INTERACTION-ALIGNED SURFACE REPRESENTATIONS FOR GENERALIZABLE BLIND DOCKING

Anonymous authors
Paper under double-blind review

ABSTRACT

Recent advances in machine learning have substantially improved protein–ligand docking, yet blind docking still generalizes poorly to unfamiliar receptors and binding environments. We argue that part of this difficulty arises from a mismatch between structural representation and interaction-level invariance: global structural novelty need not imply interaction novelty, as substantial receptor variation can leave the geometry and chemistry of relevant binding interfaces largely unchanged. Under limited supervision, however, Euclidean geometric encoders must infer such invariances implicitly and may entangle transferable interaction patterns with receptor-specific structural variation. We introduce SADock, a geometric diffusion framework that addresses this mismatch by explicitly representing both proteins and ligands through molecular surfaces, the natural boundaries on which molecular recognition occurs. Surface geometry and physicochemical fields expose interaction-facing patterns more directly, while a hierarchical heterogeneous graph integrates surface representations with residue- and atom-level context across multiple spatial scales for receptor-wide search and fine-grained pose modeling. We further establish conditions under which boundary-preserving structural changes retain binding-basin preferences, providing a theoretical characterization of why surface-aligned representations can remain transferable under structural variation. Across multiple benchmarks, SADock achieves higher docking success rates and lower pose RMSD. On challenging DockGen, our method improves Top-1/Top-5 docking success rates from 22.8%/29.6% for the strongest evaluated baseline to 27.0%/38.6%, despite using less data augmentation and without inference tuning tricks. Further analysis reveals substantial gains in pocket localization, a key bottleneck in generalizable blind docking, supporting surface-aware interaction modeling as a principled inductive bias under structural distribution shift.

1 INTRODUCTION

Accurate protein–ligand docking provides a structural basis for understanding molecular recognition and supports structure-based drug discovery (Lyu et al., 2019). Deep learning (DL) has substantially advanced biomolecular structure modeling, from protein structure prediction (Jumper et al., 2021; Lin et al., 2023) to biomolecular complex modeling (Abramson et al., 2024; team et al., 2024; Corso et al., 2023). Nevertheless, learning-based *blind docking* methods, including regression-based approaches (Lu et al., 2022; Pei et al., 2023) and diffusion-based approaches (Corso et al., 2023; 2024), still struggle to generalize to unfamiliar receptors and binding environments, where available training data sparsely cover the structural and interaction space encountered at realistic applications. Recent PoseBench evaluations further show that co-folding models trained on large-scale protein and multiple-sequence-alignment data improve overall performance but remain challenged by out-of-distribution (OOD) protein–ligand binding conformations (Morehead et al., 2026). These findings reveal a persistent gap between recent progress in biomolecular structure prediction and reliable interaction modeling beyond in-distribution (ID) binding environments, motivating more generalizable blind docking for large-scale molecular screening and design.

Beyond limited training coverage, we argue that part of this generalization gap arises from a mismatch between structural representations and interaction-level invariance. In particular, *global structural novelty does not necessarily imply interaction novelty*: substantial variation in receptor struc-

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.

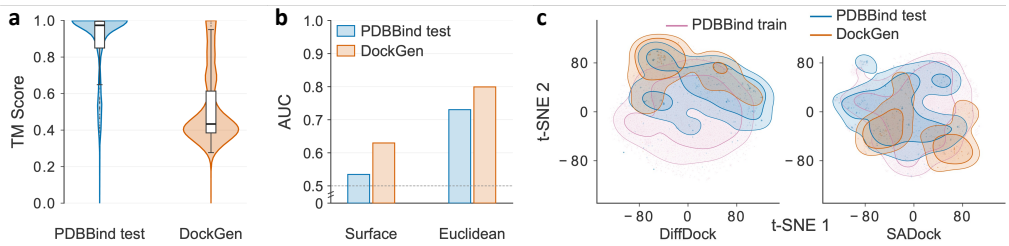

Figure 1: Surface representations exhibit reduced cross-dataset shift. **(a)** Global protein structures show a pronounced distribution shift from PDBBind to DockGen, measured by TM-score relative to the PDBBind training set. **(b)** A simple domain classifier distinguishes training and test samples more easily using Euclidean features than surface features. **(c)** Learned surface features exhibit substantially greater cross-dataset overlap than Euclidean features. Both features are extracted from the corresponding docking score models at diffusion time $t = 1$.

ture can coexist with similar local binding-site geometry and physicochemical properties (Konc & Janežič, 2010; Gainza et al., 2020). Such recurring interaction patterns create an opportunity for transfer even across structurally distinct receptors. However, most existing docking models operate primarily on Euclidean geometric representations of atoms or residues. Although highly expressive, these representations expose the model to substantial receptor-specific structural variation, leaving interaction-relevant regularities to be identified implicitly from limited supervision. Transferable interface patterns may therefore become entangled with variations that are less relevant to ligand recognition, making receptor-wide search particularly difficult in unfamiliar binding environments. This issue is especially consequential for *pocket localization*, a critical stage of blind docking whose failure can prevent subsequent pose refinement from recovering the correct binding mode (Pei et al., 2023; Corso et al., 2024). These observations motivate representations that organize structural information more directly around transferable interaction environments.

This motivates us to consider molecular surfaces as a representation more closely aligned with molecular interactions. Protein–ligand recognition occurs at molecular interfaces, where local shape and physicochemical complementarity jointly determine interaction compatibility; molecular surfaces therefore provide a natural abstraction of the geometry and chemical fields directly exposed to binding partners (McCammon et al., 2003; Konc & Janežič, 2010). By emphasizing these interaction-facing properties, surface representations may be less sensitive to receptor-specific structural variation that is weakly reflected at the binding boundary. We find preliminary evidence for this hypothesis on PDBBind and DockGen. Despite pronounced receptor-level distribution shift (Figure 1a), a simple domain classifier distinguishes training and test samples substantially more easily from Euclidean features than from surface features (Lopez-Paz & Oquab, 2016) (Figure 1b). Consistently, representations learned from the docking task show markedly greater cross-dataset overlap in the surface feature space than in the Euclidean feature space (Figure 1c). These observations align with prior molecular-surface studies on transferable pocket recognition (Gainza et al., 2020; 2023) and with SurfDock, which demonstrates the utility of surface information for pocket-conditioned pose generation (Cao et al., 2025). However, whether surface information can directly facilitate receptor-wide pocket discovery and pose generation in *blind docking* remains largely unexplored.

Motivated by these observations, we revisit blind docking from a surface-aware perspective and propose **SADock**, a geometric diffusion framework built around a dual-branch score network, SANet, that complements Euclidean geometry with explicit molecular-surface representations of both proteins and ligands. SANet introduces three key designs (Figure 2b): (1) geodesic message passing to encode local geometric and physicochemical patterns while respecting surface neighborhoods; (2) hierarchical surface message passing to enable efficient receptor-wide interaction without discarding fine-grained boundary information; and (3) heterogeneous message passing to couple surface representations with residue- and atom-level context, together with time-conditioned feature-wise linear modulation that adaptively balances surface and Euclidean information throughout denoising. We further derive sufficient conditions under which boundary-preserving receptor variations retain binding-basin preferences, providing a theoretical characterization of the interaction-level transferability captured by surface representations. Across multiple benchmarks, SADock achieves higher docking success rates and lower pose RMSD. On the challenging DockGen benchmark, it improves

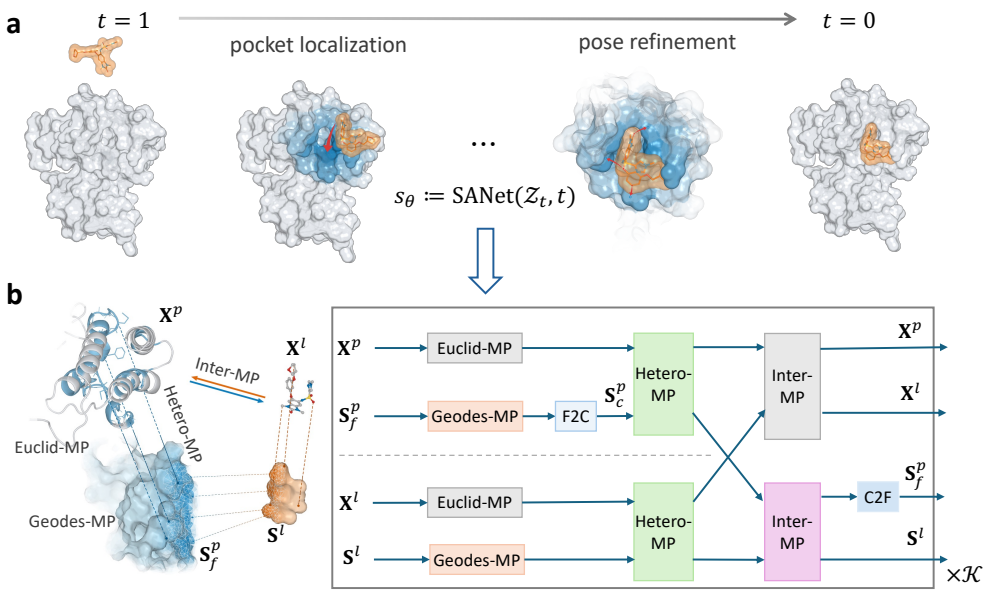

Figure 2: Overview of SADock. **(a)** SADock performs blind docking with a DiffDock-style diffusion framework, progressing from receptor-wide pocket localization to fine-grained pose refinement, with SANet as the score model. **(b)** SANet jointly models Euclidean and molecular-surface representations of both protein and ligand. Within each block, `Euclid-MP` encodes residue- and atom-level geometry, `Geodes-MP` captures local surface patterns, `F2C`/`C2F` enable efficient hierarchical surface propagation, `Hetero-MP` couples surface and Euclidean features, and `Inter-MP` exchanges protein–ligand information. The resulting features predict translation, rotation, and torsional scores; gray modules denote Euclidean components inherited from DiffDock.

docking success from 22.8% to 29.6% over the strongest evaluated baseline despite using less data augmentation and inference tuning tricks. Further analysis shows substantial gains in pocket localization, a key bottleneck in blind-docking generalization, supporting surface-aware interaction modeling as a principled inductive bias under structural distribution shift.

