SUPPLEMENTARY INFORMATION

S.1. FLOW-BASED UNIFIED TRAINING AND INFERENCE

In this section, we further clarify the training relationships among the four flows. As shown in Fig. S1, the flows do not represent separate modalities. Instead, they are jointly optimized with shared parameters on a common SE(3)-equivariant backbone and model four residue attributes: residue type (Discrete flow), spatial position (Euclidean flow), orientation on SO(3) (Spherical flow), and side-chain torsion angles (Toric flow). All flows share an SE(3)-equivariant backbone and are trained end-to-end with a joint likelihood, while being conditioned on receptor context and PLM features via a structural adapter. At inference, the flows are executed under a common timestep: the discrete flow proposes residue types with masking, and the three geometric flows update positions, orientations, and torsions. Coupling through the shared backbone enforces global consistency.

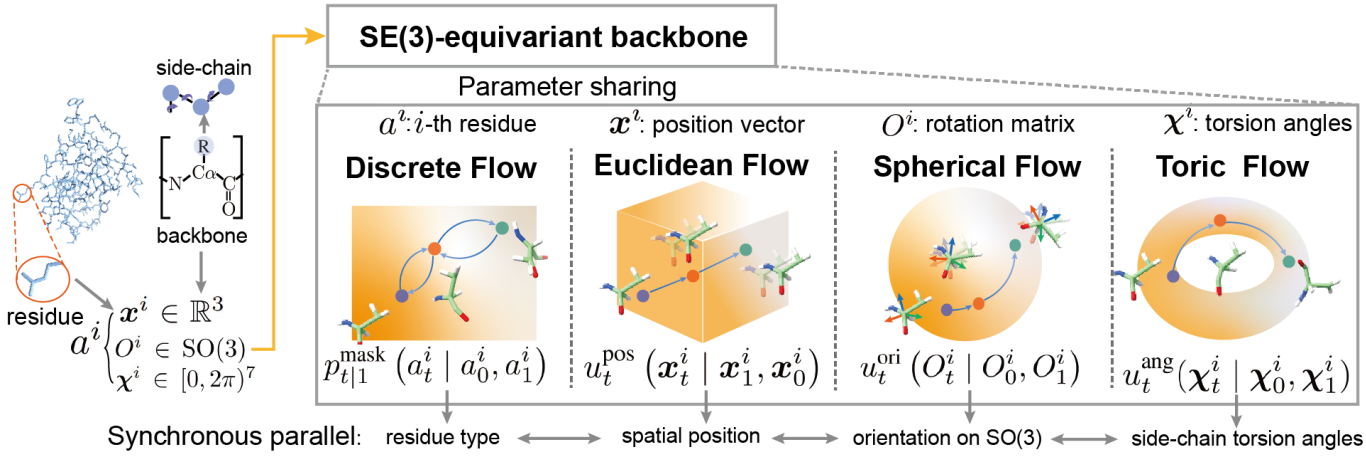

Fig. S.1: **Schematic overview of the four flows.**

S.2. TRAINING MECHANISM FOR THE STRUCTURAL ADAPTER

Our structural adapter is intentionally designed to introduce structural awareness into the PLM while ensuring stable training and minimal disruption to the pretrained knowledge encoded in the backbone. As described in Sec. IV-B, the structural adapter acts as a lightweight cross-attention module that integrates geometric features (via IPA outputs) into the PLM's hidden states. Instead of modifying PLM layers directly, the adapter injects structure-dependent signals into each layer's attention computation, enabling the PLM to refine residue-type predictions using both sequence context and intermediate structural information. During training, the PLM backbone is completely frozen, and only the adapter parameters (W_Q, W_K, W_V) and the geometric network are updated. Freezing the PLM backbone preserves the rich evolutionary priors learned from large-scale sequence corpora and avoids catastrophic forgetting. Gradients flow from the flow-matching losses through the adapter and IPA features, but stop at the PLM backbone, ensuring that structural supervision shapes only the adapter without destabilizing the pretrained model. Because only the adapter weights are optimized, fewer than $\mathbf{2-3}\%$ of the PLM parameters are trainable in our implementation. This lightweight design keeps the computational cost low and aligns with prior work demonstrating that small structural adapters are sufficient to infuse geometric awareness into large PLMs.

