Hierarchical Hybrid-Coordinate Flow Matching for Accurate and Physically Plausible Apo-to-Holo Protein Generation

Anonymous authors
Paper under double-blind review

Abstract

Predicting holo-like protein conformations from apo structures is essential in structure-based drug design. Most existing apo-to-holo models operate in Cartesian space and achieve strong global accuracy. However, high global accuracy does not ensure physically valid local backbone geometry. Internal-coordinate models offer local control and validity, but regressive generation can amplify angular errors—the lever-arm effect. We introduce CrypticFlow, a hierarchical flow-matching model that combines these two complementary representations. An ESM-3-conditioned E(3)-equivariant graph network first generates sparse Cα anchors in Cartesian space. A conditional midpoint Transformer then recursively fills the gaps in local internal coordinates. During recursive generation, the decoder constrains the distance between each midpoint and any sequence-adjacent parent anchor to the corresponding apo Cα–Cα distance. On the D3PM test split, CrypticFlow achieves state-of-the-art median and mean Cα RMSD. Across five independent training models, the median and mean direct RMSD are 1.333 ± 0.056 Å and 2.766 ± 0.041 Å, respectively. The quality of local geometry also improves substantially. Our model achieves $99.58 \pm 0.02\%$ adjacent-distance fidelity and $95.87 \pm 0.25\%$ pseudo-angle fidelity, measured within 0.2 Å and 20° of the holo reference. Then, we reconstruct the predicted Cα-only traces into all-atom structures using the cg2all coarse-to-all-atom backmapping method and evaluate their physical and geometric quality with MolProbity. Our model achieves a MolProbity score of 3.034 ± 0.054, outperforming the evaluated baselines. These results show that a hierarchical flow-matching model with hybrid coordinates can improve global conformational accuracy and structural fidelity together.

1 Introduction

Structure-based drug design (SBDD) depends on the receptor conformation used for screening. A ligand-bound, or holo, structure reveals a ligand-compatible pocket. However, obtaining one requires a suitable ligand and an experiment that captures the bound state. Screening therefore often treats an apo structure as rigid, which can hide the pocket that a drug must recognize.

This problem is particularly important for *cryptic pockets*. These sites are not apparent in a static apo structure but can become accessible through protein motion or a binding-associated conformational change (Vajda et al., 2018; Meller et al., 2023). Covalent fragments, for example, revealed the switch-II pocket of KRAS(G12C) (Ostrem et al., 2013). This discovery enabled clinically effective inhibitors such as sotorasib and adagrasib (Skoulidis et al., 2021; Jänne et al., 2022). Predicting holo-like conformations from apo structures could similarly expand receptor ensembles before experimental holo structures are available.

Recent methods learn conditional bridges or flows from paired apo–holo structures (Somnath et al., 2023; Nobis et al., 2025). They are commonly evaluated using aligned Cα RMSD. RMSD measures global accuracy, but it does not ensure local-geometry fidelity. A prediction may approach the target while distorting adjacent distances, pseudo-angles, or nonlocal contacts. These errors become important when the Cα trace is reconstructed into an all-atom structure.

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.

Cartesian and internal coordinates offer complementary advantages. Cartesian updates represent global motion and work naturally with equivariant networks (Satorras et al., 2021). However, they do not explicitly enforce local backbone geometry. Internal coordinates represent distances and angles explicitly (Wu et al., 2024; Li et al., 2023), but chain-wide generation can suffer from the lever-arm effect (Li et al., 2023; Singh et al., 2025). A small angular error can displace many downstream residues.

We therefore predict sparse Cartesian anchors, then complete their intervals in a frame defined by two current parents and an apo reference direction. The main midpoint model uses local internal coordinates and copies a radius only when its parent is sequence-adjacent. Ablations identify this copying as a small refinement and the local internal-coordinate state (ICS) representation as the stronger bias for adjacent and reconstructed bond/angle fidelity. The shallow hierarchy limits, but does not eliminate, lever-arm amplification. Its multiscale design is inspired by Protein Autoregressive Modeling (PAR) (Qu et al., 2026).

Our contributions are:

- To combine the complementary strengths of Cartesian and internal-coordinate representations, we first generate a coarse holo-like scaffold as sparse Cartesian anchors with an EGNN, then recursively complete the intervening residues with a Transformer in anchor-relative local internal coordinates.
- By representing each midpoint relative to its generated parent anchors and the apo reference, the refinement stage introduces an inductive bias toward physically plausible local geometry.
- On the fixed 150-pair D3PM test split, CrypticFlow achieves the lowest median and mean Cα RMSD among the compared methods across five independently trained models. It also achieves the strongest local-geometry fidelity among the evaluated outputs and the lowest MolProbity score among the evaluated baselines after cg2all reconstruction.

2 Related Work

Apo-to-holo transport. D3PM catalogs protein structural pairs, including apo–holo pairs, with motions from local residues to whole domains (Peng et al., 2022). Aligned Diffusion Schrödinger Bridges (SBAlign) learns transport from paired observations (Somnath et al., 2023). Augmented Bridge Matching (ABM) provides a paired transport formulation used as a baseline (De Bortoli et al., 2023). Fractional Diffusion Bridge Models (FDBM) extend this framework from Brownian to approximate fractional Brownian noise, introducing temporal correlations (Nobis et al., 2025). SBAlign, ABM, and FDBM model Cartesian Cα coordinates. Sesame instead applies flow matching to backbone SE(3) frames, combining rotations and Cartesian translations (Miñán et al., 2025). SBAlign, FDBM, and Sesame evaluate apo-to-holo prediction on D3PM. DynamicFlow generates ligand-conditioned, full-atom pocket ensembles (Zhou et al., 2025).

Equivariant Cartesian models. E(n)-equivariant graph neural networks update coordinates with invariant scalars and relative vectors (Satorras et al., 2021). This construction preserves the transformation law of a 3D structure while learning geometric messages. ESM-3 is a pretrained multimodal protein language model (PLM) that learns contextual representations over protein sequence, structure, and function tokens (Hayes et al., 2025). We use ESM-3 representations precomputed only from the apo full sequence: residue-level embeddings provide sequence context to both the anchor and midpoint models, while the beginning-of-sequence (BOS) embedding provides a global condition to the anchor model.

Transformers for conditional generation. Transformers use multi-head self-attention to model interactions among tokens (Vaswani et al., 2017). Diffusion Transformers (DiT) adapt Transformer blocks to generative denoising, with adaptive layer normalization providing a mechanism for timestep and class conditioning (Peebles & Xie, 2023).

Internal coordinates and multiscale generation. FoldingDiff generates protein backbones through torsional variables (Wu et al., 2024). Int2Cart shows that correlations among internal coordinates affect Cartesian reconstruction quality (Li et al., 2023). PAR replaces residue-wise autoregression with a sequence of successively finer protein scales (Qu et al., 2026). We combine these ideas differently. The top-level anchor state is represented in Cartesian space. At each finer level, every

midpoint is defined relative to two already generated parent anchors in a local internal-coordinate system. We evaluate the completed Cα trace directly. After cg2all backmapping (Heo & Feig, 2024), we assess the reconstructed all-atom structure with MolProbity (Chen et al., 2010).

3 Method

3.1 Problem Formulation and Hierarchical Architecture Overview

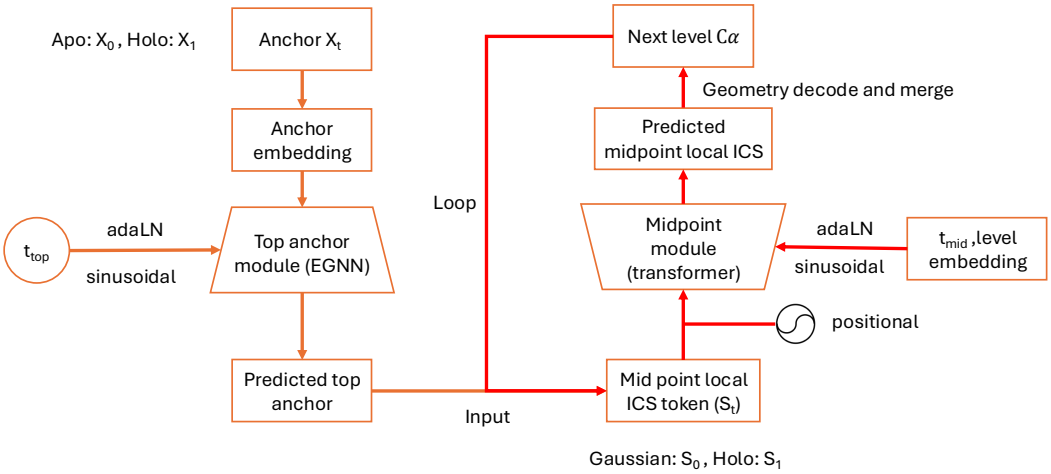


Figure 1: Overview of CrypticFlow architecture. The top-anchor module (EGNN) transports apo anchors to holo-like anchor positions in Cartesian coordinates. The midpoint module (Transformer) generates midpoints from Gaussian noise in local internal coordinates defined by their parent anchors and an apo reference direction. The geometry decoder converts these local internal-coordinate states into Cartesian midpoint coordinates. Red arrows show the recursive loop: each decoded midpoint joins the partial trace and serves as an anchor at the next level until the trace is complete. Training uses paired clean targets and teacher-forced parent-anchor contexts.

Let the residue-matched apo and holo Cα traces be $\mathbf{X}_{\text{apo}}, \mathbf{X}_{\text{holo}} \in \mathbb{R}^{L\times 3}$, and let the residue features formed by concatenating D3PM descriptors with full-sequence ESM-3 residue embeddings be $\mathbf{F}$. The top-anchor EGNN additionally uses the ESM-3 BOS embedding as a global condition. We learn a conditional model that predicts $\mathbf{X}_{\text{holo}}$ from $\mathbf{X}_{\text{apo}}$ and $\mathbf{F}$, without ligand input. CrypticFlow separates global motion from local geometry: a Cartesian flow first moves sparse anchors, and a local-coordinate flow then fills the gaps between them. Figure 1 shows the two stages. The complete training and recursive inference procedure is given in Appendix L.

D3PM DB contains only residues shared by the apo and holo structures. Removing unmatched residues can introduce gaps in the retained trace, so a single protein chain may be represented as multiple contiguous segments. We treat these gaps as chain breaks and do not allow refinement intervals to cross them. Within each segment, we select the first and last residues, together with every s-th interior residue, as the top anchor set. We use $s = 4$. After generating these anchors, we split every unresolved interval $[i, j]$ at $m = \lfloor (i + j)/2 \rfloor$, where $\lfloor \cdot \rfloor$ rounds down to the nearest integer. All midpoints at the same recursion level are generated together. They are then merged into the partial trace and become parents at the next level. A full stride-4 interval contains three missing residues and is completed in two midpoint levels. Let the top anchor set be $\mathcal{A}$, the midpoints inserted at level ℓ be $\mathcal{M}_\ell$, and the anchor residues already placed before level ℓ be $\mathcal{P}_\ell = \mathcal{A} \cup \bigcup_{h<\ell} \mathcal{M}_h$. Thus, $\mathcal{P}_\ell$ contains the original top anchors and all midpoints from earlier levels. Writing $\mathbf{X}$ for a candidate holo trace, the hierarchy gives

$$p_\theta(\mathbf{X} \mid \mathbf{X}_{\text{apo}}, \mathbf{F}) = p_{\text{top}}(\mathbf{X}_{\mathcal{A}} \mid \mathbf{X}_{\text{apo}}, \mathbf{F}) \prod_{\ell=1}^{D} p_{\text{mid}}(\mathbf{X}_{\mathcal{M}_\ell} \mid \mathbf{X}_{\mathcal{P}_\ell}, \mathbf{X}_{\text{apo}}, \mathbf{F}, \ell), \tag{1}$$

where the number of refinement levels is D. Each midpoint factor is a joint conditional distribution over a level, not a product of independent residue predictions. Generation is sequential across levels and parallel within each level.

3.2 Clean-Endpoint Conditional Flow

We use the linear conditional path of flow matching (Lipman et al., 2023) and parameterize both networks to predict the clean endpoint. For a source state $\mathbf{z}_0$, target state $\mathbf{z}_1$, and conditioning information $\mathbf{c}$,

$$\mathbf{z}_t = (1-t)\mathbf{z}_0 + t\mathbf{z}_1, \qquad t \in [0, 1). \tag{2}$$

Instead of predicting velocity directly, each network uses a clean-endpoint predictor, f_θ, to estimate the target state:

$$\mathcal{L}_{\text{clean}} = \mathbb{E}_{\mathbf{z}_0, \mathbf{z}_1, \mathbf{c}, t}\left[\|f_\theta(\mathbf{z}_t, t, \mathbf{c}) - \mathbf{z}_1\|_2^2\right]. \tag{3}$$

For the path in Equation 2, this prediction defines a velocity field

$$\widehat{\mathbf{v}}_\theta(\mathbf{z}_t, t, \mathbf{c}) = \frac{f_\theta(\mathbf{z}_t, t, \mathbf{c}) - \mathbf{z}_t}{\max(1-t, \epsilon)}. \tag{4}$$

Here the positive numerical stabilizer is $\epsilon > 0$. The anchor flow starts from apo anchor coordinates and targets their holo coordinates. The midpoint flow starts from Gaussian noise and targets a normalized holo local-coordinate state defined in Section 3.5. The networks thus share a training principle but use different representations and source distributions.

3.3 Coarse Cartesian Anchor Generation

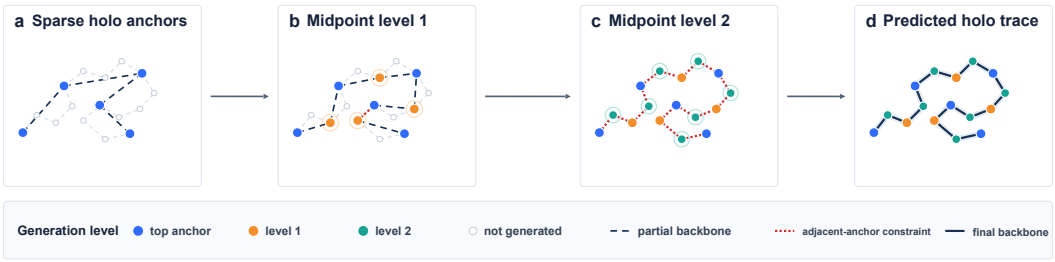


Figure 2: Hierarchical coarse-to-fine Cα generation. A Cartesian flow first predicts sparse stride-4 holo anchors. The midpoint Transformer inserts one midpoint in each active parent interval at each recursive level. Two midpoint levels fill full stride-4 intervals, while adjacent-anchor distance constraints help preserve local geometry in the completed trace.

