COMPASS: Decoupled Latent Steering for Protein Conformational Transitions

Changyu Lee [1] **Sunghee Choi** [1] **Gyu Rie Lee** [2]

Abstract

Protein function depends on conformational transitions between distinct structural states, dynamics that static predictors such as AlphaFold 3 (AF3) do not capture. Unsteered ensemble models explore conformational space broadly but lack directional control, while direct 3D-coordinate guidance accelerates transitions at the cost of geometric violations and steric clashes. We propose COMPASS, which decouples target-directed optimization from structural inference: gradients are backpropagated only to the conditional prior embeddings at an intermediate anchor timestep via Tweedie's expectation, leaving the reverse diffusion process unmodified. On 98 Apo-Holo transition pairs, COMPASS improves target-reaching efficiency by up to $+42.9\%$ (HR$<$1.0 Å) over the unsteered baseline without degrading physical clash scores. We further find that Apo-targeted steering raises the Holo hit rate from 26.5% to **41.8%** at 1.0 Å, an effect we attribute to the base model's inherent Holo-sampling bias preserved under decoupled inference. This suggests that COMPASS activates the broader transition manifold rather than steering toward a single endpoint. Downstream generative docking on standard and ternary complex benchmarks shows complementary structural coverage relative to AF3 Holo predictions.

1. Introduction

Proteins execute their functions through conformational transitions between distinct structural states, for example, from a ligand-free Apo form to a ligand-bound Holo form. While deep learning predictors such as AlphaFold 3 (AF3) (Abramson et al., 2024) provide highly accurate static structure predictions, modeling the transition pathways connecting these states remains an open problem with direct implications for understanding induced-fit binding and structure-based drug discovery (Frauenfelder et al., 1991; Boehr et al., 2009).

Current generative approaches face a core tension between coverage and control. Unsteered ensemble models such as BioEmu (Lewis et al., 2025), which inherit a Holo-sampling bias from PDB training distributions (Berman et al., 2000; Saldaño et al., 2022), sample broadly but cannot direct transitions toward a specific target state. Direct 3D-coordinate guidance methods can accelerate transitions but frequently violate the SE(3) geometric constraints of protein manifolds, producing structures with elevated steric clashes (Lu et al., 2024; Jian et al., 2026).

We propose COMPASS (Conformational Mapping of Pathways Across Structural States), which resolves this tension by decoupling target-directed optimization from structural inference. COMPASS backpropagates gradients only to the conditional prior embeddings at a discrete anchor timestep via Tweedie's expectation (Efron, 2011), then executes a clean, unmodified reverse diffusion conditioned on the updated priors. This design preserves the physical manifold of the base model while enabling directed conformational exploration, and extends recent test-time latent optimization strategies (Chung et al., 2023; Didi et al., 2026) to the dynamic transition setting.

Our primary contributions are as follows:

- We formulate an anchor-based decoupled latent steering algorithm that circumvents the geometric degradation of 3D-coordinate guidance by isolating gradient updates from the structural inference process.
- We curate a benchmark of 98 high-quality Apo-Holo transition pairs and demonstrate that COMPASS improves target-reaching efficiency over unsteered baselines without increasing steric clashes.
- We observe a *cross-conformation exploration effect*: Apo-directed steering substantially improves Holo-state coverage, an effect we attribute to the interplay between decoupled inference and the base model's inherent Holo-sampling bias.

[1]School of Computing, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea [2]Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea. Correspondence to: Gyu Rie Lee <gyurielee@kaist.ac.kr>.

Accepted at the 2026 Workshop on Generative and Agentic AI for Biology (ICML 2026)

2. Related Work

2.1. Unsteered Conformational Ensemble Generation

Generative diffusion and flow-matching models such as EigenFold (Jing et al., 2023), ESMFlow/AlphaFlow (Jing et al., 2024), and more recently FoldFlow (Bose et al., 2024), ProTDyn (Liu et al., 2026), and TEMPO (Xu et al., 2025)[1] have advanced protein conformational ensemble generation. However, these models operate unconditionally with respect to specific target states; undirected exploration leaves them prone to remaining trapped in the initial energy basin without a mechanism to preferentially navigate toward a designated endpoint.

2.2. AlphaFold 2-Based Conformational Sampling

An orthogonal line of work extracts conformational diversity directly from AlphaFold 2 by manipulating its multiple-sequence-alignment (MSA) input. MSA subsampling and depth reduction (del Alamo et al., 2022), the SPEACH_AF protocol (Stein & Mchaourab, 2022), sequence-cluster dropout (Wayment-Steele et al., 2024), and high-throughput subsampled inference (Monteiro da Silva et al., 2024) have all demonstrated the ability to recover alternative conformational states from a single sequence. Although effective at broadening structural diversity, these approaches lack a mechanism to steer the ensemble toward a designated target state, which is the core capability provided by COMPASS through latent embedding optimization.

2.3. Directed Generation via 3D Coordinate Guidance

Methods such as Str2Str (Lu et al., 2024) accelerate structural transitions via explicit coordinate-space guidance. However, iteratively perturbing 3D coordinates x_t forces the denoiser outside its training distribution, frequently introducing geometric violations and elevated steric clashes (Lu et al., 2024; Jian et al., 2026).

2.4. Experiment-Guided Ensemble Generation

Recent work has integrated experimental observables, such as SAXS profiles or cryo-EM densities, directly into generative protein models. ExEnDiff (Liu et al., 2025) conditions a diffusion process on experimental data to bias ensemble sampling, while experiment-guided AlphaFold posterior inference (Maddipatla et al., 2025) optimizes predicted structures to match measured observations. COMPASS shares the philosophy of posterior conditioning but substitutes experimental measurements with a target 3D structure, enabling directed conformational steering even when experimental data are unavailable.

[1]Empirical comparison with TEMPO was omitted as the official code is currently unavailable.

2.5. Test-Time Latent Optimization and Decoupled Inference

Diffusion Posterior Sampling (DPS) (Chung et al., 2023) and more recent test-time compute scaling strategies such as Proteína-Complexa (Didi et al., 2026) optimize internal conditional representations to incorporate task-specific objectives without direct 3D manipulation. COMPASS extends this latent optimization paradigm to the dynamic conformational transition setting, operating via a discrete anchor timestep and a strictly decoupled inference phase so that gradient updates never interfere with structural generation.

3. Background and Preliminaries

We adopt the score-based SDE framework (Ho et al., 2020; Song et al., 2021), in which a forward VP-SDE progressively noises a protein structure $x_0 \in \mathbb{R}^{N\times 3}$ to a Gaussian prior, and generation is performed by running the reverse-time SDE with a score network $s_\theta(x_t, t, E_{1D}, E_{2D})$ conditioned on 1D residue embeddings E_{1D} and 2D pairwise embeddings E_{2D} (Jing et al., 2021). A key tool is Tweedie's formula (Efron, 2011), which provides a differentiable surrogate for the clean structure at any intermediate timestep t:

$$\hat{x}_{0|t} \approx \frac{1}{\alpha_t}\left(x_t + \sigma_t^2\, s_\theta(x_t, t, E_{1D}, E_{2D})\right) \tag{1}$$

where α_t and σ_t^2 are the signal and noise scales of the VP-SDE schedule. This projection onto the clean manifold is used in Phase 3 of COMPASS without executing the remaining denoising steps.

4. Method: The COMPASS Framework

COMPASS is implemented on top of BioEmu (Lewis et al., 2025), a GVP-Transformer-based (Jing et al., 2021) $\mathrm{SE}(3)$-equivariant score model. The framework operates in four phases: (1) partial denoising to an anchor state, (2) latent optimization via Tweedie's projection, (3) embedding update, and (4) a clean full-trajectory inference conditioned on the updated embeddings. Unlike direct 3D-coordinate guidance, gradient updates never touch x_t, so the base model's geometric priors remain intact.

4.1. Prior Anchor Generation (Phases 1 & 2)

To evaluate the directional progression toward the guidance structure without executing the computationally expensive full reverse diffusion, we establish an intermediate "anchor" state. We initialize the system with pure Gaussian noise at $t = 1.0$ and perform partial denoising down to $t = 0.5$ using a DPM-Solver (Lu et al., 2022) (25 neural function evaluations, NFE) under the model's Variance Preserving SDE (VPSDE) schedule.

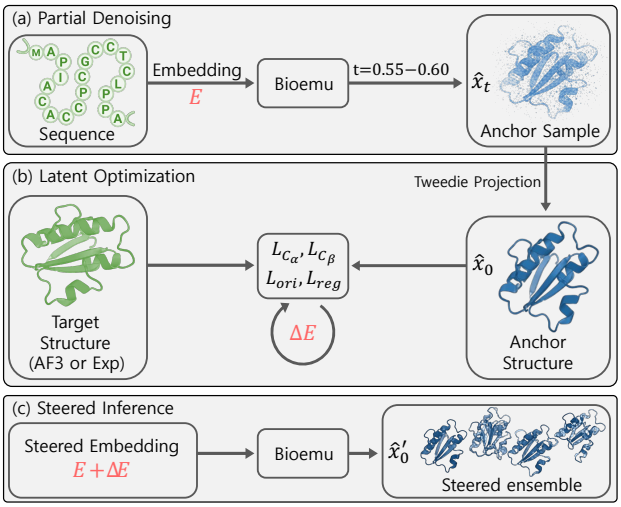

Figure 1. **Overview of the COMPASS Framework. (a)** The base BioEmu model performs partial denoising from Gaussian noise to an intermediate anchor state at $t \in [0.55, 0.60]$. **(b)** At the anchor, Tweedie's projection yields a differentiable clean-structure estimate $\hat{x}_0$; gradients are backpropagated to the conditional prior embeddings (E) via physics-aware latent optimization. **(c)** The optimized embeddings condition a fresh full-trajectory reverse diffusion, producing steered structural ensembles without any coordinate-level intervention.

