PIGEON: Pocket-Inferred Geometric Ensembles for Flexible Docking

Anonymous Authors[1]

Abstract

Molecular docking is a central tool for structure-based drug discovery, but realistic docking remains challenging because receptor conformations are flexible, a receptor-ligand pair may admit multiple stable holo states, and the true binding-pocket residues are often unknown in prospective applications. We introduce **PIGEON** (**P**ocket-**I**nferred **G**eometric **E**nsembles for Flexible D**o**cki**n**g), a manifold flow-matching framework that performs ensemble flexible docking without ground-truth pocket residues. PIGEON represents receptor-ligand complexes with a time-conditioned heterogeneous graph and predicts residual motions for ligand translation, ligand rotation, ligand torsions, receptor residue frames, and receptor side-chain torsions. To guide ligand translation in the absence of pocket labels, PIGEON constructs a receptor-intrinsic buriedness field, a ligand-centered spherical harmonic pocket field, and a compact bank of pocket-aware candidate directions, then uses a learnable staged selector to choose between pocket-ingress and local-fitting motions. Empirically, PIGEON recovers accurate ligand poses while maintaining physically plausible full-complex geometry, providing a practical route to generating coupled ligand-receptor ensembles under realistic pocket-unknown conditions.

1 Introduction

Molecular docking is a central tool in structure-based drug discovery, where predicting how a small molecule binds to a protein target supports virtual screening, lead optimization, and mechanistic interpretation (Morris et al., 2009; Trott & Olson, 2010; Friesner et al., 2004). In many therapeutic targets, rigid-receptor assumptions break down: enzyme active sites can require side-chain or loop rearrangements to admit catalytic or inhibitory binding modes (Sherman et al., 2006), GPCRs and kinases stabilize distinct ligand-specific conformations that drive different pharmacological outcomes (Wootten et al., 2018; Vijayan et al., 2015), and viral proteases such as HIV-1 protease undergo substantial flap and pocket rearrangements upon inhibitor binding (Heaslet et al., 2007). These applications expose three coupled challenges that remain difficult for current docking models: receptors are flexible and may change conformation upon ligand binding, a single receptor-ligand pair may admit multiple stable holo conformations, and the true binding-pocket residues are usually unknown in real applications.

Recent generative docking models have made substantial progress on flexible protein-ligand docking. Building on early geometric one-shot predictors (Stärk et al., 2022; Lu et al., 2022) and diffusion-based pose generation (Corso et al., 2023), methods such as DiffDock-Pocket (Plainer et al., 2023), Re-Dock (Huang et al., 2024), DynamicBind (Lu et al., 2024), FlexDock (Corso et al., 2025), and FlowDock (Morehead & Cheng, 2025) use diffusion or flow matching to predict ligand poses jointly with receptor adaptation. While these models capture important aspects of ligand-induced receptor motion, most are evaluated as single-endpoint pose generation rather than ensemble prediction.

Classical ensemble docking instead models receptor flexibility by docking ligands against multiple pre-generated conformations from crystal structures, molecular dynamics, or enhanced sampling (Lin et al., 2002; Amaro et al., 2008; 2018; Bottegoni et al., 2011). These pipelines are typically two-stage: receptor conformations are generated upstream, and docking, scoring, and ranking are performed separately against each. As a result, ensemble docking workflows focus on selecting from precomputed receptor states rather than directly generating ligand and receptor holo conformations as a coupled conditional distribution.

A further limitation is the treatment of pocket information. Site-specific docking methods generally require an explicit pocket center or residue set (Plainer et al., 2023; Huang et al., 2024), supplied by externally predicted pockets (Le Guilloux et al., 2009; Krivák & Hoksza, 2018), which is reasonable for retrospective redocking but biases the model

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

toward a fixed region in prospective use, where the holo pocket residues are unknown. Pocket-free methods such as DiffDock (Corso et al., 2023) and FlowDock (Morehead & Cheng, 2025) avoid explicit pocket inputs but do not explicitly model receptor-intrinsic pocket geometry to guide ligand motion, leaving open the challenge of inferring pocket-like directions while jointly predicting ligand and receptor conformational changes.

To address these gaps, we introduce **PIGEON**: **P**ocket-**I**nferred **G**eometric **E**nsemble for Flexible D**o**cki**n**g, a manifold flow-matching framework for ensemble flexible docking without ground-truth pocket residues. Given an apo-like receptor-ligand state, PIGEON predicts residual motion toward holo ensemble states across ligand translation, rotation, torsions, receptor residue frames, and side-chain torsions. The key idea is to replace unconstrained ligand translation regression with a pocket-inferred geometric decision process: receptor-intrinsic buriedness and a ligand-centered spherical harmonic pocket field summarize local receptor geometry, from which we derive a compact bank of pocket-aware candidate translation directions. A learnable staged selector then chooses among these candidates using atom-aware probe features measuring contact quality, clash avoidance, void avoidance, contact uniformity, and contact count. The selected direction is combined with a learned positive magnitude, while separate heads predict ligand rotation, torsions, and receptor flexibility.

Our main contributions are:

1. We formulate ensemble flexible docking as manifold residual prediction over multiple apo-holo state pairs, allowing the model to learn multiple valid holo conformations for the same receptor-ligand system.

2. We introduce a pocket-inferred geometric translation module that derives ligand motion cues from receptor-intrinsic buriedness, ligand-centered spherical harmonic fields, pocket-potential gradients, and long-range buriedness directions without using true pocket residue indices.

3. We design a learnable staged direction selector that preserves the interpretability of geometric docking rules while learning how to balance pocket-ingress and local-fitting behavior from data.

2 Problem Formulation

2.1 Flexible Docking

We consider local flexible docking for protein-ligand complexes, where both the ligand pose and the receptor conformation may change upon binding. Let the receptor contain N residues and let the ligand contain A atoms. For receptor residue i, we represent its local backbone frame by a rotation and frame origin $(R_i^{\mathrm{rec}}, \mathbf{p}_i^{\mathrm{rec}}) \in SO(3) \times \mathbb{R}^3$, and its side-chain torsions by $\boldsymbol{\chi}_i^{\mathrm{rec}} \in \mathbb{T}^{K_i}$, where $\mathbb{T}$ denotes the angular torus. For the ligand, we represent the rigid pose by a centroid $\mathbf{c}^{\mathrm{lig}} \in \mathbb{R}^3$ and rotation $R^{\mathrm{lig}} \in SO(3)$, and its internal conformation by rotatable-bond torsions $\boldsymbol{\chi}^{\mathrm{lig}} \in \mathbb{T}^{K_L}$. Since ligand local coordinates are centered at $\mathbf{c}^{\mathrm{lig}}$, this centroid serves as the ligand rigid translation coordinate. The complete flexible docking state lies on the product manifold

$$\mathcal{M} = \left(\prod_{i=1}^{N} \left(SO(3) \times \mathbb{R}^3 \times \mathbb{T}^{K_i} \right) \right) \times \mathbb{R}^3 \times SO(3) \times \mathbb{T}^{K_L}.$$

Each receptor-ligand pair may admit multiple apo and holo states, so the model learns a distribution of flexible docking transitions rather than a deterministic apo-to-holo mapping, without access to ground-truth pocket residue indices.

2.2 Manifold Flow Matching

We formulate flexible docking as flow matching on $\mathcal{M}$ (Lipman et al., 2022; 2024). For each apo-holo state pair (x_0, x_1) and scalar schedule $\alpha(t)$, the geodesic interpolant is

$$x_t = \mathrm{Exp}_{x_0}\left(\alpha(t)\,\mathrm{Log}_{x_0}(x_1)\right),$$

applied component-wise: linear interpolation in $\mathbb{R}^3$ for frame origins and the ligand centroid, $SO(3)$ geodesics for frame and ligand rotations, and wrapped angular interpolation for receptor and ligand torsions. For the ligand rigid pose, this gives

$$\mathbf{c}_t^{\mathrm{lig}} = (1-\alpha(t))\mathbf{c}_0^{\mathrm{lig}} + \alpha(t)\mathbf{c}_1^{\mathrm{lig}},$$

$$R_t^{\mathrm{lig}} = R_0^{\mathrm{lig}} \exp\Big(\alpha(t) \log\left((R_0^{\mathrm{lig}})^{-1} R_1^{\mathrm{lig}}\right)\Big).$$

We use the polynomial schedule $\alpha(t) = t^2$.

The standard flow-matching objective trains a vector field $v_\theta(x_t, t) \in T_{x_t}\mathcal{M}$ to match the target velocity $u_t^* = \frac{d}{dt}x_t$:

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}_{x_0,x_1,t}\left[\|v_\theta(x_t,t) - u_t^*\|^2_{T_{x_t}\mathcal{M}}\right].$$

For the geodesic interpolant, the path velocity is proportional to the endpoint residual $r_t^* = \mathrm{Log}_{x_t}(x_1)$ along each component, so we equivalently predict residuals from x_t to x_1:

$$\mathcal{L}_{\mathrm{res}} = \mathbb{E}_{x_0,x_1,t}\left[\left\|\widehat{r}_\theta(x_t,t) - \mathrm{Log}_{x_t}(x_1)\right\|^2_{T_{x_t}\mathcal{M}}\right].$$

Concretely, the ligand translation residual is $\Delta\mathbf{c}_{t\to1}^{\mathrm{lig},*} = \mathbf{c}_1^{\mathrm{lig}} - \mathbf{c}_t^{\mathrm{lig}}$, and the ligand rotation residual is $\boldsymbol{\omega}_{t\to1}^{\mathrm{lig},*} = \log\left((R_t^{\mathrm{lig}})^{-1}R_1^{\mathrm{lig}}\right)$. This residual formulation is equivalent to velocity matching up to a known time-dependent scaling factor.

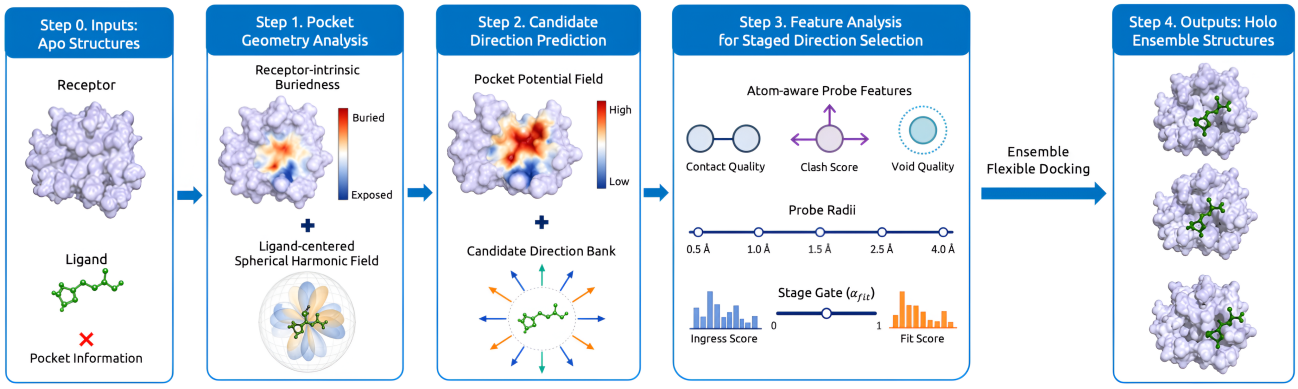

Figure 1. **Overview of PIGEON.** PIGEON performs pocket-unknown ensemble flexible docking by inferring receptor pocket geometry, constructing pocket-aware candidate translation directions, selecting a direction with a learnable staged selector, and predicting ligand and receptor residual motions.

3 PIGEON: Pocket-Inferred Geometric Ensemble Flexible Docking

3.1 Heterogeneous Complex Representation

We represent each interpolated state x_t as a heterogeneous graph spanning three structural resolutions: ligand atoms, receptor atoms, and receptor residues. Ligand-atom nodes encode atom type, chemical features, current coordinates, and time; receptor-atom nodes encode atom type, van der Waals radius, current coordinates, and time; receptor-residue nodes encode residue features, current frame, side-chain torsions, and time. The graph contains ligand bond edges, receptor atom-radius edges, receptor residue-radius edges, receptor atom-to-residue edges, and ligand-receptor cross edges. A time-conditioned message-passing network with FiLM-style time modulation and edge-type-specific projections produces final embeddings $\mathbf{h}_j^{\text{lig}}$, $\mathbf{h}_a^{\text{rec-atom}}$, and $\mathbf{h}_i^{\text{rec-res}}$, used by all downstream output heads (Perez et al., 2018). Architecture details are provided in the Appendix.