The top-anchor EGNN predicts clean holo anchor coordinates from $\mathbf{X}_t$ in eight layers. Initial node features combine D3PM descriptors and ESM-3 residue embeddings; a shared condition combines flow time with the ESM-3 BOS embedding. For each central anchor i, the graph takes the union of neighbors j selected by apo-space proximity, current-state proximity, and sparse-sequence adjacency. The neighbor list is fixed within each forward pass through the eight-layer top-anchor EGNN, while current-state distances and geometric edge features are recomputed at every EGNN layer. During inference, the graph is rebuilt at each Euler step.

At each layer, the network decides how each neighbor should influence an anchor's movement. For anchor i and neighbor j, it uses the vector from i to j in the current coordinates, $\mathbf{r}_{ij}^{\text{cur}} = \mathbf{X}_j^{(k)} - \mathbf{X}_i^{(k)}$, and the corresponding vector in the apo structure, $\mathbf{r}_{ij}^{\text{apo}} = \mathbf{X}_{\text{apo},j} - \mathbf{X}_{\text{apo},i}$. At layer k, their soft-normalized forms, $\bar{\mathbf{r}}_{ij}^{\text{cur}}$ and $\bar{\mathbf{r}}_{ij}^{\text{apo}}$, are

$$\bar{\mathbf{r}}_{ij}^{\text{cur}} = \frac{\mathbf{r}_{ij}^{\text{cur}}}{\|\mathbf{r}_{ij}^{\text{cur}}\| + 1}, \qquad \bar{\mathbf{r}}_{ij}^{\text{apo}} = \frac{\mathbf{r}_{ij}^{\text{apo}}}{\|\mathbf{r}_{ij}^{\text{apo}}\| + 1}. \tag{5}$$

The corresponding current and apo distances are $d_{ij}^{\text{cur}} = \|\mathbf{r}_{ij}^{\text{cur}}\|$ and $d_{ij}^{\text{apo}} = \|\mathbf{r}_{ij}^{\text{apo}}\|$. Each raw distance d becomes a soft-normalized magnitude $d/(d+1) < 1$; this transformation smoothly

suppresses near-coincident pairs and bounds long-range contributions. At each layer, an edge MLP forms an invariant message $\mathbf{m}_{ij}^{(k)}$ from the conditioned node features of nodes i and j and their edge geometry. This message determines a sigmoid gate $q_{ij}^{(k)}$, which controls the edge contribution, and signed weights $a_{ij}^{(k)}, b_{ij}^{(k)}$, which scale the two soft-normalized vectors. Averaging these contributions over the neighbor set $\mathcal{N}(i)$ gives the coordinate update:

$$\mathbf{X}_i^{(k+1)} = \mathbf{X}_i^{(k)} + \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} q_{ij}^{(k)} \left(a_{ij}^{(k)} \bar{\mathbf{r}}_{ij}^{\text{cur}} + b_{ij}^{(k)} \bar{\mathbf{r}}_{ij}^{\text{apo}} \right). \tag{6}$$

The weights are invariant to a joint rigid transformation of the current and apo structures, while the relative vectors transform with them. The update therefore follows the same transformation: it does not depend on the chosen coordinate frame. Therefore, every coordinate-update layer is $\mathrm{E}(3)$-equivariant. If the input structures are jointly translated, rotated, or reflected, the output transforms in the same way. Appendix G evaluates the composed generator under random proper rotations and translations and separately tests reflection.

After eight layers, let the anchor-validity mask be $m_i \in \{0, 1\}$ and the final displacement of anchor i be $\mathbf{D}_i = \mathbf{X}_i^{(8)} - (\mathbf{X}_t)_i$. The network removes the masked mean displacement, $\bar{\mathbf{D}} = \sum_i m_i \mathbf{D}_i / \sum_i m_i$, and returns the clean anchor estimate, $\widehat{\mathbf{X}}_{1,i} = (\mathbf{X}_t)_i + \mathbf{D}_i - \bar{\mathbf{D}}$, for each valid anchor. This suppresses uniform translation while preserving relative conformation and the current anchor centroid. Appendix A.1 details the top-anchor EGNN architecture, graph construction, and node and coordinate updates.

3.4 Local Internal Coordinates and Geometry Decoding

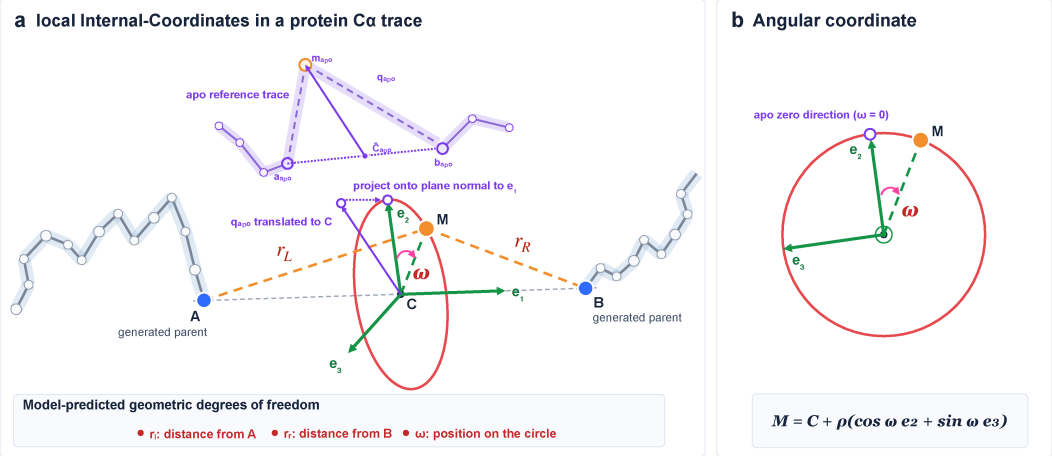


Figure 3: Local internal-coordinate geometry. (a) Parent anchors $\mathbf{A}$ and $\mathbf{B}$ define a sphere-intersection circle. The apo midpoint direction $\mathbf{q}_{\text{apo}}$ is translated to $\mathbf{C}$ and projected onto the plane normal to $\mathbf{e}_1$ to define $\mathbf{e}_2$; $\mathbf{e}_3$ completes the local frame. Predicted r_L, r_R, and ω place $\mathbf{M}$ on the circle. (b) Viewed along $+\mathbf{e}_1$, ω is measured from the apo-defined $+\mathbf{e}_2$ direction. Sequence-adjacent radii are fixed to the corresponding apo Cα–Cα distances, reducing the predicted degrees of freedom from three to two or one.

Let the left and right parent anchors be $\mathbf{A}$ and $\mathbf{B}$, and let their generated sequence midpoint be $\mathbf{M}$. Rather than predict an unconstrained Cartesian displacement, we represent $\mathbf{M}$ by its parent distances $r_L = \|\mathbf{M} - \mathbf{A}\|$, $r_R = \|\mathbf{M} - \mathbf{B}\|$, and an angle ω around the parent axis. For distinct parents, define

$$d = \|\mathbf{B} - \mathbf{A}\|, \quad \mathbf{e}_1 = \frac{\mathbf{B} - \mathbf{A}}{d}, \quad \eta = \frac{d^2 + r_L^2 - r_R^2}{2d}, \quad \mathbf{C} = \mathbf{A} + \eta \mathbf{e}_1. \tag{7}$$

Here the signed axial distance from $\mathbf{A}$ is η, the sphere-intersection circle center is $\mathbf{C}$, and the parent-axis direction is $\mathbf{e}_1$. The circle radius is $\rho = \sqrt{\max(r_L^2 - \eta^2, 0)}$. When the two distance constraints are feasible, their intersection is a circle centered at $\mathbf{C}$ with radius ρ.

The apo trace supplies the reference direction for ω. The apo coordinates of the residues represented by $\mathbf{A}, \mathbf{B}, \mathbf{M}$ are $\mathbf{A}_{\text{apo}}, \mathbf{B}_{\text{apo}}, \mathbf{M}_{\text{apo}}$, respectively. In particular, $\mathbf{M}_{\text{apo}}$ is the observed apo position of the sequence-midpoint residue, whereas $\mathbf{M}$ is its generated position. We denote the apo parents' Cartesian midpoint by $\bar{\mathbf{C}}_{\text{apo}} = (\mathbf{A}_{\text{apo}} + \mathbf{B}_{\text{apo}})/2$ and define

$$\mathbf{q}_{\text{apo}} = \mathbf{M}_{\text{apo}} - \bar{\mathbf{C}}_{\text{apo}}, \qquad \mathbf{e}_2 = \frac{(\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^\top)\mathbf{q}_{\text{apo}}}{\|(\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^\top)\mathbf{q}_{\text{apo}}\|}, \qquad \mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2. \tag{8}$$

Here the 3×3 identity matrix is $\mathbf{I}$, the apo reference vector is $\mathbf{q}_{\text{apo}}$, the zero-angle direction is $\mathbf{e}_2$, and the third frame direction is $\mathbf{e}_3$. Geometrically, we move the tail of $\mathbf{q}_{\text{apo}}$ to $\mathbf{C}$ without changing its direction or length, then project it onto the plane through $\mathbf{C}$ with normal $\mathbf{e}_1$. Normalizing the projected vector gives $\mathbf{e}_2$, the zero-angle direction on the circle. If the projected apo vector has zero or near-zero norm, it cannot provide a stable direction for $\mathbf{e}_2$. We instead project a deterministically selected least-aligned fallback axis onto the same plane and normalize the result. Appendix B.4 explains this construction and the limitation of using fixed Cartesian axes. The Cartesian midpoint is

$$\mathbf{M} = \mathbf{C} + \rho(\cos\omega\,\mathbf{e}_2 + \sin\omega\,\mathbf{e}_3). \tag{9}$$

For fixed parents, an angular change moves only this midpoint along its local circle. Figure 2 visualizes the two-level stride-4 refinement, and Figure 3 illustrates the frame used for local internal coordinates. Appendix B derives the circle and local frame step by step, including why translating the apo reference vector to $\mathbf{C}$ and projecting it onto the circle plane is equivalent to removing its axial component.

Adjacent-distance constraints. Adjacent Cα–Cα distances are tightly constrained by backbone connectivity and change little between matched apo and holo conformations, as quantified in Appendix B.6. The main configuration replaces a learned radius spanning residues i and $i+1$ with the corresponding apo distance as an explicit local prior. Let the sequence gaps from midpoint m to its left and right parents i, j be $n_L = m - i$ and $n_R = j - m$, respectively. If either gap is one, the corresponding radius is fixed to the sanitized apo Cα–Cα distance:

$$n_L = 1 \Rightarrow r_L = d^{\text{apo}}_{\mathbf{A},\mathbf{M}}, \qquad n_R = 1 \Rightarrow r_R = d^{\text{apo}}_{\mathbf{M},\mathbf{B}}. \tag{10}$$

Here the sanitized apo distances for the left and right adjacent pairs are $d^{\text{apo}}_{\mathbf{A},\mathbf{M}}$ and $d^{\text{apo}}_{\mathbf{M},\mathbf{B}}$, respectively. Sanitation retains finite apo distances in $[2.5, 5.0)$ Å and replaces other values with 3.80 Å. A midpoint therefore follows a *3/2/1-DOF policy*, with three, two, or one learned geometric variables when zero, one, or both parent radii are fixed. Across all 1,591 D3PM pairs finds that 99.8202% of 577,291 within-segment consecutive Cα pairs change by less than 0.2 Å between apo and holo structures (Appendix B.6).

Sphere feasibility. The radii and parent separation must satisfy

$$|r_L - r_R| \le d \le r_L + r_R. \tag{11}$$

For an infeasible non-final midpoint, the predicted spheres do not intersect; we keep the parent anchors fixed and minimally adjust only the learnable radii to form a valid intersection circle. If Equation 10 fixes one radius to an apo adjacent distance, only the other is adjusted. At a final leaf, both radii are fixed to apo adjacent distances and remain unchanged. If the two fixed radii are incompatible with the generated parents, their spheres do not intersect. The local-ICS geometry decoder then sets the intersection-circle radius to zero and places the midpoint at $\mathbf{C} = \mathbf{A} + \eta\mathbf{e}_1$ on the parent axis. This finite numerical fallback is not a feasibility repair, and the query remains marked infeasible. Appendix B.2 gives the exact case-wise corrections. In five-seed free test rollouts, $1.882 \pm 0.486\%$ of midpoint queries are raw-infeasible before non-final repair; Appendix B.7 reports the seed-wise counts and final-leaf events.

3.5 Recursive Midpoint Generation

The midpoint Transformer generates the local states for all active intervals at a level. Each midpoint local state, $\mathbf{z}^{\text{mid}} \in \mathbb{R}^4$, contains the left and right radial channels, u_L, u_R, and the angular channels, $\sin\omega, \cos\omega$, so $\mathbf{z}^{\text{mid}} = (u_L, u_R, \sin\omega, \cos\omega)$. The midpoint Transformer outputs local-ICS channels, whose radial components are mapped to bounded radii, while the angular pair avoids the discontinuity at $\pm\pi$. To balance the scales of the four local-ICS channels, flow matching is performed on the

channel-wise standardized state, $\widetilde{\mathbf{z}} = (\mathbf{z} - \boldsymbol{\mu}_{\text{train}})/\boldsymbol{\sigma}_{\text{train}}$, where the training-target channel mean and standard deviation are $\boldsymbol{\mu}_{\text{train}}$ and $\boldsymbol{\sigma}_{\text{train}}$. Before geometry decoding, predicted states are transformed back to raw local coordinates.

Each query concatenates its noisy local state, parent-distance features, apo local state, D3PM residue features, and ESM-3 embeddings. Each query token receives a sinusoidal residue-position embedding. Flow-time and recursion-level embeddings condition each Transformer block through adaptive layer normalization (AdaLN). A four-channel head predicts the clean local state; architectural details appear in Appendix A.2.

At each level, the local-ICS geometry decoder maps the generated local internal-coordinate states to Cartesian midpoint coordinates and merges them with the existing parent anchors to update the partial trace. The updated trace supplies the parent anchors for the next level, and this generate–decode–merge cycle repeats recursively until the Cα trace is complete.