S.3. CONTROLNET-STYLE ARCHITECTURE

Our goal in adopting ControlNet-style architecture is to enable receptor-aware generation while preserving the pretrained co-design capabilities of the base flow model. As detailed in Sec. IV-C, we follow the ControlNet paradigm by maintaining a frozen base flow model F_Θ and attaching a trainable copy F_{Θ_c} connected through zero-initialized IPA blocks $Z_{\Theta_{z1}}$,$Z_{\Theta_{z2}}$.This effectively enables receptor-aware flow matching, where the learned control branch refines the probability path without overwriting pretrained knowledge. During fine-tuning, we introduce receptor awareness through a zero-initialized control branch that leaves the pretrained flow unchanged at the start. All control pathways are set to zero, so the model initially behaves exactly like the base flow trained on general proteins. As optimization proceeds, gradients through the zero-initialized adapters allow the trainable branch to learn a receptor-conditioned modification of the flow field. The receptor structure C_{rec} then modulates the predicted velocity fields for position, orientation, and torsion $\left(v_{\text{pos}}, v_{\text{ori}}, v_{\text{ang}}\right)$ via these adapter pathways, shaping the peptide trajectory C^{pep}_t toward conformations compatible with the target receptor. This procedure yields receptor-aware flow matching in which the learned control branch refines the probability path while preserving the pretrained knowledge and avoiding destabilizing shifts. Peptide-receptor complexes are scarce, and directly fine-tuning the entire model on receptor specific data risks overfitting and performance collapse. The ControlNet mechanism provides a principled way to preserve

general structural priors learned from large PDB pretraining, inject task-specific geometric guidance, and avoid instability from full-parameter fine-tuning.

Operational mechanism in 3D (zero-IPA initialization). Our conditioning scheme follows the spirit of ControlNet but is instantiated with SE(3)-equivariant Invariant Point Attention (IPA) blocks operating on 3D protein structures. Concretely, we first pretrain an unconditioned co-design flow $F_\Theta : C_0 \mapsto \hat{C}_{pro}$ on general PDB monomers to model peptide sequence-structure pairs without any input of receptor. Then, we freeze the pretrained backbone F_Θ. Clone the backbone to a trainable copy F_{Θ_c} that ingests receptor features. Connect the two branches via two zero-initialized IPA modules $Z_{\Theta_{z1}}$,$Z_{\Theta_{z2}}$, as in Eq. (19). Both the weights and biases of the zero-IPAs are initialized to zero, so at the beginning of finetuning we have $Z_{\Theta_{z1}}(C_{rec}) = 0, Z_{\Theta_{z2}}(\cdot) = 0 \Rightarrow \hat{C}_pep = F_\Theta(C_0)$ for any receptor C_{rec} . Thus, the conditioned model is exactly equivalent to the pretrained unconditioned flow at initialization: receptor features cannot perturb the 3D latent representation until the zero-IPA weights deviate from zero. Because the ControlNet branch is added as a residual SE(3)-equivariant correction on top of the frozen backbone, the pretrained geometric features and flows remain intact, and receptor-dependent effects are "turned on" gradually as training progresses.

S.4. Details and Summary of Baseline Models

We selected five baseline models for comparison with D-Flow, their details are as follows,

- RFDiffusion [19]: an SE(3)-equivariant diffusion framework fine-tuned from RoseTTAFold that denoises residue-frame backbones under structural constraints (e.g., motifs/symmetry) to design proteins and binders.
- ProteinGenerator [39]: a RoseTTAFold-based sequence diffusion model that samples sequences while jointly predicting 3D structure, with optional conditioning on coordinates/DSSP for motif scaffolding.
- Diffusion [22]: An SE(3)-equivariant diffusion approach for CDR design that models per-residue type, $C\alpha$ position and orientation, enabling antigen-conditioned sampling.
- PPFlow [2]: a target-aware conditional flow-matching model that performs flows on heterogeneous manifolds, torus for torsions, SO(3) for rotations, $\mathbb{R}^3$ for translations, plus a categorical type flow, to generate full-atom peptides.
- PepFlow [3]: An E(3)-equivariant conditional flow-matching framework that produces full-atom peptides by flowing backbone/side-chain geometry, using a hypernetwork to inject sequence-specific parameters.

In addition, we provide a summary table comparing and contrasting D-Flow with the other models (see Table S.1).

TABLE S.1: Analytical comparison of architectures and modeling spaces. "PLM" denotes whether protein language model features are integrated. "Chirality" indicates explicit support for D-peptide design.

