# MOSAIC: CONDITIONAL STRUCTURAL MEMORY FOR ANTIBODY AND NANOBODY DESIGN

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Antibody and nanobody design requires jointly modeling local CDR sequence and structure and global antibody placement relative to a target antigen. Existing generators encode reusable structural patterns only implicitly and often struggle to coordinate these two objectives. Explicit retrieval offers reusable structural context, but similar antigens may support distinct antibody interfaces. We introduce MOSAIC, a conditional structural-memory method that augments a frozen flow-matching generator with diverse training examples selected using inference-time conditions. MOSAIC transfers retrieved sequence and layout features through separate, selectively gated pathways for CDR refinement and antibody pose correction. A null route allows the CDR adapter to abstain from memory use, while a calibrated support gate attenuates poorly supported pose corrections. Across VH–VL and VHH benchmarks, MOSAIC improves docking accuracy and success rates over its no-memory counterpart in both design and structure prediction, and achieves the best DockQ and success rates among the evaluated design methods. Full-test-set paired comparisons support the benefit of memory, while randomized-retrieval and component ablations on VH–VL support condition-dependent selection and complementary refinement pathways. These results demonstrate that explicit structural memory can improve antibody–antigen complex generation without updating the generative backbone.

## 1 INTRODUCTION

Antibody design must coordinate CDR sequence, conformation, and antigen-relative placement (Chothia & Lesk, 1987; Schroeder Jr & Cavacini, 2010). AbDesign and RosettaAntibodyDesign exploit natural structural and sequence preferences (Lapidoth et al., 2015; Adolf-Bryfogle et al., 2018), while learned sequence priors support design and antibody evolution (Dauparas et al., 2022; Hie et al., 2024). Complex design additionally requires compatible interface geometry (Jin et al., 2022). Iterative refinement (Jin et al., 2021), MEAN (Kong et al., 2023a), and dyMEAN (Kong et al., 2023b) couple sequence and structure. DiffAb (Luo et al., 2022), IgGM (Wang et al., 2025a), RFantibody (Bennett et al., 2026), BoltzGen (Stark et al., 2025), and Germinal (Mille-Fragoso et al., 2025) advance generative antibody and binder design, alongside flow-based approaches (Wu et al., 2025; Wang et al., 2026). Complex predictors complement generation (Abramson et al., 2024; Passaro et al., 2025; Zhang et al., 2026), but plausible local CDRs alone do not establish correct global placement.

Structural databases (Dunbar et al., 2014) provide reusable examples, as exploited by classical design (Lapidoth et al., 2015; Adolf-Bryfogle et al., 2018). Retrieval augments neural computation with external memory (Lewis et al., 2020; Guu et al., 2020), including diffusion models (Blattmann et al., 2022) and RADAb for antibody design (Wang et al., 2025b). We investigate memory for joint CDR refinement and pose correction in a frozen flow-matching generator.

Two considerations guide retrieval. Our queries exclude native CDR identities, antibody coordinates, and contact labels, although the overall protocol supplies a native-derived epitope. Furthermore, antigen similarity does not specify a unique interface: residue-level relevance need not justify a global pose update. This motivates diverse retrieval with separately gated pathways.

We introduce MOSAIC, a conditional structural-memory module (Figure 1). Training-example sequence and layout features form a memory bank independent of model outputs and evaluation



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.

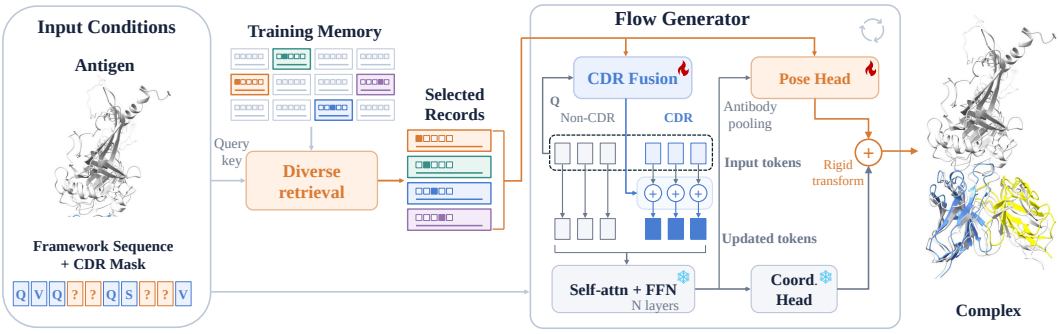



Figure 1: Overview of MOSAIC. Inference-time antigen and antibody conditions retrieve diverse training records. CDR fusion selectively updates CDR tokens, while a pose head uses pooled antibody features and retrieved context to correct global placement.

scores. Supplied antigen, framework, CDR layout, and type conditions retrieve up to four diverse records. Within a frozen backbone, cross-attention refines CDR representations and a pose head adjusts antibody placement. A null route permits residue-level abstention; a support gate attenuates weakly supported pose corrections. Disabling both pathways yields MOSAIC (off).

We evaluate design on 93 VH–VL antibodies and 50 VHH nanobodies, and structure prediction separately with supplied antibody sequences. Design retains a native-derived epitope defined by backbone contacts within 10 Å at inference; it does not establish blind antigen-only design. Paired memory-disabled comparisons use the same checkpoint, seed, and generation budget on both complete test sets. Randomized-retrieval and component ablations cover all 93 VH–VL targets.

Our contributions are:

- A conditional memory representation that separates inference-time query information from training-example content and supports diversity-aware retrieval for antibody design.
- A two-pathway memory module that refines CDR representations and antibody placement while keeping the generative backbone frozen.
- Design and structure-prediction evaluations, complemented by paired retrieval controls and analyses of model components and training-antigen similarity.

## 2 Related Work

**Antibody design and structure prediction.** AbDesign reuses natural backbone segments and sequence constraints (Lapidoth et al., 2015); RosettaAntibodyDesign combines CDR sampling and optimization (Adolf-Bryfogle et al., 2018). ProteinMPNN learns backbone-conditioned sequence design (Dauparas et al., 2022), while protein language models guide antibody evolution (Hie et al., 2024). RefineGNN jointly refines sequence and structure (Jin et al., 2021), and hierarchical refinement couples paratope generation with docking (Jin et al., 2022). Our baselines span equivariant co-design with MEAN (Kong et al., 2023a) and dyMEAN (Kong et al., 2023b), diffusion-based DiffAb (Luo et al., 2022), and antibody/nanobody generation with IgGM (Wang et al., 2025a). RFantibody adapts RFdiffusion (Bennett et al., 2026), BoltzGen targets broader binder design (Stark et al., 2025), and Germinal combines structure prediction with antibody-specific sequence modeling (Mille-Fragoso et al., 2025). We report input-protocol differences and isolate memory effects through paired comparisons. IgFold (Ruffolo et al., 2023) with HDOCK (Yan et al., 2020) separates folding from docking; tFold-Ag (Wu et al., 2026), AlphaFold 3 (Abramson et al., 2024), Boltz-2 (Passaro et al., 2025), and Protenix-v2 (Zhang et al., 2026) provide complex-prediction baselines. We distinguish prediction with supplied sequences from CDR design, and additionally refold designed sequences with Protenix-v2.

**Continuous protein generation.** Diffusion and flow matching underpin continuous generation (Ho et al., 2020; Lipman et al., 2023; Albergo & Vanden-Eijnden, 2023). RFdiffusion and Chroma support conditional protein generation (Watson et al., 2023; Ingraham et al., 2023); FrameDiff (Yim et al., 2023b), FrameFlow, and FoldFlow (Yim et al., 2023a; Bose et al., 2024) model backbone geometry.

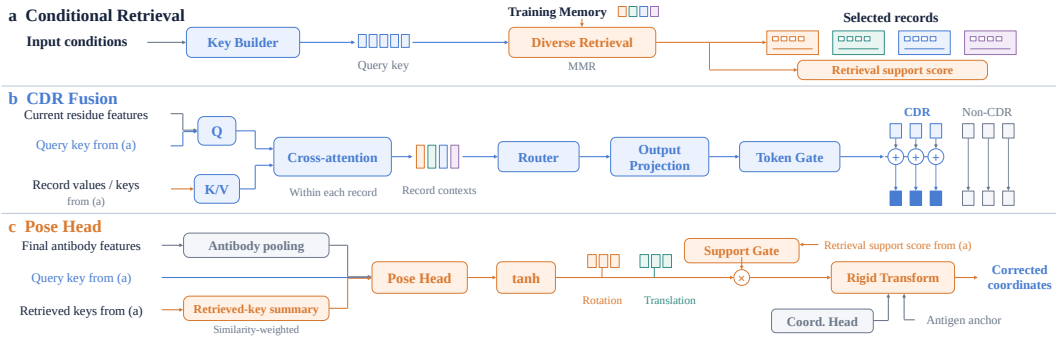



Figure 2: MOSAIC memory pathways. (a) Conditional retrieval selects diverse training records and computes a retrieval support score. (b) Current residue features and the shared query key condition record-wise cross-attention, followed by routing and gated CDR-only updates. (c) Pooled antibody features, the query key, and a retrieved-key summary drive bounded pose correction.

Extensions incorporate sequence conditioning (Huguet et al., 2024), atomistic generation through partially latent flows (Geffner et al., 2026), and antibody design (Wu et al., 2025; Wang et al., 2026). MOSAIC retains its backbone's parameters and integration procedure while adding retrieved context.

**Retrieval-augmented generation.** RAG and REALM augment parametric models with retrieval (Lewis et al., 2020; Guu et al., 2020); retrieval-augmented diffusion extends this to generative sampling (Blattmann et al., 2022). RADAb applies retrieval to structure-informed antibody design (Wang et al., 2025b). MOSAIC studies separately gated CDR refinement and bounded pose correction in a frozen flow backbone. Disabled-memory, randomized-retrieval, fixed-key value-shuffling, and parameter-matched controls assess condition-dependent memory benefits.

## 3 Method

MOSAIC augments a frozen antibody generator with explicit memory of training examples. Figure 1 gives the overall pipeline, and Figure 2 details its retrieval and memory pathways. A residue-level pathway conditions CDR representations, while a pose pathway adjusts the predicted antibody placement. Both pathways use only the supplied conditions and current model states at inference.

### 3.1 Task Definition and Generative Backbone

Let $c = (S_{\mathrm{Ag}}, X_{\mathrm{Ag}}, S_{\mathrm{FW}}, M_{\mathrm{CDR}}, \ell, \tau, e)$ denote the supplied conditions: antigen sequence and structure, antibody framework sequence, canonical CDR positions, CDR lengths, chain and antibody types, and a specified epitope description. The design task is to generate CDR identities $S_{\mathrm{CDR}}$ and complex coordinates $X$ conditioned on $c$. The CDR-position mask specifies where to design; it does not reveal the identities or native conformations of the masked residues. Native antibody coordinates are not supplied to the design model. In the reported design benchmark, $e$ is computed from backbone-atom contacts in the native complex and retained during inference, including epitope-centered preprocessing. Appendix B.1 specifies this information boundary. For structure prediction, the complete antibody sequence is supplied and only structural outputs are evaluated.

The backbone is a single-stream latent flow-matching network with antibody–antigen token representations, pair features, and coordinate recycling. Its state comprises C$\alpha$ coordinates and local structural latents. We denote its frozen parameters by $\theta$ and the trainable memory parameters by $\phi$. At flow time $t$, the network processes the current state $z_t$ and conditions $c$; MOSAIC modifies selected hidden states and the coordinate output, while retaining the backbone's flow integration and decoding procedure. With the same initial state and settings, disabling these defines MOSAIC (off).

### 3.2 Conditional Keys and Training Memory

We represent memory as

$$\mathcal{B} = \{(k_i, V_i, m_i, \mathrm{id}_i)\}_{i=1}^{N}, \tag{1}$$

where $k_i$ is a retrieval key, $V_i$ contains record-token features, $m_i$ marks valid tokens, and $\mathrm{id}_i$ supports identity exclusion. The released bank contains $N = 12{,}363$ training records, with $k_i \in \mathbb{R}^{938}$, $V_i \in \mathbb{R}^{512\times 28}$, and $m_i \in \{0,1\}^{512}$. The bank is fixed during training of the memory module.

The final query uses the layout: 160 antigen C$\alpha$ geometry features, a 768-dimensional byte-level polynomial sketch of antigen and heavy/light framework sequences, and 10 chain-composition, length, and antibody-type features. Canonical CDR masks exclude CDR identities from the sketch. The first 928 fields use training-bank mean/std normalization; the final 10 context fields remain unchanged. Appendix C.1 defines every block and the hash computation. Current model states additionally condition the CDR adapter and its routing decisions. The key builder does not read native antibody coordinates, CDR identities, epitope/paratope labels, or evaluation outputs.

Each 28-dimensional value encodes residue identity (20 dimensions), chain type (3), a canonical CDR-position flag (1), normalized antibody and CDR positions (2), and antibody type (2). Thus, the bank transfers sequence and layout patterns derived from structural training records; these values are not Cartesian coordinates. Training-example residue identities are memory content, whereas native CDR identities of the test antibody are unavailable during design. Memory construction excludes contact-derived epitope and paratope masks, clean-target coordinate tensors, model predictions, and evaluation metrics. Appendix C gives the representation and implementation details.

