IgGM2: An All-Atom Foundation Model for Adaptive Immune Receptor Design

Anonymous authors
Paper under double-blind review

Abstract

Accurate immune receptor design requires modeling the coupled variation of amino-acid sequence, full-atom conformation, and target-binding geometry across antibodies, nanobodies, and T-cell receptors (TCRs). Existing methods often address only part of this problem, either by separating structure generation from sequence design, relying on fixed-backbone inverse folding, or focusing on a single receptor class. We introduce IgGM2, a unified all-atom generative framework for structure prediction and CDR sequence–structure co-design. For structure prediction, IgGM2 models receptor monomers from sequence and receptor–target complexes conditioned on fixed target structures and epitope context. For CDR design, a two-stage sampler first jointly generates CDR sequences and backbone conformations, then refines full-atom receptor structures with the generated sequences fixed. This allows receptor frameworks to adapt to designed CDRs without separate inverse folding or external side-chain packing. On structure prediction benchmarks, IgGM2 achieves strong performance on FoldBench and in TCR–pMHC modeling. On sequence design benchmarks, IgGM2 achieves competitive amino-acid recovery and improves Rosetta-based interface preference metrics, suggesting more favorable generated binding interfaces. These results support IgGM2 as a unified all-atom framework for adaptive immune receptor structure prediction and design.

1 Introduction

Antibodies and T-cell receptors (TCRs) are key molecular mediators of adaptive immunity. Both recognize cognate targets through hypervariable complementarity-determining regions (CDRs), whose sequences and conformations define binding specificity and shape the paratope–epitope interface. Accurately modeling and engineering these immune receptors is therefore central to modern therapeutic discovery, from antibody development to TCR-based cell therapy. This naturally leads to two coupled computational settings: full-atom **structure prediction** of antibody/TCR–target complexes, and target-conditioned **sequence design** of CDRs together with their binding conformations. Despite these shared molecular principles, existing methods typically study antibodies and TCRs separately, and likewise treat prediction and design as largely independent problems.

Recent advances in protein binder design have enabled target-conditioned generation of novel binding proteins (Bennett et al., 2026; Watson et al., 2023; Zhang et al., 2025). However, many methods still follow a modular paradigm, where a binder structure is first generated and its sequence is then recovered by inverse folding (Dauparas et al., 2022; Watson et al., 2023). Although practical, this strategy treats structure generation and sequence design as separate problems, rather than modeling their coupling directly. Recent all-atom models, including RFdiffusion3, Pallatom, and SeedProteo, extend protein generation beyond backbone coordinates (Butcher et al., 2025; Qu et al., 2024; 2025). Antibody-specific methods also explore joint sequence–structure design under different input conditions: DiffAb conditions on an antigen–framework complex obtained experimentally or by docking, whereas dyMEAN takes an epitope and an incomplete antibody sequence as input (Kong et al., 2023; Luo et al., 2022). Building on these advances, we unify immune receptor structure prediction and CDR sequence–structure co-design across receptor classes through explicit residue prediction and full-atom generation, allowing framework conformations to adapt to designed CDRs.

We introduce **IgGM2**, a unified all-atom framework for immune receptor structure prediction and target-conditioned CDR sequence–structure co-design. As illustrated in Figure 1a, IgGM2 models the

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.

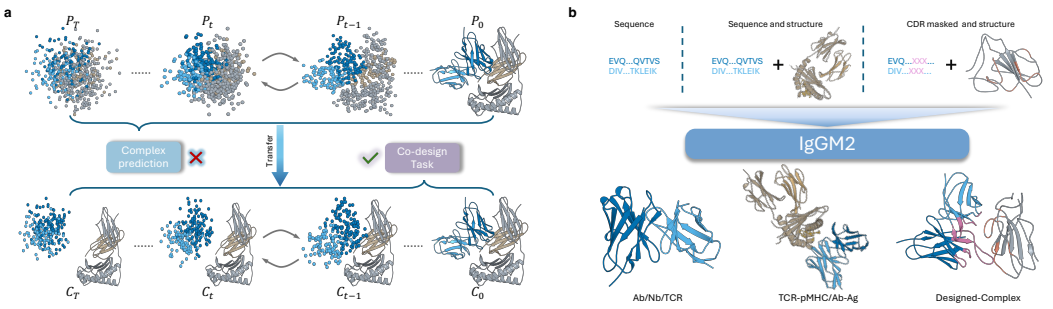


Figure 1: Overview of IgGM2. **(a)** IgGM2 performs receptor-atom diffusion conditioned on fixed structural context for both immune-complex prediction and CDR co-design. **(b)** It supports antibody/nanobody–antigen and TCR–pMHC complex prediction, antibody and TCR monomer prediction, and framework-conditioned CDR co-design.

conformations and binding poses of antibodies, nanobodies, and TCRs conditioned on fixed antigen or pMHC structures. For CDR design, IgGM2 uses two complementary design choices. First, it couples sequence generation with structural denoising through a two-stage sampler, **rather than generating a backbone followed by inverse folding**. The sampler jointly generates CDR sequences and backbone conformations, then refines full-atom receptor structures with the generated sequences fixed, allowing framework geometry to adapt to designed CDRs. Second, IgGM2 predicts residue identities explicitly and uses atom14 placeholders to maintain a fixed-size coordinate representation. Residue-specific masks restrict structural supervision to valid atom slots, while the generated sequences determine which slots are used during full-atom refinement.

As shown in Figure 1b, IgGM2 covers antibody, nanobody, TCR, and immune-complex modeling, and is instantiated as two task-specific variants: **IgGM2-P** for structure **P**rediction and **IgGM2-D** for sequence–structure co-**D**esign. Together, IgGM2 provides a structure-to-design route for immune receptor modeling, enabling full-atom CDR generation without separate inverse folding or external sidechain packing. Our main contributions are as follows:

- We develop IgGM2, a unified framework for antibody, nanobody, and TCR monomer prediction, immune-complex prediction, and target-conditioned receptor design.
- We combine noise-aligned sequence masking, mixed training, and two-stage Atom14 sampling to co-design CDR sequences and all-atom structures.
- IgGM2 achieves strong performance on structure prediction and CDR co-design benchmarks, demonstrating its effectiveness across antibody and TCR modeling tasks.

2 Related Work

Antibody modeling and design. Deep learning has rapidly advanced antibody structure prediction and generative design. Structure predictors such as IgFold and ABodyBuilder2 provide accurate variable-domain modeling (Abanades et al., 2023; Ruffolo et al., 2023), while generative models such as DiffAb and dyMEAN introduced antigen-conditioned CDR design (Kong et al., 2023; Luo et al., 2022). Later frameworks such as FlowDesign and AbFlow improved sequence-structure co-design (Wang et al., 2026; Wu et al., 2025c). More recent systems, such as Latent-X2 and Origin-1, have moved toward full-atom, developability-aware antibody design through increasingly integrated design pipelines (Kenlay et al., 2025; Levine et al., 2026). Despite this progress, most existing methods remain focused on antibodies alone rather than broader immune receptor modeling.

TCR modeling and design. TCR modeling has also progressed, though it remains less developed than antibody design. Models such as TCRModel2 and TCRBuilder2 improve structural prediction (Abanades et al., 2023; Yin et al., 2023), while TCR-VALID, TCR-TRANSLATE, TCRdesign, and TCR-epiDiff explore sequence generation and epitope-conditioned design (Karthikeyan et al., 2025; Leary et al., 2024; Li et al., 2025; Seo & Rhee, 2025). However, most current TCR approaches still focus primarily on sequence-level generation or partial structural design, with relatively few unified full-atom frameworks.

Modular and general binder design. General binder design methods commonly rely on either modular or integrated pipelines. Two-stage approaches such as RFdiffusion combined with ProteinMPNN first generate structures and then recover sequences (Dauparas et al., 2022; Watson et al., 2023), a strategy also used in RFantibody (Bennett et al., 2026). More recent frameworks, including ODesign and PPIFlow, move toward more integrated all-atom generation while still often incorporating inverse folding, refinement, or affinity maturation stages (Yu et al., 2026; Zhang et al., 2025). These methods have substantially expanded binder design capabilities, but sequence and structure generation are often still partially separated.

3 METHOD

3.1 OVERVIEW

IgGM2 supports immune receptor structure prediction and target-conditioned CDR sequence–structure co-design. For immune receptor modeling, we train **IgGM2-P** as a context-conditioned structure prediction model for antibodies, nanobodies, and TCRs. For monomer prediction, IgGM2-P generates the full immune receptor structure from sequence. For complex prediction, the antigen or pMHC is fixed as the target context, and the immune receptor structure is generated around it. **IgGM2-D** performs target-conditioned CDR co-design. In this setting, the target structure is fixed, and non-designed receptor residues keep their original amino-acid identities, while their coordinates are still generated together with those of the designed CDRs. This allows the framework geometry to adapt to CDR sequence and conformational changes, rather than being treated as a rigid scaffold. IgGM2-D couples sequence and structure generation by aligning discrete sequence masking with continuous EDM coordinate noise, and uses staged inference to first resolve CDR sequences and backbone geometry before refining the full-atom structure.

3.2 PRELIMINARY: EDM PRECONDITIONING

For coordinate diffusion, we adopt the EDM parameterization (Karras et al., 2022). Given clean atom coordinates $\boldsymbol{x}_0 \in \mathbb{R}^{N\times 3}$, noisy coordinates at noise level σ are defined as

$$\boldsymbol{x}_\sigma = \boldsymbol{x}_0 + \sigma\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}). \tag{1}$$

The denoiser is parameterized as

$$D(\boldsymbol{x}_\sigma;\sigma) = c_{\text{skip}}(\sigma)\,\boldsymbol{x}_\sigma + c_{\text{out}}(\sigma)\,F_\theta\big(c_{\text{in}}(\sigma)\,\boldsymbol{x}_\sigma;\sigma\big), \tag{2}$$

where $c_{\text{in}}(\sigma)$, $c_{\text{out}}(\sigma)$, and $c_{\text{skip}}(\sigma)$ follow the standard EDM formulation. In particular,

$$\hat{\boldsymbol{x}}_0 = D(\boldsymbol{x}_\sigma;\sigma)$$

denotes the denoised estimate of the clean coordinates used for coordinate supervision and sampling updates.

$$c_{\text{skip}}(\sigma) = \frac{\sigma_{\text{data}}^2}{\sigma^2 + \sigma_{\text{data}}^2}. \tag{3}$$

We regard $c_{\text{skip}}(\sigma)$ as a coordinate-side information level, which provides a continuous schedule for matching sequence visibility to structure denoising.