3.2 Receptor-Intrinsic Buriedness

Atoms lining a cavity are more geometrically enclosed than solvent-exposed atoms. For receptor atom a with position $\mathbf{q}_a$ and van der Waals radius r_a^{rec}, we cast K_{probe} probe directions $\{\mathbf{u}_k\}_{k=1}^{K_{\text{probe}}}$ and mark a direction occluded if another receptor atom lies within the probe radius and subtends it. With $\mathbf{d}_{ab} = \mathbf{q}_b - \mathbf{q}_a$,

$$o_{ak} = \mathbb{1}\left[\exists b : 0 < \|\mathbf{d}_{ab}\| < r_{\text{probe}} \wedge \frac{\mathbf{u}_k^\top \mathbf{d}_{ab}}{\|\mathbf{d}_{ab}\|} > \cos\left(\arcsin\frac{r_a^{\text{rec}} + r_b^{\text{rec}}}{\|\mathbf{d}_{ab}\| + \epsilon}\right)\right]. \tag{1}$$

The buriedness of atom a is

$$b_a = \frac{1}{K_{\text{probe}}} \sum_{k=1}^{K_{\text{probe}}} o_{ak}.$$

Buriedness depends only on receptor geometry, so it requires no pocket labels at inference.

3.3 Ligand-Centered Spherical Harmonic Pocket Field

To capture pocket geometry local to the current ligand pose, we construct a spherical harmonic field around $\mathbf{c}_t^{\text{lig}}$. For each receptor atom a,

$$\mathbf{r}_{a,t} = \mathbf{q}_{a,t} - \mathbf{c}_t^{\text{lig}}, \quad r_{a,t} = \|\mathbf{r}_{a,t}\|, \quad \hat{\mathbf{r}}_{a,t} = \mathbf{r}_{a,t}/(r_{a,t}+\epsilon).$$

We use Gaussian radial shells with centers $\{R_s\}_{s=1}^M$ and widths $\{\sigma_s\}_{s=1}^M$:

$$\phi_s(r_{a,t}) = \exp\left(-\frac{(r_{a,t} - R_s)^2}{2\sigma_s^2}\right),$$

with $M = 6$ shells:

$$R = (1.5, 2.5, 4.0, 8.0, 14.0, 20.0)\ \text{Å},$$
$$\sigma = (0.75, 1.0, 2.0, 3.0, 4.0, 5.0)\ \text{Å}.$$

Per shell, we compute buriedness-weighted real spherical harmonic coefficients up to order $L = 2$:

$$\mathbf{f}_\ell^{(s)}(\mathbf{c}_t^{\text{lig}}) = \sum_a b_a \phi_s(r_{a,t}) \mathbf{Y}_\ell(\hat{\mathbf{r}}_{a,t}), \quad \ell = 0, 1, 2,$$

with $L = 0$ summarizing buried mass at each radius, $L = 1$ giving a dipole direction toward asymmetric buried mass, and $L = 2$ capturing anisotropic pocket shape. The full representation is

$$\mathbf{F}_{\text{SH}}(\mathbf{c}_t^{\text{lig}}) = \bigoplus_{s=1}^{M} \bigoplus_{\ell=0}^{L} \mathbf{f}_\ell^{(s)}(\mathbf{c}_t^{\text{lig}}).$$

3.4 Pocket-Aware Translation Candidate Bank

Rather than regressing a free 3D direction, we define a compact bank of receptor-derived candidates. The first group comes from the six shellwise $L = 1$ dipoles. Letting $\mathbf{d}_s^{\text{SH}}$ denote the 3D vector associated with the $L = 1$ coefficients in shell s,

$$\mathbf{v}_s^{\text{SH}} = \mathbf{d}_s^{\text{SH}}/(\|\mathbf{d}_s^{\text{SH}}\| + \epsilon).$$

The second group comes from multi-scale pocket-potential gradients. For a preferred pocket distance r_{opt},

$$U_{r_{\mathrm{opt}}}(\mathbf{c}_t^{\mathrm{lig}}) = \sum_a b_a \exp\left(-\frac{(\|\mathbf{q}_{a,t} - \mathbf{c}_t^{\mathrm{lig}}\| - r_{\mathrm{opt}})^2}{2\sigma_{\mathrm{pot}}^2}\right),$$

$$\mathbf{v}_{r_{\mathrm{opt}}}^{\mathrm{pot}} = \nabla_{\mathbf{c}_t^{\mathrm{lig}}} U_{r_{\mathrm{opt}}}(\mathbf{c}_t^{\mathrm{lig}}) / (\|\nabla_{\mathbf{c}_t^{\mathrm{lig}}} U_{r_{\mathrm{opt}}}(\mathbf{c}_t^{\mathrm{lig}})\| + \epsilon).$$

We use five potential scales, giving five gradient candidates. We add two long-range buriedness directions: one toward the global buriedness-weighted receptor centroid,

$$\bar{\mathbf{q}}_{\mathrm{buried}} = \frac{\sum_a b_a \mathbf{q}_{a,t}}{\sum_a b_a + \epsilon},$$

$$\mathbf{v}_{\mathrm{centroid}} = \frac{\bar{\mathbf{q}}_{\mathrm{buried}} - \mathbf{c}_t^{\mathrm{lig}}}{\|\bar{\mathbf{q}}_{\mathrm{buried}} - \mathbf{c}_t^{\mathrm{lig}}\| + \epsilon},$$

and one toward the nearest highly buried receptor atom,

$$a^\star = \arg\min_{a: b_a > \tau_b} \|\mathbf{q}_{a,t} - \mathbf{c}_t^{\mathrm{lig}}\|,$$

$$\mathbf{v}_{\mathrm{nearest}} = \frac{\mathbf{q}_{a^\star,t} - \mathbf{c}_t^{\mathrm{lig}}}{\|\mathbf{q}_{a^\star,t} - \mathbf{c}_t^{\mathrm{lig}}\| + \epsilon}.$$

The 13-direction candidate bank is

$$\mathcal{V}_t = \{\mathbf{v}_1, \ldots, \mathbf{v}_{13}\}, \quad \|\mathbf{v}_k\| = 1.$$

3.5 Learnable Direction Selector

Different docking states call for different motions: when the ligand is outside the pocket, an *ingress* move toward buried receptor regions is preferred; when the ligand is already near a pocket, a *fit* move that improves local packing is preferred. We use a learnable staged selector that interpolates between an ingress and a fit scoring rule, learning only the coefficients controlling feature importance, probe-distance importance, and the ingress-fit transition.

Probe geometry. For each candidate $\mathbf{v}_k$ and probe distance ρ_s, the hypothetical translated centroid and ligand atoms are

$$\mathbf{p}_{k,s} = \mathbf{c}_t^{\mathrm{lig}} + \rho_s \mathbf{v}_k, \quad \widetilde{\mathbf{x}}_{j,k,s} = \mathbf{x}_{j,t} + \rho_s \mathbf{v}_k,$$

with probe schedule $\mathcal{R} = \{0.5, 1.0, 1.5, 2.5, 4.0\}$ Å. Letting $r_{\mathrm{lig}} = \max_j \|\widetilde{\mathbf{x}}_{j,k,s} - \mathbf{p}_{k,s}\|$, the local receptor atom set is

$$\mathcal{A}_{k,s} = \{a : \|\mathbf{q}_{a,t} - \mathbf{p}_{k,s}\| < r_{\mathrm{probe}} + r_{\mathrm{lig}} + \gamma_{\mathrm{out}} + 1\}.$$

For $a \in \mathcal{A}_{k,s}$ and ligand atom j, the surface gap is

$$g_{aj}^{(k,s)} = \|\mathbf{q}_{a,t} - \widetilde{\mathbf{x}}_{j,k,s}\| - (r_a^{\mathrm{rec}} + r_j^{\mathrm{lig}}),$$

with closest-gap reductions

$$g_{a,\min}^{(k,s)} = \min_j g_{aj}^{(k,s)}, \quad h_{j,\min}^{(k,s)} = \min_{a \in \mathcal{A}_{k,s}} g_{aj}^{(k,s)}.$$

Probe features. We define five probe features. With contact width $\sigma_{\mathrm{contact}}$ and preferred contact gap γ_{pref},

$$C_a^{(k,s)} = \exp\left(-\frac{\left(g_{a,\min}^{(k,s)} - \gamma_{\mathrm{pref}}\right)^2}{2\sigma_{\mathrm{contact}}^2}\right),$$

$$F_{\mathrm{contact}}(k,s) = \log\left(1 + \sum_{a \in \mathcal{A}_{k,s}} b_a C_a^{(k,s)}\right),$$

$$P_{\mathrm{clash}}(k,s) = \sum_{a \in \mathcal{A}_{k,s}} b_a \,\mathrm{ReLU}\left(-g_{a,\min}^{(k,s)}\right),$$

$$F_{\mathrm{clash}}(k,s) = -\log\left(1 + P_{\mathrm{clash}}(k,s)\right),$$

$$P_{\mathrm{void}}(k,s) = \frac{1}{A}\sum_{j=1}^{A} \mathrm{ReLU}\left(h_{j,\min}^{(k,s)} - \gamma_{\mathrm{void}}\right),$$

$$F_{\mathrm{void}}(k,s) = -\log\left(1 + P_{\mathrm{void}}(k,s)\right).$$

With contact-shell set

$$\mathcal{C}_{k,s} = \left\{a \in \mathcal{A}_{k,s} : \gamma_{\mathrm{in}} < g_{a,\min}^{(k,s)} < \gamma_{\mathrm{out}} \text{ and } b_a > \tau_b\right\},$$

the contact-count feature is $F_{\mathrm{count}}(k,s) = \log(1 + |\mathcal{C}_{k,s}|)$. For uniformity, define

$$\mathbf{u}_{a,k,s} = \frac{\mathbf{q}_{a,t} - \mathbf{p}_{k,s}}{\|\mathbf{q}_{a,t} - \mathbf{p}_{k,s}\| + \epsilon},$$

and, for $|\mathcal{C}_{k,s}| \geq 3$,

$$F_{\mathrm{uniform}}(k,s) = 1 - \left\|\frac{1}{|\mathcal{C}_{k,s}|}\sum_{a \in \mathcal{C}_{k,s}} \mathbf{u}_{a,k,s}\right\|,$$

with $F_{\mathrm{uniform}} = 0$ when $|\mathcal{C}_{k,s}| < 3$. Together, F_{contact} rewards good buried contacts, F_{clash} penalizes overlap, F_{void} penalizes empty space, F_{count} counts plausible buried contacts, and F_{uniform} rewards balanced contact directions.

We collect features in fixed order:

$$\mathbf{F}(k,s) = \left[F_{\mathrm{contact}}, F_{\mathrm{clash}}, F_{\mathrm{void}}, F_{\mathrm{uniform}}, F_{\mathrm{count}}\right](k,s),$$

and within-sample normalize across the $K = 13$ candidates:

$$\mu_{s,f} = \frac{1}{K}\sum_{k=1}^{K} F_f(k,s),$$

$$\sigma_{s,f} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(F_f(k,s) - \mu_{s,f})^2 + \epsilon},$$

$$Z(k,s,f) = \frac{F_f(k,s) - \mu_{s,f}}{\sigma_{s,f}}.$$

Stage gate. We summarize the current pose with a stage-feature vector $\mathbf{z}_{\text{stage}} = [z_d, z_q, z_c, z_u, z_n]$, where z_d is a smoothed pocket-distance signal and z_q, z_c, z_u, z_n are bounded versions of $F_{\text{contact}}, P_{\text{clash}}, F_{\text{uniform}}, F_{\text{count}}$ evaluated at the untranslated ligand position (full definitions in the Appendix B.9). The learnable fit gate is

$$\alpha_{\text{fit}} = \sigma\left(s_{\text{gate}}(\mathbf{w}_{\text{gate}}^{\top}\mathbf{z}_{\text{stage}} + b_{\text{gate}})\right),$$

where $\mathbf{w}_{\text{gate}}, b_{\text{gate}}, s_{\text{gate}}$ are initialized from the predefined rule and learned.