2 Methodology

2.1 Problem Setup

A target protein with L_p residues is represented as $\mathcal{P} = (\mathbf{a}^p, \mathbf{X}^p, \mathcal{C}^p)$, where $\mathbf{a}^p$ is its amino-acid sequence, $\mathbf{X}^p \in \mathbb{R}^{N_p \times 3}$ contains atomic coordinates, and $\mathcal{C}^p$ denotes chemical attributes; representative backbone coordinates $\mathbf{R}^p = \{\mathbf{r}_i^p\}_{i=1}^{L_p}$ are used for residue-level message passing. A ligand with N atoms is represented as $\mathcal{L} = (\mathcal{A}, \mathcal{E})$, where $\mathcal{A}$ and $\mathcal{E}$ denote atomic attributes and bond topology, respectively. Given $(\mathcal{P}, \mathcal{L})$, blind docking aims to sample ligand coordinates $\mathbf{R}^l \sim p_\theta(\cdot \mid \mathcal{P}, \mathcal{L})$ without access to a predefined binding pocket or bound ligand pose, while keeping the receptor fixed. As illustrated in Figure 2, we follow DiffDock's pose-diffusion formulation and training objective, but augment both interaction partners with molecular-surface representations and replace the original score network with our surface-aware SANet.

Molecular-surface representations. Alongside residue- and atom-level geometry, we represent both proteins and ligands as surface point clouds (Gainza et al., 2020; Sverrisson et al., 2021):

$$\mathcal{S}_\ell^u = \text{Surf}_\ell(\mathbf{X}^u, \mathcal{C}^u) = \{(\mathbf{s}_k^u, \mathbf{n}_k^u, \mathbf{f}_k^u)\}_{k=1}^{K_u}, \qquad u \in \{p, l\}, \tag{1}$$

where $\mathbf{s}_k^u \in \mathbb{R}^3$, $\mathbf{n}_k^u \in \mathbb{S}^2$, and $\mathbf{f}_k^u$ denote the surface position, outward unit normal, and geometric and physicochemical features, respectively; ℓ controls the surface resolution. For the ligand, $\mathbf{X}^l = \mathbf{R}^l$ and $\mathcal{C}^l$ is derived from its molecular graph and chemical attributes. The protein surface spans the entire receptor, consistent with the blind-docking setting. During denoising, the receptor surface $\mathcal{S}_\ell^p$

remains fixed, while the ligand surface $\mathcal{S}^l_{\ell,t}$ follows the current noisy ligand conformation. We omit ℓ hereafter for brevity. These surface representations complement Euclidean geometry by explicitly exposing the molecular boundary and its physicochemical context.

2.2 Diffusion-Based Pose Generation

Forward diffusion. Score-based diffusion models generate samples by learning to reverse a prescribed noising process (Song et al., 2020). Following DiffDock (Corso et al., 2023), we fix the ligand's bond lengths and bond angles and parameterize its pose by translation, rotation, and m rotatable-bond torsions:

$$q = (\mathbf{u}, Q, \boldsymbol{\tau}) \in \mathcal{Q}_{\mathcal{L}} := \mathbb{R}^3 \times \mathrm{SO}(3) \times \mathbb{T}^m. \tag{2}$$

A pose q determines the ligand coordinates as $\mathbf{R}^l = \mathcal{T}(q; \mathbf{R}^l_{\mathrm{ref}})$, where $\mathbf{R}^l_{\mathrm{ref}}$ is a reference conformer and $\mathcal{T}$ applies the corresponding rigid-body and torsional transformations. At inference, the reference conformer is generated independently of the unknown bound pose. Independent Brownian perturbations are applied to the three factors of $\mathcal{Q}_{\mathcal{L}}$ with component-specific noise schedules. Writing $\mathcal{I} = \{\mathrm{tr}, \mathrm{rot}, \mathrm{tor}\}$, the forward transition kernel factorizes as

$$K_t(q_t \mid q_0) = \prod_{a \in \mathcal{I}} K^a_t(q^a_t \mid q^a_0), \tag{3}$$

while the receptor remains fixed throughout the diffusion process.

Surface-aware score estimation. At diffusion time t, SANet takes the current molecular geometry together with the protein and ligand surface representations, summarized as $\mathcal{Z}_t = (\mathcal{P}, \mathcal{L}, \mathbf{R}^l_t, \mathcal{S}^p, \mathcal{S}^l_t)$. It predicts the conditional score on each factor of the pose manifold:

$$s^a_\theta(\mathcal{Z}_t, t) \approx \nabla_{q^a_t} \log p_t(q_t \mid \mathcal{P}, \mathcal{L}), \qquad a \in \mathcal{I}. \tag{4}$$

The gradients are defined on the corresponding manifold factors. GSDock retains DiffDock's conformer preparation, noise schedules, denoising score-matching objective, and reverse-diffusion sampler, while replacing its score network with the surface-aware SANet. Additional diffusion details are provided in Appendix A.1.

2.3 Surface-Aware Denoising with SANet

SANet jointly models Euclidean and surface representations of both interaction partners to estimate the pose score (Figure 2b). Each block combines intramolecular encoding, cross-representation fusion, and protein–ligand interaction. Surface features are first encoded at fine resolution to preserve local geometric and physicochemical patterns, and are subsequently pooled for efficient long-range interaction modeling.

Euclidean message passing. We retain DiffDock's tensor-product Euclidean message passing (`Euclid-MP`) for protein residue–residue and ligand atom–atom interactions (Corso et al., 2023; Thomas et al., 2018). For molecule $u \in \{p, l\}$, neighborhoods are defined by Euclidean distance, with radial edge features $\mathbf{e}^{E,u}_{ij} = \mathrm{RBF}(\|\mathbf{r}^u_j - \mathbf{r}^u_i\|_2)$; ligand covalent edges additionally include bond attributes. Relative directions are expanded with spherical harmonics and combined with node features through tensor products to produce rotation-covariant representations. The intermolecular module (`Inter-MP`) uses the same equivariant construction to exchange information between protein residues and ligand atoms.

Geodesic message passing. We apply geodesic message passing (`Geodes-MP`) on both molecular surfaces to encode local geometric and physicochemical patterns. For each message-passing edge (i, j), we use both the Euclidean distance $d_{E,ij} = \|\mathbf{s}_j - \mathbf{s}_i\|_2$ and a discrete geodesic distance $d_{G,ij}$ computed from shortest paths on the original NanoShaper triangulation. The edge representation combines radial encodings of $d_{E,ij}$ and $d_{G,ij}$ with normal-based geometric descriptors, while directional information is encoded through spherical harmonics. Surface features are then updated by aggregating equivariant messages from their surface neighbors. By explicitly contrasting ambient and surface distances, `Geodes-MP` distinguishes points that are spatially close but separated along

the molecular surface, preserving interaction-relevant surface topology before hierarchical pooling. Detailed graph construction and message definitions are provided in Appendix A.3.

Hierarchical message passing. To enable efficient receptor-wide surface communication, we introduce a hierarchical surface pathway. After fine-resolution `Geodes-MP`, `F2C` pools receptor surface features into coarse representatives, on which subsequent heterogeneous and intermolecular message passing is performed. The updated coarse context is then propagated back to the fine surface through `C2F` with a skip connection from the pre-pooling features. This design reduces interaction cost while preserving fine-grained surface information.