3.3 CONTEXT-CONDITIONED STRUCTURE PREDICTION

We build IgGM2-P on an AF3-like diffusion architecture for immune receptor structure prediction (Abramson et al., 2024). Unlike AlphaFold-style models that rely on MSA and template inputs, we remove both modules and use a single-sequence structure model (as shown in Figure 2a), following recent observations that MSA provides limited additional benefit for antibody and TCR modeling. Since our goal is to transfer a strong structure prediction model toward immune receptor co-design, we initialize the shared AF3-like modules from pretrained Protenix parameters (Team et al., 2025) and fine-tune them on immune receptor data. To support target-conditioned prediction, we further incorporate a structural constraint embedding following ODesign (Zhang et al., 2025).

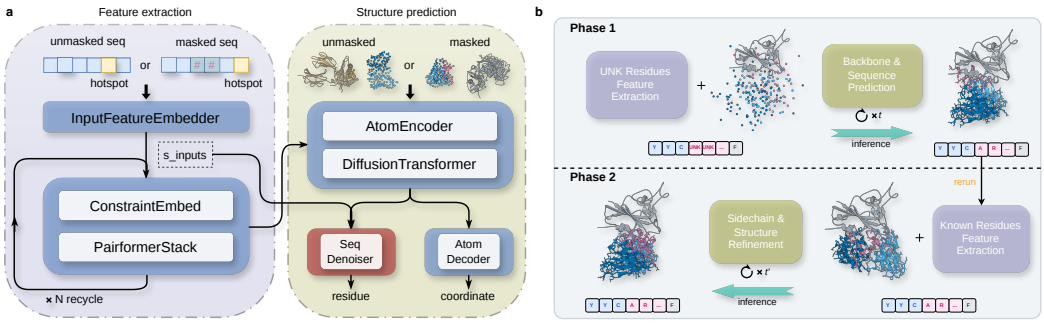


Figure 2: IgGM2 architecture. **(a)** The structure model injects fixed-context information into the Pairformer trunk and denoises only generated atoms. IgGM2-P uses this architecture without the sequence denoiser, while IgGM2-D adds the orange sequence-denoiser module for CDR sequence–structure co-design. **(b)** For CDR co-design, IgGM2-D uses a two-phase sampler: it first predicts CDR sequence and backbone geometry with masked CDR identities, then fixes the predicted sequence and refines the full-atom structure.

IgGM2-P supports monomer and complex structure prediction. For monomer prediction, including antibody, nanobody, and TCR modeling, no structural context is provided, and all receptor atoms are generated by diffusion. For complex prediction, such as antibody–antigen or TCR–pMHC modeling, the target structure is kept fixed, and only the receptor atoms are diffused and denoised. In target-conditioned settings, hotspot residues can also be provided to guide interface modeling.

The structural constraint embedding provides the network with information about fixed structural contexts. Fixed atoms are marked by binary masks, and their geometric features are projected into the pair representation. Hotspot residues are injected through the same conditioning pathway as token-level features. This allows the Pairformer trunk to use the fixed-context information, while the diffusion module updates only the generated atoms.

During training, coordinate supervision is applied only to generated atoms. Let m_{gen} denote the atom-level mask of generated regions. The coordinate loss is

$$\mathcal{L}_{\text{coord}} = \frac{1}{|m_{\text{gen}}|} \left\| m_{\text{gen}} \odot (\hat{\boldsymbol{x}}_0 - \boldsymbol{x}_0) \right\|_2^2 . \tag{4}$$

For monomer prediction, m_{gen} covers all atoms. For target-conditioned complex prediction, atoms in the fixed context are excluded from the coordinate loss. The full structure prediction objective is

$$\mathcal{L}_{\text{struct}} = \lambda_{\text{mse}}\mathcal{L}_{\text{coord}} + \lambda_{\text{lddt}}\mathcal{L}_{\text{lddt}} + \lambda_{\text{dist}}\mathcal{L}_{\text{dist}}. \tag{5}$$

3.4 Sequence Mask Rate Alignment

The EDM preconditioning in Section 3.2 provides a continuous notion of coordinate reliability through $c_{\text{skip}}(\sigma)$. We use this quantity to define a matched corruption schedule for discrete sequence tokens. We couple discrete sequence corruption to the same diffusion noise level σ. Let $p(\sigma)$ denote the masking probability for sequence tokens, such that a fraction $1 - p(\sigma)$ remains visible and the remaining tokens are masked for conditional prediction.

To maintain consistent information exposure across sequence and structure, we align the visible sequence fraction with this coordinate reliability:

$$1 - p(\sigma) = c_{\text{skip}}(\sigma). \tag{6}$$

This yields

$$p(\sigma) = \frac{\sigma^2}{\sigma^2 + \sigma_{\text{data}}^2}. \tag{7}$$

Under this schedule, low structural noise corresponds to minimal sequence masking, while high structural noise progressively increases masking toward fully de novo sequence generation. This

Algorithm 1 IgGM2-D Phase-Split Sampling

Input: trained IgGM2-D model f_θ, fixed context x_{cond}, CDR mask m_{CDR}, noise schedule $\{\sigma_i\}_{i=0}^{N}$, switch step T_{switch}
Output: designed CDR sequence $\hat{s}_{\text{CDR}}$ and full-atom structure $\hat{x}$
Initialize generated coordinates $x_0 \sim \mathcal{N}(0, \sigma_{\max}^2 I)$ and keep x_{cond} fixed
Mask CDR residue identities as UNK
Compute trunk representation h_{UNK} with masked CDR identities
▷ Phase 1: sequence and backbone generation
for $i = 0, \ldots, T_{\text{switch}} - 1$ **do**
    $(\hat{x}_{\text{bb}}, \ell_{\text{CDR}}) \leftarrow f_\theta(x_i, \sigma_i, h_{\text{UNK}}, \text{predict_seq} = \text{true})$
    Update generated backbone coordinates with the EDM sampler
    Keep fixed-context atoms unchanged
end for
▷ Phase transition: decode and fix CDR sequence
$\hat{s}_{\text{CDR}} \leftarrow \text{Decode}(\ell_{\text{CDR}})$
Recompute trunk representation $h_{\hat{s}}$ using $\hat{s}_{\text{CDR}}$
▷ Phase 2: full-atom refinement
for $i = T_{\text{switch}}, \ldots, N - 1$ **do**
    $\hat{x}_{\text{full}} \leftarrow f_\theta(x_i, \sigma_i, h_{\hat{s}}, \text{predict_seq} = \text{false})$
    Update all generated atoms with the EDM sampler
    Keep fixed-context atoms unchanged
end for
return $\hat{s}_{\text{CDR}}, \hat{x}_{\text{full}}$

defines a shared information-exposure schedule for sequence and structure, rather than assuming that sequence masking follows the same physical corruption process as coordinate diffusion. A more detailed signal-to-noise interpretation of this alignment is provided in Appendix A.2. We also compare against a linear schedule that does not align sequence masking with continuous structural noise (Yang et al., 2026), with results reported in Appendix E.1.

3.5 CDR Sequence–Structure Co-Design

We extend the structure prediction model into **IgGM2-D**, a target-conditioned CDR sequence–structure co-design model. As shown in Figure 2a, we add a sequence denoiser on top of the shared trunk to predict the residue identities of designed CDR positions.

Mixed training. To enable sequence design while preserving full-atom structural modeling, we train IgGM2-D with a mixed strategy. For each training sample, we select one of two modes according to a mixing rate ρ (see Appendix B for the mixing rate). In sequence-design mode, CDR residue identities are masked using the noise-aligned schedule in Eq. 7, and the model predicts both CDR sequences and backbone structures. In structure-prediction mode, the ground-truth CDR sequence is provided, and the model predicts full-atom structures, including sidechains, without applying a sequence loss. The training objective is

$$\mathcal{L}_{\text{mix}} = \mathcal{L}_{\text{struct}} + \lambda_{\text{seq}} \mathcal{L}_{\text{seq}}, \tag{8}$$

where $\mathcal{L}_{\text{seq}}$ is applied only in sequence-design mode, and $\mathcal{L}_{\text{struct}}$ follows the structure loss defined in Section 3.3. In sequence-design mode, the CDR residue identities are unknown, so structural supervision mainly focuses on CDR backbone atoms together with sequence recovery. In structure-prediction mode, the ground-truth CDR sequence is given, allowing full-atom supervision over both backbone and sidechain coordinates. This mixed training exposes IgGM2-D to both design and reconstruction signals, enabling it to learn CDR sequence generation while preserving accurate full-atom structural modeling.

Staged inference. At inference time, IgGM2-D uses a staged sampling procedure controlled by a noise threshold σ_{switch}, as shown in Figure 2b. In the high-noise stage, CDR identities are treated as unknown, and the model jointly predicts CDR residue types and backbone geometry. When the noise level reaches σ_{switch}, the predicted CDR sequence is decoded and fixed. The model then continues denoising in the low-noise stage, using the fixed sequence to refine the full-atom structure, including CDR sidechains. This staged procedure separates sequence determination from final full-atom

refinement, enabling IgGM2-D to generate CDR sequences and all-atom structures without a separate inverse-folding model or external sidechain-packing tool. Detailed training and sampling procedures are provided in Algorithms 1 and S1.

4 EXPERIMENT

We evaluate IgGM2 on structure prediction of antibody–antigen and TCR–pMHC complexes and on CDR sequence–structure co-design for paired antibodies, nanobodies, and TCRs. Ablation studies further assess the effects of sampling-stage transitions and epitope conditioning. Section 4.1 details the datasets, metrics, and baselines, with definitions applied consistently throughout all experiments.

4.1 SETUP

Data. Training data are curated from SAbDab (Dunbar et al., 2014) and STCRDab (Leem et al., 2018). Since our study includes both structure prediction benchmarking against models such as AlphaFold-3 and Chai-1 (Abramson et al., 2024; team et al., 2024) and sequence design experiments, we construct two separate datasets for these tasks. Both datasets are first deduplicated at the sequence level and then clustered by CDRH3 using MMseqs with `min_seq_id = 0.8` and `cov_mode = 1`. For structure prediction, the training set includes structures released before September 30, 2021. To ensure fair evaluation, we further remove training samples with full-sequence similarity above 95% to any sample in FoldBench (Xu et al., 2025) or the other test benchmarks. For sequence design, we retain only complexes with receptor–target information, as the task is to design binders under antigen-conditioned settings. We use January 1, 2024 as the temporal cutoff, assigning earlier samples to training and later samples to validation and test sets. We also apply CDRH3-based cluster splitting to ensure that CDRH3 sequence similarity between training and test sets is below 0.8.

