MoDA: A Modality-Adaptive Flow Matching Framework for Antibody Co-Design

Anonymous authors
Paper under double-blind review

Abstract

Antibody co-design methods usually model modalities uniformly, overlooking biophysical heterogeneity in scale-dependent constraints, residue roles, and modality-specific contextual requirements. To address these challenges, we propose MoDA, a modality-adaptive flow matching framework that integrates multi-scale constraints with residue role modeling. We first develop multi-scale constraint modeling to capture disentangled semantic information from local regions to global complex levels via decomposing functional unit representations. Then, we introduce cooperation-aware role modeling to infer residue roles based on residue-specific constraints and inter-residue coordination. Coupling the above constraints with inferred role assignments, we propose modality-adaptive coordinative optimization to steer asymmetric message passing and modality-specific interaction updates. Experiments on the RAbD benchmark demonstrate that MoDA achieves state-of-the-art performance, significantly improving structural fidelity (RMSD, TM-score, lDDT) and target binding affinity.

1 Introduction

Antibodies are adaptive immune proteins essential for therapeutics and diagnostics (Sela-Culang et al., 2013). A conventional antibody comprises paired heavy and light chains, each containing four conserved framework regions (FRs) that serve as structural scaffolds and three hypervariable complementarity-determining regions (CDRs) that form the antigen-binding paratope (Dunbar et al., 2014). Computational *de novo* antibody design aims to reduce the high cost of wet-lab experiments by jointly generating CDR sequences and structures with high binding affinity.

Existing co-design methods broadly fall into two modeling paradigms: (1) *Discriminative models* treat CDR design as conditional translation or iterative optimization from context. MEAN updates sequences and coordinates via equivariant graph translation (Kong et al., 2022), while dyMEAN incorporates dynamic geometric reasoning to refine flexible conformations iteratively (Kong et al., 2023). While these methods leverage interface geometric priors, their deterministic formulation restricts generative diversity and hinders exploration of the multi-modal conformational landscape for high-affinity candidates. (2) *Generative models* capture joint sequence–conformation distributions to sample diverse candidates. DiffAb pioneers antigen-specific diffusion for joint generation (Luo et al., 2022); AbDiffuser incorporates explicit physical and geometric constraints for full-atom diffusion (Martinkus et al., 2023); dyAb exploits pre-binding antigen conformations to capture flexible interaction dynamics (Tan et al., 2025), and AbFlow introduces interaction-enhanced flow matching for paratope-centric co-design (Wang et al., 2026).

Despite these advances, existing generative methods overlook inherent heterogeneity in three aspects: (1) **Functional constraint heterogeneity.** Residue-level constraints ignore antibodies' multi-scale biophysical organization, weakening overall conformational stability and paratope functionality. (2) **Residue role heterogeneity.** Homogeneous interaction treats residues uniformly, failing to capture specialized residue contributions and cooperative synergies crucial for coordinated paratope–epitope recognition and binding stability. (3) **Modality requirement heterogeneity.** Symmetric modal coupling neglects the distinct contextual demands of different semantic and generation spaces, hindering modality information preservation and cross-modal coordination.

To address these issues, we propose **MoDA**, a **Mod**ality-**a**daptive flow matching framework (Lipman et al., 2023) that integrates multi-scale constraints with residue role modeling. Specifically,

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.

we first introduce **multi-scale constraint modeling** to decouple functional unit representations across region, chain, pair, and complex levels into fundamental, inherent, and synergistic constraints. Based on these decoupled constraints, we design **cooperation-aware role modeling** to infer residue-specific functional roles and refine them via a pairwise symmetric cooperation. By integrating the constraints with calibrated roles, we propose **modality-adaptive coordinative optimization** to update sequence, position, and orientation spaces respectively via demand–supply modulated asymmetric message passing and modality-specific coordination modeling.

Evaluations on the standard RAbD benchmark (Adolf-Bryfogle et al., 2018) (60 complexes spanning diverse antigen classes and binding topologies) demonstrate that MoDA achieves state-of-the-art performance on full paratope co-design. MoDA outperforms existing baselines in structural fidelity, reducing RMSD from 1.8139 to 1.5780 while consistently improving TM-score, G-TM, lDDT, and G-lDDT. Meanwhile, MoDA effectively samples diverse candidates that achieve the highest binding-energy success rate across benchmark targets.

Our main contributions are as follows:

- We formalize antibody generative heterogeneity across scale, role, and modality, capturing them via a modality-adaptive framework with role-guided multi-level regulations.
- We introduce multi-scale constraint and cooperation-aware role modeling to construct disentangled functional constraints and infer residue role distribution.
- We propose a modality-adaptive coordinative optimization mechanism enabling asymmetric message passing and modality-specific context, adaptively coordinating updates.
- Extensive benchmarks demonstrate that MoDA establishes state-of-the-art performance in structural fidelity and binding affinity.

2 Related Work

Antibody sequence–structure design. Recent antibody design methods increasingly focus on joint modeling of sequence and structure. AbX incorporates multi-level physical and evolutionary priors into diffusion (Zhu et al., 2024). RAAD models multi-relational graphs equivariantly for epitope-unknown design (Wu et al., 2025). AbFlow leverages interaction-enhanced flow matching for paratope generation (Wang et al., 2026). DGENet iteratively co-designs full-atom CDRs under equivariant geometric and kinematic constraints (Huang et al., 2026). AAMFM learns shared sequence–structure representations conditioned on antigen geometry and epitopes (Shi et al., 2026). While these models achieve promising performance, their residue-level constraints oversimplify interactions, but our framework adopts multi-scale biophysical constraints with explicit residue functional roles for paratope design.

Multi-modal coupling in protein design. Protein sequence and structure form heterogeneous modalities whose coordinated modeling is essential for generating candidates. Pallatom employs dual-track residue and atom representations for consistent all-atom generation (Qu et al., 2024). HDProt aligns discrete sequence tokens with continuous diffusion-based structure tokens (Zhou et al., 2025). DPLM-2 tokenizes structures to perform joint sequence–structure denoising within a discrete diffusion framework (Wang et al., 2025). La-Proteina encodes sequence and atomistic details into residue latents over coarse backbones (Geffner et al., 2026). While these methods enforce sequence–structure consistency, they ignore modality-specific contextual requirements; MoDA models modality-specific contextual demands via modality-adaptive optimization.

3 Method

3.1 Problem Formulation

We formulate complete antibody CDR co-design on an antibody–antigen complex $\mathcal{V} = \mathcal{V}_{\mathrm{Ab}} \cup \mathcal{V}_{\mathrm{Ag}}$ of length $L = |\mathcal{V}|$. The antigen $\mathcal{V}_{\mathrm{Ag}}$ contains a target epitope $\mathcal{V}_E \subseteq \mathcal{V}_{\mathrm{Ag}}$, while the antibody $\mathcal{V}_{\mathrm{Ab}}$ comprises framework regions (FRs) $\mathcal{V}_F$ and n CDR loops $\mathcal{V}_C = \bigcup_{k=1}^{n} \mathcal{V}_{C_k}$. Each residue $i \in \mathcal{V}$ involves three modalities: an amino acid type s_i, a C_α coordinate $p_i \in \mathbb{R}^3$, and an orientation

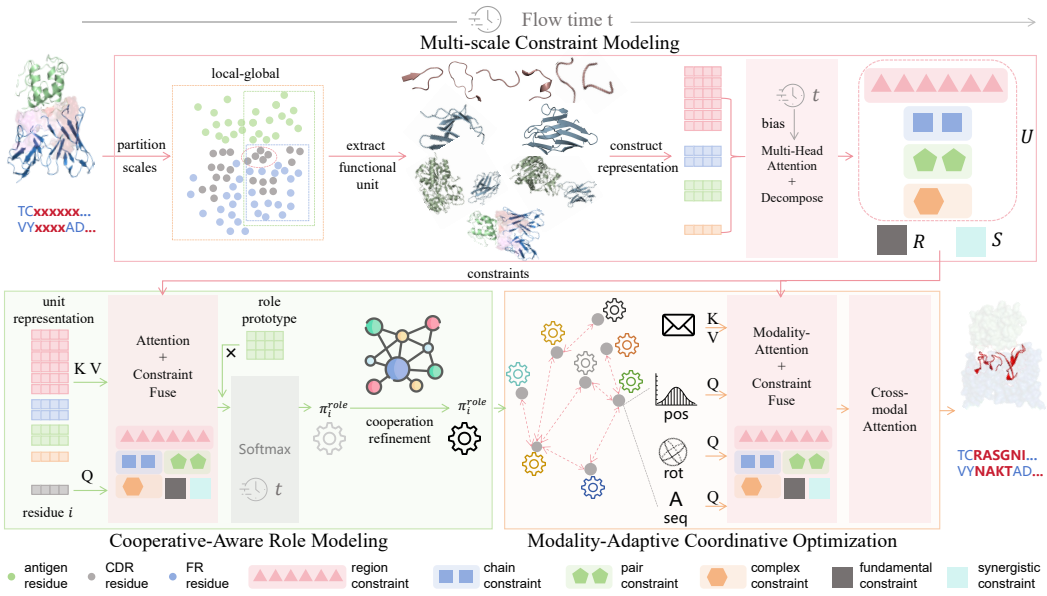


Figure 1: At timestep t, *Multi-Scale Constraint Modeling* extracts disentangled constraints to condition *Cooperative-Aware Role Modeling* to infer residue roles, which guide *Modality-Adaptive Coordinative Optimization* for asymmetric interaction and modality-specific update.

frame $R_i \in \mathrm{SO}(3)$, with backbone coordinates $x_i = [x_i^{\mathrm{N}}, x_i^{\mathrm{C}_\alpha}, x_i^{\mathrm{C}}, x_i^{\mathrm{O}}] \in \mathbb{R}^{4\times 3}$. Given the fixed context $\mathcal{C}_E = (S_{\mathcal{V}_F}, P_{\mathcal{V}_F}, R_{\mathcal{V}_F}, S_{\mathcal{V}_E}, P_{\mathcal{V}_E}, R_{\mathcal{V}_E})$, non-padding residue mask M^{valid}, and generation residue mask M^{gen}, we denote the complex state at timestep $t \in [0,1]$ as $\mathcal{C}_t = (S_t, P_t, R_t)$. Our goal is to jointly generate CDR sequences $S_{\mathcal{V}_C}$, positions $P_{\mathcal{V}_C}$, and orientations $R_{\mathcal{V}_C}$ by modeling the conditional distribution $p_\theta(S_{\mathcal{V}_C}, P_{\mathcal{V}_C}, R_{\mathcal{V}_C} \mid \mathcal{C}_E)$.