### 3.3 Diversity-Aware Retrieval

An antigen can admit several antibody-specific interfaces, so we retain multiple precedents rather than treating the closest record as a unique target (Figure 2a). Let $q$ be the query key and $\mathcal{S}_{r-1}$ the records selected before round $r$. Greedy maximum marginal relevance (MMR) selects

$$a(i;\mathcal{S}) = \cos(q,k_i) - \lambda_{\mathrm{div}} \max_{j\in\mathcal{S}} \cos(k_i,k_j), \tag{2}$$

$$i_r = \arg\max_{i\notin\mathcal{S}_{r-1}} a(i;\mathcal{S}_{r-1}), \qquad \mathcal{S}_r = \mathcal{S}_{r-1}\cup\{i_r\}. \tag{3}$$

The redundancy term is defined as zero when $\mathcal{S}$ is empty. We use $\lambda_{\mathrm{div}} = 0.35$ and at most $K = 4$ valid records, excluding the query's own record ID when available. Identity exclusion is distinct from the sequence-homology checks reported with the datasets.

For the pose pathway, retrieval support is the minimum adjusted score over the greedily ordered selected set:

$$s = \min_{1\le r\le|\mathcal{S}|} a(i_r;\mathcal{S}_{r-1}). \tag{4}$$

This statistic accounts for both query agreement and redundancy. An empty retrieval is handled by an explicit validity mask rather than interpreted as supported evidence.

### 3.4 CDR-Restricted Memory Fusion

The memory adapter is inserted before self-attention at backbone layer indices 8–11 (Figure 2b). At each such layer, the current residue states and query key form attention queries; record values and keys form attention keys, and record values form attention values. Masked cross-attention is computed within each record using eight heads and total attention width 128. A residue-dependent router then mixes the resulting record contexts with an additional null route whose context is zero.

Writing $C_{\ell,n,i}$ for the context from record $i$ at residue $n$, the fusion can be summarized as

$$\bar{C}_{\ell,n} = \sum_{i\in\mathcal{S}} \pi_{\ell,n,i} C_{\ell,n,i}, \qquad \sum_{i\in\mathcal{S}\cup\{\varnothing\}} \pi_{\ell,n,i} = 1, \tag{5}$$

$$h^{+}_{\ell,n} = h_{\ell,n} + M_{\mathrm{CDR},n}\, b_{\mathrm{tok}}\, \gamma_{\ell,n}\, W^{\mathrm{out}}_{\ell} \bar{C}_{\ell,n}. \tag{6}$$

Here $b_{\mathrm{tok}}$ indicates enabled and valid retrieval. The factor $\gamma_{\ell,n}$ combines a state- and time-dependent gate with retrieval similarity, antibody-type-specific coherence, and confidence estimated from inference-time features. The router uses current token states and query–record compatibility; its temperature is learned within $[0.5, 0.9]$, initialized at 0.7. The null route permits selective abstention.

### 3.5 SUPPORT-CALIBRATED GLOBAL POSE CORRECTION

Local CDR updates do not explicitly parameterize a rigid adjustment of the antibody as a whole. A separate pose head therefore consumes the final antibody token pool, the query key, and a similarity-weighted summary of retrieved keys (Figure 2c). Its six outputs parameterize an axis-angle rotation and a translation:

$$\omega = g_{\text{pose}}\, 0.35 \tanh(u_\omega), \qquad d = g_{\text{pose}}\, 0.50 \tanh(u_d)\ \text{nm}. \tag{7}$$

The scales bound individual components, not the Euclidean norms of the vectors. The pose support gate is

$$g_{\text{pose}} = b_{\text{pose}}\, \sigma\left(\frac{s - 0.42}{0.03}\right), \tag{8}$$

where $b_{\text{pose}}$ requires enabled memory, a nonempty retrieval, and a nonzero key for every valid selected slot. The sigmoid midpoint 0.42 is calibrated above the maximum support score, 0.404011190, observed for randomized retrieval using inference-time features. This gate provides smooth attenuation, not a hard rejection rule or a guarantee against unsupported retrieval. Exact zeros arise from explicitly disabled or invalid memory. This gate acts on the pose branch; Eq. equation 6 has its own token gates.

Let $a$ be the anchor computed from the supplied antigen and $\hat{x}_n$ an antibody clean C$\alpha$ prediction. The correction is

$$\hat{x}_n^{+} = a + \exp([\omega]_\times)(\hat{x}_n - a) + d. \tag{9}$$

The learned head receives token and key features; antigen coordinates supply the geometric anchor. For a velocity-parameterized coordinate channel, the implementation first forms $\hat{x} = x_t + (1-t)v_t$, applies Eq. equation 9, and converts back to the velocity output.

### 3.6 TRAINING AND INFERENCE

The backbone and privileged encoder are first trained on general protein data and then on antibody–antigen complexes, mixing privileged teacher and unprivileged student modes in both stages (Appendix C.3). During the subsequent MOSAIC memory-training stage, only $\phi$ is optimized; the backbone parameters $\theta$ remain frozen. This stage combines the backbone flow and structural objectives with memory-control objectives and teacher distillation:

$$\mathcal{L} = \mathcal{L}_{\text{backbone}} + \mathcal{L}_{\text{control}} + \lambda_v \mathcal{L}_v + \lambda_r \mathcal{L}_{\text{route}} + \lambda_c \mathcal{L}_{\text{context}} + \lambda_p \mathcal{L}_{\text{pose}}. \tag{10}$$

The control terms encourage consistency under randomized retrieval and distinguish conditional retrieval from randomized or disabled memory. The teacher is a training-only branch of the same flow transformer: a privileged encoder adds token/pair biases derived from clean antibody C$\alpha$ coordinates, autoencoder latents, native residue identities, and native paratope information before the shared trunk. It supplies detached targets for velocities, routing, retrieved context, and pose. During memory training, the shared backbone and privileged encoder are frozen; the conditional-memory parameters are optimized. Context distillation is weighted by student–teacher route agreement, and pose errors are normalized separately in radians and nanometers. The teacher branch is disabled at inference; the design student retains the specified epitope condition but receives no teacher-only antibody inputs. Appendix C.3 gives the teacher architecture, followed by the loss definitions and coefficients.

MOSAIC adds 7,644,034 parameters to the 166,055,748-parameter backbone (4.603% overhead). Memory-module training uses Adam at $10^{-4}$ for 8,000 steps with EMA decay 0.999. At inference, the query retrieves a set of memory records, the late-layer adapter updates CDR states, and the gated pose head corrects the coordinate prediction within the backbone's flow evaluation.

## 4 EXPERIMENTS

### 4.1 EXPERIMENTAL SETUP

**Datasets and comparisons.** The structural training corpus contains 11,119 antibody–antigen complexes available before 30 June 2025. We evaluate design on the complete test sets of 93 VH–VL antibodies and 50 VHH nanobodies, and structure prediction on the corresponding test sets. Paired memory-on/off comparisons cover both complete test sets; randomized-retrieval, value-shuffling,

Table 1: Antibody design on 93 targets.

| Model | | RFantibody (IgFold) | dyMEAN | IgGM | BoltzGen | Germinal | MOSAIC (off) | MOSAIC |
|---|---|---|---|---|---|---|---|---|
| AAR↑ | L1 | $0.185^{0.162}$ | $0.678^{0.237}$ | $0.628^{0.245}$ | $0.492^{0.217}$ | $0.420^{0.156}$ | $\mathbf{0.693}^{0.196}$ | $\underline{0.686}^{0.195}$ |
| | 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 | $0.174^{0.131}$ | $0.624^{0.174}$ | $0.585^{0.188}$ | $0.438^{0.175}$ | $0.427^{0.151}$ | $\mathbf{0.727}^{0.246}$ | $\underline{0.716}^{0.253}$ |
| | H1 | $0.262^{0.119}$ | $\mathbf{0.751}^{0.179}$ | $0.644^{0.154}$ | $0.601^{0.167}$ | $0.532^{0.166}$ | $0.696^{0.194}$ | $\underline{0.699}^{0.181}$ |
| | H2 | $0.225^{0.155}$ | $\underline{0.662}^{0.207}$ | $0.350^{0.179}$ | $0.424^{0.159}$ | $0.454^{0.183}$ | $\mathbf{0.667}^{0.226}$ | $0.660^{0.228}$ |
| | H3 | $0.109^{0.077}$ | $0.378^{0.126}$ | $0.245^{0.133}$ | $0.293^{0.108}$ | $0.277^{0.113}$ | $\underline{0.512}^{0.192}$ | $\mathbf{0.515}^{0.193}$ |
| RMSD↓ | L1 | $1.247^{1.057}$ | $0.803^{0.603}$ | $0.828^{0.796}$ | $0.777^{0.811}$ | $1.260^{0.827}$ | $\mathbf{0.629}^{0.665}$ | $\underline{0.632}^{0.662}$ |
| | L2 | $0.155^{0.160}$ | $0.263^{0.303}$ | $0.162^{0.147}$ | $\underline{0.141}^{0.169}$ | $0.258^{0.434}$ | $\mathbf{0.138}^{0.177}$ | $0.152^{0.188}$ |
| | L3 | $1.113^{0.781}$ | $0.762^{0.729}$ | $0.807^{0.788}$ | $0.788^{0.800}$ | $0.859^{0.588}$ | $\mathbf{0.728}^{0.753}$ | $\underline{0.748}^{0.765}$ |
| | H1 | $1.025^{0.694}$ | $0.796^{0.522}$ | $0.778^{0.540}$ | $\mathbf{0.740}^{0.555}$ | $0.774^{0.474}$ | $\underline{0.772}^{0.571}$ | $0.774^{0.566}$ |
| | H2 | $0.781^{0.670}$ | $0.645^{0.397}$ | $0.793^{0.488}$ | $0.569^{0.524}$ | $1.489^{0.717}$ | $\mathbf{0.538}^{0.483}$ | $\underline{0.545}^{0.496}$ |
| | H3 | $2.759^{1.009}$ | $\mathbf{2.224}^{0.865}$ | $2.572^{0.926}$ | $2.533^{1.106}$ | $2.607^{1.674}$ | $2.335^{1.035}$ | $\underline{2.315}^{0.987}$ |
| Structure↑ | TMscore | $0.907^{0.068}$ | $0.928^{0.043}$ | $0.918^{0.045}$ | $0.925^{0.053}$ | $0.924^{0.100}$ | $\underline{0.929}^{0.046}$ | $\mathbf{0.930}^{0.046}$ |
| | lDDT | $0.721^{0.061}$ | $0.777^{0.059}$ | $0.635^{0.049}$ | $0.705^{0.062}$ | $0.653^{0.058}$ | $\underline{0.783}^{0.066}$ | $\mathbf{0.790}^{0.067}$ |
| Side chain↓ | Clash rate | $0.030^{0.027}$ | $0.049^{0.020}$ | $0.025^{0.007}$ | $0.021^{0.002}$ | $0.038^{0.020}$ | $\underline{0.018}^{0.017}$ | $\mathbf{0.017}^{0.016}$ |
| | scRMSD | $17.667^{7.909}$ | $17.779^{6.808}$ | $15.055^{7.340}$ | $19.999^{7.929}$ | $20.398^{5.904}$ | $\underline{14.227}^{7.308}$ | $\mathbf{12.950}^{6.325}$ |
| Interface | Polar contact↑ | $0.004^{0.018}$ | $0.013^{0.049}$ | $0.011^{0.064}$ | $\underline{0.016}^{0.048}$ | $0.000^{0.000}$ | $0.014^{0.038}$ | $\mathbf{0.032}^{0.070}$ |
| | $\lvert\Delta\text{BSA}\rvert$↓ | $976.0^{487.0}$ | $1135.0^{1077.2}$ | $819.9^{449.3}$ | $672.1^{411.3}$ | $698.8^{548.5}$ | $\underline{248.9}^{151.3}$ | $\mathbf{177.3}^{101.6}$ |
| | Median $\Delta G$↓ | $-41.8$ | 7.4 | $\underline{-42.3}$ | $-41.6$ | $-15.0$ | $-32.2$ | $\mathbf{-48.6}$ |
| Docking | DockQ↑ | $0.069^{0.073}$ | $0.077^{0.065}$ | $0.149^{0.160}$ | $0.124^{0.114}$ | $0.122^{0.121}$ | $\underline{0.151}^{0.173}$ | $\mathbf{0.187}^{0.162}$ |
| | iRMS↓ | $10.337^{5.688}$ | $9.563^{3.936}$ | $\mathbf{8.563}^{5.405}$ | $\underline{8.707}^{4.957}$ | $8.849^{4.087}$ | $8.993^{5.383}$ | $8.737^{5.361}$ |
| | LRMS↓ | $28.854^{16.073}$ | $27.510^{11.337}$ | $\underline{21.695}^{14.729}$ | $21.697^{11.920}$ | $26.712^{12.053}$ | $23.825^{15.166}$ | $\mathbf{21.641}^{11.260}$ |
| | SR↑ | 0.043 | 0.032 | 0.237 | 0.172 | 0.172 | $\underline{0.247}$ | **0.312** |

and component ablations cover all 93 VH–VL targets. A parameter-matched baseline is additionally evaluated on both antibody types. Dataset construction and training-set similarity are described in Appendix B.