The anchor is applied at $t = 0.5$ for a fixed partial denoising pass (Phase 2), while the optimal *steering* intervention timing t^* is determined separately per guidance type via 5-fold cross-validation. Our ablation sweep across $t \in \{0.35, 0.40, \ldots, 0.70\}$ (Section F) demonstrates that performance is largely robust across this range, as all configurations substantially exceed the unsteered baseline, confirming that the anchor evaluation mechanism does not require precise tuning.

4.2. Physics-Aware Latent Optimization (Phase 3)

At the anchor state ($t = 0.5$), we apply Tweedie's formula to project the noisy intermediate coordinates onto the clean data manifold, yielding an estimated structure $\hat{x}_0$. We then compute guidance-directed gradients by comparing $\hat{x}_0$ against the Holo guidance structure.

To preserve SE(3) invariance and geometric integrity, the objective function avoids absolute 3D coordinate penalties. Instead, we formulate the loss based on internal pairwise distances and local reference frames. Let $D_{C\alpha}(\cdot)$ and $D_{C\beta}(\cdot)$ denote the operations that extract the pairwise distance matrices for $C\alpha$ atoms and mathematically derived virtual $C\beta$ atoms, respectively. The geometric loss components are computed using Mean Squared Error (MSE) between the distance matrices of the projected structure $\hat{x}_0$ and the guidance structure x_{guide}. Additionally, we apply a backbone orientation loss $\mathcal{L}_{ori}$ that maximizes the trace of the relative rotation matrix $R_{pred}R_{holo}^T$, minimizing rotational discrepancies between local coordinate frames.

The optimization targets are the additive offsets to the initial conditional embeddings: Δ_{single} for the 1D embeddings and Δ_{pair} for the 2D representations. To prevent these offsets from deviating excessively from the pretrained latent prior, we impose an L_2 regularization term ($\mathcal{L}_{reg} = \|\Delta_{single}\|_2^2 + \|\Delta_{pair}\|_2^2$), which serves as a Maximum A Posteriori (MAP) estimation constraint. The total objective function is defined as:

$$\mathcal{L}_{total} = w_{CA}\mathcal{L}_{CA} + w_{CB}\mathcal{L}_{CB} + w_{ori}\mathcal{L}_{ori} + w_{reg}\mathcal{L}_{reg} \tag{2}$$

where hyperparameters (e.g., $w_{CA} = 1.0$, $w_{CB} = 2.0$, $w_{ori} = 1.0$, $w_{reg} = 1.0$, learning rate 0.015) are determined via an independent validation set (Section 5); optimization uses Adam (Kingma & Ba, 2015) for 100 iterations with gradient clipping (norm ≤ 1.0) and offset clamping $[-0.5, 0.5]$ (see Section A for full details).

The full pseudocode of COMPASS is provided in Section A.

4.3. Clean Inference for Manifold Preservation (Phase 4)

Following latent optimization, the anchor-state 3D coordinates are discarded; the refined embeddings $\tilde{E} = E + (\Delta^*_{single}, \Delta^*_{pair})$ condition a fresh full-trajectory reverse diffusion (DPM-Solver, 50 NFE, t=0.99→0.001). Because no coordinate-level intervention occurs during this phase, the base model operates within its training distribution, ensuring physically plausible trajectories without geometric artifacts. For the cross-conformation steering experiments reported in Table 1, we additionally activated the base model's internal dropout during inference, as detailed in Section 5.4.

We formally show that 3D coordinate interpolation with a guidance structure places intermediate states outside the marginal support of p_t (Theorem B.1), while latent embedding optimization preserves SE(3)-equivariance and manifold integrity (Theorem 4.1). The L_2 regularization on Δ is equivalent to MAP estimation, and the optimal anchor $t^* \approx 0.5$ follows from maximizing $\mathcal{J}(u) \propto -u \log u$ (Section B).

Theorem 4.1 (Manifold and SE(3) Preservation)**.** *Let $\mathcal{G} = \text{SE}(3)$ act on protein coordinates $x \in \mathbb{R}^{N\times 3}$. If the score network s_θ is $\mathcal{G}$-equivariant and the loss $\mathcal{L}$ is $\mathcal{G}$-invariant, then optimizing additive embedding offsets Δ preserves SE(3)-equivariance and manifold integrity of the generated structures.*

5. Experiments

5.1. Benchmark Setup and Data Preprocessing

We constructed two benchmark datasets. The 92-pair dataset is derived solely from the Protein Structural Change

Table 1. **Cross-conformation state hit rates under decoupled latent steering** ($N = 98$ pairs, PSCDB + CoDNaS). $\mathbf{HR}_{\text{Apo}}$: fraction of pairs where at least one generated frame falls within Cα-RMSD threshold of the Apo target structure. $\mathbf{HR}_{\text{Holo}}$: same metric evaluated against the Holo target structure (Cα-RMSD). $\mathbf{HR}_{\text{Either}}$: at least one frame hits *either* target structure. The bolded values highlight the largest improvements over the unsteered baseline, notably the cross-conformation exploration effect (e.g., Apo-guided steering substantially improving Holo HR). Full four-threshold breakdown is reported in Section G.

Method	Thresh. (Å)	HR_{Apo} (%) ↑	HR_{Holo} (%) ↑	$\text{HR}_{\text{Either}}$ (%) ↑
BioEmu (Unsteered)	1.0	30.6	26.5	38.8
	1.5	61.2	57.1	69.4
	2.0	76.5	80.6	86.7
COMPASS (Apo-guided)	1.0	70.4	**41.8**	**76.5**
	1.5	86.7	**74.5**	91.8
	2.0	93.9	**87.8**	94.9
COMPASS (Holo-guided)	1.0	**41.8**	69.4	72.4
	1.5	**77.6**	89.8	**92.9**
	2.0	**85.7**	93.9	**96.9**
COMPASS (AF3 Apo-guided)	1.0	**50.0**	**52.0**	66.3
	1.5	**80.6**	**79.6**	88.8
	2.0	**87.8**	**87.8**	93.9

Table 2. Performance stratified by displacement of structurally resolved Apo and Holo state pairs. Evaluated across the 92-pair benchmark (PSCDB only). **HR**: Hit Rate (<2.0 Å). COMPASS yields the most substantial margin (+25.3%) in the $5.0 \sim 10.0$ Å spatial rearrangement regime, corresponding to functional loop/domain hinge motions.

Max Cα Disp. (Å)	Pairs (n=92)	Unsteered HR	**COMPASS** HR	Δ
< 5.0	26	92.3%	**96.2%**	+3.9%
$5.0 \sim 10.0$	37	63.9%	**89.2%**	**+25.3%**
$10.0 \sim 15.0$	20	60.0%	**70.0%**	+10.0%
> 15.0	9	44.4%	**55.6%**	+11.2%

Database (PSCDB) (Amemiya et al., 2012). The expanded 98-pair dataset, used in Table 1, includes additional structures from the Conformational Diversity of Native State (CoDNaS) database. We applied the following filtering pipeline to an initial set of 4,800 entries.

Step 1: Protein Chain Homogeneity and Ligand Free-State Verification. To prevent steering perturbations caused by multi-chain interfaces, we selected structures consisting exactly of a single unique protein sequence (i.e., true monomers or pure homo-oligomers), with a sequence length between 50 and 700 residues. Transition pairs were restricted to those sharing the same UniProt ID with a sequence length ratio of 0.85 or higher.

Apo state purity was verified by manually auditing `HETATM` records; only crystallographic solvents and phasing agents (e.g., HOH, SO4, MSE) were permitted, and false-positive Apo states were discarded.

Step 2: High-Resolution Filtering. To ensure high structural reliability and prevent mapping errors during diffusion generation, we retained only high-quality crystal structures with an experimental resolution of 2.5 Å or better for both Apo and Holo states.

Benchmark Split and Representation. This pipeline yielded **98 Apo-Holo pairs** spanning Loop/Small Hinge, Domain Hinge, and Large-Scale Allosteric motions.

Data Leakage Considerations. Hyperparameters were selected via 5-fold cross-validation, with ablation confirming robustness across the explored range (Section F). COMPASS and the unsteered baseline share an identical frozen base model, ensuring symmetric access to training information. The COMPASS–AF3 comparison is restricted to structure pairs deposited after AF3's training cutoff, providing evaluation on genuinely unseen targets.

5.2. Quantitative Evaluation Results

Cross-Conformation Steering Effect. Table 1 presents the central empirical finding of this work. A hit is defined as at least one of 20 generated frames falling within a given Cα-RMSD threshold of the target structure. Evaluating *both* the Apo and Holo target structures regardless of steering direction reveals a **cross-conformation exploration effect**: Apo-guided COMPASS raises the Holo hit rate from 26.5% to 41.8% at the 1.0 Å threshold (+15.3 pp). Steering the latent prior toward Apo increases access to Holo-proximal conformations. We attribute this to the interplay between COMPASS's decoupled inference design and the inherent Holo-sampling bias present in modern structure modeling frameworks. Empirical analysis on the 51-pair common subset confirms this bias extends to predictive models: AF3 Apo predictions natively approximate the experimental Holo ground truth (mean RMSD 1.98 Å), whereas AF3 Holo predictions exhibit a larger divergence from the Apo ground truth (2.51 Å). Because COMPASS optimizes only the conditional prior at the anchor timestep and leaves the subsequent reverse diffusion unmodified, the generative base model's native tendency toward compact, Holo-proximal states, inherited from PDB training distributions (Saldaño et al., 2022), is preserved during inference. Apo-directed latent optimization shifts the conditional prior toward the Apo basin, partially counteracting this bias; the resulting ensemble is drawn into the intermediate region of the transition manifold where conformations proximal to both endpoints become accessible.

