# AbGaze: Attentive Geometric Representation Learning for End-to-End Antibody Design

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

## Abstract

Computational antibody design requires representations that capture the geometric patterns underlying antigen–antibody interactions, yet existing approaches often rely on scalar distances or surface-intrinsic features, leaving cross-molecular geometry largely implicit. We present **AbGaze**, an end-to-end antibody design framework based on attentive geometric representation learning, which encodes distance, spatial direction, and surface-normal orientation of antigen surfaces relative to antibody-residue local frames, and adaptively aggregates these geometric interactions according to their interfacial context. The learned interaction representation is shared across multi-CDR co-design, complex structure prediction, and affinity optimization, with local-frame geometric supervision further constraining the representation. AbGaze outperforms prior methods across all three tasks: relative to the second-best method, it improves amino-acid recovery by 7.1% and reduces structural error by 14.9% on average over the six CDRs, improves interface docking quality (DockQ) by 6.6%, and raises the affinity improvement rate (IMP) by 32.5%.

## 1 Introduction

Antibodies have been widely applied in disease therapy, diagnostics, and biomedical research due to their specific molecular-recognition capabilities (Chan et al., 2025). Antibodies recognize and bind antigenic epitopes through the complementarity-determining regions (CDRs) within their variable domains, while the high sequence and conformational diversity of CDRs enables diverse antigen–antibody binding modes (Sela-Culang et al., 2013; North et al., 2011; Weitzner et al., 2015). Therefore, computational modeling of antibodies conditioned on antigen structures requires accurate characterization of antigen–antibody interactions, where modeling the local spatial relationships at the interface is central to the problem.

Existing methods have evolved from step-by-step pipelines toward end-to-end generation (Jin et al., 2022b). Early approaches decomposed antibody modeling into sequential stages such as complex structure prediction, CDR design, and side-chain assembly, where intermediate errors could propagate across the pipeline (Jin et al., 2022a; Kong et al., 2022; Luo et al., 2022; Sircar & Gray, 2010). Recent end-to-end approaches instead jointly model antibody sequences and structures conditioned on antigen information, enabling tighter sequence–structure co-generation within a unified framework (Kong et al., 2023; Wang et al., 2025; Tan et al., 2025; Wang et al., 2026). In parallel, broader studies of protein interactions have explored interface representations learned from the geometric and chemical features of molecular surfaces (Sverrisson et al., 2021). Methods such as MaSIF learn interaction fingerprints that capture patterns associated with molecular recognition and surface complementarity, applying them to interaction-site identification and binder design (Gainza et al., 2020; 2023). These studies demonstrate that local geometric and chemical patterns at molecular interfaces provide informative cues for interaction modeling.

Despite these advances, existing interface representations can capture interaction patterns at the molecular-surface or atomic level, but provide limited explicit characterization of the cross-molecular spatial relationships that depend on local conformations at antigen–antibody interfaces. This challenge is closely related to a problem in protein–protein interaction modeling: how to represent the local geometry of intermolecular contacts. Intermolecular binding is not determined by spatial proximity alone; the distance, direction, and angle to the surface normal at which a



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.

molecule approaches the opposing surface all affect local packing and complementarity (Kuroda & Gray, 2016), while the contribution of contacts can also depend on their surrounding interface context (Lawrence & Colman, 1993; McCoy et al., 1997; Yang et al., 2003). Accordingly, interaction modeling should explicitly characterize the relative spatial relationships between a paratope conformation and the antigen surface. An effective interaction representation should therefore capture three aspects: spatial proximity between local structures, their relative orientations, and the dependence of local interaction contributions on the surrounding interface context. Distance-based representations capture spatial proximity but reduce each contact to a scalar, making configurations with similar distances but different directions indistinguishable; representations based on surface geometry and chemical features capture local surface patterns, but are not expressed relative to antibody conformations. Because these interactions are defined relative to local antibody conformations, antibody-residue local frames also provide a natural basis for structural supervision: frame-aligned point error (FAPE) (Jumper et al., 2021) and dihedral-angle constraints couple atoms with residue frames, supplying global structural supervision expressed in the same local coordinates. Given the large space of CDR conformations and binding configurations and the limited number of experimentally resolved antibody–antigen complexes, requiring a model to rediscover these interaction regularities from data can increase the burden on representation learning.

Motivated by these observations, we propose **AbGaze**, a unified framework for antigen–antibody interaction modeling based on orientation-aware geometry and adaptive attention. Specifically, AbGaze constructs a local interface representation using antibody-residue coordinate frames to explicitly encode distance, spatial direction, and surface-normal information. On top of this representation, an adaptive atom–surface attention scheme, conditioned on interface geometry and fused with contextual node states, learns key local interactions and directs coordinate updates toward the attended surface region, optimized under FAPE and dihedral-angle supervision. This orientation-aware representation acts as a binding prior that unifies multi-CDR co-design, complex structure prediction, and affinity optimization under a single masked-prediction framework, where stochastic masked decoding further yields multiple plausible sequence–structure realizations.

To summarize, our main contributions are as follows:

(i) We propose an orientation-aware local interface representation using antibody-residue coordinate frames to explicitly capture relative distance, spatial direction, and surface normals of the antigen surface.

(ii) We design an adaptive atom–surface attention mechanism conditioned on interface geometry to direct spatial updates and enable geometrically constrained interaction learning under FAPE and dihedral supervision.

(iii) We establish a binding prior that unifies multi-CDR co-design, structure prediction, and affinity optimization within a single masked-prediction framework, enabling multi-candidate sequence–structure decoding.

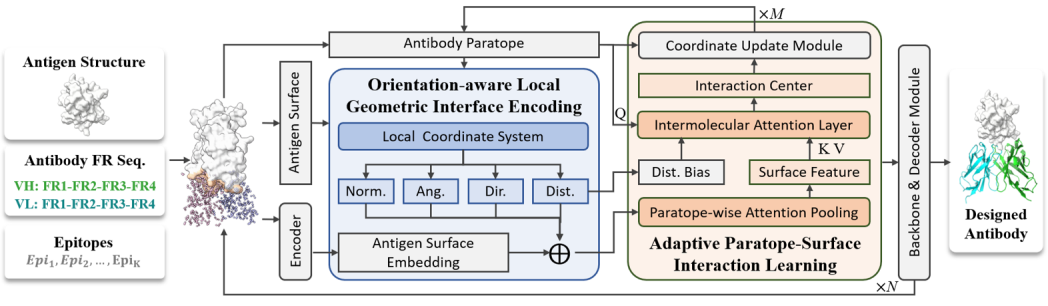



Figure 1: Overview of AbGaze. We propose AbGaze, a unified framework for antigen–antibody interaction modeling based on orientation-aware geometry and adaptive attention. The framework iteratively updates the antibody conformation by explicitly encoding the relative spatial relationships between the paratope local frames and the antigen surface, while an adaptive attention mechanism learns contextual interactions and guides structural refinement.

# 2 Related Work

## 2.1 Antigen-Conditioned Antibody Modeling

Antigen-conditioned antibody modeling has evolved from localized CDR design toward end-to-end sequence–structure co-design. Early landmark approaches, including HERN, MEAN, and DiffAb, investigated CDR generation by coupling sequence updates with backbone coordinates (Jin et al., 2022a; Kong et al., 2022; Luo et al., 2022). Subsequent end-to-end architectures, such as dyMEAN, established scalable full-atom graph dynamics and structural refinement pipelines (Kong et al., 2023), AbDiffuser paired full-atom diffusion generation with experimental antibody validation (Martinkus et al., 2023), and IgGM broadened generative modeling across diverse functional antibody and nanobody design regimes (Wang et al., 2025). More recently, flow-matching formulations have extended multi-state modeling: dyAb accommodates conformational transitions in target antigens (Tan et al., 2025), and AbFlow leverages surface-guided interaction dynamics for end-to-end design (Wang et al., 2026), alongside geometry-aware frameworks like GeoGAD (Wei et al., 2026).

## 2.2 Molecular Interface Representation

Characterizing molecular interaction interfaces requires capturing spatial proximity, local orientation, and surface complementarity. Surface-centric methods such as MaSIF and MaSIF-seed showed that geometric fingerprints and intrinsic surface curvature are informative for binding specificity (Gainza et al., 2020; 2023). This paradigm has been extended to pairwise point-cloud matching in PInet (Dai & Bailey-Kellogg, 2021), structural environment filtering in PeSTo (Krapp et al., 2023), and target-conditioned binder synthesis in ProBID-Net (Chen et al., 2024).

However, integrating such fine-grained surface topography into generative antibody architectures remains non-trivial. Existing generative models typically face a dichotomy in interface representation: they either operate at the residue level via local coordinate frames (Luo et al., 2022; Wei et al., 2026), thereby coarse-graining microscopic atomic contact geometry, or include atom-to-surface displacement vectors without decomposing them into residue-local reference frames (Wang et al., 2026). In contrast, AbGaze bridges fine-grained surface topography with frame-aligned local representations. By directly decomposing both antibody-frame approach directions and surface-normal incidence relationships into each residue's local frame, AbGaze aggregates these geometric features through a dedicated two-level surface-attention module for interaction learning and geometric supervision.

# 3 Methodology

## 3.1 Overview and Problem Setting

We represent an antigen–antibody complex as a residue graph (Jing et al., 2021) with amino-acid types and full-atom coordinates. Context and interface edges are constructed within and between the two molecules, respectively. For each epitope residue, we construct an MSMS molecular surface (Sanner et al., 1996) with vertex normals, keeping vertices within 10 Å of any epitope atom.