3.2 Overview

We propose **MoDA**, illustrated in Fig. 1, to model multi-level heterogeneities for antibody design. We first introduce **multi-scale constraint modeling** to extract functional unit representations from region-to-complex and disentangle them into fundamental, inherent, and synergistic constraints. Then **cooperation-aware role modeling** infers residue functional roles conditioned on these constraints and refines them via cooperative interaction. Jointly accounting for constraints and residue roles, we propose **modality-adaptive coordinative optimization** to guide role-directed residue interaction and adaptively coordinate cross-modal updates based on modality-specific context.

3.3 Multi-scale Constraints Modeling

To capture multi-scale biophysical requirements, we organize the antibody–antigen complex into region-, chain-, pair-, and complex-level hierarchies. We then extract and model functional constraints via dedicated attention mechanisms and reliability-aware representation decoupling, which includes three steps: **(1) Global base representation construction.** To establish a unified context, we construct a global base representation using a geometric attention encoder (GA Encoder). We first encode the complex $\mathcal{C}_t$ at generation step t using GA encoder $\mathcal{G}$, obtaining initial residue representations $H^{\mathrm{GA}} = [h_1^{\mathrm{GA}}, \ldots, h_L^{\mathrm{GA}}] \in \mathbb{R}^{L\times d}$ for each residue $i \in \mathcal{V}$. For any residue pair (i,j), we build pairwise relation features $U_{ij}^{\mathrm{relative}} \in \mathbb{R}^d$ by encoding relative amino acid type, relative distances, and relative dihedrals (detailed in App. A.1.1). As shown in Eq. 1, we aggregate relational features across valid residues j via projection W_{pair} and add temporal schedule embedding c_t to obtain residue-level base representation h_i^{base}; stacking them yields base representation H^{base}.

$$h_i^{\mathrm{base}} = h_i^{\mathrm{GA}} + W_{\mathrm{pair}} \left(\frac{\sum_{j=1}^{L} M_j^{\mathrm{valid}} U_{ij}^{\mathrm{relative}}}{\sum_{j=1}^{L} M_j^{\mathrm{valid}}} \right) + c_t, \quad H^{\mathrm{base}} = \left[h_i^{\mathrm{base}} \right]_{i\in\mathcal{V}} \tag{1}$$

(2) Functional unit representation modeling. To quantify multi-scale information, we identify functional units of each scale and model their representation based on distinct attention mechanisms. The scale levels is $\mathcal{S} = \{\text{region}, \text{chain}, \text{pair}, \text{complex}\}$, and functional units in each scale $s \in \mathcal{S}$ is $\mathcal{E}_s$ with unit indicators $m_e^s \in \{0,1\}^L$ and unit-level base representation $H_e^{\text{base}} = [h_i^{\text{base}}]_{i \in e}$.

Region-scale modeling ($e \in \mathcal{E}_{\text{region}} = \{H1\text{–}H3, L1\text{–}L3\}$). CDR loops share common biophysical properties, yet distinct CDRs serve specialized antigen-recognition roles. To capture common intra-loop features while preserving CDR-specific characteristics, we deploy multi-head self-attention ($\text{MHSA}_{\text{region}}$) (Vaswani et al., 2017) with CDR-specific low-rank adapters (Hu et al., 2022). For each CDR unit e and inner residue i, $\text{MHSA}_{\text{region}}$ aggregates contexts from CDR residues $j \in \mathcal{O}_e^{\text{region}}(i)$ to yield general CDR features H_e^{region}, while low-rank adapters $(A_e^{\text{region}}, B_e^{\text{region}})$ produce CDR-specific modulations $\Delta H_e^{\text{region}}$ (details in App. A.1.2). The region-scale representation is $Y_e^{\text{region}} = U_e^{\text{region}} + \text{FFN}_{\text{region}}(\text{LN}(U_e^{\text{region}}))$, where $U_e^{\text{region}} = H_e^{\text{base}} + \Delta H_e^{\text{region}}$.

Chain-scale modeling ($e \in \mathcal{E}_{\text{chain}} = \{H, L\}$). Intra-chain residues share backbone continuity, whereas distinct chains exhibit divergent biophysical characteristics. Similar to region-scale modeling, for a residue i, we use a chain attention encoder $\text{MHSA}_{\text{chain}}$ (Vaswani et al., 2017) over same chain domain $\mathcal{O}_e^{\text{chain}}(i) = \{j \in \mathcal{V} \mid j \in e\}$ to obtain a chain-general representation H_e^{chain}, while chain-specific low-rank adapters $(A_e^{\text{chain}}, B_e^{\text{chain}})$ extract chain characteristics $\Delta H_e^{\text{chain}}$ (Hu et al., 2022). We obtain the final chain-scale representation based on aggregated state $U_e^{\text{chain}} = H_e^{\text{base}} + H_e^{\text{chain}} + \Delta H_e^{\text{chain}}$, formulated as $Y_e^{\text{chain}} = U_e^{\text{chain}} + \text{FFN}_{\text{chain}}(\text{LN}(U_e^{\text{chain}}))$.

Pair-scale modeling ($e \in \mathcal{E}_{\text{pair}} = \{\text{HL}, \text{AbAg}\}$). Inter-chain coordination encompasses internal heavy–light chain pairing and directed antibody–antigen interfacial recognition. We model these interactions via a cross-attention encoder and a directed spatial-attention encoder with asymmetric fusion, respectively. For heavy–light coordination, each residue $i \in \mathcal{V}_{\text{Ab}}$ interacts with opposing-chain residues $j \in \mathcal{O}_{\text{HL}}^{\text{pair}}(i)$ via cross-attention and produces a context vector $c_i^{\text{HL}} = \sum_j \alpha_{ij}^{\text{HL}} W_V h_j^{\text{base}}$. We fuse base features, context, their interactive product, and discrepancy into a pair-HL representation $y_{\text{HL},i}^{\text{pair}} = f_{\text{HL}}^{\text{pair}}([h_i^{\text{base}} \| c_i^{\text{HL}} \| h_i^{\text{base}} \odot c_i^{\text{HL}} \| |h_i^{\text{base}} - c_i^{\text{HL}}|])$. For antibody–antigen interaction, we define directed relations $r \in \mathcal{R}_{\text{AbAg}} = \{\text{Ag} \to \text{Ab}, \text{H} \to \text{Ag}, \text{L} \to \text{Ag}\}$. The relation-corresponding domain $\mathcal{O}_r^{\text{pair}}(i)$ of residue i is the set of residues in the target chain (detailed in App. A.1.2). For pairs within this domain, we compute spatial coupling scores as $s_{ij} = f_{\text{sp}}([h_i^{\text{base}} \| h_j^{\text{base}} \| (p_{j,t} - p_{i,t}) \| \|p_{j,t} - p_{i,t}\|_2])$. Aggregating base representations within the domain using attention weights $\alpha_{ij}^r = \text{Softmax}_{j \in \mathcal{O}_r^{\text{pair}}(i)}(s_{ij})$ yields directed contexts $c_i^r = \sum_{j \in \mathcal{O}_r^{\text{pair}}(i)} \alpha_{ij}^r h_j^{\text{base}}$. As formulated in Eq. 2, we map directed contexts into asymmetric representations $y_{\text{AbAg},i}^{\text{pair}}$. Finally, we stack these vectors to obtain the pair-scale representations $Y_{\text{HL}}^{\text{pair}} = [y_{\text{HL},i}^{\text{pair}}]_{i \in \mathcal{V}_{\text{Ab}}}$ and $Y_{\text{AbAg}}^{\text{pair}} = [y_{\text{AbAg},i}^{\text{pair}}]_{i \in \mathcal{V}}$.

$$y_{\text{AbAg},i}^{\text{pair}} = \begin{cases} f_{\text{Ab}}^{\text{pair}}\left(\left[h_i^{\text{base}} \| c_i^{\text{Ag}\to\text{Ab}} \| c_i^{\text{HL}} \| h_i^{\text{base}} \odot c_i^{\text{Ag}\to\text{Ab}}\right]\right), & i \in \mathcal{V}_{\text{Ab}}, \\ f_{\text{Ag}}^{\text{pair}}\left(\left[h_i^{\text{base}} \| c_i^{H\to\text{Ag}} \| c_i^{L\to\text{Ag}} \| h_i^{\text{base}} \odot (c_i^{H\to\text{Ag}} + c_i^{L\to\text{Ag}})\right]\right), & i \in \mathcal{V}_{\text{Ag}}. \end{cases} \tag{2}$$

Complex-scale modeling ($e \in \mathcal{E}_{\text{complex}} = \{\text{complex}\}$). To capture complex-level stability and structural viability, we aggregate global context via component pooling and multi-head self-attention ($\text{MHSA}_{\text{complex}}$). Specifically, for each $u \in \{\text{all}, \text{H}, \text{L}, \text{Ag}\}$, masked pooling extracts component tokens $z_u^{\text{cpn}} = \frac{1}{|\mathcal{V}_u|} m_u^\top H^{\text{base}}$ and the interface token $z_{\text{int}}^{\text{combine}} = \frac{1}{|\mathcal{V}_{\text{Ab}}|} m_{\text{Ab}}^\top Y_{\text{AbAg}}^{\text{pair}}$. Based on them, we compute global context $g^{\text{complex}} = \text{MHSA}^{\text{complex}}([z_{\text{all}}^{\text{cpn}}, z_{\text{H}}^{\text{cpn}}, z_{\text{L}}^{\text{cpn}}, z_{\text{Ag}}^{\text{cpn}}, z_{\text{int}}^{\text{combine}}]^\top)$ and obtain complex-scale representation $Y^{\text{complex}} = [h_i^{\text{base}} + g^{\text{complex}}]_{i \in \mathcal{V}}$.