This effect is observed even when Apo and Holo states are structurally distinct. As shown in Table 2, the largest absolute improvement is observed in pairs with a maximum

Table 3. Transition efficiency, computational cost, and physical fidelity on the 92-pair benchmark. $\mathbf{HR}_{\mathrm{Either}}$: fraction of pairs hitting *either* the Apo or Holo endpoint (↑). **Avg. Best Cα-RMSD**: average minimum backbone Cα RMSD to target states (↓). **ETH**: Effective Time-to-Hit (↓). *ETH computed per protein pair on a single NVIDIA A5000 GPU ($L \approx 250$ residues).

Method	Guidance	$\mathrm{HR}_{\mathrm{Either}}$ (↑)			Avg. Best Cα-RMSD (↓)		NFE/sample	ETH* ↓
		<3.0 Å	<2.0 Å	<1.5 Å	To Apo (Å)	To Holo (Å)		
EigenFold	None	58.7%	51.1%	46.7%	6.55	5.41	21	∼1.82 min
ESMFlow	None	70.7%	54.3%	34.8%	5.43	4.57	10	∼0.34 min
ProTDyn	None	18.5%	12.0%	10.9%	13.62	13.12	50	∼0.28 min
BioEmu (Unsteered)	None	89.1%	83.7%	64.1%	2.96	2.70	50	∼0.23 min
COMPASS (Ours)	Apo-guided (t=0.55)	89.1%	85.9%	77.2%	**2.52**	3.23	55	∼0.21 min
COMPASS (Ours)	Holo-guided (t=0.60)	**90.2%**	**84.8%**	**79.3%**	3.37	**2.23**	55	**∼0.20 min**

local Cα displacement between the experimentally resolved Apo and Holo structures of 5.0–10.0 Å (+25.3 pp), where the two target structures are structurally distinct, rather than in the small-displacement regime (< 5.0 Å, +3.9 pp) where Apo and Holo conformations are already similar. The full four-threshold breakdown is reported in Section G.

Baseline Comparison and Computational Cost. Table 3 compares COMPASS against unsteered baselines using the $\mathrm{HR}_{\mathrm{Either}}$ metric for consistency. The unsteered BioEmu baseline achieves 83.7% $\mathrm{HR}_{\mathrm{Either}}$ at 2.0 Å, reflecting broad conformational diversity but low reliability at strict thresholds (64.1% at 1.5 Å). COMPASS improves coverage to 85.9% (Apo-gd) and 84.8% (Holo-gd) at 2.0 Å, with larger gains at the more demanding 1.5 Å criterion (+13.1 pp and +15.2 pp, respectively), while requiring only 5 additional NFE per sample.

Structural Plausibility. The steric clash score serves as an indicator of structural plausibility (additional metrics are reported in Section E). Decoupling optimization from generative inference allows COMPASS to maintain a clash score of $\mathbf{0.192}/100$ atoms, strictly comparable to the unsteered BioEmu baseline (0.191), while direct 3D-guided methods (Str2Str) exhibit an elevated $0.515/100$ clash rate driven by out-of-distribution denoiser operation. This single metric quantitatively confirms that COMPASS achieves target-directed navigation without significant geometric degradation of the base model's physical priors.

Structural Transition Manifold Metrics. To quantify how accurately generated ensembles navigate the transition pathway, we compute two metrics: *Orthogonal Deviation* (the perpendicular distance of generated structures from the linear Apo-to-Holo axis) and *Transition Progress* (the scalar projection of each generated frame onto the ideal Apo (0%) → Holo (100%) vector, averaged across the ensemble).

As shown in Table 4, the unsteered BioEmu baseline yields a mean transition progress of 52.5%, consistent with the model's Holo-sampling bias (the PDB is disproportionately populated by ligand-bound states). Rather than treating this as a confound, COMPASS leverages this bias as a structural prior: Apo-guided steering counteracts the native tendency toward the Apo basin (55.0%), while Holo-guided steering advances progress to 80.5% while reducing orthogonal deviation from 2.74 Å to 1.11 Å, indicating that the generated ensemble travels further along the transition axis while remaining within the physically valid transition tube.

Table 4. Structural transition manifold metrics ($n = 98$). **Orthogonal Deviation**: perpendicular distance from the Apo→Holo axis (↓). **Transition Progress**: scalar projection of ensemble center-of-mass from Apo (0%) toward Holo (100%) (↑).

Method (Target)	Orth. Dev. (Å) ↓	Transition Progress (%) ↑
BioEmu (Unsteered)	2.74	52.5 ± 128.9
COMPASS (Apo-gd, t=0.55)	1.11	55.0 ± 33.7
COMPASS (Holo-gd, t=0.60)	**1.11**	**80.5** ± 55.1

Steerability of Extreme Local Motions. While mean global RMSD provides an aggregate view of transition success, it can mask the difficulty of capturing large-scale local deformations (e.g., hinge openings or loop closures). To quantify performance on challenging local transitions, we stratified the 92-pair benchmark by the maximum $C\alpha$ displacement between the Apo and Holo states (Table 2). The unsteered baseline struggles as local motion magnitude increases, dropping to 44.4% for extreme displacements (> 15.0 Å). In the 5.0–10.0 Å range, which typically corresponds to ligand-binding pocket openings, COMPASS achieves an 89.2% hit rate, a $+25.3\%$ absolute improvement. For extreme deformations (> 15.0 Å), COMPASS crosses the 50% success threshold (55.6%, $+11.2\%$), where the unsteered model fails entirely.

5.3. Visualization Analysis

Qualitative Morphological Alignment. Principal Component Analysis (PCA) on the structural embeddings (Figure 2) shows that the unsteered ensemble clusters near the initial Apo energy basin, while COMPASS traces a directed structural path toward the Holo state.

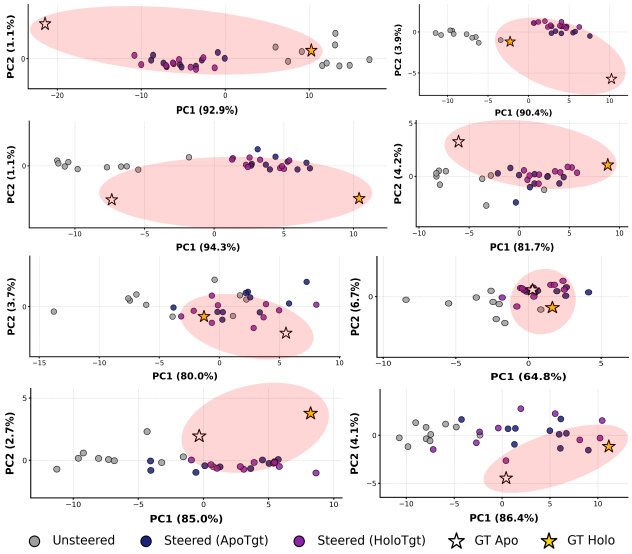

Figure 2. **Latent Space PCA Projections across 8 Target Transitions.** Structural embeddings of generated samples projected onto the principal components across 8 distinct transition pairs. The unsteered ensemble (gray) is displaced from the Apo–Holo conformational axis, while both Apo-guided (navy) and Holo-guided (purple) COMPASS samples migrate into the Apo–Holo transition ellipse (red region), indicating that latent optimization repositions conformations within the biologically relevant transition manifold. The figure panels are arranged in reading order, with the following PDB IDs representing the apo–holo pairs. (3EZM→1L5B, 2GG4→2PQC, 1A48→2CNQ, 2PRY→2PS1, 2XT4→2XT2, 2ZN9→2ZND, 3DL2→2I6T, 2CZD→2CZE)

Morphological renderings (Figure 3) further confirm alignment with the continuous induced-fit pathway.

5.4. Ablation Studies

We conducted ablation studies over both the intervention timing and the loss components (Table 5).

Intervention Timing (t). As shown in Table 5(a) and the full sweep in Section F, COMPASS maintains high hit rates across a broad range of timings ($t \in \{0.35, \ldots, 0.70\}$), substantially outperforming the unsteered baseline (64.8%). This flat performance landscape confirms that COMPASS is robust to the choice of intervention timing, a desirable property for practical application. We recommend the 5-fold CV-identified optima as balanced defaults: $t = 0.60$ for Holo-guided and $t = 0.55$ for Apo-guided steering, selected for their consistent performance across all RMSD thresholds (<1.5, <2.0, <3.0 Å) simultaneously.

Impact of Steering Loss Components. Table 5 (b) and Table A8 verify the necessity of our multi-term conditional objective. While removing either the pairwise distance regularization ($\mathcal{L}_{reg}$) or the backbone orientation constraint ($\mathcal{L}_{ori}$) still yields an +18% absolute improvement over the unsteered baseline, utilizing the full objective pushes the hit rate further to 83.7% (+19%). At the 1.5 Å threshold, combining both terms yields further improvements (Full: 78.3%, w/o $\mathcal{L}_{reg}$: 73.9%, w/o $\mathcal{L}_{ori}$: 71.7%, as detailed in Table A8), confirming that distance and orientation constraints provide orthogonal, complementary geometric information.

Overcoming Trade-offs via Dropout Ensembles. Early experiments using a single-embedding formulation on a 92-pair subset exhibited a notable steering trade-off: steering toward the Holo state substantially degraded Apo state recovery (−12.0%p), and vice-versa. Activating the Evoformer's internal dropout during inference ('deterministic=False') to generate a highly diverse conformational ensemble effectively resolves this tension. On the expanded 98-pair benchmark, the Dropout-Ensemble approach enables simultaneous performance gains across both Apo and Holo targets, pushing the Either Hit Rate to 96.9% and eliminating the cross-state performance penalty. Furthermore, at high-resolution thresholds (1.0 Å), scaling the ensemble size from 20 to 100 samples preserves significant performance margins (e.g., preventing a 9.2%p drop in Holo-guided HR), highlighting the importance of conformational stochasticity for precision steering.