**Inputs and inference.** For design, dyMEAN, IgGM, and MOSAIC use the antibody framework sequence as a condition; RFantibody uses an IgFold-initialized framework structure. MOSAIC and MOSAIC (off) additionally receive a native-derived epitope mask: an antigen residue is marked when any of its backbone atoms is within 10 Å of an antibody backbone atom in the native complex. The protocol is native-epitope-conditioned design on a fixed 143-case snapshot (Appendix B.1). Other baselines follow their respective design protocols. Unless otherwise specified, MOSAIC, MOSAIC (off), and the widened-generator control use the same generation seed (42) and 400 integration steps. The controlled comparisons and antigen-similarity analysis are reported separately. Prediction supplies the complete antibody sequence. MSA inputs, random seeds, and antigen-alignment procedures are detailed in Appendix D.9.

**Metrics and reporting.** We report amino-acid recovery (AAR), TM-Score, lDDT, backbone RMSD, DockQ, interface RMSD (iRMS), ligand RMSD (LRMS), and success rate (SR; DockQ $\geq 0.23$). RMSD, iRMS, and LRMS are in Å. Interface-quality definitions and metric-specific target counts are given in Appendix D.10 and the corresponding experimental descriptions.

### 4.2 Native-Epitope-Conditioned Antibody Design

On the 93-target VH–VL benchmark (Table 1), MOSAIC achieves DockQ 0.187 and SR 0.312, compared with 0.151 and 0.247 for MOSAIC (off). These are the highest DockQ and SR among the evaluated methods. CDR-H3 AAR is 0.515, with TM-Score 0.930 and lDDT 0.790.

On the 50-target VHH design benchmark (Table 2), MOSAIC achieves DockQ 0.156 and SR 0.240, compared with 0.114 and 0.120 for MOSAIC (off), and leads the evaluated design methods on both metrics. This improvement accompanies nearly unchanged CDR3 recovery (0.458 versus 0.457) and higher H3 RMSD (2.473 versus 2.581 Å), suggesting a benefit to interface organization despite mixed local-loop metrics. CDR1–CDR3 denote AAR; DiffAb (AF3) uses AlphaFold 3 framework initialization.

**Structural examples.** Figure 3 shows closer antigen-aligned antibody placement for selected MOSAIC designs. On VH–VL examples 9jme and 9leo, DockQ is 0.716 and 0.584, compared with 0.146 and 0.118 for MOSAIC (off). On VHH examples 9riu and 9yc3, MOSAIC scores 0.403 and 0.338, versus 0.139 and 0.057 for the strongest comparators. All four MOSAIC designs exceed the success threshold, while none of the displayed comparators does. Alignment and case-selection details appear in Appendix D.14.

Table 2: Design and structure prediction on 50 nanobody targets.

| Design of Nanobody | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Method | CDR1↑ | CDR2↑ | CDR3↑ | 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 | $0.243^{0.159}$ | $0.180^{0.125}$ | $0.107^{0.096}$ | $2.806^{1.301}$ | $0.114^{0.106}$ | $\mathbf{7.656}^{3.587}$ | $\mathbf{20.308}^{11.662}$ | 0.120 |
| BoltzGen | $0.362^{0.200}$ | $0.404^{0.158}$ | $0.204^{0.091}$ | $\mathbf{2.268}^{1.037}$ | $0.127^{0.127}$ | $8.050^{4.560}$ | $21.555^{15.806}$ | 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 |
| MOSAIC (off) | $0.478^{0.227}$ | $0.553^{0.235}$ | $\mathbf{0.458}^{0.242}$ | $2.473^{1.253}$ | $0.114^{0.112}$ | $8.417^{3.324}$ | $20.565^{8.292}$ | 0.120 |
| MOSAIC | $\mathbf{0.488}^{0.221}$ | $\mathbf{0.555}^{0.238}$ | $0.457^{0.247}$ | $2.581^{1.328}$ | $\mathbf{0.156}^{0.109}$ | $8.395^{3.496}$ | $20.532^{8.800}$ | **0.240** |
| 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‡ | $0.935^{0.055}$ | $0.844^{0.081}$ | $1.665^{1.359}$ | | $0.237^{0.308}$ | $10.456^{7.923}$ | $32.563^{25.322}$ | 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 |
| MOSAIC (off)†* | $0.913^{0.047}$ | $0.794^{0.060}$ | $2.392^{1.277}$ | | $0.181^{0.136}$ | $\mathbf{7.859}^{3.488}$ | $\mathbf{17.532}^{6.859}$ | 0.240 |
| MOSAIC†* | $0.913^{0.049}$ | $0.794^{0.062}$ | $2.416^{1.359}$ | | $0.218^{0.118}$ | $8.709^{3.703}$ | $19.737^{7.088}$ | 0.340 |

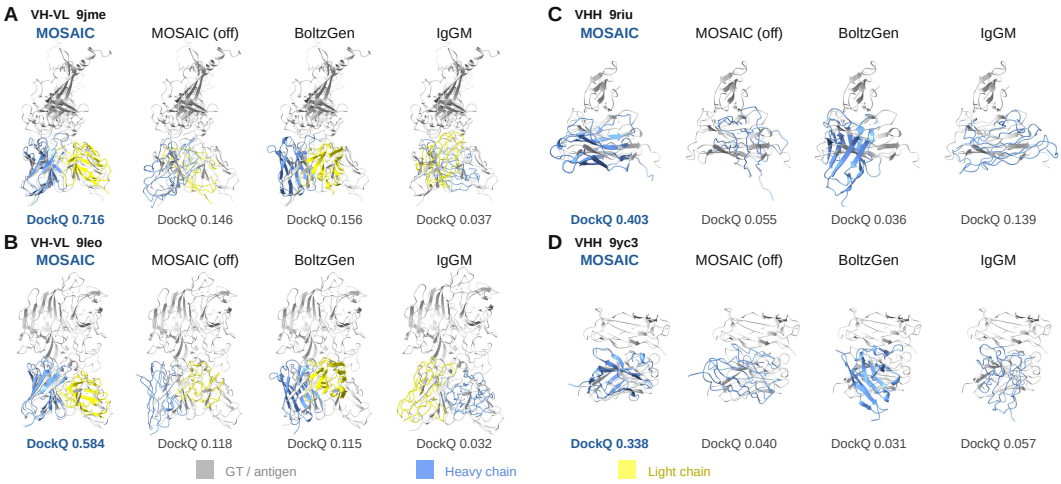



Figure 3: Selected design examples: (A) 9jme and (B) 9leo, VH–VL antibodies; (C) 9riu and (D) 9yc3, VHH nanobodies.

### 4.3 Sequence-Refolding Confidence

We refold the designed sequences with Protenix-v2 under a common protocol. On all 93 VH–VL targets (Table 3), MOSAIC has the highest mean global ipTM, pTM, ipSAE, pDockQ, and ranking score. Refolded DockQ increases from 0.088 to 0.165 over MOSAIC (off), while global ipTM increases from 0.521 to 0.525. VHH refolding DockQ likewise increases from 0.094 to 0.169 (Appendix D.13). The docking advantage persists after an independent predictor reconstructs the complexes, providing evidence beyond the generator's original coordinates; these remain computational assessments.

### 4.4 Complex Structure Prediction

On the 93-target VH–VL prediction benchmark (Table 4), MOSAIC achieves DockQ 0.244 and SR 0.398, compared with 0.165 and 0.269 for MOSAIC (off). MOSAIC also has the lowest iRMS and LRMS in the table, while MSA-based methods perform better on global antibody-structure metrics.

On the 50-target VHH prediction benchmark (Table 2), MOSAIC achieves DockQ 0.218 and SR 0.340, exceeding MOSAIC (off) on both metrics. Protenix-v2 and AlphaFold 3 have higher DockQ and stronger global structure scores. Because prediction supplies the antibody sequence, the improvement over MOSAIC (off) supports an effect on complex geometry that does not require changing the sequence. The advantage over the shared backbone does not extend to every external predictor or structural metric.

Table 3: Protenix-v2 refolding confidence and docking accuracy on 93 antibodies.

| Method | Global ipTM↑ | pLDDT↑ | pTM↑ | ipSAE↑ | pDockQ↑ | Ranking score↑ | DockQ↑ |
|---|---|---|---|---|---|---|---|
| DiffAb (IgFold) | $0.516^{0.135}$ | $\mathbf{83.211}^{5.100}$ | $0.665^{0.097}$ | $0.025^{0.073}$ | $0.119^{0.062}$ | $0.546^{0.125}$ | $0.046^{0.042}$ |
| 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 | $0.487^{0.131}$ | $80.362^{5.650}$ | $0.656^{0.088}$ | $0.022^{0.071}$ | $0.120^{0.051}$ | $0.521^{0.119}$ | $0.068^{0.069}$ |
| MOSAIC (off) | $\underline{0.521}^{0.139}$ | $82.272^{5.853}$ | $\underline{0.672}^{0.096}$ | $\underline{0.035}^{0.076}$ | $0.125^{0.056}$ | $\underline{0.551}^{0.128}$ | $\underline{0.088}^{0.090}$ |
| MOSAIC | $\mathbf{0.525}^{0.144}$ | $82.785^{5.469}$ | $\mathbf{0.674}^{0.092}$ | $\mathbf{0.042}^{0.099}$ | $\mathbf{0.139}^{0.064}$ | $\mathbf{0.554}^{0.130}$ | $\mathbf{0.165}^{0.141}$ |

Table 4: Complex structure prediction on 93 antibody targets.

| Method | Antibody Structure | | | Docking | | | |
|---|---|---|---|---|---|---|---|
| | 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‡ | $\mathbf{0.945}^{0.046}$ | $\mathbf{0.850}^{0.078}$ | $\underline{1.813}^{1.266}$ | $\underline{0.219}^{0.300}$ | $10.916^{8.566}$ | $34.082^{29.078}$ | 0.269 |
| AlphaFold 3‡ | $\underline{0.944}^{0.047}$ | $\underline{0.849}^{0.076}$ | $\mathbf{1.801}^{1.371}$ | $0.213^{0.247}$ | $11.032^{6.993}$ | $32.466^{25.263}$ | $\underline{0.280}$ |
| 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 |
| MOSAIC (off)†* | $0.932^{0.046}$ | $0.820^{0.064}$ | $2.268^{1.073}$ | $0.165^{0.158}$ | $\underline{9.087}^{4.881}$ | $\underline{22.961}^{11.717}$ | 0.269 |
| MOSAIC†* | $0.932^{0.046}$ | $0.819^{0.065}$ | $2.234^{1.122}$ | $\mathbf{0.244}^{0.166}$ | $\mathbf{8.629}^{4.977}$ | $\mathbf{21.203}^{11.213}$ | **0.398** |

### 4.5 Paired Evaluation of Conditional Memory

We compare MOSAIC with MOSAIC (off) on all 93 VH–VL and 50 VHH targets using the same checkpoint, seed 42, and 400 integration steps. Disabling both memory pathways preserves the underlying generator, isolating the effect of memory under matched generation conditions.

Figure 4 shows higher DockQ on 81/93 VH–VL and 40/50 VHH targets (87.1% and 80.0%), with mean paired gains of 0.036 and 0.035. Improvements therefore extend across most targets beyond the selected examples, although some cases deteriorate with memory.

Table 5 complements these paired comparisons with retrieval and memory-content controls on all 93 VH–VL targets.

### 4.6 Ablations and Generalization

Table 5 applies inference-time controls to all 93 VH–VL targets at a fixed checkpoint, seed, and integration budget. These disable pathways or the support gate, alter retrieved records or content, or retain one record.

Disabling pose correction or the CDR adapter reduces DockQ from 0.186 to 0.155 or 0.153; randomized retrieval gives 0.153, close to 0.150 for MOSAIC (off). With keys and support fixed, shuffled values give DockQ 0.161–0.167 across three seeds; shuffling CDR identities with keys and layout fixed gives 0.170 (Appendix D.2). These controls support a contribution from correctly associated content beyond retrieval support alone. Shuffled values still outperform MOSAIC (off), consistent with retained keys and unchanged components contributing useful information. Removing the support gate or using one record gives smaller DockQ drops to 0.181 or 0.182, with SR 0.290 versus 0.312 for MOSAIC, a difference of two successes. These settings refine the benefit of the two pathways, while some variants retain higher AAR or lower LRMS. Control settings appear in Appendix D.1.