Method	Objective	Representation	Manifolds	Conditioning	PLM	Chirality
RFdiffusion	Diffusion	Residue/backbone frames	SE(3) frames	Motif/pocket & binder	No	L-centric
ProteinGenerator	Diffusion (seq-space)	Sequence + structure (joint)	Categorical (+ SE(3) cond.)	Seq/structure constraints	No	L-centric
Antibody Diffusion	Diffusion	AA type + Cα pos./orientation	Categorical + $\mathbb{R}^3$ + SO(3)	Antigen/interface	No	L-centric
PepFlow	Flow matching (CFM)	Backbone + side-chain torsions	SE(3) + T^n (+ categorical)	Interface cues	No	L-centric
PPFlow	Flow matching	Backbone/interface + torsions	SE(3) + T^n	Protein–protein interface	No	L-centric
D-Flow (ours)[0†]	**Flow matching**	**Full-atom (type/pos./orient./torsion)**	**Categorical + $\mathbb{R}^3$ + SO(3) + T^n**	**Receptor-aware**	**Yes (adapter)**	**L & D (mirror)**

S.5. Robustness Assessment of Results

To assess the impact of the structural adapter, we compared the full D-Flow model with a variant in which the adapter is removed and only sequence features are provided to the PLM, using the large-scale dataset released by Kong et al. [24]. The results indicate that the structural adapter provides a stable improvement in structural-sequence consistency by integrating receptor-side geometry, while the effect size remains measured and realistic.

In addition, we report the mean ± s.d. over three runs for the two most relevant models: PepFlow-Ang (the strongest baseline) and D-Flow (ours). As shown in Table S.2, all metrics remain stable across seeds, with D-Flow consistently outperforming PepFlow-Ang. Across all seeds, D-Flow consistently increases AAR by 7.44% with a small s.d. ($< 0.5\%$), indicating robustness. The geometric gain is likewise stable, reducing RMSD by 0.44Å on average. Improvements in stability (+8.7%) and affinity (+2.9%) exceed run-to-run variation by more than $7\times$. These results confirm that D-Flow's gains over the strongest baseline are statistically significant, reproducible under multi-seed evaluation, and not attributable to stochastic variation. We appreciate the suggestion, which has made our results more robust and credible.

S.6. omparison of Discrete Flow Matching and Diffusion Models

As shown in Table S.3, D-Flow consistently outperforms diffusion baselines across geometric (AAR, RMSD, SSR, BSR), energetic (stability, affinity), and designability/diversity metrics, indicating that under a comparable evaluation protocol, discrete flow matching (DFM) yields higher-quality and more robust peptide designs. We further quantify optimisation stability over

[0†] D-Flow uses an SE(3)-equivariant backbone, flows are defined on the listed manifolds (categorical, $\mathbb{R}^3$, SO(3), T^n) rather than a single SE(3) variable.

TABLE S.2: **Quantitative Ablation of the Structural Adapter. AAR: amino acid recovery; C-RMSD: Cα RMSD (Å); L-RMSD: ligand RMSD (Å).**

Model	**AAR**	**C-RMSD**	**L-RMSD**
Without structural adapter	36.12%	2.98	1.74
D-Flow	39.01%	2.70	1.54

TABLE S.3: Three-run mean $\pm$ s.d. on PepMerge.

Model	AAR ↑	RMSD (Å) ↓	SSR ↑	BSR ↑	Stability ↑	Affinity ↑
PepFlow-Ang	51.25 ± 1.42	2.07 ± 0.05	83.46 ± 0.31	86.89 ± 0.28	18.15 ± 0.36	21.37 ± 0.33
D-Flow (ours)	58.69 ± 0.37	1.63 ± 0.04	89.02 ± 0.29	88.47 ± 0.27	26.85 ± 0.41	24.31 ± 0.30

three random seeds (Table S.3), where DFM shows 3–4$\times$ lower variance in both AAR and RMSD, reflecting smoother optimisation and a more stable likelihood landscape. At inference, under identical conditioning with three repeated samples, AAR variance is $\pm 0.41\%$ for DFM vs. $\pm 1.89\%$ for diffusion, confirming the numerical-stability advantage of deterministic DFM trajectories.

TABLE S.4: Validation variability under matched settings.

Metric (validation)	**DFM (ours)**	**Diffusion (matched)**
AAR std	$\pm 0.37\%$	$\pm 1.62\%$
RMSD std	$\pm 0.04\,\mathring{A}$	$\pm 0.11\,\mathring{A}$