5.5. Downstream Application: Ligand docking via PLACER

To assess the downstream biological relevance of the states generated by COMPASS, we performed ligand docking using PLACER (Anishchenko et al., 2025), a generative model trained for flexible side-chain ligand docking that jointly rebuilds binding-pocket side chains alongside pose prediction while keeping the backbone fixed.

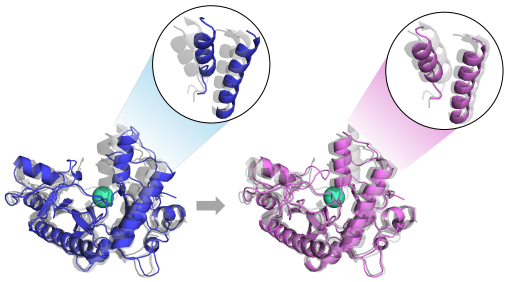

Figure 3. **Qualitative Comparison of 3D Transition Morphologies (1A48 → 2CNQ).** The left panel shows a single structure from the unsteered baseline (blue); the right panel shows the corresponding structure from COMPASS (red). The unsteered structure was selected as a favorable sample from its ensemble based on RMSD, while the COMPASS structure is an average-performing sample from its steered ensemble. Both are overlaid with the Apo state (PDB: 1A48, gray) and the target Holo state (PDB: 2CNQ, black). Mint-colored spheres mark the binding site. Close-up views highlight key structural regions for direct comparison.

Docking Pipeline. The docking process consists of two stages: 1) OpenMM-based structural relaxation and 2) PLACER generative docking. As a standard preprocessing step applied uniformly across all evaluated methods,

we first apply CPU-based OpenMM (Eastman et al., 2017) energy minimization. Raw outputs of generative diffusion models can carry residual local geometric strain at the side-chain level that may affect sensitive deep-learning docking models. During this relaxation phase, heavy atoms are fixed to preserve the predicted global backbone geometry while resolving local side-chain strain. Following relaxation, the prepared receptor structure and target ligand information are provided to the PLACER model. PLACER samples multiple highly probable binding poses within the target pocket via GPU inference. For a fair comparison, 25 structures were sampled from the AlphaFold 3 (AF3) baseline (5 random seeds $\times$ 5 samples), while 20 structures were sampled per target using BioEmu and COMPASS.

Ligand Root-Mean-Square Deviation (L-RMSD). Docking performance is evaluated using Ligand Root-Mean-Square Deviation (L-RMSD), measuring the physical displacement between the predicted ligand pose and the experimentally determined ground-truth position. To compute L-RMSD, the predicted receptor backbone is first superposed onto the ground-truth receptor backbone. The distance between corresponding ligand heavy atoms (excluding highly variable hydrogens) is then calculated. A docking attempt is considered a "hit" if the L-RMSD is ≤ 2.0 Å.

Table 5. **Ablation Studies on the 92-pair benchmark.** (a) Effect of intervention timing (t) relative to diffusion noise scale. The framework is highly robust across a wide range, with $t = 0.60$ empirically verified as optimal. (b) Impact of omitting specific structural constraints from the steering loss objective. Evaluated with Holo GT guidance.

(a) Intervention Timing Phase (t)		
Intervention Phase (t)	Hit Rate (<2.0 Å) ↑	Min Cα-RMSD (Å) ↓
Early ($t = 0.70$)	82.6%	1.97
Early-Mid ($t = 0.65$)	81.5%	2.18
Optimal ($t = 0.60$)	**83.7%**	**2.10**
Mid-Late ($t = 0.50$)	80.4%	2.23
Late ($t = 0.35$)	83.7%	2.16
(b) Intervening Loss Components (at $t = 0.60$)		
Configuration	Hit Rate (<2.0 Å) ↑	Min Cα-RMSD (Å) ↓
Baseline (Unsteered)	64.8%	2.72
w/o Dist. Reg. ($\mathcal{L}_{\text{reg}}$)	82.6%	2.16
w/o Orientation ($\mathcal{L}_{\text{ori}}$)	81.5%	2.15
Full COMPASS (Ours)	**83.7%**	**2.10**

Results and MGBench Evaluation. The 34-pair dataset covers all available Apo–Holo pairs; the 22-pair subset corresponds to targets for which AF3 Holo co-folding predictions are available, enabling direct comparison against a sequence-based co-folding baseline. As shown in Table 6, COMPASS-steered ensembles achieve higher docking hit rates compared to the unsteered baseline across both benchmarks. While Apo-guided COMPASS achieves a 63.6% hit rate on the 22-pair set, quantitatively on par with the AF3 Holo baseline, further pair-wise analysis reveals a complementary relationship. Specifically, among certain challenging targets, the two methods succeeded on disjoint subsets of pairs, indicating that latent steering provides alternative structural solutions where direct sequence-based prediction may struggle. More importantly, the observation that steering towards an Apo guidance structure successfully yields functionally viable Holo-like binding pockets highlights the framework's potential for navigating complex conformational landscapes. This parallels the cross-conformation exploration effect observed in Table 1: the same latent redistribution that broadens structural access also generates pockets compatible with the bound ligand.

Table 6. Docking Performance Summary. **Hit Rate (%)** represents the proportion of pairs where at least one generated structure achieved an L-RMSD ≤ 2.0 Å. **Avg. Best** indicates the mean of the best L-RMSD across the pair's ensemble.

Method	34-Pair Dataset		22-Pair Dataset	
	Avg. Best	Hit Rate	Avg. Best	Hit Rate
Unsteered	1.94 Å	50.0%	1.89 Å	59.1%
COMPASS (Apo)	1.87 Å	58.8%	1.81 Å	63.6%
COMPASS (Holo)	1.96 Å	55.9%	1.78 Å	68.2%
AF3 Holo	—	—	1.92 Å	63.6%

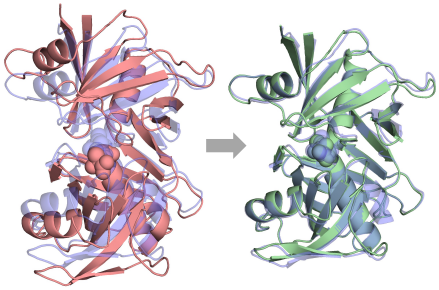

Figure 4. Visual comparison of generative docking results for the KRAS transition-state mimetic ternary complex (PDB ID: 9bgh). The experimental Ground Truth (GT) Holo structure is shown in transparent light blue. **(Left)** The AF3 Holo prediction (red) fails to open the binding pocket (receptor Cα-RMSD to GT: 4.54 Å; L-RMSD: 3.88 Å). **(Right)** The COMPASS ensemble steered by the AF3 Apo guidance structure (green) closely matches the GT backbone (receptor Cα-RMSD: 0.58 Å), achieving a substantially lower ligand L-RMSD of 1.24 Å.

We further evaluate COMPASS on 44 ternary complex structures filtered from MGBench (<2.0 Å resolution, sequence length <1000 residues), a curated benchmark of molecular glue and PPI stabilizer complexes (Table 7). We compare COMPASS steered by AF3 Apo predictions against two

static baselines: the AF3 Apo predictions themselves, and AF3 Holo predictions provided directly to PLACER. These targets are more demanding than the standard Apo–Holo benchmark. COMPASS improves upon its input guidance: while direct docking onto AF3 Apo structures yields a mean Min L-RMSD of 4.23 Å (median 3.60 Å), the COMPASS ensemble generated from those exact same Apo predictions achieves a mean Min L-RMSD of 3.31 Å (median 2.98 Å). Overall, AF3 Holo achieves the lowest aggregate error (mean 2.96 Å, median 2.71 Å), winning on 32 of 44 pairs.

Stratified analysis, however, reveals a performance split across complex types that highlights the complementary nature of these approaches. For the 9 RAS-family tricomplexes involving RMC-7977 and CypA (8tb*), AF3 Holo achieves a lower average Min L-RMSD than COMPASS (3.91 Å vs. 5.10 Å). In contrast, for the 4 KRAS complexes in which RMC-7977 and CypA pharmacologically stimulate GTP hydrolysis, modeling a catalytic transition state (Holderfield et al., 2024) (9bh*), COMPASS consistently outperforms AF3 Holo (3.71 Å vs. 4.66 Å, 3 of 4 wins). As visually demonstrated for the 9bgh complex (Figure 4), this result suggests that Apo-directed conformational sampling may access receptor conformations that static sequence-based predictors do not capture. Full pair-wise results are available in Section I.

Table 7. MGBench docking results stratified by complex type (n=44). COMPASS is steered by AF3 Apo guidance structures; AF3 Holo refers to receptor structures provided directly to PLACER. **Avg Min**: mean best L-RMSD (↓).

Complex Type	n	AF3 Holo Avg Min (Å)	COMPASS Avg Min (Å)
RAS+RMC-7977 +CypA (8tb*)	9	**3.91**	5.10
KRAS+CypA+ TS mimetic (9bh*)	4	4.66	**3.71**
14-3-3 PPI stabilizers	30	**2.51**	2.78
All MGBench	44	**2.96**	3.31

6. Conclusion

We present COMPASS, a latent steering framework for protein conformational transitions that decouples conditional prior optimization from structural inference via Tweedie's expectation, avoiding the geometric degradation of direct 3D coordinate guidance. On a curated 92-pair Apo-Holo benchmark, COMPASS improves target-state hit rates without increasing steric clashes. Notably, Apo-directed steering substantially improves Holo-state coverage, an effect we attribute to the interplay of decoupled inference and the base model's inherent Holo-sampling bias, making COMPASS useful when only Apo or predicted structures are available. Downstream docking evaluation further shows complementary structural coverage relative to AF3 Holo predictions, with COMPASS consistently outperforming AF3 on transition-state mimetic KRAS complexes where canonical structures are insufficient.