**Parameter-count control.** We widen the generator of MOSAIC (off) to match MOSAIC's parameter count, keeping training data and training and integration steps identical. At seed 42, it achieves DockQ 0.152/0.120 and SR 0.247/0.140 on VH–VL/VHH (Appendix D.3). MOSAIC improves DockQ by 0.035/0.036 and SR by 6.5/10.0 percentage points (Tables 1 and 2). This supports a benefit

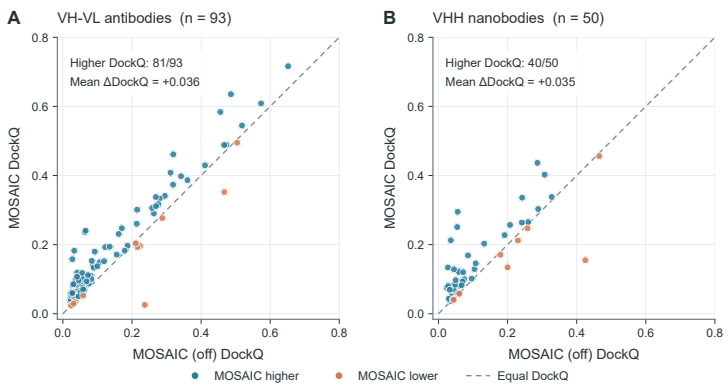



Figure 4: Paired DockQ on (A) 93 VH–VL and (B) 50 VHH targets. Blue/orange points indicate higher/lower MOSAIC scores than MOSAIC (off).

Table 5: Ablations on 93 antibody targets under native-epitope conditioning.

| Method | AAR H3↑ | RMSD H3↓ | TM-Score↑ | lDDT↑ | DockQ↑ | iRMS↓ | LRMS↓ | SR↑ |
|---|---|---|---|---|---|---|---|---|
| MOSAIC (off) | $0.513^{0.195}$ | $2.338^{1.025}$ | $0.929^{0.046}$ | $0.782^{0.067}$ | $0.150^{0.151}$ | $8.810^{5.310}$ | $22.030^{11.420}$ | 0.247 |
| w/o pose correction | $\mathbf{0.519}^{0.190}$ | $2.320^{0.989}$ | $\underline{0.930}^{0.046}$ | $0.784^{0.066}$ | $0.155^{0.157}$ | $8.750^{5.348}$ | $21.749^{11.241}$ | 0.247 |
| w/o CDR adapter | $\underline{0.518}^{0.195}$ | $2.322^{1.004}$ | $\underline{0.930}^{0.045}$ | $0.784^{0.066}$ | $0.153^{0.159}$ | $8.730^{5.236}$ | $21.765^{11.283}$ | 0.269 |
| Randomized retrieval | $0.512^{0.197}$ | $2.344^{1.031}$ | $0.929^{0.047}$ | $0.783^{0.067}$ | $0.153^{0.154}$ | $8.790^{5.340}$ | $21.980^{11.380}$ | 0.258 |
| Shuffled values (fixed keys/support) | $0.513^{0.194}$ | $2.331^{1.014}$ | $0.929^{0.046}$ | $0.786^{0.066}$ | $0.164^{0.157}$ | $8.760^{5.320}$ | $21.850^{11.350}$ | 0.269 |
| Shuffled CDR identities (fixed keys/layout) | $0.511^{0.196}$ | $2.329^{1.012}$ | $\underline{0.930}^{0.046}$ | $0.787^{0.066}$ | $0.170^{0.159}$ | $8.710^{5.300}$ | $21.740^{11.320}$ | 0.280 |
| w/o support gate | $0.514^{0.195}$ | $2.326^{0.988}$ | $\underline{0.930}^{0.046}$ | $\underline{0.789}^{0.066}$ | $0.181^{0.161}$ | $8.703^{5.367}$ | $\underline{21.499}^{11.281}$ | $\underline{0.290}$ |
| Single-record retrieval ($k = 1$) | $0.516^{0.195}$ | $\underline{2.315}^{0.998}$ | $\underline{0.930}^{0.046}$ | $\underline{0.789}^{0.066}$ | $\underline{0.182}^{0.162}$ | $\underline{8.664}^{5.330}$ | $\mathbf{21.424}^{11.265}$ | $\underline{0.290}$ |
| MOSAIC | $0.516^{0.192}$ | $\mathbf{2.310}^{0.991}$ | $\mathbf{0.931}^{0.045}$ | $\mathbf{0.790}^{0.065}$ | $\mathbf{0.186}^{0.163}$ | $\mathbf{8.650}^{5.250}$ | $21.520^{11.180}$ | **0.312** |

beyond generator width under matched data and step budgets, complementing the content-shuffling controls.

Across generation seeds 1–5, MOSAIC improves mean DockQ and SR over MOSAIC (off) on both complete test sets at every seed (Appendix D.4). The consistently positive gains across all ten seed–benchmark comparisons support robustness to generation randomness and reduce concern that the improvement depends on a favorable seed. This evidence concerns docking quality and success.

Appendix D.5 stratifies targets into $\leq 30\%$, $(30, 40]\%$, $(40, 60]\%$, and $(60, 80]\%$ training-antigen identity groups. MOSAIC improves mean DockQ in every group, including gains of 0.028 on 30 VH–VL and 0.031 on 35 VHH targets in the lowest-identity stratum. This supports improvement beyond high antigen identity, without establishing separation by antibody sequence or interface similarity. Additional budget, initialization, and interface-quality comparisons appear in Appendix D.

## 5 Conclusions

MOSAIC augments a frozen flow-matching backbone with conditional memory of structural training examples. Diversity-aware retrieval supports two complementary updates: CDR representation refinement and bounded antibody-pose correction. Paired comparisons on the complete test sets of 93 VH–VL antibodies and 50 VHH nanobodies show higher mean DockQ than MOSAIC (off). Randomized-retrieval and component ablations on all 93 VH–VL targets support condition-dependent retrieval and the contributions of both pathways. On the complete design benchmarks, MOSAIC achieves DockQ scores of 0.187 and 0.156 and success rates of 31.2% and 24.0% for VH–VL and VHH, respectively. These results support the use of conditional memory for improving generated complex geometry.

A limitation is the lack of wet-lab validation; future work will experimentally evaluate the binding activity of the designed antibodies and nanobodies.

## AI USE STATEMENT

We used generative AI tools for language editing, manuscript organization, and consistency checking. Experimental results and method specifications were checked against the source data, configurations, and machine-readable summaries. Generative AI was not used to fabricate experimental observations. The authors reviewed all AI-assisted text and take responsibility for the final manuscript.

## ETHICS STATEMENT

MOSAIC is a computational method for generating antibody-complex candidates and has potential dual-use implications. Its outputs do not establish biological activity and require experimental validation. We recommend controlled use of generation assets, explicit memory-bank provenance, institutional biosafety review, and compliance with applicable regulations. The study uses public structural data and involves no human subjects.

## REPRODUCIBILITY STATEMENT

The supplement contains source code, training and inference configurations, checkpoints, memory-bank metadata, file checksums, and evaluation scripts. Appendix A describes the software and model assets, Appendix B defines the datasets and leakage controls, and Appendix C provides implementation details. Paired memory-on/off comparisons cover the complete 93-target VH–VL and 50-target VHH test sets under matched generation settings; randomized-retrieval and component ablations cover all 93 VH–VL targets.

## 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.

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

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 Blattmann, Robin Rombach, Kaan Oktay, Jonas Müller, and Björn Ommer. Retrieval-Augmented Diffusion Models. *Advances in Neural Information Processing Systems*, 35:15309–15324, December 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/62868cc2fc1eb5cdf321d05b4b88510c-Abstract-Conference.html.

Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian FATRAS, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael M Bronstein, and Alexander Tong. Se (3)-stochastic flow matching for protein backbone generation. In *The Twelfth International Conference on Learning Representations*, 2024.

Cyrus Chothia and Arthur M Lesk. Canonical structures for the hypervariable regions of immunoglobulins. *Journal of molecular biology*, 196(4):901–917, 1987.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022.

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.

Tomas Geffner, Kieran Didi, Zhonglin Cao, Danny Reidenbach, Zuobai Zhang, Christian Dallago, Emine Kucukbenli, Karsten Kreis, and Arash Vahdat. La-proteina: Atomistic protein generation via partially latent flow matching. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=RDerF20JYT.

Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Mingwei Chang. Retrieval Augmented Language Model Pre-Training. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 3929–3938. PMLR, November 2020. URL https://proceedings.mlr.press/v119/guu20a.html. ISSN: 2640-3498.

Brian L Hie, Varun R Shanker, Duo Xu, Theodora UJ Bruun, Payton A Weidenbacher, Shaogeng Tang, Wesley Wu, John E Pak, and Peter S Kim. Efficient evolution of human antibodies from general protein language models. *Nature biotechnology*, 42(2):275–283, 2024.

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

Guillaume Huguet, James Vuckovic, Kilian Fatras, Eric Thibodeau-Laufer, Pablo Lemos, Riashat Islam, Cheng-Hao Liu, Jarrid Rector-Brooks, Tara Akhound-Sadegh, Michael Bronstein, et al. Sequence-augmented se (3)-flow matching for conditional protein backbone generation. *Advances in neural information processing systems*, 38, 2024.

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.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Antibody-antigen docking and design via hierarchical structure refinement. In *International Conference on Machine Learning*, pp. 10217–10227. PMLR, 2022.

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.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 9459–9474. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.

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.

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.

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.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, et al. Boltz-2: Towards accurate and efficient binding affinity prediction. *bioRxiv*, 2025.

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.

Harry W Schroeder Jr and Lisa Cavacini. Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology*, 125(2):S41–S52, 2010.

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.

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*, 2025a. 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.

Zichen Wang, Yaokun Ji, Jianing Tian, and Shuangjia Zheng. Retrieval augmented diffusion model for structure-informed antibody design and optimization. In *The Thirteenth International Conference on Learning Representations*, 2025b.

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.

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.

Jason Yim, Andrew Campbell, Andrew YK Foong, Michael Gastegger, José Jiménez-Luna, Sarah Lewis, Victor Garcia Satorras, Bastiaan S Veeling, Regina Barzilay, Tommi Jaakkola, et al. Fast protein backbone generation with se (3) flow matching. *arXiv preprint arXiv:2310.05297*, 2023a.

Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. Se(3) diffusion model with application to protein backbone generation. In *International Conference on Machine Learning*, 2023b.

Yuxuan Zhang, Chengyue Gong, Jinyuan Sun, Jiaqi Guan, Milong Ren, Song Xue, Hanyu Zhang, Wenzhi Ma, Zhenyu Liu, Xinshi Chen, and Wenzhi Xiao. Protenix-v2: Broadening the reach of structure prediction and biomolecular design. *bioRxiv*, 2026.

# A Code and Model Availability

The source code is available in the anonymous repository at `https://anonymous.4open.science/r/MOSAIC-D62C`.

The supplementary package contains source code for training, design, and structure prediction; model configurations; evaluation scripts; and a small VHH example for testing installation. Model assets comprise the MOSAIC checkpoint with exponential moving-average weights, the frozen generative backbone, an autoencoder checkpoint, and the memory bank.

# B Datasets and Leakage Controls

The structural training corpus contains 11,119 antibody–antigen complexes available before 30 June 2025. The memory bank contains 12,363 processed training records. Complex counts and memory-record counts describe different units and should not be equated. The antigen-similarity search uses 12,340 valid antigen profiles from the training split.

The complete design benchmarks contain 93 VH–VL antibodies and 50 VHH nanobodies, fixed as 143 unique sample IDs in a JSON file before generation rather than selected by generated quality. The public snapshot records 93 VH–VL samples. Each case includes a normalized complex PDB, antibody/antigen/complex FASTA files, a design FASTA, and metadata. Structure prediction is evaluated separately on the corresponding test sets. Paired memory-on/off comparisons include all 93 VH–VL and 50 VHH targets; randomized-retrieval and component ablations include all 93 VH–VL targets. MOSAIC uses seed 42 and 400 integration steps for the complete design benchmarks. The main design tables use seed 42 for MOSAIC (off), and the paired full-cohort, ablation, and antigen-similarity comparisons use seed 42 for both models.

CDRs follow IMGT numbering. For MOSAIC, loop backbone RMSD is calculated after fitting the corresponding CDR. Four VHH prediction targets lack a valid H3 mask, so H3 RMSD uses 46 targets without inferred loop boundaries; the complete VHH design results use all 50. DockQ, $f_{\mathrm{nat}}$, iRMS, and LRMS evaluate the antibody–antigen interface rather than the internal VH–VL interface. Method-specific prediction and design settings are detailed in Appendices D.9 and D.10.

## B.1 Native-Epitope-Conditioned Design Protocol

**Deterministic epitope construction.** The preprocessor reads each normalized native complex with epitope calculation enabled. VH–VL antibody chains are H/L, VHH uses H without L, and the antigen is mapped to A. An antigen residue $r$ is labeled as epitope if any of its backbone atoms (N, C$\alpha$, C, O) lies within 10 Å of any antibody backbone atom:

$$e_r = \mathbf{1}\left[\min_{a\in\mathcal{A}_{\mathrm{Ab}}^{\mathrm{bb}},\, b\in\mathcal{A}_r^{\mathrm{bb}}} \|x_a - x_b\|_2 < 10\,\text{Å}\right]. \tag{11}$$

This mask is recomputed from the native PDB, not supplied by an external epitope predictor. The antibody-side paratope flag instead uses the minimum antibody–antigen C$\alpha$ distance with the same strict 10 Å threshold. Antigens longer than 500 residues are cropped to a 500-residue window centered on the epitope span, retaining the corresponding mask. All 143 public cases have a nonempty processed epitope.