Ingress and fit scores. Both rules use $Z(k, s, f)$ with separate learnable feature logits $\boldsymbol{\eta}^{\text{ing}}, \boldsymbol{\eta}^{\text{fit}}$ and probe-distance logits $\boldsymbol{\zeta}^{\text{ing}}, \boldsymbol{\zeta}^{\text{fit}}$:

$$\boldsymbol{\lambda}^{\text{ing}} = \text{softmax}(\boldsymbol{\eta}^{\text{ing}}), \quad \boldsymbol{\lambda}^{\text{fit}} = \text{softmax}(\boldsymbol{\eta}^{\text{fit}}),$$
$$\boldsymbol{\rho}^{\text{ing}} = \text{softmax}(\boldsymbol{\zeta}^{\text{ing}}), \quad \boldsymbol{\rho}^{\text{fit}} = \text{softmax}(\boldsymbol{\zeta}^{\text{fit}}).$$

$$s_k^{\text{ing}} = \sum_s \rho_s^{\text{ing}} \sum_f \lambda_f^{\text{ing}} Z(k, s, f),$$
$$s_k^{\text{fit}} = \sum_s \rho_s^{\text{fit}} \sum_f \lambda_f^{\text{fit}} Z(k, s, f).$$

The stage-weighted selector logit and selected direction are

$$\ell_k = (1-\alpha_{\text{fit}})s_k^{\text{ing}} + \alpha_{\text{fit}} s_k^{\text{fit}}, \quad k^{\star} = \arg\max_k \ell_k, \quad \widehat{\mathbf{u}} = \mathbf{v}_{k^{\star}}.$$

3.6 Other Prediction Heads

A magnitude head consumes the pooled graph context, the time embedding, and invariant pocket-field summaries, and outputs a positive scalar through a Softplus readout; the final translation prediction is the selected unit direction times this magnitude. Additional heads predict the ligand $SO(3)$ log residual, wrapped ligand torsion residuals, per-residue receptor-frame residuals, and per-residue side-chain torsion residuals, all consuming the embeddings defined above. Architectural details are in the Appendix.

3.7 Training Objective

For direction selection, the normalized ground-truth direction is

$$\mathbf{u}^* = \Delta\mathbf{c}_{t\to 1}^{\text{lig},*} / (\|\Delta\mathbf{c}_{t\to 1}^{\text{lig},*}\| + \epsilon),$$

with oracle cosines $q_k^* = \mathbf{v}_k^{\top}\mathbf{u}^*$ and soft oracle distribution

$$p_k^* = \frac{\exp(q_k^*/T_{\text{oracle}})}{\sum_{j=1}^K \exp(q_j^*/T_{\text{oracle}})}.$$

The selector is trained with the listwise rank loss

$$\mathcal{L}_{\text{rank}} = -\sum_{k=1}^K p_k^* \log \frac{\exp(\ell_k)}{\sum_{j=1}^K \exp(\ell_j)},$$

Table 1. Ensemble flexible docking performance. We compare PIGEON against representative generative docking baselines using ligand RMSD, all-atom RMSD, PoseBusters validity, joint PoseBusters-valid and ligand-RMSD success rate, and runtime. Higher is better for percentage metrics marked with ↑, and lower is better for median RMSD and runtime.

Method	Ligand RMSD		All-Atom RMSD	% PB valid and	Runtime (s)
	% < 2 Å ↑	Median Å ↓	% < 1 Å ↑	L-RMSD < 2 Å ↑	
TankBind	2.39	7.14	30.39	1.05	1.77
FABFlex	13.95	3.78	40.05	0.40	0.98
FlowDock	26.28	3.17	7.46	14.83	29.4
DiffDock (rigid)	7.01	3.76	N.A.	0.99	**0.31**
FlexDock	**51.61**	**1.95**	6.07	**43.78**	11.68
PIGEON (1-step)	33.89	2.48	**49.21**	12.99	1.49
PIGEON (4-step)	36.36	2.40	**44.18**	16.52	4.78

Table 2. **Ensemble docking coverage and assignment success.** We evaluate whether each method can recover multiple holo ensemble conformations for the same receptor-ligand pair. GT coverage@1Å all-atom measures the fraction of ground-truth holo states covered by at least one prediction within 1Å all-atom RMSD. Hungarian success@1Å all-atom measures assignment-based ensemble recovery after optimal matching between predictions and ground-truth holo states. Higher is better for both metrics.

Method	Mean GT coverage@1Å all-atom ↑	Mean Hungarian success@1Å all-atom ↑
TankBind	32.69	30.59
FABFlex	39.08	40.09
FlowDock	17.71	16.92
DiffDock (rigid)	N.A.	N.A.
FlexDock	11.57	7.33
PIGEON (1-step)	36.56	**49.40**
PIGEON (4-step)	**39.49**	44.35

regularized toward the rule-based initialization θ_0:

$$\mathcal{L}_{\text{reg}} = \|\theta - \theta_0\|_2^2.$$

Magnitude is supervised in log space:

$$\mathcal{L}_{\text{mag}} = \left(\log(\widehat{m} + \epsilon) - \log(\|\Delta\mathbf{c}_{t\to 1}^{\text{lig},*}\| + \epsilon)\right)^2.$$

Non-translation outputs use fixed-scale normalized residual losses; for prediction $\widehat{\mathbf{y}}$, target $\mathbf{y}^*$, scale s,

$$\mathcal{L}_{\text{res}}(\widehat{\mathbf{y}}, \mathbf{y}^*; s) = \text{SmoothL1}(\widehat{\mathbf{y}}/s, \mathbf{y}^*/s).$$

The full objective is

$$\mathcal{L} = \lambda_{\text{rank}}\mathcal{L}_{\text{rank}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}} + \lambda_{\text{mag}}\mathcal{L}_{\text{mag}} + \lambda_{\text{frame}}\mathcal{L}_{\text{frame}} + \lambda_{\chi}\mathcal{L}_{\chi} + \lambda_{\text{rot}}\mathcal{L}_{\text{rot}} + \lambda_{\text{tor}}\mathcal{L}_{\text{tor}}.$$

4 Results

We evaluate PIGEON on an ensemble flexible docking benchmark built from apo structures from FlexDock (Corso et al., 2025) and their corresponding holo molecular-dynamics ensembles from MISATO (Siebenmorgen et al., 2024). The apo structures define the local docking inputs,

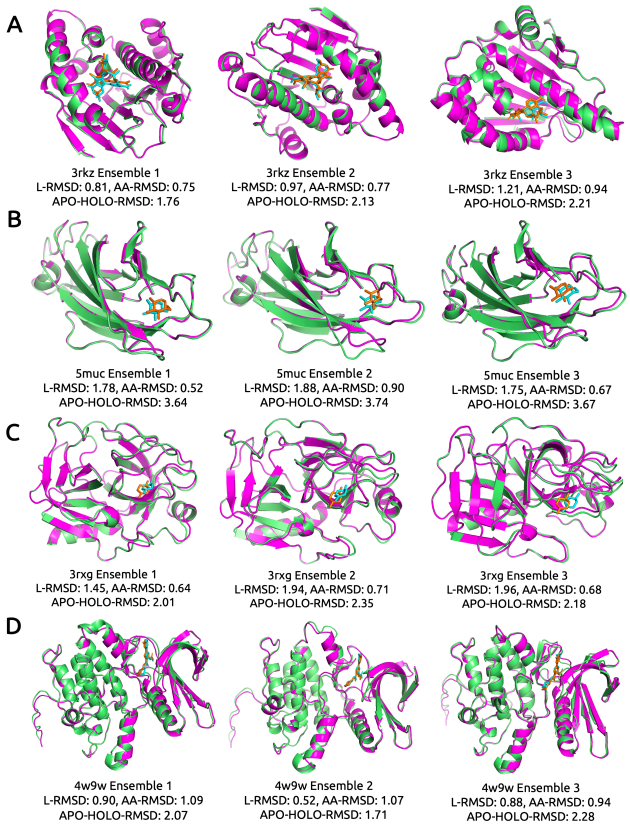

Figure 2. **Visualized ensemble flexible docking predictions.** We show representative ensemble predictions for four receptor-ligand complexes from the test set: (A) 3rkz, (B) 5muc, (C) 3rxg, and (D) 4w9w. For each complex, three holo ensemble targets are shown to illustrate that PIGEON can generate multiple flexible docking conformations for the same receptor-ligand pair. L-RMSD and AA-RMSD denote ligand RMSD and all-atom complex RMSD, respectively, after aligning the predicted and ground-truth complexes. APO-HOLO-RMSD denotes the ligand RMSD between the ligand pose in the ground-truth apo structure and the ligand pose in the corresponding ground-truth holo structure. Color code: holo protein in lime, predicted protein in magenta, true ligand in cyan, and predicted ligand in orange.

with frame-specific ligand augmentations around the binding region, while representative MISATO frames provide multiple holo targets for each receptor-ligand pair. This benchmark therefore tests whether a model can recover both ligand poses and the corresponding receptor conformations across holo ensemble states. Dataset construction details are provided in Appendix A.

4.1 Docking Accuracy, Ensemble Recovery, and Efficiency

PIGEON achieves strong performance under the pocket-unknown ensemble flexible docking setting (Table 1). Among methods that do not require explicit pocket input, PIGEON performs favorably on both ligand-level and full-complex metrics. Compared with TankBind (Lu et al., 2022), FABFlex (Zhang et al., 2025), and FlowDock (Morehead & Cheng, 2025), PIGEON improves ligand placement while also achieving substantially better all-atom complex recovery. This is important because ligand RMSD measures only pose accuracy, whereas all-atom RMSD also reflects whether the predicted receptor conformation is consistent with the holo complex. The improvement on both metrics suggests that PIGEON's pocket-inferred geometric representation and coupled residual prediction heads help recover coordinated ligand-receptor holo geometry, rather than only placing the ligand near a plausible local region.

FlexDock remains the strongest baseline for ligand RMSD success, which is expected because it receives explicit pocket information. In contrast, PIGEON must infer pocket-like directions from receptor geometry and the current ligand state. The remaining ligand-RMSD gap to FlexDock should therefore be interpreted in light of this harder input setting. However, ligand RMSD alone does not measure whether the model recovers the holo receptor conformation associated with each ensemble state. In our benchmark, all-atom RMSD is evaluated against frame-specific MISATO holo complexes, so a method can place the ligand accurately while still failing to reproduce the corresponding receptor geometry. Under this stricter full-complex criterion, PIGEON outperforms FlexDock, showing that it is better suited to recovering frame-specific coupled ligand-receptor holo conformations rather than only producing accurate pocket-localized ligand poses.

This distinction becomes clearer in the ensemble recovery metrics (Table 2). FlexDock performs well as a pocket-conditioned single-pose docking model, but its coverage remains limited because accurate docking to a supplied pocket does not necessarily produce diverse predictions matching different holo receptor-ligand conformations. Some pocket-free baselines obtain higher coverage than FlexDock, likely because their predictions vary more across runs or inputs; however, this variability is not always accompanied by accurate ligand placement or full-complex recovery. PIGEON achieves the strongest ensemble recovery because it combines pocket-inferred geometric conditioning with supervision over multiple apo-holo transitions per receptor-ligand pair, allowing it to cover distinct holo ensemble members while maintaining accurate coupled ligand-receptor geometry.

The Hungarian assignment metric provides a stricter test of ensemble generation than coverage alone (Kuhn, 1955). A model can obtain reasonable coverage by producing many predictions near a subset of easy holo states, but high Hungarian success requires one-to-one matching between predicted and ground-truth ensemble conformations. PIGEON performs strongly on this metric, showing that it does not merely generate repeated variants of a single bound state.

Table 3. Input-to-holo structural gap on the test set. Q1 and Q3 denote the 25th and 75th percentiles, respectively, and P90 denotes the 90th percentile. All values are in Å.

Metric	Mean	Median	Q1	Q3	P90
Ligand apo–holo gap					
Ligand RMSD, $x_0 \to x_1$	3.318	2.813	1.937	3.992	5.139
Ligand centroid displacement	2.673	2.230	1.539	3.328	3.980
Whole-receptor apo–holo gap					
Receptor CA RMSD	0.720	0.633	0.502	0.810	1.079
Receptor mapped-atom RMSD	0.890	0.803	0.651	1.002	1.283
Local pocket and active-site flexibility					
Pocket CA RMSD	1.112	0.974	0.763	1.280	1.713
Pocket mapped-atom RMSD	1.366	1.226	1.010	1.533	1.982
Active-site CA RMSD	1.098	0.962	0.771	1.245	1.643
Active-site mapped-atom RMSD	1.363	1.226	1.030	1.510	1.920
Aggregate complex gap					
Complex all-atom RMSD, $x_0 \to x_1$	1.013	0.893	0.727	1.125	1.503

Table 4. **Effect of frame-specific ligand augmentation.** Statistics are computed over 7,300 frame pairs after receptor alignment. x_0 denotes the unaugmented apo ligand pose, x_0^{aug} denotes the augmented ligand pose used as model input, and Δ denotes the per-frame change in each metric induced by augmentation. Q1 and Q3 are the 25th and 75th percentiles, and P90 is the 90th percentile. All values are in Å.

Metric	Input	Mean	Median	Q1	Q3	P90
Ligand RMSD	$x_0 \to x_1$	2.524	1.727	1.065	2.994	5.012
	$x_0^{\text{aug}} \to x_1$	3.318	2.813	1.937	3.992	5.139
	Δ	0.794	0.511	0.000	1.304	2.294
Ligand centroid	$x_0 \to x_1$	1.804	1.107	0.640	2.044	3.642
	$x_0^{\text{aug}} \to x_1$	2.673	2.230	1.539	3.328	3.980
	Δ	0.869	0.631	0.000	1.430	2.486
Complex all-atom RMSD	$x_0 \to x_1$	0.972	0.851	0.691	1.081	1.437
	$x_0^{\text{aug}} \to x_1$	1.013	0.893	0.727	1.125	1.503
	Δ	0.041	0.011	0.000	0.037	0.119

Instead, its predictions are distributed across multiple holo endpoints, consistent with the goal of modeling ensemble flexible docking as a conditional distribution over coupled ligand-receptor conformations.