Benchmarks. For structure prediction, we evaluate on FoldBench (Xu et al., 2025), SAbDab-22H2-AbAg (Dunbar et al., 2014), and STCRDab-22-TCR_pMHC (Leem et al., 2018). Supplementary results on SAbDab-22H2-Ab, SAbDab-22H2-Nano, and STCRDab-22-TCR are reported in Appendix C. For sequence design, the main evaluation uses a CDRH3-deduplicated SAbDab test set (203 complexes; 148 paired antibodies and 55 nanobodies) and VDJdb-2025-TCR50, comprising 50 TCR–pMHC complexes with reconstructed TCR sequences and AF3-predicted reference structures. Dataset construction is detailed in Appendices D and D.6. Full-set SAbDab results are provided in Appendix D, and additional results on five STCRDab TCR–pMHC complexes are reported in the appendix (Table S16).

Metrics. Structure prediction is evaluated using DockQ, success rate (SR; DockQ ≥ 0.23) and TM-Score. For sequence design, the main tables report CDR amino-acid recovery (AAR), framework-aligned CDR Cα RMSD, and Rosetta interface-energy metrics, including IMP and $dG_{\text{separated}}$ (`ref2015`). Co-generated DockQ / TM-Score / Cα RMSD / SR metrics on the full clean test set are reported in Appendix D. Backbone-only design baselines are first side-chain-packed with Rosetta `PackRotamersMover` and then locally relaxed before IMP is computed; full-atom outputs go directly into local `FastRelax`. For TCR design, validity is the fraction of samples for which both α- and β-chain variable regions are recognized as single domains and successfully IMGT-numbered by ANARCI (Dunbar & Deane, 2016). We also report $\beta 3$ P_{gen}, the fraction of $\beta 3$ junction sequences with zero generation probability under OLGA's (Sethna et al., 2019) default human TRB model.

Baselines. We consider a broad range of prior methods for structure prediction (AlphaFold-3 / -Multimer, Protenix, Boltz-1, HelixFold-3, Chai-1, OpenFold-3, tFold, TCRModel2, IgFold, ImmuneBuilder, ESMFold, OmegaFold, EquiFold, HelixFold-Single) (Abanades et al., 2023; Abramson et al., 2024; Evans et al., 2021; Fang et al., 2023; Lee et al., 2022; Lin et al., 2023; Liu et al., 2024; Ruffolo et al., 2023; Team et al., 2025; team et al., 2024; The OpenFold3 Team, 2025; Wohlwend et al., 2025; Wu et al., 2025a;b; 2022; Yin et al., 2023) and for CDR co-design (DiffAb, dyMEAN, AbX, RFAntibody, IgGM) (Bennett et al., 2026; Kong et al., 2023; Luo et al., 2022; Wang et al., 2025; Zhu et al., 2024). Since these models are not all evaluated on the same benchmarks, we do not enforce a unified baseline set across all experiments; instead, each benchmark is compared against the methods with publicly reported results available, following the evaluation protocol of the corresponding prior work. For CDR co-design, baselines re-trained from scratch on the training data are marked with *; baselines without * use their released checkpoints.

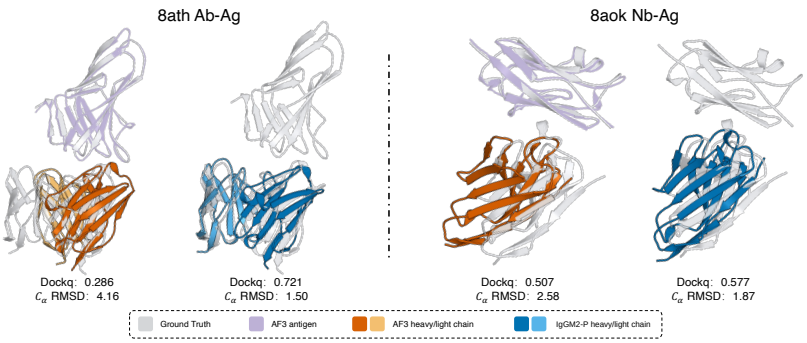


Figure 3: FoldBench predictions. Left: 8ATH antibody–antigen complex. Right: 8AOK nanobody–antigen complex.

4.2 Structure Prediction

We evaluate IgGM2-P as a binder-only structure predictor on two main immune-complex benchmarks: FoldBench for antibody/nanobody–antigen complexes and STCRDab-22-TCR_pMHC for TCR–pMHC complexes. IgGM2-P uses a single 1×1 sample without candidate selection across both benchmarks. Additional results on SAbDab-22H2-AbAg complex prediction and separate antibody, nanobody, and TCR monomer prediction benchmarks are provided in Appendix C.

Table 1: FoldBench antibody-antigen prediction. Baselines follow the official leaderboard settings, while IgGM2-P uses a single 1×1 sample without candidate ranking because it has no confidence head.

Method	SR (%) ↑	Mean DockQ ↑
Chai-1	23.6	0.17
HelixFold-3	28.4	0.21
Boltz-1	33.5	0.22
Protenix-0.5	34.1	0.23
AlphaFold-3	47.9	0.36
Protenix-1.0	48.8	0.36
IgGM2-P(epitope)	**73.8**	**0.60**

Antibody/nanobody–antigen complex prediction. On FoldBench, IgGM2-P achieves a strong success rate of 73.8% using only a single 1×1 sample without candidate ranking, outperforming the official leaderboard baselines (Table 1). This setting is also more efficient: the compared baselines follow a 5×5 sampling protocol, generating multiple candidates across five seeds and selecting the best-ranked prediction, whereas IgGM2-P produces one prediction directly without post-hoc selection. These results support the effectiveness of our binder-only formulation when the antigen is available as a fixed structural context. Representative examples further show that IgGM2-P preserves antibody/nanobody–antigen interface geometry across both receptor types (Figure 3). For the 8ATH antibody–antigen complex, IgGM2-P improves DockQ from 0.286 to 0.721 and reduces global RMSD from 4.16 Å to 1.50 Å compared with AlphaFold-3. For the 8AOK nanobody–antigen complex, it improves DockQ from 0.507 to 0.577 and reduces global RMSD from 2.58 Å to 1.87 Å. These examples indicate that fixing the antigen context helps IgGM2-P accurately place and model immune receptor binders.

Since IgGM2-P conditions on fixed target structures and can incorporate epitope-level guidance, we further evaluate it on SAbDab-22H2-AbAg against methods that use comparable contact or epitope restraints. As shown in Table S4, IgGM2-P outperforms existing epitope-conditioned baselines, demonstrating that epitope-guided structural constraints provide effective interface guidance.

STCRDab-22-TCR_pMHC complex prediction. TCR–pMHC prediction is particularly challenging because the complex contains multiple interacting components, including the TCR, peptide, and MHC molecules. Rather than predicting the full complex from scratch, IgGM2-P keeps the pMHC structure fixed and diffuses only the TCR binder. As shown in Table 2, this formulation achieves the best overall performance, with DockQ 0.746, full-complex GDT 0.887, and acceptable/medium/high success rates of $100.0/88.9/38.9\%$. Additional prediction examples are provided in Appendix F.1. These results show that binder-only diffusion is effective for TCR–pMHC modeling, where the peptide–MHC structure provides a strong conditioning context.

4.3 CDR Sequence–Structure Co-Design

Following the evaluation setup in Section 4.1, we report the main sequence-design results on the CDRH3-deduplicated SAbDab test subset. This subset includes paired antibodies and nanobodies,

Table 2: TCR–pMHC complex prediction on STCRDab-22-TCR_pMHC (18 complexes). Baselines use official FASTA/native-PDB inputs. Acc., Med., and High SR are DockQ success rates at thresholds 0.23, 0.49, and 0.80.

Method	DockQ ↑	LRMS (Å) ↓	iRMS (Å) ↓	GDT ↑	Acc. SR (%) ↑	Med. SR (%) ↑	High SR (%) ↑
TCRModel2	0.585	7.749	2.932	0.795	88.9	66.7	16.7
AlphaFold-3	0.632	5.316	2.017	0.816	94.4	77.8	33.3
tFold-TCR	0.523	5.585	2.242	0.774	100.0	55.6	5.6
AlphaFold-Multimer	0.537	6.052	2.293	0.775	100.0	44.4	16.7
IgGM2-P(epitope)	**0.746**	**2.966**	**1.262**	**0.887**	**100.0**	**88.9**	**38.9**

Table 3: Paired-antibody CDR co-design performance on the CDRH3-deduplicated SAbDab design test subset. Higher AAR and lower framework-aligned CDR Cα RMSD indicate better performance. Best values are bolded and second-best values are underlined. HDOCK is used for DiffAb and AbX when external docking is required.

Method	CDR AAR ↑							Framework-aligned Cα RMSD (Å) ↓					
	H1	H2	H3	L1	L2	L3	Avg	H1	H2	H3	L1	L2	L3
DiffAb*	0.648	0.374	0.245	0.619	0.591	0.444	0.474	2.315	3.596	8.882	4.476	5.896	6.667
AbX	**0.759**	0.678	**0.410**	**0.763**	0.588	**0.638**	**0.610**	1.607	1.462	5.454	1.548	1.187	2.134
dyMEAN*	0.713	**0.685**	0.389	0.718	**0.773**	0.601	0.596	1.356	1.282	**3.480**	1.388	1.056	1.408
IgGM	0.621	0.400	0.243	0.651	0.730	0.592	0.482	1.481	1.542	4.047	1.503	1.085	1.515
RFAntibody	0.216	0.085	0.103	0.067	0.236	0.128	0.126	4.858	5.052	6.413	6.201	8.625	3.466
IgGM2-D	0.722	0.640	0.392	0.714	0.748	0.605	0.590	**1.011**	**0.940**	3.492	**0.953**	**0.674**	**1.231**

Table 4: Nanobody CDR co-design performance on the CDRH3-deduplicated SAbDab design test subset. AAR is higher better; framework-aligned CDR Cα RMSD is lower better.

Method	CDR AAR ↑				Cα RMSD (Å) ↓		
	H1	H2	H3	Avg	H1	H2	H3
DiffAb*	0.487	0.335	0.177	0.287	2.506	1.794	4.992
IgGM	0.468	0.291	0.154	0.267	2.679	2.311	5.020
RFAntibody	0.181	0.097	0.115	0.127	5.092	4.391	7.445
IgGM2-D	**0.497**	**0.499**	**0.238**	**0.372**	**2.328**	**1.597**	**4.374**

Table 5: Rosetta interface-energy metrics for CDR co-design on the CDRH3-deduplicated SAbDab design test subset.

