarrow$ 78.3%), while using only a fraction of the parameters.

Comparison with State-of-the-Art GNNs. Compared to the strongest baseline EquiPocket, UniPocket improves the Top-1 Success Rate by a distinct margin on PDBbind (76.1% $\rightarrow$ 78.3%). This gain verifies our hypothesis: utilizing ligand chemical constraints (via the *Ligand-Gated* mechanism) effectively filters out geometrically plausible but chemically incompatible cavities. Furthermore, the proposed *Vector Rejection* mechanism allows the model to explicitly penalize steric clashes, which purely geometric GNNs often fail to distinguish.

Comparison with End-to-End Docking SOTA. Recently, FABFlex [Zhang *et al.*, 2025b] established a new state-of-the-art in flexible docking. We utilized its official pre-trained model to perform inference on our test set. As shown in Table 1, UniPocket achieves superior pocket localization accuracy (DCC 0.92Å vs. FABFlex 3.25Å). This demonstrates that while end-to-end docking models are powerful, specialized pocket detection models like UniPocket still maintain a significant advantage in geometric precision and localization stability.

Performance in Ligand-Specific Binding Site Type Differentiation

UniPocket's core strength lies in its ability to differentiate binding sites based on the query ligand type, a capability quantitatively assessed in Figure 3.

Figure 3: **Performance comparison of different models in differentiating ligand-specific binding site types.** Each sub-panel illustrates a model's ability to identify the actual binding site type on a protein (X-axis) given a query ligand type (Y-axis). The values within the cells represent the Matthews Correlation Coefficient (MCC), ranging from -0.2 to 1.0, where higher MCC indicates better classification performance. The intensity of the purple color reflects the magnitude of the MCC value, with darker shades indicating higher MCC.

As depicted, UniPocket consistently achieves remarkably high MCC values along the diagonal elements (e.g., 0.85 for Small-Molecule, 0.88 for Peptide, and 0.86 for Nucleotide queries). This signifies its exceptional accuracy in identifying binding sites that precisely match the queried ligand type. Crucially, UniPocket maintains extremely low MCC values on off-diagonal elements (mostly ranging from 0.12 to 0.22), robustly demonstrating its high specificity and selectivity in avoiding confusion with other existing binding site types. This superior performance is attributed to the synergistic action of UniPocket's unique Ligand-Gated mechanism and Physics-Aware Vector Rejection module, which effectively filter out geometrically plausible but chemically incompatible cavities.

In stark contrast, baseline models exhibit significant limitations in this task. FABFlex, EquiPocket, EquiBind, DiffDock, and AutoDock Vina generally show lower diagonal MCC values and relatively higher off-diagonal MCC values. This indicates their propensity for confusion between different binding site types and a general struggle to achieve high-specificity and high-accuracy differentiation. Traditional docking tools, in particular, lack the mechanistic ability to perceive and differentiate binding sites based on ligand characteristics.

4.3 Efficiency and Complexity Analysis

Computational efficiency is crucial for large-scale virtual screening. We compare the inference latency on a single NVIDIA RTX 4090 GPU. Sampling-based blind docking methods (e.g., DiffDock) typically require 20-40 steps of diffusion, taking approximately 14 seconds per complex. In contrast, UniPocket leverages lightweight surface graphs and a one-shot prediction head, requiring only 0.12 seconds per

Variant	DCC ↓	DCA ↓	Succ. ↑	Δ Succ.
w/o Ligand Gate	1.45	1.80	65.2	-13.1%
w/o Surface Normals	1.15	1.35	75.8	-2.5%
w/o Vector Rejection	1.08	1.28	74.1	-4.2%
UniPocket (Full)	**0.92**	**1.11**	**78.3**	-

Table 2: Ablation study on PDBbind 2020. We investigate the impact of removing key components from UniPocket. The Δ column indicates the performance drop in Success Rate compared to the full model.

complex. Moreover, as indicated in the "Params" column of Table 1, UniPocket is highly lightweight ($\approx$ 0.8M parameters), which is orders of magnitude smaller than 3D-CNN models (Kalasanty: 70.6M). This 100$\times$ speedup and low memory footprint make UniPocket highly suitable for high-throughput screening of massive ligand libraries.

