PODA: PRIVILEGED ON-POLICY DISTILLATION FOR ANTIBODY AND NANOBODY DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

Epitope-conditioned de novo antibody design seeks to generate CDR sequences and the corresponding antibody–antigen complex structure for a specified target site. Flow-matching models learn from interpolations toward native complexes, but generate through their own predicted states, creating a potential mismatch between training and inference. We introduce PODA (Privileged On-Policy Distillation for Antibody and Nanobody Design), a two-stage framework for joint CDR sequence and complex-structure generation. Given an antigen sequence and structure, a specified epitope, and a framework sequence with masked CDRs, PODA first trains a deployable student and a privileged teacher with supervised objectives on the same complexes. It then freezes the teacher and updates the student on states obtained by re-noising endpoints generated online by the evolving student policy. Distillation matches teacher-predicted clean endpoints, with residue-level weighting emphasizing CDR and interface geometry and a quality gate selecting teacher targets according to their reconstruction error. On the sequence-filtered test set, PODA increases the design success rate from 20.4% to 32.3% on antibody design and from 12.0% to 24.0% on nanobody design relative to the supervised base model. Ablations support the contribution of privileged supervision, and PODA achieves the highest mean DockQ among the evaluated post-training methods in both cohorts. These findings support privileged endpoint supervision on student-induced states as an effective approach to improving complex interface geometry.

1 INTRODUCTION

Epitope-targeted antibody and nanobody design aims to control molecular recognition by coordinating complementarity-determining region (CDR) sequences, loop conformations, and antibody–antigen placement (Kong et al., 2023b; Bennett et al., 2026). Classical approaches explore this coupled space through conformational sampling and energy optimization (Pantazes & Maranas, 2010; Lapidoth et al., 2015; Adolf-Bryfogle et al., 2018). Learning-based methods instead model sequence and geometry jointly, using graph-based co-design (Jin et al., 2021; Kong et al., 2023a), diffusion (Luo et al., 2022), or flow matching (Wu et al., 2025). Building on progress in generative protein design (Watson et al., 2023), RFdiffusion models and predictor-guided pipelines such as Germinal have demonstrated experimentally validated epitope targeting (Bennett et al., 2026; Mille-Fragoso et al., 2025).

We study epitope-conditioned de novo design for VH–VL antibodies and VHH nanobodies in the framework-sequence-conditioned setting of dyMEAN and IgGM (Kong et al., 2023b; Wang et al., 2025). Given an antigen sequence and structure, a specified epitope, and a framework sequence with masked CDRs, the model generates CDR identities and the complete antibody structure in the antigen coordinate frame, without receiving antibody coordinates or the native pose.

For flow-based generators, a central difficulty is the potential mismatch between training and inference. Flow matching learns along prescribed noise-to-data paths (Lipman et al., 2023; Liu et al., 2023), whereas generation recursively follows the model's own predictions. Accumulated errors can move sampled states beyond those well covered by supervised training, affecting CDR conformation and interface placement (Yang et al., 2026).

Post-training can refine these generators through preference optimization (Wallace et al., 2024; Zhou et al., 2024), online policy optimization (Liu et al., 2025; He et al., 2025), or reward-driven regression on noised outputs (Zheng et al., 2025; Bergmeister et al., 2026). Complete training complexes provide

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.

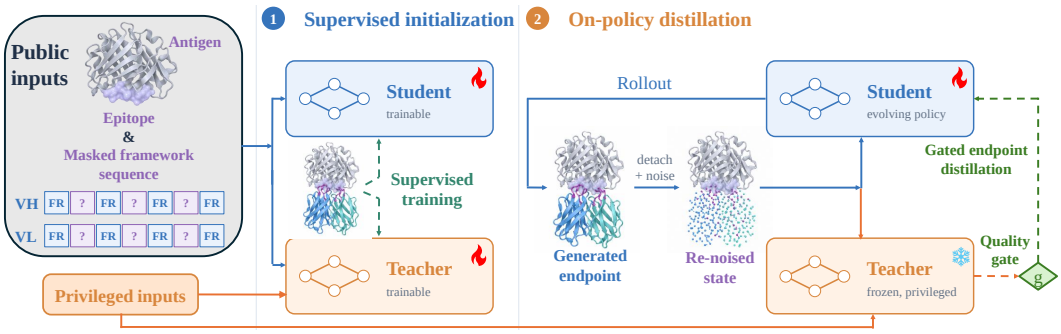

Figure 1: Overview of PODA. Stage I supervises a public-input student and a privileged teacher on the same training complexes. Stage II freezes the teacher and distills its clean-endpoint predictions on detached, re-noised student outputs. Privileged inputs condition only the teacher; the student's deployment inputs remain unchanged. Molecular views are schematic.

a complementary source of dense supervision: ABOPD uses privileged native geometry for on-policy CDR distillation (Yang et al., 2026). We investigate this principle for joint CDR sequence and complete complex generation across VH–VL and VHH formats, using teacher-predicted endpoints to guide the student on states induced by its own outputs.

We introduce PODA (Privileged On-Policy Distillation for Antibody and Nanobody Design), a two-stage sequence–structure flow framework. Stage I supervises a public-input student and a privileged teacher on the same complexes; the resulting student is PODA (base). Stage II freezes the teacher and updates the student on re-noised endpoints generated online by its evolving policy. Distillation matches teacher-predicted clean endpoints, emphasizes CDR and interface residues, and gates teacher targets by reconstruction quality. The student retains the original deployment inputs at inference.

Our contributions are threefold:

- **A unified design pipeline.** PODA jointly generates CDR sequences and complex structures for VH–VL antibodies and VHH nanobodies under the same framework-sequence-conditioned formulation.
- **On-policy privileged distillation.** We combine endpoint matching on re-noised student outputs, interface-aware weighting, and teacher-quality gating to transfer privileged structural supervision.
- **Systematic empirical validation.** We evaluate all 93 VH–VL and 50 VHH test targets through design comparisons, training-stage and teacher ablations, alternative post-training methods, and complementary structural diagnostics.

2 Related Work

Antigen-conditioned antibody design. Classical methods combine energy functions with conformational sampling (Pantazes & Maranas, 2010; Lapidoth et al., 2015; Adolf-Bryfogle et al., 2018; Warszawski et al., 2019). Learning-based approaches extend sequence modeling (Saka et al., 2021; Akbar et al., 2022) to joint sequence–structure design (Jin et al., 2021; Kong et al., 2023a; Luo et al., 2022). Like dyMEAN and IgGM (Kong et al., 2023b; Wang et al., 2025), PODA generates CDRs and antibody–antigen structures from an antigen, a specified epitope, and a masked framework sequence, without antibody coordinates at deployment. This differs from fixed-backbone inverse folding (Hsu et al., 2022; Gao et al., 2023); comparisons therefore distinguish methods by their required inputs.

Continuous generative modeling for biomolecular structure. Diffusion and flow matching (Song & Ermon, 2019; Ho et al., 2020; Lipman et al., 2023; Albergo & Vanden-Eijnden, 2023; Liu et al., 2023) support protein backbone and all-atom generation (Watson et al., 2023; Ingraham et al., 2023; Chu et al., 2024), as well as antibody design and affinity maturation (Zhu et al., 2024; Bennett et al., 2026; Wu et al., 2025; Wang et al., 2026; Tan et al., 2025; Chen et al., 2025). Predictor-guided alternatives include BindCraft (Pacesa et al., 2025) and Germinal (Mille-Fragoso et al., 2025), while filtering and increased sampling provide inference-time routes to better candidates (Stark et al.,

2025; Didi et al., 2026). PODA instead improves a joint sequence–structure flow generator through post-training, without expanding its deployment inputs or search budget.

Post-training with privileged supervision. Post-training approaches include preference optimization for diffusion models (Wallace et al., 2024), reward-driven protein generation (Lutz et al., 2023), and energy-based antibody preferences (Zhou et al., 2024). Flow-GRPO applies online policy optimization to flow matching (Liu et al., 2025), with temporal refinements in TempFlow-GRPO (He et al., 2025). DiffusionNFT uses forward-process training on generated samples (Zheng et al., 2025), whereas RAM constructs reward-corrected regression targets (Bergmeister et al., 2026). Most closely related, ABOPD uses privileged native geometry for on-policy CDR distillation (Yang et al., 2026). PODA studies privileged supervision for joint CDR sequence and complex-structure generation across VH–VL and VHH formats. A frozen teacher supplies quality-gated clean-endpoint targets on re-noised endpoints collected online from the evolving student, transferring training-only information without changing inference conditioning. Appendix C.7 details these distinctions.

3 METHOD

PODA learns a shared sequence–structure generator for epitope-conditioned antibody and nanobody design in two stages. Stage I trains a public-input student and a privileged teacher with supervised objectives on the same complexes. Stage II freezes the teacher and updates the student on states obtained by re-noising endpoints generated online by its evolving policy. The frozen teacher supplies quality-gated clean-endpoint targets on these states. The two stages differ in their supervision and sampled states, while the student's conditioning inputs remain fixed.

3.1 TASK AND JOINT SEQUENCE–STRUCTURE REPRESENTATION

Inputs and outputs. Let $\mathcal{A} = (\mathbf{C}_{ag}, \mathbf{s}_{ag})$ be the antigen structure and sequence, $\mathbf{m}_{epi}$ a specified epitope mask, and $\mathbf{s}_{fw}^{mask}$ the antibody sequence with its framework residues visible and its CDR residues masked. The deployment condition is

$$\mathbf{c} = (\mathcal{A}, \mathbf{m}_{epi}, \mathbf{s}_{fw}^{mask}). \tag{1}$$

We generate the missing CDR identities and the complete antibody structure in the coordinate frame of the supplied antigen. The antigen remains fixed. This framework-sequence-conditioned setting requires no antibody framework coordinates, native CDR identities, or reference antibody pose. Chain and CDR masks specify the VH–VL or VHH layout.

A training complex additionally supplies privileged information $\mathbf{p}$: the complete antibody sequence, native antibody coordinates, and interface annotations. These fields condition the teacher and define training losses, but are excluded from student conditioning. Ground-truth complexes still provide the student's supervised training states and targets.

Joint representation. The student and teacher operate on a joint coordinate–latent representation of antibody sequence and structure. A frozen partial autoencoder maps the training antibody to explicit C_α coordinates and per-residue sequence–side-chain latents:

$$\mathbf{u}^* = (\mathbf{X}^*, \mathbf{Z}^*), \qquad \mathbf{Z}^* = \mathrm{Enc}(\mathbf{s}_{ab}^*, \mathbf{C}_{ab}^*), \tag{2}$$

where $\mathbf{X}^* \in \mathbb{R}^{N\times 3}$ and $\mathbf{Z}^* \in \mathbb{R}^{N\times 8}$. A frozen decoder maps a generated endpoint back to residue identities and full-atom antibody geometry; the design outputs are the masked CDR identities and the antibody coordinates. VH–VL antibodies and VHH nanobodies use this same representation. The flow network maintains residue and pair features and recycles intermediate coordinate estimates into the pair representation. Architecture details are given in Appendix C.1.

3.2 STAGE I: SUPERVISED STUDENT–TEACHER TRAINING

Let $\mathbf{u}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ denote noise in the generator's coordinate–latent state space. The supervised linear path is

$$\mathbf{u}_t = (1-t)\mathbf{u}_0 + t\mathbf{u}^*, \qquad \mathbf{v}^* = \mathbf{u}^* - \mathbf{u}_0. \tag{3}$$

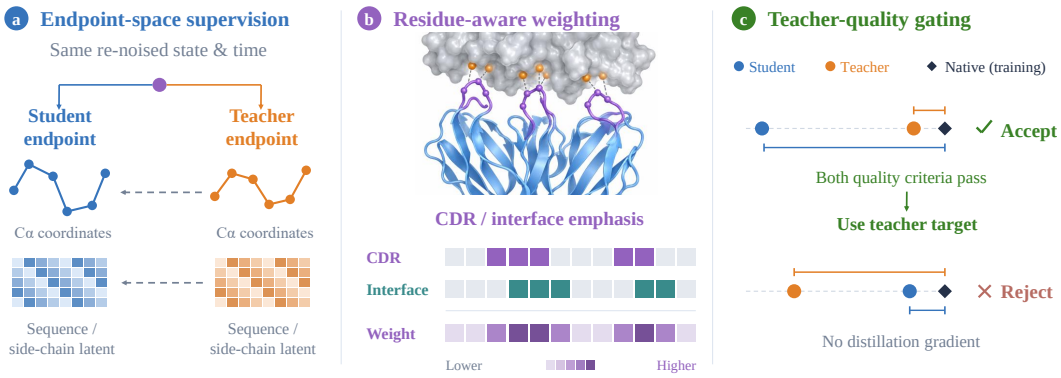

Figure 2: Mechanisms of privileged on-policy distillation. (a) Student and teacher predict clean endpoints from the same re-noised state and time. Detached teacher endpoints supervise explicit C_α coordinates and sequence–side-chain latents. (b) Residue-level weighting emphasizes CDR and interface geometry; CDR and interface masks can overlap. (c) Teacher targets are accepted only when they satisfy both the relative improvement and target-quality criteria against native training endpoints.