Heterogeneous message passing and time modulation. After receptor pooling, heterogeneous message passing (`Hetero-MP`) couples protein residues with coarse receptor surface points and ligand atoms with ligand surface points, allowing surface features to acquire task-specific molecular context while injecting interaction-boundary information into the Euclidean branch. For molecule $u \in \{p, l\}$, let $\check{\mathbf{X}}_i^u$ denote the Euclidean feature of residue or atom i after `Hetero-MP`, and let $\mathbf{c}_i^u$ denote the aggregated surface context received by this node. We use surface-conditioned FiLM (Perez et al., 2018) to modulate $\check{\mathbf{X}}_i^u$:

$$\begin{aligned}(\boldsymbol{\gamma}_i^u(t), \boldsymbol{\beta}_i^u(t)) &= \phi_{\mathrm{FiLM}}([\mathbf{c}_i^u, \mathbf{e}_t]),\\ \widetilde{\mathbf{X}}_i^u &= (\mathbf{1} + \boldsymbol{\gamma}_i^u(t)) \odot \check{\mathbf{X}}_i^u + \boldsymbol{\beta}_i^u(t),\end{aligned} \tag{5}$$

where $\mathbf{e}_t$ is the diffusion-time embedding, ϕ_{FiLM} predicts channel-wise scaling and shift parameters, and $\odot$ denotes element-wise multiplication. The modulation is applied in an equivariance-preserving manner, allowing the contribution of surface context to adapt throughout denoising. Finally, `Inter-MP` exchanges protein–ligand information within each representation branch, through residue–atom interactions in the Euclidean branch and receptor–ligand surface interactions in the surface branch. This adaptive coupling supports complementary use of surface context for coarse localization and Euclidean detail for fine-grained pose refinement. Detailed operator definitions and equivariance constraints are provided in Appendix A.

2.4 Deciphering the Transferability of Surface-Based Representations

Our hypothesis is that molecular surfaces facilitate transfer because structurally distinct receptors can present similar interaction boundaries to the same ligand. We formalize this idea at two levels: whether boundary similarity preserves binding preferences, and whether SANet responds stably to the same receptor variations.

From boundary similarity to pocket preference. For blind docking, transferability is naturally associated with preserving the relative preference among competing binding regions. Consider aligned receptors X and X' interacting with a fixed ligand $\mathcal{L}$, and let $S = \mathcal{S}_\ell^p(X)$ and $S' = \mathcal{S}_\ell^p(X')$, with surface discrepancy $\delta_S = d_S(S, S')$ measuring differences in boundary geometry and physicochemical attributes. For a microscopic configuration ξ, let $U_X(\xi)$ denote the effective protein–ligand interaction energy under receptor X. The corresponding free energy of a coarse pose basin Ω_z is

$$G_X(z) = -\beta^{-1} \log \int_{\Omega_z} e^{-\beta U_X(\xi)}\, d\mu(\xi), \tag{6}$$

where β is inverse temperature. Consider receptor variations whose interaction perturbation satisfies

$$\inf_{c\in\mathbb{R}} \sup_{\xi\in\Omega} |U_{X'}(\xi) - U_X(\xi) - c| \le K_\ell \delta_S + \varepsilon_\ell =: b_\ell, \tag{7}$$

where $\Omega = \bigcup_{z\in\mathcal{Z}} \Omega_z$ covers the competing pose basins. The constant c removes pose-independent energy shifts. The residual ε_ℓ captures interaction changes not resolved by surface representations.

Proposition 1 (Boundary stability implies basin stability)**.** Let $\eta_* = \mathrm{osc}_z[G_{X'}(z) - G_X(z)]$. Under Eq. 7, $\eta_* \le 2b_\ell$. If z_* is the preferred basin of X with margin $m_X = \min_{z\neq z_*}[G_X(z) - G_X(z_*)]$, then $2b_\ell < m_X$ guarantees that z_* remains preferred for X'.

Appendix B.2 provides proofs. This result captures the form of transferability most relevant to blind docking: global receptor structure may change substantially while the preferred pocket remains stable if the ligand-facing interaction environment changes less than the free-energy margin

between competing regions. Physically, this regime is expected when pocket geometry, boundary physicochemical fields and steric accessibility are approximately preserved. For example, boundary electrostatics can remain stable despite internal charge rearrangements, while direct short-range effects of buried structural changes decay with their distance from ligand-accessible configurations. Shared binding-site patterns across distinct folds and molecular-surface fingerprints provide empirical support for such interaction-level recurrence (Konc & Janežič, 2010; Gainza et al., 2020). The stronger distributional result and detailed physical conditions are provided in Appendices B.1-B.3.

Model-level stability of SANet. For surface-based transferability to benefit learning, SANet should respond consistently to boundary-preserving receptor variations. Fix a ligand query $v = (\mathcal{L}, q, t)$ and denote one SANet score head by $F_\theta(v; E, S)$, where E and S are the receptor's Euclidean and surface representations. If F_θ is locally $L_{S,\theta}$-Lipschitz *w.r.t.*its surface input, then

$$\|F_\theta(v; E', S') - F_\theta(v; E, S)\| \leq L_{S,\theta}\delta_S + \epsilon_{E,\theta}, \tag{8}$$

where $\epsilon_{E,\theta} = \|F_\theta(v; E', S') - F_\theta(v; E, S')\|$ measures residual sensitivity to Euclidean structural variation after fixing the surface representation. This decomposition highlights the complementary roles of SANet's two branches: surface features capture interaction-aligned variation, while Euclidean features retain fine-grained structural detail. Accordingly, similar interaction boundaries can yield stable predictions across globally distinct receptors, motivating the time-conditioned FiLM module (Eq. 5) to emphasize surface context for pocket localization while preserving Euclidean detail for pose refinement. Detailed proofs and clarifications are given in Appendix B.4.

3 Experiments

3.1 Experimental Setup

Benchmarks. We primarily evaluate on the PDBBind 2019 test set (363 complexes) and DockGen-full (189 complexes). PDBBind follows the standard time split used in prior docking studies, while DockGen evaluates transfer to held-out ligand-binding domains (Corso et al., 2023; 2024). We additionally report results on PoseBusters (428 complexes).

Implementation details. We train two configurations, SADock and SADock-L, to match the model scale and training data of DiffDock and DiffDock-L, respectively. Both models use six SANet blocks, with approximately 20M and 30M parameters. SADock is trained on PDBBind, while SADock-L adopts the BindingMOAD data extension used by DiffDock-L. Notably, SADock-L does not use the vdM augmentation or inference-time tuning employed by DiffDock-L, as the latter is not directly compatible with SANet. All models are trained on a single NVIDIA A100 GPU using the same training hyperparameters as DiffDock. More details are listed in Appendix C.1.

Baselines and metrics. We compare SADock with search-based SMINA (Koes et al., 2013) and GNINA (McNutt et al., 2021), including their P2Rank-guided variants (Kriválk & Hoksza, 2018), and regression-based baselines EquiBind (Stärk et al., 2022), TANKBind (Lu et al., 2022) and FABind+ (Gao et al., 2025). DiffDock (Corso et al., 2023), DiffDock-L (Corso et al., 2024), and ForceFM (Guo et al., 2025) serve as the diffusion baselines. We report top-1/top-5 success rates at the standard 2 Å ligand-RMSD threshold and median (Med.) RMSD in Å, where top-1 uses the highest-confidence pose and top-5 the best among the five highest-confidence poses. Detailed evaluation protocols are provided in Appendix C.2.

3.2 Blind Docking Performance

In-distribution performance. We first evaluate blind docking on the held-out PDBBind test set, which primarily measures in-distribution generalization. As shown in Table 1, SADock approaches the best top-1 docking performance among the evaluated methods, while SADock-L further reaches a top-1 success rate of 46.5% with a median RMSD of 2.2 Å, yielding the best overall performance. Compared with the most direct DiffDock baselines, SADock improves top-1 and top-5 success rates by 5.7% and 10.8%, respectively, while SADock-L improves over DiffDock-L by 3.5% and 6.7%. These results show that incorporating surface representations improves docking accuracy even within the training distribution.

Table 1: Blind docking performance on the PDBBind 2019 test set and DockGen-full. Parentheses indicate the number of generated conformations. Blank entries denote unreported results. The best and second-best results are highlighted in bold and underlined, respectively.

	PDBBind 2019 test (363)				DockGen-full (189)			
	RMSD $<$ 2Å (%)		Med. RMSD (Å)		RMSD $<$ 2Å (%)		Med. RMSD (Å)	
Method	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
Search-based								
SMINA	18.7	–	7.1	–	7.9	–	13.8	–
GNINA	22.9	–	7.7	–	14.3	–	15.2	–
P2Rank+SMINA	20.4	–	4.3	–	7.9	–	14.1	–
P2Rank+GNINA	28.8	–	4.9	–	13.8	–	16.2	–
Regression-based								
TANKBind	20.4	–	4.0	–	0.5	–	11.6	–
EquiBind	5.5	–	6.2	–	0.0	–	13.3	–
FABind+	<u>44.9</u>	–	2.4	–	1.5	–	–	–
Diffusion-based								
ForceFM (40)	41.1	–	2.3	–	6.5	–	–	–
DiffDock (40)	38.8	46.0	3.3	2.4	5.3	11.1	7.8	5.1
SADock (40)	44.6	**56.7**	2.8	**1.7**	15.3	28.6	5.2	3.9
DiffDock-L (10)	43.0	49.6	2.8	2.4	<u>22.8</u>	<u>29.6</u>	<u>4.3</u>	<u>3.5</u>
SADock-L (10)	**46.6**	<u>56.2</u>	**2.2**	**1.7**	**27.0**	**38.6**	**3.4**	**2.4**

Out-of-distribution generalization. The advantage becomes more pronounced on DockGen, which evaluates transfer to unseen ligand-binding domains. SADock-L achieves top-1/top-5 success rates of 27.0%/38.5%, outperforming the strongest competing blind-docking baseline, DiffDock-L, by 4.4% and 8.9%, respectively. Notably, this improvement is obtained with less data augmentation and without the inference-time tuning used by DiffDock-L. The gains are also reflected in pose quality: for top-1 predictions on DockGen, SADock reduces median RMSD by 33.3% relative to DiffDock, while SADock-L achieves a 21.0% reduction relative to DiffDock-L. Together, these results indicate that surface-aware interaction modeling provides particularly strong benefits under structural distribution shift, consistent with its intended role in improving transfer across unfamiliar binding environments.