**Model inputs versus evaluation references.** The design dataset masks CDR identities and zeros antibody coordinates before flow generation, while retaining framework sequence and antigen sequence/coordinates. Native antibody coordinates and sequence are retained in `gt_*` reference fields for evaluation; the student feature and retrieval-key builders do not read these fields. In contrast, `epitope_mask` remains in the model batch. Coordinates are centered on the epitope C$\alpha$ centroid; when CDR–epitope cross-attention is enabled, it directly uses the CDR mask, epitope mask, and antigen geometry. Thus the benchmark supplies a native-complex-derived interface location, even though native antibody coordinates and CDR identities are not direct model inputs.

The reported MOSAIC design protocol is therefore *native-epitope-conditioned antibody design on a fixed 143-case complex snapshot*. The retrieval-key builder itself does not consume contact

labels, but its supplied antigen can reflect epitope-dependent cropping. The teacher's clean antibody geometry, identities, and paratope information are used only during training. These distinctions prevent interpreting the design results as blind antigen-only generation. Such a claim would require a separate evaluation of all 143 targets with native-derived epitope masks and epitope-dependent preprocessing removed or replaced by explicitly specified non-native conditions.

### B.2 Training-Set Similarity and Retrieval Boundaries

Temporal separation does not guarantee sequence-homology separation. We examine exact antigen reuse and maximum CDR identity against training antibodies with the same CDR-length signature. Among 789 exact-antigen groups in the memory bank, 87.45% contain conflicting observed CDR masks, 41.83% contain a near-disjoint mask pair, and four retrieved records provide 92.13% exact-label coverage in the ambiguity analysis. These statistics motivate retaining multiple examples; they do not make a retrieved record a unique native answer.

We additionally stratify each complete test set by the highest test-to-training antigen identity returned by MMseqs2 with at least 50% query coverage. The results in Appendix D.5 retain all test targets, including highly similar antigens, and therefore should not be described as a sequence-disjoint benchmark. Native held-out CDR identities and contact labels are excluded from the retrieval-key features; this restriction does not remove the design model's native-derived epitope condition described above.

## C Implementation and Training Details

This appendix provides implementation and training details for the memory pathways in Figure 2.

**Backbone and memory features.** The frozen backbone has 14 transformer layers, token width 768, and 12 attention heads. Its flow state includes three-dimensional C$\alpha$ coordinates and an eight-dimensional local latent. The CDR adapter acts before self-attention at zero-based layer indices 8–11 and directly updates only canonical CDR positions. Retrieval uses the 938-dimensional key defined below, with a frozen training-bank transform reused at inference.

The bank contains 12,363 records with at most 512 tokens each. Value channels 0–19 encode residue identity; 20–22 encode chain type; 23 is the canonical CDR flag; 24 and 25 encode normalized antibody and CDR positions; and 26–27 encode antibody type. The value representation does not contain native query coordinates. Memory construction excludes native contact-derived epitope and paratope masks, clean-target tensors, model predictions, and evaluation metrics. Training-example residue identities and canonical CDR annotations remain part of the representation. The number of bank records should not be equated with the number of unique training complexes.

### C.1 Complete Retrieval-Key Definition

The final key concatenates geometry (160 fields), a polynomial sequence sketch (768), and public context (10), in that order. Version v10 reorders the existing geometry/context and sketch features to match the bank transform; it adds no new information source.

**Antigen geometry (160 fields).** Only supplied antigen C$\alpha$ coordinates, in nm, enter this block. For $n > 128$ valid coordinates in sequence order, deterministic sampling retains indices $\lfloor jn/128 \rfloor$, $j = 0, \ldots, 127$, into that ordered list. Otherwise all valid points are retained. The features in Table 6 are invariant to translation and proper rotation.

Histograms use equal-width bins, clamp out-of-range values to the end bins, and normalize by their total count; empty histograms are zero. Local geometry uses consecutive valid residues without bridging missing-coordinate gaps. An antigen with no valid C$\alpha$ points has a zero geometry block.

**Polynomial sequence sketch (768 fields).** Three selectors identify antigen, heavy framework, and light framework positions. Only windows lying entirely within one selector are counted, excluding CDR identities and windows crossing CDR boundaries. Each segment has four hash heads with 64 buckets per head. Orders 1–6 share the same head/bucket space, giving $3 \times 4 \times 64 = 768$ fields. A VHH has a zero light-framework block.

Table 6: Antigen geometry features in the retrieval key.

| Feature | Dim. | Definition |
|---|---|---|
| Global pair distances | 32 | Histogram of unordered sampled C$\alpha$ pairs, 0–6.0 nm. |
| Sequence-separated distances | 48 | Three 16-bin distance histograms over 0–6.0 nm, for sequence separations 1–4, 5–16, and $\geq 17$. |
| Radial neighborhoods | 24 | Distances from sampled points to all valid antigen C$\alpha$ points, 0–4.0 nm. |
| Consecutive bond lengths | 16 | Adjacent-residue C$\alpha$ distances, 0–0.8 nm. |
| Local angles | 16 | Angles from three consecutive C$\alpha$ points, 0–$\pi$. |
| Local torsions | 16 | Signed torsions from four consecutive C$\alpha$ points, $-\pi$–$\pi$. |
| Covariance eigenvalues | 3 | Nonnegative eigenvalues of the full valid point-cloud covariance, normalized by their sum. |
| Global shape statistics | 5 | Mean, population std, median, 90th percentile, and maximum distance to the full point-cloud centroid, each divided by 6.0 nm. |

For a window of amino-acid ASCII bytes $b_1, \ldots, b_n$ and head $h \in \{0, 1, 2, 3\}$, the implementation uses

$$H_0 = 0, \qquad H_j = \left[(257 + 2h)H_{j-1} + b_j + 131(h + 1)\right] \bmod 2147483647, \tag{12}$$

and increments bucket $H_n \bmod 64$. Thus hashing uses amino-acid letters rather than upstream token IDs. Let $C$ concatenate the counts from all three segments, all heads, and all orders. The implemented sketch is $4C/\max(1, \sum_j C_j)$: normalization is over the concatenated 768-dimensional count vector, not separately within each segment.

**Public context (10 fields) and bank transform.** The context tail contains three fractions of valid residues belonging to heavy, light, and antigen chains; five lengths for antigen, heavy framework, light framework, heavy CDR, and light CDR, divided by 500, 450, 450, 100, and 100, respectively; and a two-way VH–VL/VHH one-hot encoding. The frozen bank transform subtracts training means and divides by training standard deviations for the first 928 fields (geometry and sketch), leaving the context tail unchanged. Query and bank layouts agree.

The key builder accepts antigen sequence/geometry, framework sequence, canonical CDR masks, and chain/topology information. It does not read native antibody coordinates, CDR amino-acid identities, contact-derived epitope/paratope labels, generated structures, endpoint predictions, or evaluation metrics. This feature-level restriction does not remove the benchmark's native-epitope conditioning: the supplied antigen can already have undergone epitope-dependent cropping (Appendix B.1).

### C.2 Memory Pathways and Controls

**Token routing and abstention.** The CDR adapter uses LayerNorm-normalized representations and record-wise cross-attention. A learned router mixes one context per retrieved record with an additional zero-context null route. Record logits combine state-dependent compatibility with query similarity and record diversity. Null-route logits depend on residue state, antibody type, flow time, and retrieval uncertainty. The record temperature is initialized to 0.7 and learned within $[0.5, 0.9]$ from inference-time evidence; the diversity coefficient is 0.35 and the null-route temperature is 0.7. A separate state- and time-dependent gate also incorporates retrieval similarity, VHH coherence, and confidence estimated from available query features. Output projections are specific to each adapted layer.

**Pose gate and bounds.** The pose head pools valid antibody tokens and concatenates the query with a retrieved-key summary. The latter uses masked cosine-similarity weights with temperature 0.20. A LayerNorm–linear–SiLU–linear head of hidden width 128 predicts six components. Each axis-angle component is bounded by 0.35 radians and each translation component by 0.50 nm before support gating. Consequently, the corresponding vector norms are bounded by $\sqrt{3} \times 0.35$ radians and $\sqrt{3} \times 0.50$ nm, respectively.

The support statistic is the minimum greedy MMR-adjusted score of the selected keys, with redundancy coefficient 0.35. The retrieval-support sigmoid has midpoint 0.42 and temperature 0.03. The maximum randomized-retrieval support score, 0.404011190, produces a sigmoid value of approximately 0.370 before validity masking. Calibration therefore gives smooth pose modulation rather

than a hard rejection threshold. The explicit validity multiplier is zero for disabled memory, empty retrieval, or a missing key in a valid slot. Token abstention and pose support gating are different computations.

**Initialization and controls.** The token output projections and final pose-head layer are initialized to zero. They preserve the backbone prediction at initialization while allowing their output weights to learn. At inference, disabling both pathways recovers the backbone computation under identical inputs, initial noise, and solver settings. The exposure-only control removes key information and yields no effective memory update. Randomized retrieval is different: it can produce a nonzero correction. The paired comparisons use the same inputs, initial noise, and solver settings.

### C.3 Privileged Teacher and Distillation

**Shared architecture and teacher-only inputs.** The privileged teacher is a training-only branch of the same flow transformer. Starting from the ordinary token and pair representations of the corrupted training state, `PrivilegedInfoEncoder` adds a sequence bias and a pair bias before the shared transformer trunk. The downstream transformer layers, flow heads, and conditional-memory module are shared with the student; the teacher is not a separately trained large model.

The sequence encoder concatenates clean C$\alpha$ coordinates (3), clean local latents from the frozen autoencoder (8), native residue one-hot features including CDR identities (20), and a native paratope flag (1), for 32 input channels. The pair encoder concatenates 32 radial-basis features of clean C$\alpha$ distances with the interface channel $p_i e_j$, where $p_i$ and $e_j$ are paratope and epitope flags. The radial bases are $\exp[-(d-\mu_k)^2/0.08]$, with 32 equally spaced centers from 0.1 to 3.0 nm. Each encoder is LayerNorm–linear–SiLU–linear, with hidden/output width equal to the corresponding token or pair width. The final bias-free linear layers are initialized to zero, so adding the encoder initially preserves the shared base forward computation.

On the teacher branch, antibody geometry and local latents come from the clean training target; antigen coordinates remain the supplied antigen condition. Native antibody residue identities and native paratope flags are privileged inputs. Chain types, canonical CDR masks, and the specified epitope condition are also available to the student. The teacher branch leaves the public retrieval-key computation unchanged.

**Backbone pretraining and antibody adaptation.** Before training the MOSAIC memory module, the flow generator and privileged encoder are trained first on general protein data and then on antibody–antigen complexes. Both stages mix teacher-mode and student-mode training of the shared flow generator. Teacher-mode passes encode the available privileged information and supply the resulting token/pair biases as additional inputs to the generator. Student-mode passes set both privileged-encoding outputs to zero, so the generator is also trained to operate without these additional inputs. The privileged encoder learns through teacher-mode training; it is trainable during these two stages. The student setting zeros the encoder's output biases rather than merely replacing its raw inputs with zeros.

This sequence of stages yields a trained flow generator and a learned privileged encoder before conditional-memory training begins. The encoder's zero initialization is therefore followed by learning, not by immediate freezing. The antibody–antigen stage establishes the antibody-specific privileged inputs described above.

**Memory-module training and deployment.** The subsequent adapter-only stage loads the trained backbone and privileged encoder and freezes both while optimizing the conditional-memory module. Teacher forward passes enable the privileged-information flag; their targets are detached before student distillation. At inference the privileged flag is disabled and the teacher branch does not run, matching the absence of privileged encodings in student-mode training. Native antibody coordinates and native CDR identities are therefore unnecessary for student inference; the native-derived epitope condition in the reported design benchmark is a separate, retained input.

**Distillation objectives.** Let sg denote stop-gradient and let subscripts $S, T$ denote student and teacher. Velocity distillation uses masked mean squared error for both the C$\alpha$ and local-latent velocity channels. Its coefficient is 0.5 (distillation weight 0.5 times velocity weight 1.0). With $u$ denoting the memory-gate summary and $\pi$ the record/null routing distribution, router distillation is

$$\mathcal{L}_{\text{route}} = \|u_S - \text{sg}(u_T)\|^2 + D_{\text{KL}}(\text{sg}(\pi_T) \,\|\, \pi_S)\,. \tag{13}$$

Its coefficients are 0.5 for VH–VL and 1.0 for VHH. Context distillation uses one minus cosine similarity between the student context and detached teacher context, with mean squared error as a fallback for a zero teacher context. Its detached route-agreement multiplier is

$$w_{\text{agree}} = \sum_i \min(\pi_{S,i}, \pi_{T,i}) \sqrt{\left(\sum_{i \neq \varnothing} \pi_{S,i}\right)\left(\sum_{i \neq \varnothing} \pi_{T,i}\right)}. \tag{14}$$

The context coefficient is 0.5. For the gated six-component pose vector $p$, pose distillation is

$$\mathcal{L}_{\text{pose}} = \frac{1}{6} \left\| \frac{p_S - \text{sg}(p_T)}{(0.35, 0.35, 0.35, 0.50, 0.50, 0.50)} \right\|_2^2, \tag{15}$$

where division is elementwise and the coefficient is 0.5. Velocity, context, and pose distillation use the time mask $t \geq 0.4$; memory-specific terms also respect the enabled-memory masks. These targets are available during training only.