The student predicts $\mathbf{v}_\theta(\mathbf{u}_t, t \mid \mathbf{c})$; the teacher predicts $\mathbf{v}_\phi(\mathbf{u}_t, t \mid \mathbf{c}, \mathbf{p})$. Both are trained with ground-truth flow and structural supervision:

$$\mathcal{L}_{\mathrm{FM}}(\psi; \mathbf{h}) = \mathbb{E}\Big[\|\mathbf{v}_\psi(\mathbf{u}_t, t \mid \mathbf{h}) - \mathbf{v}^*\|^2_{\mathrm{mask}}\Big], \tag{4}$$

$$\mathcal{L}_{\mathrm{sup}}(\psi; \mathbf{h}) = \mathcal{L}_{\mathrm{FM}}(\psi; \mathbf{h}) + \mathcal{L}_{\mathrm{aux}}(\psi; \mathbf{h}), \tag{5}$$

$$\mathcal{L}_{\mathrm{I}} = \mathcal{L}_{\mathrm{sup}}(\theta; \mathbf{c}) + \mathcal{L}_{\mathrm{sup}}(\phi; (\mathbf{c}, \mathbf{p})). \tag{6}$$

Here ψ denotes the parameters of either model and $\mathbf{h}$ its conditioning view. The expectation covers training complexes, noise, and sampled times. Masked reductions exclude padding; auxiliary losses supervise geometry and interface contacts (Appendix C.6). Both student and teacher are updated in this stage, with the autoencoder fixed. The resulting public-input student is PODA (base). Its parameters initialize Stage II, while the teacher and its privileged encoder are frozen.

3.3 Stage II: Online Privileged Distillation

Supervised flow matching samples states from paths ending at ground-truth complexes. At inference, states instead depend on the generator's own predictions. We address this potential mismatch by constructing additional training states from endpoints generated by the current student and asking the privileged teacher to correct their predicted completions.

Online endpoint collection. At update k, the student with parameters θ_k generates

$$\mathbf{u}_1^{roll} = \mathrm{sg}[\mathrm{ODESolve}(\mathbf{v}_{\theta_k}, \boldsymbol{\xi}; \mathbf{c})], \qquad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{7}$$

where sg denotes stop-gradient. Using fresh noise $\boldsymbol{\epsilon}$ and a sampled time t, we form

$$\widetilde{\mathbf{u}}_t = (1-t)\boldsymbol{\epsilon} + t\mathbf{u}_1^{roll}. \tag{8}$$

These are *rollout-induced interpolation states*; they need not coincide with intermediate states visited by the ODE solver. Here, on-policy means that their endpoint distribution is induced by the evolving student: after an update, subsequent collection uses the updated policy. The rollout and re-noising operations carry no gradient.

Same state, different conditioning. On each re-noised state, student and teacher predict

$$\mathbf{v}_S = \mathbf{v}_\theta(\widetilde{\mathbf{u}}_t, t \mid \mathbf{c}), \qquad \mathbf{v}_T = \mathbf{v}_\phi(\widetilde{\mathbf{u}}_t, t \mid \mathbf{c}, \mathbf{p}). \tag{9}$$

The linear-path parameterization gives clean-endpoint estimates

$$\widehat{\mathbf{u}}_{1,S} = \widetilde{\mathbf{u}}_t + (1-t)\mathbf{v}_S, \qquad \widehat{\mathbf{u}}_{1,T} = \widetilde{\mathbf{u}}_t + (1-t)\mathbf{v}_T. \tag{10}$$

This expresses teacher supervision in the same space as the desired sequence–structure output. For the same input state and weighting, endpoint matching equals velocity matching scaled by $(1-t)^2$;

this makes its implicit time weighting explicit. Endpoint space also permits direct comparison with the clean training complex.

Selective, interface-aware supervision. Let e_S and e_T denote masked endpoint errors against $\mathbf{u}^*$. We select teacher targets with the quality gate

$$g = \mathrm{sg}(\mathbf{1}[e_T \le (1-m)e_S]\,\mathbf{1}[\mathrm{relL2}(\widehat{\mathbf{u}}_{1,T}) \le \tau]), \tag{11}$$

where m sets the required improvement margin and τ bounds the teacher endpoint's relative-error diagnostic. The gate uses training targets only and is not evaluated during generation. For each mode $q \in \{C_\alpha, \mathrm{latent}\}$, define a weighted reduction

$$D_q(\mathbf{a}, \mathbf{b}) = \frac{\sum_i w_i^q \|\mathbf{a}_i - \mathbf{b}_i\|_2^2}{\sum_i w_i^q}. \tag{12}$$

The weights mask invalid residues and emphasize CDR/interface coordinates, so that framework residues do not dominate the structural correction. The distillation objective is

$$\mathcal{L}_{\mathrm{OPD}} = \mathbb{E}\left[g \sum_q D_q\Big(\widehat{\mathbf{u}}_{1,S}^q, \mathrm{sg}(\widehat{\mathbf{u}}_{1,T}^q)\Big)\right]. \tag{13}$$

Native interface annotations supply loss weights and teacher features, not additional student conditioning. Rejected examples contribute no distillation gradient.

Auxiliary RAM correction and supervised anchor. The post-training objective also includes an auxiliary RAM correction (Bergmeister et al., 2026). A frozen public-input reference predicts $\mathbf{v}_{ref}$ at the same state. With a clipped, detached structural score r, its regression target and loss are

$$\mathbf{v}_{\mathrm{RAM}} = \mathrm{sg}\big[\mathbf{v}_{ref} + \eta r\big((\mathbf{u}_1^{roll} - \boldsymbol{\epsilon}) - \mathbf{v}_S\big)\big], \tag{14}$$

$$\mathcal{L}_{\mathrm{RAM}} = \mathbb{E}\big[g\|\mathbf{v}_S - \mathbf{v}_{\mathrm{RAM}}\|_{\mathrm{mask}}^2\big]. \tag{15}$$

Stop-gradient applies to the *entire* regression target, including its student-velocity term. The score supplies an auxiliary structural signal; it is not a measured binding affinity. The complete objective is

$$\mathcal{L}_{\mathrm{II}} = \mathcal{L}_{\mathrm{OPD}} + \lambda_{\mathrm{RAM}}\mathcal{L}_{\mathrm{RAM}} + \lambda_{\mathrm{anchor}}\mathcal{L}_{\mathrm{sup}}(\theta; \mathbf{c}). \tag{16}$$

The anchor is evaluated on ground-truth interpolation states as in Eq. equation 3; it preserves supervised reconstruction while the additional losses act on re-noised student endpoints. Only the student is optimized in Stage II. Setting $\lambda_{\mathrm{RAM}} = 0$ gives the RAM-off control; omitting Stage II gives PODA (base).

3.4 Teacher-Free Generation

At inference, PODA integrates its learned velocity field from noise using only $\mathbf{c}$ and decodes the terminal coordinate–latent state. The requested CDR sequence and antibody structure are returned with the supplied antigen fixed. The teacher, privileged encoder, quality gate, RAM score, and supervised losses are training components and are not needed for generation. No native antibody structure is required for candidate generation or selection. Reference complexes are used afterwards to evaluate the generated outputs. In the separate structure-prediction setting, the complete antibody sequence replaces the CDR-masked sequence.

4 Experiments

We evaluate PODA along three axes: joint CDR sequence–structure design, the effect of privileged post-training, and complex prediction when the complete antibody sequence is available.

4.1 Experimental Setup

Benchmark and tasks. The SAbDab benchmark (Dunbar et al., 2014) contains 12,759 complexes: 12,364 for training, 252 for validation, and 143 for testing. We report the 93 VH–VL antibodies and 50 VHH nanobodies separately, with CDRs defined by IMGT numbering (Lefranc et al., 2003; 2009).

Table 1: Results for de novo antibody design on the 93-antibody test set. dyMEAN, IgGM, and our models operate in the framework-sequence-conditioned setting; RFantibody instead uses an IgFold-initialized framework structure. Other systems follow their documented design protocols.

Model		RFantibody (IgFold)	dyMEAN	IgGM	BoltzGen	Germinal	PODA (Base)	**PODA**
AAR↑	L1	$0.185^{0.162}$	$0.678^{0.237}$	$0.628^{0.245}$	$0.492^{0.217}$	$0.420^{0.156}$	$\mathbf{0.693}^{0.196}$	$\underline{0.692}^{0.203}$
	L2	$0.266^{0.233}$	$0.710^{0.332}$	$0.629^{0.343}$	$0.551^{0.252}$	$0.583^{0.211}$	$\underline{0.821}^{0.252}$	$\mathbf{0.824}^{0.261}$
	L3	$0.174^{0.131}$	$0.624^{0.174}$	$0.585^{0.188}$	$0.438^{0.175}$	$0.427^{0.151}$	$\mathbf{0.727}^{0.246}$	$\underline{0.723}^{0.245}$
	H1	$0.262^{0.119}$	$\mathbf{0.751}^{0.179}$	$0.644^{0.154}$	$0.601^{0.167}$	$0.532^{0.166}$	$0.696^{0.194}$	$\underline{0.702}^{0.193}$
	H2	$0.225^{0.155}$	$\underline{0.662}^{0.207}$	$0.350^{0.179}$	$0.424^{0.159}$	$0.454^{0.183}$	$\mathbf{0.667}^{0.226}$	$0.658^{0.238}$
	H3	$0.109^{0.077}$	$0.378^{0.126}$	$0.245^{0.133}$	$0.293^{0.108}$	$0.277^{0.113}$	$\mathbf{0.512}^{0.192}$	$\underline{0.506}^{0.196}$
RMSD↓	L1	$1.247^{1.057}$	$0.803^{0.603}$	$0.828^{0.796}$	$0.777^{0.811}$	$1.260^{0.827}$	$\mathbf{0.629}^{0.665}$	$\mathbf{0.629}^{0.660}$
	L2	$0.155^{0.160}$	$0.263^{0.303}$	$0.162^{0.147}$	$0.141^{0.169}$	$0.258^{0.434}$	$\underline{0.138}^{0.177}$	$\mathbf{0.135}^{0.190}$
	L3	$1.113^{0.781}$	$0.762^{0.729}$	$0.807^{0.788}$	$0.788^{0.800}$	$0.859^{0.588}$	$\mathbf{0.728}^{0.753}$	$\underline{0.742}^{0.764}$
	H1	$1.025^{0.694}$	$0.796^{0.522}$	$0.778^{0.540}$	$\mathbf{0.740}^{0.555}$	$0.774^{0.474}$	$0.772^{0.571}$	$\underline{0.758}^{0.559}$
	H2	$0.781^{0.670}$	$0.645^{0.397}$	$0.793^{0.488}$	$0.569^{0.524}$	$1.489^{0.717}$	$\mathbf{0.538}^{0.483}$	$\underline{0.555}^{0.491}$
	H3	$2.759^{1.009}$	$\mathbf{2.224}^{0.865}$	$2.572^{0.926}$	$2.533^{1.106}$	$2.607^{1.674}$	$\underline{2.335}^{1.035}$	$2.349^{0.965}$
Structure↑	TMscore	$0.907^{0.068}$	$\underline{0.928}^{0.043}$	$0.918^{0.045}$	$0.925^{0.053}$	$0.924^{0.100}$	$\mathbf{0.929}^{0.046}$	$0.925^{0.049}$
	lDDT	$0.721^{0.061}$	$0.777^{0.059}$	$0.635^{0.049}$	$0.705^{0.062}$	$0.653^{0.058}$	$\mathbf{0.783}^{0.066}$	$\underline{0.778}^{0.066}$
Side chain↓	Clash rate	$0.030^{0.027}$	$0.049^{0.020}$	$0.025^{0.007}$	$0.021^{0.002}$	$0.038^{0.020}$	$\underline{0.018}^{0.017}$	$\mathbf{0.013}^{0.014}$
	scRMSD	$17.667^{7.909}$	$17.779^{6.808}$	$15.055^{7.340}$	$19.999^{7.929}$	$20.398^{5.904}$	$\underline{14.227}^{7.308}$	$\mathbf{12.294}^{5.939}$
Interface	Polar contact↑	$0.004^{0.018}$	$0.013^{0.049}$	$0.011^{0.064}$	$\underline{0.016}^{0.048}$	$0.000^{0.000}$	$0.014^{0.038}$	$\mathbf{0.024}^{0.065}$
	$\lvert\Delta\text{BSA}\rvert$↓	$976.0^{487.0}$	$1135.0^{1077.2}$	$819.9^{449.3}$	$672.1^{411.3}$	$698.8^{548.5}$	$\underline{248.9}^{151.3}$	$\mathbf{210.7}^{142.6}$
	Median ΔG↓	-41.8	7.4	$\underline{-42.3}$	-41.6	-15.0	-32.2	$\mathbf{-48.3}$
Docking	DockQ↑	$0.069^{0.073}$	$0.077^{0.065}$	$0.139^{0.160}$	$0.124^{0.114}$	$0.122^{0.121}$	$\underline{0.141}^{0.173}$	$\mathbf{0.170}^{0.177}$
	iRMS↓	$10.337^{5.688}$	$9.563^{3.936}$	$\underline{8.563}^{5.405}$	$8.707^{4.957}$	$8.849^{4.087}$	$8.993^{5.383}$	$\mathbf{8.533}^{5.390}$
	LRMS↓	$28.854^{16.073}$	$27.510^{11.337}$	$\underline{21.695}^{14.729}$	$21.697^{11.920}$	$26.712^{12.053}$	$23.825^{15.166}$	$\mathbf{21.440}^{11.719}$
	SR↑	0.043	0.032	0.194	0.172	0.172	$\underline{0.204}$	**0.323**

The test set is held out from model development. Design conditions on the antigen sequence and structure, a specified epitope, and a CDR-masked framework sequence; prediction instead reveals the complete antibody sequence. For this benchmark, a fixed contact rule defines the epitope, supplied identically to methods that accept it. The sequence-filtering rule is detailed in Appendix B.