Given the antigen sequence and structure and an antibody with masked target regions, AbGaze jointly generates antibody sequences and full-atom structures using a shared interface representation and atom–surface interaction learning. We adopt the general graph-based modeling and decoding framework of Kong et al. (2023).

## 3.2 Orientation-Aware Local Interface Representation

To capture microscopic interactions at the antibody–antigen interface, we construct an orientation-aware representation. For each antibody–antigen residue pair $e = (i, b_e)$, we relate every heavy atom of antibody residue $b_e$ to every vertex of the paired antigen surface patch $\mathcal{V}_i$. Both antibody and antigen heavy atoms are mapped into a standardized 14-slot vocabulary schema $p \in \{1, \ldots, 14\}$ (slots 1–4 hold the backbone atoms; slots 5–14 hold side-chain heavy atoms in a fixed per-type order, with unused slots padded and masked).

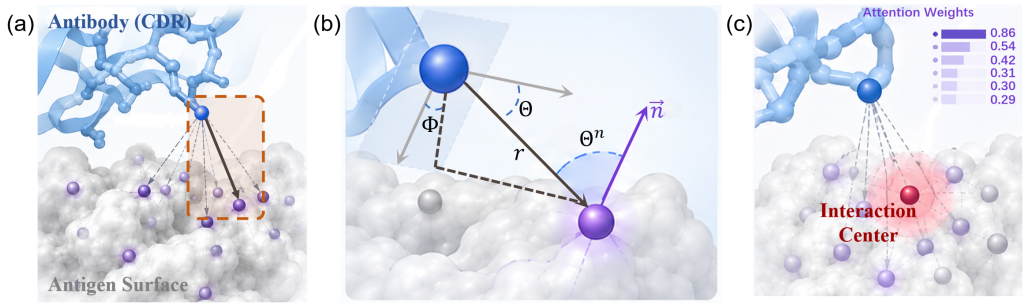



Figure 2: **Overview of local interface representation and adaptive attention. (a)** Global interaction view between the antibody (CDR) residue and the antigen surface mesh. **(b)** Orientation-aware geometric encoding expressed in the antibody-residue local frame, capturing distance ($r$), spatial direction ($\Theta, \Phi$), and surface normal ($\Theta^n, \vec{n}$). **(c)** Adaptive atom–surface attention weighting used to dynamically locate the interaction center on the antigen surface.

**Residue local frame.** For an antibody residue $a$, we construct a local coordinate frame $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ using its backbone heavy atoms $\mathbf{x}_{a,\mathrm{N}}$, $\mathbf{x}_{a,\mathrm{C}_\alpha}$, and $\mathbf{x}_{a,\mathrm{C}}$. Assuming non-collinear backbone atoms ($\mathbf{w} \times \mathbf{u} \neq \mathbf{0}$), we define the translation origin as $\mathbf{t}_a = \mathbf{x}_{a,\mathrm{C}_\alpha}$ and the proper rotation matrix $\mathbf{R}_a = [\mathbf{e}_1\ \mathbf{e}_2\ \mathbf{e}_3] \in \mathrm{SO}(3)$ via Gram–Schmidt orthogonalization:

$$\begin{aligned} \mathbf{u} &= \mathbf{x}_{a,\mathrm{C}} - \mathbf{x}_{a,\mathrm{C}_\alpha}, & \mathbf{e}_1 &= \frac{\mathbf{u}}{\max(\|\mathbf{u}\|_2, \epsilon)}, \\ \mathbf{w} &= \mathbf{x}_{a,\mathrm{N}} - \mathbf{x}_{a,\mathrm{C}_\alpha}, & \mathbf{e}_2 &= \frac{\mathbf{w} - (\mathbf{e}_1^\top \mathbf{w})\mathbf{e}_1}{\max(\|\mathbf{w} - (\mathbf{e}_1^\top \mathbf{w})\mathbf{e}_1\|_2, \epsilon)}, \\ & & \mathbf{e}_3 &= \mathbf{e}_1 \times \mathbf{e}_2, \end{aligned} \tag{1}$$

where $\epsilon = 10^{-12}$ prevents division by zero. Expressing local spatial quantities in this residue frame via $\mathbf{y}^{\mathrm{loc}} = \mathbf{R}_a^\top(\mathbf{y} - \mathbf{t}_a)$ guarantees that all downstream frame-aligned quantities, namely the geometric descriptors, attention queries, and attention weights (Eqs. 2–4), are strictly SE(3)-invariant to global rigid transformations of the molecular complex.

**Surface-aware geometric encoding.** For antibody atom slot $p$ and antigen surface vertex $j$, the relative displacement and surface normal expressed in the local frame of residue $a \equiv b_e$ are computed as $\mathbf{d}_{pj} = \mathbf{R}_a^\top(\mathbf{x}_{a,p} - \mathbf{v}_j)$ and $\vec{n}_j^{\mathrm{loc}} = \mathbf{R}_a^\top \vec{n}_j$, where $\mathbf{d}_{pj}$ points from surface vertex $j$ toward antibody atom $p$. The descriptor combines the antibody-frame approach direction with a surface-normal incidence cosine:

$$\mathbf{g}_{pj} = [\phi_r(r_{pj}),\ \mathbf{A}(\Theta_{pj}, \Phi_{pj}),\ \hat{\mathbf{u}}_{pj},\ c_{pj},\ \mathbf{a}_p] \in \mathbb{R}^{40}, \tag{2}$$

where $r_{pj} = \|\mathbf{d}_{pj}\|_2$, $\hat{\mathbf{u}}_{pj} = \mathbf{d}_{pj}/\max(r_{pj}, \epsilon)$, and $(r_{pj}, \Theta_{pj}, \Phi_{pj})$ are the spherical coordinates of $\mathbf{d}_{pj}$; $\phi_r : \mathbb{R} \to \mathbb{R}^{16}$ is a Gaussian radial basis (Schütt et al., 2017), $\mathbf{A}(\Theta, \Phi) = [\sin\Theta, \cos\Theta, \sin\Phi, \cos\Phi]^\top$ a periodic directional encoding, smooth away from the polar axis of the residue frame ($\Theta = 0, \pi$, where $\Phi$ is undefined), $\mathbf{a}_p \in \mathbb{R}^{16}$ the atom-type embedding, and $c_{pj} = \hat{\mathbf{u}}_{pj}^\top \vec{n}_j^{\mathrm{loc}} = \cos\Theta_{pj}^n$ the frame-independent approach–normal cosine, with $\Theta_{pj}^n$ the angle between the approach direction and the surface normal. The direction channels expose what scalar distances cannot: two contacts with $\mathbf{d}_1 = (r, 0, 0)^\top$ and $\mathbf{d}_2 = (0, r, 0)^\top$ have equal distance but different orientation in the antibody frame, and are separated by $\mathbf{A}(\Theta, \Phi)$ and $\hat{\mathbf{u}}_{pj}$; and while a single $c_{pj}$ does not encode the surface-normal azimuth, aggregating $c_{pj}$ across the contacts of the patch provides incidence information about the local surface orientation. All atom–surface quantities are *edge-local*: computed independently for each pairing $(i, b_e)$ in the frame of $a$, never pooled across antibody residues or local frames, and mixed across edges only through the invariant edge messages $\mathbf{m}_e$ at the node level.

### 3.3 Adaptive Atom–Surface Interaction and Gated Coordinate Update

We aggregate fine-grained atom–surface interactions using a two-level attention scheme.

**Atom-level aggregation.** Atom-wise interaction features are pooled into surface vertices via invariant attention weighting (Fuchs et al., 2020):

$$\bar{\mathbf{g}}_j^{(e)} = \sum_{p \in a} \alpha_{pj} \mathbf{g}_{pj}, \qquad \alpha_{pj} = \mathrm{softmax}_p \left( f(\mathbf{g}_{pj}) \right), \tag{3}$$

where $f$ is a multi-layer perceptron.

**Surface-level attention and interface edge readout.** Content queries (Vaswani et al., 2017) are constructed from the frame-aligned local atom positions $\mathbf{q}_p^{\mathrm{loc}} = \mathbf{R}_a^\top (\mathbf{x}_{a,p} - \mathbf{t}_a) \in \mathbb{R}^3$:

$$\beta_{pj}^{(e)} = \mathrm{softmax}_j \left( \frac{(\mathbf{W}_q \mathbf{q}_p^{\mathrm{loc}})^\top \mathbf{W}_k \bar{\mathbf{g}}_j^{(e)}}{\sqrt{d}} \right). \tag{4}$$

Here, both the query $\mathbf{q}_p^{\mathrm{loc}}$ and the keys $\bar{\mathbf{g}}_j^{(e)}$ are defined within the residue-local frame of $a$.

Each atom slot retrieves a surface feature vector $\mathbf{f}_p = \sum_j \beta_{pj}^{(e)} \mathbf{W}_v \bar{\mathbf{g}}_j^{(e)}$, which is averaged to form a residue-level interaction summary $\bar{\mathbf{f}} = \frac{1}{|a|} \sum_{p \in a} \mathbf{f}_p$. Its normalized embedding $\mathbf{r}_e = \mathbf{W}_r \left( \bar{\mathbf{f}} / (\|\bar{\mathbf{f}}\|_2 + \epsilon) \right)$ is integrated into the interface edge message:

$$\mathbf{m}_e = \mathrm{MLP}_e \left( [\mathbf{h}_i \parallel \mathbf{h}_{b_e} \parallel \mathbf{r}_e] \right). \tag{5}$$

The invariant message $\mathbf{m}_e$ updates residue node states $\mathbf{h}_i$ and parameterizes scalar gating functions for spatial updates (Satorras et al., 2021).