**Optimization and reproducibility.** Memory-module training combines flow and structural losses with randomized-memory consistency, margins between conditional retrieval and its randomized or disabled controls, confidence contrast, routing contrast, and distillation. The supplementary training configuration specifies antibody-type-specific coefficients and masks for these terms. This stage uses Adam with learning rate $10^{-4}$, 8,000 steps, training seed 20260903, and EMA decay 0.999; these settings do not specify the preceding general-protein or antibody–antigen training stages. The token-memory and pose-head parameter counts are 7,299,412 and 344,622, respectively. Their sum is 7,644,034, bringing the total to 173,699,782 parameters. The bank is excluded from trainable parameter counts.

**Paired evaluation.** The paired memory-on/off comparison covers all 93 VH–VL and 50 VHH design targets using the same checkpoint, seed 42, and 400 integration steps. Table 5 reports randomized-retrieval and component controls on all 93 VH–VL targets, with population target-level standard deviations and mean-only SR. The exposure-only behavior described above follows from disabling effective memory updates; no separate full-test-set performance claim is made for that condition. These accuracy comparisons are distinct from sequence-refolding confidence.

# D Additional Experiments

The following experiments complement the main comparisons on 93 VH–VL antibodies and 50 VHH nanobodies. Unless noted otherwise, standard deviations describe variation across targets, not across random seeds. Experimental settings that differ between tables are stated explicitly. MOSAIC and MOSAIC (off) design results retain the native-derived epitope conditions described in Appendix B.1.

## D.1 Full-Test-Set Comparisons and Component Ablations

Figure 4 compares MOSAIC and MOSAIC (off) on the complete 93-target VH–VL and 50-target VHH design sets using seed 42 and 400 integration steps. Table 5 reports randomized-retrieval, value-shuffling, and component controls on all 93 VH–VL targets under the same checkpoint and generation settings. Its updated MOSAIC (off) row has DockQ 0.150 and SR 0.247. The updated ablation summary is reported separately from the original main-table and paired-figure results.

Within the VH–VL ablation, MOSAIC has DockQ 0.186 and SR 0.312. Removing pose correction or the CDR adapter gives DockQ 0.155 or 0.153, while randomized retrieval gives 0.153. Removing the support gate or retaining one record gives DockQ 0.181 or 0.182, with SR 0.290 for both. Exposure-only behavior is described as an implementation property in Appendix C, not as an additional full-test-set experimental result.

## D.2 Fixed-Key Memory-Content Controls

Table 7 reports three value-shuffling seeds on the 93-target VH–VL benchmark. The intervention replaces the retrieved values while retaining the keys and retrieval-support scores from conditional retrieval, thereby preserving the support-dependent pose-gate coefficient. The shuffling seeds 0–2

Table 7: Memory-content controls on 93 antibody targets.

| Method | AAR H3↑ | RMSD H3↓ | TM-Score↑ | lDDT↑ | DockQ↑ | iRMS↓ | LRMS↓ | SR↑ |
|---|---|---|---|---|---|---|---|---|
| MOSAIC (off) | $0.513^{0.195}$ | $2.338^{1.025}$ | $0.929^{0.046}$ | $0.782^{0.067}$ | $0.150^{0.151}$ | $8.810^{5.310}$ | $22.030^{11.420}$ | 0.247 |
| Randomized retrieval | $0.512^{0.197}$ | $2.344^{1.031}$ | $0.929^{0.047}$ | $0.783^{0.067}$ | $0.153^{0.154}$ | $8.790^{5.340}$ | $21.980^{11.380}$ | 0.258 |
| Shuffled values (seed 0) | $0.513^{0.194}$ | $2.331^{1.014}$ | $0.929^{0.046}$ | $0.786^{0.066}$ | $0.164^{0.157}$ | $8.760^{5.320}$ | $21.850^{11.350}$ | 0.269 |
| Shuffled values (seed 1) | $0.512^{0.196}$ | $2.339^{1.020}$ | $0.928^{0.047}$ | $0.785^{0.067}$ | $0.161^{0.155}$ | $8.800^{5.360}$ | $21.930^{11.410}$ | 0.258 |
| Shuffled values (seed 2) | $\underline{0.514}^{0.193}$ | $\underline{2.326}^{1.008}$ | $\underline{0.930}^{0.046}$ | $\underline{0.787}^{0.066}$ | $0.167^{0.160}$ | $8.730^{5.290}$ | $21.780^{11.300}$ | $\underline{0.280}$ |
| Shuffled CDR identities (fixed keys/layout) | $0.511^{0.196}$ | $2.329^{1.012}$ | $\underline{0.930}^{0.046}$ | $\underline{0.787}^{0.066}$ | $\underline{0.170}^{0.159}$ | $\underline{8.710}^{5.300}$ | $\underline{21.740}^{11.320}$ | $\underline{0.280}$ |
| MOSAIC | $\mathbf{0.516}^{0.192}$ | $\mathbf{2.310}^{0.991}$ | $\mathbf{0.931}^{0.045}$ | $\mathbf{0.790}^{0.065}$ | $\mathbf{0.186}^{0.163}$ | $\mathbf{8.650}^{5.250}$ | $\mathbf{21.520}^{11.180}$ | **0.312** |

Table 8: Parameter-matched baseline for antibody and nanobody design.

| Test set | AAR H3↑ | RMSD H3↓ | TM-Score↑ | lDDT↑ | DockQ↑ | iRMS↓ | LRMS↓ | SR↑ |
|---|---|---|---|---|---|---|---|---|
| VH–VL ($N = 93$) | $0.508^{0.198}$ | $2.378^{1.029}$ | $0.927^{0.048}$ | $0.784^{0.067}$ | $0.152^{0.163}$ | $8.845^{5.132}$ | $23.037^{15.431}$ | 0.247 |
| VHH ($N = 50$) | $0.452^{0.251}$ | $2.499^{1.312}$ | $0.906^{0.054}$ | $0.750^{0.069}$ | $0.120^{0.114}$ | $8.439^{3.484}$ | $20.431^{8.835}$ | 0.140 |

index content randomization, rather than independent generation seeds. A separate control shuffles CDR identities while holding keys and layout fixed. These controls assess sensitivity to memory content without retraining.

Value shuffling gives DockQ 0.164, 0.161, and 0.167, with SR 0.269, 0.258, and 0.280, respectively. MOSAIC has DockQ 0.186 and SR 0.312, corresponding to a mean-DockQ advantage of 0.019–0.025 over these controls. Shuffling CDR identities gives DockQ 0.170 and SR 0.280. The shuffled controls retain higher mean DockQ than MOSAIC (off) (0.150), consistent with retained key-conditioned information and other unchanged model components. These interventions support the relevance of correctly associated values, but can also introduce key–value mismatch; they do not isolate the performance of a separately trained key-only model.

In the updated ablation and memory-content tables, SR is rounded to the nearest feasible fraction with denominator 93 before reporting three decimals. The corresponding success counts are reconstructed from summary rates. Per-target outcomes are needed to verify these counts and compute paired confidence intervals; target-level standard deviations alone do not provide those intervals.

### D.3 Parameter-Matched Baseline

Table 8 reports a no-memory baseline constructed by increasing the generator width of MOSAIC (off) to match the parameter count of the full MOSAIC model. It uses the same training data and number of training steps as MOSAIC, and is evaluated with generation seed 42 and the same 400 integration steps on 93 VH–VL and 50 VHH targets. MOSAIC contains 166,055,748 backbone parameters and 7,644,034 memory-module parameters, totaling 173,699,782; memory-bank records are excluded from this count. Entries report means with target-level standard deviations in superscript, except SR, which is reported as a fraction. Unlike the inference-time interventions in Table 5, this experiment compares models with different parameterizations.

The baseline gives DockQ 0.152 on VH–VL and 0.120 on VHH, compared with the main-benchmark MOSAIC results of 0.187 and 0.156. Its SR is 0.247 on VH–VL and 0.140 on VHH, compared with 0.312 and 0.240 for MOSAIC. Together with the content-shuffling controls, these results support a benefit from conditional memory beyond widening the generator at matched parameter count and training and inference step budgets. The controls address complementary explanations: generator capacity and the association between retrieved keys and values.

### D.4 Multi-Seed Design Evaluation

Table 9 evaluates MOSAIC (off) and MOSAIC using generation seeds 1–5, fixed model weights, and 400 integration steps. All ten runs completed generation and evaluation on all 143 targets (93 VH–VL and 50 VHH), with no invalid cases. Each run generates one design per target with batch size one. Target-specific random seeds are assigned deterministically, with the same target-seeding namespace for both methods at each generation seed. This matched evaluation assesses robustness across

Table 9: Design robustness on antibody and nanobody benchmarks.

| Method | Seed | AAR H3↑ | RMSD H3↓ | TM-Score↑ | lDDT↑ | DockQ↑ | iRMS↓ | LRMS↓ | SR↑ |
|---|---|---|---|---|---|---|---|---|---|
| VH–VL antibodies ($N = 93$) | | | | | | | | | |
| MOSAIC (off) | 1 | $0.508^{0.192}$ | $2.387^{0.996}$ | $0.929^{0.046}$ | $0.783^{0.065}$ | $0.130^{0.135}$ | $9.065^{5.025}$ | $22.829^{10.711}$ | 0.183 |
| MOSAIC | 1 | $0.511^{0.186}$ | $2.363^{0.992}$ | $0.929^{0.046}$ | $0.789^{0.065}$ | $0.168^{0.138}$ | $9.047^{4.989}$ | $22.702^{10.798}$ | 0.237 |
| MOSAIC (off) | 2 | $0.499^{0.202}$ | $2.401^{1.076}$ | $0.929^{0.045}$ | $0.783^{0.065}$ | $0.144^{0.165}$ | $9.177^{5.169}$ | $24.278^{15.606}$ | 0.226 |
| MOSAIC | 2 | $0.497^{0.201}$ | $2.373^{1.051}$ | $0.929^{0.045}$ | $0.790^{0.065}$ | $0.184^{0.170}$ | $9.170^{5.296}$ | $23.956^{15.570}$ | 0.247 |
| MOSAIC (off) | 3 | $0.515^{0.199}$ | $2.343^{1.013}$ | $0.929^{0.046}$ | $0.784^{0.066}$ | $0.141^{0.151}$ | $8.974^{5.185}$ | $23.803^{14.909}$ | 0.247 |
| MOSAIC | 3 | $0.510^{0.202}$ | $2.361^{1.032}$ | $0.929^{0.046}$ | $0.789^{0.066}$ | $0.190^{0.165}$ | $8.537^{5.290}$ | $22.844^{15.175}$ | 0.312 |
| MOSAIC (off) | 4 | $0.509^{0.200}$ | $2.347^{1.018}$ | $0.929^{0.046}$ | $0.783^{0.066}$ | $0.138^{0.152}$ | $9.218^{5.204}$ | $24.244^{15.352}$ | 0.204 |
| MOSAIC | 4 | $0.506^{0.200}$ | $2.342^{1.007}$ | $0.929^{0.046}$ | $0.789^{0.065}$ | $0.179^{0.156}$ | $9.045^{5.327}$ | $23.766^{15.486}$ | 0.269 |
| MOSAIC (off) | 5 | $0.504^{0.196}$ | $2.377^{1.038}$ | $0.928^{0.046}$ | $0.783^{0.066}$ | $0.148^{0.160}$ | $8.848^{5.142}$ | $23.039^{15.446}$ | 0.226 |
| MOSAIC | 5 | $0.506^{0.195}$ | $2.378^{1.036}$ | $0.928^{0.046}$ | $0.789^{0.066}$ | $0.192^{0.167}$ | $8.813^{5.164}$ | $22.848^{15.709}$ | 0.301 |
| MOSAIC (off) | Across seeds | $0.507^{0.005}$ | $2.371^{0.023}$ | $0.929^{0.000}$ | $0.783^{0.000}$ | $0.140^{0.006}$ | $9.056^{0.135}$ | $23.639^{0.603}$ | 0.217 |
| MOSAIC | Across seeds | $0.506^{0.005}$ | $2.363^{0.012}$ | $0.929^{0.000}$ | $0.789^{0.000}$ | $0.183^{0.009}$ | $8.922^{0.225}$ | $23.223^{0.527}$ | 0.273 |
| VHH nanobodies ($N = 50$) | | | | | | | | | |
| MOSAIC (off) | 1 | $0.461^{0.251}$ | $2.500^{1.296}$ | $0.906^{0.055}$ | $0.749^{0.069}$ | $0.116^{0.103}$ | $8.144^{3.271}$ | $19.525^{7.974}$ | 0.100 |
| MOSAIC | 1 | $0.468^{0.244}$ | $2.547^{1.300}$ | $0.906^{0.055}$ | $0.757^{0.068}$ | $0.155^{0.117}$ | $8.136^{3.341}$ | $19.331^{7.854}$ | 0.180 |
| MOSAIC (off) | 2 | $0.458^{0.242}$ | $2.514^{1.288}$ | $0.905^{0.054}$ | $0.747^{0.068}$ | $0.095^{0.078}$ | $8.563^{3.165}$ | $21.439^{8.035}$ | 0.080 |
| MOSAIC | 2 | $0.451^{0.242}$ | $2.529^{1.278}$ | $0.905^{0.054}$ | $0.756^{0.067}$ | $0.130^{0.080}$ | $8.887^{3.272}$ | $21.892^{7.809}$ | 0.100 |
| MOSAIC (off) | 3 | $0.448^{0.245}$ | $2.471^{1.319}$ | $0.906^{0.054}$ | $0.748^{0.069}$ | $0.117^{0.106}$ | $8.241^{3.482}$ | $20.179^{8.675}$ | 0.200 |
| MOSAIC | 3 | $0.458^{0.243}$ | $2.476^{1.307}$ | $0.905^{0.054}$ | $0.756^{0.068}$ | $0.159^{0.119}$ | $8.203^{3.666}$ | $19.803^{8.538}$ | 0.220 |
| MOSAIC (off) | 4 | $0.452^{0.247}$ | $2.481^{1.260}$ | $0.906^{0.053}$ | $0.748^{0.069}$ | $0.113^{0.098}$ | $8.298^{3.524}$ | $20.247^{8.148}$ | 0.180 |
| MOSAIC | 4 | $0.462^{0.242}$ | $2.498^{1.291}$ | $0.906^{0.054}$ | $0.757^{0.069}$ | $0.159^{0.108}$ | $8.064^{3.504}$ | $19.871^{8.360}$ | 0.240 |
| MOSAIC (off) | 5 | $0.449^{0.249}$ | $2.506^{1.315}$ | $0.906^{0.053}$ | $0.749^{0.069}$ | $0.119^{0.115}$ | $8.438^{3.494}$ | $20.450^{8.925}$ | 0.140 |
| MOSAIC | 5 | $0.448^{0.250}$ | $2.482^{1.274}$ | $0.906^{0.053}$ | $0.756^{0.068}$ | $0.151^{0.122}$ | $8.649^{3.409}$ | $20.927^{8.637}$ | 0.180 |
| MOSAIC (off) | Across seeds | $0.454^{0.005}$ | $2.494^{0.016}$ | $0.906^{0.000}$ | $0.748^{0.001}$ | $0.112^{0.008}$ | $8.337^{0.148}$ | $20.368^{0.619}$ | 0.140 |
| MOSAIC | Across seeds | $0.458^{0.008}$ | $2.506^{0.027}$ | $0.906^{0.000}$ | $0.756^{0.001}$ | $0.151^{0.011}$ | $8.388^{0.322}$ | $20.365^{0.925}$ | 0.184 |