Method	Paired antibodies		Nanobodies	
	IMP ↑	Mean dG ↓	IMP ↑	Mean dG ↓
DiffAb*	0.392	29.86	0.236	10.52
AbX	0.318	2.70	–	–
dyMEAN*	0.327	2191.46	–	–
IgGM	0.541	-17.83	0.273	-9.46
RFAntibody	0.466	118.90	0.327	1.25
IgGM2-D	**0.615**	**-23.27**	**0.527**	**-24.84**

with CDR recovery and local structural accuracy reported in Tables 3 and 4. We further report Rosetta-based interface metrics in Table 5. Full-set results and additional complex-quality metrics are provided in Appendices D.5 and D.

Paired antibody design performance. We first evaluate IgGM2-D on paired antibody CDR design, considering both sequence recovery and generated structural quality (Table 3). IgGM2-D does not always achieve the highest AAR, as AbX and dyMEAN recover more residues on some CDRs. However, AAR should be interpreted together with the design setting and the generated structure quality. AbX requires a completed antibody structure and external docking in our evaluation pipeline, rather than generating the bound complex end to end. It also mainly edits CDR residues while keeping the framework conformation fixed, which can be restrictive because CDR mutations may require framework-level adjustment to support the final binding geometry. IgGM shows lower AAR in our split, partly because many test antigens contain multiple chains, whereas IgGM was originally designed for single-chain antigen inputs. Although these antigens are merged into a single chain during evaluation, this input mismatch may still affect its performance. dyMEAN can generate antibody structures, but its interface-energy results are much weaker than IgGM2-D. As shown in Table 5, IgGM2-D achieves the best IMP and a mean dG of -23.27, suggesting more favorable binding geometries. Appendix D.4 provides further analysis of the interface-energy distribution and the effects of FastRelax. It also obtains the best framework-aligned CDR Cα RMSD on five of six paired-antibody CDRs, indicating that its sequence designs are supported by more coherent local conformations.

Table 6: All-six-CDR TCR co-design on VDJdb-2025-TCR50, comprising 50 AF3-predicted reference complexes (ipTM ≥ 0.8). Inverse-folding baselines receive reference TCR backbones, whereas IgGM2-D generates receptor structures. Validity measures successful ANARCI numbering of both chains. $\beta 3$ Pgen=0 is the fraction of $\beta 3$ sequences with zero OLGA generation probability.

Method	All-CDR ↑	$\alpha 1$ ↑	$\alpha 2$ ↑	$\alpha 3$ ↑	$\beta 1$ ↑	$\beta 2$ ↑	$\beta 3$ ↑	Validity ↑	$\beta 3$ Pgen=0 ↓
Rosetta FastDesign	0.390	0.355	0.576	0.315	0.229	0.281	0.528	100%	1.8%
ESM-IF1	0.421	0.267	0.353	**0.610**	0.188	0.302	0.516	100%	34.6%
ProteinMPNN	0.476	0.449	0.414	0.516	0.394	0.427	**0.548**	100%	9.6%
IgGM2-D	**0.563**	**0.670**	**0.648**	0.381	**0.802**	**0.752**	0.443	100%	**0.0%**

Table 7: Phase-transition ablation at seq:struct $30/70$ and backbone-noise 0.1. The selected step is bolded.

Metric	25	50	**75**	200
All AAR ↑	0.523	0.528	**0.530**	0.522
Paired AAR ↑	0.579	0.589	**0.587**	0.582
Nano AAR ↑	0.373	0.369	**0.379**	0.362
HCDR3 RMSD (Å) ↓	3.846	3.830	**3.702**	3.618
IMP ↑	0.595	0.561	**0.584**	0.543

Table 8: Epitope ablation and target-context comparisons. IgGM2-P fixes the target; Protenix-0.5 uses a pocket, AF3/Protenix-1.0 use GT templates.

Setting	SR ↑	Mean DockQ ↑	Median DockQ ↑
Protenix-0.5 + pocket	0.523	0.411	0.324
AF3 + GT template	0.517	0.421	0.271
Protenix-1.0 + GT template	0.517	0.401	0.300
IgGM2-P w/o epitope	0.610	0.498	0.605
IgGM2-P w/ epitope	**0.738**	**0.604**	**0.755**

Nanobody design performance. We next evaluate IgGM2-D on nanobody CDR co-design, as shown in Table 4. Nanobody design is more challenging because nanobodies usually contain a longer HCDR3 loop, making sequence recovery and local structure generation harder. IgGM2-D achieves the best average AAR, improving from 0.287 for the strongest baseline DiffAb to 0.372. On HCDR3, IgGM2-D improves AAR from 0.154 for IgGM to 0.238, corresponding to a 54.55% relative gain. It also achieves the best framework-aligned CDR Cα RMSD on all three nanobody CDRs. Consistently, Table 5 shows that IgGM2-D improves nanobody IMP from 0.327 to 0.527 and reaches a mean dG of -24.84, suggesting more favorable generated interfaces.

TCR design performance. On VDJdb-2025-TCR50, IgGM2-D achieves the highest all-six-CDR AAR (0.563; Table 6), with the best recovery on $\alpha 1$, $\alpha 2$, $\beta 1$, and $\beta 2$. ESM-IF1 and ProteinMPNN achieve the highest recovery on $\alpha 3$ and $\beta 3$, respectively. All methods achieve 100% validity. The fraction of $\beta 3$ sequences with $P_{\text{gen}} = 0$ is 0.0% for IgGM2-D versus 1.8% for Rosetta FastDesign, 9.6% for ProteinMPNN, and 34.6% for ESM-IF1. Inverse-folding baselines use reference TCR backbones, whereas IgGM2-D generates receptor structures under fixed-target conditioning. Results on five leak-free STCRDab TCR–pMHC targets are reported in Table S16.

4.4 Ablations

Phase-transition step. We first study the phase-transition step under a $30/70$ mixed-training setting, where structure-prediction samples are used more frequently to stabilize full-atom generation. With this ratio fixed, we vary T_{switch} to determine when the sampler should switch from sequence-and-backbone generation to full-atom refinement. We select the transition step mainly by All AAR, which measures overall sequence recovery across paired antibody and nanobody designs. As shown in Table 7, T_{switch}=75 achieves the highest All AAR of 0.530, although it does not give the best IMP.

Epitope conditioning and target-context comparison. On FoldBench, adding epitope information to target-fixed IgGM2-P improves SR from 0.610 to 0.738 and mean DockQ from 0.498 to 0.604 (Table 8). Protenix-0.5 with pocket input and AF3/Protenix-1.0 with GT antigen templates achieve SRs of 0.517–0.523 and mean DockQ scores of 0.401–0.421. Across these distinct conditioning settings, IgGM2-P achieves the highest SR and mean DockQ, highlighting the effectiveness of explicit target-conditioned receptor modeling and the additional benefit of epitope information.

5 Discussion

IgGM2 unifies immune receptor structure prediction and CDR sequence–structure co-design within a single all-atom diffusion framework. By transferring structural modeling to target-conditioned design, IgGM2 generates CDR sequences and all-atom structures without separate inverse folding or external sidechain packing. We discuss remaining limitations, including sequence recovery and sidechain refinement, in Appendix B. Future work will explore stronger sequence priors, improved all-atom optimization, developability-aware objectives, and experimental validation.

AI USE STATEMENT

Generative AI tools were used to assist with code implementation and debugging and to improve the wording, clarity, and style of the manuscript. They were not used to create experimental data or fabricate results. All scientific claims and conclusions were based on the authors' own analysis and judgment. We reviewed all AI-assisted text and code and take full responsibility for the research and the final manuscript.

REPRODUCIBILITY STATEMENT

All results reported in this paper are reproducible from the model, data, and evaluation settings specified in the manuscript.

REFERENCES

Brennan Abanades, Wing Ki Wong, Fergus Boyles, Guy Georges, Alexander Bujotzek, and Charlotte M Deane. Immunebuilder: Deep-learning models for predicting the structures of immune proteins. *Communications Biology*, 6(1):575, 2023.

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*, 630(8016):493–500, 2024.

Nathaniel R Bennett, Joseph L Watson, Robert J Ragotte, Andrew J Borst, DéJenaé L See, Connor Weidle, Riti Biswas, Yutong Yu, Ellen L Shrock, Russell Ault, et al. Atomically accurate de novo design of antibodies with rfdiffusion. *Nature*, 649(8095):183–193, 2026.

Jasper Butcher, Rohith Krishna, Raktim Mitra, Rafael I Brent, Yanjing Li, Nathaniel Corley, Paul T Kim, Jonathan Funk, Simon Mathis, Saman Salike, et al. De novo design of all-atom biomolecular interactions with rfdiffusion3. *bioRxiv*, 2025.

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 and Charlotte M Deane. Anarci: antigen receptor numbering and receptor classification. *Bioinformatics*, 32(2):298–300, 2016.

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.

Richard Evans, Michael O'neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, et al. Protein complex prediction with alphafold-multimer. *biorxiv*, pp. 2021–10, 2021.

Xiaomin Fang, Fan Wang, Lihang Liu, Jingzhou He, Dayong Lin, Yingfei Xiang, Kunrui Zhu, Xiaonan Zhang, Hua Wu, Hui Li, et al. A method for multiple-sequence-alignment-free protein structure prediction using a protein language model. *Nature Machine Intelligence*, 5(10):1087–1096, 2023.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022.

Dhuvarakesh Karthikeyan, Sarah N Bennett, Amy G Reynolds, Benjamin G Vincent, and Alex Rubinsteyn. Conditional generation of real antigen-specific t cell receptor sequences. *Nature Machine Intelligence*, 7(9): 1494–1509, 2025.

Henry Kenlay, Daniella Pretorius, Jonathan Crabbé, Alex Bridgland, Sebastian M Schmon, Agrin Hilmkil, James Vuckovic, Simon Mathis, Tomas Matteson, Rebecca Bartke-Croughan, et al. Drug-like antibodies with low immunogenicity in human panels designed with latent-x2. *arXiv e-prints*, pp. arXiv–2512, 2025.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. End-to-end full-atom antibody design. *arXiv preprint arXiv:2302.00203*, 2023.

Allen Y Leary, Darius Scott, Namita T Gupta, Janelle C Waite, Dimitris Skokos, Gurinder S Atwal, and Peter G Hawkins. Designing meaningful continuous representations of t cell receptor sequences with deep generative models. *Nature Communications*, 15(1):4271, 2024.

Jae Hyeon Lee, Payman Yadollahpour, Andrew Watkins, Nathan C Frey, Andrew Leaver-Fay, Stephen Ra, Kyunghyun Cho, Vladimir Gligorijević, Aviv Regev, and Richard Bonneau. Equifold: Protein structure prediction with a novel coarse-grained structure representation. *Biorxiv*, pp. 2022–10, 2022.