**Attended surface centroid and transient coordinate update.** Averaging $\beta_{pj}^{(e)}$ over the valid (unmasked) atoms of residue $a$ yields the marginal vertex distribution $\beta_j^{(e)} = \frac{1}{|a|} \sum_{p \in a} \beta_{pj}^{(e)}$ and the attended surface centroid $\bar{\mathbf{v}}_e = \sum_j \beta_j^{(e)} \mathbf{v}_j$.

Within the local interaction encoder, atomic positions are maintained as 14-slot coordinates: antigen atoms enter from the observed context, while all antibody coordinates are initialized from the conserved framework template and refined during decoding (full state provenance in Appendix A.4). For an epitope residue $i$, its transient atomic coordinates $\tilde{\mathbf{x}}_{i,p} \in \mathbb{R}^3$ are updated across encoder layers via a gated message-passing scheme:

$$\tilde{\mathbf{x}}_{i,p} \leftarrow \tilde{\mathbf{x}}_{i,p} + \frac{1}{|\mathcal{N}_i|} \sum_{e \in \mathcal{N}_i} \psi_p(\mathbf{m}_e) \big( \mathbf{x}_{a,p} - \bar{\mathbf{v}}_e \big), \tag{6}$$

where $\mathbf{x}_{a,p}$ is the coordinate of atom slot $p$ and $\psi_p(\mathbf{m}_e)$ a slot-specific scalar gate projected from the invariant edge message $\mathbf{m}_e$. Here, $\tilde{\mathbf{x}}_{i,p}$ is a transient state that conditions downstream pairwise spatial contexts, not an explicit structure prediction. Unlike the residue-local descriptors and queries of Eqs. (2)–(4), the quantities of Eq. (6) live in the shared global frame of the current complex; the update acts on relative displacements and co-transforms with the complex under rigid motions (Appendix A.5).

**Layer-wise geometric recurrence.** The frames $(\mathbf{R}_a, \mathbf{t}_a)$, the descriptors $\mathbf{g}_{pj}$, the queries $\mathbf{q}_p^{\mathrm{loc}}$, and the attended centroids $\bar{\mathbf{v}}_e$ are all recomputed at each encoder layer from the current transient coordinates, while the mesh $\{\mathbf{v}_j\}$ stays anchored to the fixed antigen context. Each layer interleaves the surface-attention update with an equivariant inter-chain update whose pair-distance features are conditioned on the refined transient states $\tilde{\mathbf{x}}^{(\ell)}$, and the update of Eq. (6) is $\mathrm{SE}(3)$-equivariant.

### 3.4 Geometric Supervision

We use geometric objectives in the same residue-local frames as the interface representation: FAPE aligns every atom in every residue frame and thereby constrains global structural consistency, while the torsion terms regularize local backbone geometry; together they act as a training-time geometric prior that stabilizes what the interface representation learns.

**Frame-aligned point error.** Following Jumper et al. (2021),

$$\mathcal{L}_{\mathrm{FAPE}} = \frac{1}{|\mathcal{F}||\mathcal{P}|} \sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{P}} \min \left( d_{\max}, \left| \mathbf{R}_i^\top (\hat{\mathbf{x}}_j - \mathbf{t}_i) - \mathbf{R}_i^{\star\top} (\mathbf{x}_j^\star - \mathbf{t}_i^\star) \right|_2 \right). \tag{7}$$

**Torsion and angle loss.** For backbone dihedrals $\theta \in \{\phi, \psi, \omega\}$ and bond angles $\alpha$, $\mathcal{L}_{\text{torsion}}$ penalizes errors of the predicted cosines with the smooth-$\ell_1$ penalty $\text{s}\ell_1$:

$$\mathcal{L}_{\text{torsion}} = \sum_{\theta} \text{s}\ell_1\Big(\cos\hat{\theta}, \cos\theta\Big) + \sum_{\alpha} \text{s}\ell_1(\cos\hat{\alpha}, \cos\alpha)\,. \tag{8}$$

**Total objective.** $\mathcal{L}_{\text{seq}}$ is the cross-entropy of masked-token predictions summed over decoding rounds. The structure term comprises a Kabsch-aligned (Kabsch, 1976) full-atom coordinate loss $\mathcal{L}_x$, smooth-$\ell_1$ bond-length losses $\mathcal{L}_{\text{bond}}$ on backbone and sidechain bonds, $\mathcal{L}_{\text{FAPE}}$, and $\mathcal{L}_{\text{torsion}}$. The docking term supervises the interface coordinates of the shadow paratope ($\mathcal{L}_{\text{SP}}$) and the predicted inter-chain edge-distance map ($\mathcal{L}_{\text{ed}}$). These terms follow the implementation of Kong et al. (2023). An auxiliary smooth-$\ell_1$ term $\mathcal{L}_{\text{pRMSD}}$ additionally trains a per-residue RMSD prediction head on the antibody residues. The total objective is

$$\begin{aligned}\mathcal{L} = \lambda_{\text{seq}}\mathcal{L}_{\text{seq}} + \underbrace{\lambda_x\mathcal{L}_x + \lambda_{\text{bond}}\mathcal{L}_{\text{bond}} + \lambda_{\text{FAPE}}\mathcal{L}_{\text{FAPE}} + \lambda_{\text{torsion}}\mathcal{L}_{\text{torsion}}}_{\text{structure}} \\ + \underbrace{\lambda_{\text{SP}}\mathcal{L}_{\text{SP}} + \lambda_{\text{ed}}\mathcal{L}_{\text{ed}}}_{\text{docking}} + \underbrace{\lambda_{\text{pRMSD}}\mathcal{L}_{\text{pRMSD}}}_{\text{auxiliary}},\end{aligned} \tag{9}$$

### 3.5 Generation across Modeling Tasks

AbGaze decodes the sequence by stochastically revealing masked positions (Ghazvininejad et al., 2019) under a linear unmasking schedule, refining the coordinates conditioned on the partially revealed sequence: the target region starts fully masked with antibody coordinates from the conserved framework template, and each of the $K$ reveal rounds refines the structure, predicts every still-masked position, and commits each independently,

$$p_\theta\Big(x_i^{(s)} \,\Big|\, x^{(s-1)}, c\Big) = (1-\rho_s)\,\delta_{\text{m}} + \rho_s\,\text{softmax}\Big(\mathbf{z}_i^{(s-1)}/\tau\Big)\,, \qquad \rho_s = \frac{1}{K-s+1}, \quad \rho_K = 1, \tag{10}$$

where $\text{m}$ is the mask token, $x^{(s)}$ the decoding state after reveal round $s$, $c$ the antigen and committed context, $\mathbf{z}_i^{(s-1)}$ the round-$s$ decoder logits, and $K$ the number of reveal rounds. Revealed tokens are never re-masked, the final round commits the rest, and the temperature $\tau$ controls candidate diversity. Training uses the masked-prediction objective, a member of the simplified masked-cross-entropy family used to train absorbing-state diffusion (Sahoo et al., 2024), with the corruption level set by an annealed context curriculum rather than a uniform prior; Appendix A.6 states its precise scope.

Masking any subset of CDRs configures joint sequence–structure design; observing the full antibody sequence yields complex structure prediction from template initialization; and affinity optimization tunes the template-initialization noise through the frozen generator, guided by a pretrained $\Delta\Delta G$ regressor as a differentiable proxy with a KL trust region toward $\mathcal{N}(0, \mathbf{I})$. Across all four tasks, the same interface representation, attention, and equivariant refinement apply: within the shared generator there is no task-specific component or loss term, and the tasks differ only in masking, initialization, and sampling or optimization configurations; the sole component fitted outside the generator is the frozen $\Delta\Delta G$ regressor, which contributes no term to the generator's objective (Appendix B).

## 4 Experiments

We evaluate AbGaze on the four settings served by the shared interface representation: all-CDR design, CDR-H3 design, complex structure prediction, and affinity optimization.

### 4.1 Setup

**Data and benchmark.** Training and evaluation follow the data pipeline of Kong et al. (2023) (details in Appendix C): SAbDab (Dunbar et al., 2014) complexes clustered by CDR sequence identity (Kong et al., 2022). To rule out cross-task leakage, we hold out the single benchmark RAbD (Adolf-Bryfogle et al., 2018) (60 complexes), remove every overlapping cluster from training, and

evaluate *all four* tasks exclusively on it. No published numbers are copied; all baselines are re-run under this protocol, with the structure-prediction and affinity baselines locally re-evaluated (Appendix C).

**Baselines.** We compare against RosettaAb (Adolf-Bryfogle et al., 2018), DiffAb (Luo et al., 2022), MEAN (Kong et al., 2022), HERN (Jin et al., 2022a), dyMEAN (Kong et al., 2023), and AbFlow (Wang et al., 2026). Models that fill CDRs on a docked backbone are standardized into the four-stage IgFold (Ruffolo et al., 2023), HDock (Yan et al., 2020), generation, and Rosetta pipeline of Kong et al. (2023); Wang et al. (2026).

**Metrics.** Sequence quality is measured by AAR and its contact-restricted variant CAAR; structure quality by TM-score (Zhang & Skolnick, 2004), lDDT (Mariani et al., 2013), and $C_\alpha$ RMSD after alignment; interface quality by DockQ (Basu & Wallner, 2016); and affinity by the best $\Delta\Delta G$ (via a shared pretrained regressor), the improvement percentage (IMP, fraction of candidates with $\Delta\Delta G < 0$), and the mutation count $\Delta L$. Evaluation must specify how the distribution over designs is consumed; the protocol is fixed per task, with every method sampled and selected identically within the design and affinity comparisons. *Design* (all-CDR and CDR-H3): five samples per target at $\tau$=0.5, reporting the candidate with the highest AAR against the native, an oracle selection applied identically to every method. *Complex structure prediction*: ten stochastic draws (random shadow-paratope initialization), keeping the one with the lowest model-predicted per-residue RMSD; no ground truth is involved. *Affinity optimization*: thirty optimized candidates per target, reporting the best $\Delta\Delta G$; IMP is computed over all candidates. The budget contributes little: best-of-five adds only 2.2 AAR points over a single deterministic design (Appendix F).