PIGEON also maintains competitive physical validity. Its PoseBusters-valid rate is comparable to other methods that do not use explicit pocket input, indicating that the structured translation and flexible receptor prediction do not come at the cost of severe geometric invalidity (Buttenschoen et al., 2024). The joint PoseBusters-valid and ligand-RMSD success metric remains more challenging, because it requires both chemically valid ligand geometry and accurate ligand placement. FlexDock performs strongly on this metric, likely benefiting from explicit pocket conditioning and stronger pose-level refinement. This suggests a clear direction for improving PIGEON: adding a lightweight local relaxation module, stronger stereochemical regularization, or validity-aware losses could improve chemically strict ligand success while preserving the model's ensemble recovery advantages.

Finally, PIGEON is efficient at inference time. Even one-step generation already gives strong docking and ensemble recovery performance, while four-step generation improves ligand placement and coverage-related metrics with only a modest runtime increase (Table 1). This efficiency comes from the factorized translation design: PIGEON selects from a compact bank of pocket-aware directions and predicts a scalar magnitude, rather than relying on expensive unconstrained pose sampling. Thus, PIGEON provides a practical tradeoff between speed, ligand accuracy, full-complex recovery, and ensemble diversity under the realistic setting where the true pocket is not provided.

4.2 Qualitative Analysis

The qualitative examples further illustrate PIGEON's ability to recover multiple holo-like conformations for the same receptor-ligand pair (Figure 2). Across 3rkz, 5muc, 3rxg, and 4w9w, the three predicted ensemble members are not identical copies of a single docked pose; instead, they follow different holo targets with distinct ligand placements and receptor conformations, which is consistent with the ensemble recovery metrics. The visualized examples also show that the task is not a trivial reconstruction of the input structure. The apo-holo ligand displacements are often substantial, and the test-set statistics confirm that the initial ligand and receptor states have a non-negligible gap to the corresponding holo targets (Table 3). Moreover, Table 4 shows that the frame-specific ligand augmentation in the dataset construction increases the ligand input-to-holo displacement while keeping the overall complex perturbation controlled. Thus, successful predictions require PIGEON to move the ligand away from its augmented apo-like initialization, adjust its orientation and internal torsions, and predict receptor conformational changes that remain consistent with the corresponding MISATO holo frame. In the examples, PIGEON frequently reduces the apo-holo ligand gap while maintaining low all-atom complex error, indicating that the model learns meaningful residual motions for both ligand and receptor components rather than merely preserving the starting pose.

5 Discussion

Here, we have presented PIGEON, a pocket-inferred manifold flow-matching framework for ensemble flexible docking under the realistic setting where the true binding pocket is not provided. By integrating receptor-intrinsic buriedness fields, ligand-centered spherical harmonic pocket representations, pocket-aware translation candidates, and coupled

residual prediction heads for ligand and receptor degrees of freedom, PIGEON generates holo-like receptor-ligand ensembles without relying on ground-truth pocket residues. Across our ensemble docking benchmark, PIGEON demonstrates strong full-complex recovery, competitive ligand placement among pocket-free methods, efficient few-step generation, and improved recovery of multiple holo conformations for the same receptor-ligand pair. These results highlight the value of treating flexible docking as coupled ensemble generation rather than as single-pose ligand placement, providing a practical route toward modeling receptor adaptation and conformational heterogeneity in structure-based drug discovery.

Looking forward, several directions could further strengthen pocket-unknown ensemble flexible docking. While PIGEON already recovers diverse and physically plausible holo complexes, incorporating lightweight local relaxation, stronger stereochemical or torsional regularization, and validity-aware training objectives may further improve chemically strict ligand recovery. In addition, uncertainty-aware sampling and learned ranking could help prioritize the most relevant generated ensemble states for downstream virtual screening and prospective design. Extending the framework to broader blind-docking settings, larger receptor ensembles, and experimental validation would further test its generality. Nevertheless, PIGEON lays a solid foundation for pocket-unknown ensemble flexible docking, showing that receptor-inferred geometric cues can support efficient generation of diverse coupled ligand-receptor holo conformations without requiring explicit pocket inputs.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning, specifically generative modeling for structure-based drug discovery. PIGEON predicts protein-ligand binding geometries and does not generate novel chemical entities, synthesis routes, or activity predictions. The societal consequences of advancing flexible docking methods are largely those associated with computational drug discovery as a whole, none of which we feel must be specifically highlighted here.

References

Amaro, R. E., Baron, R., and McCammon, J. A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of Computer-Aided Molecular Design*, 22(9):693–705, 2008. doi: 10.1007/s10822-007-9159-2.

Amaro, R. E., Baudry, J., Chodera, J., Demir, Ö., McCammon, J. A., Miao, Y., and Smith, J. C. Ensemble docking in drug discovery. *Biophysical Journal*, 114(10):2271–2278, 2018. doi: 10.1016/j.bpj.2018.02.038.

Bottegoni, G., Rocchia, W., Rueda, M., Abagyan, R., and Cavalli, A. Systematic exploitation of multiple receptor conformations for virtual ligand screening. *PLOS ONE*, 6(5):e18845, 2011. doi: 10.1371/journal.pone.0018845.

Buttenschoen, M., Morris, G. M., and Deane, C. M. Posebusters: Ai-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15(9):3130–3139, 2024.

Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T. DiffDock: Diffusion steps, twists, and turns for molecular docking. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023. URL https://arxiv.org/abs/2210.01776.

Corso, G., Somnath, V. R., Getz, N., Barzilay, R., Jaakkola, T., and Krause, A. Composing unbalanced flows for flexible docking and relaxation. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025. URL https://openreview.net/forum?id=gHLWTzKiZV.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, 47 (7):1739–1749, 2004. doi: 10.1021/jm0306430.

Heaslet, H., Rosenfeld, R., Giffin, M., Lin, Y.-C., Tam, K., Torbett, B. E., Elder, J. H., McRee, D. E., and Stout, C. D. Conformational flexibility in the flap domains of ligand-free HIV protease. *Acta Crystallographica Section D: Biological Crystallography*, 63(8):866–875, 2007. doi: 10.1107/S0907444907029125.

Huang, Y., Zhang, O., Wu, L., Tan, C., Lin, H., Gao, Z., Li, S., and Li, S. Z. Re-Dock: Towards flexible and realistic molecular docking with diffusion bridge. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, pp. 20474–20489. PMLR, 2024.

Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Foundations of Crystallography*, 32(5): 922–923, 1976.

Krivák, R. and Hoksza, D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of Cheminformatics*, 10(1):39, 2018. doi: 10.1186/s13321-018-0285-8.

Kuhn, H. W. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

Le Guilloux, V., Schmidtke, P., and Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, 10(1):168, 2009. doi: 10.1186/1471-2105-10-168.

Lin, J.-H., Perryman, A. L., Schames, J. R., and McCammon, J. A. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *Journal of the American Chemical Society*, 124(20):5632–5633, 2002. doi: 10.1021/ja0260162.

Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Lipman, Y., Havasi, M., Holderrieth, P., Shaul, N., Le, M., Karrer, B., Chen, R. T., Lopez-Paz, D., Ben-Hamu, H., and Gat, I. Flow matching guide and code. *arXiv preprint arXiv:2412.06264*, 2024.

Lu, W., Wu, Q., Zhang, J., Rao, J., Li, C., and Zheng, S. Tankbind: Trigonometry-aware neural networks for drug-protein binding structure prediction. *Advances in neural information processing systems*, 35:7236–7249, 2022.

Lu, W., Zhang, J., Huang, W., Zhang, Z., Jia, X., Wang, Z., Shi, L., Li, C., Wolynes, P. G., and Zheng, S. DynamicBind: predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nature Communications*, 15(1):1071, 2024. doi: 10.1038/s41467-024-45461-2.

Morehead, A. and Cheng, J. FlowDock: Geometric flow matching for generative protein–ligand docking and affinity prediction. *Bioinformatics*, 41(Supplement_1):i198–i206, 2025. doi: 10.1093/bioinformatics/btaf187.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16):2785–2791, 2009. doi: 10.1002/jcc.21256.

Perez, E., Strub, F., De Vries, H., Dumoulin, V., and Courville, A. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Plainer, M., Toth, M., Dobers, S., Stärk, H., Corso, G., Marquet, C., and Barzilay, R. DiffDock-Pocket: Diffusion for pocket-level docking with sidechain flexibility. In *NeurIPS 2023 Workshop on Machine Learning in Structural Biology (MLSB)*, 2023.

Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., and Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *Journal of Medicinal Chemistry*, 49 (2):534–553, 2006. doi: 10.1021/jm050540c.

Siebenmorgen, T., Menezes, F., Benassou, S., Merdivan, E., Didi, K., Mourão, A. S. D., Kitel, R., Liò, P., Kesselheim, S., Piraud, M., et al. Misato: machine learning dataset of protein–ligand complexes for structure-based drug discovery. *Nature computational science*, 4(5):367–378, 2024.

Stärk, H., Ganea, O.-E., Pattanaik, L., Barzilay, R., and Jaakkola, T. EquiBind: Geometric deep learning for drug binding structure prediction. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, volume 162 of *Proceedings of Machine Learning Research*, pp. 20503–20521. PMLR, 2022.

Trott, O. and Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, 2010. doi: 10.1002/jcc.21334.

Vijayan, R. S. K., He, P., Modi, V., Duong-Ly, K. C., Ma, H., Peterson, J. R., Dunbrack, Jr., R. L., and Levy, R. M. Conformational analysis of the DFG-out kinase motif and biochemical profiling of structurally validated type II inhibitors. *Journal of Medicinal Chemistry*, 58(1): 466–479, 2015. doi: 10.1021/jm501603h.

Wootten, D., Christopoulos, A., Marti-Solano, M., Babu, M. M., and Sexton, P. M. Mechanisms of signalling and biased agonism in G protein-coupled receptors. *Nature Reviews Molecular Cell Biology*, 19(10):638–653, 2018. doi: 10.1038/s41580-018-0049-3.

Zhang, Z., Wu, L., Gao, K., Yao, J., Qin, T., and Han, B. Fast and accurate blind flexible docking. *arXiv preprint arXiv:2502.14934*, 2025.

Appendix

A Dataset Preparation

We construct an ensemble flexible docking dataset by combining apo receptor-ligand complexes from FlexDock (Corso et al., 2025) with holo molecular-dynamics ensemble structures from MISATO (Siebenmorgen et al., 2024). The FlexDock data provide apo-like receptor structures and initial ligand information, while MISATO provides trajectory coordinates for the corresponding holo complexes. For each receptor-ligand pair, we treat the 100 MISATO trajectory frames as a set of possible holo ensemble states. Since training on all 100 frames for every complex is computationally expensive and contains redundant conformations, we select 5 representative holo frames for each PDB entry and use these frames as the ensemble targets during training.

Representative holo-frame selection. For each PDB entry, we first load the 100 MISATO trajectory frames and use the receptor and ligand atom mappings to extract the matched receptor and ligand atoms. To select representative ensemble states, we compute pairwise distances between trajectory frames using both receptor-pocket motion and ligand-pose variation. Specifically, for two frames, we align the receptor using a stable receptor subset and compute a weighted distance that combines receptor RMSD in the local pocket region and ligand RMSD after receptor alignment. The combined distance is defined as

$$D(i,j) = w_{\mathrm{rec}}\,\mathrm{RMSD}_{\mathrm{rec}}(i,j) + w_{\mathrm{lig}}\,\mathrm{RMSD}_{\mathrm{lig}}(i,j),$$

where $w_{\mathrm{rec}} = 0.4$ and $w_{\mathrm{lig}} = 0.6$ in our preprocessing. We then run k-medoids clustering with $k = 5$ on this distance matrix and select the five medoid frames as representative holo ensemble states. This procedure preserves diverse ligand placements and receptor conformational changes while avoiding the redundancy of using all trajectory frames.

Apo-gauge alignment of holo frames. For each selected holo frame, we align the MISATO holo coordinates into the apo coordinate system. This step is necessary because the model should learn the conformational transition from an apo-like input state to a holo target state in a consistent global frame. We identify a stable receptor core using mapped receptor atoms, excluding chain termini and optionally refining the core by receptor RMSF across sampled trajectory frames. A Kabsch alignment is then computed between the selected holo frame and the apo receptor using this stable core (Kabsch, 1976). The same rigid transformation is applied to both receptor and ligand holo coordinates, producing apo-gauge holo receptor coordinates P_1 and ligand coordinates L_1. We also save the corresponding apo receptor and ligand coordinates P_0 and L_0, together with alignment metrics and the selected core-anchor metadata.