Jinwoo Leem, Saulo H P de Oliveira, Konrad Krawczyk, and Charlotte M Deane. Stcrdab: the structural t-cell receptor database. *Nucleic acids research*, 46(D1):D406–D412, 2018.

Simon Levine, Jonathan Edward King, Jacob Stern, David Grayson, Raymond Wang, Rui Yin, Umberto Lupo, Paulina Kulytė, Ryan Matthew Brand, Tristan Bertin, et al. Origin-1: a generative ai platform for de novo antibody design against novel epitopes. *bioRxiv*, pp. 2026–01, 2026.

Xiaokun Li, Qiang Yang, Long Xu, Weihe Dong, Kuanquan Wang, Suyu Dong, Wei Wang, Gongning Luo, Xianyu Zhang, Tiansong Yang, et al. Tcrdesign: an antigen-specific generative language model for de novo design of t-cell receptors. *Briefings in Bioinformatics*, 26(6):bbaf691, 2025.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

Lihang Liu, Shanzhuo Zhang, Yang Xue, Xianbin Ye, Kunrui Zhu, Yuxin Li, Yang Liu, Jie Gao, Wenlai Zhao, Hongkun Yu, et al. Technical report of helixfold3 for biomolecular structure prediction. *arXiv preprint arXiv:2408.16975*, 2024.

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.

Wei Qu, Jiawei Guan, Rui Ma, Ke Zhai, Weikun Wu, and Haobo Wang. P (all-atom) is unlocking new path for protein design. *bioRxiv*, pp. 2024–08, 2024.

Wei Qu, Yiming Ma, Fei Ye, Chan Lu, Yi Zhou, Kexin Zhang, Lan Wang, Minrui Gui, and Quanquan Gu. Seedproteo: Accurate de novo all-atom design of protein binders. *arXiv preprint arXiv:2512.24192*, 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.

Se Yeon Seo and Je-Keun Rhee. Tcr-epidiff: solving dual challenges of tcr generation and binding prediction. *Bioinformatics*, 41(Supplement_1):i125–i132, 2025.

Zachary Sethna, Yuval Elhanati, Curtis G Callan Jr, Aleksandra M Walczak, and Thierry Mora. Olga: fast computation of generation probabilities of b-and t-cell receptor amino acid sequences and motifs. *Bioinformatics*, 35(17):2974–2981, 2019.

ByteDance AML AI4Science Team, Xinshi Chen, Yuxuan Zhang, Chan Lu, Wenzhi Ma, Jiaqi Guan, Chengyue Gong, Jincai Yang, Hanyu Zhang, Ke Zhang, et al. Protenix-advancing structure prediction through a comprehensive alphafold3 reproduction. *BioRxiv*, pp. 2025–01, 2025.

Chai Discovery team, Jacques Boitreaud, Jack Dent, Matthew McPartlon, Joshua Meier, Vinicius Reis, Alex Rogozhonikov, and Kevin Wu. Chai-1: Decoding the molecular interactions of life. *BioRxiv*, pp. 2024–10, 2024.

The OpenFold3 Team. Openfold3-preview, 2025. URL https://github.com/aqlaboratory/openfold-3.

Rubo Wang, Fandi Wu, Xingyu Gao, Jiaxiang Wu, Peilin Zhao, and Jianhua Yao. Iggm: A generative model for functional antibody and nanobody design. In *International Conference on Learning Representations*, volume 2025, pp. 100801–100827, 2025.

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

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, July 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06415-8. URL http://dx.doi.org/10.1038/s41586-023-06415-8.

Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Noah Getz, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Liam Atkinson, Tally Portnoi, Itamar Chinn, et al. Boltz-1 democratizing biomolecular interaction modeling. *BioRxiv*, pp. 2024–11, 2025.

Fandi Wu, Yu Zhao, JiaXiang Wu, Biaobin Jiang, Bing He, Longkai Huang, Chenchen Qin, Yang Xiao, Fan Yang, Rubo Wang, et al. De novo design of epitope-specific antibodies via a structure-driven computational workflow. *Nature Communications*, 17(1):625, 2025a.

Fandi Wu, Yu Zhao, Yang Xiao, Chenchen Qin, Fang Wang, Zihan Wu, Long-Kai Huang, Xiao Liu, Jiangning Song, Bing He, et al. Fast and accurate modeling of tcr-peptide-mhc complexes using tfold-tcr. *bioRxiv*, pp. 2025–01, 2025b.

Jun Wu, Xiangzhe Kong, Ningguan Sun, Jing Wei, Sisi Shan, Fuli Feng, Feng Wu, Jian Peng, Linqi Zhang, Yang Liu, et al. Flowdesign: Improved design of antibody cdrs through flow matching and better prior distributions. *Cell Systems*, 16(6), 2025c.

Ruidong Wu, Fan Ding, Rui Wang, Rui Shen, Xiwen Zhang, Shitong Luo, Chenpeng Su, Zuofan Wu, Qi Xie, Bonnie Berger, et al. High-resolution de novo structure prediction from primary sequence. *BioRxiv*, pp. 2022–07, 2022.

Sheng Xu, Qiantai Feng, Lifeng Qiao, Hao Wu, Tao Shen, Yu Cheng, Shuangjia Zheng, and Siqi Sun. Benchmarking all-atom biomolecular structure prediction with foldbench. *Nature Communications*, 17(1):442, 2025.

Nianzu Yang, Songlin Jiang, Jian Ma, Huaijin Wu, Shuangjia Zheng, Wengong Jin, and Junchi Yan. Repurposing alphafold3-like protein folding models for antibody sequence and structure co-design. *Advances in Neural Information Processing Systems*, 38:215–255, 2026.

Rui Yin, Helder V Ribeiro-Filho, Valerie Lin, Ragul Gowthaman, Melyssa Cheung, and Brian G Pierce. Tcrmodel2: high-resolution modeling of t cell receptor recognition using deep learning. *Nucleic Acids Research*, 51(W1):W569–W576, 2023.

Qilin Yu, Liangyue Guo, Xiayan Qin, Xikun Huang, Baihui Tian, Hongzhun Wang, Yu Liu, Yunzhi Lang, Di Wang, Zhouhanyu Shen, et al. High-affinity protein binder design via flow matching and in silico maturation. *bioRxiv*, pp. 2026–01, 2026.

Odin Zhang, Xujun Zhang, Haitao Lin, Cheng Tan, Qinghan Wang, Yuanle Mo, Qiantai Feng, Gang Du, Yuntao Yu, Zichang Jin, et al. Odesign: A world model for biomolecular interaction design. *arXiv preprint arXiv:2510.22304*, 2025.

Tian Zhu, Milong Ren, and Haicang Zhang. Antibody design using a score-based diffusion model guided by evolutionary, physical and geometric constraints. In *Forty-first International Conference on Machine Learning*, 2024.

Appendix

CONTENTS

A Algorithm Details **14**

A.1 Mixed Training and Phase-Split Sampling 14

A.2 Effective SNR Interpretation of the Sequence Mask Schedule 14

A.3 Antibody-Centered Cropping 15

A.4 Hotspot Conditioning 16

B Training details and limitations **17**

B.1 Training details 17

B.2 Training–Test Leakage Control 17

B.3 Runtime Comparison 17

B.4 Limitations 18

C Supplementary Structure Prediction Results **18**

C.1 FoldBench Per-Subset Breakdown 20

C.2 SAbDab-22H2-AbAg Complex Prediction 20

C.3 SAbDab-22H2-Ab 20

C.4 SAbDab-22H2-Nano 20

C.5 STCRDab-22-TCR (unliganded TCR) 20

D Supplementary Sequence-Design Results **21**

D.1 SAbDab test-set construction 21

D.2 Co-Generated Complex Quality 21

D.3 Orthogonal Refolding of Designed Sequences 21

D.4 Interface-Energy Distribution and the Effect of Relaxation 22

D.5 Full-Set Sequence-Design Results 23

D.6 TCR Sequence–Structure Co-Design 23

E Supplementary Ablations **24**

E.1 Sequence-Masking Schedule 24

E.2 Backbone Noise and Sampling-Stage Ablation 24

F Visualization **25**

F.1 TCR–pMHC Complexes 25

F.2 Antibody 25

F.3 Nanobody 26

F.4 Unliganded TCR 26

A Algorithm Details

A.1 Mixed Training and Phase-Split Sampling

In the implementation, the noise-level threshold σ_{switch} is the only user-facing knob; the equivalent step index T_{switch} used in the main text is computed from σ_{switch} via Algorithm 1, and changes if the number of denoising steps N or the PC strength γ_0 changes. The training hyperparameter `trunk_cache.sigma_threshold` equals the inference-time σ_{switch}, so the phase transition at inference is placed exactly where sequence-design supervision hands over to structure-prediction supervision during training.

Algorithm S1 IgGM2-D Mixed Training

Input: minibatch $\{(\mathcal{I}^{(b)}, x_{\text{gt}}^{(b)}, s_{\text{gt}}^{(b)}, m_{\text{CDR}}^{(b)}, m_{\text{gen}}^{(b)})\}_{b=1}^{B}$
  sequence-design ratio ρ_{seq}, sequence masker $\mathcal{M}$, model f_θ
Output: updated model parameters θ
for each sample b in the minibatch **do**
  Sample a mode $z^{(b)} \in \{\text{design}, \text{structure}\}$ with $\Pr[z^{(b)} = \text{design}] = \rho_{\text{seq}}$
  Sample noise level σ and construct noisy coordinates only on generated atoms:

$$x_\sigma^{(b)} = x_{\text{gt}}^{(b)} + m_{\text{gen}}^{(b)} \odot \sigma\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I)$$

  if $z^{(b)} = \text{design}$ **then**
    Mask CDR residue identities with $\mathcal{M}(s_{\text{gt}}^{(b)}, \sigma, m_{\text{CDR}}^{(b)})$ to obtain $\mathcal{I}_{\text{mask}}^{(b)}$
    Predict denoised coordinates and CDR residue logits:

$$(\hat{x}^{(b)}, \ell_{\text{CDR}}^{(b)}) \leftarrow f_\theta(x_\sigma^{(b)}, \sigma, \mathcal{I}_{\text{mask}}^{(b)})$$

    Compute design-mode loss:

$$\mathcal{L}^{(b)} = \mathcal{L}_{\text{struct}}^{\text{bb}}(\hat{x}^{(b)}, x_{\text{gt}}^{(b)}) + \lambda_{\text{seq}}\text{CE}(\ell_{\text{CDR}}^{(b)}, s_{\text{CDR}}^{(b)})$$

  else
    Use ground-truth CDR residue identities to construct $\mathcal{I}_{\text{gt}}^{(b)}$
    Predict denoised full-atom coordinates:

$$\hat{x}^{(b)} \leftarrow f_\theta(x_\sigma^{(b)}, \sigma, \mathcal{I}_{\text{gt}}^{(b)})$$

    Compute structure-mode loss:

$$\mathcal{L}^{(b)} = \mathcal{L}_{\text{struct}}^{\text{full}}(\hat{x}^{(b)}, x_{\text{gt}}^{(b)})$$

  end if
end for
$\mathcal{L}_{\text{total}} \leftarrow \frac{1}{B}\sum_{b=1}^{B}\mathcal{L}^{(b)}$
Update θ by backpropagation on $\mathcal{L}_{\text{total}}$

A.2 Effective SNR Interpretation of the Sequence Mask Schedule

In this appendix, we provide an alternative interpretation of the masking schedule in Eq. (7) through an effective signal-to-noise matching argument.