Models and evaluation. PODA (base) is the Stage-I supervised student, and PODA is the model obtained after Stage II. The generator has 171M parameters, 14 transformer layers, token width 384, pair width 256, and 12 attention heads. Unless otherwise stated, inference uses 400 integration steps and one candidate per target. We measure CDR recovery and RMSD for local design, TM-score and lDDT for global structure, and DockQ, iRMS, LRMS, and success rate (SR; DockQ> 0.23) for docking. Additional side-chain and interface diagnostics, sequence refolding, and integration-budget and timing analyses are provided in Appendices D.6, D.3, and D.4.

Comparison protocol. All baseline models are faithfully implemented following their respective papers and use the same training and testing settings as PODA. Method-specific architectures, objectives, and required inputs follow the original methods; the data splits and evaluation protocol are shared. We distinguish framework-sequence conditioning from methods requiring framework coordinates, MSAs, or other inputs. Method-specific inputs are identified alongside the comparisons; metric definitions and additional evaluation details appear in Appendix D.1 and the corresponding experimental subsections.

4.2 De Novo Antibody and Nanobody Design

Antibody design. Table 1 compares PODA with the framework-sequence-conditioned dyMEAN and IgGM models (Kong et al., 2023b; Wang et al., 2025), as well as RFantibody (Bennett et al., 2026), BoltzGen (Stark et al., 2025), and Germinal (Mille-Fragoso et al., 2025) under their respective input requirements. PODA achieves the strongest docking results across the reported metrics, while other methods lead on some loop-reconstruction and global-fold metrics. Relative to PODA (base), PODA achieves higher DockQ and SR alongside lower interface and ligand RMSDs, while CDR recovery and fold scores change only modestly. This pattern points to improved antibody–antigen placement as the main benefit of post-training: a native-like interface requires accurate relative placement beyond reconstruction of individual loops or the antibody fold.

Nanobody design. Table 2 reports the separate 50-target VHH cohort. On this single-domain cohort, PODA improves DockQ and SR over PODA (base), supporting the applicability of post-training beyond the VH–VL architecture. PODA leads in DockQ, SR, and LRMS, yet other methods obtain lower iRMS or CDR-H3 RMSD. The separation between local loop accuracy and complex-level

Table 2: Results for de novo design on the 50-nanobody test set. DiffAb (AF3) uses an AlphaFold 3-initialized framework structure; IgGM, PODA (base), and PODA use framework-sequence conditioning. AAR H1–H3 denote heavy-chain CDR recovery, and RMSD H3 measures CDR-H3 geometry.

Method	AAR H1↑	AAR H2↑	AAR H3↑	RMSD H3↓	DockQ↑	iRMS↓	LRMS↓	SR↑
DiffAb (AF3)	$0.419^{0.207}$	$0.216^{0.160}$	$0.228^{0.177}$	$2.773^{1.963}$	$0.053^{0.031}$	$16.874^{6.567}$	$43.632^{23.630}$	0.020
IgGM	$0.243^{0.159}$	$0.180^{0.125}$	$0.107^{0.096}$	$2.806^{1.301}$	$0.114^{0.106}$	$\mathbf{7.656}^{3.587}$	$\underline{20.308}^{11.662}$	$\underline{0.120}$
BoltzGen	$0.362^{0.200}$	$0.404^{0.158}$	$0.204^{0.091}$	$\mathbf{2.268}^{1.037}$	$\underline{0.127}^{0.127}$	$8.050^{4.560}$	$21.555^{15.806}$	$\underline{0.120}$
Germinal	$0.347^{0.163}$	$0.256^{0.167}$	$0.240^{0.147}$	$2.639^{1.068}$	$0.112^{0.102}$	$8.656^{2.772}$	$22.373^{10.256}$	0.100
PODA (base)	$\underline{0.478}^{0.227}$	$\mathbf{0.553}^{0.235}$	$\mathbf{0.458}^{0.242}$	$\underline{2.473}^{1.253}$	$0.114^{0.112}$	$8.417^{3.324}$	$20.565^{8.292}$	$\underline{0.120}$
PODA	$\mathbf{0.493}^{0.234}$	$\underline{0.552}^{0.244}$	$\underline{0.455}^{0.249}$	$2.520^{1.305}$	$\mathbf{0.150}^{0.143}$	$\underline{8.015}^{3.497}$	$\mathbf{18.655}^{7.898}$	$\mathbf{0.240}$

Table 3: Side-chain and interface quality of de novo designs on 50 VHH targets.

Method	Clash rate (%)↓	Interface scRMSD (Å)↓	Polar-contact recovery (%)↑	Median ΔG (REU)↓	Mean $\lvert\Delta\text{BSA}\rvert$ (Å^2)↓
DiffAb (AF3)	$\mathbf{0.390}^{0.810}$	$\underline{15.694}^{7.186}$	$0.000^{0.000}$	-18.915	$303.274^{275.900}$
IgGM	$1.280^{1.560}$	$16.699^{6.952}$	$0.120^{0.820}$	-12.957	$212.271^{154.013}$
BoltzGen	$1.720^{1.400}$	$15.722^{6.952}$	$\underline{1.290}^{4.220}$	-7.561	$\underline{200.270}^{146.672}$
Germinal	$1.900^{1.800}$	$18.539^{3.722}$	$0.100^{0.700}$	-17.464	$292.042^{237.260}$
PODA (base)	$1.920^{2.030}$	$16.001^{8.208}$	$0.920^{2.660}$	$\underline{-21.193}$	$\mathbf{130.714}^{104.794}$
PODA	$\underline{1.067}^{1.194}$	$\mathbf{15.466}^{6.928}$	$\mathbf{1.685}^{4.432}$	$\mathbf{-27.651}$	$217.402^{177.238}$

agreement therefore persists across both formats; the benefit extends beyond native-sequence recovery alone. Additional DiffAb and MEAN initialization variants appear in Table 9. Appendix D.8 provides a complementary per-target DockQ comparison between PODA (base) and PODA.

Side-chain and interface quality. Table 3 evaluates all 50 VHH targets. PODA achieves the lowest interface scRMSD, highest polar-contact recovery, and most favorable median Rosetta interface energy among the compared methods. Post-training also reduces clashes relative to PODA (base), although the buried-area discrepancy increases. Evaluation details are provided in Appendix D.6.

Qualitative comparison. Figure 3 examines complex placement on two VH–VL and two VHH targets. Each complete complex is aligned to the reference using antigen C_α atoms, with the same rigid transformation applied to the antibody, so the comparison retains errors in relative placement. The selected VH–VL cases illustrate large improvements over the base model, whereas the VHH cases show refinement of already stronger base predictions. Within each format, the two cases with the largest PODA-minus-strongest-comparator DockQ margins were selected after evaluation; they illustrate selected favorable outcomes, with typical performance outside the scope of these examples.

4.3 Post-Training Comparisons and Ablations

Alternative post-training methods. Figure 4 compares PODA with DiffusionNFT (Zheng et al., 2025), Flow-GRPO (Liu et al., 2025), RAM (Bergmeister et al., 2026), DRTune (Wu et al., 2024), and TempFlow-GRPO (He et al., 2025) using 400 steps, one candidate per target, and a common evaluator. PODA has the highest mean DockQ and SR and the lowest mean iRMS and LRMS in both cohorts. Agreement across these metrics suggests that the advantage extends beyond moving predictions across the SR threshold. The post-training baselines follow their original papers and use the same training and testing settings as PODA, as specified in the comparison protocol. The comparison supports quality-gated endpoint supervision as an effective post-training strategy under these shared settings. Individual component effects are examined through the component ablations.

Training-stage and teacher ablations. Table 4 evaluates all 93 VH–VL targets. Removing Stage II recovers PODA (base); removing the teacher retains endpoint collection and re-noising but omits privileged-teacher supervision. Removing the teacher lowers both docking and global fold quality below PODA (base) levels. Thus, this recipe requires appropriate supervision alongside collection and re-noising of student outputs. By contrast, removing Stage II preserves slightly better CDR-H3 and global reconstruction scores but weaker docking. Together, these controls distinguish the role of supervised initialization in reconstruction from the contribution of teacher-guided post-training to interface placement.

Supervision targets and training states. Direct GT regresses native endpoints and yields stronger reconstruction but weaker docking than PODA, distinguishing direct supervision from teacher-guided

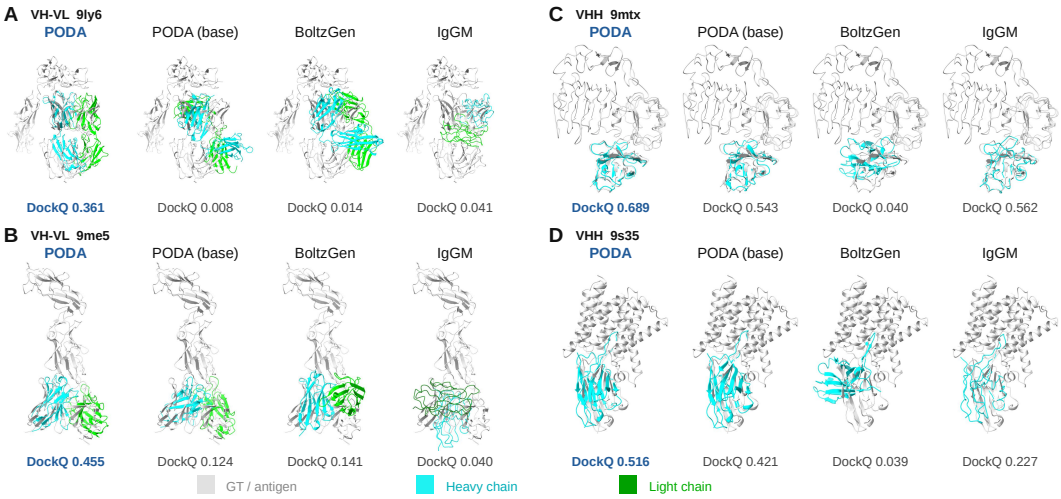

Figure 3: Structural comparisons on VH–VL targets and VHH targets.

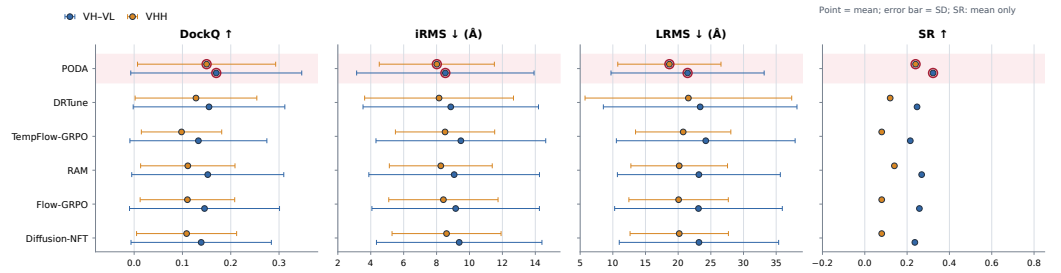

Figure 4: Comparison with post-training methods on the complete test cohort.

interface refinement. On VH–VL, applying the quality gate only to distillation or only to RAM yields DockQ of 0.166 and 0.167, respectively, compared with 0.165 without either gate and 0.170 with both. SR is 28.0% for the ungated and both single-gate variants, versus 32.3% for PODA. The complete gated objective therefore achieves higher DockQ and SR than either selective-gating control. Offline rollout retains the privileged teacher and gate but uses previously generated endpoints; its lower docking quality supports refreshing the endpoint distribution with the evolving student. Public-teacher and GT-path controls, together with VHH results, are reported in Appendix D.7.

Auxiliary RAM control. Removing RAM under the same training and evaluation settings weakens DockQ and SR, although docking remains better than for PODA (base). The auxiliary correction therefore contributes to the complete recipe alongside gains retained in the RAM-off control. Its removal slightly improves some reconstruction metrics.

4.4 Complex Structure Prediction

We next reveal the complete antibody sequence to examine complex prediction separately from CDR design. Table 5 groups the 93-antibody and 50-nanobody results into separate panels. Baselines include IgFold–HDOCK (Ruffolo et al., 2023; Yan et al., 2020), tFold-Ag (Wu et al., 2026), dyMEAN (Kong et al., 2023b), IgGM (Wang et al., 2025), Boltz-2, Protenix-v2, and AlphaFold 3 (Abramson et al., 2024), where available for each cohort. Their method-specific conditioning requirements are retained, and all predictions are assessed using the common evaluation protocol.

Antibody prediction. On VH–VL targets, PODA improves docking and antibody RMSD over PODA (base) even when CDR identities are supplied. The benefit therefore extends beyond sequence generation to the structural component of the task. PODA has the highest SR and lowest mean iRMS and LRMS in this comparison, while Protenix-v2 and AlphaFold 3 achieve higher mean DockQ and global fold scores. These different rankings underscore that successful-target coverage, average interface agreement, and fold accuracy capture distinct properties.