Limitations and Future Work. Several limitations remain. First, steering accuracy depends on guidance quality: if the static predictor proposes an incorrect conformation, COMPASS follows that guidance; weighting signals by epistemic uncertainty (e.g., pLDDT or PAE) is a natural extension. Second, gradient backpropagation through the score network at the anchor timestep is memory-intensive for large complexes (exceeding 1,000 residues), motivating gradient checkpointing or latent-space distillation at that scale. Third, the method uses protein-only targets; incorporating ligand features into the optimization objective may improve docking-oriented steering when binding geometry is known.

Software and Data

The complete source code, the 92-pair Apo-Holo benchmark, and pre-computed AF3 guidance structures are publicly available at `https://github.com/Chan-gyuLee/COMPASS`.

Acknowledgements

This work was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-01158, Development of a Framework for 3D Geometric Model Processing). This research was also supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (RS-2025-00558381, RS-2025-14383304), the Samsung Research Funding & Incubation Center of Samsung Electronics (SRFC-MA2502-07), and the Korea National Supercomputing Center with supercomputing resources (KSC-2025-CRE-0120) to GR Lee.

Impact Statement

This paper presents work whose goal is to advance generative machine learning in structural biology. By providing a computationally efficient method to simulate protein conformational transitions from static structures, this research can assist in structure-based drug discovery, particularly in identifying cryptic pockets and enabling flexible-receptor docking on conformationally diverse structure ensembles. While we acknowledge the dual-use nature of biological modeling tools, our framework operates by conditionally steering existing sequences without altering the underlying

generative capabilities of the base model. We do not foresee specific biosecurity risks beyond those already discussed in the broader context of protein structure prediction.

References

Abramson, J., Adler, J., Dunger, J., et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

Amemiya, T., Koike, R., Kidera, A., and Ota, M. Pscdb: a database for protein structural change upon ligand binding. *Nucleic acids research*, 40(D1):D554–D558, 2012.

Anishchenko, I. et al. Modeling protein–small molecule conformational ensembles with PLACER. *Proceedings of the National Academy of Sciences*, 122(45):e2427161122, 2025.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. The protein data bank. *Nucleic acids research*, 28(1): 235–242, 2000.

Boehr, D. D., Nussinov, R., and Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nature chemical biology*, 5(11):789–796, 2009.

Bose, A. J., Akhound-Sadegh, T., et al. Se(3)-stochastic flow matching for protein backbone generation. In *International Conference on Learning Representations*, 2024.

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., and Ye, J. C. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023.

del Alamo, D., Sala, D., Mchaourab, H. S., and Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife*, 11:e75751, 2022.

Didi, K., Zhang, Z., Zhou, G., Reidenbach, D., Cao, Z., Cha, S., Geffner, T., Dallago, C., Tang, J., Bronstein, M. M., Steinegger, M., Kucukbenli, E., Vahdat, A., and Kreis, K. Scaling atomistic protein binder design with generative pretraining and test-time compute. *arXiv preprint arXiv:2603.27950*, 2026.

Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L.-P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*, 13 (7):e1005659, 2017.

Efron, B. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Frauenfelder, H., Sligar, S. G., and Wolynes, P. G. The energy landscapes and motions of proteins. *Science*, 254 (5038):1598–1603, 1991.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Holderfield, M., Lee, B. J., Jiang, J., et al. Concurrent inhibition of oncogenic and wild-type RAS-GTP for cancer therapy. *Nature*, 629:919–926, 2024.

Jian, Y., Wu, C., Reidenbach, D., and Krishnapriyan, A. S. General binding affinity guidance for diffusion models in structure-based drug design. *Journal of Chemical Information and Modeling*, 2026.

Jing, B., Eismann, S., Suriana, P., Townshend, R. J., and Dror, R. O. Learning from protein structure with geometric vector perceptrons. In *International Conference on Learning Representations*, 2021.

Jing, B., Barzilay, R., and Jaakkola, T. Eigenfold: Generative protein structure prediction with diffusion models. In *International Conference on Learning Representations*, 2023.

Jing, B. et al. Alphafold meets flow matching for generating protein ensembles. In *International Conference on Machine Learning*, 2024.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Lewis, S. et al. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, 389 (6761):eadv9817, 2025.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637): 1123–1130, 2023.

Liu, Y., Yu, Z., Lindsay, R. J., Lin, G., Chen, M., Sahoo, A., and Hanson, S. M. ExEnDiff: An experiment-guided diffusion model for protein conformational ensemble generation. *PRX Life*, 3:023013, 2025.

Liu, Y., Zheng, H., Mao, L., Wang, Y., Chen, M., and Lin, G. Protdyn: a foundation protein language model for thermodynamics and dynamics generation. In *International Conference on Learning Representations*, 2026.

Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., and Zhu, J. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural*

Information Processing Systems, volume 35, pp. 5775–5787, 2022.

Lu, J., Zhong, B., Zhang, Z., and Tang, J. Str2str: A score-based framework for zero-shot protein conformation sampling. In *International Conference on Learning Representations*, 2024.

Maddipatla, S. A., Sellam, N. E., Bojan, M. I., Vedula, S., Schanda, P., Marx, A., and Bronstein, A. M. Inverse problems with experiment-guided AlphaFold. In *International Conference on Machine Learning*, 2025.

Monteiro da Silva, G., Cui, J. Y., Dalgarno, D. C., Lisi, G. P., and Rubenstein, B. M. High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. *Nature Communications*, 15:2464, 2024.

Saldaño, T. E., Escobedo, N., Martino, J., Heck, M., et al. Impact of protein conformational diversity on alphafold predictions. *Bioinformatics*, 38(10):2742–2748, 2022.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Stein, R. A. and Mchaourab, H. S. SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. *PLOS Computational Biology*, 18 (8):e1010483, 2022.

Wayment-Steele, H. K., Ojoawo, A., Otten, R., Apitz, J. M., Pitsawong, W., Hoemberger, M., Ovchinnikov, S., Colwell, L., and Kern, D. Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature*, 625 (7996):832–839, 2024.

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*, 27(1):293–315, 2018.

Xu, Y. et al. Tempo: Temporal multi-scale autoregressive generation of protein conformational ensembles. *arXiv preprint arXiv:2511.05510*, 2025.

A. COMPASS Algorithm

Algorithm 1 COMPASS: Decoupled Target-Directed Latent Steering

Input: Sequence S, Guidance Structure P_{guide} (e.g., Apo, Holo, or AF3 Prediction)
Output: Steered structure ensemble X_{fin}

// Phase 1: Alignment & Guidance Prep
$idx \leftarrow \text{Align}(S, P_{guide})$
$D_{ca}, D_{cb}, R_{guide} \leftarrow \text{Extract}(P_{guide}, idx)$
$E_s, E_p \leftarrow \text{ESM}(S)$ ▷ *Initial 1D/2D embeddings from ESM (Lin et al., 2023)*

// Phase 2: Prior Anchor Generation
$z_1 \sim \mathcal{N}(0, \mathbf{I})$
$x_{0.5}, R_{0.5} \leftarrow \text{Diff}(z_1 \mid E_s, E_p)_{1.0 \to 0.5}$ ▷ *Partial denoising to anchor* $t = 0.5$

// Phase 3: Physics-Aware Latent Optimization
Init offsets: $\Delta_s \leftarrow \mathbf{0},\ \Delta_p \leftarrow \mathbf{0}$
for $step = 1$ **to** 100 **do**
  $\tilde{E}_s \leftarrow E_s + \Delta_s, \quad \tilde{E}_p \leftarrow E_p + \Delta_p$
  $\hat{x}_0, \hat{R}_0 \leftarrow \text{PredX}_0(x_{0.5}, R_{0.5} \mid \tilde{E}_s, \tilde{E}_p)$ ▷ *Tweedie projection onto clean manifold*
  $\hat{x}_{cb} \leftarrow \text{VirtCB}(\hat{x}_0, \hat{R}_0)$
  $\mathcal{L}_{ca} \leftarrow \text{MSE}(\text{Dist}(\hat{x}_0),\ D_{ca})$
  $\mathcal{L}_{cb} \leftarrow \text{MSE}(\text{Dist}(\hat{x}_{cb}),\ D_{cb})$
  $\mathcal{L}_{ori} \leftarrow \text{Loss}_{ori}(\hat{R}_0, R_{guide})$
  $\mathcal{L}_{reg} \leftarrow \lambda_1 \|\Delta_s\|_2^2 + \lambda_2 \|\Delta_p\|_2^2$
  $\mathcal{L}_{tot} \leftarrow \mathcal{L}_{ca} + \mathcal{L}_{cb} + \mathcal{L}_{ori} + \mathcal{L}_{reg}$
  Update Δ_s, Δ_p via Adam w.r.t. $\mathcal{L}_{tot}$; clip grad norm ≤ 1.0; clamp $\Delta \in [-0.5, 0.5]$
end for

// Phase 4: Final Inference (No Coordinate Intervention)
$z_1' \sim \mathcal{N}(0, \mathbf{I})$
$X_{fin} \leftarrow \text{Diff}(z_1' \mid \tilde{E}_s, \tilde{E}_p)_{1.0 \to 0.001}$ ▷ *Fresh reverse diffusion on updated embeddings*
return X_{fin}

B. Extended Theoretical Proofs and Derivations

This section provides extended mathematical details for the propositions and assumptions introduced in Section 4.

B.1. Proof of Theorem B.1: Out-of-Distribution Shift via 3D Interpolation

Proposition B.1 (OOD Shift via 3D Interpolation)**.** *Let $x_t \sim p_t(x_t)$ be a valid sample from the forward VP-SDE marginal at timestep t. For any interpolation $x'_t = (1-\alpha)x_t + \alpha x_{\text{holo}}$ with step size $\alpha \in (0, 1]$, the modified state x'_t lies outside the marginal support of p_t.*

Let the marginal transition of the forward Variance-Preserving (VP) SDE be defined as $q(x_t \mid x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I)$. Any valid intermediate sample $x_t \sim p_t(x_t)$ inherently contains a stochastic noise component $\epsilon_t \sim \mathcal{N}(0, I)$ scaled precisely by $\sqrt{1-\bar{\alpha}_t}$.