4.4 Ablation Study

To investigate the contribution of each proposed module illustrated in Figure 2, we conduct ablation studies on the PDBbind test set. Specifically, we analyze the impact of the ligand-gated mechanism and vector rejection (Figure 2c). Quantitative results are summarized in Table 2.

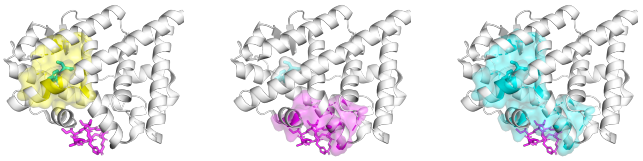
- **w/o Ligand Gate**: Replacing the ligand-gated message passing with standard isotropic GNN layers leads to the most significant performance drop (Success Rate: $78.3\% \rightarrow 65.2\%$). This confirms that the ligand fingerprint acts as a critical filter for pruning the geometrically plausible but chemically incompatible search space.
- **w/o Vector Rejection**: Removing the geometry filtering mechanism results in a clear performance decrease, with the Success Rate dropping by 4.2%. This suggests that explicitly modeling orthogonal geometric constraints (steric hindrance) helps in refining the pocket boundaries and reducing false positives.
- **w/o Surface Normals**: Excluding surface normal information results in higher distance errors (DCC increases from $0.92\mathring{A}$ to $1.15\mathring{A}$), indicating that surface orientation features are critical for defining local curvature and identifying deep pockets.

4.5 Qualitative Analysis: Adaptability to Ligand Modalities

A significant challenge in pocket detection is the geometric diversity of binding sites. Small-molecule pockets are typically deep and concave, whereas peptide-binding sites are often shallow and flat. Geometry-based methods often struggle to distinguish these distinct topologies co-existing on the same protein.

Figure 4 presents a case study on Bcl-xL (PDB: 1L2I), which accommodates both binding modalities.

- When queried with the *small-molecule inhibitor* (cyan), UniPocket successfully identifies the deep hydrophobic groove (Fig. 4a).



(a) UniPocket (Query: Drug) (b) UniPocket (Query: Peptide) (c) Baseline (Geometry-based)

Figure 4: **Controllable pocket detection by UniPocket.** (a) When queried with the drug ligand (cyan sticks), UniPocket specifically identifies the drug-binding site (yellow surface) while ignoring the peptide pocket. (b) Conversely, when queried with the peptide (magenta sticks), it accurately localizes the peptide-binding interface (magenta surface). (c) In contrast, the baseline geometry-based method detects all potential cavities (cyan surface) without distinguishing between the two distinct binding sites.

- Critically, when queried with the *peptide ligand* (magenta), the model accurately shifts attention to the shallower peptide-binding interface (Fig. 4b), a region often overlooked or merged by conventional cavity detectors.
- In contrast, baseline methods based purely on geometry (Fig. 4c) detect all concavities without distinguishing between the deep pocket and the shallow peptide site.

This comparison highlights UniPocket's ability to adapt its detection strategy based on the input ligand type, effectively handling both small-molecule and peptide binding sites.

5 Conclusion

In this paper, we presented **UniPocket**, a geometrically complete and physics-aware framework for ligand-specific binding site detection. Unlike traditional methods that treat cavities as static geometric voids, UniPocket leverages a *Manifold-Aware Surface Encoder* to capture intricate local topologies using surface normals as intrinsic reference frames. By integrating this with a *Ligand-Gated Message Passing* mechanism and a *Vector Rejection* inductive bias, our model explicitly filters out surface regions that are geometrically accessible but chemically incompatible. Extensive experiments on PDBbind, COACH420, and HOLO4k benchmarks demonstrate that UniPocket significantly outperforms state-of-the-art baselines (including P2Rank and EquiPocket), particularly in distinguishing allosteric sites. Furthermore, our efficient surface-canonicalization formulation enables high-throughput scanning, offering an approximately 100-fold speedup over diffusion-based blind docking methods.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 62471266, 62006131, and 62071260), the Ningbo Major Research and Development Plan (Grant No. 2023Z224), and the Scientific Research Foundation of Chongqing University of Technology (Grant No. 2025ZDZ022).