generation seeds 1–5, complementing the seed-42 comparisons. H3 RMSD uses self-alignment, and SR is the fraction of targets with DockQ $\geq 0.23$.

Individual-seed rows report target-level means and population standard deviations. The summary rows report the mean and population standard deviation of the five seed-level means, rather than treating repeated designs of the same targets as independent test objects. SR is shown as a mean only in both types of rows.

MOSAIC has higher mean DockQ, lDDT, and SR than MOSAIC (off) at every seed for both antibody types. Across seeds, VH–VL DockQ increases from $0.140 \pm 0.006$ to $0.183 \pm 0.009$, with mean SR increasing from 0.217 to 0.273. For VHH, DockQ increases from $0.112 \pm 0.008$ to $0.151 \pm 0.011$, and mean SR increases from 0.140 to 0.184. Other metrics show mixed changes: mean H3 RMSD decreases slightly for VH–VL but increases for VHH, and VHH iRMS does not improve on average. These results support the consistency of the docking-score and success-rate gains across generation seeds.

### D.5 Design Performance across Antigen Similarity Strata

Each test antigen is assigned by the highest identity returned from an MMseqs2 local-alignment search against 12,340 valid training-antigen profiles. Search uses coverage mode 2, at least 50% query coverage, and one retained match per query. The identity intervals are $\leq 30\%$, $(30, 40]\%$, $(40, 60]\%$, and $(60, 80]\%$. Their sizes are 30, 6, 8, and 49 for VH–VL, and 35, 6, 5, and 4 for VHH. In Table 10, $n$ denotes targets per stratum; bold marks the better displayed mean within each stratum, and displayed ties are not highlighted. Both methods use seed 42 and 400 integration steps here.

MOSAIC has higher mean DockQ in every stratum for both antibody types (Table 10). In the $\leq 30\%$ groups, DockQ increases from 0.196 to 0.224 for VH–VL and from 0.114 to 0.145 for VHH. SR remains 0.400 for VH–VL and increases from 0.171 to 0.200 for VHH. The remaining metrics show mixed changes. These strata describe sensitivity to antigen similarity rather than a sequence-disjoint evaluation, and groups containing only 4–8 targets warrant caution.

Table 10: Design performance across training-antigen similarity strata.

| Identity | $n$ | Method | AAR H3↑ | RMSD H3↓ | TM-Score↑ | lDDT↑ | DockQ↑ | iRMS↓ | LRMS↓ | SR↑ |
|---|---|---|---|---|---|---|---|---|---|---|
| VH–VL antibodies ($N = 93$) | | | | | | | | | | |
| $\leq 30\%$ | 30 | MOSAIC (off) | $0.524^{0.200}$ | $2.213^{1.212}$ | $0.929^{0.069}$ | $0.797^{0.085}$ | $0.196^{0.180}$ | $\mathbf{7.849}^{5.724}$ | $\mathbf{20.424}^{12.725}$ | 0.400 |
| | | MOSAIC | $\mathbf{0.529}^{0.203}$ | $\mathbf{2.190}^{1.157}$ | $\mathbf{0.930}^{0.070}$ | $\mathbf{0.804}^{0.086}$ | $\mathbf{0.224}^{0.195}$ | $7.913^{5.799}$ | $20.606^{12.445}$ | 0.400 |
| $(30, 40]\%$ | 6 | MOSAIC (off) | $\mathbf{0.413}^{0.115}$ | $\mathbf{2.229}^{0.756}$ | $0.929^{0.035}$ | $0.772^{0.060}$ | $0.132^{0.104}$ | $6.600^{2.456}$ | $21.643^{9.668}$ | 0.167 |
| | | MOSAIC | $0.391^{0.120}$ | $2.257^{0.722}$ | $0.929^{0.034}$ | $\mathbf{0.778}^{0.061}$ | $\mathbf{0.203}^{0.131}$ | $\mathbf{6.532}^{2.727}$ | $\mathbf{21.586}^{10.610}$ | **0.333** |
| $(40, 60]\%$ | 8 | MOSAIC (off) | $\mathbf{0.559}^{0.200}$ | $\mathbf{2.069}^{0.895}$ | $\mathbf{0.947}^{0.018}$ | $0.804^{0.031}$ | $0.176^{0.198}$ | $\mathbf{8.638}^{5.462}$ | $24.472^{15.362}$ | 0.250 |
| | | MOSAIC | $0.551^{0.206}$ | $2.085^{0.913}$ | $0.946^{0.019}$ | $\mathbf{0.808}^{0.033}$ | $\mathbf{0.211}^{0.212}$ | $8.691^{5.474}$ | $\mathbf{24.163}^{15.623}$ | 0.250 |
| $(60, 80]\%$ | 49 | MOSAIC (off) | $\mathbf{0.517}^{0.192}$ | $2.449^{0.908}$ | $0.927^{0.027}$ | $0.774^{0.056}$ | $0.123^{0.124}$ | $9.549^{4.990}$ | $22.315^{9.500}$ | 0.184 |
| | | MOSAIC | $0.516^{0.187}$ | $\mathbf{2.437}^{0.892}$ | $\mathbf{0.928}^{0.027}$ | $\mathbf{0.780}^{0.054}$ | $\mathbf{0.159}^{0.124}$ | $\mathbf{9.518}^{5.152}$ | $\mathbf{21.869}^{9.510}$ | **0.265** |
| VHH nanobodies ($N = 50$) | | | | | | | | | | |
| $\leq 30\%$ | 35 | MOSAIC (off) | $0.436^{0.259}$ | $\mathbf{2.330}^{1.219}$ | $0.904^{0.056}$ | $0.739^{0.068}$ | $0.114^{0.109}$ | $\mathbf{8.599}^{3.358}$ | $\mathbf{20.911}^{8.809}$ | 0.171 |
| | | MOSAIC | $\mathbf{0.443}^{0.256}$ | $2.346^{1.278}$ | $0.904^{0.056}$ | $\mathbf{0.747}^{0.067}$ | $\mathbf{0.145}^{0.098}$ | $8.754^{3.297}$ | $21.256^{8.418}$ | **0.200** |
| $(30, 40]\%$ | 6 | MOSAIC (off) | $0.418^{0.207}$ | $3.299^{1.465}$ | $0.902^{0.061}$ | $0.760^{0.087}$ | $0.124^{0.098}$ | $8.379^{3.176}$ | $19.121^{6.250}$ | 0.333 |
| | | MOSAIC | $0.418^{0.207}$ | $\mathbf{3.267}^{1.485}$ | $\mathbf{0.903}^{0.062}$ | $\mathbf{0.769}^{0.088}$ | $\mathbf{0.163}^{0.096}$ | $\mathbf{7.516}^{3.474}$ | $\mathbf{16.657}^{6.895}$ | 0.333 |
| $(40, 60]\%$ | 5 | MOSAIC (off) | $\mathbf{0.447}^{0.069}$ | $\mathbf{2.550}^{0.356}$ | $0.919^{0.031}$ | $0.773^{0.038}$ | $0.191^{0.115}$ | $\mathbf{5.597}^{2.912}$ | $\mathbf{16.117}^{8.612}$ | 0.400 |
| | | MOSAIC | $0.429^{0.058}$ | $2.552^{0.371}$ | $\mathbf{0.921}^{0.029}$ | $\mathbf{0.781}^{0.035}$ | $\mathbf{0.211}^{0.134}$ | $6.077^{2.767}$ | $17.176^{8.268}$ | 0.400 |
| $(60, 80]\%$ | 4 | MOSAIC (off) | $0.667^{0.272}$ | $\mathbf{3.592}^{1.341}$ | $0.891^{0.038}$ | $0.763^{0.039}$ | $0.096^{0.110}$ | $9.718^{4.262}$ | $24.993^{11.265}$ | 0.250 |
| | | MOSAIC | $0.667^{0.272}$ | $3.640^{1.403}$ | $0.891^{0.039}$ | $\mathbf{0.772}^{0.037}$ | $\mathbf{0.173}^{0.155}$ | $\mathbf{9.470}^{4.452}$ | $\mathbf{24.206}^{11.682}$ | 0.250 |

Table 11: Integration-step ablation on 93 antibody targets.

| Steps | AAR H3↑ | RMSD H3↓ | TM-Score↑ | lDDT↑ | DockQ↑ | iRMS↓ | LRMS↓ | SR↑ |
|---|---|---|---|---|---|---|---|---|
| 20 | $0.507^{0.197}$ | $6.919^{1.065}$ | $0.257^{0.036}$ | $0.147^{0.033}$ | $0.075^{0.046}$ | $13.915^{2.799}$ | $35.811^{16.104}$ | 0.000 |
| 50 | $0.502^{0.203}$ | $3.869^{0.828}$ | $0.487^{0.095}$ | $0.296^{0.042}$ | $0.119^{0.081}$ | $10.591^{3.920}$ | $27.808^{15.310}$ | 0.075 |
| 100 | $0.510^{0.201}$ | $2.511^{0.933}$ | $0.853^{0.039}$ | $0.562^{0.032}$ | $0.181^{0.145}$ | $8.903^{4.901}$ | $22.564^{13.950}$ | 0.290 |
| 200 | $0.510^{0.197}$ | $2.337^{0.992}$ | $0.929^{0.045}$ | $0.789^{0.065}$ | $0.192^{0.158}$ | $8.541^{5.178}$ | $21.564^{11.584}$ | 0.312 |
| 400 | $0.515^{0.193}$ | $2.315^{0.987}$ | $0.930^{0.046}$ | $0.790^{0.066}$ | $0.187^{0.162}$ | $8.737^{5.361}$ | $21.641^{11.260}$ | 0.312 |
| 800 | $0.510^{0.199}$ | $2.332^{1.039}$ | $0.929^{0.046}$ | $0.790^{0.066}$ | $0.177^{0.153}$ | $9.156^{5.337}$ | $22.639^{11.969}$ | 0.269 |
| 1000 | $0.505^{0.200}$ | $2.329^{1.041}$ | $0.929^{0.046}$ | $0.790^{0.066}$ | $0.188^{0.161}$ | $8.734^{5.340}$ | $21.008^{10.820}$ | 0.301 |

### D.6 Number of Integration Steps

Table 11 compares 20, 50, 100, 200, 400, 800, and 1000 integration steps on 93 VH–VL targets using seed 42.

Increasing the budget from 20 to 200 steps improves DockQ from 0.075 to 0.192 and reduces self-aligned H3 RMSD from 6.919 to 2.337 Å. Larger budgets do not yield monotonic docking gains: 800 and 1000 steps give DockQ 0.177 and 0.188. The main-table 400-step reference replaces the separate single-record retrieval result in the original comparison. Its retrieval configuration has not been verified as matched to the other rows, so differences involving this reference should not be attributed solely to integration-step count.