If a 3D coordinate guidance method forcibly interpolates x_t with a static target x_{holo} using a step size $\alpha \in (0, 1]$:

$$x'_t = (1-\alpha)x_t + \alpha x_{\text{holo}} \tag{3}$$

The variance of the stochastic component in the modified state x'_t becomes:

$$\text{Var}(x'_t) = (1-\alpha)^2\text{Var}(x_t) = (1-\alpha)^2(1-\bar{\alpha}_t)I \tag{4}$$

Because $(1-\alpha)^2 < 1$ for any $\alpha > 0$, the artificially constructed x'_t exhibits an "over-smoothed" property with a strictly smaller noise variance than the mathematically required scale $(1-\bar{\alpha}_t)I$ at timestep t. Consequently, passing x'_t to the pre-trained score network s_θ forces the model to evaluate an out-of-distribution (OOD) input, empirically leading to structural artifacts and geometric violations during the denoising process.

B.2. Derivation of the Optimal Anchor Timestep ($t = 0.5$)

Assumption B.2 (Structural Steering Capacity)**.** The joint capacity for structural steering $\mathcal{J}(u)$ is proportional to the product of noise variance u (steerability) and retained signal information $\log(u^{-1})$ (projection fidelity), i.e., $\mathcal{J}(u) \propto -u \log u$.

The choice of the anchor timestep t^* governs the trade-off between structural plasticity (the ability to optimize embeddings) and structural coherence (the accuracy of the Tweedie projection). Let $u = 1 - \bar{\alpha}_t$ denote the noise variance. As stated in Theorem B.2, we heuristically model the joint capacity for structural steering, $\mathcal{J}(u)$, as proportional to the product of available steerability (which scales with the noise variance u) and the structural faithfulness of the projection (which correlates with the retained signal information $\log(u^{-1})$):

$$\mathcal{J}(u) \propto u \cdot \log(u^{-1}) = -u \log u \tag{5}$$

To find the optimal noise variance u^*, we set the derivative of $\mathcal{J}(u)$ with respect to u to zero:

$$\frac{d\mathcal{J}}{du} = -\log u - 1 = 0 \implies \log u = -1 \implies u^* = e^{-1} \approx 0.368 \tag{6}$$

Under the standard VP SDE schedule, the theoretically optimal signal retention is $\bar{\alpha}_{t^*} = 1 - 0.368 = 0.632$. In the discrete noise schedule used by the DPM-Solver, the temporal midpoint $t = 0.5$ yields $\bar{\alpha}_{0.5} \approx 0.75$, conceptually approximating this theoretical optimum while ensuring stable structural projections.

C. Benchmark Curation Details

To ensure objective evaluation free from selection bias, the 92-pair dataset was curated exclusively from the Protein Structural Change Database (PSCDB) (Amemiya et al., 2012). The expanded 98-pair dataset additionally includes structures from the Conformational Diversity of Native State (CoDNaS) database.

- **Initial Pool:** Started from heavily curated structural changes in PSCDB and CoDNaS.
- **Biological Relevance:** Excluded complexes with DNA/RNA or multiple heteromers (retained only valid Monomers and pure Homo-dimers). The Apo state must genuinely lack ligands, and the Holo state must contain a valid biological ligand.
- **Quality Control:** Excluded structures with missing backbone atoms or experimental resolutions worse than 2.5 Å.
- **Motion Diversity:** Unlike previous benchmarks restricted strictly to unhinged domain motions, our dataset encompasses a broad spectrum of structural rearrangements, including Loop Hinge, Domain Hinge, and Allosteric deformations.
- **Displacement Threshold:** Pairs with an Apo-to-Holo Cα-RMSD < 2.0 Å ($n = 51$), which represent near-equilibrium fluctuations, were intentionally retained to show that latent steering does not destabilize structurally stable proteins.

This filtering yielded 92 conformational transition pairs (27 Monomeric, 65 Dimeric). All structural evaluation hyperparameters (e.g., $t = 0.60$ for Holo GT) were determined via 5-fold cross-validation on the full benchmark.

D. Hyperparameters and Computational Setup

All experiments were conducted on a single NVIDIA A5000 (24GB) GPU. The base generative model was the BioEmu v1.1 checkpoint with frozen weights.

Table A1. Hyperparameter settings for Phase 3 (Physics-Aware Latent Optimization).

Hyperparameter	Description	Value
Anchor Timestep (t)	Fixed diffusion step for Tweedie projection	0.50
Optimization Steps	Iterations for Adam optimizer	100
Learning Rate	Adam optimizer learning rate	0.015
w_{CA}	Weight for $C\alpha$ pairwise distance loss	1.0
w_{CB}	Weight for virtual $C\beta$ pairwise distance loss	2.0
w_{ori}	Weight for backbone orientation loss	1.0
w_{reg}	Weight for L_2 offset regularization (λ_1, λ_2)	1.0
Gradient Clip Norm	Maximum norm for gradients	1.0
Offset Clamp Range	Mathematical bounds for Δ_s and Δ_p	$[-0.5, 0.5]$

E. Structural Plausibility Metrics

The raw structural plausibility of generated ensembles is evaluated using three standard MolProbity (Williams et al., 2018) metrics: steric clash score, Ramachandran outlier rate, and bond length violation rate. As reported in Table A2, COMPASS (both Apo- and Holo-guided) maintains clash scores strictly comparable to the unsteered BioEmu baseline, confirming that the decoupled inference preserves the base model's physical priors. All BioEmu-based methods exhibit 0.00% angle violations due to the invariant $\mathrm{SE}(3)$ local framing architecture of the model, and are thus omitted.

Note on Relaxation. Raw outputs of all diffusion-based generative models, including BioEmu and COMPASS, are expected to undergo a standard energy minimization (relaxation) step (e.g., OpenMM or Rosetta) prior to downstream applications such as docking. The metrics reported here reflect the raw generation quality before relaxation. The Ramachandran and bond violation rates are therefore expected to improve uniformly across all compared methods after relaxation, and the relative ordering of clash scores is expected to remain consistent.

Table A2. Full structural plausibility metrics on the test benchmark (raw, pre-relaxation). Lower is better for all metrics.

Method (Target)	Clash (/100) ↓	Rama Out. (%) ↓	Bond Viol. (%) ↓
BioEmu (Unsteered)	**0.191**	4.886	25.087
Str2Str (Holo-like)	0.515	**2.260**	**18.760**
COMPASS (Apo-gd)	0.192	4.259	23.652
COMPASS (Holo-gd)	0.243	4.260	23.525

F. Ablation Study: Full Intervention Timing Sweep

The following tables report the complete t-value sweep for all four steering configurations. The recommended t per configuration (bolded) was selected via 5-fold cross-validation on the training folds. Full cross-condition summary in Table A7.

Table A3. Effect of intervention timing t for Holo-guided COMPASS with experimental GT Holo structures ($n = 92$). Best per column in **bold**.

Timing (t)	HR (<1.5 Å)	HR (<2.0 Å)	HR (<3.0 Å)	Avg. Min. (Å)
$t = 0.30$	70.7%	82.6%	87.0%	2.24
$t = 0.35$	71.7%	83.7%	85.9%	2.16
$t = 0.40$	73.9%	83.7%	84.8%	2.16
$t = 0.45$	70.7%	81.5%	87.0%	2.14
$t = 0.50$	70.7%	80.4%	87.0%	2.23
$t = 0.55$	72.8%	81.5%	85.9%	2.11
$t = 0.60$	**78.3%**	**83.7%**	**88.0%**	**2.10**
$t = 0.65$	75.0%	81.5%	88.0%	2.18
$t = 0.70$	78.3%	82.6%	85.9%	1.97

Table A4. Effect of intervention timing t for Apo-guided COMPASS with experimental GT Apo structures ($n = 92$). Best per column in **bold**.

Timing (t)	HR (<1.5 Å)	HR (<2.0 Å)	HR (<3.0 Å)	Avg. Min. (Å)
$t = 0.30$	63.0%	76.1%	87.0%	2.55
$t = 0.35$	70.7%	79.3%	87.0%	2.48
$t = 0.40$	66.3%	76.1%	85.9%	2.50
$t = 0.45$	69.6%	78.3%	84.8%	2.49
$t = 0.50$	68.5%	80.4%	87.0%	2.52
$t = 0.55$	67.4%	79.3%	85.9%	**2.43**
$t = 0.60$	70.7%	**80.4%**	**87.0%**	2.36
$t = 0.65$	65.2%	78.3%	83.7%	2.42
$t = 0.70$	**71.7%**	78.3%	85.9%	2.33

Table A5. Effect of intervention timing t for Holo-guided COMPASS using AF3-predicted Holo structures ($n = 51$, pairs with available AF3 Holo prediction). Best per column in **bold**.

Timing (t)	HR (<1.5 Å)	HR (<2.0 Å)	HR (<3.0 Å)	Avg. Min. (Å)
$t = 0.30$	66.7%	78.4%	88.2%	1.82
$t = 0.35$	64.7%	74.5%	88.2%	1.98
$t = 0.40$	62.7%	82.4%	88.2%	1.82
$t = 0.45$	66.7%	82.4%	86.3%	1.73
$t = 0.50$	68.6%	84.3%	88.2%	**1.71**
$t = 0.55$	72.5%	82.4%	86.3%	1.84
$t = 0.60$	70.6%	82.4%	86.3%	1.80
$t = 0.65$	**74.5%**	82.4%	88.2%	1.85
$t = 0.70$	72.5%	**84.3%**	**90.2%**	1.68

F.1. Impact of Steering Loss Components

We perform a detailed ablation over the loss objective during the latent optimization phase. As shown in Table A8, removing either the pairwise distance constraint ($\mathcal{L}_{\text{reg}}$) or the backbone orientation constraint ($\mathcal{L}_{\text{ori}}$) slightly degrades the final hit rate. The combination of both terms in the full COMPASS objective is required to achieve the peak hit rate, demonstrating that distance and orientation provide complementary physical information for precise latent steering.