### 4.2 ALL-CDR DESIGN

Table 1 evaluates the joint design of all six CDRs. The *All* AAR is pooled over the residues of the six CDRs, while the *All* RMSD is the $C_\alpha$ RMSD of the entire antibody. The pooled AAR reaches 66.2%, a 10% relative gain over the strongest baseline. Relative to the strongest baseline on each CDR, AbGaze improves AAR by 7.1% and reduces RMSD by 14.9% on average over the six loops, with the largest gains on the conformationally variable L3 and H3: L3 AAR reaches 0.69 ($+19\%$) and the H3 $C_\alpha$ RMSD falls to 1.65 Å ($-10\%$). Interface quality follows, with a DockQ of 0.422 vs. 0.396 ($+6.6\%$).

Table 1: Performance comparison on all-CDR antibody design. ↑ indicates higher is better, while ↓ indicates lower is better. Bold denotes the best performance.

| Metric | Item | AbGaze | AbFlow | dyMEAN |
|---|---|---|---|---|
| AAR↑ | L1 | **0.77** | 0.69 | 0.76 |
| | L2 | **0.85** | 0.82 | 0.83 |
| | L3 | **0.69** | 0.58 | 0.52 |
| | H1 | **0.79** | 0.74 | 0.76 |
| | H2 | **0.71** | 0.65 | 0.69 |
| | H3 | **0.43** | 0.38 | 0.38 |
| | All | **0.662** | 0.597 | 0.601 |
| RMSD (CA)↓ | L1 | **0.44** | 0.64 | 0.86 |
| | L2 | **0.21** | 0.25 | 0.48 |
| | L3 | **0.65** | 0.65 | 0.94 |
| | H1 | **0.52** | 0.63 | 0.63 |
| | H2 | **0.47** | 0.55 | 0.71 |
| | H3 | **1.65** | 1.83 | 2.45 |
| | All | **1.052** | 1.104 | 1.357 |
| Structure | DockQ↑ | **0.422** | 0.379 | 0.396 |
| | LDDT↑ | **0.831** | 0.815 | 0.803 |
| | TM-Score↑ | **0.973** | 0.971 | 0.965 |

The concurrent improvements across heavy- and light-chain CDRs suggest that explicit orientation and distance cues help coordinate multiple flexible loops: resolving each residue's contacts in its

own local frame against the epitope surface supports joint optimization of paratope sequence and backbone geometry, keeping adjacent loops coherent while aligning the paratope with the antigen.

### 4.3 CDR-H3 Design

Table 2 evaluates CDR-H3 design on RAbD, the most widely optimized loop given H3's decisive role in antigen recognition. On global backbone metrics, most end-to-end baselines reach comparably high performance ($\geq 0.84$ lDDT, $\geq 0.97$ TM-score), indicating that coarse loop topology is largely well-captured. AbGaze leads across sequence, interface, and structural metrics, reaching an AAR of 45.60% ($+4.5\%$), a contact-restricted recovery of 32.20% ($+11.8\%$), a DockQ of 0.443 ($+3.5\%$), and the lowest $C_\alpha$ RMSD (8.06), all relative to the strongest baseline. The gains concentrate at the interface itself, suggesting that encoding the distance, direction, and angle to the surface normal at which each atom approaches the epitope lets the model align local paratope geometry with epitope constraints.

Table 2: Performance comparison on CDR-H3 antibody design on the RAbD benchmark.

| Method | AAR ↑ | TMscore ↑ | lDDT ↑ | CAAR ↑ | RMSD ↓ | DockQ ↑ |
|---|---|---|---|---|---|---|
| RosettaAb | 32.31% | 0.9717 | 0.8272 | 14.58% | 17.70 | 0.137 |
| DiffAb | 35.31% | 0.9695 | 0.8281 | 22.17% | 23.24 | 0.158 |
| MEAN | 37.38% | 0.9688 | 0.8252 | 24.11% | 17.30 | 0.162 |
| HERN | 32.65% | – | – | 19.27% | 9.15 | 0.294 |
| dyMEAN | 43.65% | 0.9726 | 0.8454 | 28.11% | 8.11 | 0.409 |
| AbFlow | 42.10% | **0.9735** | **0.8518** | 28.80% | 8.45 | 0.428 |
| **AbGaze** | **45.60%** | 0.9730 | 0.8455 | **32.20%** | **8.06** | **0.443** |

### 4.4 Complex Structure Prediction

Table 3 evaluates prediction from sequences alone against the docking pipeline HDock, the hierarchical refinement of HERN, and the end-to-end baselines. Global antibody structure is near-saturated for all end-to-end methods (TM-score $\geq 0.97$); the differences appear precisely at the interface, where AbGaze attains the best DockQ (0.435) and RMSD (8.03). The simultaneous DockQ and RMSD gains indicate that the representation better resolves the paratope's spatial arrangement relative to the antigen surface: the distance-, direction-, and orientation-structure of the static binding mode on this benchmark. These gains show that the representation characterizes the binding mode accurately.

Table 3: Performance comparison on antigen–antibody complex structure prediction.

| Model | TMscore ↑ | lDDT ↑ | RMSD ↓ | DockQ ↑ |
|---|---|---|---|---|
| HDock | 0.9723 | 0.8503 | 18.46 | 0.170 |
| HERN | 0.9722 | 0.8441 | 10.19 | <u>0.424</u> |
| dyMEAN | <u>0.9730</u> | <u>0.8568</u> | 9.04 | 0.409 |
| AbFlow | 0.9720 | 0.8526 | <u>8.66</u> | 0.419 |
| **AbGaze** | **0.9735** | **0.8572** | **8.03** | **0.435** |

### 4.5 Affinity Optimization

Table 4 compares affinity optimization on RAbD with the same frozen generator and the same $\Delta\Delta G$ regressor for all methods. AbGaze reaches the best $\Delta\Delta G$ ($-11.10$) and by far the highest improvement rate (IMP 70.6%, $+17.3$ points over dyMEAN and $+17.8$ over AbFlow): almost three quarters of the optimized candidates receive a negative predicted $\Delta\Delta G$ from the shared scorer. The mutations it requests are more numerous ($\Delta L$ 8.07 vs. 4.25 for dyMEAN); we regard this as an acceptable trade, since IMP already scores success per candidate. That gradient ascent through the shared representation alone suffices to steer affinity supports the central claim of §3.2: the orientation-aware encoding captures the interfacial context—proximity, relative orientation, and the contribution of

each contact under its surrounding geometry—well enough that optimizing toward this representation, rather than merely fitting the design objective, transfers to a downstream affinity criterion. Together with the binding-mode results in prediction, this is consistent with the representation characterizing how the antibody binds, and how strongly, more faithfully than distance- or surface-based encodings.

Table 4: Comparison of affinity optimization performance.

| **Method** | **Best** $\Delta\Delta G \downarrow$ | **IMP (%)** $\uparrow$ | $\Delta L \downarrow$ |
|---|---|---|---|
| DiffAb | -3.29 | 38.8 | 5.62 |
| dyMEAN | -4.47 | 53.3 | **4.25** |
| AbFlow | -9.31 | 52.8 | 6.98 |
| **AbGaze** | **-11.10** | **70.6** | 8.07 |

## 5 Ablation

We ablate the orientation-aware interaction module (Variant A) and the residue-local geometric loss (Variant B) on RAbD all-CDR design.

**Orientation-aware interaction.** Variant (A) replaces the local interface representation (§3.2–3.3) with standard distance-based message passing. Removing explicit spatial and normal directionality primarily weakens interface packing (DockQ $-0.040$, antibody C$\alpha$ RMSD $+0.061$ Å) and reduces overall sequence recovery (AAR $-4.1\%$), showing that atom–surface orientation carries information beyond scalar distances.

**Local-frame geometric supervision.** Variant (B) omits the residue-local FAPE and dihedral terms (§3.4), relying solely on global coordinate and auxiliary losses. While interface docking is essentially unchanged (DockQ $+0.012$), backbone structural quality degrades significantly (antibody C$\alpha$ RMSD $+0.147$ Å, lDDT $-0.043$, TM-Score $-0.009$). The local-frame terms thus act as global geometric regularizers that constrain the relative orientation of adjacent residues and stabilize the overall fold.

Table 5: Ablation study on all-CDR antibody design.

| **Variant** | **AAR** $\uparrow$ | **RMSD** $\downarrow$ | **DockQ** $\uparrow$ | **lDDT** $\uparrow$ | **TMscore** $\uparrow$ |
|---|---|---|---|---|---|
| **AbGaze (Full)** | **66.2%** | **1.052** | **0.422** | **0.831** | **0.973** |
| (A) w/o Orient.-Aware Attn. | $-4.1\%$ | $+0.061$ | $-0.040$ | $-0.007$ | $-0.003$ |
| (B) w/o Local-Frame Supv. | $-1.2\%$ | $+0.147$ | $+0.012$ | $-0.043$ | $-0.009$ |

## 6 Conclusion

We presented AbGaze, a unified framework for antigen–antibody interface modeling built on an orientation-aware local representation and adaptive atom–surface attention. By anchoring distance, direction, and surface-normal geometry to residue-local frames, AbGaze explicitly encodes how each antibody atom approaches the antigen surface, and its two-level attention learns which local interactions matter under each interfacial context. A single model covers all four tasks: relative to the second-best method, it improves amino-acid recovery by 7.1% and reduces per-CDR $C_\alpha$ RMSD by 14.9% on average over the six CDRs, improves DockQ by 6.6%, and raises the affinity improvement rate (IMP) by 32.5%. These results support the premise that explicitly characterizing the relative spatial relationships between paratope and epitope in antibody-residue local frames outperforms leaving them implicit in scalar distances or surface-intrinsic features. Next steps include extending the representation to flexible antigens, validating designs experimentally, and incorporating developability into the optimization objective.