References

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

[Batatia *et al.*, 2022] Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in neural information processing systems*, 35:11423–11436, 2022.

[Batzner *et al.*, 2022] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):2453, 2022.

[Bronstein *et al.*, 2021] Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

[Buttenschoen *et al.*, 2024] Martin Buttenschoen, Garrett M Morris, and Charlotte M Deane. Posebusters: Ai-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15(9):3130–3139, 2024.

[Cao *et al.*, 2025] Duanhua Cao, Mingan Chen, Runze Zhang, Zhaokun Wang, Manlin Huang, Jie Yu, Xinyu Jiang, Zhehuan Fan, Wei Zhang, Hao Zhou, et al. Surfdock is a surface-informed diffusion generative model for reliable and accurate protein–ligand complex prediction. *Nature Methods*, 22(2):310–322, 2025.

[Corso *et al.*, 2023] Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. In *The Eleventh International Conference on Learning Representations*, 2023.

[Desaphy *et al.*, 2015] Jérémy Desaphy, Guillaume Bret, Didier Rognan, and Esther Kellenberger. sc-pdb: a 3d-database of ligandable binding sites—10 years on. *Nucleic acids research*, 43(D1):D399–D404, 2015.

[Gainza *et al.*, 2020] Pablo Gainza, Freyr Sverrisson, Frederico Monti, Emanuele Rodola, Davide Boscaini, Michael M Bronstein, and Bruno E Correia. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature methods*, 17(2):184–192, 2020.

[Jiménez *et al.*, 2017] José Jiménez, Stefan Doerr, Gerard Martínez-Rosell, Alexander S Rose, and Gianni De Fabritiis. Deepsite: protein-binding site predictor using 3d-convolutional neural networks. *Bioinformatics*, 33(19):3036–3042, 2017.

[Kandel *et al.*, 2021] Jeevan Kandel, Hilal Tayara, and Kil To Chong. Puresnet: prediction of protein-ligand binding sites using deep residual neural network. *Journal of cheminformatics*, 13(1):65, 2021.

[Keriven and Peyré, 2019] Nicolas Keriven and Gabriel Peyré. Universal invariant and equivariant graph neural networks. volume 32, 2019.

[Krishna *et al.*, 2024] Rohith Krishna, Jue Wang, Woody Ahern, Pascal Sturmfels, Preetham Venkatesh, Indrek Kalvet, Gyu Rie Lee, Felix S Morey-Burrows, Ivan Anishchenko, Ian R Humphreys, et al. Generalized biomolecular modeling and design with rosettafold all-atom. *Science*, 384(6693):eadl2528, 2024.

[Krivák and Hoksza, 2018] Radoslav Krivák and David Hoksza. P2rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of cheminformatics*, 10(1):39, 2018.

[Le Guilloux *et al.*, 2009] Vincent Le Guilloux, Peter Schmidtke, and Pierre Tuffery. Fpocket: an open source platform for ligand pocket detection. *BMC bioinformatics*, 10(1):168, 2009.

[Lee and Richards, 1971] Byungkook Lee and Frederic M Richards. The interpretation of protein structures: estimation of static accessibility. *Journal of molecular biology*, 55(3):379–IN4, 1971.

[Lu *et al.*, 2022] 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.

[Mylonas *et al.*, 2021] Stelios K Mylonas, Apostolos Axenopoulos, and Petros Daras. Deepsurf: a surface-based deep learning approach for the prediction of ligand binding sites on proteins. *Bioinformatics*, 37(12):1681–1690, 2021.