For coordinate diffusion, the noisy input is

$$\boldsymbol{x}_\sigma = \boldsymbol{x}_0 + \sigma\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}). \tag{9}$$

Under the EDM parameterization, the skip coefficient is

$$c_{\text{skip}}(\sigma) = \frac{\sigma_{\text{data}}^2}{\sigma^2 + \sigma_{\text{data}}^2}. \tag{10}$$

Correspondingly, the coordinate signal-to-noise ratio scales as

$$\text{SNR}_{\text{coord}} = \frac{\sigma_{\text{data}}^2}{\sigma^2}. \tag{11}$$

For sequence corruption, let $p(\sigma)$ be the masking probability. Then the visible fraction is $1 - p(\sigma)$ and the masked fraction is $p(\sigma)$. We define an effective sequence information ratio as

$$\mathrm{SNR}_{\mathrm{seq}} := \frac{1 - p(\sigma)}{p(\sigma)}. \tag{12}$$

Although this quantity is not a Gaussian SNR in the strict sense, it provides a simple measure of how much visible token information remains relative to the masked portion.

Matching this effective sequence ratio to the coordinate SNR gives

$$\frac{1 - p(\sigma)}{p(\sigma)} = \frac{\sigma_{\mathrm{data}}^2}{\sigma^2}. \tag{13}$$

Solving for $p(\sigma)$ yields

$$p(\sigma) = \frac{\sigma^2}{\sigma^2 + \sigma_{\mathrm{data}}^2}. \tag{14}$$

Using the definition of $c_{\mathrm{skip}}(\sigma)$, this can be equivalently written as

$$p(\sigma) = 1 - c_{\mathrm{skip}}(\sigma), \qquad 1 - p(\sigma) = c_{\mathrm{skip}}(\sigma). \tag{15}$$

Therefore, the masking schedule adopted in the main text can be understood in two equivalent ways: directly as a visible-fraction alignment rule using the EDM skip coefficient, or as an effective SNR matching rule between sequence and coordinate corruption.

A.3 Antibody-Centered Cropping

For antibody/TCR–antigen complexes, the total number of tokens often exceeds the model crop size (e.g., 512 tokens). A naive contiguous or spatial crop, as commonly used in general protein structure prediction, is not ideal for immune complexes: it may discard CDR residues or retain antigen regions that are far from the binding interface. We therefore use an *antibody-centered* cropping strategy tailored to immune complex training (Algorithm S2).

The strategy follows four principles. First, all antibody/TCR tokens are always retained, since they provide the generation target and its structural context. Second, antigen tokens are ranked by distance to the antibody center of mass, so the crop focuses on the interface region. Third, within each antigen chain, selected spans are completed when possible to improve structural continuity. Fourth, during training, the number of antigen tokens is uniformly sampled between a minimum $N_{\mathrm{min}}^{\mathrm{ag}}$ and the maximum allowed by the crop budget, which serves as antigen-context augmentation. At inference time, the crop budget is fixed.

Compared with the mixture of contiguous, spatial, and spatial-interface crops used in general structure prediction, this strategy is deterministic in which chains are preserved and stochastic only in how much antigen context is kept.

Algorithm S2 Antibody-Centered Cropping

Input: token array with Cα coordinates $\{\mathbf{r}_j\}_{j=1}^{N}$
antibody chain IDs $\mathcal{C}_{\text{ab}}$, antigen chain IDs $\mathcal{C}_{\text{ag}}$
max token budget T, minimum antigen tokens $N_{\min}^{\text{ag}}$ (default 100)
Output: selected token index set $\mathcal{S}$

▷ Step 1: Identify antibody and antigen tokens

$\mathcal{I}_{\text{ab}} \leftarrow \{j : \text{chain}(j) \in \mathcal{C}_{\text{ab}}\}$
$\mathcal{I}_{\text{ag}} \leftarrow \{j : \text{chain}(j) \in \mathcal{C}_{\text{ag}} \text{ and } j \text{ is resolved}\}$
$N_{\text{ab}} \leftarrow |\mathcal{I}_{\text{ab}}|$

▷ Step 2: Compute antibody center

$\mathcal{I}_{\text{ab}}^{\text{res}} \leftarrow \{j \in \mathcal{I}_{\text{ab}} : j \text{ is resolved}\}$
$\bar{\mathbf{r}} \leftarrow \frac{1}{|\mathcal{I}_{\text{ab}}^{\text{res}}|} \sum_{j \in \mathcal{I}_{\text{ab}}^{\text{res}}} \mathbf{r}_j$

▷ Step 3: Determine antigen budget

$N_{\text{ag}}^{\max} \leftarrow \min\big(T - N_{\text{ab}},\ |\mathcal{I}_{\text{ag}}|\big)$
if $T < 0$ **then** ▷ No crop limit
    $N_{\text{ag}} \leftarrow |\mathcal{I}_{\text{ag}}|$
else
    $N_{\text{ag}} \sim \mathcal{U}\big[\min(N_{\min}^{\text{ag}},\ |\mathcal{I}_{\text{ag}}|,\ N_{\text{ag}}^{\max}),\ N_{\text{ag}}^{\max}\big]$
end if

▷ Step 4: Rank antigen tokens by distance

for each $j \in \mathcal{I}_{\text{ag}}$ **do**
    $d_j \leftarrow \|\mathbf{r}_j - \bar{\mathbf{r}}\|$
end for
Sort $\mathcal{I}_{\text{ag}}$ by d_j in ascending order

▷ Step 5: Greedy selection with span completion

$\mathcal{S}_{\text{ag}} \leftarrow \emptyset$
for each j in sorted $\mathcal{I}_{\text{ag}}$ **do**
    if $|\mathcal{S}_{\text{ag}}| \geq N_{\text{ag}}$ **then**
        break
    end if
    if $j \in \mathcal{S}_{\text{ag}}$ **then**
        continue
    end if
    $\mathcal{S}_{\text{ag}} \leftarrow \mathcal{S}_{\text{ag}} \cup \{j\}$

▷ Complete the selected span within the same chain

    $c \leftarrow \text{chain}(j)$
    $P_c \leftarrow \{\text{pos_in_chain}(i) : i \in \mathcal{S}_{\text{ag}},\ \text{chain}(i) = c\}$
    if $|P_c| > 1$ **and** $\max(P_c) - \min(P_c) + 1 \leq \lfloor N_{\text{ag}}/2 \rfloor$ **then**
        for $p = \min(P_c)$ to $\max(P_c)$ **do**
            if $|\mathcal{S}_{\text{ag}}| < N_{\text{ag}}$ **then**
                $\mathcal{S}_{\text{ag}} \leftarrow \mathcal{S}_{\text{ag}} \cup \{\text{token_at}(c, p)\}$
            end if
        end for
    end if
end for

▷ Step 6: Combine and sort

$\mathcal{S} \leftarrow \text{Sort}(\mathcal{I}_{\text{ab}} \cup \mathcal{S}_{\text{ag}})$ ▷ By original token index
return $\mathcal{S}$

A.4 Hotspot Conditioning

At training time, epitope residues are computed on the fly from the bound complex and mapped to the model's `is_hotspot_residue` input feature. This supplies target-side epitope information and guides receptor generation toward the specified target region.

To avoid over-reliance on fully observed interface annotations, we randomly mask the epitope signal during training. At inference time, the user may provide either the full epitope or a partial subset. The overall procedure is summarized in Algorithm S3.

Algorithm S3 Epitope Identification and Hotspot Masking

Input: bound complex with Cα coordinates and immune-receptor/target partition
distance threshold $\tau = 6.0$ Å
geometric masking parameter $p = 0.2$, minimum keep count $n_{\min} = 1$
Output: token-level hotspot mask $\mathbf{h} \in \{0, 1\}^{N_{\text{token}}}$
▷ Part I: Identify epitope residues from the bound complex
Extract immune-receptor and target Cα coordinates, $\mathbf{R}_{\text{imm}}$ and $\mathbf{R}_{\text{tgt}}$
$D \leftarrow \text{PairwiseDistance}(\mathbf{R}_{\text{imm}}, \mathbf{R}_{\text{tgt}})$
$\text{epitope_CA}[i] \leftarrow \mathbf{1}[\min_j D_{j,i} < \tau] \quad \forall i$ on the target side
$\text{is_epitope} \leftarrow \mathbf{0}^{N_{\text{atom}}}$
for each target Cα index i with $\text{epitope_CA}[i] = 1$ **do**
    $\text{is_epitope}[\text{residue}(i)] \leftarrow 1$
end for
for each token k **do**
    $\text{token_epitope}[k] \leftarrow \text{is_epitope}[\text{centre_atom}(k)]$
end for
▷ Part II: Randomly mask hotspot residues
$\mathcal{H} \leftarrow \{k : \text{token_epitope}[k] = 1\}$
$K \leftarrow |\mathcal{H}|$
if $K = 0$ **then**
    return $\mathbf{0}^{N_{\text{token}}}$
end if
$n_{\text{keep}} \leftarrow \text{Geometric}(p) - 1$
$n_{\text{keep}} \leftarrow \text{clamp}(n_{\text{keep}}, n_{\min}, K)$
$\mathcal{H}_{\text{keep}} \leftarrow \text{RandomSample}(\mathcal{H}, n_{\text{keep}}, \text{replace} = \text{false})$
$\mathbf{h} \leftarrow \mathbf{0}^{N_{\text{token}}}$
$\mathbf{h}[\mathcal{H}_{\text{keep}}] \leftarrow 1$
return $\mathbf{h}$

B Training details and limitations

B.1 Training details

All IgGM2 models are trained on 16 A100 GPUs using the atom14 all-atom representation, a max crop size of 512 tokens. We optimize the model with Adam using a learning rate of 1.8×10^{-3}, $\beta_1 = 0.9$, $\beta_2 = 0.95$, and weight decay 1×10^{-8}. The learning rate follows an AF3-style schedule with 3,000 warmup steps, followed by a decay factor of 0.95 every 1,000 steps.