3.6 Training and Inference

Training. Paired apo–holo structures provide clean targets for both flows. Midpoint training uses teacher-forced parent contexts; inference instead uses the parents generated at earlier levels. The anchor clean-endpoint loss, $\mathcal{L}_{\text{top}}$, and midpoint clean-endpoint loss, $\mathcal{L}_{\text{mid}}$, apply the objective to anchor coordinates and local midpoint states, respectively. Radius channels fixed by Equation 10 are excluded from the midpoint state loss. The full objective is

$$\mathcal{L} = \mathcal{L}_{\text{top}} + \mathcal{L}_{\text{mid}} + \alpha_s(0.2\mathcal{L}_{\text{xyz}} + 0.1\mathcal{L}_{\text{adj}} + 0.1\mathcal{L}_{\text{sparse}} + 0.05\mathcal{L}_{\text{feas}})\,. \tag{12}$$

The auxiliary losses penalize errors in decoded midpoint coordinates, $\mathcal{L}_{\text{xyz}}$; adjacent distances, $\mathcal{L}_{\text{adj}}$; sparse pair distances, $\mathcal{L}_{\text{sparse}}$; and sphere infeasibility, $\mathcal{L}_{\text{feas}}$. The auxiliary-loss scale, α_s, increases linearly from zero to one over the first 500 optimization steps. Flow times are drawn in equal proportions from $t = 0$, the five-point Euler grid, and a continuous uniform distribution on $[0, 1)$. Appendix D gives the exact masks, reductions, and Smooth-L1 definitions used by every objective term.

Inference. The anchor flow first transforms apo anchor coordinates into holo-like anchors. These anchors initialize a partial trace that is recursively completed by the midpoint flow. Unlike teacher-forced training, inference builds the hierarchy from its own predictions: each decoded midpoint is added to the partial trace and becomes a fixed parent anchor for the next level. Each flow uses five Euler steps, and evaluation draws one prediction per test protein.

4 Experimental Setup

Data. We use the fixed D3PM-Large split introduced by SBAlign: 1,291 apo–holo pairs for training, 150 for validation, and 150 for testing (Somnath et al., 2023; Peng et al., 2022). We center and Kabsch-align each pair during preprocessing (Kabsch, 1976); segment labels prevent midpoint generation across mask-defined gaps.

Full-sequence ESM-3 features. We run ESM-3 on the full apo SEQRES sequence and gather embeddings for observed D3PM residues, without holo input. Appendix F gives the windowing, residue-mapping, and validation details.

Model and training. Complete architecture settings are reported in Appendix C. We train for 1,000 epochs with AdamW (Loshchilov & Hutter, 2019), an effective batch size of eight, and top and midpoint learning rates of 10^{-4} and 5×10^{-4}. For each seed in $\{1, 3, 42, 88, 99\}$, we evaluate the checkpoint with the lowest validation mean Kabsch RMSD and report mean $\pm$ sample standard deviation. Appendices I and E report model selection and computational cost.

Baselines and protocol. For global RMSD, we use published D3PM results for SBAlign and Sesame and FDBM-reported results for EGNN, ABM, and FDBM (Somnath et al., 2023; Miñán et al., 2025; Nobis et al., 2025). Local and all-atom analyses use stored predictions for direct-regression EGNN, SBAlign, ABM, and FDBM on the same 150 test identifiers; Sesame is omitted because

reproducible per-target outputs were unavailable. We report one prediction from each of five training seeds.

Metrics. Following the D3PM protocol (Somnath et al., 2023; Peng et al., 2022), we measure RMSD directly in the pre-aligned frame without a prediction-to-target fit and report success below 2, 5, and 10 Å. Local fidelity uses 0.2 Å distance and 20° pseudo-angle thresholds; Cα clashes are non-adjacent pairs below 3 Å per million eligible pairs.

Relaxation-free all-atom validation. Fixed-Cα cg2all reconstructs all predictions, and MolProbity scores the unrelaxed outputs (Heo & Feig, 2024; Chen et al., 2010). All reconstructions succeed; GT-holo applies the same pipeline to reference Cα traces.

5 Results

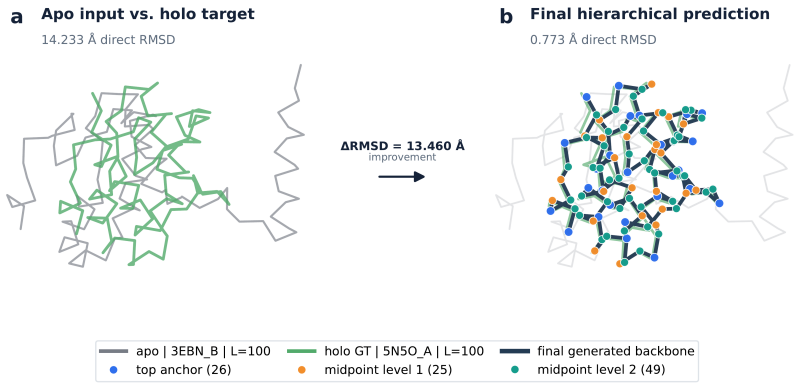


Figure 4: Example for the 100-residue pair 3EBN_B–5N5O_A. (a) Apo input (gray) and holo target (green), with 14.233 Å direct Cα RMSD. (b) The prediction (dark blue) reaches 0.773 Å RMSD; markers show 26 top anchors (blue), 25 level-1 midpoints (orange), and 49 level-2 midpoints (teal).

Figure 4 visualizes how sparse anchors and two midpoint levels resolve a large apo–holo displacement. A corresponding hierarchical inferance animation is provided in the supplementary archive under `videos/crypticflow_hierarchical_inference`.

5.1 Apo-to-Holo Accuracy

Table 1: Cα accuracy on the fixed D3PM test split. Ours is mean ± sample standard deviation over five models; baseline values are taken from the cited reports (Somnath et al., 2023; Miñán et al., 2025; Nobis et al., 2025). Test std. is the dispersion across test proteins; our row summarizes this statistic across five models.

	RMSD (Å) ↓			Success rate (%) ↑		
Method	Median	Mean	Test std.	< 2 Å	< 5 Å	< 10 Å
EGNN (reported)	19.99	21.37	8.21	1.0	1.0	3.0
SBAlign (10, 10)	3.80	4.98	3.95	0.0	69.0	93.0
SBAlign (100, 100)	3.81	5.02	3.96	0.0	70.0	93.0
Sesame	2.87	3.65	2.95	38.0	82.0	96.0
ABM	2.40	3.49	3.54	43.0	84.0	96.0
FDBM ($H = 0.3$)	2.33	3.42	3.42	43.0	85.0	97.0
FDBM ($H = 0.2$)	2.12	3.34	3.59	48.0	**86.0**	96.0
FDBM ($H = 0.1$)	2.20	3.44	3.57	46.0	83.0	97.0
CrypticFlow (ours)	**1.333 ±0.056**	**2.766 ±0.041**	3.646 ±0.026	**58.53 ±0.73**	84.40 ±1.12	96.53 ±0.30

5.2 Cα and All-Atom Structural Validity

Table 2: Core validity metrics. CrypticFlow reports mean ± sample standard deviation over five models; baselines are single stored predictions.

Method	Dist. ≤ 0.2 Å (%) ↑	Angle $\leq 20°$ (%) ↑	MP score ↓
GT-holo Cα control	100.0000	100.0000	2.291
Direct-regression EGNN	8.7252	37.8666	4.861
SBAlign	11.5265	46.1836	4.624
FDBM	32.0681	82.1678	4.085
ABM	30.9050	82.4563	4.089
CrypticFlow (ours)	$\mathbf{99.5825 \pm 0.0200}$	$\mathbf{95.8652 \pm 0.2543}$	$\mathbf{3.034 \pm 0.054}$

Tables 1 and 2 show gains in both global and local quality. Relative to FDBM ($H = 0.2$), CrypticFlow lowers median/mean direct RMSD from 2.12/3.34 Å to 1.333 ± 0.056/2.766 ± 0.041 Å and raises sub-2 Å success from 48.0% to $58.53 \pm 0.73\%$. It also achieves $99.58 \pm 0.02\%$ distance fidelity, $95.87 \pm 0.25\%$ angle fidelity, and the lowest learned-model MolProbity score (3.034 ± 0.054). Detailed metrics appear in Appendix H.

6 Discussion and Limitations

Validation controls (Appendix J) show that ESM-3 conditioning and the selected time sampler improve global accuracy. Local ICS chiefly preserves adjacent distances and reconstructed bond/angle geometry; a local-Cartesian control is slightly better on RMSD and some other metrics. Fixed apo radii remove one learned radial degree of freedom for each adjacent parent and provide a deterministic boundary condition, although their measured aggregate gain is modest. The shallow hierarchy restricts but cannot eliminate error propagation: top-anchor errors can reach descendants, but stride four limits every dependency path to at most two midpoint refinements. Full-chain controls and perturbation diagnostics in Appendix K further illustrate this difference from a chain-wide internal-coordinate rollout.

Adjacent-distance and pseudo-angle scores measure fidelity to the paired holo reference, not physical validity by themselves; clashes and relaxation-free cg2all–MolProbity provide complementary evidence. Our 34.08M parameter model is also substantially larger than the published bridge baselines, so the comparison is not capacity matched. Despite its larger size, however, its per-target inference time is shorter than those of all three bridge baselines under the hardware-controlled comparison in Appendix Table 5. Finally, CrypticFlow produces one ligand-free Cα trace for the paired target, rather than an energy-minimized, ligand-specific ensemble. D3PM does not evaluate pocket recovery or docking utility.

7 Conclusion

We introduced CrypticFlow, a hierarchical flow-matching model that predicts sparse anchors in Cartesian space and completes the trace with recursive local internal coordinates. Across five models on the fixed 150-pair D3PM split, it achieves the lowest compared median and mean direct Cα RMSD while retaining high local-geometry fidelity and the lowest compared cg2all–MolProbity score. Full-chain controls support shallow, anchor-bounded refinement as a way to restrict lever-arm amplification. These results establish CrypticFlow as a strong apo-conditioned holo-like predictor; evaluation on external cryptic-pocket datasets and ligand-conditioned all-atom generation remain future work. More broadly, the framework could be extended to model conformational changes induced by protein–protein interactions or post-translational modifications. Its hierarchical hybrid-coordinate design could also provide a useful structural bias for unconditional protein generation, potentially improving physical plausibility over purely Cartesian generation.

Reproducibility statement

Sections 3 and 4 specify the fixed split, representations, model dimensions, training objective, inference procedure, and evaluation metrics. Appendix C records the complete main configuration and seed protocol; Appendices I, J, K, and F document model selection, post-selection controls, perturbation diagnostics, and sequence mapping. The submitted anonymous supplementary archive contains the implementation, fixed split manifest, five sanitized training/evaluation configuration pairs, per-target metric records, and a standard-library script that recomputes the CrypticFlow rows in Tables 1 and 2. Its `README.md` and `REPRODUCIBILITY.md` document the environment, data schema, training, inference, and all-atom evaluation procedures. Upstream D3PM structures, ESM-3 assets, and large checkpoints are not redistributed; the archive documents their preparation and the commands needed to retrain or evaluate a checkpoint.

AI use statement

Generative AI tools, including OpenAI Prism and Codex, assisted with literature discovery; manuscript drafting, editing, and translation; method implementation and software review; dataset validation and formatting; experimental design; figure and artifact development; and interpretation of results. They were not used to generate synthetic datasets, formulate or prove mathematical claims, or propose scientific hypotheses; qualitative and thematic data analysis was not applicable. The authors manually verified the cited literature, implementation, experimental artifacts, numerical claims, and final wording, and take responsibility for the complete content of the submission.

References

Vincent B. Chen, W. Bryan Arendall, III, Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson, and David C. Richardson. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(1):12–21, 2010. doi: 10.1107/S0907444909042073.

Valentin De Bortoli, Guan-Horng Liu, Tianrong Chen, Evangelos A. Theodorou, and Weilie Nie. Augmented bridge matching. *arXiv preprint arXiv:2311.06978*, 2023. URL `https://arxiv.org/abs/2311.06978`.

Thomas Hayes, Roshan Rao, Halil Akin, Nicholas J. Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, et al. Simulating 500 million years of evolution with a language model. *Science*, 387 (6736):850–858, 2025. doi: 10.1126/science.ads0018.

Lim Heo and Michael Feig. One bead per residue can describe all-atom protein structures. *Structure*, 32(1):97–111.e6, 2024. doi: 10.1016/j.str.2023.10.013.

Pasi A. Jänne, Gregory J. Riely, Shirish M. Gadgeel, Rebecca S. Heist, Sai-Hong Ignatius Ou, Jessica M. Pacheco, et al. Adagrasib in non–small-cell lung cancer harboring a KRASG12C mutation. *The New England Journal of Medicine*, 387(2):120–131, 2022. doi: 10.1056/NEJMoa2204619.

Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, 1976. doi: 10.1107/S0567739476001873.

Jie Li, Oufan Zhang, Seokyoung Lee, Ashley Namini, Zi Hao Liu, João M. C. Teixeira, Julie D. Forman-Kay, and Teresa Head-Gordon. Learning correlations between internal coordinates to improve 3d cartesian coordinates for proteins. *Journal of Chemical Theory and Computation*, 19 (14):4689–4700, 2023. doi: 10.1021/acs.jctc.2c01270.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. URL `https://openreview.net/forum?id=PqvMRDCJT9t`.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=Bkg6RiCqY7`.

Artur Meller, Michael Ward, Jonathan Borowsky, Meghana Kshirsagar, Jeffrey M. Lotthammer, Felipe Oviedo, Juan Lavista Ferres, and Gregory R. Bowman. Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network. *Nature Communications*, 14: 1177, 2023. doi: 10.1038/s41467-023-36699-3.

Raúl Miñán, Carles Perez-Lopez, Javier Iglesias, Álvaro Ciudad, and Alexis Molina. Sesame: Opening the door to protein pockets. *arXiv preprint arXiv:2509.05302*, 2025. URL https://arxiv.org/abs/2509.05302.

Gabriel Nobis, Maximilian Springenberg, Arina Belova, Rembert Daems, Christoph Knochenhauer, Manfred Opper, Tolga Birdal, and Wojciech Samek. Fractional diffusion bridge models. In *Advances in Neural Information Processing Systems 38*, 2025. doi: 10.52202/085713-3819. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/a66caa1703fe34705a4368c3014c1966-Abstract-Conference.html.

Jonathan M. Ostrem, Ulf Peters, Martin L. Sos, James A. Wells, and Kevan M. Shokat. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature*, 503(7477):548–551, 2013. doi: 10.1038/nature12796.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4195–4205, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Peebles_Scalable_Diffusion_Models_with_Transformers_ICCV_2023_paper.html.

Cheng Peng, Xinben Zhang, Zhijian Xu, Zhaoqiang Chen, Yanqing Yang, Tingting Cai, and Weiliang Zhu. D3PM: A comprehensive database for protein motions ranging from residue to domain. *BMC Bioinformatics*, 23:70, 2022. doi: 10.1186/s12859-022-04595-0.

Yanru Qu, Cheng-Yen Hsieh, Zaixiang Zheng, Ge Liu, and Quanquan Gu. Protein autoregressive modeling via multiscale structure generation. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2602.04883. Oral presentation.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. $E(n)$ Equivariant Graph Neural Networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 2021. URL https://proceedings.mlr.press/v139/satorras21a.html.