(3) Reliability-aware constraint decomposition. To prevent representation redundancy and coupling induced by cross-scale residue sharing, we introduce a reliability-aware constraint decomposition that disentangles representations into fundamental (R), unit-specific inherent (U), and synergistic (S) constraints. We first introduce a power-law fidelity schedule β_t^γ to measure semantic uncertainty under noise perturbation, where the learnable parameter γ controls noise sensitivity. Then we construct unit reliability $q_e(t) = \frac{1}{|\mathcal{V}_e|} m_e^\top (\mathbf{1} - M^{\text{gen}} + M^{\text{gen}} \beta_t^\gamma)$ to suppress interference. As shown in Eq. 3, for each unit $e \in \mathcal{E}$ with pooled token $z_e^s = \frac{1}{|\mathcal{V}_e|} \sum_{i \in \mathcal{V}_e} Y_{e,i}^s$, we design a reliability-guided cross-unit multi-head attention mechanism incorporating the reliability bias

$\log q_{e'}(t)$ from another unit e'. After aggregating and projecting across heads to obtain the updated unit tokens $\widehat{z}_e^s$, we stack them into a contextualized matrix Z^{cross} and extract the decoupled constraints: $R = f_R(Z^{\text{cross}}), \quad S = f_S(Z^{\text{cross}}), \quad U_e^s = f_{U_e^s}(\widehat{z}_e^s)$.

$$\alpha_{ee'}^m = \text{Softmax}_{e' \in \mathcal{E}} \left(\frac{(W_Q^{\text{RUS},m} \,\text{LN}(z_e^s))^\top (W_K^{\text{RUS},m} \,\text{LN}(z_{e'}^s))}{\sqrt{d_h}} + \log q_{e'}(t) \right) \tag{3}$$

3.4 Cooperation-Aware Role Modeling

Residue functional role characterizes its specific contribution to overall antibody activity. We model residue role based on residue-specific constraints and inter-residue cooperation. Specifically, it involves two steps: **(1) intrinsic role inference based on specific constraints.** Residues within the same functional unit take distinct functional roles. For a residue i, we first use an attention encoder to aggregate intra-unit context $c_{e,i}^s = \sum_{j \in e} \alpha_{ij}^{s,e} W_V Y_{e,j}^s$. Then we fuse it with unit inherent constraints U_e^s, residue base representation, and reliability $q_e(t)$ to obtain residue single-unit constraint $x_{e,i}^s = f_{\text{read},e}^s([h_i^{\text{base}} \| c_{e,i}^s \| W_{U,e}^s U_e^s \| q_e(t)])$. By concatenating these constraints across all units covering residue i, we obtain residue-specific unit constraint $x_i^{\text{AU}} = [x_{\text{complex},i}^{\text{complex}} \| x_{\text{HL},i}^{\text{pair}} \| x_{\text{AbAg},i}^{\text{pair}} \| x_{\text{chain}(i),i}^{\text{chain}} \| x_{\text{CDR}(i),i}^{\text{region}}]$. We introduce a learnable role prototype matrix $P = [p_1, \ldots, p_K]^\top \in \mathbb{R}^{K \times d}$ with K candidate functional roles. By normalizing the fused features among residue base representation, fundamental and residue-specific unit constraint, we obtain the role representation $x_i^{\text{role}} = \text{Norm}(f_{\text{self}}([h_i^{\text{base}} \| R \| x_i^{\text{AU}}]))$. Combining the relevance with the prototype matrix P, we compute compatibility logits $\ell_i^{\text{self}} = x_i^{\text{role}} P^\top + f_R(R) + f_U(x_i^{\text{AU}})$. Standardizing logits via complex variance s^{self} (details in App. A.1.3) and scaling by reliability $q_{\text{complex}}(t)$, we obtain the intrinsic role distribution $\pi_i^{\text{role}} = \text{Softmax}\left(q_{\text{complex}}(t) \widehat{\ell}_i^{\text{self}}\right), \pi_i^{\text{role}} \in \mathbb{R}^K$.

(2) Role refinement based on pairwise cooperation. A residue's function relies on coordinated interactions with its neighbors to sustain interface stability. Therefore, we refine the residue role distribution by interacting with others. We first calculate cooperative weights over residues. The cooperative influence of residue j on residue i depends on their physical distance, relative geometry, and synergistic semantics. We compute the mean distance $\bar{d}_t$ across all residues and normalize individual distances as $d_{ij}^{\text{norm}} = \|p_{i,t} - p_{j,t}\|_2 / \max(\bar{d}_t, \varepsilon)$. As shown in Eq. 4, compatibility score e_{ij}^{coop} balances physical proximity, geometry features U_{ij}^{relative}, and synergistic constraint S, which is normalized across valid neighbors v to produce weight w_{ij}. Then, based on residue base representation and functional constraints, we construct a unified cooperation feature $z_i^{\text{coop}} = f_{\text{coop}}([R \| x_i^{\text{AU}} \| S])$ for residue i. Considering the base representations and compatibility score, we compute a symmetric cooperation matrix $\Phi_{ij} = \frac{1}{2}[f_\Phi(h_i^{\text{base}} \| h_j^{\text{base}} \| z_i^{\text{coop}} \| z_j^{\text{coop}}) + f_\Phi(h_j^{\text{base}} \| h_i^{\text{base}} \| z_j^{\text{coop}} \| z_i^{\text{coop}})^\top]$. By projecting centered neighbor preferences $\pi_j^{\text{role}} - \frac{1}{K}\mathbf{1}_K$ via this cooperation matrix, we calculate role correction $\Delta\ell_i = \sum_{j=1}^L w_{ij} \Phi_{ij} (\pi_j^{\text{role}} - \frac{1}{K}\mathbf{1}_K)$, and update logits via $\widehat{\ell}_i^{\text{self}} \leftarrow \widehat{\ell}_i^{\text{self}} + \Delta\ell_i$. Finally, we recompute the refined role distribution $\pi_i^{\text{role}} = \text{Softmax}\left(q_{\text{complex}}(t) \widehat{\ell}_i^{\text{self}}\right)$.

$$e_{ij}^{\text{coop}} = -d_{ij}^{\text{norm}} + f_{\text{pair}}\left(U_{ij}^{\text{relative}}\right) + q_{\text{complex}}(t) f_S\left(S \| h_i^{\text{base}}\right), \quad w_{ij} = \frac{M_{ij}^{\text{pair}} \exp\left(e_{ij}^{\text{coop}}\right)}{\sum_{v=1}^L M_{iv}^{\text{pair}} \exp\left(e_{iv}^{\text{coop}}\right)} \tag{4}$$

3.5 Modality-Adaptive Coordinative Optimization

Heterogeneity in residue roles and modality semantics requires differentiated information interaction at residue and modality levels. To model these distinctions, we propose modality-adaptive coordination optimization unifying role-directed message passing with modality-specific feature updates.

(1) Role-directed message passing. To capture directional interaction across residues, we combine SE(3)-invariant local geometric relations $\delta_{ij} = [\|p_{j,t} - p_{i,t}\|_2 \| R_{i,t}^\top (p_{j,t} - p_{i,t}) \| R_{i,t}^\top R_{j,t}]$ between residues i and j by fusing Euclidean distance, local relative translation, and orientation. For each residue i, we project its base representation into modality states $h_i^m = f^m(h_i^{\text{base}}), m \in \mathcal{M} = \{\text{seq}, \text{pos}, \text{rot}\}$ and average them into a unified state $\bar{h}_i = \frac{1}{3} \sum_{m \in \mathcal{M}} h_i^m$. We map target state $\bar{h}_i$,

source state $\bar{h}_j$, relational prior U_{ij}^{relative}, and geometry δ_{ij} to construct preliminary message $b_{ij} = f_{\text{rel}}(W_{\text{tgt}}\bar{h}_i + W_{\text{src}}\bar{h}_j + W_p U_{ij}^{relative} + W_g\delta_{ij})$. As shown in Eq. 5, we modulate message b_{ij} based on a demand–supply mechanism that considers target residue role demand $f_{\text{demand}}(\pi_i^{\text{role}} - \frac{1}{K}\mathbf{1}_K)$ and source residue supply $f_{\text{supply}}(\pi_j^{\text{role}} - \frac{1}{K}\mathbf{1}_K)$, yielding the role-directed message r_{ij}^{dir}.

$$r_{ij}^{\text{dir}} = b_{ij} + f_{\text{mod}}\left(b_{ij} \odot \tanh\left(f_{\text{demand}}\left(\pi_i^{\text{role}} - \tfrac{1}{K}\mathbf{1}_K\right)\right) \odot \tanh\left(f_{\text{supply}}\left(\pi_j^{\text{role}} - \tfrac{1}{K}\mathbf{1}_K\right)\right)\right) \quad (5)$$

(2) Modality-specific feature update. For modality $m \in \mathcal{M}$ of residue i, we first compute the modality match weight ω_{ij}^m by regarding the modality feature h_i^m as query and message r_{ij}^{dir} across neighbors $j \in \mathcal{V}$ as key and value (detailed in App A.1.4). We weighted aggregate messages r_{ij}^{dir} into modality-specific context $c_i^m = W_O^m \sum_{j=1}^{L} \omega_{ij}^m W_V^m r_{ij}^{\text{dir}}$. Then, based on functional constraints, we calculate the modality-specific constraint $k_i^m = W_H^m f_{\text{hier}}([R\|x_i^{\text{AU}}\|S])$. By fusing the modality feature, modality-specific context, modality-specific constraint, and related residue role preferences, we obtain consolidated modality information $e_i^m = f_m([h_i^m\|c_i^m\|k_i^m\|\pi_i^{\text{role}} - \frac{1}{K}\mathbf{1}_K])$. As shown in Eq. 6, using an asymmetric cross-modal attention mechanism, the target modality m adaptively gates incoming information from source modality $n \in \mathcal{M} \setminus \{m\}$ via directional gating coefficient $\alpha_i^{n\to m} \in (-1, 1)$. We aggregate gated source features across modalities $n \neq m$, normalize and scale by complex reliability $q_{\text{complex}}(t)$ to produce update $\Delta h_i^m = q_{\text{complex}}(t)\,\text{LN}(e_i^m + \sum_n \alpha_i^{n\to m} W_V[e_i^n\|\pi_i^{\text{role}} - \frac{1}{K}\mathbf{1}_K])$. Finally, we obtain optimized modality representations via residual projection $h_i^m = h_i^{\text{base}} + W_{\text{out}}^m \Delta h_i^m$.

$$\alpha_i^{n\to m} = \tanh\left(\frac{(W_Q^m e_i^m)^\top W_K\left[e_i^n\|\pi_i^{\text{role}} - \frac{1}{K}\mathbf{1}_K\right]}{\sqrt{d}}\right) \quad (6)$$

3.6 Training Objective

We formulate antibody co-design as multimodal conditional flow matching over time $t \sim \mathcal{U}(0, 1)$. For positional representations h_i^{pos}, we predict velocity fields $\widehat{u}_{i,t}$ for linear Gaussian flows, where target velocity $u_{i,t}^\star = p_{i,0} - p_i^{\text{noise}}$ with $p_i^{\text{noise}} \sim \mathcal{N}(0, I)$, supervised by matching loss $\ell_{\text{pos}} = \|\widehat{u}_{i,t} - u_{i,t}^\star\|_2^2$. For rotational representations h_i^{rot}, we predict native frames $\widehat{R}_{i,0}$ and define SO(3) geodesic flows, penalized by alignment loss $\ell_{\text{rot}}(\widehat{R}, R) = \sum_{a=1}^{3}\left(1 - \frac{(\widehat{R}^{(:,a)})^\top R^{(:,a)}}{\|\widehat{R}^{(:,a)}\|\|R^{(:,a)}\|}\right)$. For sequential representations h_i^{seq}, we predict residue distributions $\widehat{s}_{i,0}$ for absorbing categorical flows, optimized via masked cross-entropy $\ell_{\text{seq}} = -\mathbb{I}_{[s_{i,t}=\texttt{MASK}]} \log \widehat{s}_{i,0}(s_{i,0})$. The unified multimodal flow-matching objective is $\mathcal{L}_{\text{gen}} = \mathbb{E}_{t,\mathcal{C}_0,\mathcal{C}_t}[\sum_{i=1}^{L} M_i^{\text{gen}}(\ell_{\text{pos}} + \ell_{\text{rot}}(\widehat{R}_{i,0}, R_{i,0}) + \ell_{\text{seq}})]$.