Active-region and pocket labels for supervision and analysis. Although PIGEON does not receive true pocket residue indices as model input, we still compute pocket and active-region labels during preprocessing for supervision, filtering, and analysis. For each selected holo frame, we define pocket residues based on the minimum heavy-atom distance between receptor residues and the holo ligand. A residue is labeled as pocket if any of its heavy atoms lies within a fixed cutoff of the ligand heavy atoms. We then define an active region by expanding the pocket through local residue-neighborhood rules, including one-hop sequence or structural shells and distance-based residue-center expansion. The resulting masks identify residues that directly contact the ligand and residues in the nearby flexible binding region. These labels are saved as per-frame masks and are used only as training targets or diagnostics, not as inference-time pocket inputs.

Receptor flexible-state construction. For every selected holo frame, we convert receptor coordinates into residue-level flexible degrees of freedom. Each residue is represented by a local backbone frame computed from its N, C_α, and C atoms, together with side-chain torsions computed from standard amino-acid chi definitions. We first construct target receptor latents from the aligned holo frame. We then perform frame-based receptor alignment and side-chain conformational matching to obtain receptor coordinates that preserve the target holo conformation while maintaining chemically valid local geometry. In particular, side-chain torsions are matched by rotating downstream atoms around chi axes, with clash-aware acceptance criteria. A subsequent backbone cleanup stage repairs local peptide-continuity issues and reduces artifacts introduced by frame-wise reconstruction. The final receptor target state contains post-cleanup receptor atom coordinates, residue-frame rotations and translations, frame-valid masks, receptor chi values, and receptor chi masks.

Ligand topology, pose, and torsion matching. We also construct a consistent ligand representation for each receptor-ligand pair. From the mapped ligand structure, we identify heavy atoms, atomic numbers, rotatable bonds, torsion-defining atom quadruplets, anchor atoms, and downstream masks for torsion rotations. The apo ligand coordinates are stored as the reference ligand state, while each selected holo frame provides a frame-specific target ligand conformation. For each target frame, we fit a rigid ligand pose and torsion vector so that the reconstructed ligand coordinates match the aligned holo ligand coordinates. The saved ligand state therefore contains the target ligand atom coordinates, ligand pose rotation and translation, ligand torsion values, torsion masks, and reconstruction diagnostics. This representation is consistent with the decoupled ligand-pose parameterization used by the model.

Frame-specific ligand augmentation. To better represent the ensemble local docking setting, we augment the initial ligand state separately for each selected holo frame. Rather than using only one fixed ligand initialization for all holo states of a receptor-ligand pair, we generate frame-specific apo-like ligand starts by applying controlled rigid perturbations and selecting poses that remain near the binding region while avoiding severe receptor-ligand clashes. Candidate augmented poses are evaluated using center-distance criteria, ligand RMSD improvement criteria, and clash constraints against the local receptor environment. The selected augmented pose is saved as the frame-specific ligand initial state, including augmented atom coordinates, pose rotation and translation, and torsion values. Thus, each receptor-ligand pair can contribute multiple apo-like initial states and multiple holo target states, matching our ensemble flexible docking formulation.

Final dataset package. After preprocessing, each training item corresponds to a PDB entry and one selected holo ensemble frame. The per-frame directory contains the aligned apo and holo coordinates, receptor active-region masks, receptor target latents, cleaned receptor target coordinates, ligand target pose and torsions, and optional frame-specific ligand augmentation. Static per-complex files contain ligand topology, receptor topology, receptor local templates, receptor chi topology, atomic numbers, and apo receptor latents. During training, the dataloader samples these processed apo-holo state pairs and constructs online interpolated states x_t between the frame-specific initial state and the selected holo target. This dataset construction enables PIGEON to learn ensemble flexible docking transitions while preserving the realistic constraint that the true pocket residue indices are not provided as model input.

B Model Implementation

This section provides implementation details for PIGEON that are omitted from the main Method section for clarity. We describe the heterogeneous graph encoder, time conditioning, pocket-field construction, learnable selector implementation, output heads, online interpolation, and training protocol.

B.1 Heterogeneous Graph Construction

For every interpolated state x_t, we reconstruct the current receptor atom coordinates and ligand atom coordinates before graph construction. The graph contains three node types: ligand atoms, receptor atoms, and receptor residues. Ligand atoms represent the chemically detailed ligand geometry. Receptor atoms provide atom-level steric and contact information. Receptor residues provide coarse flexible receptor state information, including residue identity, local frame, and side-chain torsions.

The graph uses five edge types. Ligand bond edges connect covalently bonded ligand atoms and carry ligand bond features. Receptor atom-radius edges connect receptor atoms within a fixed spatial cutoff and encode local atom-level receptor geometry. Receptor residue-radius edges connect nearby residues and encode residue-level geometry. Receptor atom-to-residue edges connect receptor atoms to their parent residues. Ligand-receptor cross edges connect ligand atoms to nearby receptor atoms and provide the main local interaction channel between ligand and receptor.

For a generic spatial edge from node u to node v, with positions $\mathbf{p}_u$ and $\mathbf{p}_v$, we compute the relative displacement, distance, and normalized direction as $\mathbf{d}_{uv} = \mathbf{p}_v - \mathbf{p}_u$, $r_{uv} = \|\mathbf{d}_{uv}\|$, and $\widehat{\mathbf{d}}_{uv} = \mathbf{d}_{uv}/(r_{uv} + \epsilon)$. Distances are further encoded with radial basis functions $\phi_\ell(r_{uv}) = \exp(-(r_{uv} - \mu_\ell)^2/(2\sigma_{\mathrm{rbf}}^2))$. The final edge feature concatenates the scalar distance, unit direction, radial basis encoding, and any edge-type-specific attributes.

B.2 Node and Edge Feature Projections

Each node type has its own input projection into the shared hidden dimension. Ligand atom features include atom identity, ligand atom descriptors, atomic number, van der Waals radius, current atom coordinate information, and time features. Receptor atom features include atom identity, atomic number, van der Waals radius, current atom coordinates, and time features. Receptor residue features include residue identity features, residue-level embeddings, current receptor frame origin, current receptor frame rotation, encoded side-chain torsion values, and time features.

For torsion features, angular values are encoded with sine and cosine channels rather than raw angles. This avoids discontinuities at the $-\pi$ and π boundary. For receptor frames, rotations are represented in flattened matrix form in the input feature vector, while output frame residuals are predicted as 6D local rigid-frame residuals.

Each edge type also has a separate projection into the shared edge hidden dimension. This allows ligand bonds, receptor atom contacts, receptor residue contacts, atom-residue edges, and ligand-receptor cross edges to retain distinct geometric and chemical meanings while being processed by the same heterogeneous interaction backbone.

B.3 Time Conditioning with FiLM Modulation

PIGEON uses explicit time conditioning throughout the encoder. The scalar interpolation time t is first embedded with a sinusoidal embedding and projected to the model hidden dimension. Let $\mathbf{e}_t$ denote the resulting time embedding. For each node type, the initial projected node state is modulated by a FiLM transformation:

$$\mathrm{FiLM}(\mathbf{h}, \mathbf{e}_t) = \mathrm{LN}(\mathbf{h}) \odot (1 + \gamma(\mathbf{e}_t)) + \beta(\mathbf{e}_t).$$

Here $\gamma(\cdot)$ and $\beta(\cdot)$ are learned linear maps specific to the node type. We also apply analogous FiLM-style time modulation after each heterogeneous interaction block. This ensures that the same molecular geometry can be interpreted differently depending on whether the state is close to the apo start, close to the holo target, or in the middle of the interpolation path.

B.4 Heterogeneous Interaction Blocks

The encoder consists of multiple heterogeneous interaction blocks. Each block updates ligand atom states, receptor atom states, and receptor residue states using the corresponding projected edge features and edge indices. Within each block, messages are computed separately for different edge types and aggregated into the appropriate target node type. Residual connections and layer normalization are applied to stabilize training.

After the final interaction block, we denote the output embeddings by $\mathbf{h}_j^{\mathrm{lig}}$ for ligand atom j, $\mathbf{h}_a^{\text{rec-atom}}$ for receptor atom a, and $\mathbf{h}_i^{\text{rec-res}}$ for receptor residue i. These embeddings are used by all prediction heads. We also compute pooled graph-level summaries by mean pooling ligand atom embeddings, receptor atom embeddings, and receptor residue embeddings over each complex in the batch. These pooled summaries are normalized and concatenated with the time embedding to form the global complex context.

B.5 Local Ligand-Receptor Context

In addition to global pooling, PIGEON builds a local ligand-receptor context from ligand-receptor cross edges. This context is designed to summarize receptor atoms near the current ligand placement. A ligand-derived query attends over nearby receptor atom embeddings, producing a local pocket-context vector. This vector is combined with pooled ligand and receptor summaries and the explicit time embedding to form the sample-level feature used by several output heads. This local context is important because translation, rotation, and receptor flexibility are primarily determined by the ligand's immediate receptor environment rather than by the entire protein equally.

B.6 Pocket Field Tensor and Invariant Summaries

The graph builder computes a ligand-centered pocket field from receptor geometry before running the translation module. For receptor atom i with current position $\mathbf{q}_{i,t}$, let b_i^{local} denote the local receptor-intrinsic buriedness score computed by probe occlusion. We use the normalized pocket score

$$s_i^{\mathrm{pocket}} = \mathrm{clip}\left(\frac{b_i^{\mathrm{local}}}{b_{\mathrm{ref}}}, 0, 1\right), \tag{2}$$

where b_{ref} is a fixed reference buriedness value.

At the current ligand centroid $\mathbf{c}_t^{\mathrm{lig}}$, define

$$\mathbf{d}_{i,t} = \mathbf{q}_{i,t} - \mathbf{c}_t^{\mathrm{lig}}, \qquad r_{i,t} = \|\mathbf{d}_{i,t}\|, \qquad \widehat{\mathbf{d}}_{i,t} = \frac{\mathbf{d}_{i,t}}{r_{i,t} + \epsilon}. \tag{3}$$

We use six radial shells with centers

$$R = (1.5, 2.5, 4.0, 8.0, 14.0, 20.0) \tag{4}$$

and widths

$$\sigma = (0.75, 1.0, 2.0, 3.0, 4.0, 5.0)\,. \tag{5}$$

The soft shell membership is

$$\phi_s(r_{i,t}) = \exp\!\left(-\frac{(r_{i,t} - R_s)^2}{2\sigma_s^2}\right). \tag{6}$$

For outer shells, the atom weight is

$$w_{is}^{\mathrm{std}} = s_i^{\mathrm{pocket}} \phi_s(r_{i,t}), \tag{7}$$

while for the two inner shells we use a gradient-like radial upweighting

$$w_{is}^{\mathrm{grad}} = \frac{s_i^{\mathrm{pocket}} \phi_s(r_{i,t})}{r_{i,t} + \epsilon_r}. \tag{8}$$

The implementation uses $\epsilon_r = 0.5$ Å. The effective shell weight is

$$w_{is} = w_{is}^{\mathrm{grad}} \mathbf{1}[s \in \{1,2\}] + w_{is}^{\mathrm{std}} \mathbf{1}[s \in \{3,4,5,6\}]. \tag{9}$$

For inner shells, weights are normalized by the total shell mass, while outer shells use raw weights:

$$\widetilde{w}_{is} = \frac{w_{is}}{\sum_i w_{is} + \epsilon} \mathbf{1}[s \in \{1,2\}] + w_{is} \mathbf{1}[s \in \{3,4,5,6\}]. \tag{10}$$

For the real $l = 2$ components, we use

$$Y_2(\widehat{\mathbf{d}}) = [xy, yz, 3z^2 - 1, xz, x^2 - y^2], \tag{11}$$

where $\widehat{\mathbf{d}} = (x, y, z)$. The shellwise spherical harmonic coefficients are

$$f_0^{(s)} = \sum_i \widetilde{w}_{is}, \tag{12}$$

$$\mathbf{f}_1^{(s)} = \sum_i \widetilde{w}_{is} \widehat{\mathbf{d}}_{i,t}, \tag{13}$$

$$\mathbf{f}_2^{(s)} = \sum_i \widetilde{w}_{is} Y_2(\widehat{\mathbf{d}}_{i,t}). \tag{14}$$

The shellwise dipole norm, dipole direction, and quadrupole power are

$$n_1^{(s)} = \|\mathbf{f}_1^{(s)}\|, \qquad \widehat{\mathbf{f}}_1^{(s)} = \frac{\mathbf{f}_1^{(s)}}{n_1^{(s)} + \epsilon}, \qquad P_2^{(s)} = \|\mathbf{f}_2^{(s)}\|_2^2. \tag{15}$$

The full shellwise spherical harmonic tensor is

$$\mathbf{F}_{\mathrm{SH}}(\mathbf{c}_t^{\mathrm{lig}}) = \bigoplus_{s=1}^{6} \left[f_0^{(s)}, \mathbf{f}_1^{(s)}, \mathbf{f}_2^{(s)} \right] \in \mathbb{R}^{54}. \tag{16}$$

The six shellwise dipole directions $\{\widehat{\mathbf{f}}_1^{(s)}\}_{s=1}^{6}$ form the first six translation candidates.