Lakshaditya Singh, Adwait Shelke, and Divyansh Agrawal. Torsion-space diffusion for protein backbone generation with geometric refinement. *arXiv preprint arXiv:2511.19184*, 2025. URL https://arxiv.org/abs/2511.19184.

Ferdinand Skoulidis, Bob T. Li, Grace K. Dy, Timothy J. Price, Gerald S. Falchook, Jürgen Wolf, et al. Sotorasib for lung cancers with KRAS p.G12C mutation. *The New England Journal of Medicine*, 384(25):2371–2381, 2021. doi: 10.1056/NEJMoa2103695.

Vignesh Ram Somnath, Matteo Pariset, Ya-Ping Hsieh, Maria Rodriguez Martinez, Andreas Krause, and Charlotte Bunne. Aligned Diffusion Schrödinger Bridges. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pp. 1985–1995. PMLR, 2023. URL https://proceedings.mlr.press/v216/somnath23a.html.

Sandor Vajda, Dmitri Beglov, Amanda E. Wakefield, Matthew Egbert, and Adrian Whitty. Cryptic binding sites on proteins: Definition, detection, and druggability. *Current Opinion in Chemical Biology*, 44:1–8, 2018. doi: 10.1016/j.cbpa.2018.05.003.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Kevin E. Wu, Kevin K. Yang, Rianne van den Berg, Sarah Alamdari, James Y. Zou, Alex X. Lu, and Ava P. Amini. Protein structure generation via folding diffusion. *Nature Communications*, 15:1059, 2024. doi: 10.1038/s41467-024-45051-2.

Xiangxin Zhou, Yi Xiao, Haowei Lin, Xinheng He, Jiaqi Guan, Yang Wang, Qiang Liu, Feng Zhou, Liang Wang, and Jianzhu Ma. Integrating protein dynamics into structure-based drug design via full-atom stochastic flows. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/2eea2b3ce7266e83785692cebbe1381d-Abstract-Conference.html.

Supplementary Benchmark Visualizations

Figures 5 and 6 visualize the results reported in main Tables 1 and 2, respectively.

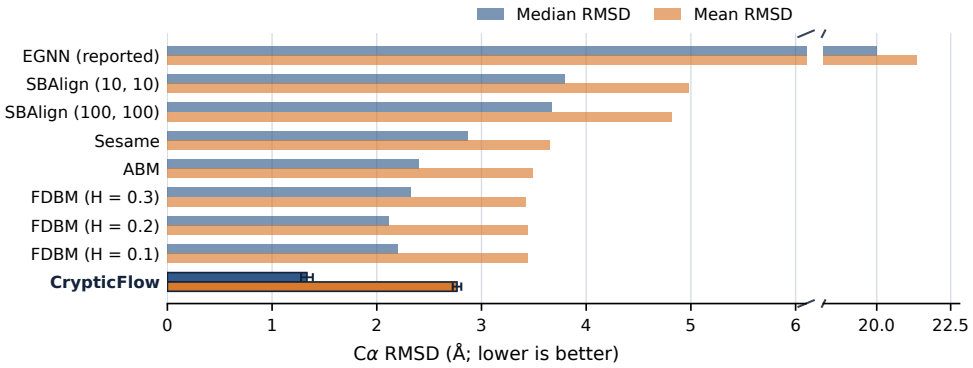


Figure 5: Median and mean direct Cα RMSD on the fixed D3PM test split after the dataset's apo-to-holo preprocessing alignment. The broken axis retains legacy EGNN; CrypticFlow error bars show sample standard deviations over five models. Baseline values are taken from the cited reports.

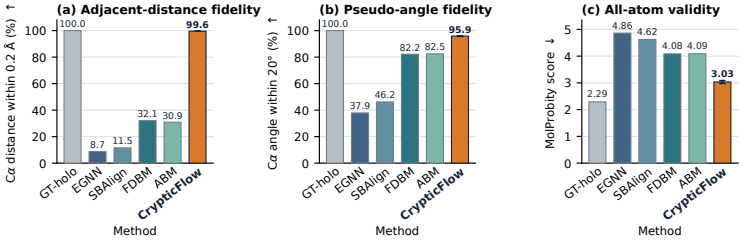


Figure 6: Local and all-atom validity. Error bars show sample standard deviations over five CrypticFlow models. Sesame is omitted because reproducible per-target predictions were unavailable.

A Detailed Network Architectures

A.1 Anchor Network Details

Each anchor receives a 145-dimensional D3PM descriptor and a 1,536-dimensional ESM-3 residue representation. A projection produces the initial invariant node state $\mathbf{H}^{(0)} \in \mathbb{R}^{|\mathcal{A}|\times 320}$. A sinusoidal time embedding and the projected ESM-3 BOS representation produce a 320-dimensional condition $\mathbf{c}$, broadcast to all anchors.

The graph unites three directed neighborhoods: 32-nearest neighbors in the apo structure, 16-nearest neighbors in the flow state $\mathbf{X}_t$ at the start of the forward pass, and adjacent entries in the sparse anchor sequence whose original residue index gap is at most 16. Graph indices stay fixed during that pass. Layer k recomputes the invariant edge features, $\mathbf{G}_{ij}^{(k)}$,

$$\mathbf{G}_{ij}^{(k)} = \mathrm{RBF}(\|\mathbf{X}_j^{(k)} - \mathbf{X}_i^{(k)}\|) \,\|\, \mathrm{RBF}(\|\mathbf{X}_{\mathrm{apo},j} - \mathbf{X}_{\mathrm{apo},i}\|) \,\|\, \mathrm{Fourier}(p_j - p_i), \tag{13}$$

with 32, 32, and 17 channels, respectively. Here the original residue index of anchor i is p_i, and the concatenation operator is $\|$.

Each of eight EGNN layers computes the conditioned node state, $\widetilde{\mathbf{H}}_i^{(k)}$; edge message, $\mathbf{m}_{ij}^{(k)}$; and message gate, $q_{ij}^{(k)}$:

$$\widetilde{\mathbf{H}}_i^{(k)} = \mathrm{AdaLN}(\mathbf{H}_i^{(k)}, \mathbf{c}), \tag{14}$$

$$\mathbf{m}_{ij}^{(k)} = \phi_e\Big([\widetilde{\mathbf{H}}_i^{(k)} \| \widetilde{\mathbf{H}}_j^{(k)} \| \mathbf{G}_{ij}^{(k)}]\Big), \tag{15}$$

$$q_{ij}^{(k)} = \sigma\Big(\phi_q(\mathbf{m}_{ij}^{(k)})\Big). \tag{16}$$

The 320-dimensional edge message is multiplied by $q_{ij}^{(k)}$ and averaged over valid neighbors. Node features receive a local message update, a masked-global-mean update, and a SwiGLU transition. Each residual is modulated channel-wise by the condition:

$$\mathbf{H} \leftarrow \mathbf{H} + g_s(\mathbf{c}) \odot \Delta\mathbf{H}_s, \qquad s \in \{\mathrm{local}, \mathrm{global}, \mathrm{tr}\}. \tag{17}$$

Here $\Delta\mathbf{H}_s$ is the candidate update from branch s, and $g_s(\mathbf{c})$ is its condition-dependent channel gate. The edge messages also produce the coordinate weights in Equation 6. Figure 7 shows the complete layer-level data flow.

Reading the figure from left to right, D3PM and ESM-3 residue features first initialize $\mathbf{H}^{(0)}$, while flow time and the ESM-3 BOS embedding produce the condition $\mathbf{c}$. Current and apo coordinates define the neighbor union and edge geometry. In every EGNN layer, AdaLN conditions the node features; the center feature, neighbor feature, and $\mathbf{G}_{ij}^{(k)}$ enter the edge MLP to form $\mathbf{m}_{ij}^{(k)}$ and the gate $q_{ij}^{(k)}$. The gated messages update node features through successive local, global, and transition residuals. In parallel, the same message predicts signed weights for the soft-normalized current and apo relative vectors, and the gated mixture updates the coordinates. The updated node states and coordinates enter the next layer. Neighbor indices stay fixed, whereas current-state geometry is recomputed at each layer. After eight layers, the final displacement is mean-centered to produce the clean holo-like anchor estimate.

A.2 Midpoint Network Details

For each active interval, the input token concatenates the noisy local state (4 channels), current parent-distance features (17), apo local state (4), and D3PM plus ESM-3 residue features (1,681). Parent-distance features comprise a normalized distance and 16 Gaussian RBF channels. Layer normalization and a linear map project the 1,706 input channels to 256, after which a sinusoidal residue-position embedding is added.

Two learned register tokens are prepended. A sinusoidal flow-time embedding and a learned recursion-level embedding form the 128-dimensional condition. Eight non-causal Transformer blocks apply AdaLN-conditioned eight-head self-attention and an AdaLN-conditioned SwiGLU transition, each with a residual connection. The registers are removed before a four-channel head predicts the clean local state. Figure 8 shows the tensor dimensions and decoding path.

B Local Internal Coordinates: A Step-by-Step Derivation

This section explains Equations 7–9 using right triangles, vector projections, and circle coordinates. The parents $\mathbf{A}, \mathbf{B}$ are already placed. The new residue $\mathbf{M}$ is a midpoint in *sequence index*, not necessarily the spatial midpoint $(\mathbf{A} + \mathbf{B})/2$. We assume distinct parents and nonnegative radii throughout the derivation.

B.1 From Two Distances to a Circle

Fixing $r_L = \|\mathbf{M} - \mathbf{A}\|$ restricts $\mathbf{M}$ to a sphere centered at $\mathbf{A}$. Fixing $r_R = \|\mathbf{M} - \mathbf{B}\|$ restricts it to a second sphere centered at $\mathbf{B}$. Their intersection is generally a circle: the distances determine its position and radius, but leave the position of $\mathbf{M}$ around that circle free.

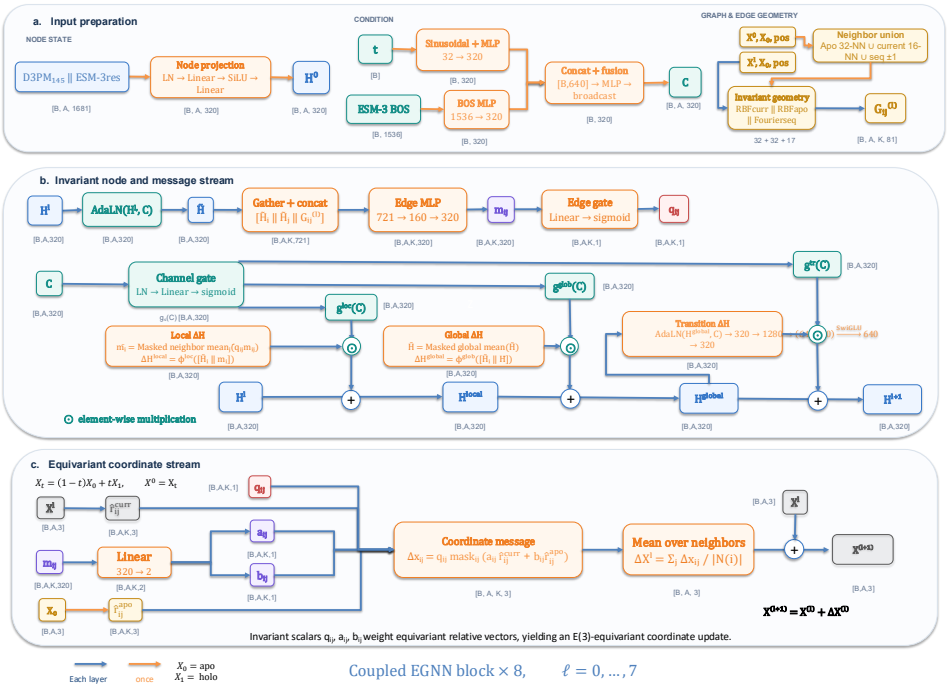


Figure 7: Detailed top-anchor Cartesian flow. Precomputed ESM-3 residue and BOS features condition an eight-layer EGNN. A fixed neighbor union supports invariant messages and equivariant coordinate updates. Orange paths are prepared once; blue paths are recomputed in each layer. Tensor dimensions are shown below the blocks.

Let the parent separation be $d = \|\mathbf{B} - \mathbf{A}\|$ and the unit vector from $\mathbf{A}$ to $\mathbf{B}$ be $\mathbf{e}_1 = (\mathbf{B} - \mathbf{A})/d$. For any candidate $\mathbf{M}$, write

$$\mathbf{M} = \mathbf{A} + \eta \mathbf{e}_1 + \mathbf{v}, \qquad \mathbf{v} \cdot \mathbf{e}_1 = 0.$$

Here the signed distance along the parent axis is η, and the remaining displacement perpendicular to it is $\mathbf{v}$. By the Pythagorean theorem,

$$r_L^2 = \eta^2 + \|\mathbf{v}\|^2, \qquad r_R^2 = (d - \eta)^2 + \|\mathbf{v}\|^2.$$

Subtracting the second equation from the first eliminates $\|\mathbf{v}\|^2$:

$$r_L^2 - r_R^2 = \eta^2 - (d - \eta)^2 = 2d\eta - d^2.$$

Thus every point satisfying both distances has the same axial coordinate,

$$\eta = \frac{d^2 + r_L^2 - r_R^2}{2d}, \qquad \mathbf{C} = \mathbf{A} + \eta \mathbf{e}_1.$$

Its perpendicular distance from $\mathbf{C}$ is also fixed:

$$\rho = \|\mathbf{v}\| = \sqrt{r_L^2 - \eta^2}.$$

All such points therefore lie on a circle of radius ρ, centered at $\mathbf{C}$ in a plane perpendicular to $\mathbf{e}_1$. If $r_L = r_R$, then $\eta = d/2$ and $\mathbf{C}$ is the spatial midpoint of the parents. For unequal radii, $\mathbf{C}$ shifts along the axis and can lie outside the segment AB.

The two spheres intersect only when $|r_L - r_R| \le d \le r_L + r_R$. At equality, $\rho = 0$ and the circle reduces to a point. The clamped square root used in the local-ICS geometry decoder prevents taking the square root of a negative value; it does not by itself make incompatible distance constraints feasible.