To regularize multi-scale constraints and prevent representation collapse, we design an auxiliary objective $\mathcal{L}_{\text{RUS}} = \mathcal{L}_{\text{fca}} + \mathcal{L}_{\text{field}} + \mathcal{L}_{\text{self}} + \mathcal{L}_{\text{coop}} + \mathcal{L}_{\text{reg}}$, where $\mathcal{L}_{\text{fca}}$ aligns unit tokens to the fundamental constraint R with reliability weighting; $\mathcal{L}_{\text{field}}$ imposes scale-wise physicochemical and SE(3)-invariant geometric supervision via cross-entropy and SmoothL1 losses. Role assignments are supervised by $\mathcal{L}_{\text{self}}$ (intrinsic properties and regional/chain geometry) and $\mathcal{L}_{\text{coop}}$ (inter-partner matching and complex geometry). $\mathcal{L}_{\text{reg}}$ prevents information redundancy and feature collapse (details in App. A.1.5). The overall training loss is $\mathcal{L} = \mathcal{L}_{\text{gen}} + 0.05\mathcal{L}_{\text{RUS}}$.

4 Experiments

Datasets. Following prior antibody co-design works (Kong et al., 2023; Luo et al., 2022; Zhu et al., 2024; Wang et al., 2026), we train MoDA on SAbDab dataset (Dunbar et al., 2014) and evaluate on the RAbD benchmark (Adolf-Bryfogle et al., 2018). We retain complexes with resolution better than 4.0 Å. To prevent data leakage, we cluster CDR-H3 sequences via MMseqs2 (Steinegger & S"oding, 2017) (50% sequence identity, 80% coverage) and discard clusters overlapping RAbD.

Baselines. We compare MoDA with representative antibody co-design methods for all-CDR design. DiffAb employs diffusion for antigen-conditioned co-generation of antibody sequences and structures (Luo et al., 2022). dyMEAN performs end-to-end full-atom design via equivariant geometric reasoning (Kong et al., 2023). AbX regularizes score-based diffusion with evolutionary, physical,

Table 1: Performance on the RAbD all-CDR co-design.

Method	RMSD ↓	TM ↑	G-TM ↑	lDDT ↑	G-lDDT ↑	AAR ↑	ddG-SR ↑
DiffAb (Luo et al., 2022)	1.9971	0.6231	0.8224	0.6121	0.5318	48.9000	45.0%
dyMEAN (Kong et al., 2023)	2.4715	0.9455	0.7598	0.9063	0.7433	65.1798	33.3%
AbX (Zhu et al., 2024)	1.9061	0.9713	0.7912	0.9502	0.7702	65.5887	50.0%
AAMFM (Shi et al., 2026)	2.2386	0.8989	0.8082	0.8672	0.7929	**70.5000**	38.3%
AbFlow (Wang et al., 2026)	1.8139	0.9545	0.7822	0.9078	0.7485	63.8898	40.0%
MoDA	**1.5780**	**0.9745**	**0.8485**	**0.9677**	**0.8341**	62.2509	**58.3%**

Table 2: Ablation study of MoDA on the RAbD benchmark.

Variant	RMSD ↓	TM ↑	G-TM ↑	lDDT ↑	G-lDDT ↑	AAR ↑
MoDA	**1.5780**	**0.9745**	0.8485	**0.9677**	**0.8341**	**62.2509**
w/o MSC	1.6940	0.9720	0.8203	0.9561	0.7877	61.5920
- RUS Loss	1.6031	0.9742	**0.8500**	0.9616	0.8192	61.9483
w/o CAR	1.6649	0.9724	0.8345	0.9580	0.8033	60.8254
- Role Inference	1.6811	0.9726	0.8317	0.9580	0.7993	59.9383
- Role Refinement	1.6356	0.9734	0.8401	0.9596	0.8102	61.9697
w/o MACP	1.6461	0.9729	0.8280	0.9573	0.7973	61.6452

and geometric priors (Zhu et al., 2024). AAMFM builds on ESM3 using a lightweight antigen-aware adapter and Cal-DPO for antigen-specific functional antibody co-design (Shi et al., 2026). AbFlow realizes paratope-centered design with interaction-enhanced flow matching (Wang et al., 2026).

Evaluation metrics. We evaluate generated antibodies on sequence, structural, and binding quality. AAR measures CDR sequence recovery; RMSD (Kabsch, 1976) reports CDR structural accuracy under fixed frameworks; TM-score (Zhang & Skolnick, 2004) and G-TM characterize global and CDR-only fold similarity; lDDT (Mariani et al., 2013) and G-lDDT quantify local distance preservation at the full complex and CDR levels; ddG-SR denotes the fraction of targets where at least one generated candidate yields $\Delta\Delta G < 0$ (details in App. A.2).

4.1 Antibody Paratope Co-design

To evaluate MoDA on full paratope co-design, we benchmark it against representative baselines on the RAbD dataset. Following prior design protocol (Luo et al., 2022; Kong et al., 2023; Zhu et al., 2024), we sample 100 candidate antibodies per antibody–antigen complex. As shown in Tab. 1, MoDA achieves the best performance across RMSD, TM-score, G-TM, lDDT, G-lDDT, and $\Delta\Delta G$-SR alongside competitive sequence recovery. Specifically, MoDA decreases the all-CDR RMSD from 1.8139 (the previous best baseline) to 1.5780, validating the superior geometric accuracy of jointly generated CDRs under our modality-adaptive framework. MoDA also improves global structural metrics, increasing the TM-score to 0.9745 and the lDDT to 0.9677. To better reflect the CDR design quality, we compute the region-wise metrics. On G-TM and G-lDDT, MoDA advances the strongest baseline from 0.8224 to 0.8485 and from 0.7929 to 0.8341, respectively. These results confirm that multi-scale constraints, inferred functional roles, and modality-adaptive optimization collectively enhance both global compatibility and local conformational fidelity. Beyond structural quality, MoDA delivers competitive AAR and the highest $\Delta\Delta G$-SR at **58.3%** ($+8.33\%$ over the top baseline), demonstrating that our method samples diverse sequences with favorable binding potential rather than non-functional random variations. App. A.3.1 shows per-CDR metrics.

4.2 Ablation Study

To validate the necessity of MoDA's core modules, we conduct both module-level and operation-level ablations on multi-scale constraint modeling, cooperation-aware role modeling, and modality-adaptive coordinative optimization. As shown in Tab. 2, removing multi-scale constraint modeling (**w/o MSC**) increases CDR RMSD from 1.5780 to 1.6940 and decreases G-lDDT from 0.8341 to 0.7877, demonstrating that multi-scale functional constraints provide important structural regularization for coordinated multi-CDR generation. Removing RUS supervision (**-RUS Loss**) increases

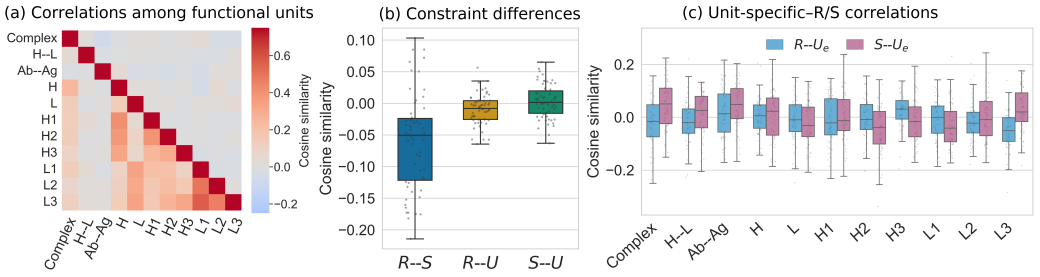


Figure 2: Disentanglement analysis of multi-scale constraints. (a) Pairwise cosine similarities among functional units before (lower triangle) and after (upper triangle) decomposition. (b–c) Independence among fundamental (R), synergistic (S), and unit-specific (U) constraints.

RMSD to 1.6031 and decreases G-lDDT and AAR to 0.8192 and 61.9483, respectively. Although G-TM marginally increases, the decline in the other metrics suggests that RUS supervision provides an overall benefit to generation quality. Disabling cooperation-aware role modeling (**w/o CAR**) consistently impairs structural accuracy and sequence recovery. Further, removing either intrinsic role inference or cooperation-aware role refinement increases RMSD to 1.6811 and 1.6356 and decreases G-lDDT to 0.7993 and 0.8102, respectively, demonstrating their complementarity. When eliminating modality-adaptive coordinative optimization (**w/o MACP**), increasing RMSD to 1.6461 and lowering G-TM and G-lDDT to 0.8280 and 0.7973 indicate that differentiated residue interaction and modality coordination facilitate sequence–structure co-design.

4.3 Disentanglement Analysis of Multi-Scale Constraints

To investigate whether multi-scale constraint modeling effectively resolves feature overlap and extracts independent physical constraints, we analyze both inter-unit redundancy and constraint orthogonality: (1) **Semantic redundancy elimination of functional units.** We compute pairwise cosine similarities between functional unit representations before and after decomposition. As shown in Fig. 2(a), raw unit representations exhibit substantial semantic overlap before decomposition (lower triangle): the average cosine similarity of same-scale units reaches 0.4363, with an overall global mean of 0.2404. After decomposition (upper triangle), the average cosine similarity across all unit pairs drops sharply to near zero (global: -0.0024; same-scale pairs: -0.0132). This demonstrates that we effectively extract differentiated features of functional units. (2) **Independence assessment of disentangled constraints.** We evaluate the cosine similarities among fundamental (R), synergistic (S), and unit-specific (U) constraints. As shown in Fig. 2(b–c), the pairwise cosine similarities both across distinct constraint categories (R, S, and U) and between diverse unit-specific (U_e) and global constraints (R, S) remain consistently around zero. These near-orthogonal relations confirm that MoDA successfully partitions universal, cooperative, and unit-inherent constraints into independent feature subspaces without representational collapse.