Table A6. Effect of intervention timing t for Apo-guided COMPASS using AF3-predicted Apo structures ($n = 92$). Best per column in **bold**.

Timing (t)	HR (<1.5 Å)	HR (<2.0 Å)	HR (<3.0 Å)	Avg. Min. (Å)
$t = 0.30$	51.1%	67.4%	81.5%	2.93
$t = 0.35$	51.1%	68.5%	77.2%	2.99
$t = 0.40$	47.8%	64.1%	77.2%	3.13
$t = 0.45$	53.3%	67.4%	81.5%	2.94
$t = 0.50$	48.9%	66.3%	**82.6%**	2.94
$t = 0.55$	52.2%	66.3%	77.2%	3.11
$t = 0.60$	51.1%	65.2%	79.3%	2.98
$t = 0.65$	52.2%	69.6%	**82.6%**	2.78
$t = 0.70$	**55.4%**	**71.7%**	80.4%	**2.90**

Table A7. HR (<2.0 Å) across all four steering configurations and the full t-value sweep. **Bold** marks the recommended t per configuration (selected by 5-fold CV). "GT" = experimental ground truth; "AF3" = AlphaFold 3 predicted guidance structure.

	HR (<2.0 Å) ↑			
t	Holo-GT (n=92)	Apo-GT (n=92)	Holo-AF3 (n=51)	Apo-AF3 (n=92)
0.30	82.6%	76.1%	78.4%	67.4%
0.35	83.7%	79.3%	74.5%	68.5%
0.40	83.7%	76.1%	82.4%	64.1%
0.45	81.5%	78.3%	82.4%	67.4%
0.50	80.4%	80.4%	**84.3%**	66.3%
0.55	81.5%	**79.3%**	82.4%	66.3%
0.60	**83.7%**	80.4%	82.4%	65.2%
0.65	81.5%	78.3%	82.4%	69.6%
0.70	82.6%	78.3%	84.3%	**71.7%**
Unsteered: 64.8% (n=92) / 80.4% (AF3 subset, n=51)				

Table A8. Impact of steering loss components on Holo-guided COMPASS ($n = 92$). $\mathcal{L}_{\text{reg}}$: pairwise distance regularization. $\mathcal{L}_{\text{ori}}$: backbone orientation constraint.

Configuration	HR (<1.5 Å)	HR (<2.0 Å)	HR (<3.0 Å)	Avg. Min. (Å)
Unsteered	44.0%	64.8%	82.4%	2.72
w/o $\mathcal{L}_{\text{reg}}$	73.9%	82.6%	85.9%	2.16
w/o $\mathcal{L}_{\text{ori}}$	71.7%	81.5%	88.0%	2.15
Full COMPASS (Ours)	**78.3%**	**83.7%**	**88.0%**	**2.10**

F.2. Stratification by Oligomeric State

The full benchmark consists of 27 monomeric and 65 dimeric protein pairs. Table A9 reports the target-reaching hit rates separately for these subsets. The base BioEmu model naturally struggles more with highly flexible monomers compared to tightly constrained dimers. COMPASS consistently improves the hit rates across both categories, with Holo-guided steering yielding a substantial absolute boost ($+12.6\%$ at < 2.0 Å) for the highly challenging monomeric subset.

Table A9. COMPASS performance stratified by oligomeric state. HR (<2.0 Å) and HR (<3.0 Å) are reported for Monomeric ($n = 27$) and Dimeric ($n = 65$) subsets.

Method	Guidance	Monomeric (n=27)	Dimeric (n=65)	All (n=92)
		HR (<2.0 Å)		
BioEmu Unsteered	None	61.5%	66.2%	64.8%
COMPASS	Apo-gd (t=0.55)	63.0%	**86.2%**	79.3%
COMPASS	Holo-gd (t=0.60)	**74.1%**	**87.7%**	**83.7%**
		HR (<3.0 Å)		
BioEmu Unsteered	None	73.1%	86.2%	82.4%
COMPASS	Apo-gd (t=0.55)	77.8%	89.2%	85.9%
COMPASS	Holo-gd (t=0.60)	**77.8%**	**92.3%**	**88.0%**

F.3. Stratification by Global Conformational Displacement

Table A10 categorizes performance by global Apo-to-Holo Cα RMSD (d_{conf}). The Unsteered model reaches 0% for extreme global rearrangements ($d_{\text{conf}} \geq 8$ Å), while COMPASS retains a 50% hit rate under the same conditions by directing the manifold trajectory toward the distant Holo basin via $t = 0.5$ latent gradients.

Table A10. COMPASS performance stratified by ground-truth conformational displacement d_{conf} (Apo→Holo Cα RMSD). HR (<2.0 Å) reported. Pair counts per bin in parentheses.

Method	Guidance	Small d_{conf}<2 Å (n=51)	Medium 2≤d_{conf}<4 Å (n=23)	Large 4≤d_{conf}<8 Å (n=14)	Very Large d_{conf}≥8 Å (n=4)
BioEmu Unsteered	None	76%	65%	43%	0%
COMPASS	Apo (t=0.55)	**90%**	78%	64%	0%
COMPASS	Holo (t=0.60)	**96%**	**78%**	**57%**	**50%**

G. Full Cross-Conformation Hit Rate Breakdown

Table A11 reports the complete bidirectional hit rates at all four RMSD thresholds (1.0, 1.5, 2.0, 3.0 Å), supplementing the two-threshold summary in Table 1 of the main text. A key observation across all thresholds is the *asymmetric cross-effect*: Apo-guided steering yields a disproportionately large gain in Holo coverage (e.g., +29.3 pp at 1.0 Å), while Holo-guided steering produces only a modest change in Apo coverage. This asymmetry is consistent with the base model's inherent Holo-sampling tendency: latent optimization toward the Apo target structure expands the ensemble into the transition region, which is structurally proximal to Holo.

Table A11. Full bidirectional conformational hit rates at all four RMSD thresholds ($N = 92$ pairs). $\mathbf{HR}_{\mathrm{Apo}}$: fraction of pairs where at least one generated frame falls within threshold of the Apo endpoint. $\mathbf{HR}_{\mathrm{Holo}}$: same for the Holo endpoint. $\mathbf{HR}_{\mathrm{Both}}$: hits *both* endpoints simultaneously. $\mathbf{HR}_{\mathrm{Either}}$: hits at least one endpoint. Δ_{Apo} and Δ_{Holo} denote absolute gains over the Unsteered baseline.

Method (Target)	Thresh.	$\mathrm{HR}_{\mathrm{Apo}}$	Δ_{Apo}	$\mathrm{HR}_{\mathrm{Holo}}$	Δ_{Holo}	$\mathrm{HR}_{\mathrm{Both}}$	$\mathrm{HR}_{\mathrm{Either}}$
EigenFold (Baseline)	1.0 Å	9.9	—	14.3	—	5.5	18.7
	1.5 Å	23.1	—	35.2	—	12.1	46.7
	2.0 Å	30.8	—	40.7	—	20.9	51.1
	3.0 Å	40.7	—	52.7	—	35.2	58.7
ProTDyn (Baseline)	1.0 Å	3.3	—	4.3	—	1.1	6.5
	1.5 Å	7.6	—	4.3	—	1.1	10.9
	2.0 Å	8.7	—	6.5	—	3.3	12.0
	3.0 Å	10.9	—	16.3	—	8.7	18.5
ESMFlow (Baseline)	1.0 Å	3.3	—	2.2	—	1.1	4.3
	1.5 Å	23.9	—	23.9	—	13.0	34.8
	2.0 Å	38.0	—	42.4	—	26.1	54.3
	3.0 Å	58.7	—	59.8	—	47.8	70.7
BioEmu (Unsteered)	1.0 Å	21.7	—	20.7	—	9.8	32.6
	1.5 Å	46.7	—	52.2	—	34.8	64.1
	2.0 Å	68.5	—	70.7	—	55.4	83.7
	3.0 Å	84.8	—	80.4	—	76.1	89.1
COMPASS (Apo-gd, t=0.55)	1.0 Å	16.3	−5.4	**50.0**	**+29.3**	12.0	54.3
	1.5 Å	42.4	−4.3	**70.7**	**+18.5**	35.9	77.2
	2.0 Å	58.7	−9.8	**78.3**	**+7.6**	51.1	**85.9**
	3.0 Å	76.1	−8.7	81.5	+1.1	68.5	89.1
COMPASS (Holo-gd, t=0.60)	1.0 Å	16.3	−5.4	**52.2**	**+31.5**	13.0	55.4
	1.5 Å	37.0	−9.7	**73.9**	**+21.7**	31.5	79.3
	2.0 Å	59.8	−8.7	**78.3**	**+7.6**	53.3	84.8
	3.0 Å	73.9	−10.9	**83.7**	**+3.3**	67.4	**90.2**

H. Steering Fidelity: Full Threshold Breakdown

Table A12 reports the complete conformational distribution shift for all steering configurations across four RMSD thresholds ($\leq$1.0, $\leq$1.5, $\leq$2.0, $\leq$3.0 Å). This supplements the stratified analysis in Table 1 of the main text and demonstrates that COMPASS consistently shifts probability mass toward the designated target endpoint across all stringencies.

Table A12. Steering Fidelity and Generalization to Predicted Targets. We compare the conformational distribution of the unsteered baseline against COMPASS steered by experimental ground-truth (GT) targets and AlphaFold 3 (AF3) predicted targets. The table demonstrates how COMPASS shifts the probability mass (Hit Apo vs. Hit Holo) toward the designated target across all stringencies ($1.0 \sim 3.0$ Å).