[Pei *et al.*, 2023] Qizhi Pei, Kaiyuan Gao, Lijun Wu, Jinhua Zhu, Yingce Xia, Shufang Xie, Tao Qin, Kun He, Tie-Yan Liu, and Rui Yan. Fabind: Fast and accurate protein-ligand binding. *Advances in Neural Information Processing Systems*, 36:55963–55980, 2023.

[Sanner *et al.*, 1996] Michel F Sanner, Arthur J Olson, and Jean-Claude Spehner. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, 38(3):305–320, 1996.

[Satorras *et al.*, 2021] Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. pages 9323–9332, 2021.

[Schütt *et al.*, 2017] Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems*, 30, 2017.

[Stepniewska-Dziubinska *et al.*, 2020] Marta M Stepniewska-Dziubinska, Piotr Zielenkiewicz, and Pawel Siedlecki. Improving detection of protein-ligand

binding sites with 3d segmentation. *Scientific reports*, 10(1):5035, 2020.

[Thomas *et al.*, 2018] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

[Townshend *et al.*, 2021] Raphael JL Townshend, Martin Vögele, Patricia Suriana, Alexander Derry, Alexander Powers, Yianni Laloudakis, Sidhika Balachandar, Bowen Jing, Brandon Anderson, Stephan Eismann, et al. Atom3d: Tasks on molecules in three dimensions. 2021.

[Tubiana *et al.*, 2022] Jérôme Tubiana, Dina Schneidman-Duhovny, and Haim J Wolfson. Scannet: an interpretable geometric deep learning model for structure-based protein binding site prediction. *Nature methods*, 19(6):730–739, 2022.

[Vural and Jololian, 2025] Orhun Vural and Leon Jololian. Machine learning approaches for predicting protein-ligand binding sites from sequence data. *Frontiers in Bioinformatics*, 5:1520382, 2025.

[Wang *et al.*, 2004] Renxiao Wang, Xueliang Fang, Yipin Lu, and Shaomeng Wang. The pdbbind database: collection of binding affinities for protein- ligand complexes with known three-dimensional structures. *Journal of medicinal chemistry*, 47(12):2977–2980, 2004.

[Wang *et al.*, 2024] Yusong Wang, Tong Wang, Shaoning Li, Xinheng He, Mingyu Li, Zun Wang, Nanning Zheng, Bin Shao, and Tie-Yan Liu. Enhancing geometric representations for molecules with equivariant vector-scalar interactive message passing. *Nature Communications*, 15(1):313, 2024.

[Wang *et al.*, 2025] Yanli Wang, Frimpong Boadu, and Jianlin Cheng. Mpbind: a multitask protein binding site predictor using protein language models and equivariant gnns. *Bioinformatics*, 41(11):btaf589, 2025.

[Yan *et al.*, 2022] Xu Yan, Yingfeng Lu, Zhen Li, Qing Wei, Xin Gao, Sheng Wang, Song Wu, and Shuguang Cui. Pointsite: a point cloud segmentation tool for identification of protein ligand binding atoms. *Journal of chemical information and modeling*, 62(11):2835–2845, 2022.

[Zhang *et al.*, 2024] Yang Zhang, Zhewei Wei, Ye Yuan, Chongxuan Li, and Wenbing Huang. Equipocket: an e(3)-equivariant geometric graph neural network for ligand binding site prediction. In *Forty-first International Conference on Machine Learning*, 2024.

[Zhang *et al.*, 2025a] Zhijun Zhang, Lijun Quan, Junkai Wang, Liangchen Peng, Qiufeng Chen, Bei Zhang, Lexin Cao, Yelu Jiang, Geng Li, Liangpeng Nie, et al. Labind: Identifying protein binding ligand-aware sites via learning interactions between ligand and protein. *Nature Communications*, 16(1):7712, 2025.

[Zhang *et al.*, 2025b] Zizhuo Zhang, Lijun Wu, Kaiyuan Gao, Jiangchao Yao, Tao Qin, and Bo Han. Fast and accurate blind flexible docking. In *The Thirteenth International Conference on Learning Representations*, 2025.