4.4 Functional Specialization in Cooperation-Aware Role Modeling

To explore whether cooperation-aware role modeling provides meaningful residue functional specialization, we group complex residues by structural identity into CDR interface (CDR-I), CDR non-interface (CDR-NI), and framework anchor (FR anchor) residues. Then we conduct analyses from two perspectives: (1) **Non-redundancy of role refinement**. We compute the cosine similarity between inferred intrinsic roles and refined roles throughout the generation process. As shown in Fig. 3(a), the cosine similarity of role logits exhibits a continuous decrease across all timesteps, indicating that inter-residue cooperation imposes context-aware, non-redundant information on residue role preferences. (2) **Biological relevance of role distributions.** We evaluate intra-group role consistency using cosine similarity and inter-group differences using a covariance-standardized distance on role distributions. As shown in Fig. 3(b–c), residues within each group attain high role consistency (0.9277 for CDR interface residues and 0.9665 for FR anchors), whereas all cross-group comparisons exhibit differences in role distributions, especially between CDR non-interface and FR-anchor residues, demonstrating that our method identifies biologically relevant residue roles.

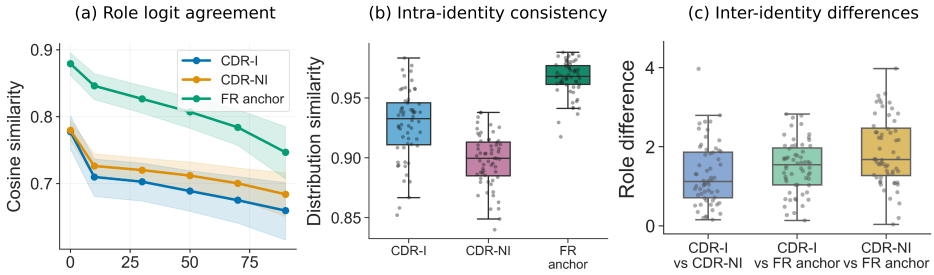


Figure 3: Role specialization analysis. (a) Role similarity between inference and cooperative refinement over generation; (b) role consistency within identity; (c) role divergence between identities.

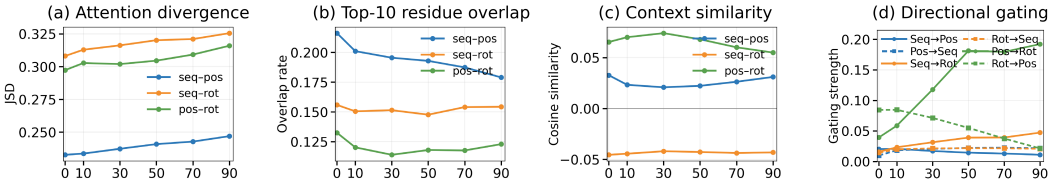


Figure 4: Modality heterogeneity analysis. (a) Cross-modality attention divergence; (b) top-10 residue overlap; (c) cross-modality context similarity; (d) inter-modality asymmetric coupling.

4.5 Heterogeneity Analysis in Modality-Adaptive Optimization

To examine whether MoDA effectively captures modality-heterogeneous requirements, we analyze both local context selections and the directional dependencies across sequence, position, and orientation: **(1) Distinct contextual preferences across modalities.** As shown in Fig. 4(a–c), we first compute the Jensen–Shannon divergence (Lin, 1991) (JSD) of residue-level attention distributions for each modality pair. Pairwise JSD stays high (0.2328–0.3256) with divergent evolution, demonstrating modality-specific preferences for neighboring residues. We further calculate the overlap ratio of each modality's top-10 attended residues. The low overlap (0.1139–0.2163) confirms distinct neighborhood selection across modalities. In addition, we measure the cosine similarity of the learned modality-specific contexts c_i^m. The near-orthogonal relations ($[-0.0455, 0.0741]$, peak < 0.1) reveal cross-modal contextual decoupling. These complementary results verify that sequence, position, and orientation possess heterogeneous contextual requirements during antibody modeling. **(2) Asymmetry in cross-modal interaction.** We track the directional gating coefficients among sequence, position, and orientation during generation. As shown in Fig. 4(d), position exhibits the strongest growing influence on orientation, with average interaction strength rising from 0.0393 to 0.1922 as modality-specific semantics clarify. Conversely, the reverse orientation-to-position interaction decays from 0.0844 to 0.0213, showing strong asymmetry between the two geometric modalities. The sequence-to-orientation interaction also rises steadily (0.0158 to 0.0472), whereas its reverse stays slight. The sequence–position bidirectional interaction remains the weakest overall. These varying interaction strengths show that the MoDA adaptively modulates cross-modal dependencies and confirm asymmetric relationships across sequence, position, and orientation.

5 Conclusion

In this study, we propose **MoDA**, a modality-adaptive flow matching framework that integrates multi-scale constraints with residue role modeling. We design multi-scale constraint modeling to capture disentangled constraints from local-to-global constraints via dedicated attention. Based on constraints, we introduce cooperation-aware role modeling to infer the residue role distribution, which guides modality-adaptive coordinative optimization for modality-specific feature updates. Experiments on the RAbD benchmark show that MoDA enhances structural fidelity and antibody binding potential, demonstrating the effectiveness of role-guided modality-adaptive interactions.

AI USE STATEMENT

In this work, we used generative AI tools for code writing and sentence polishing. We did not use generative AI tools to identify research questions, create the core idea, or design the model architecture. We reviewed all AI-assisted work, and LLM-generated code was verified and tested for correctness. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work uses publicly available antibody–antigen structures and does not involve human participants or the collection of personally identifiable information. The proposed method is intended for computational antibody design; any generated candidates require appropriate experimental validation and biosafety assessment before practical use.

REPRODUCIBILITY STATEMENT

Our code is released at `https://anonymous.4open.science/r/MoDA-D2CE`.

REFERENCES

Jared Adolf-Bryfogle, Oleksandr Kalyuzhniy, Matthew Kubitz, Brian D. Weitzner, Xiaozhen Hu, Yasushi Adachi, William R. Schief, and Roland L. Dunbrack Jr. RosettaAntibodyDesign (RAbD): A general framework for computational antibody design. *PLOS Computational Biology*, 14(4): e1006112, 2018.

Rebecca F. Alford, Andrew Leaver-Fay, Jeliazko R. Jeliazkov, et al. The rosetta all-atom energy function for macromolecular modeling and design. *Journal of Chemical Theory and Computation*, 13(6):3031–3048, 2017. doi: 10.1021/acs.jctc.7b00125.

S. Chaudhury, S. Lyskov, and J. J. Gray. Pyrosetta: A script-based interface for implementing molecular modeling algorithms using rosetta. *Bioinformatics*, 26(5):689–691, 2010. doi: 10.1093/bioinformatics/btq007.

James Dunbar, Konrad Krawczyk, Jinwoo Leem, Terry Baker, Angelika Fuchs, Guy Georges, Jiye Shi, and Charlotte M. Deane. Sabdab: the structural antibody database. *Nucleic Acids Research*, 42(D1):D1140–D1146, 2014. doi: 10.1093/nar/gkt1043.

Tomas Geffner, Kieran Didi, Zhonglin Cao, Danny Reidenbach, Zuobai Zhang, Christian Dallago, Emine Kucukbenli, Karsten Kreis, and Arash Vahdat. La-proteina: Atomistic protein generation via partially latent flow matching. In *International Conference on Learning Representations*, volume 2026, pp. 67350–67395, 2026.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

Weihong Huang, Feng Yang, Qiang Zhang, and Juan Liu. Dynamic geometric equivariant network for full-atom antibody design. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2026.

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

Xiangzhe Kong, Wenbing Huang, and Yang Liu. Conditional antibody design as 3d equivariant graph translation. *arXiv preprint arXiv:2208.06073*, 2022.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. End-to-end full-atom antibody design. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 17409–17429. PMLR, 23–29 Jul 2023.

Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991. doi: 10.1109/18.61115.

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.

Shitong Luo, Yufeng Su, Xingang Peng, Sheng Wang, Jian Peng, and Jianzhu Ma. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. In *Advances in Neural Information Processing Systems*, volume 35, pp. 9754–9767, 2022.

Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. lDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 2013.

Karolis Martinkus, Jan Ludwiczak, Kyunghyun Cho, Wei-Chun Chen, Rohit Midha, Deniz Akpinaroglu, Sisi Tian, Maxim V. Prigorovskiy, Jian Peng, Nathan C. Rockwell, et al. AbDiffuser: Full-atom generation of in-vitro functioning antibodies. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Wei Qu, Jiawei Guan, Rui Ma, Ke Zhai, Weikun Wu, et al. P(all-atom) is unlocking new path for protein design. *bioRxiv*, 2024. doi: 10.1101/2024.08.16.608235.

Idit Sela-Culang, Vered Kunik, and Yair Ofran. The structural basis of antibody-antigen recognition. *Frontiers in Immunology*, 4:302, 2013. doi: 10.3389/fimmu.2013.00302.

Xiaoliang Shi, Zichen Wang, Runze Ma, Zhongyue Zhang, and Shuangjia Zheng. Antigen-specific antibody multi-modal foundation model for functional antibody design. *arXiv preprint arXiv:2607.20057*, 2026.

Martin Steinegger and Johannes S"oding. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35:1026–1028, 2017.

Cheng Tan, Yijie Zhang, Zhangyang Gao, Yufei Huang, Haitao Lin, Lirong Wu, Fandi Wu, Mathieu Blanchette, and Stan Z. Li. dyAb: Flow matching for flexible antibody design with AlphaFold-driven pre-binding antigen. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

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, 2017.

Wenda Wang, Yang Zhang, Zhewei Wei, and Wenbing Huang. Abflow: End-to-end paratope-centric antibody design by interaction enhanced flow matching. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 1471–1482, 2026.

Xinyou Wang, Zaixiang Zheng, Fei Ye, Dongyu Xue, Shujian Huang, and Quanquan Gu. DPLM-2: A multimodal diffusion protein language model. In *International Conference on Learning Representations*, 2025.

Lirong Wu, Haitao Lin, Yufei Huang, Zhangyang Gao, Cheng Tan, Yunfan Liu, Tailin Wu, and Stan Z. Li. Relation-aware equivariant graph networks for epitope-unknown antibody design and specificity optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

Yang Zhang and Jeffrey Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4):702–710, 2004.

Yi Zhou, Haohao Qu, Yunqing Liu, Shanru Lin, Le Song, and Wenqi Fan. HD-Prot: A protein language model for joint sequence-structure modeling with continuous structure tokens. *arXiv preprint arXiv:2512.15133*, 2025.

Tian Zhu, Milong Ren, and Haicang Zhang. Antibody design using a score-based diffusion model guided by evolutionary, physical and geometric constraints. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 62531–62548, 2024.