Method	Target	Thresh $\leq$ 1.0 Å		Thresh $\leq$ 1.5 Å		Thresh $\leq$ 2.0 Å		Thresh $\leq$ 3.0 Å	
		Hit Apo	Hit Holo	Hit Apo	Hit Holo	Hit Apo	Hit Holo	Hit Apo	Hit Holo
BioEmu (Unsteered)	None (n=92)	21.7%	20.7%	46.7%	52.2%	68.5%	70.7%	84.8%	80.4%
COMPASS (GT)	Apo GT (n=92)	16.3%	**50.0%**	42.4%	**70.7%**	58.7%	**78.3%**	76.1%	81.5%
COMPASS (GT)	Holo GT (n=92)	16.3%	**52.2%**	37.0%	**73.9%**	59.8%	**78.3%**	73.9%	**83.7%**
COMPASS (AF3)	AF3 Apo (n=92)	**25.0%**	25.0%	48.9%	55.4%	**66.3%**	65.2%	**82.6%**	76.1%
COMPASS (AF3)	AF3 Holo (n=51)	19.6%	**39.2%**	31.4%	**68.6%**	54.9%	**84.3%**	76.5%	**88.2%**

I. Full Generative Docking Results

This section provides the complete pair-wise evaluation data for the downstream generative docking task using PLACER. Table A13 reports the performance on the 34-pair dataset comparing the unsteered baseline with COMPASS. Table A14 presents the 22-pair benchmark results, explicitly comparing COMPASS against AlphaFold 3 (AF3) Holo predictions.

Table A13. Full Docking L-RMSD Results for the 34-Pair Benchmark.

Pair	Type	Motion	Best L-RMSD (Å) [Hit Rate %]			Status
			Unsteered	Apo-guided (COMPASS)	Holo-guided (COMPASS)	
Dimeric Targets						
71	Dimer	Loop	0.69 [6%]	1.79 [15%]	1.84 [5%]	Loss
146	Dimer	Loop	2.92 [0%]	2.53 [0%]	1.61 [5%]	Win
160	Dimer	Domain Hinge	1.92 [5%]	0.95 [40%]	0.79 [20%]	Win
2938	Dimer	Domain Hinge	2.22 [0%]	1.95 [5%]	1.49 [15%]	Win
2948	Dimer	Loop	1.01 [35%]	0.73 [25%]	0.83 [40%]	Win
4243	Dimer	Domain Hinge	2.21 [0%]	2.10 [0%]	2.52 [0%]	Win
4361	Dimer	Domain Hinge	2.17 [0%]	0.72 [30%]	0.64 [20%]	Win
Monomeric Targets						
7	Monomer	Large Scale	1.33 [10%]	1.48 [35%]	1.74 [10%]	Loss
11	Monomer	Loop	1.26 [25%]	1.24 [10%]	1.10 [25%]	Win
17	Monomer	Large Scale	0.47 [50%]	0.40 [35%]	0.53 [70%]	Win
20	Monomer	Domain Hinge	3.63 [0%]	3.87 [0%]	3.94 [0%]	Loss
41	Monomer	Loop	4.10 [0%]	0.88 [0%]	1.86 [5%]	Win
69	Monomer	Large Scale	2.03 [0%]	3.92 [0%]	2.69 [0%]	Loss
77	Monomer	Large Scale	4.77 [0%]	3.74 [0%]	3.99 [0%]	Win
96	Monomer	Large Scale	1.08 [15%]	1.08 [20%]	0.48 [30%]	Win
150	Monomer	Large Scale	4.23 [0%]	4.94 [0%]	5.00 [0%]	Loss
152	Monomer	Large Scale	0.53 [40%]	0.45 [30%]	0.68 [40%]	Win
155	Monomer	Large Scale	1.85 [5%]	1.66 [21%]	1.09 [45%]	Win
156	Monomer	Large Scale	1.87 [5%]	1.58 [15%]	1.39 [15%]	Win
157	Monomer	Large Scale	1.72 [0%]	1.48 [15%]	1.80 [0%]	Win
162	Monomer	Domain Hinge	0.72 [30%]	0.97 [35%]	1.09 [35%]	Loss
169	Monomer	Loop	1.50 [25%]	1.50 [35%]	1.78 [10%]	Loss
172	Monomer	Loop	0.78 [30%]	0.74 [55%]	0.87 [90%]	Win
173	Monomer	Loop	2.16 [0%]	2.31 [0%]	2.23 [0%]	Loss
174	Monomer	Loop	2.78 [0%]	1.63 [5%]	4.18 [0%]	Win
176	Monomer	Loop	1.31 [5%]	1.99 [5%]	3.63 [0%]	Loss
189	Monomer	Large Scale	2.56 [0%]	2.38 [0%]	2.74 [0%]	Win
196	Monomer	Loop	2.21 [0%]	2.19 [0%]	2.08 [0%]	Win
202	Monomer	Loop	1.59 [5%]	1.48 [20%]	1.04 [5%]	Win
1916	Monomer	Domain Hinge	1.88 [5%]	2.53 [0%]	2.61 [0%]	Loss
2600	Monomer	Loop	-	-	-	Loss
4035	Monomer	Loop	1.51 [0%]	3.46 [0%]	3.69 [0%]	Loss
4255	Monomer	Loop	1.14 [20%]	1.33 [20%]	0.65 [30%]	Win
4699	Monomer	Loop	-	-	-	Loss

Table A14. Full Docking L-RMSD Results for the 22-Pair Benchmark comparing against AF3 Holo Predictions.

Pair	Type	Best L-RMSD (Å) [Hit Rate %]			AF3 Holo (Å)	Status
		Unsteered	Apo-guided (COMPASS)	Holo-guided (COMPASS)		
Dimeric Targets						
71	Dimer	0.69 [6%]	1.79 [15%]	1.84 [5%]	0.50	Loss
146	Dimer	2.92 [0%]	2.53 [0%]	1.61 [5%]	0.57	Win
160	Dimer	1.92 [5%]	0.95 [40%]	0.79 [20%]	0.50	Win
2938	Dimer	2.22 [0%]	1.95 [5%]	1.49 [15%]	0.50	Win
2948	Dimer	1.01 [35%]	0.73 [25%]	0.83 [40%]	0.67	Win
4243	Dimer	2.21 [0%]	2.10 [0%]	2.52 [0%]	4.16	Win
4361	Dimer	2.17 [0%]	0.72 [30%]	0.64 [20%]	0.50	Win
Monomeric Targets						
7	Monomer	1.33 [10%]	1.48 [35%]	1.74 [10%]	0.63	Loss
11	Monomer	1.26 [25%]	1.24 [10%]	1.10 [25%]	3.88	Win
17	Monomer	0.47 [50%]	0.40 [35%]	0.53 [70%]	0.16	Win
20	Monomer	3.63 [0%]	3.87 [0%]	3.94 [0%]	0.34	Loss
77	Monomer	4.77 [0%]	3.74 [0%]	3.99 [0%]	8.06	Win
96	Monomer	1.08 [15%]	1.08 [20%]	0.48 [30%]	2.34	Win
150	Monomer	4.23 [0%]	4.94 [0%]	5.00 [0%]	1.26	Loss
152	Monomer	0.53 [40%]	0.45 [30%]	0.68 [40%]	0.18	Win
156	Monomer	1.87 [5%]	1.58 [15%]	1.39 [15%]	2.45	Win
162	Monomer	0.72 [30%]	0.97 [35%]	1.09 [35%]	2.01	Loss
169	Monomer	1.50 [25%]	1.50 [35%]	1.78 [10%]	0.84	Loss
172	Monomer	0.78 [30%]	0.74 [55%]	0.87 [90%]	0.29	Win
173	Monomer	2.16 [0%]	2.31 [0%]	2.23 [0%]	1.50	Loss
196	Monomer	2.21 [0%]	2.19 [0%]	2.08 [0%]	3.51	Win
1916	Monomer	1.88 [5%]	2.53 [0%]	2.61 [0%]	7.47	Loss

Table A15. Full Docking L-RMSD Results for the 44-Pair MGBench Benchmark comparing against AF3 Holo Predictions. The better (lower) L-RMSD value for each target is highlighted in **bold**.

Target	Min L-RMSD (Å)		Target	Min L-RMSD (Å)	
	AF3 Holo	Apo-Tgt (COMPASS)		AF3 Holo	Apo-Tgt (COMPASS)
7te8	**0.77**	1.33	8byb	**1.98**	2.00
8a62	2.57	**2.21**	8byc	2.57	**2.21**
8a65	**1.31**	2.79	8byd	**2.22**	2.41
8a68	**2.75**	3.05	8bye	**2.24**	2.43
8a6f	**2.43**	3.44	8byf	2.62	**1.83**
8a6h	**2.96**	3.63	8byg	**2.22**	2.71
8adm	**2.57**	2.75	8byo	**2.50**	2.93
8afn	3.67	**2.71**	8byy	**2.39**	2.79
8av0	**2.96**	3.15	8c0k	**2.01**	2.63
8bwj	**2.09**	2.52	8tbf	**2.79**	4.80
8bwx	**1.91**	2.77	8tbg	**4.35**	5.41
8bwz	**2.44**	3.37	8tbh	**3.93**	6.04
8bx0	**2.74**	2.86	8tbi	**3.85**	4.89
8bx3	**2.27**	3.03	8tbj	**3.62**	6.68
8bx4	**2.26**	2.54	8tbk	**3.55**	4.02
8bxi	2.45	**2.11**	8tbl	**4.64**	5.02
8bxm	**2.68**	2.76	8tbm	4.16	**3.56**
8bxn	**2.97**	3.40	8tbn	**4.33**	5.51
8bxo	3.27	**3.21**	9bgh	5.22	**3.73**
8bxq	3.42	**3.29**	9bho	5.43	**3.53**
8bxs	**1.89**	2.36	9bhq	3.46	**2.58**
8by9	**3.06**	3.49	9bi1	**4.56**	5.01