Table 4: Ablation study for de novo design on the 93-antibody test set. The no-gate variant gates neither loss; each gate-only variant retains both losses and gates only the named term. PODA gates both terms. Bold and underlined means rank all variants jointly.

Variant	AAR H3↑	RMSD H3↓	TM-Score↑	lDDT↑	DockQ↑	iRMS↓	LRMS↓	SR↑
PODA w/o Teacher	$0.496^{0.194}$	$2.831^{1.066}$	$0.743^{0.252}$	$0.613^{0.041}$	$0.114^{0.115}$	$9.606^{4.637}$	$24.012^{15.243}$	0.129
PODA w/o Stage II	$0.512^{0.192}$	$\underline{2.335}^{1.035}$	$\underline{0.929}^{0.046}$	$\underline{0.783}^{0.066}$	$0.141^{0.173}$	$8.993^{5.383}$	$23.825^{15.166}$	0.204
Direct GT	$\mathbf{0.514}^{0.191}$	$\mathbf{2.321}^{1.018}$	$\mathbf{0.930}^{0.045}$	$\mathbf{0.784}^{0.064}$	$0.159^{0.173}$	$8.812^{5.284}$	$22.916^{14.382}$	0.269
Privileged teacher (no gate)	$0.509^{0.193}$	$2.367^{1.006}$	$0.924^{0.050}$	$0.777^{0.068}$	$0.165^{0.176}$	$8.674^{5.321}$	$22.184^{12.946}$	$\underline{0.280}$
Gate on distillation only	$0.511^{0.189}$	$2.359^{1.017}$	$0.925^{0.049}$	$0.778^{0.071}$	$0.166^{0.174}$	$8.684^{5.278}$	$22.214^{12.893}$	$\underline{0.280}$
Gate on RAM only	$0.512^{0.194}$	$2.357^{1.004}$	$0.925^{0.051}$	$0.779^{0.067}$	$\underline{0.167}^{0.175}$	$8.673^{5.251}$	$\underline{22.174}^{12.922}$	$\underline{0.280}$
Offline rollout	$\underline{0.513}^{0.192}$	$2.339^{1.024}$	$\underline{0.929}^{0.046}$	$0.782^{0.066}$	$0.165^{0.175}$	$8.716^{5.298}$	$22.437^{13.684}$	$\underline{0.280}$
PODA w/o RAM	$0.512^{0.193}$	$2.374^{0.984}$	$0.927^{0.046}$	$0.780^{0.066}$	$0.163^{0.176}$	$\underline{8.587}^{5.337}$	$22.918^{15.026}$	0.269
PODA	$0.506^{0.196}$	$2.349^{0.965}$	$0.925^{0.049}$	$0.778^{0.066}$	$\mathbf{0.170}^{0.177}$	$\mathbf{8.533}^{5.390}$	$\mathbf{21.440}^{11.719}$	**0.323**

Table 5: Complex structure prediction with complete antibody sequences: 93 VH–VL antibodies (upper panel) and 50 VHH nanobodies (lower panel). †, ∗, and ‡ indicate antigen-structure input, epitope conditioning, and MSAs, respectively.

Structure prediction of Antibody							
Method	TM-Score↑	lDDT↑	RMSD H3↓	DockQ↑	iRMS↓	LRMS↓	SR↑
IgFold→ HDock†	$0.918^{0.119}$	$0.796^{0.047}$	$3.831^{1.094}$	$0.017^{0.031}$	$19.129^{9.861}$	$57.511^{23.135}$	0.000
tFold-Ag∗‡	$0.929^{0.047}$	$0.813^{0.070}$	$2.061^{0.868}$	$0.061^{0.042}$	$12.658^{3.248}$	$37.851^{21.680}$	0.042
Boltz-2‡	$0.934^{0.061}$	$0.843^{0.072}$	$2.009^{1.139}$	$0.078^{0.119}$	$13.918^{7.232}$	$39.319^{22.507}$	0.075
Protenix-v2‡	$\mathbf{0.945}^{0.046}$	$\mathbf{0.850}^{0.078}$	$1.813^{1.266}$	$\mathbf{0.219}^{0.300}$	$10.916^{8.566}$	$34.082^{29.078}$	0.269
AlphaFold 3‡	$\underline{0.944}^{0.047}$	$\underline{0.849}^{0.076}$	$\underline{1.801}^{1.371}$	$\underline{0.213}^{0.247}$	$11.032^{6.993}$	$32.466^{25.263}$	$\underline{0.280}$
dyMEAN†∗	$0.916^{0.044}$	$0.801^{0.062}$	$2.153^{0.852}$	$0.107^{0.092}$	$9.359^{3.860}$	$24.325^{9.057}$	0.118
IgGM†∗	$0.892^{0.125}$	$0.751^{0.047}$	$2.417^{1.025}$	$0.170^{0.164}$	$10.005^{5.732}$	$23.943^{16.170}$	0.258
PODA (base)†∗	$0.932^{0.046}$	$0.820^{0.064}$	$2.268^{1.073}$	$0.165^{0.158}$	$\underline{9.087}^{4.881}$	$\underline{22.961}^{11.717}$	0.269
PODA†∗	$0.928^{0.049}$	$0.815^{0.065}$	$\mathbf{1.628}^{0.933}$	$0.197^{0.193}$	$\mathbf{8.615}^{5.104}$	$\mathbf{21.504}^{11.495}$	**0.301**
Structure prediction of Nanobody							
Method	TM-Score↑	lDDT↑	RMSD H3↓	DockQ↑	iRMS↓	LRMS↓	SR↑
tFold-Ag∗‡	$0.855^{0.204}$	$0.740^{0.218}$	$2.297^{1.666}$	$0.052^{0.072}$	$13.782^{5.010}$	$41.721^{18.821}$	0.040
Boltz-2‡	$0.926^{0.054}$	$0.829^{0.076}$	$1.926^{1.372}$	$0.142^{0.201}$	$11.012^{6.132}$	$32.516^{20.307}$	0.220
Protenix-v2‡	$\mathbf{0.938}^{0.055}$	$\mathbf{0.848}^{0.084}$	$\mathbf{1.481}^{1.306}$	$\mathbf{0.258}^{0.312}$	$9.420^{7.017}$	$28.073^{23.016}$	**0.380**
AlphaFold 3‡	$\underline{0.935}^{0.055}$	$\underline{0.844}^{0.081}$	$\underline{1.665}^{1.359}$	$\underline{0.237}^{0.308}$	$10.456^{7.923}$	$32.563^{25.322}$	$\underline{0.340}$
IgGM†∗	$0.899^{0.058}$	$0.733^{0.012}$	$2.598^{1.204}$	$0.112^{0.094}$	$8.482^{3.233}$	$19.599^{8.776}$	0.100
PODA (base)†∗	$0.913^{0.047}$	$0.794^{0.060}$	$2.392^{1.277}$	$0.181^{0.136}$	$\mathbf{7.859}^{3.488}$	$\mathbf{17.532}^{6.859}$	0.240
PODA†∗	$0.910^{0.047}$	$0.794^{0.061}$	$1.968^{0.993}$	$0.189^{0.169}$	$\underline{8.029}^{3.994}$	$\underline{17.577}^{8.018}$	0.320

Nanobody prediction. On VHH targets, higher DockQ and SR coexist with essentially unchanged fold scores and slightly worse mean iRMS and LRMS relative to PODA (base). Post-training therefore improves the fraction of predictions satisfying the DockQ success criterion, while changes in structural deviations remain mixed. The prediction results support benefits of privileged post-training for the shared generator.

5 Conclusions

We presented PODA, a two-stage framework for epitope-conditioned antibody and nanobody design from an antigen, a specified epitope, and a CDR-masked framework sequence. Supervised student–teacher training establishes the base generator and a privileged teacher. Post-training freezes the teacher and updates the student on re-noised endpoints generated online by the student's evolving policy. Clean-endpoint matching, residue-level weighting, and teacher-quality gating transfer dense structural supervision while preserving the student's deployment inputs. The teacher is not used at inference. On the complete VH–VL and VHH benchmark cohorts, post-training improves DockQ and success rate relative to PODA (base). Teacher ablations and comparisons with alternative post-training methods further support this approach to improving interface geometry.

A limitation of this study is that the evaluation is computational, and the designed antibodies and nanobodies have not yet been validated in wet-lab experiments. Future work will experimentally evaluate their binding affinity, specificity, and developability to assess whether the observed structural improvements translate into functional binders.

AI USE STATEMENT

Generative AI tools were used to assist with language editing, manuscript organization. All scientific claims, equations, experimental configurations, numerical results, citations, and AI-assisted artifacts were reviewed by the authors against the underlying code and experiment records. The authors take responsibility for the complete content of the submission.

ETHICS STATEMENT

This work is computational and uses publicly available structural data; it does not involve human subjects or clinical experiments. Antibody-design models are dual-use technologies. PODA produces computational candidates and structural scores, not validated biological activity, and any downstream use requires appropriate institutional review, biosafety practices, and experimental validation. We report the model's information boundary and limitations to discourage interpreting structural agreement as evidence of efficacy or safety.

REPRODUCIBILITY STATEMENT

Section 3 specifies the model inputs, information boundary, and distillation objective. Section 4 reports the data split, inference protocol, metrics, and baseline contracts. The appendix provides dataset processing, architectural and optimization details, additional baseline comparisons, and evaluation protocols. The supplementary package will include the frozen configurations, target manifests, checkpoint identifiers, and table generation inputs needed to distinguish PODA (base) from PODA.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pp. 1–3, 2024.

Jared Adolf-Bryfogle, Oleks Kalyuzhniy, Michael Kubitz, Brian D Weitzner, Xiaozhen Hu, Yumiko 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.

Rahmad Akbar, Philippe A. Robert, Cédric R. Weber, Michael Widrich, Robert Frank, Milena Pavlović, Lonneke Scheffer, Maria Chernigovskaya, Igor Snapkov, Andrei Slabodkin, Brij Bhushan Mehta, Enkelejda Miho, Fridtjof Lund-Johansen, Jan Terje Andersen, Sepp Hochreiter, Ingrid Hobæk Haff, Günter Klambauer, Geir Kjetil Sandve, and Victor Greiff. In silico proof of principle of machine learning-based antibody design at unconstrained scale. *mAbs*, 14(1): 2031482, 2022.

Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023.

Sankar Basu and Björn Wallner. Dockq: a quality measure for protein-protein docking models. *PloS one*, 11(8):e0161879, 2016.

Nathaniel R. Bennett, Joseph L. Watson, Robert J. Ragotte, et al. Atomically accurate de novo design of antibodies with rfdiffusion. *Nature*, 649:183–193, 2026. doi: 10.1038/s41586-025-09721-5.

Andreas Bergmeister, Stefanie Jegelka, Nikolas Nüsken, Carles Domingo-Enrich, and Jakiw Pidstrigach. Reinforce adjoint matching: Scaling rl post-training of diffusion and flow-matching models, 2026.

Can Chen, Karla-Luise Herpoldt, Chenchao Zhao, Zichen Wang, Marcus D. Collins, Shang Shang, and Ron Benson. Affinityflow: Guided flows for antibody affinity maturation. In *International Conference on Machine Learning*, 2025.

Alexander E Chu, Jinho Kim, Lucy Cheng, Gina El Nesr, Minkai Xu, Richard W Shuai, and Po-Ssu Huang. An all-atom protein generative model. *Proceedings of the National Academy of Sciences*, 121(27):e2311500121, 2024.

Kieran Didi, Zuobai Zhang, Guoqing Zhou, Danny Reidenbach, Zhonglin Cao, Sooyoung Cha, Tomas Geffner, Christian Dallago, Jian Tang, Michael M. Bronstein, Martin Steinegger, Emine Kucukbenli, Arash Vahdat, and Karsten Kreis. Scaling atomistic protein binder design with generative pretraining and test-time compute. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=qmCpJtFZra.

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.

Zhangyang Gao, Cheng Tan, and Stan Z. Li. Pifold: Toward effective and efficient protein inverse folding. In *International Conference on Learning Representations*, 2023.

Xiaoxuan He, Siming Fu, Yuke Zhao, Wanli Li, Jian Yang, Dacheng Yin, Fengyun Rao, and Bo Zhang. Tempflow-grpo: When timing matters for grpo in flow models, 2025.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. In *International Conference on Machine Learning*, pp. 8946–8970. PMLR, 2022.

John B Ingraham, Max Baranov, Zak Costello, Karl W Barber, Wujie Wang, Ahmed Ismail, Vincent Frappier, Dana M Lord, Christopher Ng-Thow-Hing, Erik R Van Vlack, et al. Illuminating protein space with a programmable generative model. *Nature*, 623(7989):1070–1078, 2023.

Wengong Jin, Jeremy Wohlwend, Regina Barzilay, and Tommi Jaakkola. Iterative refinement graph neural network for antibody sequence-structure co-design. *arXiv preprint arXiv:2110.04624*, 2021.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. Conditional antibody design as 3d equivariant graph translation. In *The Eleventh International Conference on Learning Representations*, 2023a.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. End-to-end full-atom antibody design. In *International Conference on Machine Learning*, pp. 17409–17429. PMLR, 2023b.