We train IgGM2-D with the mixed objective described in Section 3.5, using 30% CDR sequence-design samples and 70% structure-only denoising samples. For sequence-design samples, CDR amino-acid identities are replaced with UNK tokens. We additionally add 0.1 Å Gaussian noise to CDR backbone atoms during training.

B.2 Training–Test Leakage Control

FoldBench structures postdate the September 30, 2021 cutoff for IgGM2-P structure-prediction training. The January 1, 2024 design-training cutoff applies to a separate dataset and task. In addition to our 95% full-sequence identity filter, we remove every training sample whose CDR-H3 shares more than 80% identity with a test antibody. Table S1 reports the resulting distribution of maximum CDR-H3 identity between each test case and the training set. Roughly three quarters of maximum identities lie between 40% and 70% (77.3% of paired and 76.9% of nanobody cases), indicating that the evaluation is not dominated by close CDR-H3 homologues.

B.3 Runtime Comparison

Table S2 compares end-to-end runtime against AlphaFold-3 on the same FoldBench antibody–antigen test set, measured over MSA generation, feature preparation, and one inference sample. IgGM2-P requires approximately 14 s per complex versus approximately 30 min for AlphaFold-3, a speedup of

Table S1: Maximum CDR-H3 identity of each test antibody to the training set.

Maximum CDR-H3 identity	Paired	Nanobody
$[0.7, 0.8)$	17.0%	3.8%
$[0.6, 0.7)$	30.2%	9.6%
$[0.5, 0.6)$	24.5%	13.5%
$[0.4, 0.5)$	22.6%	53.8%
$[0.3, 0.4)$	5.7%	19.2%

Table S2: End-to-end FoldBench runtime.

Method	Total	Per complex
AlphaFold-3	$\sim$56.5 h	$\sim$30 min
IgGM2-P	**$\sim$0.44 h**	**$\sim$14 s**

roughly $130\times$. This end-to-end comparison includes MSA generation for AlphaFold-3; the measured ratio is protocol-specific and should not be interpreted as a hardware-normalized sampler speedup.

B.4 Limitations

IgGM2 provides a unified full-atom framework for immune receptor structure prediction and CDR sequence–structure co-design, but several limitations remain.

Sequence recovery. IgGM2-D still has room to improve amino acid recovery, especially for challenging CDR3 regions and nanobody designs. Although native sequence recovery is not the only goal of de novo design, higher AAR is still useful when preserving native-like binding motifs or functionally important residue patterns. The lower recovery on long HCDR3 loops suggests that stronger sequence priors may help better capture the diversity and constraints of immune receptor repertoires.

Staged full-atom refinement. IgGM2-D uses a staged inference procedure, where CDR sequence and backbone geometry are determined before sidechain refinement. This improves sampling stability and avoids separate inverse-folding or external sidechain-packing tools. However, sequence, backbone, and sidechain conformations are tightly coupled, so early errors may affect later sidechain packing. Future work could explore iterative refinement strategies that jointly update sequence, backbone, and sidechains.

Dependence on target information. The current design setting assumes that the target structure, and in some cases epitope-level information, is available. This is useful for structure-based design, but performance may depend on the accuracy of the target structure and interface annotations. Predicted antigens, uncertain epitopes, or flexible targets may introduce additional challenges in practical discovery settings.

Computational evaluation. Our evaluation mainly relies on computational metrics, including AAR, local Cα RMSD, DockQ, and Rosetta-based interface energy. These metrics capture sequence recovery, structural plausibility, and interface quality, but cannot fully determine binding affinity, specificity, stability, or developability.

C Supplementary Structure Prediction Results

The immune-receptor benchmarks include those studied with tFold-Ag and tFold-TCR (Wu et al., 2025a;b). General and monomer predictors include OpenFold-3, ESMFold, OmegaFold, EquiFold, and HelixFold-Single (Fang et al., 2023; Lee et al., 2022; Lin et al., 2023; The OpenFold3 Team, 2025; Wu et al., 2022). FoldBench results follow its official benchmark protocol.

Table S4: SAbDab-22H2-AbAg antibody–antigen complex prediction. Unconstrained methods are listed first, followed by methods using contact or epitope restraints.

Method	Restraint	DockQ ↑	SR (%) ↑	TM-Score ↑
AlphaFold-3	–	0.257	32.3	0.710
tFold-Ag	–	0.217	28.3	0.708
Protenix	–	0.218	29.3	–
Chai-1	–	0.178	23.2	–
AlphaFold-Multimer	–	0.158	18.2	0.665
Boltz-1	–	0.144	17.2	–
ColabFold	–	0.117	14.1	0.648
tFold-Ag	contact	0.703	97.0	0.918
tFold-Ag-ppi	epitope & paratope	0.416	62.6	0.814
tFold-Ag-ppi	epitope	0.303	46.5	0.761
ColabDock	contact	0.266	41.4	0.722
DyMEAN	epitope	0.160	28.3	0.700
IgGM2-P	epitope	**0.602**	**75.8**	**0.895**

Table S5: SAbDab-22H2-Ab paired-antibody H–L complex prediction. DockQ measures the heavy/light interface quality.

Method	DockQ ↑	LRMS (Å) ↓	iRMS (Å) ↓
AlphaFold-3	0.781	1.806	1.343
Chai-1	0.773	1.881	1.403
AlphaFold-Multimer	0.773	1.863	1.358
tFold-Ab	0.770	**1.678**	**1.284**
Uni-Fold MuSSe	0.756	1.989	1.459
EquiFold	0.751	2.041	1.467
ImmuneBuilder	0.749	1.923	1.451
DeepAb	0.722	2.079	1.564
IgFold	0.715	2.120	1.541
IgGM2-P	**0.782**	1.729	1.294

Table S6: SAbDab-22H2-Nano nanobody structure prediction. Per-CDR FR-aligned backbone RMSD (Å, lower is better) under Chothia numbering.

Method	Fr ↓	CDR-1 ↓	CDR-2 ↓	CDR-3 ↓
tFold-Ab	**0.67**	1.95	1.20	3.57
AlphaFold-3	0.68	**1.73**	1.14	3.57
Chai-1	0.73	1.80	**1.13**	3.57
AlphaFold	0.71	1.96	1.20	3.96
ESMFold	0.70	1.87	1.34	3.80
OmegaFold	0.73	1.77	1.34	3.63
IgFold	0.73	1.97	1.29	4.64
ImmuneBuilder	0.75	1.91	1.23	3.79
HelixFold-Single	0.76	1.92	1.27	4.16
DeepAb	0.93	2.36	1.51	9.03
IgGM2-P	0.75	2.09	1.32	**3.41**

Table S3: IgGM2-P per-subset DockQ / RMSD breakdown on FoldBench.

Subset	DockQ mean	DockQ median	SR	iRMSD / LRMSD (Å)
Ab–Ag	0.611	0.751	76.23% (93/122)	3.04 / 8.37
Nano–Ag	0.574	0.762	66.67% (32/48)	4.08 / 9.24
scFv–Ag	0.843	0.843	100% (2/2)	0.85 / 1.66
Overall	**0.604**	**0.755**	**73.84% (127/172)**	**3.31 / 8.53**

Bin	Range	Count (%)
High	DockQ ≥ 0.80	78 (45.3%)
Medium	$0.49 \leq$ DockQ < 0.80	37 (21.5%)
Acceptable	$0.23 \leq$ DockQ < 0.49	12 (7.0%)
Incorrect	DockQ < 0.23	45 (26.2%)

C.1 FoldBench Per-Subset Breakdown

We report the per-subset breakdown and DockQ-quality distribution behind the main FoldBench result in Table S3. The benchmark contains 172 interfaces from 113 PDB biological assemblies, including 122 Ab–Ag, 48 Nano–Ag, and 2 scFv–Ag interfaces. Target-fixed, epitope-conditioned IgGM2-P achieves a mean DockQ of 0.604 on the overall split, with 45.3% of interfaces falling into the high-quality bucket (DockQ ≥ 0.80) and 26.2% into the incorrect bucket (DockQ < 0.23). Unlike the main FoldBench comparison, this analysis provides a more detailed view of performance across receptor types and DockQ-quality levels.

C.2 SAbDab-22H2-AbAg Complex Prediction

Table S4 reports the paired antibody–antigen complex prediction results discussed in Section 4.2.

C.3 SAbDab-22H2-Ab

On standalone paired-antibody modeling (Table S5), IgGM2-P achieves the best H–L DockQ among all compared methods, reaching 0.782 and slightly outperforming AlphaFold-3 at 0.781. Across 169 samples, IgGM2-P obtains LRMS of 1.729 Å and iRMS of 1.294 Å, with a success rate of 100%. It also maintains strong backbone quality, with TM-score 0.951, GDT 0.948, and Cα-RMSD 1.222 Å. Its LRMS is second only to the specialized tFold-Ab model, which reaches 1.678 Å.

C.4 SAbDab-22H2-Nano

Table S6 reports per-CDR FR-aligned backbone RMSD on the 73-sample nanobody benchmark using Chothia numbering. IgGM2-P achieves the best HCDR3 RMSD, reaching 3.41 Å compared with 3.57 Å for both tFold-Ab and AlphaFold-3. It also obtains a framework RMSD of 0.75 Å and strong overall backbone quality, with TM-score 0.928, GDT 0.919, and Cα-RMSD 1.538 Å. Since the input is a single VHH domain without a light chain, interface DockQ is not defined.

Table S7: STCRDab-22 unliganded TCR structure prediction (24 samples; 23 $\alpha\beta$-TCR + 1 $\gamma\delta$-TCR). Overall α–β interface docking quality. TCR-specific baselines include tFold-TCR (Wu et al., 2025b) and TCRModel2 (Yin et al., 2023).

Method	DockQ ↑	LRMS (Å) ↓	iRMS (Å) ↓	GDT ↑
tFold-TCR	**0.797**	**1.638**	**1.070**	0.956
TCRModel2	0.794	1.795	1.120	0.948
AlphaFold-Multimer	0.785	1.816	1.186	0.931
AlphaFold-3	0.773	2.047	1.224	0.928
IgGM2-P	0.740	4.324	2.744	**0.961**

C.5 STCRDab-22-TCR (unliganded TCR)

Table S7 reports overall α–β interface quality on the 24-sample unliganded TCR benchmark. The benchmark mainly consists of $\alpha\beta$ TCRs, with one $\gamma\delta$ TCR included following the same chain-mapping convention used by prior baselines. IgGM2-P achieves the best GDT of 0.961, while trailing tFold-TCR (Wu et al., 2025b) and TCRModel2 (Yin et al., 2023) on DockQ and iRMSD. The interface metrics are affected by a small number of outlier cases, but IgGM2-P maintains strong overall backbone quality, with TM-score 0.958, DockQ 0.740, and SR 95.83%.