For the multi-scale pocket-potential branch, we use five preferred distances

$$r_{\mathrm{opt}} = (2.0, 3.5, 5.0, 8.0, 12.0) \tag{17}$$

with widths

$$\sigma_{\mathrm{pot}} = (1.0, 1.5, 2.0, 3.0, 5.0)\ . \tag{18}$$

For potential scale a, define

$$\psi_a(r_{i,t}) = \exp\!\left(-\frac{(r_{i,t} - r_{\mathrm{opt},a})^2}{2\sigma_{\mathrm{pot},a}^2}\right). \tag{19}$$

The raw pocket-potential gradient is

$$\mathbf{g}_a = \sum_i s_i^{\mathrm{pocket}} \psi_a(r_{i,t}) \frac{r_{i,t} - r_{\mathrm{opt},a}}{\sigma_{\mathrm{pot},a}^2} \widehat{\mathbf{d}}_{i,t}. \tag{20}$$

We store its norm, unit direction, and log-norm feature:

$$n_a^{\mathrm{pot}} = \|\mathbf{g}_a\|, \qquad \widehat{\mathbf{g}}_a = \frac{\mathbf{g}_a}{n_a^{\mathrm{pot}} + \epsilon}, \qquad \eta_a = \log(1 + n_a^{\mathrm{pot}}). \tag{21}$$

The five normalized vectors $\{\widehat{\mathbf{g}}_a\}_{a=1}^5$ form the next five translation candidates.

The pocket field also includes four distance-to-surface features. Let τ_b be a buriedness threshold. The nearest-buried distance is

$$d_{\mathrm{buried}} = \min_{i: s_i^{\mathrm{pocket}} > \tau_b} r_{i,t}. \tag{22}$$

If no receptor atom passes the threshold, we use the centroid-distance fallback. For a finite radius R_c, define the local pocket-score-weighted centroid

$$\bar{\mathbf{q}}_{R_c} = \frac{\sum_{i: r_{i,t} < R_c} s_i^{\mathrm{pocket}} \mathbf{q}_{i,t}}{\sum_{i: r_{i,t} < R_c} s_i^{\mathrm{pocket}} + \epsilon}, \qquad d_{\mathrm{centroid}} = \|\mathbf{c}_t^{\mathrm{lig}} - \bar{\mathbf{q}}_{R_c}\|. \tag{23}$$

For a near radius R_n and Gaussian weight

$$\varphi_i = \exp\!\left(-\frac{1}{2}\left(\frac{r_{i,t}}{\sigma_n}\right)^2\right), \tag{24}$$

the local buriedness mean and variance are

$$\mu_b = \frac{\sum_{i: r_{i,t} < R_n} \varphi_i s_i^{\mathrm{pocket}}}{\sum_{i: r_{i,t} < R_n} \varphi_i + \epsilon}, \tag{25}$$

$$\mathrm{Var}_{\mathrm{local}} = \frac{\sum_{i: r_{i,t} < R_n} \varphi_i (s_i^{\mathrm{pocket}} - \mu_b)^2}{\sum_{i: r_{i,t} < R_n} \varphi_i + \epsilon}. \tag{26}$$

For local anisotropy, within radius R_a we compute

$$\boldsymbol{\mu}_a = \frac{\sum_{i: r_{i,t} < R_a} s_i^{\mathrm{pocket}} \mathbf{q}_{i,t}}{\sum_{i: r_{i,t} < R_a} s_i^{\mathrm{pocket}} + \epsilon}, \tag{27}$$

$$C_a = \frac{\sum_{i: r_{i,t} < R_a} s_i^{\mathrm{pocket}} (\mathbf{q}_{i,t} - \boldsymbol{\mu}_a)(\mathbf{q}_{i,t} - \boldsymbol{\mu}_a)^\top}{\sum_{i: r_{i,t} < R_a} s_i^{\mathrm{pocket}} + \epsilon}. \tag{28}$$

If $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of C_a, then

$$A_{\mathrm{local}} = 1 - \frac{\lambda_{\min}}{\lambda_{\max} + \epsilon}. \tag{29}$$

The distance-to-surface feature vector is

$$\mathbf{d}_{\text{surf}} = [d_{\text{buried}}, d_{\text{centroid}}, \text{Var}_{\text{local}}, A_{\text{local}}]. \tag{30}$$

The invariant pocket summary used by the magnitude head is

$$\mathbf{z}_{\text{pocket inv}} = [f_0^{(1)}, \ldots, f_0^{(6)}, P_2^{(1)}, \ldots, P_2^{(6)}, \eta_1, \ldots, \eta_5, \mathbf{d}_{\text{surf}}]. \tag{31}$$

Thus, the translation module uses both directional pocket cues, through $\widehat{\mathbf{f}}_1^{(s)}$ and $\widehat{\mathbf{g}}_a$, and invariant pocket summaries, through $\mathbf{z}_{\text{pocket inv}}$.

B.7 Candidate Direction Bank Implementation

The candidate direction bank is constructed deterministically from the pocket geometry at the current ligand centroid. Six directions are obtained from the $L = 1$ spherical harmonic dipoles, one per radial shell. Five directions are obtained from the multi-scale pocket-potential gradients. Two directions are obtained from long-range buriedness cues: the direction to the buriedness-weighted receptor centroid and the direction to the nearest highly buried receptor atom. All candidate directions are normalized to unit length with an ϵ stabilizer. If a direction has near-zero norm, the implementation uses numerical stabilization to avoid invalid values.

The final candidate bank has 13 directions. This bank is recomputed for every interpolated state x_t, so the candidate directions are ligand-position dependent. As the ligand moves along the interpolation path, the spherical harmonic field and pocket-potential gradients change accordingly, allowing the candidate bank to adapt to the current local docking geometry.

B.8 Learnable Staged Selector Implementation

The staged direction selector is implemented as a small set of learnable coefficients rather than a free neural network over candidate directions. For each candidate and probe radius, the graph builder computes five atom-aware features: contact quality, clash score, void quality, uniformity, and contact count. These features are z-scored across the 13 candidates at each probe radius. The selector then learns feature weights and probe-radius weights separately for the ingress and fit regimes.

The selector contains four softmax-normalized coefficient vectors: ingress feature weights, fit feature weights, ingress probe-radius weights, and fit probe-radius weights. The selector also contains a learnable stage gate that maps the current-state stage-feature vector to a scalar $\alpha_{\text{fit}} \in [0, 1]$. The final candidate logit is the convex interpolation between ingress and fit scores. All selector coefficients are initialized from the hand-designed geometric rule, and the training objective includes a regularization term that penalizes drift away from the initialization. This design lets the model learn from data while preserving the interpretability of the original rule.

B.9 Stage-Feature Definitions

The learnable staged selector uses a five-dimensional current-state feature vector

$$\mathbf{z}_{\text{stage}} = [z_d, z_q, z_c, z_u, z_n]. \tag{32}$$

These features indicate whether the current ligand state is in an ingress-like regime, where the ligand should move toward a pocket, or a local-fit regime, where the ligand should refine contacts near the current pose.

We first compute a local pocket centroid around the current ligand centroid. For receptor atom i, define

$$\omega_i^{\text{stage}} = s_i^{\text{pocket}} \exp\left(-\frac{1}{2}\left(\frac{\|\mathbf{q}_{i,t} - \mathbf{c}_t^{\text{lig}}\|}{\sigma_{\text{stage}}}\right)^2\right) \mathbf{1}[\|\mathbf{q}_{i,t} - \mathbf{c}_t^{\text{lig}}\| < R_{\text{stage}}]. \tag{33}$$

The stage pocket centroid is

$$\bar{\mathbf{q}}_{\text{stage}} = \frac{\sum_i \omega_i^{\text{stage}} \mathbf{q}_{i,t}}{\sum_i \omega_i^{\text{stage}} + \epsilon}. \tag{34}$$

If the total weight is numerically zero, we set $\bar{\mathbf{q}}_{\text{stage}} = \mathbf{c}_t^{\text{lig}}$. Let the current ligand radius be

$$r_{\text{lig}} = \max_j \|\mathbf{x}_{j,t} - \mathbf{c}_t^{\text{lig}}\|. \tag{35}$$

The normalized pocket-distance feature is

$$z_d = \mathrm{clip}\left(\frac{\|\mathbf{c}_t^{\mathrm{lig}} - \bar{\mathbf{q}}_{\mathrm{stage}}\|}{r_{\mathrm{lig}} + 1}, 0, 2\right). \tag{36}$$

The remaining four stage features are computed by evaluating the same atom-aware probe feature function at the current untranslated ligand pose, i.e., at destination centroid $\mathbf{p} = \mathbf{c}_t^{\mathrm{lig}}$. Let

$$[Q_{@\mathbf{c}_t}, C_{@\mathbf{c}_t}, V_{@\mathbf{c}_t}, U_{@\mathbf{c}_t}, N_{@\mathbf{c}_t}] \tag{37}$$

denote contact quality, clash score, void quality, uniformity, and contact count at the current pose. We also retain the raw clash penalty $P_{\mathrm{clash},@\mathbf{c}_t}$ before the negative log transformation. The bounded stage signals are

$$z_q = \mathrm{clip}\left(\frac{Q_{@\mathbf{c}_t} - 2.5}{1.5}, 0, 1\right), \tag{38}$$

$$z_c = \mathrm{clip}\left(\frac{P_{\mathrm{clash},@\mathbf{c}_t}}{1.5}, 0, 1\right), \tag{39}$$

$$z_u = \mathrm{clip}(U_{@\mathbf{c}_t}, 0, 1), \tag{40}$$

$$z_n = \mathrm{clip}\left(\frac{N_{@\mathbf{c}_t} - 2.5}{2.0}, 0, 1\right). \tag{41}$$

Thus, z_q measures current contact quality, z_c measures current clash, z_u measures current contact uniformity, and z_n measures current contact count.

B.10 Translation Magnitude Head

After the selector chooses a unit direction $\widehat{\mathbf{u}} = \mathbf{v}_{k^\star}$, PIGEON predicts a nonnegative translation magnitude from graph-level and pocket-invariant features. Let $\mathbf{h}^{\mathrm{graph}}$ denote the pooled graph context and let $\mathbf{h}^{\mathrm{time}}$ denote the time embedding. We first project the invariant pocket summary from Eq. 31:

$$\mathbf{h}^{\mathrm{pocket\ inv}} = \phi_{\mathrm{inv}}(\mathbf{z}_{\mathrm{pocket\ inv}}). \tag{42}$$

The magnitude head input is

$$\mathbf{h}^{\mathrm{mag}} = [\mathbf{h}^{\mathrm{graph}}, \mathbf{h}^{\mathrm{time}}, \mathbf{h}^{\mathrm{pocket\ inv}}]. \tag{43}$$

The predicted magnitude is

$$\widehat{m} = \mathrm{softplus}(f_{\mathrm{mag}}(\mathbf{h}^{\mathrm{mag}})). \tag{44}$$

The final ligand translation residual is

$$\widehat{\Delta\mathbf{c}}_{t\to 1}^{\mathrm{lig}} = \widehat{m}\mathbf{v}_{k^\star}. \tag{45}$$

This factorization separates the direction decision from the distance prediction: the selected direction comes from receptor-derived geometric candidates, while the scalar magnitude is learned from pooled complex context and invariant pocket-field summaries.

B.11 Ligand Rotation and Torsion Heads

The ligand rotation head predicts a 3D $SO(3)$ log residual. It uses pooled ligand atom embeddings, the sample-level complex context, and explicit time features. The output is interpreted as the remaining rotation vector from R_t^{lig} to R_1^{lig}.

The ligand torsion head predicts wrapped residuals for ligand rotatable bonds. For each rotatable bond, the model builds a torsion feature from the embeddings of the atoms defining the torsion and from torsion-local geometric information. A shared torsion MLP predicts one residual per valid torsion. The output is bounded with a smooth angular nonlinearity so that the predicted torsion residual remains in a physically meaningful angular range. Invalid or padded torsions are masked out of the loss.

B.12 Receptor Frame and Side-Chain Heads

The receptor frame head predicts one 6D residual for each receptor residue. The output represents the remaining local rigid-frame motion from the current receptor frame to the target holo receptor frame. The receptor chi head predicts side-chain torsion residuals for each residue. The number of chi angles varies by residue type, so the receptor chi loss uses a residue-specific chi mask. Both receptor heads condition on receptor residue embeddings and sample-level context, allowing receptor flexibility to be predicted in a ligand-aware and time-aware manner.