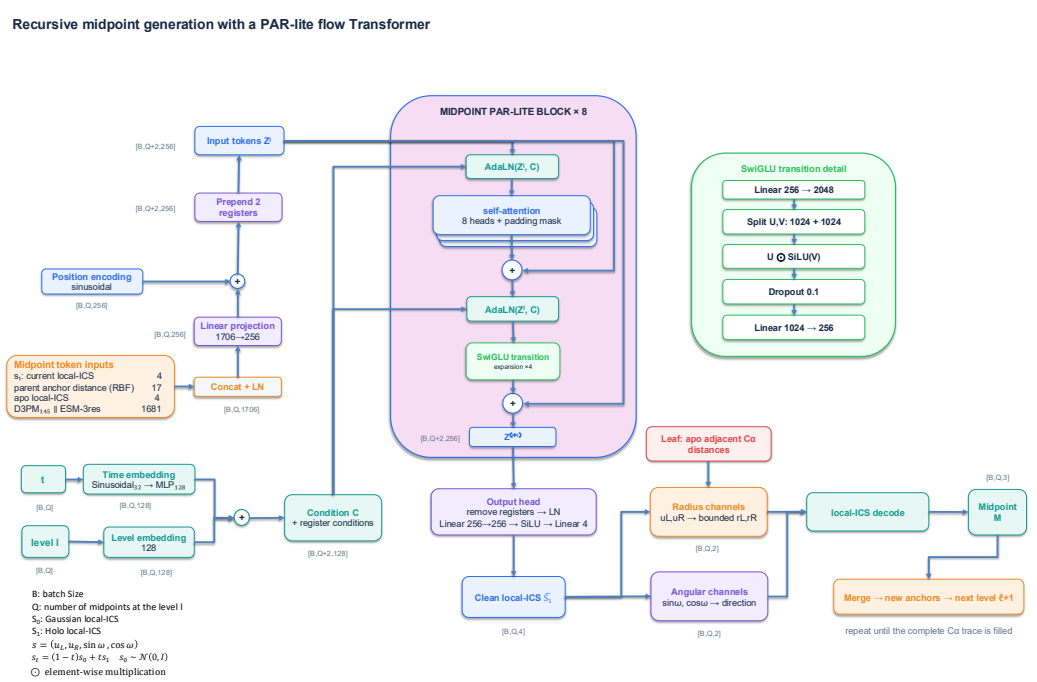


Figure 8: Detailed midpoint flow at recursion level ℓ. The Q active midpoint queries are processed in parallel with two learned registers. Self-attention is non-causal. After clean-state integration, bounded radii and an angular direction are decoded in local internal coordinates and merged into the trace.

B.2 Feasibility Repair for Non-intersecting Spheres

The main configuration enables geometric repair for non-final midpoint queries. Let the small interior margin be $\delta = 10^{-3}$. If the two spheres are externally separated, $d > r_L + r_R$, we set

$$\Delta_{\text{sep}} = \frac{d + \delta - r_L - r_R}{2}, \qquad (\widetilde{r}_L, \widetilde{r}_R) = (r_L + \Delta_{\text{sep}}, r_R + \Delta_{\text{sep}}).$$

Here the separation correction is Δ_{sep}, and the repaired radii are $\widetilde{r}_L, \widetilde{r}_R$. This equally shares the minimum Euclidean correction between the two radii and moves their sum just inside the feasible region. If the left sphere contains the right one, $r_L - r_R > d$, we instead use

$$\Delta_{\text{con}} = \frac{r_L - r_R - \max(d - \delta, 0)}{2}, \qquad (\widetilde{r}_L, \widetilde{r}_R) = (r_L - \Delta_{\text{con}}, r_R + \Delta_{\text{con}}),$$

where the containment correction is Δ_{con}. We apply the symmetric update when $r_R - r_L > d$. This transfers half of the excess from the larger radius to the smaller one until their difference lies just inside the containment boundary. Parent coordinates are never moved.

An adjacent-distance constraint takes priority over the symmetric repair. If, for example, r_L is fixed to its sanitized apo value and r_R is learned, the implementation leaves r_L unchanged and clamps only the free radius:

$$\widetilde{r}_L = r_L, \qquad \widetilde{r}_R = \text{clip}(r_R,\ |d - r_L| + \delta,\ d + r_L - \delta)\,,$$

with the roles reversed when only r_R is fixed. The repaired radii define η, $\mathbf{C}$, and ρ in Equation 7; the angular state is normalized before Equation 9 is applied. We retain the raw feasibility flag and record whether a projection was applied and its magnitude. The repair does not re-impose the learned contour-length caps, so a large correction remains visible as evidence of an upstream parent–radius mismatch.

Final leaves are handled separately. Both radii are apo-fixed and are never altered by the non-final repair. If predicted parents make these two constraints incompatible, the local-ICS geometry decoder

evaluates $\rho = \sqrt{\max(r_L^2 - \eta^2, 0)}$. Hence $\rho = 0$ and $\mathbf{M} = \mathbf{C}$ lies on the parent axis, while the raw query remains labeled infeasible. This is a finite numerical fallback, not a claim that both incompatible distances have been satisfied. During training, the raw violations are additionally penalized by

$$\mathcal{L}_{\text{feas}} = \mathbb{E}\big[\text{ReLU}(d - r_L - r_R)^2 + \text{ReLU}(|r_L - r_R| - d)^2\big] .$$

Training and inference use the same local-ICS geometry-decoder repair.

B.3 From an Apo Direction to a Local Frame

Choose a reference vector. The circle does not yet have a zero-angle direction. We obtain one from the apo trace by defining

$$\bar{\mathbf{C}}_{\text{apo}} = \frac{\mathbf{A}_{\text{apo}} + \mathbf{B}_{\text{apo}}}{2}, \qquad \mathbf{q}_{\text{apo}} = \mathbf{M}_{\text{apo}} - \bar{\mathbf{C}}_{\text{apo}}.$$

The vector from the apo parents' spatial midpoint toward the apo midpoint residue is $\mathbf{q}_{\text{apo}}$. The apo parents' spatial midpoint, $\bar{\mathbf{C}}_{\text{apo}}$, is not the circle center $\mathbf{C}$: the former uses apo coordinates, while the latter uses the current parents and the two prescribed radii.

Remove the component along the parent axis. Since $\mathbf{e}_1$ has unit length, the scalar $\mathbf{q}_{\text{apo}} \cdot \mathbf{e}_1$ is the signed length of the component along the axis. Multiplying this scalar by $\mathbf{e}_1$ gives the axial vector itself. Subtracting it leaves only the component parallel to the circle plane:

$$\mathbf{q}_{\parallel} = (\mathbf{q}_{\text{apo}} \cdot \mathbf{e}_1)\mathbf{e}_1, \qquad \mathbf{q}_{\perp} = \mathbf{q}_{\text{apo}} - \mathbf{q}_{\parallel}.$$

Indeed, $\mathbf{e}_1 \cdot \mathbf{q}_{\perp} = \mathbf{e}_1 \cdot \mathbf{q}_{\text{apo}} - (\mathbf{q}_{\text{apo}} \cdot \mathbf{e}_1)(\mathbf{e}_1 \cdot \mathbf{e}_1) = 0$. In matrix notation, the same operation is

$$\mathbf{q}_{\perp} = (\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^{\top})\mathbf{q}_{\text{apo}},$$

where $\mathbf{I}$ is the 3×3 identity matrix. The outer product $\mathbf{e}_1\mathbf{e}_1^{\top}$ extracts the axial component; subtracting it from $\mathbf{I}$ keeps the perpendicular component.

Why translation followed by projection is the same construction. The target plane passes through $\mathbf{C}$ and is *perpendicular* to $\mathbf{e}_1$; it is not a plane containing the parent axis. Denote this circle plane by $\Pi_{\mathbf{C}}$:

$$\Pi_{\mathbf{C}} = \{\mathbf{P} : (\mathbf{P} - \mathbf{C}) \cdot \mathbf{e}_1 = 0\}.$$

Translate the arrow representing $\mathbf{q}_{\text{apo}}$ so that its tail moves from $\bar{\mathbf{C}}_{\text{apo}}$ to $\mathbf{C}$. Its length and direction do not change. Its translated head is $\mathbf{T} = \mathbf{C} + \mathbf{q}_{\text{apo}}$. The tail already lies in $\Pi_{\mathbf{C}}$. To project the head onto that plane, subtract its signed normal displacement:

$$\begin{aligned}\mathbf{T}_{\text{proj}} &= \mathbf{T} - \big[(\mathbf{T} - \mathbf{C}) \cdot \mathbf{e}_1\big]\mathbf{e}_1 \\ &= \mathbf{C} + \mathbf{q}_{\text{apo}} - (\mathbf{q}_{\text{apo}} \cdot \mathbf{e}_1)\mathbf{e}_1 \\ &= \mathbf{C} + \mathbf{q}_{\perp}.\end{aligned}$$

The projected arrow from $\mathbf{C}$ to $\mathbf{T}_{\text{proj}}$ is therefore exactly $\mathbf{q}_{\perp}$. Translating the arrow chooses its location; removing the axial component determines its projection. No rotation of the apo vector is involved. Both constructions use the axis of the *current* parents, not the apo parent axis.

Normalize the direction and complete the frame. The projected arrow need not reach the circle. Only its direction is needed:

$$\mathbf{e}_2 = \frac{\mathbf{q}_{\perp}}{\|\mathbf{q}_{\perp}\|}, \qquad \mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2.$$

The vectors $\mathbf{e}_2, \mathbf{e}_3$ are perpendicular unit vectors in the circle plane; the cross product fixes their orientation by the right-hand rule. The zero-angle point is $\mathbf{C} + \rho\mathbf{e}_2$, not necessarily $\mathbf{M}_{\text{apo}}$ or $\mathbf{T}_{\text{proj}}$. This normalization requires a nonzero projected vector; the degenerate case is discussed next.

B.4 Degenerate References and the Axis Fallback

Why a fallback is needed. If $\mathbf{q}_{\text{apo}}$ is parallel to $\mathbf{e}_1$, removing its axial component leaves $\mathbf{q}_\perp = 0$. The same happens if $\mathbf{q}_{\text{apo}}$ itself is zero. There is then no apo-defined direction in the circle plane, and dividing by $\|\mathbf{q}_\perp\|$ is undefined. A near-zero projection also makes the normalized direction sensitive to small numerical perturbations. The implementation activates the fallback when the projected norm is at most 10^{-6} during target encoding or at most 10^{-8} during Cartesian decoding. The fallback supplies a reference direction for these cases; it does not change the circle center or the prescribed radii.

Least-aligned coordinate-axis construction. For a fallback based on the fixed Cartesian unit axes $\mathbf{a}_x = (1,0,0)$, $\mathbf{a}_y = (0,1,0)$, and $\mathbf{a}_z = (0,0,1)$, the selection rule is

$$\mathbf{a}_* \in \operatorname*{arg\,min}_{\mathbf{a}\in\{\mathbf{a}_x,\mathbf{a}_y,\mathbf{a}_z\}} |\mathbf{a}\cdot\mathbf{e}_1|.$$

The absolute dot product measures how closely an axis aligns with $\mathbf{e}_1$, regardless of sign. Choosing the smallest value selects the axis closest to being perpendicular to $\mathbf{e}_1$. The implementation tests the axes in x, y, z order, and the first minimum is selected in a tie. The replacement frame is then

$$\mathbf{v}_{\text{fallback}} = \mathbf{a}_* - (\mathbf{a}_*\cdot\mathbf{e}_1)\mathbf{e}_1, \qquad \mathbf{e}_2 = \frac{\mathbf{v}_{\text{fallback}}}{\|\mathbf{v}_{\text{fallback}}\|}, \qquad \mathbf{e}_3 = \mathbf{e}_1\times\mathbf{e}_2.$$

This projection cannot vanish for a unit $\mathbf{e}_1$. Its three squared Cartesian components sum to one, so the smallest is at most $1/3$. Consequently,

$$\|\mathbf{v}_{\text{fallback}}\|^2 = 1 - (\mathbf{a}_*\cdot\mathbf{e}_1)^2 \geq \frac{2}{3}.$$

For example, when $\mathbf{e}_1$ points along the x-axis, either the y- or z-axis provides a perpendicular reference direction.

Determinism is not rotation equivariance. For the nondegenerate apo-based frame, rotating both structures rotates $\mathbf{e}_1$ and $\mathbf{q}_{\text{apo}}$ together, so their projection rotates with them. Fixed Cartesian axes do not rotate with the protein. A fallback chosen from those axes is reproducible in a given coordinate frame, but need not produce the rotated version of the original $\mathbf{e}_2$ after the protein is rotated. It therefore does not, in general, preserve rotation equivariance of the local decoder when activated. This limitation is distinct from the equivariance of the Cartesian anchor network in Section 3.3.

Coincident parents ($d = 0$) are a separate degeneracy: the parent axis itself is undefined, so replacing the apo reference vector alone does not resolve that case. At sphere tangency ($\rho = 0$), the decoded point is simply $\mathbf{C}$, independent of the angular reference.

B.5 Placing the Midpoint: A Worked Example

In a two-dimensional plane, a circle of radius ρ has coordinates $(\rho\cos\omega, \rho\sin\omega)$. Replacing the two coordinate axes by $\mathbf{e}_2, \mathbf{e}_3$ and placing the origin at $\mathbf{C}$ gives

$$\mathbf{M} = \mathbf{C} + \rho(\cos\omega\,\mathbf{e}_2 + \sin\omega\,\mathbf{e}_3).$$

At $\omega = 0$, the displacement points along $+\mathbf{e}_2$; at $\omega = \pi/2$, it points along $+\mathbf{e}_3$. Positive angles thus run from $+\mathbf{e}_2$ toward $+\mathbf{e}_3$. Changing ω leaves both parent distances unchanged because $\cos^2\omega + \sin^2\omega = 1$.

For a simple numerical example, let $\mathbf{A} = (0,0,0)$, $\mathbf{B} = (6,0,0)$, and $r_L = r_R = 5$, in arbitrary length units. Then

$$d = 6, \quad \mathbf{e}_1 = (1,0,0), \quad \eta = 3, \quad \mathbf{C} = (3,0,0), \quad \rho = \sqrt{25-9} = 4.$$

The circle lies in the plane $x = 3$. Suppose the apo reference vector is $\mathbf{q}_{\text{apo}} = (2,3,4)$. Moving its tail to $\mathbf{C}$ places its head at $\mathbf{T} = (5,3,4)$. Orthogonal projection onto $x = 3$ gives $\mathbf{T}_{\text{proj}} = (3,3,4)$, so the projected arrow is $(0,3,4)$. Subtracting the axial component gives the same result directly:

$$\mathbf{q}_\perp = (2,3,4) - 2(1,0,0) = (0,3,4).$$

Normalizing and taking the cross product yields $\mathbf{e}_2 = (0, 3/5, 4/5)$ and $\mathbf{e}_3 = (0, -4/5, 3/5)$. At $\omega = 0$,

$$\mathbf{M} = \mathbf{C} + 4\mathbf{e}_2 = (3, 12/5, 16/5).$$

Both squared parent distances are $3^2 + (12/5)^2 + (16/5)^2 = 25$, as required. Notice that the projected apo arrow has length 5, while the circle radius is 4: projection determines the reference direction, and the two parent distances determine the radius.