Comparison with localization-oriented methods. We further compare with methods that explicitly target pocket localization. FABind+ strengthens geometric pocket prediction through dynamic pocket sizing and permutation-invariant pose learning (Gao et al., 2025), achieving strong in-distribution performance but degrading sharply on DockGen. By contrast, P2Rank-guided conventional docking remains competitive under distribution shift despite its simpler localization mechanism. P2Rank predicts ligandability from geometric and chemical descriptors centered on solvent-accessible surface points (Kriv\'ak & Hoksza, 2018), providing an independent example of a transferable surface-centered localization prior. These results are consistent with our hypothesis that interaction-facing surface features provide a more robust bias for pocket localization under limited supervision. Unlike P2Rank-based two-stage pipelines, SADock integrates this prior directly into receptor-wide score modeling, jointly optimizing pocket discovery and pose generation.

Further analysis on OOD generalization. To better understand the source of the gains on DockGen, we analyze both pose accuracy and pocket localization for SADock and DiffDock. To reduce sensitivity to individual samples, we evaluate the five highest-confidence poses for each complex and report success curves over increasing RMSD and pocket-centroid-distance thresholds. As shown in Figure 3, SADock consistently outperforms DiffDock across a wide range of thresholds for both metrics, indicating that its advantage is not tied to a particular success cutoff. The improvement is especially pronounced in pocket localization, supporting our hypothesis that surface-aware representations facilitate receptor-wide search under structural distribution shift.

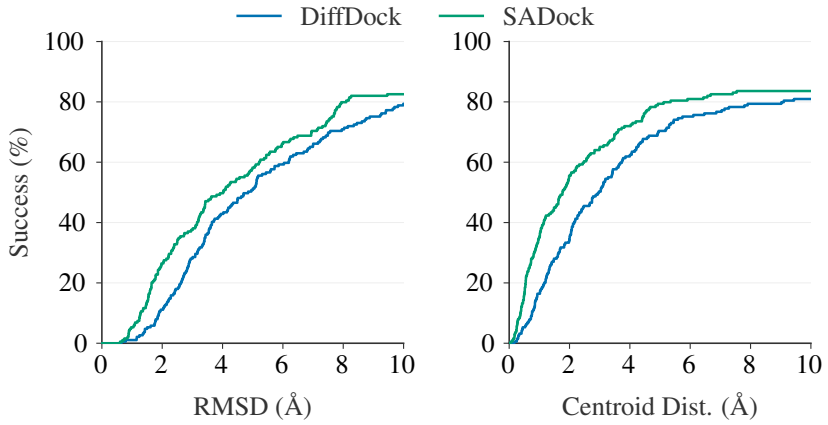

Figure 3: Pose recovery and pocket localization on DockGen-full, using the five highest-confidence candidates per complex. Curves show cumulative success against ligand heavy-atom RMSD (left) and heavy-atom centroid distance to the reference pose (right).

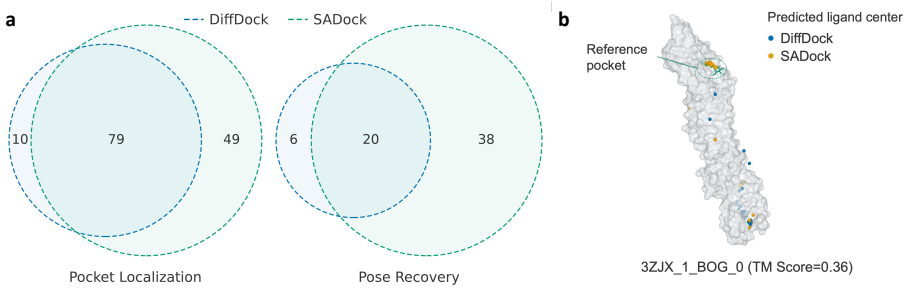

Figure 4: Pocket localization and pose recovery on DockGen with 40 samples. **(a)** Overlap of oracle success sets over 189 DockGen complexes. Pocket localization and pose recovery are considered successful if any candidate achieves centroid distance < 2 Å or ligand RMSD < 2 Å, respectively. Numbers denote cases solved exclusively by either method or by both; unsuccessful cases are omitted. All statistics use the full candidate pool, independent of confidence ranking. **(b)** Candidate ligand centroids for `3ZJX_1_BOG_0` (maximum receptor TM Score to PDBBind training: 0.36). The protein is shown in gray, the reference ligand in green, and predicted centroids in blue (DiffDock) or orange (SADock).

We further examine which complexes are uniquely recovered by SADock. As shown in Figure 4a, SADock covers most of the successful cases solved by DiffDock while additionally recovering targets with substantially lower maximum TM Score to the training set, i.e., structurally less familiar receptors. This trend is consistent with our motivation that interaction-facing surface patterns can remain transferable even when global receptor structure changes considerably. A representative example is shown in Figure 4b: the target has a maximum TM Score of only 0.36 to the training set, for which SADock reaches the native pocket with a minimum centroid distance of 1.36 Å, compared with 14.09 Å for DiffDock. Together, these analyses suggest that the OOD gains of SADock are closely associated with more reliable pocket localization on structurally unfamiliar receptors.

3.3 Ablation on surface representations.

To disentangle the contribution of surface representations, we ablate both the surface branch and the time-conditioned FiLM modulation. As shown in Table 2, introducing surface information consistently improves docking success and pose accuracy over the Euclidean-only model, with substantially larger gains on DockGen than on the in-distribution PDBBind test set. This indicates that surface representations are particularly beneficial under structural distribution shift. Adding FiLM further improves performance, suggesting that adaptively balancing surface and Euclidean features across diffusion time is more effective than using a fixed fusion scheme. These results are consistent

Table 2: Ablation results on surface representations.

		PDBBind		DockGen	
Surface	FiLM	Top-1	Top-5	Top-1	Top-5
		38.8	45.7	6.3	11.6
✓		39.4	50.7	12.2	19.6
✓	✓	44.6	56.7	15.3	28.6

with our theoretical analysis, which motivates surface features as a transferable interaction-aligned representation while retaining Euclidean detail for fine-grained pose modeling.

4 Related Work

Pocket localization and blind docking. Blind docking requires both locating a compatible binding region and predicting a ligand pose within it. Several learning-based methods explicitly decompose these objectives. TANKBind reconstructs poses from candidate binding regions and learned geometric constraints (Lu et al., 2022), while FABind jointly learns ligand-conditioned pocket prediction and pose regression (Pei et al., 2023). Subsequent methods further improve localization and refinement through adaptive pocket modeling, candidate selection, hierarchical refinement, or flexible receptor updates (Gao et al., 2025; Yan et al., 2024; Zhang et al., 2025). An alternative is to combine a pocket detector with a downstream docking engine. P2Rank predicts ligandability from geometric and chemical descriptors centered on solvent-accessible surface points (Kriv\'ak & Hoksza, 2018), and has been paired with conventional docking programs in DockGen (Corso et al., 2024). These studies highlight pocket localization as a central component of blind docking, while also showing that improved localization does not by itself determine end-to-end docking accuracy.

Surface representations for molecular interactions. Molecular surfaces provide an interaction-centered representation complementary to atom- and residue-level geometry. MaSIF learns geometric and physicochemical surface fingerprints for interaction-site recognition (Gainza et al., 2020), and dMaSIF improves their efficiency through on-the-fly surface construction and geometric convolutions (Sverrisson et al., 2021). Surface information has also been incorporated into pose prediction: DeepDock combines surface-aware target representations with learned distance potentials (Méndez-Lucio et al., 2021), while SurfDock integrates surface, residue, and sequence features into diffusion-based pose generation (Cao et al., 2025). Notably, SurfDock conditions pose generation on a specified protein pocket; therefore, the role of surface information in receptor-wide localization and pose generation under blind docking remains less explored.

Unlike prior pocket-conditioned surface methods, SADock maintains protein and ligand surface representations throughout receptor-wide blind docking, coupling them with Euclidean geometry for localization and pose refinement, with theory characterizing when such boundary-preserving transfer is expected to hold.