Gideon D Lapidoth, Dror Baran, Gabriele M Pszolla, Christoffer Norn, Assaf Alon, Michael D Tyka, and Sarel J Fleishman. AbDesign: An algorithm for combinatorial backbone design guided by natural conformations and sequences. *Proteins: Structure, Function, and Bioinformatics*, 83(8): 1385–1406, 2015.

Marie-Paule Lefranc, Christelle Pommié, Manuel Ruiz, Véronique Giudicelli, Elodie Foulquier, Lisa Truong, Valérie Thouvenin-Contet, and Gérard Lefranc. Imgt unique numbering for immunoglobulin and t cell receptor variable domains and ig superfamily v-like domains. *Developmental & Comparative Immunology*, 27(1):55–77, 2003.

Marie-Paule Lefranc, Veronique Giudicelli, Chantal Ginestoux, Joumana Jabado-Michaloud, Geraldine Folch, Fatena Bellahcene, Yan Wu, Elodie Gemrot, Xavier Brochet, Jeroˆme Lane, et al. Imgt®, the international immunogenetics information system®. *Nucleic acids research*, 37 (suppl_1):D1006–D1012, 2009.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

Jie Liu, Gongye Liu, Jiajun Liang, Yangguang Li, Jiaheng Liu, Xintao Wang, Pengfei Wan, Di Zhang, and Wanli Ouyang. Flow-grpo: Training flow matching models via online rl. *arXiv preprint arXiv:2505.05470*, 2025.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. 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. *Advances in Neural Information Processing Systems*, 35:9754–9767, 2022.

Isaac D Lutz, Shunzhi Wang, Christoffer Norn, Alexis Courbet, Andrew J Borst, Yan Ting Zhao, Annie Dosey, Longxing Cao, Jinwei Xu, Elizabeth M Leaf, et al. Top-down design of protein architectures with reinforcement learning. *Science*, 380(6642):266–273, 2023.

Luis Santiago Mille-Fragoso, John N. Wang, Claudia L. Driscoll, Haoyu Dai, Talal M. Widatalla, Xiaowe Zhang, Brian L. Hie, and Xiaojing J. Gao. Efficient generation of epitope-targeted de novo antibodies with germinal. *bioRxiv*, 2025. URL https://www.biorxiv.org/content/10.1101/2025.09.19.677421v1.

Saul B Needleman and Christian D Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3):443–453, 1970.

Martin Pacesa, Lennart Nickel, Christian Schellhaas, Joseph Schmidt, Ekaterina Pyatova, Lucas Kissling, Patrick Barendse, et al. One-shot design of functional protein binders with BindCraft. *Nature*, 2025.

Robert J. Pantazes and Costas D. Maranas. OptCDR: a general computational method for the design of antibody complementarity determining regions for targeted epitope binding. *Protein Engineering, Design and Selection*, 23(11):849–858, 2010.

Jeffrey A. Ruffolo, Lee-Shin Chu, Sai Pooja Mahajan, and Jeffrey J. Gray. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nature Communications*, 14(1):2389, 2023.

Koichiro Saka, Taro Kakuzaki, Shoichi Metsugi, Daiki Kashiwagi, Kenji Yoshida, Manabu Wada, Hiroyuki Tsunoda, and Reiji Teramoto. Antibody design using lstm based deep generative model from phage display library for affinity maturation. *Scientific reports*, 11(1):5852, 2021.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Hannes Stark, Felix Faltings, MinGyu Choi, Yuxin Xie, Eunsu Hur, Timothy John O'Donnell, Anton Bushuiev, Talip Uçar, Saro Passaro, Weian Mao, Mateo Reveiz, Roman Bushuiev, Tomáš Pluskal, Josef Sivic, Karsten Kreis, Arash Vahdat, Shamayeeta Ray, Jonathan T. Goldstein, Andrew Savinov, Jacob A. Hambalek, Anshika Gupta, Diego A. Taquiri-Diaz, Yaotian Zhang, A. Katherine Hatstat, Angelika Arada, Nam Hyeong Kim, Ethel Tackie-Yarboi, Dylan Boselli, Lee Schnaider, Chang C. Liu, Gene-Wei Li, Denes Hnisz, David M. Sabatini, William F. DeGrado, Jeremy Wohlwend, Gabriele Corso, Regina Barzilay, and Tommi Jaakkola. Boltzgen: Toward universal binder design. *bioRxiv*, 2025. doi: 10.1101/2025.11.20.689494.

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.

Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. Diffusion model alignment using direct preference optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

Rubo Wang, Fandi Wu, Xingyu Gao, Jiaxiang Wu, Peilin Zhao, and Jianhua Yao. IgGM: A generative model for functional antibody and nanobody design. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=zmmfsJpYcq.

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 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2026.

Shira Warszawski, Alon Borenstein Katz, Rosalie Lipsh, Lev Khmelnitsky, Gili Ben Nissan, Gali Javitt, Orly Dym, Tamar Unger, Orly Knop, Shira Albeck, et al. Optimizing antibody affinity and stability by the automated design of the variable light-heavy chain interfaces. *PLOS Computational Biology*, 15(8):e1007207, 2019.

Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

Fandi Wu, Yu Zhao, Jiaxiang Wu, et al. De novo design of epitope-specific antibodies via a structure-driven computational workflow. *Nature Communications*, 17:625, 2026. doi: 10.1038/s41467-025-67361-9.

Jun Wu, Xiangzhe Kong, Ningguan Sun, Jing Wei, Sisi Shan, Fuli Feng, Feng Wu, Jian Peng, Linqi Zhang, Yang Liu, and Jianzhu Ma. Flowdesign: Improved design of antibody cdrs through flow matching and better prior distributions. *Cell Systems*, 16, 2025.

Xiaoshi Wu, Yiming Hao, Manyuan Zhang, Keqiang Sun, Zhaoyang Huang, Guanglu Song, Yu Liu, and Hongsheng Li. Deep reward supervisions for tuning text-to-image diffusion models. In *European Conference on Computer Vision*, 2024.

Yumeng Yan, Huanyu Tao, Jiahua He, and Sheng-You Huang. The HDOCK server for integrated protein–protein docking. *Nature Protocols*, 15(5):1829–1852, 2020.

Zhuo Yang, Jiaying He, Jiaqing Xie, Daolang Wang, Xipeng Qiu, Yuxin Wang, Tianfan Fu, and Beilun Wang. Abopd: Antibody cdr design via on-policy distillation. In *arXiv preprint arXiv:2607.18835*, 2026.

Kaiwen Zheng, Huayu Chen, Haotian Ye, Haoxiang Wang, Qinsheng Zhang, Kai Jiang, Hang Su, Stefano Ermon, Jun Zhu, and Ming-Yu Liu. Diffusionnft: Online diffusion reinforcement with forward process. *arXiv preprint arXiv:2509.16117*, 2025.

Xiangxin Zhou, Dongyu Xue, Ruizhe Chen, Zaixiang Zheng, Liang Wang, and Quanquan Gu. Antigen-specific antibody design via direct energy-based preference optimization. In *Advances in Neural Information Processing Systems*, 2024.

Tian Zhu, Milong Ren, and Haicang Zhang. Antibody design using a score-based diffusion model guided by evolutionary, physical and geometric constraints. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=1YsQI04KaN.

A CODE

The source code for PODA, including model architecture, training scripts, and evaluation pipelines, is available at:

`https://anonymous.4open.science/r/PODA-F8C5`

B DATASET CURATION AND FILTERING

Source and temporal split. All antibody-antigen complexes are sourced from SAbDab (Dunbar et al., 2014). Structures deposited on or before 2025-06-30 form the train/validation pool (12,616 raw entries); structures deposited strictly after this date form the candidate test pool. The temporal cutoff ensures that no test structure could have been seen during training.

Preprocessing. Each complex undergoes a six-step preprocessing pipeline: (1) multi-antigen complexes are split into one sample per antigen chain; (2) FASTA files are generated for all chains; (3) merged H+L chains are detected and split using sequence anchoring; (4) PDB and FASTA sequences are reconciled (PDB is ground truth); (5) train/valid/test split indices are written; (6) all samples are validated for chain completeness and sequence consistency. Samples failing any validation check are excluded.

Sequence-based redundancy filtering for the test set. To reduce sequence redundancy with the training set, candidate test samples are filtered against the full training set. For each candidate, both the antigen sequence(s) and the antibody sequence(s) are compared against every corresponding sequence in the training pool using global pairwise alignment (Needleman & Wunsch, 1970) with match score $+2$, mismatch score -1, gap open -0.5, and gap extend -0.1. The normalized alignment score is computed as

$$s_{\text{norm}} = \frac{\text{alignment score}}{\max_\text{len} \times 2} \times 100\%, \tag{17}$$

where $\max_\text{len}$ is the length of the longer sequence and 2 is the per-residue match score. This score includes mismatch and gap penalties and is distinct from sequence identity, which measures the fraction of matching residues in an alignment. A candidate is retained only if *all* of its antigen sequences and *all* of its antibody sequences have $s_{\text{norm}} \leq 80\%$ with every training sequence of the same type. Candidates exceeding this normalized-score threshold on any chain are excluded; the threshold does not imply sequence identity $\leq 80\%$. After filtering, the final test set contains 93 conventional antibodies and 50 nanobodies (143 total).

C METHOD DETAILS

We use the joint coordinate–latent state $\mathbf{u}$ defined in Section 3; $\mathbf{u}^*$ denotes the encoded ground-truth endpoint, whereas $\mathbf{u}_1^{roll}$ denotes a detached student-generated endpoint. These play different roles in supervised training and post-training.

C.1 PARTIAL AUTOENCODER AND BASE GENERATOR

The frozen partial autoencoder maps each antibody residue to an explicit C_α coordinate and an 8-dimensional latent. Its decoder reconstructs amino-acid identity and full-atom geometry. The base flow network contains 14 transformer layers, token width 384, pair width 256, and 12 attention heads. It is shared across VH–VL antibodies and VHH nanobodies. The Stage-I public-input student defines PODA (base) and initializes the Stage-II student. Both stages retain this sequence–structure representation.

The public-view student and privileged-view teacher are first optimized in parallel with supervised objectives on the same training complexes. The teacher receives the additional complete-complex fields described in Section 3. Its parameters are frozen before online OPD begins; only the student is updated during the second stage.

For the linear path

$$\mathbf{u}_t = (1-t)\mathbf{u}_0 + t\mathbf{u}^*, \tag{18}$$

a predicted velocity implies the endpoint

$$\widehat{\mathbf{u}}_1 = \mathbf{u}_t + (1-t)\mathbf{v}_\theta(\mathbf{u}_t, t). \tag{19}$$

This identity is used both by intermediate geometric recycling in PODA (base) and by endpoint-space distillation in PODA.

C.2 Online Endpoint Collection and Re-noising

The rollout collector uses the current public-input student policy. It receives the antigen, specified epitope, and masked framework sequence and keeps the antigen coordinates fixed throughout integration. Terminal rollout endpoints are detached and re-noised to sampled times using fresh noise, as in Eq. equation 8. These interpolation states are constructed from generated endpoints, not extracted from the ODE trajectory. The same resulting $(\widetilde{\mathbf{u}}_t, t)$ is passed to the student and teacher; only the teacher call receives complete-complex training fields. After the student parameters are updated, later trajectories are sampled from the updated student policy. Thus, "on-policy" refers to online state collection from the evolving deployable policy; gradients are not propagated through the rollout sampler.

C.3 Endpoint Distillation and Gating

For coordinate and latent modes q, OPD uses

$$\mathcal{L}_{\mathrm{OPD}} = g \sum_q \frac{1}{\sum_i w_i^q} \sum_i w_i^q \left\| \widehat{\mathbf{u}}^q_{1,S,i} - \mathrm{sg}(\widehat{\mathbf{u}}^q_{1,T,i}) \right\|_2^2. \tag{20}$$

CDR and interface residues receive larger w_i^q so that the designed region is not overwhelmed by framework residues. The gate is evaluated only from training targets:

$$g = \mathbf{1}[e_T \le (1-m)e_S]\, \mathbf{1}[\mathrm{relL2}(\widehat{\mathbf{u}}_{1,T}) \le \tau]. \tag{21}$$

The first term requires a relative improvement of at least m (or no worse error when $m = 0$); the second bounds the relative-error diagnostic. The gate and teacher targets are detached. Endpoint matching at a common state obeys

$$\|\widehat{\mathbf{u}}_{1,S} - \widehat{\mathbf{u}}_{1,T}\|_W^2 = (1-t)^2 \|\mathbf{v}_S - \mathbf{v}_T\|_W^2. \tag{22}$$

Thus the endpoint objective includes an implicit time weighting. Re-noised endpoint states need not equal intermediate states from the ODE trajectory.

C.4 Detached RAM Correction

The auxiliary RAM term uses a frozen public reference velocity and a clipped scalar structural score. The entire target in Section 3.3 is detached, including the student velocity appearing inside the correction. Consequently the loss differentiates only through the student prediction being regressed, not through the rollout, teacher, reference model, or score. The same teacher-quality gate masks this auxiliary loss.

C.5 Information-Flow Audit

The student's conditioning view contains the antigen, specified epitope, masked framework sequence, and public chain/CDR masks. Complete antibody information is used for supervised paths, teacher conditioning, and training losses. It is excluded from deployment conditioning, where neither the privileged encoder nor teacher is required. In particular, ground-truth training endpoints should not be confused with additional inference inputs. VH–VL antibodies and VHH nanobodies use the same coordinate–latent representation, with chain and CDR masks specifying their layouts. The reported post-training model is the PODA checkpoint, evaluated separately on the two cohorts.