B.6 Empirical Support for the Adjacent-Distance Prior

We characterize the apo-distance assumption over all 1,591 D3PM pairs as a post-hoc dataset audit; these statistics were not used to select a model or a hyperparameter. Segment boundaries are essential: treating consecutive rows of a compressed tensor as bonded residues would introduce false adjacencies across gaps. We restore the curated residue masks for 1,587 pairs and recover the four remaining 6XR4 pairs from their source PDB files. This yields 577,291 within-segment consecutive Cα pairs and excludes 3,960 transitions across mask-defined gaps.

Among these pairs, 576,253/577,291 (99.8202%) satisfy $|d_{\text{holo}} - d_{\text{apo}}| < 0.2$ Å. The mean and median absolute differences are 0.0226 and 0.0147 Å, respectively. Figure 9 shows the full adjacent-distance distributions and their signed differences.

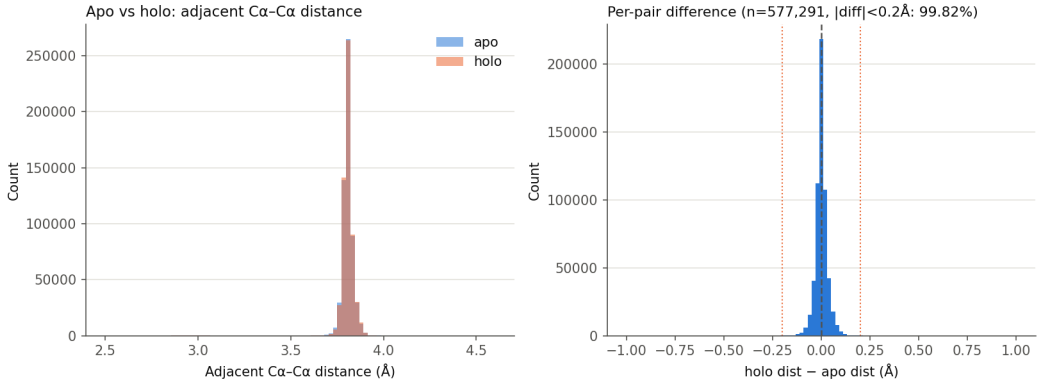


Figure 9: Adjacent Cα–Cα distances in matched apo and holo structures. The left panel overlays the two distance distributions. The right panel shows $d_{\text{holo}} - d_{\text{apo}}$; orange dotted lines mark ± 0.2 Å. The analysis contains 577,291 within-segment consecutive Cα pairs from all 1,591 D3PM pairs, of which 99.8202% lie within the marked interval.

B.7 Test-Time Sphere-Feasibility Diagnostics

Table 3 reports raw sphere-intersection violations during free recursive rollout on the fixed 150-pair test split. The `feasible` flag is evaluated from the predicted radii before geometric repair. A non-final query can therefore be counted here even though its radii are subsequently projected before the midpoint is placed.

Of the 3,833 raw violations, 3,696 (96.43%) are non-final queries and 137 (3.57%) are final leaves. Final-leaf violations constitute 0.0673% of all generated midpoint queries. The saved evaluation artifacts do not contain the total number of final-leaf opportunities, so we do not report a category-conditional final-leaf rate.

C Main Configuration

The top distance RBF centers span 0–6 normalized coordinate units; this is not a 6 Å graph cutoff. The midpoint parent-distance RBF centers span 0–64 Å. The top graph indices are fixed within a forward pass, but current edge distances are recomputed after each coordinate update. Self-conditioning, scheduled sampling, and a separate previous-level memory module are disabled in the main configuration.

Training seed	Midpoint queries	Raw infeasible	Rate (%)	Non-final / final-leaf events
1	40,724	564	1.3849	532 / 32
3	40,724	685	1.6821	646 / 39
42	40,724	996	2.4457	972 / 24
88	40,724	960	2.3573	940 / 20
99	40,724	628	1.5421	606 / 22
Pooled	203,620	3,833	1.8824	3,696 / 137

Table 3: Raw pre-repair sphere infeasibility for the five base models. The per-seed rate is $1.882 \pm 0.486\%$ (mean $\pm$ sample standard deviation). Non-final radii are eligible for the repair in Appendix B.2; final leaves retain both apo-fixed radii and use the finite axial fallback if needed.

Table 4: Configuration used for the five-seed main result.

Component	Setting
Data split	1,291 / 150 / 150 apo–holo pairs
Residue features	D3PM 145 + precomputed ESM-3 1,536
Top anchor stride	4
Top model	EGNN flow, 8 layers, node 320, edge bottleneck 160
Top graph	apo 32-NN $\cup$ initial-state 16-NN $\cup$ sequence edges
Top RBF / sequence features	32 current + 32 apo + 17 Fourier channels
Global node context	masked mean over valid anchors
Midpoint model	Transformer, 8 layers, 8 heads, token 256, condition 128
Midpoint registers	2
Parent feature	normalized distance + 16 Gaussian RBF channels
Midpoint geometry	local internal coordinates with 3/2/1-DOF policy
Flow parameterization	clean endpoint prediction
Time sampling	$t = 0$ / Euler grid / uniform = 1:1:1
Inference	5 Euler steps per active flow
Training	AdamW, batch 8, 1,000 epochs
Checkpoint selection	minimum validation mean Kabsch RMSD
Training seeds	1, 3, 42, 88, 99

Coordinate normalizers are estimated from training-set holo top anchors: $\widetilde{\mathbf{X}} = (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})/\boldsymbol{\sigma}_{\mathbf{X}}$, with per-channel population mean and standard deviation. Midpoint states use the analogous channel-wise statistics over all training targets, $\widetilde{\mathbf{S}} = (\mathbf{S} - \boldsymbol{\mu}_{\mathbf{S}})/\boldsymbol{\sigma}_{\mathbf{S}}$. Predictions are mapped back with the inverse transformations before geometry decoding and evaluation. The midpoint parent feature concatenates $d/64$ with 16 Gaussian RBF values whose centers are uniformly spaced from 0 to 64 Å.

D Training Objective Details

The main configuration predicts clean endpoints. Let the normalized holo anchor coordinates be $\widetilde{\mathbf{X}}_1$, and let the standardized four-channel local state for query q at midpoint level ℓ be $\widetilde{\mathbf{S}}^{(\ell)}_{1,q}$. The two primary losses are

$$\mathcal{L}_{\text{top}} = \frac{1}{3|\mathcal{A}|} \sum_{i \in \mathcal{A}} \left\| f^{\text{top}}_\theta(\mathbf{X}_t, t, \mathbf{F})_i - \widetilde{\mathbf{X}}_{1,i} \right\|_2^2, \tag{18}$$

$$\mathcal{L}_{\text{mid}} = \frac{1}{D} \sum_{\ell=1}^{D} \frac{1}{Q_\ell} \sum_{q=1}^{Q_\ell} \left[\frac{1}{2} \sum_{c=1}^{2} a_{q,c} e^2_{q,c} + \frac{1}{2} \sum_{c=3}^{4} e^2_{q,c} \right], \tag{19}$$

where the number of active midpoint queries at level ℓ is Q_ℓ, and the clean-state prediction error for query q and channel c is $e_{q,c}$. The first two midpoint channels encode radii and the last two encode $(\sin\omega, \cos\omega)$. The mask $a_{q,c}$ is zero when Equation 10 fixes that radius and one otherwise. Thus the model receives no state-loss gradient for a radius that the decoder replaces by an apo distance. The loss is averaged over all queries within each level and equally over levels; apo-fixed radius channels contribute zero but retain the two-channel normalization.

For the Cartesian auxiliary terms, define the Smooth-L1 penalty

$$h_\beta(e) = \begin{cases} e^2/(2\beta), & |e| < \beta, \\ |e| - \beta/2, & \text{otherwise.} \end{cases} \tag{20}$$

The geometric auxiliary losses below are evaluated on the teacher-forced partial reconstruction used during training. The decoded-coordinate loss $\mathcal{L}_{\text{xyz}}$ averages $h_{0.1}(\widehat{\mathbf{M}} - \mathbf{M}_{\text{holo}})$ over Cartesian components, queries, and midpoint levels. For a contiguous segment g and sequence offset Δ, let

$$\ell_{\Delta,g}(\widehat{\mathbf{X}}, \mathbf{X}_1) = \frac{1}{|\mathcal{P}_{\Delta,g}|} \sum_{i \in \mathcal{P}_{\Delta,g}} h_\beta\Big(\|\widehat{\mathbf{X}}_{i+\Delta} - \widehat{\mathbf{X}}_i\| - \|\mathbf{X}_{1,i+\Delta} - \mathbf{X}_{1,i}\| \Big), \tag{21}$$

where the set of valid index pairs separated by Δ within segment g is $\mathcal{P}_{\Delta,g}$. Then $\mathcal{L}_{\text{adj}}$ is the mean of $\ell_{1,g}$ over segments with $\beta = 0.05$. The sparse multiscale term is

$$\mathcal{L}_{\text{sparse}} = \frac{\sum_{\Delta \in \{1,2,4\}} \sum_g w_\Delta \ell_{\Delta,g}}{\sum_{\Delta \in \{1,2,4\}} \sum_g w_\Delta}, \qquad (w_1, w_2, w_4) = (1, 1, 0.5), \tag{22}$$

using $\beta = 0.1$. Offsets one, two, and four supervise connectivity, local bending, and slightly wider local shape without constructing all L^2 pairs.

Before any non-final repair, let the parent separation and two decoded radii for query q be d_q, $r_{L,q}$, and $r_{R,q}$, respectively. The feasibility penalty is averaged within each midpoint level and then equally across levels:

$$\mathcal{L}_{\text{feas}} = \frac{1}{D} \sum_{\ell=1}^{D} \frac{1}{Q_\ell} \sum_{q=1}^{Q_\ell} \left[\text{ReLU}(d_q - r_{L,q} - r_{R,q})^2 + \text{ReLU}(|r_{L,q} - r_{R,q}| - d_q)^2 \right]. \tag{23}$$

The full implemented objective is

$$\mathcal{L} = \mathcal{L}_{\text{top}} + \mathcal{L}_{\text{mid}} + \alpha_s \left(0.2\mathcal{L}_{\text{xyz}} + 0.1\mathcal{L}_{\text{adj}} + 0.1\mathcal{L}_{\text{sparse}} + 0.05\mathcal{L}_{\text{feas}} \right), \tag{24}$$

with a linear auxiliary warm-up $\alpha_s = \min(\max(s/500, 0), 1)$ at optimization step s. Optional top-coordinate, pairwise, endpoint, and RMSD auxiliaries are implemented but have zero weight in the reported configuration.

E Model Size and Inference Cost

Method	Parameters	Solver steps	Hardware	Test wall time	Time/target
CrypticFlow (ours)	34.080M	5 per active stage	RTX 2080 Ti 11GB	36.20 s	0.241 s
Direct-regression EGNN (local control)	0.165M	1	RTX 2080 Ti 11GB	2.30 s	0.015 s
SBAlign	0.545M	100	RTX 2080 Ti 11GB	245.86 s	1.639 s
ABM (FDBM implementation)	0.538M	100	RTX 2080 Ti 11GB	200.86 s	1.339 s
FDBM	0.538M	100	RTX 2080 Ti 11GB	208.24 s	1.388 s

Table 5: Hardware-controlled inference cost on the same 150-pair D3PM test split, with one generated structure per target. Every method was run sequentially on one NVIDIA GeForce RTX 2080 Ti. Wall time is measured around the CUDA-synchronized inference loop and includes per-target data transfer, sampling, and Cα metric computation; it excludes Python startup, checkpoint and dataset initialization, prediction serialization, ESM-3 preprocessing, cg2all reconstruction, and MolProbity.

CrypticFlow has 34,080,066 registered trainable parameters: 23,562,666 in the top-anchor EGNN and 10,517,400 in the midpoint Transformer. The count excludes ESM-3 because embeddings are precomputed. In the hardware-controlled run, CrypticFlow, direct-regression EGNN, SBAlign, ABM, and FDBM take 36.20, 2.30, 245.86, 200.86, and 208.24 seconds, respectively, for the complete 150-target test inference loop. These are single-checkpoint runtime measurements rather than seed-averaged accuracy estimates. CrypticFlow performs five Euler updates for the top-anchor stage and five updates at each active recursive midpoint level, whereas the bridge baselines use 100 integration steps and the direct EGNN uses one forward pass. The local direct-EGNN control is distinct from the legacy `EGNN (reported)` row in Table 1.

F Full-Sequence ESM-3 Mapping

We map each canonical D3PM row first to an ordered apo ATOM residue and then to the deposited polymer sequence. Author residue numbers and insertion codes are used when available. If insertion codes were lost during dataset construction, positions with the same residue number are disambiguated by amino-acid identity. Four pairs require a global-alignment fallback. All 1,591 pairs map without changing a deposited amino acid. Eight pairs contain a non-monotone gather order because of insertion-code ordering or circular permutation.

We first run ESM-3 on the entire deposited apo polymer sequence (SEQRES), including residues that are unresolved in the apo coordinates. For sequences of at most 600 residues, ESM-3 is applied once. Longer sequences use length-600 windows with stride 300. If a residue occurs in several windows, we retain the representation from the window in which it is farthest from a boundary. BOS and EOS representations are averaged across windows. We then gather only the embeddings corresponding to observed D3PM residues, in D3PM graph order.

G Rotation, Translation, and Reflection Consistency

An equivariant model should give the same structure when the input protein is shown in a different position or orientation. The output should only move with the input. We test this property on all 150 test pairs using each of the five trained models. First, we generate a prediction from the original input. We then transform every Cartesian input and run the model again. We use the same Gaussian midpoint-noise seed in both runs, pairing the stochastic draws channel-wise. Finally, we apply the inverse transform to the second prediction. If the model is equivariant to that transform, the two predictions should coincide without structural alignment.

We test rotation, translation, rotation plus translation, and reflection. A proper rotation is obtained by QR-orthogonalizing a 3×3 Gaussian matrix, correcting its column signs, and enforcing determinant $+1$. For reflection, we independently sample a proper rotation and flip its last column, producing a Haar-random improper orthogonal matrix $\mathbf{Q}$ with $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}$ and $\det \mathbf{Q} = -1$. We draw one transform per test pair with a fixed random seed. Each translation component is sampled from $[-25, 25]$ Å. All coordinate fields, including the apo reference vectors used by the local decoder, receive the same transform.

The two RMSD columns answer different questions. The *full-trace difference* directly compares the original prediction with the inverse-transformed new prediction, without Kabsch alignment. It is the strict equivariance check. The *change in holo RMSD* first computes the Kabsch-aligned RMSD of each prediction to its corresponding holo target and then reports the absolute difference between those two errors. It checks whether prediction accuracy is preserved, but does not by itself prove that the two generated structures are identical.