## REFERENCES


Jared Adolf-Bryfogle, Oleks Kalyuzhniy, Michael Kubitz, Brian D. Weitzner, Xiaozhen Hu, Yumiko Adachi, William R. Schief, and Roland L. Dunbrack. Rosettaantibodydesign (rabd): A general framework for computational antibody design. *PLOS Computational Biology*, 14(4):e1006112, 2018. doi: 10.1371/journal.pcbi.1006112.

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17981–17993, 2021.

Sankar Basu and Björn Wallner. Dockq: A quality measure for protein-protein docking models. *PLOS ONE*, 11(8):e0161879, 2016. doi: 10.1371/journal.pone.0161879.

Andrew C. Chan, Greg D. Martyn, and Paul J. Carter. Fifty years of monoclonals: the past, present and future of antibody therapeutics. *Nature Reviews Immunology*, 25:745–765, 2025. doi: 10.1038/s41577-025-01207-9.

Zhihang Chen, Menglin Ji, Jie Qian, Zhe Zhang, Xiangying Zhang, Haotian Gao, Haojie Wang, Renxiao Wang, and Yifei Qi. Probid-net: a deep learning model for protein-protein binding interface design. *Chemical Science*, 15(47):19977–19990, 2024. doi: 10.1039/D4SC02233E.

Bowen Dai and Chris Bailey-Kellogg. Protein interaction interface region prediction by geometric deep learning. *Bioinformatics*, 37(17):2580–2588, 2021. doi: 10.1093/bioinformatics/btab154.

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

Fabian B. Fuchs, Daniel E. Worrall, Volker Fischer, and Max Welling. SE(3)-transformers: 3D roto-translation equivariant attention networks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 1970–1981, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/15231a7ce4ba789d13b722cc5c955834-Abstract.html.

Pablo Gainza, Freyr Sverrisson, Federico Monti, Emanuele Rodolà, Davide Boscaini, Michael M. Bronstein, and Bruno E. Correia. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, 17:184–192, 2020. doi: 10.1038/s41592-019-0666-6.

Pablo Gainza, Sarah Wehrle, Alexandra Van Hall-Beauvais, Anthony Marchand, Andreas Scheck, Zander Harteveld, Stephen Buckley, Dongchun Ni, Shuguang Tan, Freyr Sverrisson, Casper Goverde, Priscilla Turelli, Charlène Raclot, Alexandra Teslenko, Martin Pacesa, Stéphane Rosset, Sandrine Georgeon, Jane Marsden, Aaron Petruzzella, Kefang Liu, Zepeng Xu, Yan Chai, Pu Han, George F. Gao, Elisa Oricchio, Beat Fierz, Didier Trono, Henning Stahlberg, Michael Bronstein, and Bruno E. Correia. De novo design of protein interactions with learned surface fingerprints. *Nature*, 617:176–184, 2023. doi: 10.1038/s41586-023-05993-x.

Marjan Ghazvininejad, Omer Levy, Yinhan Liu, and Luke Zettlemoyer. Mask-Predict: Parallel decoding of conditional masked language models. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 6112–6121. Association for Computational Linguistics, 2019. doi: 10.18653/v1/D19-1633. URL https://aclanthology.org/D19-1633/.

Wengong Jin, Regina Barzilay, and Tommi S. Jaakkola. Antibody-antigen docking and design via hierarchical equivariant refinement. *arXiv preprint arXiv:2207.06616*, 2022a.

Wengong Jin, Jeremy Wohlwend, Regina Barzilay, and Tommi Jaakkola. Iterative refinement graph neural network for antibody sequence-structure co-design. In *International Conference on Learning Representations*, 2022b. URL https://arxiv.org/abs/2110.04624.

Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael J. L. Townshend, and Ron Dror. Learning from protein structure with geometric vector perceptrons. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2009.01411.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2.

Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, 1976. doi: 10.1107/S0567739476001873.

Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. URL https://arxiv.org/abs/1412.6980.

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

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

Lucien F. Krapp, Luciano A. Abriata, Fabio Cortés Rodriguez, and Matteo Dal Peraro. Pesto: parameter-free geometric deep learning for accurate prediction of protein binding interfaces. *Nature Communications*, 14(1):2175, 2023. doi: 10.1038/s41467-023-37701-8.

Daisuke Kuroda and Jeffrey J. Gray. Shape complementarity and hydrogen bond preferences in protein–protein interfaces: implications for antibody modeling and protein–protein docking. *Bioinformatics*, 32(16):2451–2456, 2016. doi: 10.1093/bioinformatics/btw197.

Michael C. Lawrence and Peter M. Colman. Shape complementarity at protein/protein interfaces. *Journal of Molecular Biology*, 234(4):946–950, 1993. doi: 10.1006/jmbi.1993.1648.

Marie-Paule Lefranc, Christelle Pommié, Manuel Ruiz, Véronique Giudicelli, Elodie Foulquier, Lisa Truong, Valérie Thouvenin-Contet, and Gérard Lefranc. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Developmental and Comparative Immunology*, 27(1):55–77, 2003. doi: 10.1016/S0145-305X(02)00039-3.

Shitong Luo, Kevin K. Yang, Minkai Xu, Zhuoran Wu, Pengtao Xie, Wengong Jin, Bowen Tang, Jian Peng, and Jianzhu Ma. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. In *Advances in Neural Information Processing Systems*, volume 35, pp. 9754–9767, 2022.

Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 2013. doi: 10.1093/bioinformatics/btt473.

Karolis Martinkus, Jan Ludwiczak, Wei-Ching Liang, Julien Lafrance-Vanasse, Isidro Hotzel, Arvind Rajpal, Yan Wu, Kyunghyun Cho, Richard Bonneau, Vladimir Gligorijevic, and Andreas Loukas. AbDiffuser: Full-atom generation of in-vitro functioning antibodies. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/801ec05b0aae9fcd2ef35c168bd538e0-Abstract-Conference.html.

A. J. McCoy, V. Chandana Epa, and P. M. Colman. Electrostatic complementarity at protein/protein interfaces. *Journal of Molecular Biology*, 268(2):570–584, 1997. doi: 10.1006/jmbi.1997.0987.

Benjamin North, Andreas Lehmann, and Roland L. Dunbrack Jr. A new clustering of antibody CDR loop conformations. *Journal of Molecular Biology*, 406(2):228–256, 2011. doi: 10.1016/j.jmb.2010.10.030.

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*, 14:2389, 2023. doi: 10.1038/s41467-023-38063-x.

Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T. Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 130136–130184, 2024. doi: 10.52202/079017-4135.

Michel F. Sanner, Arthur J. Olson, and Jean-Claude Spehner. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers*, 38(3):305–320, 1996. doi: 10.1002/(SICI)1097-0282(199603)38:3⟨305::AID-BIP4⟩3.0.CO;2-Y.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 2021. URL https://proceedings.mlr.press/v139/satorras21a.html.

Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/303ed4c69846ab36c2904d3ba8573050-Abstract.html.

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

Aroop Sircar and Jeffrey J. Gray. SnugDock: Paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models. *PLOS Computational Biology*, 6(1):e1000644, 2010. doi: 10.1371/journal.pcbi.1000644.

Freyr Sverrisson, Jean Feydy, Bruno E. Correia, and Michael M. Bronstein. Fast end-to-end learning on protein surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15272–15281, 2021. URL https://openaccess.thecvf.com/content/CVPR2021/html/Sverrisson_Fast_End-to-End_Learning_on_Protein_Surfaces_CVPR_2021_paper.html.

Cheng Tan, Yijie Zhang, Zhangyang Gao, Yufei Huang, Haitao Lin, Lirong Wu, Fandi Wu, Mathieu Blanchette, and Stan Z. Li. dyab: Flow matching for flexible antibody design with alphafold-driven pre-binding antigen. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 782–790, 2025. doi: 10.1609/aaai.v39i1.32061.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, 2017. URL https://papers.nips.cc/paper/7181-attention-is-all-you-need.

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

Wenda Wang, Yang Zhang, Zhewei Wei, and Wenbing Huang. Abflow: End-to-end paratope-centric antibody design by interaction enhanced flow matching. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2026. doi: 10.1145/3770854.3780296.

Songjian Wei, Jinxiong Zhang, Yan Chen, Chunyan Tang, and Jiayang Tan. Geogad: geometry-aware antibody design framework for complementarity-determining region precision engineering. *Bioinformatics*, 42(2):btag042, 2026. doi: 10.1093/bioinformatics/btag042.

Brian D. Weitzner, Roland L. Dunbrack Jr., and Jeffrey J. Gray. The origin of cdr h3 structural diversity. *Structure*, 23(2):302–311, 2015. doi: 10.1016/j.str.2014.11.010.

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. doi: 10.1038/s41596-020-0312-x.

Jianying Yang, Chittoor P. Swaminathan, Yuping Huang, Rongjin Guan, Sangwoo Cho, Michele C. Kieke, David M. Kranz, Roy A. Mariuzza, and Eric J. Sundberg. Dissecting cooperative and additive binding energetics in the affinity maturation pathway of a protein-protein interface. *Journal of Biological Chemistry*, 278(50):50412–50421, 2003. doi: 10.1074/jbc.M306848200.