5 Conclusions

We presented SADock, a surface-aware geometric diffusion framework for generalizable blind protein–ligand docking. By complementing Euclidean representations with molecular-surface geometry and physicochemical context, SADock emphasizes interaction-facing patterns while retaining residue- and atom-level detail for accurate pose modeling. Our theoretical analysis characterizes sufficient conditions that explain the transferability of surface representations, with extensive experiments validating their effectiveness. While SADock mitigates generalization failures arising from the mismatch between structural variation and interaction-level invariance, its advantages naturally diminish when the binding interface itself becomes structurally or chemically novel. Future work will therefore explore ways to expand the range of transferable interaction patterns, including corruption-based augmentation (Corso et al., 2024) and physics-guided modeling (Guo et al., 2025). We hope this work encourages interaction-aligned representations as a broader direction for transferable geometric learning in molecular systems.

AI USE STATEMENT

We used conversational generative AI tools to assist with language polishing. No experimental data were generated by these tools. Multiple authors reviewed the AI-assisted work. AI assistance does not replace scientific validation, and the authors take full responsibility for the final content of this work, including its text, claims, mathematics, code, and figures.

ETHICS STATEMENT

This work studies computational protein–ligand docking using existing, publicly available datasets in accordance with their terms of use. No new datasets were collected, and the study does not involve human participants or animal experiments. To the best of our knowledge, this work raises no additional ethical concerns.

REPRODUCIBILITY STATEMENT

The model and algorithms are described in Section 2 and Appendix A, with theoretical assumptions and proofs provided in Appendix B. The datasets and experimental setup are specified in Section 3.1, while implementation, training, and evaluation details are given in Appendix C.1 and Appendix C.2. Upon acceptance, we will publicly release the source code, model weights, environment specifications and dependencies, training, inference, and evaluation scripts, configuration files, and data splits. Settings not explicitly specified in the paper or appendix follow the defaults in our implementation and configuration files, which will be included in the release to support reproduction of the reported results.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024. doi: 10.1038/s41586-024-07487-w.

Duanhua Cao, Mingan Chen, Runze Zhang, Zhaokun Wang, Manlin Huang, Jie Yu, Xinyu Jiang, Zhehuan Fan, Wei Zhang, Hao Zhou, Xutong Li, Zunyun Fu, Sulin Zhang, and Mingyue Zheng. SurfDock is a surface-informed diffusion generative model for reliable and accurate protein–ligand complex prediction. *Nature Methods*, 22(2):310–322, 2025. doi: 10.1038/s41592-024-02516-y.

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. URL https://openreview.net/forum?id=kKF8_K-mBbS.

Gabriele Corso, Arthur Deng, Nicholas Polizzi, Regina Barzilay, and Tommi Jaakkola. Deep confident steps to new pockets: Strategies for docking generalization. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/db334db287337b2a365120b524300ef3-Abstract-Conference.html.

Pablo Gainza, Freyr Sverrisson, Federico Monti, Emanuele Rodolà, 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. doi: 10.1038/s41592-019-0666-6.

Pablo Gainza, Sarah Wehrle, Alexandra Van Hall-Beauvais, Anthony Marchand, Andreas Scheck, Zander Harteveld, Stephen Buckley, Dongchun Ni, Shuguang Tan, Freyr Sverrisson, et al. De novo design of protein interactions with learned surface fingerprints. *Nature*, 617(7959):176–184, 2023.

Kaiyuan Gao, Qizhi Pei, Gongbo Zhang, Jinhua Zhu, Kun He, and Lijun Wu. FABind+: Enhancing molecular docking through improved pocket prediction and pose generation. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2025. doi: 10.1145/3690624.3709253. URL https://arxiv.org/abs/2403.20261v4.

Huanlei Guo, Song Liu, and Bingyi Jing. ForceFM: Enhancing protein-ligand predictions through force-guided flow matching. In *Advances in Neural Information Processing Systems*, volume 38, pp. 124268–124296. Curran Associates, Inc., 2025. doi: 10.52202/085713-4140. URL https://papers.nips.cc/paper_files/paper/2025/hash/b3ee701cea4872c41356f32592a72289-Abstract-Conference.html.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2.

David Ryan Koes, Matthew P. Baumgartner, and Carlos J. Camacho. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *Journal of Chemical Information and Modeling*, 53(8):1893–1904, 2013. doi: 10.1021/ci300604z.

Janez Konc and Dušanka Janežič. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics*, 26(9):1160–1168, 2010. doi: 10.1093/bioinformatics/btq100.

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*, 10:39, 2018. doi: 10.1186/s13321-018-0285-8.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

David Lopez-Paz and Maxime Oquab. Revisiting classifier two-sample tests. *arXiv preprint arXiv:1610.06545*, 2016.

Wei Lu, Qifeng Wu, Jixian Zhang, Jiahua Rao, Chengtao Li, and Shuangjia Zheng. TANKBind: Trigonometry-aware neural networks for drug-protein binding structure prediction. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://www.biorxiv.org/content/10.1101/2022.06.06.495043v3.

Jiankun Lyu, Sheng Wang, Trent E. Balius, Isha Singh, Anat Levit, Yurii S. Moroz, Matthew J. O'Meara, Tao Che, Enkhjargal Algaa, Kateryna Tolmachova, Andrey A. Tolmachev, Brian K. Shoichet, Bryan L. Roth, and John J. Irwin. Ultra-large library docking for discovering new chemotypes. *Nature*, 566(7743):224–229, 2019. doi: 10.1038/s41586-019-0917-9.

J. Andrew McCammon et al. Profiling charge complementarity and selectivity for binding at the protein surface. *Biophysical Journal*, 84(4):2712–2718, 2003.

Andrew T. McNutt, Paul Francoeur, Rishal Aggarwal, Tomohide Masuda, Rocco Meli, Matthew Ragoza, Jocelyn Sunseri, and David Ryan Koes. GNINA 1.0: molecular docking with deep learning. *Journal of Cheminformatics*, 13:43, 2021. doi: 10.1186/s13321-021-00522-2.

Oscar Méndez-Lucio, Mazen Ahmad, Ehecatl Antonio del Rio-Chanona, and Jörg Kurt Wegner. A geometric deep learning approach to predict binding conformations of bioactive molecules. *Nature Machine Intelligence*, 3(12):1033–1039, 2021. doi: 10.1038/s42256-021-00409-9.

Alex Morehead, Nabin Giri, Jian Liu, Pawan Neupane, and Jianlin Cheng. Assessing the potential of deep learning for protein–ligand docking. *Nature Machine Intelligence*, 8(1):32–41, 2026. doi: 10.1038/s42256-025-01160-1.

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. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://arxiv.org/abs/2310.06763.

Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Stefan D. Search, Christopher D. Cooper, and Elwin van 't Wout. Towards optimal boundary integral formulations of the Poisson–Boltzmann equation for molecular electrostatics. *Journal of Computational Chemistry*, 43(10):674–691, 2022. doi: 10.1002/jcc.26825.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Hannes Stärk, Octavian-Eugen Ganea, Lagnajit Pattanaik, Regina Barzilay, and Tommi Jaakkola. EquiBind: Geometric deep learning for drug binding structure prediction. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 20503–20521. PMLR, 2022. URL https://proceedings.mlr.press/v162/stark22b.html.

Freyr Sverrisson, Jean Feydy, Bruno E. Correia, and Michael M. Bronstein. Fast end-to-end learning on protein surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. doi: 10.1109/CVPR46437.2021.01502. URL https://ieeexplore.ieee.org/document/9577686/.

Chai Discovery team, Jacques Boitreaud, Jack Dent, Matthew McPartlon, Joshua Meier, Vinicius Reis, Alex Rogozhonikov, and Kevin Wu. Chai-1: Decoding the molecular interactions of life. *BioRxiv*, pp. 2024–10, 2024.

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. URL https://arxiv.org/abs/1802.08219.

Jiaxian Yan, Zaixi Zhang, Jintao Zhu, Kai Zhang, Jianfeng Pei, and Qi Liu. DeltaDock: A unified framework for accurate, efficient, and physically reliable molecular docking, 2024. URL https://arxiv.org/abs/2410.11224v1. Version 1.

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. URL https://arxiv.org/abs/2502.14934.

APPENDIX

A ADDITIONAL METHODOLOGICAL DETAILS

Here we provide more methodological implementation details, as an extension to Section 2.

A.1 Diffusion Models

Training and sampling. We train SANet with the same component-wise denoising score-matching objective as DiffDock:

$$\mathcal{L}_{\mathrm{DSM}}(\theta) = \mathbb{E}\left[\sum_{a\in\mathcal{I}} \|s^a_\theta(\mathcal{Z}_t, t) - \nabla_{q^a_t} \log K^a_t(q^a_t \mid q^a_0)\|_2^2\right]. \tag{9}$$

The expectation is over training complexes, sampled diffusion times, and forward perturbations, with $q_t \sim K_t(\cdot \mid q_0)$. We retain DiffDock's conformer preparation, noise schedules, and reverse-diffusion sampling procedure, replacing only the score network with SANet. Starting from the same noise prior, sampling iteratively updates ligand translation, rotation, and torsions to generate poses across the receptor. Our contribution therefore lies in surface-aware interaction modeling within the denoising network, rather than in a modified diffusion objective or sampler. The following section details the message-passing and time-modulation modules that implement this representation.