A Appendix

A.1 Method Details

A.1.1 Global Pairwise and Base Representation Construction

To comprehensively encode spatial, sequence, and geometric relations across all residue pairs (i, j), we compute four complementary pairwise feature representations:

- **Relative residue type embedding (u_{ij}^{aa}):** Residue type $s_{i,t}$ and $s_{j,t}$ at generation step t are embedded and concatenated via a linear pair projection layer: $u_{ij}^{\text{aa}} = \text{Emb}_{\text{pair}}([s_{i,t} \| s_{j,t}])$.
- **Relative sequence position embedding (u_{ij}^{rel}):** Based on intra-chain sequence indices r_i and r_j, we project the positional relation: $u_{ij}^{\text{rel}} = \text{Emb}_{\text{rel}}([r_i \| r_j])$.
- **Continuous radial distance feature (u_{ij}^{dist}):** Euclidean distances across all pairwise backbone atom pairs $a, b \in \{\text{N}, \text{C}_\alpha, \text{C}, \text{O}\}$ are expanded via Gaussian Radial Basis Functions (RBF) with centers μ_k and variance σ: $u_{ij}^{\text{dist}} = f_{\text{dist}}(\text{RBF}(\|x_i^a - x_j^b\|_2))$.
- **Multi-frequency dihedral angle feature (u_{ij}^{ang}):** Inter-residue relative orientations are captured by the pseudo-dihedral angles $\phi_{ij}^{\text{inter}} = \text{Dihedral}(\text{C}_i, \text{N}_j, \text{C}_{\alpha,j}, \text{C}_j)$ and $\psi_{ij}^{\text{inter}} = \text{Dihedral}(\text{N}_i, \text{C}_{\alpha,i}, \text{C}_i, \text{N}_j)$. Using the multi-frequency set $\Omega = \{1, 2, 3, \frac{1}{2}, \frac{1}{3}\}$, we encode angular periodicity as:

$$u_{ij}^{\text{ang}} = [(\sin(\omega_k \phi_{ij}^{\text{inter}}), \cos(\omega_k \phi_{ij}^{\text{inter}}))_{\omega_k \in \Omega} \parallel (\sin(\omega_k \psi_{ij}^{\text{inter}}), \cos(\omega_k \psi_{ij}^{\text{inter}}))_{\omega_k \in \Omega}]. \quad (7)$$

These pairwise features are concatenated and mapped via an MLP f_{relative} into $U_{ij}^{\text{relative}} = f_{\text{relative}}([u_{ij}^{\text{aa}} \| u_{ij}^{\text{rel}} \| u_{ij}^{\text{dist}} \| u_{ij}^{\text{ang}}])$. We compute the temporal schedule embedding c_t under the continuous schedule β_t, i.e., $c_t = f_t([\beta_t \parallel \sin(\beta_t) \parallel \cos(\beta_t)])$.

A.1.2 Multi-Scale Functional Unit Modeling Details

Region-scale and Chain-scale Low-Rank Multi-Head Attention. For functional unit $e \in \mathcal{E}_{\text{region}}$ (or $\mathcal{E}_{\text{chain}}$) over interaction domain $\mathcal{O}_e(i) = \{j \in \mathcal{V} \mid j \in e\}$, we normalize attention weights across head $m \in \{1, \ldots, M\}$ and aggregate contexts as:

$$\alpha_{ij}^{e,m} = \frac{\exp\left(\frac{(W_Q^m \,\text{LN}(h_i^{\text{base}}))^\top (W_K^m \,\text{LN}(h_j^{\text{base}}))}{\sqrt{d_h}}\right)}{\sum_{k \in \mathcal{O}_e(i)} \exp\left(\frac{(W_Q^m \,\text{LN}(h_i^{\text{base}}))^\top (W_K^m \,\text{LN}(h_k^{\text{base}}))}{\sqrt{d_h}}\right)}, \quad \widetilde{h}_{e,i}^m = \sum_{j \in \mathcal{O}_e(i)} \alpha_{ij}^{e,m} W_V^m \,\text{LN}(h_j^{\text{base}}). \quad (8)$$

The unit-general representation is constructed as $H_e^{\text{region}} = [W_O[\widetilde{h}_{e,i}^1 \| \ldots \| \widetilde{h}_{e,i}^M]]_{i \in e}$ (similarly H_e^{chain} for chain scale). Concurrently, low-rank adapters parameterized by $A_e \in \mathbb{R}^{r \times d}$ and $B_e \in \mathbb{R}^{d \times r}$ ($r \ll d$) compute unit-specific modulations:

$$\Delta H_e = \frac{1}{\sqrt{r}} (\text{LN}(H_e^{\text{base}}) A_e^\top) B_e^\top. \quad (9)$$

The aggregated state $U_e = H_e^{\text{base}} + H_e + \Delta H_e$ is subsequently processed by a feed-forward network to yield $Y_e = U_e + \text{FFN}(\text{LN}(U_e))$.

Pair-scale Inter-Chain and Interfacial Interaction. For antibody heavy–light chain pairing ($\mathcal{E}_{\text{pair}} = \{\text{HL}\}$), each residue queries the opposite chain domain $\mathcal{O}_{\text{HL}}^{\text{pair}}(i)$ (where $\mathcal{O}_{\text{HL}}^{\text{pair}}(i) = \mathcal{V}_L$ if $i \in \mathcal{V}_H$, and vice versa) via cross-attention with normalized weights α_{ij}^{HL} to aggregate context $c_i^{\text{HL}} = \sum_j \alpha_{ij}^{\text{HL}} W_V h_j^{\text{base}}$. The HL-pair encoder $f_{\text{HL}}^{\text{pair}}$ incorporates both interactive association and discrepancy:

$$y_{\text{HL},i}^{\text{pair}} = f_{\text{HL}}^{\text{pair}}([h_i^{\text{base}} \parallel c_i^{\text{HL}} \parallel h_i^{\text{base}} \odot c_i^{\text{HL}} \parallel |h_i^{\text{base}} - c_i^{\text{HL}}|]), \quad i \in \mathcal{V}_{\text{Ab}}. \quad (10)$$

For antibody–antigen interfacial binding ($\mathcal{E}_{\text{pair}} = \{\text{AbAg}\}$), directed relations $r \in \mathcal{R}_{\text{AbAg}} = \{\text{Ag} \to \text{Ab}, H \to \text{Ag}, L \to \text{Ag}\}$ define the interaction domains:

$$\mathcal{O}_r^{\text{pair}}(i) = \begin{cases} \{j \in \mathcal{V} \mid j \in \mathcal{V}_{\text{Ag}}\}, & r = \text{Ag} \to \text{Ab},\ i \in \mathcal{V}_{\text{Ab}}, \\ \{j \in \mathcal{V} \mid j \in \mathcal{V}_H\}, & r = H \to \text{Ag},\ i \in \mathcal{V}_{\text{Ag}}, \\ \{j \in \mathcal{V} \mid j \in \mathcal{V}_L\}, & r = L \to \text{Ag},\ i \in \mathcal{V}_{\text{Ag}}. \end{cases} \tag{11}$$

The spatial scoring network f_{sp} computes continuous geometric coupling scores:

$$s_{ij} = f_{\text{sp}}([h_i^{\text{base}} \parallel h_j^{\text{base}} \parallel (p_{j,t} - p_{i,t}) \parallel \|p_{j,t} - p_{i,t}\|_2]). \tag{12}$$

Directed spatial attention weights $\alpha_{ij}^r = \exp(s_{ij}) / \sum_{k \in \mathcal{O}_r^{\text{pair}}(i)} \exp(s_{ik})$ aggregate source contexts $c_i^r = \sum_{j \in \mathcal{O}_r^{\text{pair}}(i)} \alpha_{ij}^r h_j^{\text{base}}$, which are then transformed into the pair-scale representation $Y_{\text{AbAg}}^{\text{pair}} = [y_{\text{AbAg},i}^{\text{pair}}]_{i \in \mathcal{V}}$ via Eq. 2.

A.1.3 Role Logits Centering and Scale Calibration.

Raw role compatibility logits $\ell_i^{\text{self}} = x_i^{\text{role}} P^\top + f_R(R) + f_U(x_i^{\text{AU}}) \in \mathbb{R}^K$ are centered by subtracting the mean activation: $\bar{\ell}_i^{\text{self}} = \ell_i^{\text{self}} - \frac{1}{K}(\mathbf{1}_K^\top \ell_i^{\text{self}})\mathbf{1}_K$. The complex-level root-mean-square scale factor s^{self} is computed across all valid residues:

$$s^{\text{self}} = \sqrt{\max\left(\frac{\sum_{i=1}^L M_i^{\text{valid}} \|\bar{\ell}_i^{\text{self}}\|_2^2}{K \max(\sum_{i=1}^L M_i^{\text{valid}}, 1)}, \varepsilon\right)}, \qquad \widehat{\ell}_i^{\text{self}} = \frac{\bar{\ell}_i^{\text{self}}}{s^{\text{self}}}. \tag{13}$$

A.1.4 Modality Context and Heterogeneous Coordination Details

In the modality-adaptive coordination update, modality-specific context c_i^m aggregates incoming directed messages r_{ij}^{dir} via query projection of single-modality state h_i^m:

$$\omega_{ij}^m = \text{Softmax}_{j \in \mathcal{V}}\left(\frac{(W_Q^m h_i^m)^\top (W_K^m r_{ij}^{\text{dir}})}{\sqrt{d}}\right), \qquad c_i^m = W_O^m \sum_{j=1}^L \omega_{ij}^m W_V^m r_{ij}^{\text{dir}}. \tag{14}$$

A.1.5 Detailed Auxiliary Semantic Loss Formulations

To regularize multi-scale constraints and prevent representation collapse, we propose an auxiliary objective $\mathcal{L}_{\text{RUS}} = \mathcal{L}_{\text{fca}} + \mathcal{L}_{\text{field}} + \mathcal{L}_{\text{self}} + \mathcal{L}_{\text{coop}} + \mathcal{L}_{\text{reg}}$.