C.6 Base Training Objective

The Stage-I supervised objective, denoted $\mathcal{L}_{\text{sup}}$ in the main text, is

$$\mathcal{L}_{\text{sup}} = \mathcal{L}_{\text{FM}} + \omega(t) \sum_i \lambda_i \mathcal{L}_i + \lambda_c \mathcal{L}_{\text{contact}}, \tag{23}$$

where the auxiliary terms cover local geometry, steric validity, coordinate alignment, and interface contacts. During OPD this objective is retained only as a small anchor that limits drift from the pretrained generator. Exact weights and the frozen release configuration are supplied with the reproducibility artifacts.

C.7 Distinctions from Prior Privileged Distillation

PODA combines privileged supervision with a specific design objective: improving the joint generation of CDR sequences and antibody–antigen geometry under a fixed deployment interface. Its contribution lies in how the generated state distribution, teacher targets, and structural selection criterion are coupled to this objective. The following distinctions connect these design choices to the empirical advantages observed in our benchmark.

From CDR refinement to complete-complex generation. ABOPD (Yang et al., 2026) studies CDR sequence and structure generation conditioned on an antigen and non-target antibody residues, and distills C_α coordinate transitions on student denoising trajectories. PODA generates CDR identities and the complete antibody structure in the antigen coordinate frame, conditioned on the antigen, epitope, and masked framework sequence. Antibody coordinates and the native binding pose are not deployment inputs. Consequently, the student must resolve both local CDR geometry and the antibody's placement relative to the antigen. A shared coordinate–latent representation supports this task for both VH–VL and VHH. DockQ, iRMS, LRMS, and SR directly assess the resulting interface geometry, complementing local sequence and loop-recovery metrics.

Online endpoint distributions and joint endpoint targets. ABOPD supervises states visited along the reverse diffusion trajectory. In PODA, an online rollout first produces a detached endpoint; fresh noise and a sampled time then construct the state used for distillation. This separates the distillation time from the rollout solver's visited states while retaining dependence on the current student's output distribution. Updating the collector as the student changes refreshes the generated endpoints used for supervision. The teacher and student receive the same re-noised state, and matching acts on clean endpoints in both coordinate and sequence–structure latent modes. Thus, teacher feedback addresses the joint generated representation, with residue weights emphasizing CDR and interface geometry. The endpoint formulation also makes the quality comparison against native training endpoints explicit; its equivalence to velocity matching with a $(1-t)^2$ weight is derived above.

Quality selection makes privileged feedback conditional on usefulness. Privileged information does not guarantee that a teacher prediction is better at every student-induced state. PODA admits a teacher target only when it satisfies both the relative-improvement criterion against the student and the relative-error bound. This gives the training procedure an explicit rule for rejecting unsuitable teacher feedback. Direct GT instead regresses the native endpoint, while an ungated teacher supplies its prediction without this selection. These choices are empirically distinct in Table 12. The implemented gate also masks the RAM correction. Selective-gating controls in Table 4 separate gating the distillation loss from gating RAM, while retaining both losses. The supervised anchor remains active independently of this selection.

Empirical advantages over simpler supervision and state choices. The experiments use baselines implemented according to their original methods under the shared training and testing settings in Section 4.1. The following comparisons support specific benefits of the complete recipe; all SR differences below are percentage points computed from the displayed table values.

- **Beyond additional ground-truth regression.** Relative to Direct GT, PODA increases DockQ from 0.159 to 0.170 on VH–VL and from 0.130 to 0.150 on VHH, with SR gains of 5.4 and 8.0 points, respectively (Table 12). Teacher-guided post-training therefore achieves better docking than directly regressing native endpoints on the generated states.

Table 6: Autoencoder reconstruction on 143 held-out complexes: AFDB-pretrained La-Proteina versus the antibody-finetuned autoencoder.

AE checkpoint	CDR AAR↑	Full-Ab AAR↑	CDR scRMSD↓	H3 scRMSD↓	Ab scRMSD↓	Atom coverage↑
La-Proteina (AFDB pretrained)	$0.992^{0.039}$	$0.989^{0.044}$	$0.300^{0.125}$	$0.308^{0.213}$	$0.347^{0.094}$	99.960%
Antibody-finetuned	$\mathbf{1.000}^{0.000}$	$\mathbf{1.000}^{0.000}$	$\mathbf{0.187}^{0.064}$	$\mathbf{0.188}^{0.091}$	$\mathbf{0.180}^{0.045}$	**100.000%**

Table 7: Effect of integration steps on PODA for 93 VH–VL targets.

Steps	AAR H3↑	RMSD H3↓	TM-Score↑	lDDT↑	DockQ↑	iRMS↓	LRMS↓	SR↑
20	$0.492^{0.199}$	$6.766^{1.114}$	$0.269^{0.037}$	$0.156^{0.035}$	$0.041^{0.031}$	$13.472^{3.088}$	$33.691^{14.290}$	0.000
50	$0.500^{0.194}$	$3.738^{0.636}$	$0.535^{0.077}$	$0.301^{0.033}$	$0.088^{0.086}$	$10.955^{4.232}$	$26.957^{13.080}$	0.097
100	$0.512^{0.191}$	$2.512^{0.891}$	$0.859^{0.041}$	$0.571^{0.033}$	$0.132^{0.146}$	$9.139^{4.708}$	$23.785^{12.714}$	0.215
200	$0.511^{0.195}$	$2.381^{1.058}$	$0.926^{0.047}$	$0.779^{0.066}$	$0.148^{0.173}$	$9.172^{5.123}$	$23.375^{13.758}$	0.204
400	$0.506^{0.196}$	$2.349^{0.965}$	$0.925^{0.049}$	$0.778^{0.066}$	$0.170^{0.177}$	$8.533^{5.390}$	$21.440^{11.719}$	0.323
800	$0.520^{0.193}$	$2.313^{1.015}$	$0.926^{0.047}$	$0.780^{0.066}$	$0.169^{0.184}$	$9.033^{5.440}$	$21.206^{14.261}$	0.301
1000	$0.521^{0.192}$	$2.326^{1.002}$	$0.926^{0.047}$	$0.780^{0.066}$	$0.171^{0.186}$	$8.962^{5.466}$	$21.160^{14.318}$	0.323

- **Value of privileged context and quality selection.** Replacing the public teacher with the ungated privileged teacher raises DockQ from 0.153 to 0.165 on VH–VL and from 0.118 to 0.136 on VHH. The complete gated method further raises these values to 0.170 and 0.150, with additional SR gains of 4.3 and 6.0 points (Table 12). This ordering supports both richer training context and quality selection within the full objective.
- **Value of gating both loss terms.** On VH–VL, gating only distillation yields DockQ 0.166 and gating only RAM yields 0.167, compared with 0.170 for PODA. Both selective-gating controls have SR 28.0%, compared with 32.3% for the full method (Table 4). Holding either gate active, enabling the other improves both reported docking measures. The observed gain supports quality selection across the two loss terms in the complete objective.
- **Value of refreshing generated endpoints.** With the privileged teacher and gate fixed, online collection improves DockQ over offline rollout from 0.165 to 0.170 on VH–VL and from 0.134 to 0.150 on VHH, while SR rises from 28.0% to 32.3% and from 16.0% to 24.0%, respectively (Table 13). These results support updating the supervision-state distribution with the evolving student.
- **Gains retained without the auxiliary RAM correction.** On VH–VL, the RAM-off variant still improves DockQ over PODA (base) from 0.141 to 0.163 and SR from 20.4% to 26.9%; the full method reaches 0.170 and 32.3% (Table 4). The observed improvement therefore extends beyond the auxiliary reward correction.

Deployment advantage and scope of the evidence. The resulting gains require no additional inference inputs, teacher calls, or candidate search: PODA (base) and PODA use the same 400-step integration budget and one candidate per target. Across five inference seeds, mean DockQ rises from 0.141 to 0.170 on VH–VL and from 0.114 to 0.150 on VHH; SR rises from 20.4% to 32.3% and from 12.0% to 24.0%, respectively (Table 14). This demonstrates the practical value of transferring training-only structural information into the deployable generator. The comparisons establish observed docking advantages in the evaluated setting; the methodological comparison with ABOPD concerns task and objective differences, since the two methods have not been evaluated head-to-head under a common protocol here.

D Additional Experiments

D.1 Evaluation Metrics

We report amino-acid recovery (AAR) for CDR sequence reconstruction; CDR C_α RMSD for local geometry; TM-score and lDDT for global fold agreement; and DockQ (Basu & Wallner, 2016), iRMS, LRMS, and success rate (SR) for antibody–antigen placement. Success is defined as DockQ> 0.23. CDR-H3 RMSD is measured after H3 self-alignment and therefore does not measure antibody–

Table 8: End-to-end sampling cost on a single NVIDIA H20 GPU with batch size 4. Time is reported in seconds per sample.

ODE steps	20	50	100	200	400
Seconds/sample↓	1.963	3.813	6.531	11.346	21.546

Table 9: Additional design comparisons on 93 VH–VL targets. DiffAb and MEAN use IgFold- or AlphaFold 3-initialized framework structures; PODA (base) and PODA use framework-sequence conditioning.

Model		DiffAb (IgFold)	MEAN (IgFold)	DiffAb (AF3)	MEAN (AF3)	PODA (base)	**PODA**
AAR↑	L1	$0.500^{0.247}$	–	$0.462^{0.249}$	–	$\mathbf{0.693}^{0.196}$	$\underline{0.692}^{0.203}$
	L2	$0.489^{0.301}$	–	$0.430^{0.323}$	–	$\underline{0.821}^{0.252}$	$\mathbf{0.824}^{0.261}$
	L3	$0.406^{0.152}$	–	$0.369^{0.177}$	–	$\mathbf{0.727}^{0.246}$	$\underline{0.723}^{0.245}$
	H1	$0.625^{0.168}$	–	$0.513^{0.177}$	–	$\underline{0.696}^{0.194}$	$\mathbf{0.702}^{0.193}$
	H2	$0.271^{0.142}$	–	$0.255^{0.136}$	–	$\mathbf{0.667}^{0.226}$	$\underline{0.658}^{0.238}$
	H3	$0.256^{0.107}$	$0.356^{0.123}$	$0.253^{0.111}$	$0.343^{0.128}$	$\mathbf{0.512}^{0.192}$	$\underline{0.506}^{0.196}$
RMSD↓	L1	$1.005^{0.922}$	–	$1.245^{1.032}$	–	$\mathbf{0.629}^{0.665}$	$\mathbf{0.629}^{0.660}$
	L2	$0.192^{0.179}$	–	$0.213^{0.180}$	–	$\underline{0.138}^{0.177}$	$\mathbf{0.135}^{0.190}$
	L3	$0.892^{0.873}$	–	$1.038^{0.842}$	–	$\mathbf{0.728}^{0.753}$	$\underline{0.742}^{0.764}$
	H1	$0.900^{0.749}$	–	$1.664^{2.724}$	–	$\underline{0.772}^{0.571}$	$\mathbf{0.758}^{0.559}$
	H2	$0.705^{0.604}$	–	$1.129^{1.798}$	–	$\mathbf{0.538}^{0.483}$	$\underline{0.555}^{0.491}$
	H3	$2.667^{1.245}$	$\mathbf{2.262}^{0.894}$	$2.951^{2.700}$	$2.657^{0.910}$	$\underline{2.335}^{1.035}$	$2.349^{0.965}$
Structure↑	TMscore	$0.910^{0.112}$	$0.914^{0.113}$	$0.867^{0.099}$	$0.851^{0.096}$	$\mathbf{0.929}^{0.046}$	$\underline{0.925}^{0.049}$
	lDDT	$0.337^{0.044}$	$0.342^{0.044}$	$0.641^{0.097}$	$0.637^{0.097}$	$\mathbf{0.783}^{0.066}$	$\underline{0.778}^{0.066}$
Side chain↓	Clash rate	$\underline{0.002}^{0.002}$	$\mathbf{0.002}^{0.002}$	$0.018^{0.020}$	$0.003^{0.007}$	$0.018^{0.017}$	$0.013^{0.014}$
	scRMSD	$\underline{13.320}^{8.284}$	$17.572^{9.235}$	$17.148^{9.724}$	$19.907^{7.626}$	$14.227^{7.308}$	$\mathbf{12.294}^{5.939}$
Interface	Polar contact↑	$0.000^{0.000}$	$0.000^{0.000}$	$0.002^{0.017}$	$0.000^{0.000}$	$\underline{0.014}^{0.038}$	$\mathbf{0.024}^{0.065}$
	$\lvert\Delta\mathrm{BSA}\rvert$↓	$645.1^{439.3}$	$639.5^{432.6}$	$705.2^{645.6}$	$762.5^{768.3}$	$\underline{248.9}^{151.3}$	$\mathbf{210.7}^{142.6}$
	Median ΔG↓	$\mathbf{-53.1}$	-39.1	-44.1	-43.8	-32.2	$\underline{-48.3}$
Docking	DockQ↑	$0.011^{0.008}$	$0.014^{0.014}$	$0.019^{0.017}$	$0.019^{0.017}$	$\underline{0.141}^{0.173}$	$\mathbf{0.170}^{0.177}$
	iRMS↓	$21.055^{9.076}$	$19.894^{9.811}$	$16.881^{5.240}$	$16.409^{5.238}$	$\underline{8.993}^{5.383}$	$\mathbf{8.533}^{5.390}$
	LRMS↓	$61.100^{20.423}$	$60.930^{22.769}$	$50.232^{19.456}$	$50.819^{19.624}$	$\underline{23.825}^{15.166}$	$\mathbf{21.440}^{11.719}$
	SR↑	0.000	0.000	0.000	0.000	$\underline{0.204}$	**0.323**

antigen placement. Native references are used for evaluation only after generation and native-free ranking are complete.