A.2 Block Ordering and Feature Types

We write $u \in \{p, l\}$ for the protein or ligand. Euclidean features are organized into rotation-order blocks $\mathbf{X}^{u,(r)}_i \in \mathbb{R}^{C_r\times(2r+1)}$, while surface hidden features $\mathbf{S}^u_k$ are rotation-invariant scalar channels. Within each SANet block, the receptor surface follows

$$\texttt{Geodes-MP} \to \texttt{F2C} \to \texttt{Hetero-MP} \to \texttt{Inter-MP} \to \texttt{C2F}.$$

The ligand surface follows the same interaction pathway without hierarchical pooling. In parallel, the Euclidean branches perform intramolecular message passing before `Hetero-MP`, followed by FiLM modulation and `Inter-MP`. Importantly, receptor pooling is applied to the features learned by fine-resolution `Geodes-MP`, rather than to the raw surface descriptors.

A.3 Geodesic Message Passing

Discrete geodesic distance. We compute surface distances from the original NanoShaper triangulation before any hierarchical pooling. The underlying mesh graph contains one node for each triangular face, represented by its centroid, and connects two nodes only when the corresponding faces share a mesh edge. For adjacent faces f and f' with centroids $\mathbf{c}_f$ and $\mathbf{c}_{f'}$ and shared edge $e_{ff'}$, the edge weight is the shortest piecewise-linear path passing through their common edge:

$$w_{ff'} = \min_{\mathbf{x}\in e_{ff'}} \left(\|\mathbf{c}_f - \mathbf{x}\|_2 + \|\mathbf{c}_{f'} - \mathbf{x}\|_2\right). \tag{10}$$

For each pair of surface model nodes i and j, we associate them with representative mesh faces and define the discrete geodesic distance as the weighted shortest-path distance on this graph:

$$d_{G,ij} = \mathrm{ShortestPath}_{\mathcal{G}_{\mathrm{mesh}}}(f_i, f_j), \tag{11}$$

which is computed using Dijkstra's algorithm. This construction follows the connectivity of the molecular surface and therefore avoids shortcuts through nearby but disconnected surface regions.

Geometric edge encoding. For a directed message $j \to i$, let

$$d_{E,ij} = \|\mathbf{s}_j - \mathbf{s}_i\|_2, \qquad \widehat{\mathbf{r}}_{ij} = \frac{\mathbf{s}_j - \mathbf{s}_i}{d_{E,ij}}. \tag{12}$$

We encode the geometric relation between the two surface nodes as

$$\mathbf{g}_{ij} = \left[\mathrm{RBF}(d_{E,ij}), \mathrm{RBF}(d_{G,ij}), \Delta_{ij}, \mathbf{n}_i^\top \mathbf{n}_j, \mathbf{n}_i^\top \widehat{\mathbf{r}}_{ij}, \mathbf{n}_j^\top \widehat{\mathbf{r}}_{ij}\right], \tag{13}$$

where

$$\Delta_{ij} = \log(1 + d_{G,ij}) - \log(1 + d_{E,ij}). \tag{14}$$

The two radial encodings describe ambient and intrinsic separation, while Δ_{ij} explicitly measures their discrepancy. The remaining terms characterize the relative orientation of the two local surface normals and the inter-node direction.

Equivariant surface messages. Surface nodes carry geometric and physicochemical attributes such as curvature, represented area, hydrophobicity, and other chemical descriptors. Their hidden states contain scalar and vector $O(3)$ irreducible representations up to degree one. Directional dependence is represented using spherical harmonics up to degree two, which are combined with $\mathbf{g}_{ij}$ and neighboring surface features through equivariant tensor-product operations. Denoting the resulting message from j to i by $\mathbf{m}_{j\to i}$, the update uses mean aggregation:

$$\overline{\mathbf{m}}_i = \frac{1}{|\mathcal{N}_i^G|} \sum_{j\in\mathcal{N}_i^G} \mathbf{m}_{j\to i}. \tag{15}$$

Here $\mathcal{N}_i^G$ denotes the surface neighbors participating in message passing. The aggregated message is combined with the current surface state through a residual update.

By jointly encoding d_E and d_G, the operator can distinguish two geometrically different cases that appear similar in ambient space: points that are genuinely close along the surface and points that are Euclidean-near but separated by the molecular boundary. This distinction is particularly important around pockets, grooves, and folded surface regions.

A.4 Hierarchical Surface Operators

The receptor surface typically contains substantially more nodes than the residue graph. We therefore preserve fine resolution for local surface encoding and use a coarse representation for subsequent cross-representation and protein–ligand interactions.

Let the receptor surface contain K_p fine points and $M_p \ll K_p$ coarse representatives. Denote by $\mathbf{S}_k^p$ the encoded feature of fine surface point k after `Geodes-MP`, and by A_k^p its represented surface area. Given nonnegative assignment weights π_{kc} from fine point k to coarse representative c, `F2C` performs area-weighted aggregation:

$$A_c^p = \sum_k A_k^p \pi_{kc}, \qquad \mathbf{S}_c^p = \frac{\sum_k A_k^p \pi_{kc} \mathbf{S}_k^p}{A_c^p}. \tag{16}$$

The coarse features are then updated by `Hetero-MP` and `Inter-MP`; we denote the resulting interaction-aware features by $\mathbf{S}_c^{p,\mathrm{int}}$. Finally, `C2F` propagates this coarse context back to the fine surface:

$$\mathbf{S}_k^{p,\mathrm{out}} = \mathbf{S}_k^p + \phi_{\mathrm{up}}\left(\left[\mathbf{S}_k^p, \sum_c \pi_{kc}\mathbf{S}_c^{p,\mathrm{int}}\right]\right). \tag{17}$$

where ϕ_{up} is a learnable MLP that projects the concatenated fine-scale feature and interpolated coarse context back to the surface feature dimension. The skip connection preserves the local patterns learned before pooling, while the interpolated coarse features inject broader interaction context. The ligand surface remains at its original resolution throughout this procedure.

A.5 Cross-Representation and Intermolecular Message Passing

We distinguish two types of cross-type interaction. `Hetero-MP` exchanges information within each molecule,

$$E_p \leftrightarrow S_p, \qquad E_l \leftrightarrow S_l,$$

whereas `Inter-MP` communicates across the two interaction partners,

$$E_p \leftrightarrow E_l, \qquad S_p \leftrightarrow S_l.$$

Here E_p and E_l denote protein residues and ligand atoms, and S_p and S_l denote their surface nodes.

For a directed relation $A \leftarrow B$, messages take the generic form

$$\mathbf{m}_i^{A\leftarrow B} = \mathrm{AGG}_{j\in\mathcal{N}_i^{A\leftarrow B}} \Phi_{A\leftarrow B}\left(\mathbf{v}_i^A, \mathbf{v}_j^B, \mathrm{RBF}(d_{ij}^{AB}), \widehat{\mathbf{d}}_{ij}^{AB}, \mathbf{e}_t\right), \tag{18}$$

where each relation uses separate parameters and neighborhood cutoffs. Tensor-product mappings are used when the receiver carries Euclidean equivariant features, while invariant mappings produce scalar surface features. This bidirectional coupling enriches surface descriptors with molecular context and propagates interaction-boundary information back to the Euclidean branch.

For `Inter-MP`, cross-partner relations are constructed in the shared ambient three-dimensional coordinate system. In particular, protein and ligand surfaces are treated as two distinct manifolds, so their intermolecular interaction uses Euclidean relative geometry rather than a common intrinsic surface distance.

A.6 Equivariant Time Modulation

FiLM modulates Euclidean features using both the incoming surface context and diffusion time. Let

$$\mathbf{c}_i^u = (\mathbf{m}_i^{E_u \leftarrow S_u})^{(0)} \tag{19}$$

denote the invariant surface message received by Euclidean node i. For each rotation order r, the modulation parameters are generated as

$$\left(\{\boldsymbol{\gamma}_i^{u,(r)}\}_r, \boldsymbol{\beta}_i^{u,(0)}\right) = \mathrm{MLP}^u_{\mathrm{FiLM}}([\mathbf{c}_i^u, \mathbf{e}_t]), \tag{20}$$

and applied through

$$\widetilde{\mathbf{X}}_i^{u,(r)} = (\mathbf{1} + \boldsymbol{\gamma}_i^{u,(r)}) \odot \check{\mathbf{X}}_i^{u,(r)} + \boldsymbol{\beta}_i^{u,(r)}. \tag{21}$$

For $r > 0$, the additive bias is set to zero, while each scaling coefficient is shared across the $2r + 1$ components of the corresponding irreducible representation. The modulation therefore preserves rotational equivariance while allowing the influence of surface context to vary with diffusion time.

B Analysis of Surface-Based Transferability

B.1 Comparison Domain and Surface Discrepancy

We compare receptors under a fixed ligand identity, chemical preparation, temperature, and environment. The receptors and ligand queries are expressed in a common coordinate frame, and the basin partition $\{\Omega_z\}_{z\in\mathcal{Z}}$ is fixed before comparison. The basins cover the competing binding regions relevant to the docking problem, using a common configuration representation and reference measure as required by Eq. 6.