Fundamental Constraint Alignment Loss ($\mathcal{L}_{\text{fca}}$). To enforce semantic consistency between localized functional unit tokens and the universal representation, the alignment loss aligns the contextualized unit tokens $\widehat{z}_e^s$ across all levels $s \in \mathcal{S}$ with the fundamental constraint R. A projection network f_{fca} maps unit representations into the fundamental latent space. To prevent unreliable, noisy generation steps from misleading optimization, we dynamically scale each unit's contribution by its temporal reliability weight $q_e(t)$:

$$\mathcal{L}_{\text{fca}} = \frac{\sum_{s \in \mathcal{S}} \sum_{e \in \mathcal{E}_s} q_e(t) \|f_{\text{fca}}(\widehat{z}_e^s) - R\|_2^2}{\max(\sum_{s \in \mathcal{S}} \sum_{e \in \mathcal{E}_s} q_e(t), 1)}. \tag{15}$$

To guide the model to learn biophysical information and promote favorable inter-residue and inter-facial interactions, we construct physicochemical categories and scale-aware geometric features as supervision targets, respectively:

(1) **Physicochemical category**. We group the amino acid types into six categories according to their side-chain function, charge states, and conformational characteristics $y_i = \rho(s_{i,0}) \in \{0, 1, \ldots, 5\}$. The categorical assignment $\rho(\cdot)$ partitions the 20 standard amino acids according to their side-chain biophysical characteristics:

Index	Category	Amino Acids
0	Acidic	Asp (D), Glu (E)
1	Basic	His (H), Lys (K), Arg (R)
2	Polar neutral	Asn (N), Gln (Q), Ser (S), Thr (T)
3	Aliphatic	Ala (A), Ile (I), Leu (L), Met (M), Val (V)
4	Aromatic	Phe (F), Trp (W), Tyr (Y)
5	Special Conformation	Cys (C), Gly (G), Pro (P)

(2) $\mathrm{SE}(3)$**-invariant geometric targets**. We supervise relative spatial relations within different scales. We first define scale-specific reference coordinates $\bar{p}^s_{i,0}$ based on the representative residues of each scale s:

- *Region scale* ($s = \mathrm{region}$)*:* $\bar{p}^{\mathrm{region}}_{i,0} = \frac{1}{|\mathcal{N}^{\mathrm{local}}_i|}\sum_{j\in\mathcal{N}^{\mathrm{local}}_i} p_{j,0}$, which corresponds to the spatial centroid of residue i's local k-nearest sequence/spatial neighborhood within the loop.
- *Chain scale* ($s = \mathrm{chain}$)*:* $\bar{p}^{\mathrm{chain}}_{i,0} = p_{j_{\mathrm{chain}}(i),0}$, where $j_{\mathrm{chain}}(i)$ represents the closest partner residue along the same backbone chain ($j \neq i$).
- *Pair-HL scale* ($s = \mathrm{HL}$)*:* $\bar{p}^{\mathrm{HL}}_{i,0} = p_{j_{\mathrm{HL}}(i),0}$, where $j_{\mathrm{HL}}(i) = \arg\min_{j\in\mathcal{O}^{\mathrm{pair}}_{\mathrm{HL}}(i)} \|p_{i,0} - p_{j,0}\|_2$ denotes the closest native inter-chain residue in the opposing variable domain ($\mathcal{V}_L$ for $i \in \mathcal{V}_H$, and $\mathcal{V}_H$ for $i \in \mathcal{V}_L$).
- *Pair-AbAg scale* ($s = \mathrm{AbAg}$)*:* $\bar{p}^{\mathrm{AbAg}}_{i,0} = p_{j_{\mathrm{AbAg}}(i),0}$, where $j_{\mathrm{AbAg}}(i) = \arg\min_{j\in\mathcal{V}_{\mathrm{Ag}}} \|p_{i,0} - p_{j,0}\|_2$ denotes the closest native epitope antigen residue for paratope residue $i \in \mathcal{V}_{\mathrm{Ab}}$ (or vice versa).
- *Complex scale* ($s = \mathrm{complex}$)*:* $\bar{p}^{\mathrm{complex}}_{i,0} = \frac{1}{|\mathcal{V}_C|}\sum_{j\in\mathcal{V}_C} p_{j,0}$, which represents the overall geometric center of all generative CDR loops $\mathcal{V}_C$.

We then project native spatial displacement vectors into the residue's native local coordinate frame $(p_{i,0}, R_{i,0})$, which ensures that invariant geometric targets $\Delta^s_i = R^\top_{i,0}(p_{i,0} - \bar{p}^s_{i,0})$.

Multi-Scale Field Constraint Loss ($\mathcal{L}_{\mathrm{field}}$). To preserve scale-specific chemical semantics and recover local rigid-invariant geometric offsets, we construct the field supervision loss to regularize single-unit constraint representations $x^s_{e,i}$. We use a dedicated network f^s_{chem} to predict the physicochemical class distributions, while the geometric regression network f^s_{geo} predicts local relative coordinate offsets $\hat{\Delta}^s_i$:

$$\begin{aligned}\mathcal{L}_{\mathrm{field}} = \; & \mathrm{CE}(f^{\mathrm{region}}_{\mathrm{chem}}(x^{\mathrm{region}}_{\mathrm{CDR}(i),i}), y^{\mathrm{region}}_i) \\ & + \mathrm{CE}(f^{\mathrm{HL}}_{\mathrm{chem}}(x^{\mathrm{pair}}_{\mathrm{HL},i}), y^{\mathrm{HL}}_i) \\ & + \mathrm{CE}(f^{\mathrm{AbAg}}_{\mathrm{chem}}(x^{\mathrm{pair}}_{\mathrm{AbAg},i}), y^{\mathrm{AbAg}}_i) \\ & + \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} \mathrm{SmoothL1}(f^s_{\mathrm{geo}}(x^s_{e(s,i),i}), \Delta^s_i),\end{aligned} \tag{16}$$

where $y^{\mathrm{region}}_i = \rho(s_{i,0})$ supervises residue property, while $y^{\mathrm{HL}}_i = \rho(s_{j_{\mathrm{HL}}(i),0})$ and $y^{\mathrm{AbAg}}_i = \rho(s_{j_{\mathrm{AbAg}}(i),0})$ supervise the model to implicitly deduce the biochemical properties of the nearest pairing and binding partners, respectively, for promoting biophysically complementary interactions across inter-chain and antibody–antigen interfaces.

Intrinsic Role Loss ($\mathcal{L}_{\mathrm{self}}$) and Cooperative Role Loss ($\mathcal{L}_{\mathrm{coop}}$). To ensure that inferred functional role assignments accurately reflect biophysical properties rather than arbitrary cluster allocations, we supervise both the intrinsic role and the pairwise-refined role: (1) **Intrinsic Role Supervision ($\mathcal{L}_{\mathrm{self}}$):** We supervise it using the residue's self physicochemical identity y_i alongside intra-chain regional geometric displacements using the prototype representation $h^{\mathrm{self,proto}}_i = (\pi^{\mathrm{role}}_i)^\top P$, where π^{role}_i is the unrefined intrinsic role distribution.:

$$\mathcal{L}_{\mathrm{self}} = \mathrm{CE}(f^{\mathrm{self}}_{\mathrm{chem}}(h^{\mathrm{self,proto}}_i), y_i) + \mathrm{SmoothL1}(f^{\mathrm{self}}_{\mathrm{geo}}((h^{\mathrm{self,proto}}_i), [\Delta^{\mathrm{region}}_i \parallel \Delta^{\mathrm{chain}}_i]). \tag{17}$$

(2) **Cooperative role supervision ($\mathcal{L}_{\mathrm{coop}}$).** Because role refinement via interaction matrix Φ_{ij} explicitly models inter-residue and cross-chain interaction, we penalize mismatches against interfacial

binding partners and complex-level spatial offsets based on the cooperatively refined prototype representation $h_i^{\text{role,proto}} = (\pi_i^{\text{role}})^\top P$:

$$\begin{aligned}\mathcal{L}_{\text{coop}} = \ & \text{CE}(f_{\text{chem}}^{\text{coop,HL}}((h_i^{\text{role,proto}}), y_i^{\text{HL}}) + \text{CE}(f_{\text{chem}}^{\text{coop,AbAg}}((h_i^{\text{role,proto}}), y_i^{\text{AbAg}}) \\ & + \text{SmoothL1}(f_{\text{geo}}^{\text{coop}}((h_i^{\text{role,proto}}), [\Delta_i^{\text{HL}} \parallel \Delta_i^{\text{AbAg}} \parallel \Delta_i^{\text{complex}}]).\end{aligned} \tag{18}$$

Representation Regularization Loss ($\mathcal{L}_{\text{reg}}$). Due to inherent coupling between fundamental R and unit-specific $\{U_e^s\}$ constraints, to avoid degenerate representations and enforce their functional separation, we adopt a composite regularization loss with variance–covariance and prototype orthogonality constraints, ensuring they capture distinct biophysical attributes:

$$\mathcal{L}_{\text{reg}} = \lambda_{\text{var}}\mathcal{L}_{\text{var}} + \lambda_{\text{cov}}\mathcal{L}_{\text{cov}} + \lambda_{\text{cross}}\mathcal{L}_{\text{cross}} + \lambda_{\text{proto}}\mathcal{L}_{\text{proto}}. \tag{19}$$

We define $Z_R \in \mathbb{R}^{B\times d}$ and $Z_{U,e} \in \mathbb{R}^{B\times d}$ as batch feature matrices for R and U_e^s, where B is the batch size and d is the feature dimension. Let $\bar{Z}$ denote the batch mean of Z. The empirical variance along each dimension n is $\text{Var}(Z)_n = \frac{1}{B-1}\sum_{b=1}^{B}(Z_{b,n} - \bar{Z}_n)^2$. The sample covariance matrix is $\text{Cov}(Z) = \frac{1}{B-1}\sum_{b=1}^{B}(Z_b - \bar{Z})(Z_b - \bar{Z})^\top$, and the cross-covariance between Z_R and $Z_{U,e}$ is $\text{Cov}(Z_R, Z_{U,e}) = \frac{1}{B-1}\sum_{b=1}^{B}(Z_{R,b} - \bar{Z}_R)(Z_{U,e,b} - \bar{Z}_{U,e})^\top \in \mathbb{R}^{d\times d}$.

Based on these variances and covariances, we construct four regularization terms as:

(1) **Variance regularization ($\mathcal{L}_{\text{var}}$).** To avoid representation collapse, we constrain the batch-wise standard deviation of each feature dimension to stay near the target value of 1:$\mathcal{L}_{\text{var}} = \frac{1}{|\mathcal{E}|}\sum_{e\in\mathcal{E}}(\frac{1}{d}\sum_{k=1}^{d}\max(0, 1 - \sqrt{\text{Var}(Z_{U,e})_k + \epsilon})) + \frac{1}{d}\sum_{k=1}^{d}\max(0, 1 - \sqrt{\text{Var}(Z_R)_k + \epsilon})$.

(2) **Covariance regularization ($\mathcal{L}_{\text{cov}}$).** To avoid redundant encoding across feature dimensions, we penalize off-diagonal covariance entries to maintain distinct information per dimension: $\mathcal{L}_{\text{cov}} = \frac{1}{|\mathcal{E}|}\sum_{e\in\mathcal{E}}(\frac{1}{d}\sum_{j\neq k}[\text{Cov}(Z_{U,e})]_{jk}^2) + \frac{1}{d}\sum_{j\neq k}[\text{Cov}(Z_R)]_{jk}^2$.