Unless otherwise specified, entries are mean^{std} across targets with population SD; SR is reported as a mean only. Bold and underlined means mark the best and second-best available performance within each table or cohort. RMSD, iRMS, LRMS, and scRMSD are in Å. Interface scRMSD uses alignment on common antibody backbone atoms; $|\Delta\mathrm{BSA}|$ measures the absolute generated–native buried-surface-area difference in Å^2. Interface ΔG is a median in Rosetta energy units (REU), not a binding free-energy estimate. Clash rate and polar-contact recovery are fractions in the VH–VL tables and percentages in the VHH interface table.

D.2 Autoencoder Reconstruction Quality

The frozen partial autoencoder defines the coordinate–latent representation shared by PODA (base) and PODA. On the same 143 held-out complexes, antibody finetuning improves side-chain reconstruction in both the CDRs and the full antibody relative to the released AFDB-pretrained La-Proteina autoencoder (Table 6), while retaining complete atom coverage and sequence recovery at the reported precision. This supports its use as an antibody-specific representation, but does not establish generative quality: reconstruction supplies the decoder with an encoding of the reference structure, whereas generation must predict that representation.

D.3 ODE Integration Budget

Table 7 evaluates a fixed PODA model. Low budgets impair both fold reconstruction and docking. Fold scores largely stabilize by 200 steps, while docking continues to improve up to 400 steps. This

Table 10: Protenix-v2 refolding confidence and native-reference docking quality for designed sequences on 93 VH–VL targets. DockQ compares each sequence-refolded complex with its native reference complex. Entries show means with standard deviations in superscript.

Method	Global ipTM↑	pLDDT↑	pTM↑	ipSAE↑	pDockQ↑	Ranking score↑	DockQ↑
DiffAb (IgFold)	$0.516^{0.135}$	$\mathbf{83.211}^{5.100}$	$0.665^{0.097}$	$0.025^{0.073}$	$0.119^{0.062}$	$\underline{0.546}^{0.125}$	$0.046^{0.042}$
MEAN (IgFold)	$0.496^{0.140}$	$81.988^{5.138}$	$0.660^{0.092}$	$0.023^{0.061}$	$0.125^{0.071}$	$0.529^{0.128}$	$0.047^{0.043}$
DiffAb (AF3)	$0.507^{0.142}$	$81.680^{5.919}$	$0.659^{0.104}$	$0.030^{0.079}$	$0.119^{0.054}$	$0.538^{0.132}$	$0.047^{0.043}$
MEAN (AF3)	$0.494^{0.136}$	$81.759^{4.918}$	$0.661^{0.090}$	$0.014^{0.034}$	$0.125^{0.064}$	$0.528^{0.123}$	$0.047^{0.042}$
RFantibody (IgFold)	$0.491^{0.133}$	$82.508^{5.534}$	$0.657^{0.085}$	$0.017^{0.060}$	$0.133^{0.074}$	$0.524^{0.120}$	$0.048^{0.054}$
dyMEAN	$0.504^{0.135}$	$\underline{82.978}^{6.023}$	$0.661^{0.096}$	$0.019^{0.043}$	$0.110^{0.049}$	$0.535^{0.125}$	$0.058^{0.062}$
IgGM	$0.505^{0.133}$	$82.829^{5.797}$	$0.657^{0.097}$	$0.020^{0.056}$	$0.115^{0.058}$	$0.535^{0.124}$	$0.063^{0.063}$
BoltzGen	$0.511^{0.147}$	$82.670^{5.526}$	$0.666^{0.100}$	$0.031^{0.104}$	$\underline{0.134}^{0.078}$	$0.542^{0.135}$	$0.052^{0.054}$
Germinal	$0.487^{0.131}$	$80.362^{5.650}$	$0.656^{0.088}$	$0.022^{0.071}$	$0.120^{0.051}$	$0.521^{0.119}$	$0.068^{0.069}$
PODA (base)	$\underline{0.521}^{0.139}$	$82.272^{5.853}$	$\underline{0.672}^{0.096}$	$\underline{0.035}^{0.076}$	$0.125^{0.056}$	$\underline{0.551}^{0.128}$	$\underline{0.088}^{0.090}$
PODA	$\mathbf{0.526}^{0.158}$	$82.711^{8.253}$	$\mathbf{0.699}^{0.102}$	$\mathbf{0.041}^{0.114}$	$\mathbf{0.180}^{0.088}$	$\mathbf{0.553}^{0.142}$	$\mathbf{0.160}^{0.142}$

Table 11: Protenix-v2 refolding confidence and native-reference docking quality for designed sequences on 50 VHH targets, using the same evaluation protocol as the VH–VL comparison. DockQ compares each sequence-refolded complex with its native reference complex. Entries show means with standard deviations in superscript.

Method	Global ipTM↑	pLDDT↑	pTM↑	ipSAE↑	pDockQ↑	Ranking score↑	DockQ↑
DiffAb (AF3)	$0.349^{0.165}$	$79.179^{8.006}$	$0.697^{0.107}$	$0.054^{0.143}$	$\underline{0.194}^{0.101}$	$0.418^{0.147}$	$0.057^{0.038}$
IgGM	$0.353^{0.182}$	$80.052^{6.868}$	$0.689^{0.102}$	$\underline{0.061}^{0.151}$	$0.181^{0.099}$	$0.421^{0.160}$	$0.074^{0.058}$
BoltzGen	$0.347^{0.152}$	$\underline{82.795}^{6.891}$	$0.700^{0.096}$	$0.043^{0.121}$	$0.174^{0.083}$	$0.418^{0.134}$	$0.082^{0.063}$
Germinal	$0.330^{0.133}$	$78.349^{7.575}$	$0.691^{0.103}$	$0.029^{0.120}$	$0.145^{0.091}$	$0.402^{0.117}$	$0.078^{0.059}$
PODA (base)	$\underline{0.370}^{0.163}$	$\mathbf{82.865}^{6.870}$	$\underline{0.710}^{0.094}$	$0.057^{0.153}$	$0.189^{0.093}$	$\underline{0.438}^{0.143}$	$\underline{0.094}^{0.068}$
PODA	$\mathbf{0.426}^{0.158}$	$82.711^{8.253}$	$\mathbf{0.719}^{0.102}$	$\mathbf{0.081}^{0.114}$	$\mathbf{0.210}^{0.088}$	$\mathbf{0.484}^{0.142}$	$\mathbf{0.166}^{0.132}$

separation suggests that judging convergence by global fold quality alone would overlook residual errors in interface placement.

Beyond 400 steps, the docking metrics do not improve uniformly: 1000 steps slightly lowers LRMS but increases iRMS, with essentially unchanged DockQ and SR. Thus, the sweep supports 400 steps as a practical operating point, not a universal optimum.

D.4 Inference Cost

Table 8 reports reference timings for the generator pipeline underlying PODA (base) on an NVIDIA H20 GPU with batch size 4. The increase in sampling time with integration budget motivates evaluating additional steps against their limited quality gains, rather than treating a larger budget as cost-free.

D.5 Alternative Baseline Initializations

Table 9 compares DiffAb and MEAN with IgFold- and AlphaFold 3-initialized framework structures; the IgFold variants use HDOCKlite docking. MEAN designs CDR-H3 only, so its other CDR entries are unavailable. PODA (base) and PODA instead receive the masked framework sequence without an antibody pose. All interface metrics cover the complete 93-target cohort.

Changing initialization substantially affects the reconstruction and docking metrics of these pipelines, so their results should be interpreted as end-to-end system comparisons rather than isolated generator comparisons. Neither initialization closes the docking gap to PODA in this evaluation. At the same time, low clash rates or favorable Rosetta energies can coexist with poor DockQ, showing why local validity or energy scores alone are insufficient to assess native-like complex placement.

D.6 Interface Quality and Sequence Refolding

Side-chain evaluation. For PODA (base) and PODA, interface-geometry evaluation uses fixed-backbone, fixed-sequence Rosetta side-chain repacking of the antibody while retaining the antigen.

Table 12: Endpoint-supervision ablation on 93 VH–VL and 50 VHH targets. Privileged teacher + gate corresponds to PODA.

Supervision	AAR H3↑	RMSD H3↓	TM-Score↑	lDDT↑	DockQ↑	iRMS↓	LRMS↓	SR↑
VH–VL ($N = 93$)								
Direct GT	$\mathbf{0.514}^{0.191}$	$\mathbf{2.321}^{1.018}$	$\mathbf{0.930}^{0.045}$	$\mathbf{0.784}^{0.064}$	$0.159^{0.173}$	$8.812^{5.284}$	$22.916^{14.382}$	0.269
Public teacher	$\underline{0.510}^{0.195}$	$2.352^{1.042}$	$\underline{0.927}^{0.048}$	$\underline{0.780}^{0.067}$	$0.153^{0.171}$	$8.951^{5.416}$	$23.487^{15.027}$	0.247
Privileged teacher	$0.509^{0.193}$	$2.367^{1.006}$	$0.924^{0.050}$	$0.777^{0.068}$	$\underline{0.165}^{0.176}$	$\underline{8.674}^{5.321}$	$\underline{22.184}^{12.946}$	$\underline{0.280}$
Privileged teacher + gate (PODA)	$0.506^{0.196}$	$\underline{2.349}^{0.965}$	$0.925^{0.049}$	$0.778^{0.066}$	$\mathbf{0.170}^{0.177}$	$\mathbf{8.533}^{5.390}$	$\mathbf{21.440}^{11.719}$	$\mathbf{0.323}$
VHH ($N = 50$)								
Direct GT	$\mathbf{0.462}^{0.240}$	$\mathbf{2.451}^{1.218}$	$\mathbf{0.915}^{0.045}$	$\mathbf{0.796}^{0.057}$	$0.130^{0.128}$	$8.246^{3.408}$	$19.734^{8.251}$	0.160
Public teacher	$\underline{0.456}^{0.245}$	$\underline{2.498}^{1.271}$	$\underline{0.910}^{0.050}$	$\underline{0.791}^{0.063}$	$0.118^{0.116}$	$8.394^{3.462}$	$20.381^{8.617}$	0.120
Privileged teacher	$0.453^{0.248}$	$2.537^{1.298}$	$0.906^{0.054}$	$0.787^{0.067}$	$\underline{0.136}^{0.135}$	$\underline{8.137}^{3.531}$	$\underline{19.182}^{8.034}$	$\underline{0.180}$
Privileged teacher + gate (PODA)	$0.455^{0.249}$	$2.520^{1.305}$	$0.905^{0.053}$	$0.749^{0.069}$	$\mathbf{0.150}^{0.143}$	$\mathbf{8.015}^{3.497}$	$\mathbf{18.655}^{7.898}$	$\mathbf{0.240}$

Table 13: Training-state source ablation on 93 VH–VL and 50 VHH targets, with privileged-teacher supervision and the quality gate fixed. Online rollout corresponds to PODA.

Training-state source	AAR H3↑	RMSD H3↓	TM-Score↑	lDDT↑	DockQ↑	iRMS↓	LRMS↓	SR↑
VH–VL ($N = 93$)								
GT-path states	$\mathbf{0.515}^{0.190}$	$\mathbf{2.328}^{1.012}$	$\mathbf{0.930}^{0.044}$	$\mathbf{0.784}^{0.064}$	$0.158^{0.173}$	$8.857^{5.361}$	$23.116^{14.528}$	0.258
Offline rollout	$\underline{0.513}^{0.192}$	$\underline{2.339}^{1.024}$	$\underline{0.929}^{0.046}$	$\underline{0.782}^{0.066}$	$\underline{0.165}^{0.175}$	$\underline{8.716}^{5.298}$	$\underline{22.437}^{13.684}$	$\underline{0.280}$
Online rollout (PODA)	$0.506^{0.196}$	$2.349^{0.965}$	$0.925^{0.049}$	$0.778^{0.066}$	$\mathbf{0.170}^{0.177}$	$\mathbf{8.533}^{5.390}$	$\mathbf{21.440}^{11.719}$	$\mathbf{0.323}$
VHH ($N = 50$)								
GT-path states	$\mathbf{0.464}^{0.239}$	$\mathbf{2.459}^{1.231}$	$\mathbf{0.916}^{0.045}$	$\mathbf{0.797}^{0.057}$	$0.126^{0.124}$	$8.291^{3.389}$	$20.026^{8.418}$	$\underline{0.160}$
Offline rollout	$\underline{0.459}^{0.243}$	$\underline{2.482}^{1.249}$	$\underline{0.913}^{0.047}$	$\underline{0.794}^{0.059}$	$\underline{0.134}^{0.133}$	$\underline{8.173}^{3.447}$	$\underline{19.317}^{8.126}$	$\underline{0.160}$
Online rollout (PODA)	$0.455^{0.249}$	$2.520^{1.305}$	$0.905^{0.053}$	$0.749^{0.069}$	$\mathbf{0.150}^{0.143}$	$\mathbf{8.015}^{3.497}$	$\mathbf{18.655}^{7.898}$	$\mathbf{0.240}$

The VHH comparison applies this procedure to IgGM as well (Table 3). In VHH, fewer clashes and better polar-contact recovery after post-training coexist with a larger buried-area discrepancy. Recovering local contacts therefore does not guarantee recovery of the overall extent of the interface.