Transform	Top-anchor difference (Å)	Full-trace difference (Å)	Change in holo RMSD (Å)	Max. full-trace difference (Å)
Rotation	$3.245\times10^{-4}\pm1.9\times10^{-5}$	$4.998\times10^{-4}\pm2.5\times10^{-5}$	$3.708\times10^{-5}\pm3.7\times10^{-6}$	5.375×10^{-3}
Translation	$3.231\times10^{-4}\pm3.2\times10^{-5}$	$4.872\times10^{-4}\pm3.6\times10^{-5}$	$3.811\times10^{-5}\pm4.2\times10^{-6}$	5.788×10^{-3}
Rotation + translation	$3.252\times10^{-4}\pm2.1\times10^{-5}$	$4.904\times10^{-4}\pm3.7\times10^{-5}$	$3.625\times10^{-5}\pm5.5\times10^{-6}$	5.554×10^{-3}
Reflection	$2.576\times10^{-4}\pm1.4\times10^{-4}$	1.633 ± 0.027	0.725 ± 0.014	4.442

Table 6: Consistency under rigid transforms and reflection on the fixed 150-pair test split. We invert each input transform before comparing the new prediction with the original prediction; no Kabsch alignment is used. Values are means ± sample standard deviations across five trained models. The last column is the largest difference among all 5×150 predictions.

For proper rotations and translations, the inverse-transformed full traces differ by about 0.0005 Å on average, and the largest difference is below 0.0058 Å. Holo RMSD changes by less than 0.000039 Å on average. Thus, the composed generator is numerically consistent under the SE(3) transforms that preserve molecular handedness.

Reflection gives a different result. The top-anchor difference remains only $(2.576 \pm 1.4) \times 10^{-4}$ Å, confirming numerical reflection equivariance of the Cartesian EGNN. However, the complete trace

differs by 1.633 ± 0.027 Å, and holo RMSD changes by 0.725 ± 0.014 Å. The adjacent- distance MAE changes by only $(6.49 \pm 3.37) \times 10^{-7}$ Å because the radial constraints remain distance based, whereas pseudo-angle MAE changes by 3.915 ± 0.024 degrees. The composed model is therefore not reflection- equivariant. This behavior arises after the anchor stage: reflection reverses handedness, changes the parity of the cross-product axis $\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2$, and is not an enforced symmetry of the midpoint Transformer.

At the Cα-trace level, reflection produces the mirror geometry associated with the opposite handedness rather than another pose of the same L-protein. L- and D-proteins are biologically distinct because chirality changes their stereochemistry and molecular interactions. Reflection therefore lies outside the intended SE(3) symmetry group for L-protein conformations. We require consistency under proper rotations and translations, which preserve handedness. This fixed-seed experiment is a samplewise check, not a test of distribution-level symmetry under alternative latent-noise couplings. It is also empirical rather than a proof for every input. In addition, the degenerate-frame fallback in Appendix B.4 uses a fixed global axis and remains an exception to exact rotation equivariance.

H Detailed Structural Validity

Main Table 2 retains the three summary measures used in Figure 6. Table 7 reports the remaining Cα and all-atom diagnostics. Our values are mean $\pm$ sample standard deviation over five independently trained models.

Table 7: Detailed Cα and relaxation-free cg2all–MolProbity metrics. Cα clashes count non-bonded pairs; only direct sequence neighbours within the same restored segment are excluded.

Method	Cα clash/10^6 ↓	Clashscore ↓	Rama fav. (%) ↑	Rama out. (%) ↓	Bond RMSD (Å) ↓	Angle RMSD (deg) ↓
GT-holo Cα control	–	38.3	96.83	0.35	0.0205	2.00
Direct-regression EGNN	2794.18	507.7	26.65	51.05	0.1773	14.48
SBAlign	270.95	415.8	35.78	39.98	0.1680	11.69
FDBM	113.09	261.7	63.77	18.72	0.1014	9.15
ABM	121.43	264.6	63.15	18.97	0.1019	9.30
CrypticFlow (ours)	$\mathbf{20.01 \pm 3.27}$	$\mathbf{95.35 \pm 6.64}$	$\mathbf{90.14 \pm 0.45}$	$\mathbf{3.12 \pm 0.26}$	$\mathbf{0.0242 \pm 0.0006}$	$\mathbf{2.817 \pm 0.071}$

I Validation-Based Hyperparameter Selection

We selected time sampling, training-grid resolution, ESM-3 conditioning, and anchor stride on the validation split. The original selection matrix contains 19 conditions and five training seeds per condition, for 95 trained models in total. Additional post-selection mechanism controls are reported separately in Appendix J and were not used to revise the test-set results. Every model is evaluated on the same 150 validation pairs. Fixed-Cα cg2all reconstruction and MolProbity use the same 149 eligible pairs per run; the excluded pair, `4P7G_C_30E4_A`, consists of four disconnected singleton segments and does not define a local frame for backmapping. The RMSD columns use the public-protocol direct metric in the pre-aligned D3PM frame. Checkpoints were nevertheless selected by the prespecified validation Kabsch RMSD. Local metrics pool the same 55,064 within-segment adjacent pairs and 54,571 contiguous triplets for every run. Table 8 reports representative global, local, and all-atom metrics. Figure 10 visualizes the factor-wise comparisons and the accuracy–validity trade-off. The held-out test results are not used in this analysis.

Time sampling. The selected sampler draws equal proportions from the exact source endpoint $t = 0$, the discrete Euler grid, and the continuous uniform distribution. It has the lowest validation direct mean and median RMSD among the sampled time distributions. Grid-only reaches a direct mean/median RMSD of 3.165/1.644 Å, versus 3.083/1.595 Å for the selected sampler, and also has lower pseudo-angle validity and a higher MolProbity score. Uniform-only and SimpleFold-style sampling increase direct mean RMSD by 1.593 and 2.308 Å, respectively. This result suggests that explicit exposure to the source endpoint and inference-grid times is useful for this conditional transport problem.

Training time-grid resolution. Here the training-grid resolution, K_{train}, controls the support $\{0, 1/K_{\text{train}}, \ldots, (K_{\text{train}} - 1)/K_{\text{train}}\}$ of the discrete component of the training-time sampler. It is

Table 8: Validation ablations used for model selection. Values are mean ± sample SD over five training seeds. The selected base (marked †) uses equal-probability $t = 0$/Euler-grid/uniform sampling, $K_{\text{train}} = 5$, stride 4, and ESM-3. Distance and angle validity denote errors within 0.2 Å and 20°, respectively. Direct RMSD is measured in the shared pre-aligned D3PM frame without a post-hoc prediction-to-target Kabsch fit. All conditions use five Euler integration steps at evaluation.

Setting	Direct mean RMSD (Å) ↓	Distance valid (%) ↑	Angle valid (%) ↑	Clash/10^6 ↓	MP score ↓
Selected base†	3.083 ± 0.011	99.632 ± 0.021	95.261 ± 0.225	25.18 ± 3.73	3.044 ± 0.064
Time sampling					
t_0/grid/uniform = 10/40/50	3.148 ± 0.022	99.618 ± 0.033	94.745 ± 0.200	27.85 ± 3.92	3.133 ± 0.020
Euler grid only	3.165 ± 0.042	99.617 ± 0.013	94.974 ± 0.153	25.37 ± 4.78	3.119 ± 0.065
Grid/uniform = 50/50	3.232 ± 0.023	99.627 ± 0.033	94.440 ± 0.164	27.38 ± 1.69	3.202 ± 0.029
Curriculum: t_0 →grid→uniform	3.280 ± 0.051	99.630 ± 0.014	94.062 ± 0.262	30.14 ± 1.66	3.181 ± 0.030
Curriculum: uniform→grid	3.260 ± 0.060	99.593 ± 0.046	94.213 ± 0.269	32.96 ± 2.60	3.204 ± 0.045
Curriculum: grid→uniform	3.405 ± 0.033	99.562 ± 0.032	93.409 ± 0.314	40.27 ± 5.66	3.268 ± 0.056
Uniform only	4.677 ± 0.077	99.180 ± 0.148	89.986 ± 0.701	108.46 ± 17.00	3.637 ± 0.071
Curriculum: logit-normal→uniform	4.806 ± 0.071	99.006 ± 0.178	88.911 ± 0.513	133.00 ± 14.32	3.703 ± 0.040
SimpleFold-style	5.391 ± 0.025	99.171 ± 0.078	89.163 ± 0.945	141.95 ± 25.92	3.697 ± 0.104
Training time-grid resolution					
$K_{\text{train}} = 10$	3.097 ± 0.023	99.625 ± 0.026	95.195 ± 0.102	24.61 ± 2.18	3.030 ± 0.021
$K_{\text{train}} = 20$	3.092 ± 0.019	99.636 ± 0.011	95.278 ± 0.165	25.07 ± 2.71	3.066 ± 0.039
$K_{\text{train}} = 40$	3.089 ± 0.024	99.636 ± 0.019	95.243 ± 0.086	24.82 ± 2.15	3.091 ± 0.042
$K_{\text{train}} = 100$	3.102 ± 0.011	99.614 ± 0.018	95.192 ± 0.210	23.71 ± 3.14	3.049 ± 0.020
ESM-3 conditioning					
Without ESM-3 residue and BOS features	3.770 ± 0.052	99.564 ± 0.012	90.554 ± 0.313	70.32 ± 4.27	3.399 ± 0.023
Anchor stride					
Stride 1	2.888 ± 0.042	64.296 ± 2.231	94.556 ± 0.236	8.45 ± 0.99	3.137 ± 0.075
Stride 2	3.161 ± 0.032	99.328 ± 0.064	82.955 ± 0.974	42.48 ± 7.42	3.924 ± 0.042
Stride 8	3.254 ± 0.020	99.728 ± 0.010	94.133 ± 0.264	59.50 ± 7.90	3.233 ± 0.034
Stride 16	3.906 ± 0.008	99.734 ± 0.016	92.621 ± 0.097	120.76 ± 5.63	3.370 ± 0.023

not the number of integration steps used at evaluation, which remains five for every row. The selected $K_{\text{train}} = 5$ has the lowest direct mean RMSD (3.083 Å); $K_{\text{train}} = 20$ and 40 give 3.092 and 3.089 Å, respectively, differences that are small relative to seed variation. A grid of 10 gives a slightly lower MolProbity score (3.030 versus 3.044). We therefore interpret this ablation as showing no consistent benefit from a finer training grid, rather than as evidence that every metric is uniquely optimized at five.

ESM-3 conditioning. Removing both residue-level and BOS ESM-3 features increases direct mean RMSD by 0.687 Å, decreases pseudo-angle validity by 4.71 percentage points, and increases the MolProbity score by 0.355. The Cα clash rate also rises from 25.18 to 70.32 per million eligible pairs. Thus sequence-level context supports both conformational prediction and structural fidelity. This comparison ablates the two ESM-3 inputs together and does not isolate their individual contributions.

Anchor stride. Stride 1 predicts every residue in the Cartesian top flow and obtains the lowest direct mean RMSD, 2.888 Å. It removes the recursive local-coordinate stage, however, and its adjacent-distance validity falls from 99.63% to 64.30%. Stride 8 and stride 16 preserve adjacent distances but increase global error, clashes, and all-atom MolProbity score. Stride 2 gives lower pseudo-angle and all-atom validity than stride 4. We therefore select stride 4 as the best observed balance between global accuracy and local and all-atom validity, not as the optimum for RMSD alone. Stride 1 also uses the legacy execution path with state normalization disabled, so it is an architectural boundary case rather than a perfectly isolated stride-only perturbation.

Overall, the validation results support the equal three-way time mixture, ESM-3 conditioning, a five-point training grid, and stride 4 for the final model. The uncertainty values are seed-level sample standard deviations, not confidence intervals. We perform no significance test and avoid interpreting small differences as statistically resolved.

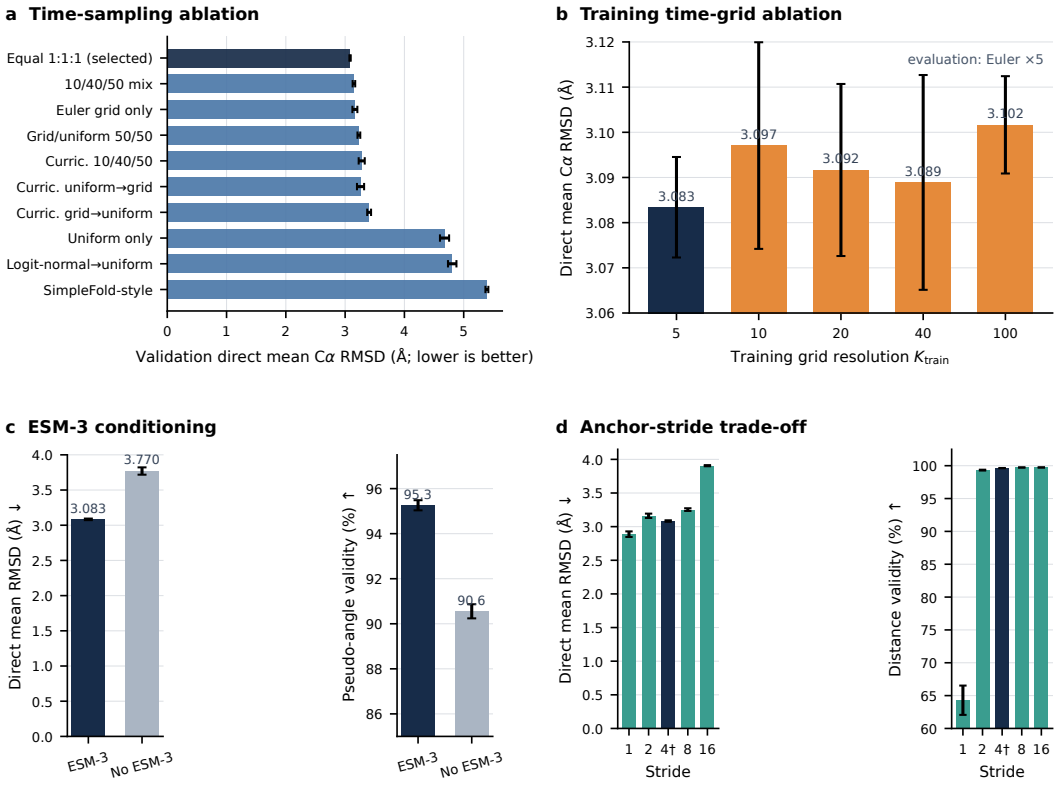


Figure 10: Validation ablations used for model selection. Bars and error bars show means and sample standard deviations across five independently trained models. (a) Absolute direct mean RMSD for every time sampler, including the selected equal 1:1:1 mixture of $t = 0$, Euler-grid, and uniform samples. (b) Training-grid resolution; evaluation uses five Euler steps in every condition. (c) ESM-3 conditioning improves direct RMSD and pseudo-angle fidelity. (d) Stride 4 balances direct RMSD and adjacent-distance fidelity. Dark bars and daggers denote the selected configuration.