B.13 Output Dictionary and Ligand Pose Assembly

For compatibility with the loss and evaluation code, the predicted ligand pose is assembled as a 6D vector by concatenating the translation and rotation predictions:

$$\widehat{\mathbf{y}}^{\mathrm{lig}}_{\mathrm{pose}} = [\widehat{\Delta\mathbf{c}}^{\mathrm{lig}}_{t\to 1}; \widehat{\boldsymbol{\omega}}^{\mathrm{lig}}_{t\to 1}].$$

The translation portion is produced by the direction selector and magnitude head, while the rotation portion is produced by the ligand rotation head. The model output dictionary also stores diagnostic quantities such as candidate logits, selector probabilities, selected candidate index, oracle candidate cosine values, predicted translation magnitude, and selector-stage gate value. These diagnostics are used to monitor whether the model is learning the intended geometric behavior.

B.14 Loss Implementation Details

The selector is trained with a soft listwise ranking objective. For each sample, the oracle score for candidate k is the cosine similarity between candidate direction $\mathbf{v}_k$ and the normalized ground-truth translation direction. A low-temperature softmax over oracle cosines produces the target distribution. The selector logits are trained with cross entropy against this soft oracle distribution. This supervision gives high probability to all candidates close to the true direction, rather than treating direction selection as a brittle one-hot classification problem.

The magnitude head is trained in log space against the ground-truth residual translation norm. This reduces sensitivity to the absolute scale of the translation and encourages relative magnitude accuracy. Receptor frame residuals, receptor chi residuals, ligand rotation residuals, and ligand torsion residuals are trained with fixed-scale normalized Smooth L1 losses. Fixed scales are used instead of target-dependent normalization so that the loss remains stable when the target residual norm is small.

The complete loss is a weighted sum of selector ranking loss, selector regularization, translation log-magnitude loss, receptor frame loss, receptor chi loss, ligand rotation loss, and ligand torsion loss. The final weights are selected based on validation performance and diagnostic stability.

Selector regularization. To keep the selector close to the intended geometric rule family, we regularize the learnable feature logits, radius logits, and stage-gate parameters toward their initialized values. Let $\boldsymbol{\theta}^{\mathrm{fit}}_{\mathrm{feat}}$ and $\boldsymbol{\theta}^{\mathrm{ing}}_{\mathrm{feat}}$ denote the fit and ingress feature logits, and let $\boldsymbol{\theta}^{\mathrm{fit}}_{\mathrm{rad}}$ and $\boldsymbol{\theta}^{\mathrm{ing}}_{\mathrm{rad}}$ denote the fit and ingress probe-radius logits. The selector regularization is

$$\begin{aligned}\mathcal{L}_{\text{selector reg}} =& \|\boldsymbol{\theta}^{\mathrm{fit}}_{\mathrm{feat}} - \boldsymbol{\theta}^{\mathrm{fit}}_{\mathrm{feat},0}\|_2^2 + \|\boldsymbol{\theta}^{\mathrm{ing}}_{\mathrm{feat}} - \boldsymbol{\theta}^{\mathrm{ing}}_{\mathrm{feat},0}\|_2^2 \\ &+ \|\boldsymbol{\theta}^{\mathrm{fit}}_{\mathrm{rad}} - \boldsymbol{\theta}^{\mathrm{fit}}_{\mathrm{rad},0}\|_2^2 + \|\boldsymbol{\theta}^{\mathrm{ing}}_{\mathrm{rad}} - \boldsymbol{\theta}^{\mathrm{ing}}_{\mathrm{rad},0}\|_2^2 + \mathcal{L}_{\text{gate reg}},\end{aligned} \tag{46}$$

where

$$\mathcal{L}_{\text{gate reg}} = \|\mathbf{w}_{\mathrm{gate}} - \mathbf{w}_{\mathrm{gate},0}\|_2^2 + (b_{\mathrm{gate}} - b_{\mathrm{gate},0})^2 + (\log s_{\mathrm{gate}} - \log s_{\mathrm{gate},0})^2. \tag{47}$$

This regularizer allows the selector to adapt to data while preserving the inductive bias of the hand-designed ingress and fit rules.

B.15 Training and Inference Details

We train PIGEON using the selector configuration described in the main text. All experiments use random seed 42 and are trained on four NVIDIA RTX A6000 GPUs with 48GB memory each. Training uses PyTorch Lightning distributed data parallelism with `ddp_find_unused_parameters_true`, bf16 mixed precision, batch size 1, gradient accumulation over 8 steps, and gradient clipping with global norm threshold 1.0. The model is trained for 20 epochs, with validation performed once per epoch. We use AdamW with learning rate 1.0×10^{-4}, weight decay 0.01, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and

$\epsilon = 1.0 \times 10^{-8}$. The dataset is split by PDB ID into train, validation, and test partitions with fractions $0.70/0.15/0.15$ using split seed 42.

The model uses hidden dimension 256, edge hidden dimension 256, time embedding dimension 128, and 6 heterogeneous interaction blocks with dropout 0.0. The receptor-residue, receptor-atom, and ligand-atom input feature dimensions are 1307, 6, and 523, respectively. The edge feature dimensions are 20 for receptor atom-radius edges, receptor residue-radius edges, receptor atom-to-residue edges, and ligand-receptor cross edges, and 23 for ligand bond edges. The graph builder uses a receptor atom-radius cutoff of 6.0 Å, receptor residue-radius cutoff of 12.0 Å, and ligand-receptor cross cutoff of 15.0 Å. The maximum numbers of receptor atom neighbors, receptor residue neighbors, and cross-edge neighbors are 64, 32, and 128, respectively. Distances are encoded with 16 radial basis functions, and dense attention masks and dense node features are included.

For the pocket geometry module, the spherical harmonic field uses 6 radial shells with radii $(1.5, 2.5, 4.0, 8.0, 14.0, 20.0)$ Å and widths $(0.75, 1.0, 2.0, 3.0, 4.0, 5.0)$ Å. We use real spherical harmonics up to $L = 2$, giving a 54-dimensional shell field. Buriedness is computed with 48 probe directions, a buriedness probe radius of 6.0 Å, local and global probe radii of 5.0 Å and 10.0 Å, an occlusion radius scale of 0.8, neighbor exclusion margin 0.5 Å, and buriedness reference value 0.3. The inner two shells are used for gradient-like SH weighting with $\epsilon = 0.5$. The pocket-potential branch uses five preferred radii $(2.0, 3.5, 5.0, 8.0, 12.0)$ Å with Gaussian widths $(1.0, 1.5, 2.0, 3.0, 5.0)$ Å. The distance-to-surface branch uses buriedness threshold 0.25, centroid radius 20.0 Å, near radius 6.0 Å, near sigma 2.0 Å, and anisotropy radius 6.0 Å.

The translation module uses pocket-guided translation with a 13-direction candidate bank. The selector probe radii are $(0.5, 1.0, 1.5, 2.5, 4.0)$ Å. The selector uses contact-probe radius 10.0 Å, buriedness threshold 0.15, preferred contact gap 0.5 Å, contact sigma 0.75 Å, contact-shell inner and outer gaps 0.0 Å and 1.5 Å, void-gap threshold 2.5 Å, stage radius 12.0 Å, and stage sigma 6.0 Å. Candidate-type bias is disabled. The translation magnitude head uses translation hidden dimension 256, pocket SH field dimension 74, pocket distance feature dimension 4, pocket invariant projection dimension 256, and translation time dropout 0.25.

The interpolant uses residual targets with polynomial schedule $\alpha(t) = t^2$ and time stabilizer 10^{-4}. Intermediate states are sampled with uniform-alpha time sampling. Clearance projection is disabled in the current setting. Although declashing and cleanup hyperparameters remain specified in the configuration for compatibility, they are not used as a clearance-projection step during training because `use_clearance_projection` is set to false.

The training objective uses Smooth L1 residual losses and the structured selector losses described in the main text. The receptor frame, receptor chi, and ligand torsion losses each have weight 1.0. The ligand pose loss weight is set to 0.0, while ligand rotation diagnostics are still computed with a rotation-specific weight parameter. Direct Cartesian translation loss, translation direction cosine loss, geometry-magnitude loss, steer-direction loss, geometry-direction loss, and residual regularization are disabled. Translation is trained through the selector rank loss and log-magnitude loss: the selector rank loss has weight 1.0 with oracle temperature 0.05, the selector coefficient regularization has weight 100, and the log-magnitude loss has weight 1.0 with a 300-step warmup and $\epsilon = 10^{-4}$. Fixed normalization scales are 5.0 for receptor frame residuals, 0.05 for receptor chi residuals, 0.25 for ligand rotation residuals, and 0.25 for ligand torsion residuals.

At inference time, we use the same configuration, and model predictions as the residual source. Unless otherwise stated, inference is performed on the test split with 4 integration steps and terminal time $t_{\text{end}} = 0.85$. We run PoseBusters v2 during inference to compute physical-validity diagnostics. The ODE inference command uses one GPU and writes predicted structures and evaluation outputs to the specified checkpoint output directory.

C Algorithms

Algorithm 1 Pocket-Aware Translation Direction Selection

Input: interpolated complex state x_t at time t
  receptor atom positions $\{\mathbf{q}_{a,t}\}$ with vdW radii $\{r_a^{\mathrm{rec}}\}$
  ligand atom positions $\{\mathbf{x}_{j,t}\}$ with vdW radii $\{r_j^{\mathrm{lig}}\}$, centroid $\mathbf{c}_t^{\mathrm{lig}}$
  learnable selector parameters $\theta_{\mathrm{sel}} = (\boldsymbol{\eta}^{\mathrm{ing}}, \boldsymbol{\eta}^{\mathrm{fit}}, \boldsymbol{\zeta}^{\mathrm{ing}}, \boldsymbol{\zeta}^{\mathrm{fit}}, \mathbf{w}_{\mathrm{gate}}, b_{\mathrm{gate}}, s_{\mathrm{gate}})$
Hyperparameters: probe directions K_{probe}, shells $M = 6$, harmonic order $L = 2$, probe schedule $\mathcal{R} = \{0.5, 1.0, 1.5, 2.5, 4.0\}$ Å, potential scales, buriedness threshold τ_b, contact gaps $(\gamma_{\mathrm{in}}, \gamma_{\mathrm{out}}, \gamma_{\mathrm{pref}}, \gamma_{\mathrm{void}})$
▷ *Step 1: receptor-intrinsic buriedness, independent of ligand pose*
for each receptor atom a **do**
  Cast K_{probe} probe directions $\{\mathbf{u}_k\}$ around $\mathbf{q}_{a,t}$
  Compute occlusion indicators $\{o_{ak}\}$ ▷ *Eq. 1*
  $b_a \leftarrow \frac{1}{K_{\mathrm{probe}}} \sum_k o_{ak}$
end for
▷ *Step 2: ligand-centered spherical harmonic pocket field*
for each shell $s = 1, \ldots, M$ **do**
  for each harmonic degree $\ell = 0, 1, 2$ **do**
    $\mathbf{f}_\ell^{(s)}(\mathbf{c}_t^{\mathrm{lig}}) \leftarrow \sum_a b_a \, \phi_s(r_{a,t}) \, \mathbf{Y}_\ell(\widehat{\mathbf{r}}_{a,t})$
  end for
end for
Construct invariant pocket summaries $\mathbf{F}_{\mathrm{pocket}}$ from shell masses, SH powers, potential strengths, and distance-to-surface features
▷ *Step 3: build 13-direction candidate bank $\mathcal{V}_t$*
$\mathcal{V}_t \leftarrow \emptyset$
for $s = 1, \ldots, 6$ **do** ▷ *shellwise dipole directions*
  $\mathbf{v}_s^{\mathrm{SH}} \leftarrow \mathbf{d}_s^{\mathrm{SH}} / (\|\mathbf{d}_s^{\mathrm{SH}}\| + \epsilon)$
  $\mathcal{V}_t \leftarrow \mathcal{V}_t \cup \{\mathbf{v}_s^{\mathrm{SH}}\}$
end for
for each potential scale r_{opt} **do** ▷ *five gradient candidates*
  $\mathbf{v}_{r_{\mathrm{opt}}}^{\mathrm{pot}} \leftarrow \nabla_{\mathbf{c}_t^{\mathrm{lig}}} U_{r_{\mathrm{opt}}} / (\|\nabla_{\mathbf{c}_t^{\mathrm{lig}}} U_{r_{\mathrm{opt}}}\| + \epsilon)$
  $\mathcal{V}_t \leftarrow \mathcal{V}_t \cup \{\mathbf{v}_{r_{\mathrm{opt}}}^{\mathrm{pot}}\}$
end for
Add long-range directions $\mathbf{v}_{\mathrm{centroid}}$ and $\mathbf{v}_{\mathrm{nearest}}$ to $\mathcal{V}_t$
▷ *Step 4: probe each candidate at every probe distance*
for each candidate $\mathbf{v}_k \in \mathcal{V}_t$ **do**
  for each probe distance $\rho_s \in \mathcal{R}$ **do**
    $\mathbf{p}_{k,s} \leftarrow \mathbf{c}_t^{\mathrm{lig}} + \rho_s \mathbf{v}_k$
    $\widetilde{\mathbf{x}}_{j,k,s} \leftarrow \mathbf{x}_{j,t} + \rho_s \mathbf{v}_k$ for all ligand atoms j
    Build local receptor set $\mathcal{A}_{k,s}$ and compute surface gaps $g_{aj}^{(k,s)}$
    Compute features $F_{\mathrm{contact}}, F_{\mathrm{clash}}, F_{\mathrm{void}}, F_{\mathrm{uniform}}, F_{\mathrm{count}}$
  end for
end for
▷ *Step 5: within-sample normalization across candidates*
for each probe distance s and feature f **do**
  $Z(k, s, f) \leftarrow (F_f(k, s) - \mu_{s,f}) / \sigma_{s,f}$ for all candidates k
end for
▷ *Step 6: stage gate from current untranslated ligand pose*
Compute $\mathbf{z}_{\mathrm{stage}} = [z_d, z_q, z_c, z_u, z_n]$ ▷ *App. B.9*
$\alpha_{\mathrm{fit}} \leftarrow \sigma\big(s_{\mathrm{gate}}(\mathbf{w}_{\mathrm{gate}}^\top \mathbf{z}_{\mathrm{stage}} + b_{\mathrm{gate}})\big)$
▷ *Step 7: compute ingress/fit scores and select direction*
$\boldsymbol{\lambda}^{\mathrm{ing}}, \boldsymbol{\lambda}^{\mathrm{fit}} \leftarrow \mathrm{softmax}(\boldsymbol{\eta}^{\mathrm{ing}}), \mathrm{softmax}(\boldsymbol{\eta}^{\mathrm{fit}})$
$\boldsymbol{\rho}^{\mathrm{ing}}, \boldsymbol{\rho}^{\mathrm{fit}} \leftarrow \mathrm{softmax}(\boldsymbol{\zeta}^{\mathrm{ing}}), \mathrm{softmax}(\boldsymbol{\zeta}^{\mathrm{fit}})$
for each candidate k **do**
  $s_k^{\mathrm{ing}} \leftarrow \sum_s \rho_s^{\mathrm{ing}} \sum_f \lambda_f^{\mathrm{ing}} Z(k, s, f)$
  $s_k^{\mathrm{fit}} \leftarrow \sum_s \rho_s^{\mathrm{fit}} \sum_f \lambda_f^{\mathrm{fit}} Z(k, s, f)$
  $\ell_k \leftarrow (1 - \alpha_{\mathrm{fit}}) s_k^{\mathrm{ing}} + \alpha_{\mathrm{fit}} s_k^{\mathrm{fit}}$
end for
$k^\star \leftarrow \arg\max_k \ell_k; \quad \widehat{\mathbf{u}} \leftarrow \mathbf{v}_{k^\star}$
return selected direction $\widehat{\mathbf{u}}$, candidate logits $\{\ell_k\}$, candidate bank $\mathcal{V}_t$, pocket summaries $\mathbf{F}_{\mathrm{pocket}}$