Yang Zhang and Jeffrey Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4):702–710, 2004. doi: 10.1002/prot.20264.

# A Extended Method Details

This section supplements §3.2–§3.3 with the implementation-level details behind the interface representation: the atom-slot schema and the construction of interface edges (Appendix A.1), the surface-patch construction (Appendix A.2), the composition of the geometric descriptor and the normalization axes of the two-level attention (Appendix A.3), and the transient coordinate states, their layer-wise recurrence, and the $\mathrm{SE}(3)$-equivariance of the interface update (Appendices A.4 and A.5). Notation follows the main text.

## A.1 Atom-Slot Schema and Interface Edges

**Atom slots.** Both molecules use the standardized 14-slot vocabulary of §3.2: slots 1–4 hold the backbone atoms $\mathrm{N}, \mathrm{C}_\alpha, \mathrm{C}, \mathrm{O}$, and slots 5–14 hold the side-chain heavy atoms in the fixed per-type order of Table 6 (at most ten; tryptophan is the only residue type that uses all of them). When a residue provides fewer atoms than its type allows, which is always the case for generated residues before side-chain completion and for any unresolved atom in an observed structure, the unused slots are padded at the $\mathrm{C}_\alpha$ coordinate and flagged as padding. Padded slots are excluded wherever atom slots are enumerated: from the $k$-NN edge distances below, from the atom-level softmax of Eq. 3, from the atom averages $\bar{\mathbf{f}}$ and $\beta_j^{(e)}$, and from the gate indices $\psi_p$ of Eq. 6.

Table 6: Side-chain heavy atoms occupying slots 5–14, in slot order; slots 1–4 hold the backbone atoms.

| Res. | Slots 5–14 | Res. | Slots 5–14 |
|---|---|---|---|
| Gly | — | Ser | CB, OG |
| Ala | CB | Thr | CB, OG1, CG2 |
| Val | CB, CG1, CG2 | Cys | CB, SG |
| Leu | CB, CG, CD1, CD2 | Pro | CB, CG, CD |
| Ile | CB, CG1, CG2, CD1 | Phe | CB, CG, CD1, CD2, CE1, CE2, CZ |
| Asp | CB, CG, OD1, OD2 | Tyr | CB, CG, CD1, CD2, CE1, CE2, CZ, OH |
| Asn | CB, CG, OD1, ND2 | His | CB, CG, ND1, CD2, CE1, NE2 |
| Glu | CB, CG, CD, OE1, OE2 | Met | CB, CG, SD, CE |
| Gln | CB, CG, CD, OE1, NE2 | Trp | CB, CG, CD1, CD2, NE1, CE2, CE3, CZ2, CZ3, CH2 |
| Lys | CB, CG, CD, CE, NZ | Arg | CB, CG, CD, NE, CZ, NH1, NH2 |

**Edges.** Context edges are constructed within each molecule and interface edges between the two, both by $k$-NN over residues with $k$=9 (Table 12); the distance between two residues is the minimum Euclidean distance over all pairs of non-padded atom slots, so a single resolved contact suffices to rank a pair. Each interface edge $e = (i, b_e)$ pairs an epitope residue $i$ with an antibody residue $b_e$ and carries the surface patch $\mathcal{V}_i$ of the epitope residue.

## A.2 Surface Patch Construction

The antigen surface is computed once per context with MSMS (Sanner et al., 1996) (1.5 Å probe radius); each vertex carries the MSMS normal, oriented outward from the antigen. Following §3.1,

only vertices within $10\,\mathring{A}$ of any epitope atom are retained. Each retained vertex is assigned to its nearest epitope residue, which yields per-residue patches with ≈72 raw vertices on average; each patch is randomly subsampled, or zero-padded and masked, to a fixed size of $M{=}50$ vertices (Table 12) before batching. The mesh $\{\mathbf{v}_j\}$ is therefore part of the fixed antigen context: it is precomputed before refinement, never moves, and the same vertex assignment is reused at every decoding step. Padded vertices are excluded from the vertex-level softmax of Eq. 4 and from the attended centroid $\bar{\mathbf{v}}_e$, so they can neither receive nor contribute attention mass.

### A.3 Descriptor Composition and Attention Axes

Eq. 2 is evaluated, per interface edge, on the grid of valid atom slots × real patch vertices, that is, non-padded slots $p$ of $a$ against non-padded vertices $j$ of $\mathcal{V}_i$, in the frame of $a$; its 40 channels tally as $16 + 4 + 3 + 1 + 16$ across $\phi_r$, $\mathbf{A}$, $\hat{\mathbf{u}}$, $c_{pj}$, and $\mathbf{a}_p$. The cosine $c_{pj}$ is frame-independent by construction (Eq. 2) and its sign is interpretable: $c_{pj} > 0$ when atom $p$ approaches vertex $j$ from the solvent side along the outward normal, $c_{pj} < 0$ when the approach direction points into the antigen.

The two attention levels of §3.3 normalize over the two axes of this grid separately: the atom-level weights $\alpha_{pj}$ of Eq. 3 softmax over the valid slots $p$ for each vertex, and the surface-level weights $\beta_{pj}^{(e)}$ of Eq. 4 softmax over the real vertices $j$ for each slot. Masked entries are excluded from their normalization rather than zeroed after it, so no probability mass leaks onto padding. The readout $\mathbf{f}_p$, its valid-slot average $\bar{\mathbf{f}}$, the normalized embedding $\mathbf{r}_e$, and the edge message $\mathbf{m}_e$ (Eq. 5) then follow the main text; all are strictly $\mathrm{SE}(3)$-invariant because every input is frame-aligned (§3.2).

### A.4 Transient Coordinates and Layer-Wise Recurrence

Inside the interaction encoder, coordinates live in a *local complex*: the antibody residues together with the epitope residues of their interface edges. At the start of each refinement round, every coordinate in the local complex is re-imposed from the current states, observed on the antigen side and predicted on the antibody side; within the round, the encoder maintains transient atomic states $\tilde{\mathbf{x}}_{i,p}$ on the epitope side that evolve across layers through Eq. 6, while antibody-side coordinates enter as the current predictions. The transient states exist only inside the round: they are re-initialized at the next round, enter no loss, and never persist across decoding steps, so the ground-truth antigen remains the fixed context everywhere it is consumed.

Formally, let $\tilde{\mathbf{x}}^{(\ell)}$ collect the transient coordinates entering encoder layer $\ell$ and $\mathbf{h}^{(\ell)}$ the node states. Each layer interleaves two operators:

1. an *equivariant inter-chain update*: a multi-channel EGNN layer (Satorras et al., 2021) over the inter-chain edges whose radial (pair-distance) features are computed from the current coordinates of all local nodes, including the epitope-side states refined at earlier layers;
2. the *surface-attention update* of Eq. 6, which refines the epitope-side states. Every quantity it consumes is recomputed from the states entering the layer: the residue frames $(\mathbf{R}_a, \mathbf{t}_a)$ (Eq. 1), the descriptors $\mathbf{g}_{pj}$ and queries $\mathbf{q}_p^{\mathrm{loc}}$ (Eqs. 2–4), the atom-level weights $\alpha_{pj}$, and the attended centroids $\bar{\mathbf{v}}_e$.

The composition realizes the recurrence $\tilde{\mathbf{x}}^{(\ell+1)} = F_\ell(\tilde{\mathbf{x}}^{(\ell)}, \{\mathbf{v}_j\}, \mathbf{h}^{(\ell)})$. The epitope-side states written by operator 2 at layer $\ell$ are read by operator 1 at layers $\ell{+}1, \ell{+}2, \ldots$ exclusively through their pair-distance features, which is how the update conditions downstream distance contexts, whereas operator 2 itself consumes only the current antibody coordinates and the fixed mesh. No layer therefore mixes quantities defined in different geometric states.

Table 7 summarizes where every state consumed by the encoder comes from, how it evolves, and which training losses touch it. Complex-structure prediction is the degenerate configuration with an empty mask (the full antibody sequence is observed), and affinity optimization adds Gaussian noise to the template initialization.

### A.5 SE(3)-Equivariance of the Interface Update

**Proof of $\mathrm{SE}(3)$-equivariance.** Let $g = (\mathbf{R}, \mathbf{t}) \in \mathrm{SE}(3)$ act on *all* input point sets, atomic coordinates and the precomputed mesh alike, as $\mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{t}$ and $\mathbf{v}'_j = \mathbf{R}\mathbf{v}_j + \mathbf{t}$.