J Local-Geometry Ablations

These controls were performed after the test-reported configuration had been fixed and use only the 150-pair validation split. They are explanatory rather than a second round of test-set model selection. Five-seed radius/loss controls use seeds $\{1, 3, 42, 88, 99\}$; the representation and auxiliary/projection controls use matched seeds $\{1, 3, 42\}$. Cα statistics come from the original validation predictions (150 pairs per seed) and use 55,064 within-segment adjacent pairs and 54,571 contiguous triplets. Fixed-Cα cg2all and MolProbity use the identical 149 successful pairs per seed; the excluded pair is `4P7G_C_30E4_A`. Values are means and seed-level sample standard deviations, not confidence intervals.

Table 9: Fixed-radius and distance-loss controls on validation. Values are mean $\pm$ sample SD over five seeds. Fixed denotes the reported apo-adjacent and final-leaf radius rules, not fixed radii everywhere. All rows retain coordinate and feasibility losses and non-final feasibility projection. Cα metrics use 150 pairs per seed and direct RMSD is measured in the shared pre-aligned D3PM frame without a post-hoc target fit; MolProbity (MP) uses the identical 149 successful pairs.

Setting	Direct mean RMSD (Å) ↓	Dist. valid (%) ↑	Angle valid (%) ↑	Cα clash/10^6 ↓	MP ↓
Fixed apo/final-leaf radii, full losses	3.083 ± 0.011	99.632 ± 0.021	95.261 ± 0.225	25.18 ± 3.73	3.044 ± 0.064
Learned all radii, full losses	3.072 ± 0.016	99.482 ± 0.019	95.319 ± 0.152	24.95 ± 1.77	3.009 ± 0.015
Fixed apo/final-leaf radii, no adjacent loss	3.074 ± 0.034	99.634 ± 0.026	95.364 ± 0.178	24.08 ± 4.09	3.064 ± 0.050
Learned all radii, no adjacent loss	3.075 ± 0.027	99.472 ± 0.019	95.396 ± 0.168	24.51 ± 2.56	3.032 ± 0.035
Fixed apo/final-leaf radii, no distance losses	3.075 ± 0.019	99.635 ± 0.016	95.311 ± 0.170	25.07 ± 1.97	3.076 ± 0.033
Learned all radii, no distance losses	3.071 ± 0.025	99.475 ± 0.018	95.435 ± 0.103	22.95 ± 2.78	2.990 ± 0.037

Table 10: Three-seed local-representation and auxiliary/projection controls on validation. RMSD is direct in the shared pre-aligned D3PM frame. Minimal Local ICS and Local Cartesian disable fixed radii, all geometric auxiliary losses, and feasibility projection. Local ICS + auxiliaries/projection retains coordinate and feasibility losses and the non-final projection. All rows learn every radius, disable both distance losses, and use the same flow-training schedule.

Setting	Median direct RMSD (Å) ↓	Mean direct RMSD (Å) ↓	Dist. valid (%) ↑	Angle valid (%) ↑	Cα clash/10^6 ↓	MP ↓
Minimal Local ICS	1.559 ± 0.010	3.064 ± 0.014	99.486 ± 0.018	95.207 ± 0.049	26.16 ± 2.28	3.025 ± 0.042
Local Cartesian	1.601 ± 0.091	3.037 ± 0.010	91.735 ± 0.459	96.364 ± 0.259	23.07 ± 5.69	2.931 ± 0.042
Local ICS + auxiliaries/projection	1.565 ± 0.028	3.060 ± 0.026	99.486 ± 0.010	95.449 ± 0.141	21.83 ± 1.84	2.980 ± 0.043

Table 11: All-atom details for the three-seed controls in Table 10. MolProbity evaluates the same 149 relaxation-free cg2all reconstructions per seed.

Setting	Clashscore ↓	Rama favored (%) ↑	Rama outlier (%) ↓	Bond RMSD (Å) ↓	Angle RMSD (deg) ↓
Minimal Local ICS	95.44 ± 5.73	90.174 ± 0.188	3.319 ± 0.020	0.02484 ± 0.00051	2.866 ± 0.031
Local Cartesian	94.51 ± 6.64	91.355 ± 0.185	3.086 ± 0.099	0.03188 ± 0.00061	3.327 ± 0.065
Local ICS + auxiliaries/projection	90.18 ± 4.45	90.404 ± 0.399	2.992 ± 0.162	0.02423 ± 0.00067	2.840 ± 0.082

Fixed radii and distance losses. Table 9 compares the reported fixed apo-adjacent/final-leaf radius rules with learned-all-radii variants, then removes the adjacent-distance loss or both adjacent and sparse-distance losses. Feasibility projection remains enabled in all six rows. Learning all radii under the full objective lowers distance fidelity by only 0.150 percentage points (99.632% to 99.482%), while direct mean RMSD and MolProbity slightly improve. Removing either set of distance losses changes the metrics only weakly and inconsistently relative to seed variation. The fixed-radius rule and distance auxiliaries therefore act as minor refinements rather than the primary source of local fidelity.

Local representation. The representation control retains the same stride-4 hierarchy, generated parent-anchor frame, apo reference, network capacity, conditioning, and training schedule. Both variants learn every radius and disable all geometric auxiliary losses and infeasibility projection. Local Cartesian predicts the three components of $\mathbf{M} - (\mathbf{A} + \mathbf{B})/2$ in the same $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ parent frame, padded by one zero channel to preserve the model interface. It has lower direct mean RMSD but higher direct median RMSD, higher pseudo-angle validity, and lower MolProbity. Local ICS, however, raises adjacent-distance fidelity by 7.751 percentage points and reduces reconstructed bond

and angle RMSDs. The shared frame defines each new midpoint relative to its generated parent anchors and the apo reference; the ICS distance parameterization adds a stronger bond-scale geometric bias. This supports a specific role in local geometry, not uniform superiority on every metric.

Other geometric auxiliaries and projection. Minimal Local ICS uses the same local-ICS architecture as the reported model, but learns every radius, removes all four geometric auxiliary losses ($\mathcal{L}_{\mathrm{xyz}}$, $\mathcal{L}_{\mathrm{adj}}$, $\mathcal{L}_{\mathrm{sparse}}$, and $\mathcal{L}_{\mathrm{feas}}$), and disables non-final sphere-feasibility projection. The *Local ICS + auxiliaries/projection* control also learns every radius and disables the two distance losses, but retains coordinate and feasibility losses and the non-final projection. Their direct mean RMSD and adjacent-distance fidelity are essentially unchanged. Removing these terms modestly worsens pseudo-angle validity, Cα clashes, and MolProbity, indicating that they are mild regularizers rather than the primary source of accuracy or local fidelity.

K Full-Chain Internal Coordinates and Lever-Arm Diagnostics

The shallow midpoint hierarchy does not make generated-parent errors vanish. Instead, each local prediction depends on one parent interval, whereas a sequential full-chain internal-coordinate rollout can transmit an early angular error through every downstream reconstruction. We examine this mechanism with two previously trained full-chain controls on the fixed 150-pair D3PM test split and with a controlled reconstruction diagnostic.

Model	Median direct RMSD (Å) ↓	Mean direct RMSD (Å) ↓	Dist. ≤ 0.2 Å (%) ↑	Angle $\leq 20°$ (%) ↑
Full-chain ICS Transformer	7.559	10.377	**99.699**	**98.833**
Full-chain ICS NeRF–EGNN	7.618	10.489	**99.699**	98.793
CrypticFlow (seed 42)	**1.338**	**2.780**	99.587	95.599

Table 12: Diagnostic comparison on the fixed 150-pair D3PM test split. The full-chain controls sequentially reconstruct the Cα trace from global internal coordinates. Because their NeRF root pose is undefined, we align each decoded trace to the pre-aligned apo input and then compute RMSD to holo without a prediction-to-holo fit. Local metrics pool 54,509 within-segment adjacent pairs and 54,082 contiguous triplets. All rows use seed 42, but the controls are not parameter-matched ablations; they provide supporting rather than causal evidence for the hierarchy.

This conclusion also replicates on the separate fixed 150-pair D3PM validation split using the same seed-42 checkpoints and metric protocols. The full-chain ICS Transformer, full-chain ICS NeRF–EGNN, and CrypticFlow obtain median/mean direct RMSDs of $8.900/11.677$, $8.300/11.396$, and $1.632/3.074$ Å, respectively. Their adjacent-distance/pseudo-angle fidelities are $99.809/98.824\%$, $99.809/98.668\%$, and $99.599/94.891\%$, respectively. Validation local metrics pool 55,064 within-segment adjacent pairs and 54,571 contiguous triplets.

The full-chain models match or exceed CrypticFlow on the local threshold metrics yet have substantially larger direct RMSD. This decoupling is consistent with the lever-arm effect: small local angular errors can accumulate into large downstream Cartesian displacements. To separate this effect from a decoder implementation failure, an oracle encode–decode round trip gives 1.39×10^{-4} Å RMSD. We then add zero-mean Gaussian noise to the angular channels of a 265-residue trace while keeping its distance channels fixed. Angular standard deviations of 0.001, 0.005, 0.010, and 0.020 radians ($0.057°$, $0.286°$, $0.573°$, and $1.146°$) produce Cartesian RMSDs of 0.242, 1.231, 2.284, and 4.946 Å, respectively (Figure 11). Upstream torsions have up to $328.8\times$ larger coordinate sensitivity than the least influential torsions. These diagnostics show chain-length amplification in sequential ICS; they do not claim that the hierarchical model eliminates parent-error propagation.

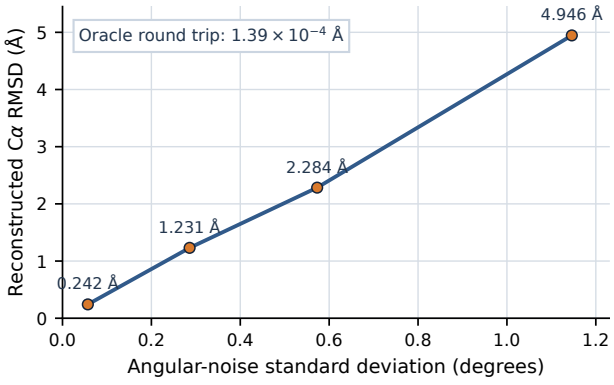


Figure 11: Controlled full-chain reconstruction diagnostic. Sub-degree angular noise produces increasing Cartesian displacement despite fixed distance channels and a numerically exact oracle round trip.

L Training and Inference Algorithm

Algorithm 1 makes explicit the different contexts used during training and generation. During training, the top-anchor module regresses the clean holo endpoint from interpolated apo–holo anchors. The midpoint module is trained level by level with ground-truth parent coordinates and clean local-coordinate targets. This teacher-forced construction prevents errors from earlier generated levels from changing the target distribution seen by the midpoint Transformer during optimization. The adjacent-distance rule is also active during training: radius channels next to a sequence-adjacent parent are fixed to the sanitized apo distance before decoding. Non-final decoded states use the same sphere-feasibility repair as inference, and their Cartesian coordinates contribute to geometric auxiliary losses, including an adjacent-distance term.

At inference, the two stages are composed autoregressively across hierarchy levels. The top-anchor flow is integrated first, and its generated anchors initialize a partial Cα trace. At each midpoint level, all active intervals are processed in parallel from fresh Gaussian local states. The integrated states are decoded with the adjacent-distance and sphere-feasibility rules, merged into the partial trace, and reused as parents for the next level. The procedure is therefore sequential across levels but parallel within each level. For stride four, two midpoint levels complete every full interval.

Algorithm 1 Training and inference with hierarchical flow matching.

Model components:

f_θ^{top}	ESM-3-conditioned E(3)-equivariant anchor EGNN;
f_θ^{mid}	conditional midpoint Transformer;
$\mathcal{D}_{\mathrm{ICS}}$	local-ICS geometry decoder and merge operator;
Euler_K	clean-endpoint flow integration with $K = 5$ steps.

Stage I: Data preparation

Split each paired trace into contiguous segments; select stride-4 anchors $\mathcal{A}$ and recursive midpoint levels $\{\mathcal{M}_\ell\}_{\ell=1}^{D}$.
Build residue features $\mathbf{F}$, apo-referenced local frames, and clean local targets $\mathbf{S}_1$.

Stage II: Training

Sample $(\mathbf{X}_{\mathrm{apo}}, \mathbf{X}_{\mathrm{holo}}, \mathbf{F})$ and flow time t.
Set $\mathbf{X}_t = (1-t)\mathbf{X}_{\mathrm{apo},\mathcal{A}} + t\mathbf{X}_{\mathrm{holo},\mathcal{A}}$ and regress $f_\theta^{\mathrm{top}}(\mathbf{X}_t, t, \mathbf{F})$ to $\mathbf{X}_{\mathrm{holo},\mathcal{A}}$.
for level $\ell = 1, \ldots, D$ with teacher-forced parent coordinates **do**
  Draw $\mathbf{S}_0 \sim \mathcal{N}(0, \mathbf{I})$ and set $\mathbf{S}_t = (1-t)\mathbf{S}_0 + t\mathbf{S}_1$.
  Regress $f_\theta^{\mathrm{mid}}(\mathbf{S}_t, t, \ell, \mathbf{F}, \mathrm{parents})$ to $\mathbf{S}_1$.
  Apply adjacent-distance and sphere-feasibility rules to the decoded training state; evaluate geometric auxiliary losses.
end for
Update θ with the joint clean-endpoint and geometric objective.

Stage III: Inference

Generate top anchors $\mathbf{X}_{\mathcal{A}} \leftarrow \mathrm{Euler}_K(f_\theta^{\mathrm{top}}, \mathbf{X}_{\mathrm{apo},\mathcal{A}}, \mathbf{F})$.
Initialize the partial trace $\mathcal{P} \leftarrow \mathbf{X}_{\mathcal{A}}$.
for level $\ell = 1, \ldots, D$ **do**
  Build unresolved parent intervals and draw fresh $\mathbf{S}_0 \sim \mathcal{N}(0, \mathbf{I})$.
  Set $\widehat{\mathbf{S}}_1 \leftarrow \mathrm{Euler}_K(f_\theta^{\mathrm{mid}}, \mathbf{S}_0, \ell, \mathbf{F}, \mathcal{P})$.
  Apply adjacent-distance and sphere-feasibility rules; update $\mathcal{P} \leftarrow \mathcal{P} \cup \mathcal{D}_{\mathrm{ICS}}(\widehat{\mathbf{S}}_1)$.
end for
return the completed holo-like Cα trace $\widehat{\mathbf{X}}_1 \leftarrow \mathcal{P}$.