### D.7 Autoencoder Reconstruction

Table 12 compares an AFDB-pretrained La-Proteina autoencoder with its antibody-finetuned counterpart on the same 143 held-out complexes (93 VH–VL antibodies and 50 VHH nanobodies). Entries with superscripts report the mean and population standard deviation across targets; scRMSD is in Å, and atom coverage is reported as a percentage. Bold marks the better result. Antibody finetuning improves CDR and full-antibody sequence recovery, reduces reconstruction scRMSD, and reaches complete atom coverage. Reconstruction scRMSD is distinct from the interface scRMSD used in the design comparisons.

### D.8 Inference Cost

Table 13 reports end-to-end MOSAIC sampling times on a single NVIDIA H20 GPU with batch size 4, including memory retrieval and memory-module computation. The cost increases from 1.963 seconds per sample at 20 integration steps to 21.546 seconds at 400 steps.

Table 12: Autoencoder reconstruction on 143 held-out complexes.

| 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 13: End-to-end sampling cost across integration budgets.

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

### D.9 Structure-Prediction Protocol

Prediction uses the complete antibody sequence. Methods marked † receive the antigen structure, $*$ denotes an epitope condition, and ‡ denotes MSA input. tFold-Ag uses antigen-only UniRef50 MSAs. AlphaFold 3, Boltz-2, and Protenix-v2 use per-chain unpaired MSAs without templates. For VH–VL, the predicted antigen is aligned to the native antigen, the same transform is applied to the antibody, and the predicted antigen is then replaced by the native antigen for evaluation. For VHH, the predicted antigen is retained during scoring. VHH MSAs cover the observed antibody and antigen chains and contain no paired alignments.

MOSAIC and MOSAIC (off) use 400 integration steps. Prediction AAR is omitted because the antibody sequence is supplied.

### D.10 Design Metrics and Interface Evaluation

Design AAR is the fraction of recovered native residues in the indicated CDR. MOSAIC CDR backbone RMSD is evaluated after fitting the corresponding loop; the VHH Germinal H3 value also uses self-alignment. Baseline metrics follow their respective evaluation protocols. Interface scRMSD measures side-chain geometry after fitting common antibody backbone atoms. Buried-surface-area error, $|\Delta\mathrm{BSA}|$, is the absolute difference between generated and native complexes. Interface $\Delta G$ is reported as a median in Rosetta energy units (REU), not as a measured binding free energy. The VH–VL tables report $\Delta G$ and $|\Delta\mathrm{BSA}|$ to one decimal place. All other design entries use three decimal places. VH–VL clash rate and polar-contact recovery are fractions; the VHH interface table reports them as percentages. Buried-surface-area differences are in $\mathrm{Å}^2$.

For the main VH–VL interface comparison, IgGM, BoltzGen, MOSAIC (off), and MOSAIC use fixed-backbone Rosetta side-chain repacking. For the VHH interface comparison, IgGM and MOSAIC use fixed-backbone, fixed-sequence antibody repacking while retaining the antigen. Energy estimates use coordinate-constrained FastRelax and InterfaceAnalyzer.

### D.11 Initialization-Controlled Design Comparison

DiffAb and MEAN use framework structures initialized by IgFold or AlphaFold 3. The IgFold variants use HDOCKlite for docking. In contrast, MOSAIC (off) and MOSAIC receive framework sequences without native framework coordinates. MEAN designs CDR-H3 only, so other CDR entries are unavailable.

### D.12 Nanobody Interface Quality

Table 15 reports interface geometry and energy proxies for the 50 VHH designs. These measurements complement backbone and docking accuracy and do not establish thermodynamic binding affinity.

### D.13 Sequence-Refolding Confidence

Protenix-v2 refolds the supplied design sequences without regenerating them. Both antibody types use independent per-chain unpaired MSAs from local UniRef50, with no paired MSA or templates,

Table 14: Antibody design with alternative framework initializations.

| Model | | DiffAb (IgFold) | MEAN (IgFold) | DiffAb (AF3) | MEAN (AF3) | MOSAIC (off) | MOSAIC |
|---|---|---|---|---|---|---|---|
| AAR↑ | L1 | $0.500^{0.247}$ | – | $0.462^{0.249}$ | – | $\mathbf{0.693}^{0.196}$ | $\underline{0.686}^{0.195}$ |
| | L2 | $0.489^{0.301}$ | – | $0.430^{0.323}$ | – | $\underline{0.821}^{0.252}$ | $\mathbf{0.824}^{0.261}$ |
| | L3 | $0.406^{0.152}$ | – | $0.369^{0.177}$ | – | $\mathbf{0.727}^{0.246}$ | $\underline{0.716}^{0.253}$ |
| | H1 | $0.625^{0.168}$ | – | $0.513^{0.177}$ | – | $\underline{0.696}^{0.194}$ | $\mathbf{0.699}^{0.181}$ |
| | H2 | $0.271^{0.142}$ | – | $0.255^{0.136}$ | – | $\mathbf{0.667}^{0.226}$ | $\underline{0.660}^{0.228}$ |
| | H3 | $0.256^{0.107}$ | $0.356^{0.123}$ | $0.253^{0.111}$ | $0.343^{0.128}$ | $\underline{0.512}^{0.192}$ | $\mathbf{0.515}^{0.193}$ |
| RMSD↓ | L1 | $1.005^{0.922}$ | – | $1.245^{1.032}$ | – | $\mathbf{0.629}^{0.665}$ | $\underline{0.632}^{0.662}$ |
| | L2 | $0.192^{0.179}$ | – | $0.213^{0.180}$ | – | $\mathbf{0.138}^{0.177}$ | $\underline{0.152}^{0.188}$ |
| | L3 | $0.892^{0.873}$ | – | $1.038^{0.842}$ | – | $\mathbf{0.728}^{0.753}$ | $\underline{0.748}^{0.765}$ |
| | H1 | $0.900^{0.749}$ | – | $1.664^{2.724}$ | – | $\mathbf{0.772}^{0.571}$ | $\underline{0.774}^{0.566}$ |
| | H2 | $0.705^{0.604}$ | – | $1.129^{1.798}$ | – | $\mathbf{0.538}^{0.483}$ | $\underline{0.545}^{0.496}$ |
| | H3 | $2.667^{1.245}$ | $\mathbf{2.262}^{0.894}$ | $2.951^{2.700}$ | $2.657^{0.910}$ | $2.335^{1.035}$ | $\underline{2.315}^{0.987}$ |
| Structure↑ | TMscore | $0.910^{0.112}$ | $0.914^{0.113}$ | $0.867^{0.099}$ | $0.851^{0.096}$ | $\underline{0.929}^{0.046}$ | $\mathbf{0.930}^{0.046}$ |
| | lDDT | $0.337^{0.044}$ | $0.342^{0.044}$ | $0.641^{0.097}$ | $0.637^{0.097}$ | $\underline{0.783}^{0.066}$ | $\mathbf{0.790}^{0.067}$ |
| Side chain↓ | Clash rate | $\underline{0.002}^{0.002}$ | $\mathbf{0.002}^{0.002}$ | $0.018^{0.020}$ | $0.003^{0.007}$ | $0.018^{0.017}$ | $0.017^{0.016}$ |
| | scRMSD | $\underline{13.320}^{8.284}$ | $17.572^{9.235}$ | $17.148^{9.724}$ | $19.907^{7.626}$ | $14.227^{7.308}$ | $\mathbf{12.950}^{6.325}$ |
| Interface | Polar contact↑ | $0.000^{0.000}$ | $0.000^{0.000}$ | $0.002^{0.017}$ | $0.000^{0.000}$ | $\underline{0.014}^{0.038}$ | $\mathbf{0.032}^{0.070}$ |
| | $\lvert\Delta\mathrm{BSA}\rvert$↓ | $645.1^{439.3}$ | $639.5^{432.6}$ | $705.2^{645.6}$ | $762.5^{768.3}$ | $\underline{248.9}^{151.3}$ | $\mathbf{177.3}^{101.6}$ |
| | Median $\Delta G$↓ | $\mathbf{-53.1}$ | $-39.1$ | $-44.1$ | $-43.8$ | $-32.2$ | $\underline{-48.6}$ |
| Docking | DockQ↑ | $0.011^{0.008}$ | $0.014^{0.014}$ | $0.019^{0.017}$ | $0.019^{0.017}$ | $\underline{0.151}^{0.173}$ | $\mathbf{0.187}^{0.162}$ |
| | iRMS↓ | $21.055^{9.076}$ | $19.894^{9.811}$ | $16.881^{5.240}$ | $16.409^{5.238}$ | $\underline{8.993}^{5.383}$ | $\mathbf{8.737}^{5.361}$ |
| | LRMS↓ | $61.100^{20.423}$ | $60.930^{22.769}$ | $50.232^{19.456}$ | $50.819^{19.624}$ | $\underline{23.825}^{15.166}$ | $\mathbf{21.641}^{11.260}$ |
| | SR↑ | 0.000 | 0.000 | 0.000 | 0.000 | $\underline{0.247}$ | **0.312** |

Table 15: Interface quality of nanobody designs on 50 targets.

| Method | Clash rate (%)↓ | Interface scRMSD (Å)↓ | Polar-contact recovery (%)↑ | Median $\Delta G$ (REU)↓ | Mean $\lvert\Delta\mathrm{BSA}\rvert$ (Å$^2$)↓ |
|---|---|---|---|---|---|
| DiffAb (AF3) | $\underline{0.390}^{0.810}$ | $\underline{15.694}^{7.186}$ | $0.000^{0.000}$ | $-18.915$ | $303.274^{275.900}$ |
| IgGM | $1.280^{1.560}$ | $16.699^{6.952}$ | $0.120^{0.820}$ | $-12.957$ | $212.271^{154.013}$ |
| BoltzGen | $1.720^{1.400}$ | $15.722^{6.952}$ | $\underline{1.290}^{4.220}$ | $-7.561$ | $200.270^{146.672}$ |
| Germinal | $1.900^{1.800}$ | $18.539^{3.722}$ | $0.100^{0.700}$ | $-17.464$ | $292.042^{237.260}$ |
| MOSAIC (off) | $1.920^{2.030}$ | $16.001^{8.208}$ | $0.920^{2.660}$ | $\underline{-21.193}$ | $\mathbf{130.714}^{104.794}$ |
| MOSAIC | $\mathbf{0.150}^{0.300}$ | $\mathbf{13.104}^{7.975}$ | $\mathbf{1.400}^{5.250}$ | $\mathbf{-28.346}$ | $\underline{160.151}^{120.551}$ |

seed 5, one sample, three Pairformer cycles, and 20 diffusion steps. The VHH search uses the January 2026 UniRef50 release. All methods within each table use the same refolding settings.

We report global ipTM, pLDDT, pTM, ipSAE, pDockQ, and ranking score. pLDDT uses the 0–100 scale; the other reported scores use the 0–1 scale. For VH–VL, ipSAE and pDockQ are the maximum antibody–antigen scores across heavy–antigen and light–antigen chain pairs. For VHH, they use the heavy–antigen pair. Statistics cover all 93 VH–VL or 50 VHH targets. Rankings use unrounded means. These confidence scores do not measure agreement with the native complex; the separately reported DockQ evaluates each Protenix-v2 refolded complex against its native reference.

The complete VH–VL refolding results are reported in the main text (Table 3); the VHH results are given in Table 16.

### D.14 Structural Case Selection and Alignment

The four complexes in Figure 3 correspond to 9jme antibody 2, 9leo antibody 4, 9riu nanobody 3, and 9yc3 nanobody 1. Each predicted antigen is rigidly aligned to its native counterpart, and the transform is applied to the entire predicted complex without a separate antibody fit or antigen replacement. DockQ uses the same native reference for all methods. The margins for panels A–D, computed before rounding, are 0.560, 0.466, 0.264, and 0.281.

Table 16: Protenix-v2 refolding confidence and docking accuracy on 50 nanobodies.

| 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 | $0.347^{0.152}$ | $82.795^{6.891}$ | $0.700^{0.096}$ | $0.043^{0.121}$ | $0.174^{0.083}$ | $0.418^{0.134}$ | $0.082^{0.063}$ |
| Germinal | $0.330^{0.133}$ | $78.349^{7.575}$ | $0.691^{0.103}$ | $0.029^{0.120}$ | $0.145^{0.091}$ | $0.402^{0.117}$ | $0.078^{0.059}$ |
| MOSAIC (off) | $\underline{0.370}^{0.163}$ | $\underline{82.865}^{6.870}$ | $\underline{0.710}^{0.094}$ | $0.057^{0.153}$ | $0.189^{0.093}$ | $\underline{0.438}^{0.143}$ | $\underline{0.094}^{0.068}$ |
| MOSAIC | $\mathbf{0.435}^{0.191}$ | $\mathbf{84.102}^{7.375}$ | $\mathbf{0.726}^{0.107}$ | $\mathbf{0.098}^{0.203}$ | $\mathbf{0.217}^{0.096}$ | $\mathbf{0.494}^{0.169}$ | $\mathbf{0.169}^{0.152}$ |