Table 7: Provenance of the coordinate and sequence states consumed by the interaction encoder.

| State | Initialization | Evolution | Training loss |
|---|---|---|---|
| antigen atoms | observed context | fixed | — |
| mesh $\{\mathbf{v}_j\}$, normals | precomputed (MSMS) | fixed | — |
| sequence (framework) | observed | fixed | — |
| sequence (CDRs) | [MASK] | committed per round | $\mathcal{L}_{\text{seq}}$ |
| coordinates (all) | framework template | refined per round | structure terms |
| transient states $\tilde{\mathbf{x}}_{i,p}$ | from observed antigen | Eq. 6 only | none |
| shadow paratope | predicted | predicted per round | $\mathcal{L}_{\text{SP}}$ |
| inter-chain edge distances | predicted | predicted per round | $\mathcal{L}_{\text{ed}}$ |
| RMSD head | predicted | predicted per round | $\mathcal{L}_{\text{pRMSD}}$ |

1. *Frames.* The Gram–Schmidt construction of Eq. 1 is equivariant: $\mathbf{R}'_a = \mathbf{R}\mathbf{R}_a$ and $\mathbf{t}'_a = \mathbf{R}\mathbf{t}_a + \mathbf{t}$.
2. *Descriptors and queries.* Hence $\mathbf{d}'_{pj} = \mathbf{R}'^{\top}_a(\mathbf{x}'_{a,p} - \mathbf{v}'_j) = \mathbf{d}_{pj}$ and $\vec{n}^{\text{loc}\prime}_j = \vec{n}^{\text{loc}}_j$, so $\mathbf{g}_{pj}$, the atom-level weights $\alpha_{pj}$, the vertex summaries $\bar{\mathbf{g}}^{(e)}_j$, and the queries $\mathbf{q}^{\text{loc}}_p$ are strictly invariant.
3. *Attention and gates.* Both arguments of the logit in Eq. 4 are invariant, so $\beta^{(e)\prime}_{pj} = \beta^{(e)}_{pj}$, and consequently the readout $\mathbf{f}_p$, its average $\bar{\mathbf{f}}$, the normalized embedding $\mathbf{r}_e$, the edge message $\mathbf{m}_e$, and the gates $\psi_p(\mathbf{m}_e)$ are invariant.
4. *Centroid and displacement basis.* The attended centroid co-transforms, $\bar{\mathbf{v}}'_e = \sum_j \beta^{(e)}_j(\mathbf{R}\mathbf{v}_j + \mathbf{t}) = \mathbf{R}\bar{\mathbf{v}}_e + \mathbf{t}$, so translation cancels in the displacement basis: $(\mathbf{x}'_{a,p} - \bar{\mathbf{v}}'_e) = \mathbf{R}(\mathbf{x}_{a,p} - \bar{\mathbf{v}}_e)$.
5. *Update.* Therefore $\tilde{\mathbf{x}}'_{i,p} = \mathbf{R}\tilde{\mathbf{x}}_{i,p} + \mathbf{t} + \mathbf{R}\,\Delta\tilde{\mathbf{x}}_{i,p}$: the update of Eq. 6 transforms covariantly under $g$.
6. *Induction.* The inter-chain operator of step 1 is an EGNN layer built from pair distances (invariant radial features) and point differences (covariant directions) with invariant messages, and is likewise equivariant. Taking the co-transforming input point sets as the base case, equivariance composes through the full encoder by induction over layers.

### A.6 Generation Procedure

Training and decoding share the same masked-prediction core. Each training step draws a masking configuration (which determines the task, §3.5), reveals a random, annealed fraction of the masked residues as conditioning, initializes the masked sequence tokens to [MASK] and the masked coordinates from the template, and runs $R$ refinement rounds of the interface encoder with the node-state memory carried across rounds; parameters are then updated under Eq. 9. Decoding mirrors this construction: starting from the same masked initialization, each reveal round runs the $R$ refinement rounds with the memory carried over and commits tokens as in §3.5; the antibody is then rigidly aligned to the predicted paratope (Kabsch). Candidate selection at evaluation time is fixed per task in §4.1 (oracle-AAR best-of-five for the design tasks; lowest predicted per-residue RMSD over ten draws for structure prediction).

Because each committed token is drawn from $\mathrm{softmax}(\mathbf{z}/\tau)$ and the commit order is randomized, decoding defines a distribution over sequence–structure designs whose concentration is controlled by $\tau$; $\tau \to 0$ recovers deterministic greedy decoding. Diversity is thus a property of the decoding procedure by construction, and we measure it empirically in Appendix E. The paragraph below makes the correspondence with absorbing-state discrete diffusion (Austin et al., 2021) precise.

**Correspondence with absorbing-state discrete diffusion.** Let $x = (x_1, \ldots, x_n)$ be the target-region sequence over the amino-acid vocabulary $\mathcal{V}$, extended with the absorbing state $\mathrm{m} = [\mathrm{MASK}]$, and let $c$ collect the conditioning context (antigen, framework, and already committed residues). Write $x^{(s)}$ for the decoding state after reveal round $s$, with $x^{(0)} = \mathrm{m}^n$. The decoder logits entering round $s$ are produced with the recurrent node-state memory carried across rounds and the

coordinates refined so far; both are deterministic functions of the committed tokens and the fixed context $c$. Conditioned on $c$ and this decoder state, the one-step transition factorizes over positions: still-masked positions follow Eq. 10, and committed positions are never re-masked: revealed states are absorbing along the chain. By telescoping, the schedule induces an exactly linear unmasking trajectory,

$$\mathbb{E}[\gamma_s] = \prod_{u=1}^{s} (1 - \rho_u) = \frac{K - s}{K}, \qquad s = 0, \ldots, K, \tag{11}$$

where $\gamma_s$ is the masked fraction, the discrete counterpart of the linear schedule of absorbing-state diffusion (Austin et al., 2021). Three boundaries of the correspondence follow. **Kernel:** the logits condition on the recurrent hidden state and the refined coordinates as above, so the process is Markov in the sequence only jointly with this decoder state; the correspondence is therefore drawn at the level of the induced marginal unmasking schedule of the sequence channel (Eq. 11), not as an ELBO decomposition of the joint sequence–structure process. **Training:** the masked-prediction objective of §3.5 lies in the simplified masked-cross-entropy family used to train absorbing-state diffusion (Sahoo et al., 2024), with the corruption level swept by a curriculum instead of sampled from a uniform prior; no ELBO identity is claimed under this curriculum. **Scope:** only the sequence carries absorbing-state semantics; coordinates are initialized from a template (or from optimized noise, for affinity optimization) and refined by the geometric updates of §3.3, not diffused.

## B Training and Task Unification

**Surface construction** follows Appendix A.2 (MSMS, per-residue patches subsampled to $M{=}50$).

**Training details.** We train with Adam (Kingma & Ba, 2015) and an exponentially decayed learning rate. Following standard teacher-forcing annealing, each training step reveals a random fraction of the masked residues as conditioning, where the fraction is annealed from near zero toward larger values, so that the model learns to decode from every intermediate state of the generative process. All tasks are trained jointly in a single run: the masking configuration of each step determines the task (§3.5), the same objective and schedules apply to every configuration, and no task is fine-tuned or adapted separately. The affinity-scoring pathway likewise adds only a lightweight prediction head on the shared interface representation; this head is the single component fitted outside the unified run (see **Task unification**). All schedules and loss weights are listed in Table 12; none is tuned per task.

**Task unification.** All tasks share a single architecture, a single interface operator set, and the single training objective of Eq. 9, evaluated with one shared checkpoint; no task-specific adaptation is performed, and within the shared generator no task-specific architectural component or loss term exists. Which terms of the objective are active at each step is determined entirely by the masking configuration. In complex structure prediction the antibody sequence is fully observed (no position is masked), so the masked-token term $\mathcal{L}_{\text{seq}}$ is vacuous by construction, and this fully observed configuration corresponds to the conditioned endpoint of the context annealing described above, the same spectrum of masking states the model is trained on. In affinity optimization the same generator is used frozen and unchanged; the $\Delta\Delta G$ regressor is the single component fitted outside the shared model: a lightweight prediction head on the shared interface representation, fit beforehand on designed variants of the training complexes only (Appendix C) and injected frozen as the differentiable affinity proxy. Its supervision and usage are external to the generator: ascent steps differentiate the frozen regressor through the frozen generator with respect to the template-initialization noise, updating neither the regressor nor the generator, and the identical frozen regressor scores every compared method (§4.5), introducing no task-specific advantage. Task adaptation thus operates entirely through masking, initialization, and sampling or optimization configurations, consistent with §3.5.

**Loss terms and weights.** Table 8 lists the terms of Eq. 9.

## C Evaluation Details

**Data.** Both the training pool and the benchmark follow the official data pipeline of Kong et al. (2023). Antibody–antigen complexes are obtained from SAbDab (Dunbar et al., 2014) (snapshot

Table 8: Loss terms of Eq. 9; structure, docking, and auxiliary terms are smooth-$\ell_1$ penalties.

| Term | Definition (one line) | Weight | Source |
|---|---|---|---|
| $\mathcal{L}_{\text{seq}}$ | per-round CE, masked residues | 1 | standard |
| $\mathcal{L}_{x}$ | Kabsch-aligned coordinates | 1 | Kong et al. (2023) |
| $\mathcal{L}_{\text{bond}}$ | backbone/side-chain bond lengths | 1 | Kong et al. (2023) |
| $\mathcal{L}_{\text{SP}}$ | shadow-paratope coordinates | 1 | Kong et al. (2023) |
| $\mathcal{L}_{\text{ed}}$ | predicted inter-edge distances | 1 | Kong et al. (2023) |
| $\mathcal{L}_{\text{FAPE}}$ | frame-aligned point error, $d_{\max}$ clamp | 0.5 | Jumper et al. (2021) |
| $\mathcal{L}_{\text{torsion}}$ | dihedral/bond-angle cosines | 0.2 | Kong et al. (2023) |
| $\mathcal{L}_{\text{pRMSD}}$ | per-residue RMSD prediction | 1 | ours (auxiliary) |

of November 12, 2022) as IMGT-renumbered (Lefranc et al., 2003) structures with paired heavy–light chains; entries with mis-annotated chains or malformed structures are dropped during cleaning, leaving 5,370 valid complexes, whose CDR boundaries are read directly from the IMGT numbering. For the benchmark split, CDR-H3 sequences are clustered with MMseqs2 at 40% sequence identity, and every cluster containing a complex of the hold-out benchmark is removed from the training and validation pools, so no training complex shares a CDR-H3 cluster with any test complex; 10% of the remaining complexes are held out for validation. A conserved framework template is extracted from the training structures for template initialization. The benchmark itself is the 60-complex RAbD set (Adolf-Bryfogle et al., 2018), likewise IMGT-renumbered. We note that SAbDab and SKEMPI share complexes (e.g., `1a2y`, `2b2x` appear in both), which is why per-task test sets cannot be retained under multi-task training (§4.1).