The surface discrepancy d_S measures changes in both embedded boundary geometry and physicochemical attributes. For matched surface point sets, we define an area-weighted discrepancy

$$d_S^2(S, S') = \frac{1}{A_S} \sum_k A_k \left[\frac{\|\mathbf{s}_k - \mathbf{s}'_k\|_2^2}{r_0^2} + \lambda_n \|\mathbf{n}_k - \mathbf{n}'_k\|_2^2 + \|W_f(\mathbf{f}_k - \mathbf{f}'_k)\|_2^2 \right], \tag{22}$$

where $A_S = \sum_k A_k$, A_k is the represented surface area of point k, $r_0 > 0$ sets the geometric length scale, and λ_n and W_f balance normal and physicochemical differences. This definition can be viewed as an expectation over the normalized surface-area measure and therefore characterizes the typical boundary variation rather than the single most perturbed surface point.

When a uniform physical bound is required, we additionally use the corresponding L_∞ discrepancy

$$d_{S,\infty}(S, S') = \max_k \left[\frac{\|\mathbf{s}_k - \mathbf{s}'_k\|_2}{r_0} + \sqrt{\lambda_n} \|\mathbf{n}_k - \mathbf{n}'_k\|_2 + \|W_f(\mathbf{f}_k - \mathbf{f}'_k)\|_2 \right]. \tag{23}$$

Different surface samplings are compared through a common resampling or correspondence, with the associated approximation error included in the analysis. Near-contact comparisons additionally require preservation of steric accessibility and minimum intermolecular separation, as discussed in Appendix B.3.

B.2 Proof of Boundary-to-Basin Stability

Let

$$Z_X(z) = \int_{\Omega_z} e^{-\beta U_X(\xi)}\, d\mu(\xi), \qquad G_X(z) = -\beta^{-1} \log Z_X(z), \tag{24}$$

and write $b = b_\ell$. By Eq. 7, for any $\delta > 0$ there exists $c_\delta \in \mathbb{R}$ such that

$$|U_{X'}(\xi) - U_X(\xi) - c_\delta| \le b + \delta, \qquad \forall\, \xi \in \Omega. \tag{25}$$

Therefore, for every basin z,

$$e^{-\beta(c_\delta + b + \delta)} Z_X(z) \le Z_{X'}(z) \le e^{-\beta(c_\delta - b - \delta)} Z_X(z). \tag{26}$$

Taking $-\beta^{-1}\log(\cdot)$ gives

$$c_\delta - b - \delta \le G_{X'}(z) - G_X(z) \le c_\delta + b + \delta. \tag{27}$$

Since Eq. 27 holds uniformly over z and c_δ is independent of the basin index, taking the oscillation over z gives

$$\eta_* = \operatorname{osc}_z \left[G_{X'}(z) - G_X(z)\right] \le 2(b + \delta), \tag{28}$$

where, for a function $f : \mathcal{Z} \to \mathbb{R}$, $\operatorname{osc}_z f := \sup_{z\in\mathcal{Z}} f(z) - \inf_{z\in\mathcal{Z}} f(z)$. Letting $\delta \to 0$ yields

$$\eta_* \le 2b_\ell. \tag{29}$$

Distributional stability. The stronger statement in terms of the complete basin distribution follows by defining

$$\pi_X(z) = \frac{e^{-\beta G_X(z)}}{\sum_{z'\in\mathcal{Z}} e^{-\beta G_X(z')}}. \tag{30}$$

We measure the discrepancy between the two basin distributions using the total variation distance,

$$\mathrm{TV}(\pi_X, \pi_{X'}) = \frac{1}{2}\sum_{z\in\mathcal{Z}} \left|\pi_X(z) - \pi_{X'}(z)\right|. \tag{31}$$

Equivalently,

$$\mathrm{TV}(\pi_X, \pi_{X'}) = \sup_{A\subseteq\mathcal{Z}} \left|\pi_X(A) - \pi_{X'}(A)\right|. \tag{32}$$

Thus, total variation measures the largest change in probability assigned to any collection of binding basins, making it suitable for assessing transferability beyond the identity of the most probable pocket.

Let

$$d(z) = G_{X'}(z) - G_X(z), \qquad w(z) = e^{-\beta d(z)}. \tag{33}$$

Then

$$\pi_{X'}(z) = \frac{w(z)\pi_X(z)}{\mathbb{E}_{\pi_X}[w]}. \tag{34}$$

Let $\bar{w} = \mathbb{E}_{\pi_X}[w]$ and suppose that $0 < w_{\min} \le w(z) \le w_{\max}$. From Eq. 34,

$$\mathrm{TV}(\pi_X, \pi_{X'}) = \frac{1}{2\bar{w}}\mathbb{E}_{\pi_X}\left[|w - \bar{w}|\right] \tag{35}$$

$$\le \frac{(w_{\max} - \bar{w})(\bar{w} - w_{\min})}{\bar{w}(w_{\max} - w_{\min})} \tag{36}$$

$$\le \frac{\sqrt{w_{\max}} - \sqrt{w_{\min}}}{\sqrt{w_{\max}} + \sqrt{w_{\min}}}. \tag{37}$$

The first inequality follows from the maximal mean absolute deviation of a variable supported on $[w_{\min}, w_{\max}]$. The second follows by maximizing over $\bar{w} \in [w_{\min}, w_{\max}]$, with equality at $\bar{w} = \sqrt{w_{\min} w_{\max}}$.

Since $w_{\max}/w_{\min} = e^{\beta\eta_*}$,

$$\mathrm{TV}(\pi_X, \pi_{X'}) \le \tanh\left(\frac{\beta\eta_*}{4}\right) \le \tanh\left(\frac{\beta b_\ell}{2}\right). \tag{38}$$

The first inequality is tight when only η_* is specified.

Preservation of the preferred basin. Let z_* be the unique preferred basin of X with margin

$$m_X = \min_{z \neq z_*} [G_X(z) - G_X(z_*)]. \tag{39}$$

For any $z \neq z_*$,

$$G_{X'}(z) - G_{X'}(z_*) \geq m_X - \eta_*. \tag{40}$$

Thus,

$$\eta_* < m_X \quad \Longrightarrow \quad \arg\min_z G_{X'}(z) = z_*. \tag{41}$$

With $\eta_* \leq 2b_\ell$, we obtain the sufficient condition

$$\boxed{2b_\ell < m_X \quad \Longrightarrow \quad z_* \text{ remains the preferred binding basin.}} \tag{42}$$

The total-variation result controls the complete coarse basin distribution, whereas Eq. 42 isolates the condition most directly relevant to pocket localization in blind docking.

B.3 Physical Interpretation of the Boundary Condition

The quantity b_ℓ in Eq. 7 summarizes pose-dependent interaction changes induced by receptor variation. We relate it to three physically interpretable contributions: boundary electrostatics, direct short-range interactions, and residual response terms.

Electrostatic perturbations. We first isolate the electrostatic contribution for rigid receptors with a common smooth dielectric boundary Γ and homogeneous exterior domain D. The internal charge distributions may differ, while the exterior potentials satisfy

$$(-\Delta + \kappa^2)\phi_X = 0 \quad \text{in } D, \qquad \phi_X|_\Gamma = b_X, \qquad \phi_X \to 0 \quad \text{at infinity.} \tag{43}$$

Here $\kappa \geq 0$, and all receptor charge sources lie inside Γ. Boundary-integral formulations provide the corresponding exterior-field representation (Search et al., 2022). Applying the maximum principle to the difference between two solutions gives

$$\|\phi_{X'} - \phi_X\|_{\infty,D} \leq \|b_{X'} - b_X\|_{\infty,\Gamma} =: \delta_\phi. \tag{44}$$

Under a fixed receptor field,

$$U_{\mathrm{el}}^X(\xi) = \sum_{a \in \mathcal{L}} Q_a \phi_X(\mathbf{r}_a(\xi)), \tag{45}$$

and therefore

$$\sup_{\xi \in \Omega} |U_{\mathrm{el}}^{X'}(\xi) - U_{\mathrm{el}}^X(\xi)| \leq Q_{\mathrm{abs}} \delta_\phi, \qquad Q_{\mathrm{abs}} = \sum_{a \in \mathcal{L}} |Q_a|. \tag{46}$$

Thus, internal charge rearrangements have a limited electrostatic effect whenever the ligand-facing boundary field remains similar.

Short-range perturbations. Consider perturbed receptor atoms interacting with ligand atoms through fixed-parameter Lennard–Jones potentials

$$V_{ia}(r) = C_{12,ia} r^{-12} - C_{6,ia} r^{-6}. \tag{47}$$

Let receptor atom i move along a path of length at most u_i, while remaining at least $h_{ia} > 0$ from ligand atom a over all configurations of interest. The mean-value theorem gives

$$B_{\mathrm{sr}} \leq \sum_{i \in \mathrm{pert}} \sum_{a \in \mathcal{L}} u_i \left(\frac{12 C_{12,ia}}{h_{ia}^{13}} + \frac{6 C_{6,ia}}{h_{ia}^{7}} \right). \tag{48}$$

The strong decay with h_{ia} captures the limited direct influence of sufficiently buried structural changes when pocket accessibility and the outer interaction shell are preserved.