Algorithm 2 PIGEON Training Step on a Sampled Apo-Holo Pair

Input: dataset $\mathcal{D}$ of apo–holo state pairs (x_0, x_1)
heterogeneous message-passing network f_θ with output heads
rule-based selector initialization θ_0
Hyperparameters: schedule $\alpha(t) = t^2$, oracle temperature T_{oracle}, loss weights $\{\lambda_{\text{rank}}, \lambda_{\text{reg}}, \lambda_{\text{mag}}, \lambda_{\text{frame}}, \lambda_\chi, \lambda_{\text{rot}}, \lambda_{\text{tor}}\}$, residual scales $\{s_{\text{frame}}, s_\chi, s_{\text{rot}}, s_{\text{tor}}\}$
▷ *Sample apo–holo pair and interpolation time*
Sample $(x_0, x_1) \sim \mathcal{D}$; $\alpha \sim \mathcal{U}(0,1)$; $t \leftarrow \sqrt{\alpha}$
▷ *Geodesic interpolation on the product manifold*
$x_t \leftarrow \text{Exp}_{x_0}\big(\alpha \,\text{Log}_{x_0}(x_1)\big)$ ▷ *component-wise on* $\mathcal{M}$
▷ *Encode complex with time-conditioned heterogeneous GNN*
$\{\mathbf{h}_j^{\text{lig}}, \mathbf{h}_a^{\text{rec-atom}}, \mathbf{h}_i^{\text{rec-res}}\} \leftarrow f_\theta(x_t, t)$
Let $\mathbf{h}$ denote pooled ligand, receptor-atom, receptor-residue, local-context, and time-conditioned features
▷ *Predict ligand translation via pocket-aware selector*
$\widehat{\mathbf{u}}, \{\ell_k\}_{k=1}^K, \{\mathbf{v}_k\}_{k=1}^K, \mathbf{F}_{\text{pocket}} \leftarrow$ Algorithm 1$(x_t, t, \theta_{\text{sel}})$
$\widehat{m} \leftarrow \text{Softplus}(\text{MagnitudeHead}(\mathbf{h}, t, \mathbf{F}_{\text{pocket}}))$
$\widehat{\Delta\mathbf{c}}_{t\to1}^{\text{lig}} \leftarrow \widehat{m}\,\widehat{\mathbf{u}}$
▷ *Predict remaining residuals from output heads*
$\widehat{\boldsymbol{\omega}}^{\text{lig}} \leftarrow \text{LigandRotationHead}(\mathbf{h})$
$\widehat{\boldsymbol{\chi}}^{\text{lig}} \leftarrow \text{LigandTorsionHead}(\mathbf{h})$
$\widehat{\boldsymbol{\Delta}}^{\text{frame}} \leftarrow \text{FrameHead}(\mathbf{h})$
$\widehat{\boldsymbol{\chi}}^{\text{rec}} \leftarrow \text{ChiHead}(\mathbf{h})$
▷ *Compute ground-truth residuals from* x_t *to* x_1
$\Delta\mathbf{c}_{t\to1}^{\text{lig},*} \leftarrow \mathbf{c}_1^{\text{lig}} - \mathbf{c}_t^{\text{lig}}$
$\mathbf{u}^* \leftarrow \Delta\mathbf{c}_{t\to1}^{\text{lig},*} / (\|\Delta\mathbf{c}_{t\to1}^{\text{lig},*}\| + \epsilon)$
Compute $\boldsymbol{\omega}_{t\to1}^{\text{lig},*}$, ligand torsion residuals, receptor frame residuals, and receptor side-chain residuals
▷ *Selector loss: listwise rank against soft oracle*
for each candidate k **do**
  $q_k^* \leftarrow \mathbf{v}_k^\top \mathbf{u}^*$
end for
for each candidate k **do**
  $p_k^* \leftarrow \exp(q_k^*/T_{\text{oracle}}) / \sum_j \exp(q_j^*/T_{\text{oracle}})$
end for
$\mathcal{L}_{\text{rank}} \leftarrow -\sum_k p_k^* \log\Big(\exp(\ell_k)/\sum_j \exp(\ell_j)\Big)$
$\mathcal{L}_{\text{reg}} \leftarrow \|\theta_{\text{sel}} - \theta_0\|_2^2$
▷ *Magnitude loss in log space*
$\mathcal{L}_{\text{mag}} \leftarrow \Big(\log(\widehat{m} + \epsilon) - \log(\|\Delta\mathbf{c}_{t\to1}^{\text{lig},*}\| + \epsilon)\Big)^2$
▷ *Fixed-scale residual losses for non-translation outputs*
$\mathcal{L}_{\text{rot}} \leftarrow \text{SmoothL1}(\widehat{\boldsymbol{\omega}}^{\text{lig}}/s_{\text{rot}}, \boldsymbol{\omega}_{t\to1}^{\text{lig},*}/s_{\text{rot}})$
$\mathcal{L}_{\text{tor}}, \mathcal{L}_{\text{frame}}, \mathcal{L}_\chi \leftarrow$ analogous SmoothL1 residual losses with fixed scales
▷ *Backpropagate combined objective*
$\mathcal{L} \leftarrow \lambda_{\text{rank}}\mathcal{L}_{\text{rank}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}} + \lambda_{\text{mag}}\mathcal{L}_{\text{mag}} + \lambda_{\text{frame}}\mathcal{L}_{\text{frame}} + \lambda_\chi\mathcal{L}_\chi + \lambda_{\text{rot}}\mathcal{L}_{\text{rot}} + \lambda_{\text{tor}}\mathcal{L}_{\text{tor}}$
Take an optimizer step on $\nabla_\theta \mathcal{L}$
return updated parameters θ

Algorithm 3 PIGEON Inference: Apo-to-Holo Sampling via Manifold Flow Integration

Input: initial apo-like state $x_0 \in \mathcal{M}$, trained network f_θ
Hyperparameters: number of integration steps T, terminal time t_{end}, schedule $\alpha(t) = t^2$, ensemble size S
$\mathcal{E} \leftarrow \emptyset$ ▷ *predicted holo ensemble*
for $s = 1, \ldots, S$ **do**
  Sample or choose apo-like initialization $x^{(0)}$ from x_0
  Set $\alpha_n \leftarrow \frac{n}{T}\alpha(t_{\text{end}})$ and $t_n \leftarrow \sqrt{\alpha_n}$ for $n = 0, \ldots, T$
  ▷ *Iteratively integrate the manifold residual field*
  for $n = 0, \ldots, T-1$ **do**
    $\{\mathbf{h}_j^{\text{lig}}, \mathbf{h}_a^{\text{rec-atom}}, \mathbf{h}_i^{\text{rec-res}}\} \leftarrow f_\theta(x^{(n)}, t_n)$
    Let $\mathbf{h}$ denote pooled ligand, receptor-atom, receptor-residue, local-context, and time-conditioned features
    ▷ *Predict ligand translation residual*
    $\widehat{\mathbf{u}}, \cdot, \cdot, \mathbf{F}_{\text{pocket}} \leftarrow$ Algorithm 1$(x^{(n)}, t_n, \theta_{\text{sel}})$
    $\widehat{m} \leftarrow \text{Softplus}(\text{MagnitudeHead}(\mathbf{h}, t_n, \mathbf{F}_{\text{pocket}}))$
    $\widehat{\Delta\mathbf{c}}^{\text{lig}} \leftarrow \widehat{m}\,\widehat{\mathbf{u}}$
    ▷ *Predict remaining endpoint residuals*
    $\widehat{\boldsymbol{\omega}}^{\text{lig}} \leftarrow \text{LigandRotationHead}(\mathbf{h})$
    $\widehat{\boldsymbol{\chi}}^{\text{lig}} \leftarrow \text{LigandTorsionHead}(\mathbf{h})$
    $\widehat{\boldsymbol{\Delta}}^{\text{frame}} \leftarrow \text{FrameHead}(\mathbf{h})$
    $\widehat{\boldsymbol{\chi}}^{\text{rec}} \leftarrow \text{ChiHead}(\mathbf{h})$
    ▷ *Convert endpoint residuals to this integration step*
    $\gamma_n \leftarrow (\alpha_{n+1} - \alpha_n)/(1 - \alpha_n + \epsilon)$
    ▷ *Update each manifold component*
    $\mathbf{c}_{\text{lig}}^{(n+1)} \leftarrow \mathbf{c}_{\text{lig}}^{(n)} + \gamma_n \widehat{\Delta\mathbf{c}}^{\text{lig}}$
    $R_{\text{lig}}^{(n+1)} \leftarrow R_{\text{lig}}^{(n)} \exp(\gamma_n \widehat{\boldsymbol{\omega}}^{\text{lig}})$
    $\boldsymbol{\chi}_{\text{lig}}^{(n+1)} \leftarrow \text{wrap}(\boldsymbol{\chi}_{\text{lig}}^{(n)} + \gamma_n \widehat{\boldsymbol{\chi}}^{\text{lig}})$
    Update receptor residue frames and side-chain torsions analogously using $\widehat{\boldsymbol{\Delta}}^{\text{frame}}$ and $\widehat{\boldsymbol{\chi}}^{\text{rec}}$
    Assemble $x^{(n+1)}$ from updated components
  end for
  $\mathcal{E} \leftarrow \mathcal{E} \cup \{x^{(T)}\}$
end for
return predicted holo ensemble $\mathcal{E}$