**Baselines.** All baselines are standardized under the unified protocol of §4.1: same benchmark, same evaluator implementations, and same per-task selection rules. For structure prediction and affinity optimization, some baselines are our locally re-trained or re-reproduced versions on RAbD, run with the identical inputs and initialization as AbGaze.

$\Delta\Delta G$ **regressor.** The affinity proxy is the simple prediction head on the shared interface representation described above: its weights are fit beforehand on designed variants of the training complexes only (no test complex is involved) and are frozen during optimization; the identical frozen regressor scores every method.

**Metrics.** DockQ is computed with the reference implementation of Basu & Wallner (2016) (`github.com/bjornwallner/DockQ`), applied to the CDR-H3–antigen interface consistent with Jin et al. (2022a); RMSD is the $C_\alpha$ RMSD of the full antibody (heavy and light chains) after Kabsch alignment, with per-CDR RMSDs aligned per loop; CAAR restricts AAR to binding residues within 6.6 Å of the epitope. Experiments run on two NVIDIA A800 GPUs (80 GB each).

## D Comparison with IgGM

IgGM (Wang et al., 2025) is a generative model for functional antibody design that reports CDR-level AAR on its own SAbDab split. We compare against it in Table 9 rather than in the main experiments, for three reasons. First, the test sets differ: IgGM evaluates on its own SAbDab split (post-2023 entries), while our unified protocol evaluates all tasks on RAbD (§4.1); placing IgGM in the main table would require either re-training on a different split or comparing numbers across incompatible test sets. Second, our training set is smaller: IgGM trains on a larger set of antibody–antigen complexes collected with additional filtering, so AbGaze is disadvantaged in this comparison. Third, IgGM models backbone atoms only, which precludes comparison on full-atom metrics such as side-chain lDDT or atom-level interface quality.

Despite these handicaps, AbGaze outperforms IgGM on five of six CDRs, with the largest margins on the binding-critical loops H2 (+5.0 points) and L3 (+4.2 points); IgGM leads only on L1 (+2.2 points).

Table 9: CDR-level AAR comparison with IgGM on its own SAbDab split (post-2023); IgGM numbers are from its published Table 2.

| CDR | AbGaze | IgGM | AbGaze wins |
|---|---|---|---|
| H1 | **.752** | .740 | ✓ |
| H2 | **.694** | .644 | ✓ |
| H3 | **.397** | .360 | ✓ |
| L1 | .728 | **.750** | × |
| L2 | **.753** | .743 | ✓ |
| L3 | **.677** | .635 | ✓ |

## E Diversity Analysis

As stated in Appendix A.6, decoding defines a distribution over designs whose concentration is controlled by $\tau$. We quantify this on the all-CDR design setting (five samples per target, $\tau = 0.5$): Table 10 reports, per CDR, the number of distinct sequences among the five samples and the mean pairwise normalized Hamming distance between them.

Table 10: Design diversity on all-CDR design (five samples per target, $\tau = 0.5$).

| CDR | Distinct (↑) | Diversity (↑) | % all-same |
|---|---|---|---|
| H1 | 2.60 | 0.092 | 21.7% |
| H2 | 2.93 | 0.127 | 16.7% |
| H3 | **4.92** | **0.458** | **1.7%** |
| L1 | 2.00 | 0.070 | 50.0% |
| L2 | 1.42 | 0.062 | 70.0% |
| L3 | 3.93 | 0.181 | 3.3% |
| All | 2.97 | 0.165 | 27.2% |

Diversity differs strongly across loops: CDR-H3, the loop that dominates binding specificity and tolerates the most sequence variation, is nearly saturated (4.92/5 distinct, pairwise Hamming 0.458, and only 1.7% of targets produce identical H3 across all five samples), while the framework-proximal L2 is the most conservative (1.42 distinct, 70% identical). This pattern is consistent with the greater sequence tolerance of H3 relative to the framework-proximal loops.

## F Gains without Diverse Sampling

The improvements reported in the main experiments do not rely on sampling diverse candidates. To show this, we replace the diverse decoding of Appendix A.6 with a single deterministic regression-style design: all masked positions of the target region are predicted and committed in one forward pass from the fully masked state, without the iterative commit-and-resample schedule or the temperature $\tau$. Table 11 evaluates this single-draw instantiation on all-CDR design under otherwise identical conditions (RAbD, same evaluator), against the full generative results of AbFlow and dyMEAN from Table 1. AAR/CAAR are pooled over CDR residues, and RMSD is the $C_\alpha$ RMSD of the full antibody (heavy and light chains) after Kabsch alignment.

Table 11: All-CDR design on RAbD without diverse sampling; notation as in Table 1.

| **Method** | **AAR↑** | **CAAR↑** | **RMSD↓** | **DockQ↑** | **lDDT↑** | **TM-Score↑** |
|---|---|---|---|---|---|---|
| dyMEAN | 60.1% | 50.3% | 1.357 | 0.396 | 0.803 | 0.965 |
| AbFlow | 59.7% | 49.8% | 1.104 | 0.379 | 0.815 | 0.971 |
| **AbGaze (Reg.)** | 64.0% | 54.9% | **1.021** | 0.407 | **0.834** | **0.974** |
| AbGaze (Full) | **66.2%** | **57.1%** | 1.052 | **0.422** | 0.831 | 0.973 |

Switching to the single deterministic design has little cost. Only AAR drops slightly, from 66.2% to 64.0%, and its contact-restricted variant likewise, while still leading the stronger baseline by

3.9 points; every other metric is essentially unchanged: mean RMSD improves (1.052 to 1.021 Å), lDDT and TM-Score are comparable, and DockQ shifts from 0.422 to 0.407, above both baselines (0.396/0.379). A single deterministic regression-style design therefore already outperforms the full generative models of AbFlow and dyMEAN on every metric: the advantage stems from the orientation-aware interface representation itself, not from diverse sampling.

We nevertheless retain the diverse absorbing-state decoding in the main protocol, for reasons orthogonal to benchmark performance. First, parity: every compared method is sampled and selected identically (§4.1), so the stochastic protocol is the like-for-like configuration. Second, practice: antibody design workflows consume a *pool* of diverse candidates for screening, filtering, and best-of-$N$ retrieval, whereas a deterministic regressor collapses the pool to a single mode per target; diversity concentrates on the loops with the most design freedom (Appendix E). Diverse generative decoding is thus adopted because it matches the actual requirements of the design task.

## G Hyperparameters

Table 12 lists all hyperparameters; none is tuned per task.

Table 12: Hyperparameters.

| Hyperparameter | Value | Description |
|---|---|---|
| embed dim | 64 | residue/atom embedding |
| hidden size | 128 | message-passing hidden size |
| encoder layers | 3 | equivariant encoder layers |
| refinement rounds $R$ | 3 | rounds per decoding step |
| $k$ (neighbors) | 9 | KNN neighbors per residue |
| atom channels $C$ | 14 | 4 backbone + $\leq$10 side-chain atoms |
| surface size $M$ | 50 | vertices per epitope residue |
| RBF bases / cutoff | 16 / 10 Å | distance embedding |
| $\lambda_{\text{FAPE}}$ / $\lambda_{\text{torsion}}$ | 0.5 / 0.2 | loss weights |
| $d_{\max}$ | 10 Å | FAPE clamping |
| reveal rounds $K$ | 9 | unmasking steps ($T{=}K{+}1{=}10$) |
| temperature $\tau$ | 0.5 | candidate diversity |
| prediction retries | 10 | draws ranked by predicted RMSD |
| affinity candidates | 30 | optimized designs per target |
| optimization steps / lr | 5 / 1.0 | ascent on template noise |
| optimizer / lr | Adam / $10^{-3} \to 10^{-4}$ | exponential decay |
| gradient clipping | 1.0 | max global norm |
| sequence-loss warmup | 10 epochs | linear ramp of $\lambda_{\text{seq}}$ |
| batch size / epochs | 16 / 500 | |
| hardware | 2$\times$ NVIDIA A800 (80 GB) | |

## Reproducibility Statement

AbGaze is trained on public data (SAbDab; RAbD for evaluation) with public tools (MSMS for surfaces). All baselines are re-run under the unified protocol of §4.1. Code and pretrained checkpoints will be released upon publication.

## Ethics Statement

This work is computational; no wet-lab experiments are involved. Learned antibody models are dual-use in principle; we evaluate only on public benchmarks and release artifacts for reproducibility.

## THE USE OF LARGE LANGUAGE MODELS (LLMS)

In preparing this manuscript we used LLMs (ChatGPT and Claude) in the following ways, in each case under author direction, with the authors verifying all outputs and taking full responsibility for the final content.

**Writing aid.** LLMs were used to polish and edit the manuscript prose, including grammar, clarity, and cross-section consistency. All suggestions were reviewed and revised by the authors.

**Retrieval and discovery.** LLMs were used to survey related work and suggest candidate references. Every retained citation was verified by the authors against official bibliographic records (Crossref, the ACL Anthology, PMLR, and OpenReview); this audit also identified and corrected several mismatched entries in our reference list.

**Drafting.** LLMs were used to draft portions of the manuscript from author-provided technical content, outlines, and experimental results. All drafted text was checked against the codebase and experimental logs and revised by the authors.

All experiments were designed and executed by the authors, all reported numbers come from author-run experiments, and all scientific decisions rest with the authors.