Sequence-refolding protocol. Tables 10 and 11 evaluate existing designed sequences with Protenix-v2, without regeneration or candidate selection. Refolding uses one sample, three cycles, and 20 diffusion steps, with independent unpaired per-chain UniRef50 MSAs and no paired MSA or templates; the VHH runs use release 2026_01. For VH–VL, ipSAE and pDockQ take the maximum over the H–A and L–A antibody–antigen chain pairs; VHH pDockQ uses the H–A pair. pLDDT uses the 0–100 scale and the other reported confidence scores use the 0–1 scale. Entries are rounded to three decimals; VH–VL rankings use unrounded values.

Post-training improves interface-related refolding confidence in both formats, while VHH pLDDT changes little. This is consistent with the distinction between interface placement and local fold quality seen in the reference-based evaluation. Refolding supplies a complementary, sequence-conditioned check, but confidence scores alone establish neither recovery of the native binding mode nor experimental binding.

D.7 Supervision and Training-State Ablations

Supervision targets. Table 12 compares direct regression to native endpoints, a public-input teacher, a privileged teacher without gating, and the complete gated method. Direct GT changes the supervision target, not the source of training states. The final row reuses the reported PODA design results rather than an additional run.

Direct GT achieves stronger local and global reconstruction than the full method, but weaker docking. Simply regressing native endpoints on these states is therefore not equivalent to learning from the privileged teacher's predictions. Privileged conditioning improves docking over the public-input teacher, and quality gating adds a further improvement in both cohorts. This pattern supports both access to training-only context and selection of teacher targets, rather than attributing the gain to teacher supervision alone.

Table 14: Five-seed design results on 93 VH–VL and 50 VHH targets. The primary tables report averages over seeds 5–9; SDs here describe target-to-target variation within each seed.

Cohort	Seed	PODA (base) DockQ↑	PODA DockQ↑	PODA (base) successes / SR↑	PODA successes / SR↑
VH–VL ($N = 93$)	5	0.1410 ± 0.1734	0.1726 ± 0.1780	19 / 20.43%	32 / 34.41%
	6	0.1360 ± 0.1664	0.1666 ± 0.1720	18 / 19.35%	29 / 31.18%
	7	0.1460 ± 0.1804	0.1756 ± 0.1850	20 / 21.51%	31 / 33.33%
	8	0.1390 ± 0.1694	0.1636 ± 0.1690	18 / 19.35%	28 / 30.11%
	9	0.1430 ± 0.1754	0.1716 ± 0.1810	20 / 21.51%	30 / 32.26%
VHH ($N = 50$)	5	0.1136 ± 0.1112	0.1532 ± 0.1458	6 / 12.00%	13 / 26.00%
	6	0.1186 ± 0.1182	0.1462 ± 0.1388	7 / 14.00%	11 / 22.00%
	7	0.1096 ± 0.1072	0.1562 ± 0.1518	5 / 10.00%	12 / 24.00%
	8	0.1156 ± 0.1142	0.1432 ± 0.1348	6 / 12.00%	11 / 22.00%
	9	0.1126 ± 0.1092	0.1512 ± 0.1438	6 / 12.00%	13 / 26.00%

Table 15: Design performance stratified by MMseqs2 antigen identity.

Identity	n	Method	AAR H3↑	RMSD H3↓	TM-Score↑	lDDT↑	DockQ↑	iRMS↓	LRMS↓	SR↑
VH–VL ($N = 93$)										
$< 30\%$	19	PODA (base)	$0.524^{0.114}$	$2.301^{1.165}$	$0.916^{0.082}$	$0.784^{0.098}$	$0.169^{0.155}$	$8.500^{5.326}$	$21.939^{14.457}$	0.211
		PODA	$0.563^{0.182}$	$2.263^{1.130}$	$0.914^{0.085}$	$0.784^{0.096}$	$0.175^{0.187}$	$9.458^{6.112}$	$22.397^{13.894}$	0.368
30–40%	17	PODA (base)	$0.474^{0.129}$	$2.059^{1.078}$	$0.944^{0.031}$	$0.800^{0.055}$	$0.105^{0.138}$	$10.275^{4.596}$	$28.948^{14.573}$	0.118
		PODA	$0.402^{0.189}$	$2.283^{0.970}$	$0.929^{0.044}$	$0.788^{0.055}$	$0.151^{0.163}$	$7.890^{4.540}$	$22.865^{12.389}$	0.353
40–60%	7	PODA (base)	$0.568^{0.052}$	$1.972^{0.643}$	$0.949^{0.020}$	$0.803^{0.039}$	$0.244^{0.217}$	$6.071^{4.430}$	$17.990^{11.344}$	0.286
		PODA	$0.505^{0.190}$	$2.329^{1.098}$	$0.938^{0.031}$	$0.792^{0.042}$	$0.115^{0.093}$	$8.917^{4.900}$	$25.444^{9.191}$	0.286
60–80%	50	PODA (base)	$0.513^{0.126}$	$2.493^{0.938}$	$0.926^{0.027}$	$0.774^{0.055}$	$0.129^{0.135}$	$9.154^{4.640}$	$23.617^{12.205}$	0.220
		PODA	$0.519^{0.190}$	$2.406^{0.867}$	$0.926^{0.028}$	$0.770^{0.056}$	$0.183^{0.184}$	$8.347^{5.381}$	$20.032^{10.621}$	0.300
VHH ($N = 50$)										
$< 30\%$	25	PODA (base)	$0.464^{0.130}$	$2.055^{0.823}$	$0.919^{0.022}$	$0.753^{0.044}$	$0.106^{0.074}$	$9.114^{2.964}$	$20.216^{6.010}$	0.120
		PODA	$0.487^{0.244}$	$2.183^{0.866}$	$0.916^{0.022}$	$0.752^{0.048}$	$0.159^{0.142}$	$7.934^{3.526}$	$17.403^{7.415}$	0.240
30–40%	11	PODA (base)	$0.466^{0.097}$	$2.985^{2.041}$	$0.900^{0.061}$	$0.759^{0.073}$	$0.097^{0.036}$	$7.951^{2.492}$	$21.437^{7.074}$	0.091
		PODA	$0.394^{0.287}$	$2.846^{1.964}$	$0.897^{0.072}$	$0.749^{0.091}$	$0.118^{0.089}$	$8.314^{3.462}$	$21.143^{8.350}$	0.182
40–60%	7	PODA (base)	$0.509^{0.110}$	$2.479^{0.444}$	$0.886^{0.091}$	$0.738^{0.103}$	$0.099^{0.058}$	$7.573^{2.699}$	$21.280^{6.164}$	0.143
		PODA	$0.370^{0.081}$	$2.693^{0.642}$	$0.889^{0.090}$	$0.734^{0.102}$	$0.122^{0.100}$	$8.193^{2.760}$	$19.972^{5.021}$	0.286
60–80%	7	PODA (base)	$0.375^{0.167}$	$3.157^{1.432}$	$0.893^{0.024}$	$0.745^{0.030}$	$0.187^{0.188}$	$7.506^{4.423}$	$19.728^{11.765}$	0.143
		PODA	$0.523^{0.267}$	$3.035^{1.464}$	$0.889^{0.032}$	$0.756^{0.040}$	$0.198^{0.217}$	$7.659^{4.025}$	$17.899^{9.837}$	0.286

Where the quality gate acts. Table 4 separates the two gate placements on VH–VL. Writing the per-example ungated losses as ℓ_{OPD} and ℓ_{RAM}, these controls use

$$\mathcal{L}_{\mathrm{II}} = \mathbb{E}[g_D \ell_{\mathrm{OPD}} + \lambda_{\mathrm{RAM}} g_R \ell_{\mathrm{RAM}}] + \lambda_{\mathrm{anchor}} \mathcal{L}_{\mathrm{sup}}, \qquad g_D, g_R \in \{1, g\}. \tag{24}$$

The no-gate, distillation-only, RAM-only, and full variants use $(g_D, g_R) = (1,1), (g,1), (1,g), (g,g)$, respectively; all four retain both losses and the supervised anchor. Holding RAM gating active, adding the distillation gate raises DockQ from 0.167 to 0.170. Holding distillation gating active, adding the RAM gate raises DockQ from 0.166 to 0.170. In both comparisons, SR rises from 28.0% to 32.3%. With neither gate, DockQ is 0.165 and SR is 28.0%; either gate alone gives a small DockQ increase but no SR increase at the reported precision. Thus, the complete method's observed advantage over either selective-gating control supports using quality selection on both losses. These comparisons distinguish gate placement from removing RAM altogether.

Training-state sources. Table 13 fixes the privileged teacher and quality gate while changing the endpoints used to construct training states: native endpoints (GT-path), offline-generated endpoints, or online endpoints from the evolving student. The online row reuses the same PODA reference as the supervision comparison.

Offline rollout improves docking over GT-path states, and online collection provides a further gain. This ordering is consistent with the motivation for training on student-induced states and refreshing their distribution as the policy changes. However, online collection also reduces some reconstruction scores, especially VHH lDDT. The evidence thus supports better interface placement, not uniform improvement in native reconstruction or direct proof that the train–inference distribution gap has been eliminated.

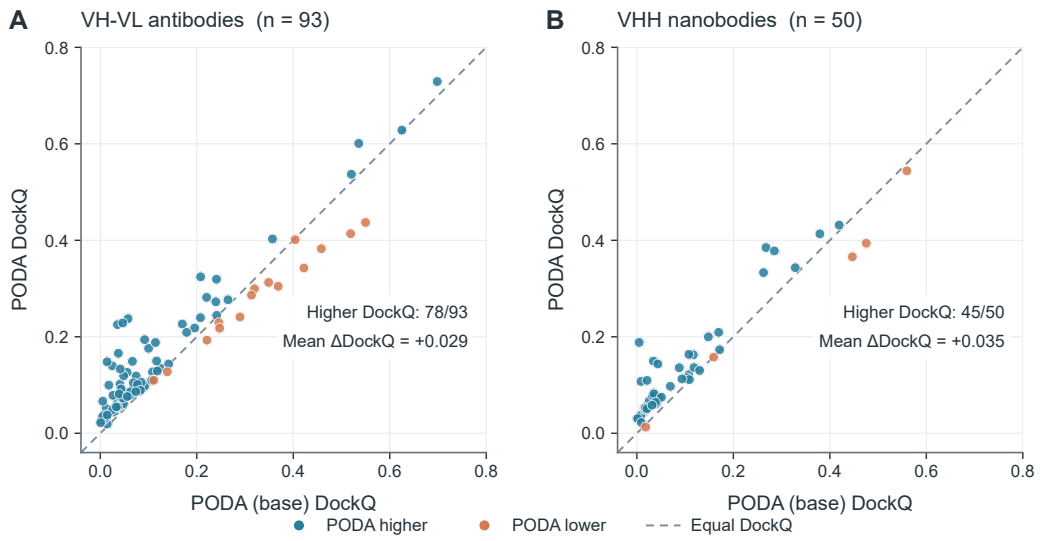

Figure 5: Paired per-target DockQ comparison between PODA (base) (horizontal axis) and PODA (vertical axis) on (A) 93 VH–VL targets and (B) 50 VHH targets. Each point represents one target. Blue points indicate higher DockQ for PODA, and orange points indicate lower DockQ. The dashed diagonal marks equal DockQ. Panel annotations report the number of improved targets and the mean DockQ difference for the plotted pairs.

D.8 Paired Target-Level DockQ Comparison

Figure 5 compares PODA (base) and PODA on matching targets in the supplied paired evaluation. Points above the diagonal indicate higher DockQ after post-training; points below it indicate lower DockQ. The annotations summarize the plotted target pairs. This target-level view complements the five-seed summaries in Appendix D.9.

D.9 Multi-Seed Design Evaluation

Table 14 reports seeds 5–9 for both cohorts, whose mean DockQ, within-seed target SD, and SR are each averaged over the five seeds in the primary design tables. PODA improves mean DockQ and SR over PODA (base) in every reported seed. The gain therefore persists across the sampled inference randomness rather than depending on a single favorable seed, although its magnitude varies.

D.10 Antigen-Similarity-Stratified Evaluation

Table 15 groups the existing test designs by the antigen identity reported for the highest-bitscore MMseqs2 alignment against the training-antigen library. No coverage threshold is applied. This is not the maximum identity over all hits, nor the normalized global alignment score used for dataset filtering (Appendix B); it should not be interpreted as a strict low-homology benchmark. The four strata are $< 30\%$, 30–40%, 40–60%, and 60–80% antigen identity, containing 19/17/7/50 VH–VL targets and 25/11/7/7 VHH targets, respectively.

The pattern varies by topology and similarity stratum. For VHH, PODA has higher mean DockQ and higher reported SR in every stratum. For VH–VL, PODA achieves higher mean DockQ and SR than PODA (base) in three of the four antigen-identity strata: $< 30\%$, 30–40%, and 60–80%.