(3) **Cross-correlation regularization ($\mathcal{L}_{\text{cross}}$).** To mitigate inter-branch information overlap between R and U_e^s, we minimize the squared Frobenius norm of their cross-covariance matrix to reduce linear dependence: $\mathcal{L}_{\text{cross}} = \frac{1}{|\mathcal{E}|}\sum_{e\in\mathcal{E}}\frac{1}{d^2}\|\text{Cov}(Z_R, Z_{U,e})\|_F^2$.

(4) **Prototype orthogonality ($\mathcal{L}_{\text{proto}}$):** To prevent learnable functional-role prototypes from collapsing to identical directions, we enforce geometric orthogonality among K predefined role prototypes in matrix $\mathcal{L}_{\text{proto}} = \frac{1}{K^2}\|\bar{P}\bar{P}^\top - I_K\|_F^2$, where $\bar{P} = [\frac{p_1}{\|p_1\|_2}, \ldots, \frac{p_K}{\|p_K\|_2}]^\top$ represents the row-normalized prototype basis matrix, and $I_K \in \mathbb{R}^{K\times K}$ denotes the identity matrix.

A.2 Evaluation Metrics

We evaluate co-designed antibody paratopes across sequence recovery, backbone fold accuracy, local conformational agreement, and binding affinity.

Amino Acid Recovery (AAR). AAR quantifies the sequence identity between generated and native CDR loops. Let $\mathcal{V}_C$ denote the subset of generated CDR residues:

$$\text{AAR} = \frac{1}{|\mathcal{V}_C|}\sum_{i\in\mathcal{V}_C}\mathbf{1}[\hat{s}_i = s_{i,0}], \tag{20}$$

where $\hat{s}_i$ and $s_{i,0}$ denote the generated and ground-truth amino acid type, respectively.

Backbone RMSD. We align the generated antibody to the native structure using the framework ($\mathcal{V}_F$) C_α atoms, and then compute the root-mean-square deviation across all designed CDR loops:

$$\text{RMSD} = \sqrt{\frac{1}{|\mathcal{V}_C|}\sum_{i\in\mathcal{V}_C}\|\hat{p}_i^{\text{C}_\alpha} - p_{i,0}^{\text{C}_\alpha}\|_2^2}. \tag{21}$$

Global and Region-wise TM-score (TM and G-TM). TM-score measures structural topology consistency independent of protein length (Zhang & Skolnick, 2004). We report the global TM-score computed across the full complex, as well as region-wise G-TM evaluated exclusively on the CDR loops $\mathcal{V}_C$ using TM-score.

Table 3: Performance of each CDR on RMSD and AAR. The best result is shown in bold, and the second-best result is underlined.

Heavy-chain CDRs				Light-chain CDRs			
Region	Method	RMSD ↓ (Å)	AAR ↑ (%)	Region	Method	RMSD ↓ (Å)	AAR ↑ (%)
HCDR1	DiffAb	1.0554	66.3709	LCDR1	DiffAb	1.0136	59.3854
	dyMEAN	1.0745	75.5644		dyMEAN	0.9466	<u>75.5535</u>
	AbX	<u>0.8932</u>	<u>78.0968</u>		AbX	<u>0.8626</u>	75.4752
	AAMFM	1.4234	**79.7783**		AAMFM	0.9057	**84.0634**
	AbFlow	1.1926	73.9589		AbFlow	1.1788	68.8894
	MoDA	**0.8657**	75.2401		MoDA	**0.8013**	71.1917
HCDR2	DiffAb	**0.7481**	42.4751	LCDR2	DiffAb	0.6890	54.6997
	dyMEAN	0.9982	**68.3598**		dyMEAN	0.6387	<u>83.2341</u>
	AbX	0.8402	65.3724		AbX	**0.5416**	73.9944
	AAMFM	1.5858	<u>66.4024</u>		AAMFM	0.9292	**88.6755**
	AbFlow	1.0741	65.1444		AbFlow	0.8006	81.6071
	MoDA	<u>0.7929</u>	62.1222		MoDA	<u>0.5733</u>	75.1551
HCDR3	DiffAb	3.3461	23.6599	LCDR3	DiffAb	1.3738	46.7942
	dyMEAN	4.0864	<u>36.4426</u>		dyMEAN	1.4133	51.9242
	AbX	2.9618	**38.8927**		AbX	1.2949	61.7005
	AAMFM	3.7407	33.3906		AAMFM	**1.1234**	**70.6895**
	AbFlow	**2.6850**	36.1899		AbFlow	1.1876	57.5493
	MoDA	<u>2.7131</u>	27.9128		MoDA	<u>1.1252</u>	<u>61.8832</u>

Global and Region-wise lDDT (lDDT and G-lDDT). Local Distance Difference Test (lDDT) measures local distance matrix preservation under multiple distance thresholds (0.5 Å, 1.0 Å, 2.0 Å, 4.0 Å) (Mariani et al., 2013). We report full-complex lDDT and generated CDR-specific G-lDDT.

Binding Energy Success Rate (ddG-SR). To estimate binding affinity, all generated complexes undergo side-chain repacking and local energy minimization via PyRosetta InterfaceAnalyzer using the `ref2015` energy function (Chaudhury et al., 2010; Alford et al., 2017). For each candidate $m \in \mathcal{C}_n$ sampled for target complex n, the binding free energy change relative to the wild-type native complex is defined as:

$$\Delta\Delta G_{n,m} = \Delta G_{\text{gen},n,m} - \Delta G_{\text{native},n}. \tag{22}$$

Candidates with $\Delta\Delta G < 0$ represent predicted affinity-enhancing binders. The success rate ddG-SR across N targets is computed as:

$$\text{ddG-SR} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}[\min_{m\in\mathcal{C}_n} \Delta\Delta G_{n,m} < 0]. \tag{23}$$

A.3 Experiments

A.3.1 Antibody Paratope Co-design

As shown in Tab. 3, we further report the RMSD and AAR for each CDR. MoDA exhibits consistently strong structural accuracy across all six CDRs, achieving either the best or second-best RMSD in every region. Specifically, MoDA ranks first on HCDR1 and LCDR1, and second on HCDR2, HCDR3, LCDR2, and LCDR3. This consistent performance demonstrates its robust generalization across heavy- and light-chain regions with varying degrees of conformational flexibility. AAMFM achieves higher native-sequence recovery in several regions and obtains the lowest RMSD on LCDR3, but its structural accuracy is less consistent across the remaining CDRs.

Notably, region-wise evaluation measures the local reconstruction accuracy of each CDR separately, whereas the overall metrics additionally reflect whether all generated CDRs form a geometrically coherent paratope under antigen and cross-region constraints. Comparing the results in Tab. 1 and

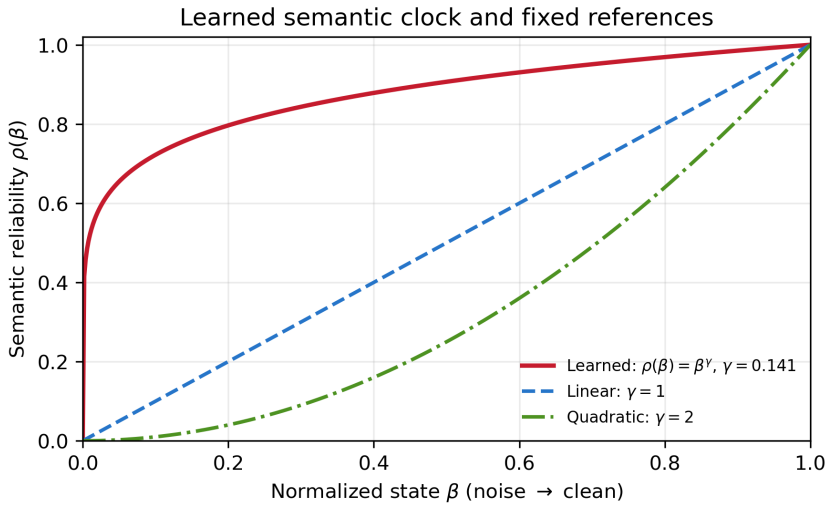


Figure 5: Temporal dynamics of reliability scheduling functions under learnable and fixed schedules ($\gamma = 1, \gamma = 2$).

Table 4: Effect of the reliability schedule on CDRs co-design. Fixed Linear and Fixed Quadratic respectively use $\rho(t) = \beta_t$ and $\rho(t) = \beta_t^2$, while Learnable schedule is $\rho(t) = \beta_t^{\gamma}, \gamma = 0.141$.

Schedule	RMSD ↓	TM ↑	G-TM ↑	lDDT ↑	G-lDDT ↑	AAR ↑	ddG-SR ↑
Fixed Linear	1.6116	0.9737	0.8484	0.9585	0.7657	**63.0096**	55.0%
Fixed Quadratic	1.6521	0.9733	0.8415	0.9564	0.7608	61.6468	56.7%
Learnable	**1.5776**	**0.9745**	**0.8485**	**0.9586**	**0.7668**	62.2509	**58.3%**

Tab. 3, MoDA may not attain the lowest RMSD or highest AAR in every individual region, yet it achieves the best overall RMSD, G-TM, and G-lDDT by better coordinating the relative structures of multiple CDRs during simultaneous generation. These results also align with MoDA's objective to maintain overall paratope coherence rather than optimize isolated CDRs. Moreover, CDR sequences are intrinsically diverse, where different sequences may adopt similar structures and recognize the same antigen. AAR with respect to a single native sequence does not fully characterize the quality of generated candidates. The highest $\Delta\Delta G$-SR, together with its high sequence diversity, demonstrates that MoDA effectively explores diverse yet high-affinity antibody candidates.

A.3.2 Effect of Learnable Reliability Scheduling

To validate the necessity of learnable scheduling, we conduct a comparison against two fixed schedules: the *linear schedule* ($\gamma = 1$) and the *quadratic schedule* ($\gamma = 2$). The reliability of intermediate residues evolves dynamically along flow trajectories: representations carry high uncertainty at early generation stages and become increasingly reliable as structural and sequential refinement proceeds. Fig. 5 illustrates the distinct temporal trends among our learnable schedule $\rho(t) = \beta_t^{\gamma}$ and the two fixed alternatives. With these distinct scheduling schemes, we perform CDR co-design under identical settings and compare their performance. As shown in Tab. 4, the learnable schedule achieves the lowest structural RMSD of 1.5776 Å, outperforming the linear (1.6116 Å) and quadratic (1.6521 Å) counterparts, along with consistently higher G-TM and G-lDDT scores. Although the linear schedule yields marginally higher AAR, our learnable strategy improves overall structural quality and boosts the binding success rate from 55.0% to 58.3%. These results confirm that dynamically learnable reliability scheduling better characterizes temporal semantic evolution and benefits the exploration of high-affinity antibodies.