Combined perturbation bound. Let B_{other} denote the contribution from pose-dependent interaction effects not included in the electrostatic and short-range terms. Absorbing pose-independent

receptor energy changes into c gives

$$\inf_{c\in\mathbb{R}} \sup_{\xi\in\Omega} |U_{X'}(\xi) - U_X(\xi) - c| \le Q_{\mathrm{abs}}\delta_\phi + B_{\mathrm{sr}} + B_{\mathrm{other}} =: B_{\mathrm{phys}}. \tag{49}$$

Hence the basin-stability result applies whenever B_{phys} is controlled by the surface discrepancy:

$$B_{\mathrm{phys}} \le K_\ell \delta_S + \varepsilon_\ell. \tag{50}$$

A sufficient condition for pocket preservation is therefore

$$2B_{\mathrm{phys}} < m_X. \tag{51}$$

From physical fields to finite surface features. SANet observes finite geometric and physicochemical descriptors rather than complete physical boundary fields. For a chosen perturbation class, suppose

$$\delta_\phi \le C_\phi \delta_S + \epsilon_\phi, \qquad B_{\mathrm{sr}} + B_{\mathrm{other}} \le C_{\mathrm{rem}}\delta_S + \epsilon_{\mathrm{rem}}. \tag{52}$$

Then

$$B_{\mathrm{phys}} \le (Q_{\mathrm{abs}}C_\phi + C_{\mathrm{rem}})\delta_S + Q_{\mathrm{abs}}\epsilon_\phi + \epsilon_{\mathrm{rem}}, \tag{53}$$

recovering Eq. 7 with

$$K_\ell = Q_{\mathrm{abs}}C_\phi + C_{\mathrm{rem}}, \qquad \varepsilon_\ell = Q_{\mathrm{abs}}\epsilon_\phi + \epsilon_{\mathrm{rem}}. \tag{54}$$

The residual terms account for finite surface resolution, incomplete physicochemical descriptors, and interaction mechanisms not explicitly represented by the chosen surface features.

Expected regime of applicability. The preceding conditions favor receptor variations that preserve pocket shape, boundary physicochemical fields, steric accessibility, and the local chemistry of direct contacts. Such recurring local interaction environments across globally distinct proteins are consistent with structural binding-site comparisons and molecular-surface fingerprinting studies (Konc & Janežič, 2010; Gainza et al., 2020). Changes in pocket opening, protonation, metal coordination, direct-contact residues, or large-scale conformational equilibria increase the residual terms and therefore weaken the expected transferability.

B.4 Fixed-Query Stability of SANet

The physical analysis identifies receptor variations that preserve the underlying interaction preference. We next characterize whether SANet responds stably to the same variations.

Fix a ligand query $v = (\mathcal{L}, q, t)$ and write one SANet score head as $F_\theta(v; E, S)$, where E and S denote the receptor's Euclidean and surface representations. The notation includes the full SANet computation, including heterogeneous message passing, surface–Euclidean feedback, FiLM modulation, and score readout. Assume that, on the receptor pairs of interest,

$$\|F_\theta(v; E, S') - F_\theta(v; E, S)\| \le L_{S,\theta} d_S(S, S'). \tag{55}$$

Introducing the intermediate evaluation $F_\theta(v; E, S')$ gives

$$\|F_\theta(v; E', S') - F_\theta(v; E, S)\| \le L_{S,\theta}\delta_S + \epsilon_{E,\theta}, \tag{56}$$

where

$$\epsilon_{E,\theta} = \|F_\theta(v; E', S') - F_\theta(v; E, S')\|. \tag{57}$$

The first term measures sensitivity to changes in the interaction boundary. The second isolates additional sensitivity to Euclidean structural variation while holding the surface condition fixed. This decomposition directly reflects SANet's dual representation: surface features organize interaction-aligned variation, while Euclidean features retain complementary structural information.

Operator-level conditions. The local constant $L_{S,\theta}$ can be obtained by composing sensitivities of SANet's constituent operators. On bounded feature domains with stable graph neighborhoods and pooling assignments, the RBF mappings, tensor products, scalar networks, residual updates, and weighted aggregations are locally Lipschitz.

For the tangent projector $\mathbf{P}(\mathbf{n}) = \mathbf{I} - \mathbf{n}\mathbf{n}^\top$,

$$\|\mathbf{P}(\mathbf{n}) - \mathbf{P}(\mathbf{n}')\|_2 \le 2\|\mathbf{n} - \mathbf{n}'\|_2. \tag{58}$$

For $\rho = \|\mathbf{P}(\mathbf{n})\mathbf{d}\|$ and its perturbed counterpart,

$$|\rho - \rho'| \le \|\mathbf{d} - \mathbf{d}'\|_2 + 2\|\mathbf{d}'\|_2\|\mathbf{n} - \mathbf{n}'\|_2. \tag{59}$$

Thus, local geometric perturbations produce controlled changes in the geodesic message-passing kernel when neighborhood connectivity is unchanged.

Area-weighted pooling and interpolation with fixed nonnegative normalized weights are nonexpansive in the maximum feature norm. Variations in assignment weights contribute an additional local sensitivity term.

For a FiLM block $\widetilde{\mathbf{X}} = (1 + \boldsymbol{\gamma}) \odot \mathbf{X} + \boldsymbol{\beta}$, assuming $\|\boldsymbol{\gamma}\|_\infty \le G$ and $\|\mathbf{X}\| \le H$ gives

$$\|\widetilde{\mathbf{X}}' - \widetilde{\mathbf{X}}\| \le (1 + G)\|\mathbf{X}' - \mathbf{X}\| + H\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_\infty + \|\boldsymbol{\beta}' - \boldsymbol{\beta}\|. \tag{60}$$

Composing these bounds yields a finite local $L_{S,\theta}$ determined by the learned weights, feature ranges, and graph construction. Changes in cutoff neighborhoods, nearest-neighbor assignments, or coarse-point sampling contribute additional graph-perturbation terms.

C Additional Experimental Results

C.1 Implementation Details

Both SADock and SADock-L use six SANet interaction blocks. For SADock, the hidden representations of $\mathbf{X}^u$ and $\mathbf{S}^u$ ($u \in \{p, l\}$) are configured as

$$48 \times 0\mathrm{e} + 10 \times 1\mathrm{o} + 10 \times 1\mathrm{e} + 48 \times 0\mathrm{o},$$

resulting in approximately 20M parameters. SADock-L increases the channel capacity to

$$64 \times 0\mathrm{e} + 16 \times 1\mathrm{o} + 16 \times 1\mathrm{e} + 64 \times 0\mathrm{o},$$

yielding approximately 30M parameters. Here, $l_g p_g$ denotes an $O(3)$ irreducible representation of angular degree l_g and parity $p_g \in \{\mathrm{e}, \mathrm{o}\}$, with e and o indicating even and odd parity under spatial inversion. Accordingly, $0\mathrm{e}/0\mathrm{o}$ denote scalar/pseudoscalar channels and $1\mathrm{o}/1\mathrm{e}$ denote polar/axial vector channels; the leading coefficient specifies the channel multiplicity.

SADock is trained on PDBBind for 300K optimization steps, while SADock-L is trained for 400K steps with the additional Binding MOAD data. These training budgets are substantially smaller than those of DiffDock (850K) and DiffDock-L (1250K). Both models follow the optimizer, learning-rate schedule, and training hyperparameters of their corresponding DiffDock baselines, and are trained on a single NVIDIA A100 GPU.

C.2 Evaluation Protocol

For SADock and SADock-L, we use a 20-step diffusion schedule with 19 reverse updates and omit noise injection at the final step. All generated poses are ranked using the DiffDock confidence model. For the diffusion-based methods, we evaluate sampling budgets of 10 and 40 candidate poses per complex.

Ligand RMSD is computed over all heavy atoms in the protein coordinate frame with symmetry correction and without additional rigid-body alignment. When multiple valid reference poses are available, we take the minimum RMSD across references. For each complex m, let $\mathcal{D}_m$ denote the RMSD selected according to the evaluation protocol. We report

$$\text{Success} = \frac{100}{M}\sum_{m=1}^{M} \mathbf{1}[\mathcal{D}_m < 2\,\text{Å}], \qquad \text{Med.RMSD} = \text{median}\{\mathcal{D}_m\}_{m=1}^{M},$$

where M is the size of the complete fixed test cohort. Top-1 uses the highest-confidence candidate, whereas top-5 uses the lowest-RMSD pose among the five highest-confidence candidates. Missing

or failed predictions are assigned $\mathcal{D}_m = +\infty$ and retained in all statistics. For candidate-quality analysis, oracle success uses the lowest-RMSD pose among all candidates generated within the specified sampling budget.

Regarding the empirical results in Section 3, we re-evaluate DiffDock and DiffDock-L using their publicly released checkpoints and regenerate docking poses to ensure a consistent comparison of docking accuracy and pocket localization. Our reproduced results are comparable to, and in some cases better than, those reported in the original papers. Results for the other methods are taken from Corso et al. (2024), except for FABind+ and ForceFM, for which we report the results from the original ForceFM paper (Guo et al., 2025). All tables and figures consistently use our re-evaluated results for DiffDock and DiffDock-L.