IgGM2-P is weaker than specialized methods on several unliganded metrics in Tables S5–S7, including unliganded TCR DockQ and iRMSD, where tFold-TCR and TCRModel2 remain ahead. Its performance is nonetheless not uniformly inferior: it attains the best paired-antibody H–L DockQ (0.782, Table S5), the lowest nanobody HCDR3 RMSD (3.41 Å, Table S6), and the best unliganded TCR GDT (0.961). IgGM2-P therefore remains competitive in the unconditioned setting, while its clearest advantages lie in the target-conditioned prediction and co-design settings it is designed for.

Table S8: Co-generated complex quality on the full paired-antibody clean test set. Baselines marked with * are re-trained on the leak-free training data. HDOCK is used for DiffAb and AbX where external docking is required.

Method	DockQ ↑	TM-Score ↑	CA-RMSD (Å) ↓	SR ↑
DiffAb*	0.022	0.027	3.241	0.010
AbX	0.025	0.863	1.921	0.005
dyMEAN*	0.087	0.864	1.425	0.082
IgGM	0.170	0.815	1.608	0.262
RFAntibody	0.065	0.537	11.707	0.021
IgGM2-D	**0.292**	**0.867**	1.575	**0.503**

Table S9: Co-generated complex quality on the full nanobody clean test set. Baselines marked with * are re-trained on the leak-free training data. HDOCK is used for DiffAb where external docking is required.

Method	DockQ ↑	TM-Score ↑	CA-RMSD (Å) ↓	SR ↑
DiffAb*	0.023	0.006	2.146	0.000
IgGM	0.137	0.771	2.244	0.162
RFAntibody	0.085	0.661	8.939	0.041
IgGM2-D	**0.216**	**0.808**	**2.001**	**0.284**

D Supplementary Sequence-Design Results

D.1 SAbDab test-set construction

The full clean SAbDab sequence-design test split (Dunbar et al., 2014) contains 269 complexes (195 paired antibodies and 74 nanobodies; `indices_design_sab_test_clean.csv`). The train–test cluster split controls similarity between training and test complexes but does not remove redundancy within the held-out test split. We therefore construct a subset with pairwise CDRH3 similarity below 0.80, retaining 203 complexes (148 paired antibodies and 55 nanobodies) for the main evaluation.

D.2 Co-Generated Complex Quality

Paired antibody complex quality. Table S8 reports bound-complex quality on the full paired-antibody clean test set by comparing the complexes generated from designed CDR sequences with the ground-truth antibody–antigen complexes. IgGM2-D achieves the best DockQ (0.292), TM-score (0.867), and SR (0.503) among all compared methods. Compared with IgGM, IgGM2-D improves DockQ from 0.170 to 0.292 and nearly doubles SR from 0.262 to 0.503, while maintaining comparable Cα RMSD. These results complement the AAR/RMSD and interface-energy analyses, showing that IgGM2-D better preserves the native antigen-bound geometry after CDR sequence design.

Nanobody complex quality. Table S9 reports complex-level quality on the full nanobody clean test set by comparing the co-generated nanobody–antigen complexes with the ground-truth complexes. IgGM2-D achieves the best performance across all reported metrics. Compared with IgGM, it improves DockQ from 0.137 to 0.216 and SR from 0.162 to 0.284, while reducing Cα RMSD to 2.001 Å and increasing TM-score to 0.808. These results suggest that the improvements in nanobody AAR and local CDR geometry also translate into better global antigen-bound complex geometry, even in the more challenging long-HCDR3 setting.

D.3 Orthogonal Refolding of Designed Sequences

As an orthogonal check on the designed sequences, we refolded them with AlphaFold-3 and compared the refolded models both against the IgGM2-D outputs and against the ground-truth complexes (Table S10). The refolded receptors closely reproduce the IgGM2-D receptor structures, with a mean Cα-RMSD of 1.52 Å, which supports receptor-fold consistency.

Table S10: Orthogonal AlphaFold-3 refolding of IgGM2-D designs.

Type	AF3 ipTM ↑	Ab Cα-RMSD (Å) ↓	DockQ vs. IgGM2-D ↑	DockQ vs. GT ↑
All	0.451	1.52	0.061	0.058
Paired	0.539	1.33	0.063	0.058
Nanobody	0.213	2.04	0.058	0.058

Table S11: Rosetta $dG_{\text{separated}}$ distribution for paired antibodies and nanobodies on the CDRH3-deduplicated SAbDab design test subset, after a shared FastRelax protocol. Baselines marked with * are re-trained on the leak-free training data.

Paired antibodies

Method	Mean ↓	Median ↓	Q25	Q75	Max	> 1000	Median $\Delta\Delta G$ ↓	IMP ↑
DiffAb*	29.86	-11.08	-18.57	3.16	3041.77	1	7.94	0.392
AbX	2.70	0.00	-10.37	0.00	375.43	**0**	10.15	0.318
dyMEAN*	2191.46	-3.55	-15.51	98.15	95289.59	23	19.19	0.327
IgGM	-17.83	-15.50	-26.93	-8.82	81.30	**0**	-2.44	0.541
RFAntibody	118.90	-8.28	-17.72	-2.60	5019.33	7	2.41	0.466
IgGM2-D	**-23.27**	**-21.60**	**-28.93**	**-13.13**	**52.95**	**0**	**-5.03**	**0.615**

Nanobodies

Method	Mean ↓	Median ↓	Q25	Q75	Max	> 1000	Median $\Delta\Delta G$ ↓	IMP ↑
DiffAb*	10.52	-12.15	-17.79	-1.75	435.56	**0**	18.09	0.236
IgGM	-9.46	-15.43	-21.47	-4.51	143.57	**0**	11.58	0.273
RFAntibody	1.25	-12.38	-18.35	-4.88	899.67	**0**	11.77	0.327
IgGM2-D	**-24.84**	**-24.73**	**-34.42**	**-13.81**	**0.00**	**0**	**-2.18**	**0.527**

Table S12: Effect of Rosetta FastRelax (FR) on interface IMP and mean $dG_{\text{separated}}$.

	With FR				No FR			
Method	Paired IMP ↑	Paired dG ↓	Nano IMP ↑	Nano dG ↓	Paired IMP ↑	Paired dG ↓	Nano IMP ↑	Nano dG ↓
DiffAb*	0.392	29.86	0.236	10.52	0.007	+1361	0.000	+1035
AbX	0.318	2.70	–	–	0.034	+551	–	–
dyMEAN*	0.327	2191.46	–	–	0.014	+17689	–	–
IgGM	0.541	−17.83	0.273	−9.46	0.027	+1141	0.000	+1145
RFAntibody	0.466	118.90	0.327	1.25	0.020	+4001	0.000	+5826
IgGM2-D	**0.615**	**−23.27**	**0.527**	**−24.84**	0.007	+5095	0.018	+4634

The binding poses, however, agree poorly (DockQ 0.061 against IgGM2-D). AlphaFold-3 also shows similarly low DockQ against the ground-truth complexes under the same protocol (0.058), so it does not reliably recover the native interfaces here. Low AlphaFold-3–IgGM2-D interface agreement therefore does not by itself invalidate the generated binding poses, and we treat this experiment as evidence of receptor-fold consistency only rather than as validation of interface quality.

D.4 Interface-Energy Distribution and the Effect of Relaxation

Table S11 summarizes the Rosetta $dG_{\text{separated}}$ distributions for paired antibodies and nanobodies. IgGM2-D achieves the lowest median $dG_{\text{separated}}$, the lowest median $\Delta\Delta G$, and the highest IMP in both settings, with no samples exceeding $dG_{\text{separated}} = 1000$. For paired antibodies, dyMEAN and RFAntibody have 23 and 7 samples above this threshold, respectively, substantially inflating their mean energies.

We also quantify how much the shared FastRelax stage contributes. Table S12 compares interface energies with and without FastRelax, reconstructing missing sidechains beforehand only for methods that require it. For IgGM2-D, FastRelax raises paired-antibody IMP from 0.007 to 0.615 and improves mean $dG_{\text{separated}}$ from $+5095$ to -23.27; for nanobodies, IMP rises from 0.018 to 0.527 and mean $dG_{\text{separated}}$ from $+4634$ to -24.84.

Without relaxation, every method exhibits near-zero IMP and strongly positive mean $dG_{\text{separated}}$, indicating that direct Rosetta evaluation of packed but unrelaxed structures is dominated by local

Table S13: Paired-antibody CDR co-design performance on the full SAbDab design test set before internal test-set deduplication. Higher AAR and lower local Cα RMSD indicate better performance. Best values are bolded and second-best values are underlined. HDOCK is used for DiffAb and AbX when external docking is required.

Method	CDR AAR ↑							Local Cα RMSD (Å) ↓					
	H1	H2	H3	L1	L2	L3	Avg	H1	H2	H3	L1	L2	L3
DiffAb*	0.667	0.379	0.247	0.621	0.593	0.446	0.477	1.140	1.885	4.508	1.450	0.902	2.332
AbX	**0.767**	0.670	**0.403**	**0.754**	0.568	**0.634**	**0.604**	0.756	0.597	3.214	0.725	0.317	1.310
dyMEAN*	0.724	**0.684**	0.387	0.717	**0.763**	0.604	0.597	0.700	0.618	**2.150**	0.697	0.207	0.803
IgGM	0.625	0.397	0.241	0.655	0.720	0.591	0.480	0.689	0.710	2.618	0.675	0.125	0.791
RFAntibody	0.210	0.084	0.103	0.061	0.222	0.135	0.124	1.900	2.246	3.510	2.716	0.307	1.161
IgGM2-D	0.722	0.643	0.384	0.708	0.740	0.605	0.587	**0.603**	**0.438**	2.288	**0.531**	**0.078**	**0.755**

Table S14: Nanobody CDR co-design performance on the full SAbDab design test set before internal test-set deduplication. AAR is higher better; local Cα RMSD is lower better.

Method	CDR AAR ↑				Local Cα RMSD (Å) ↓		
	H1	H2	H3	Avg	H1	H2	H3
DiffAb*	**0.496**	0.339	0.200	0.298	1.668	0.910	2.919
IgGM	0.442	0.284	0.171	0.267	1.695	1.160	2.814
RFAntibody	0.176	0.103	0.103	0.124	2.382	2.044	3.534
IgGM2-D	0.479	**0.515**	**0.246**	**0.379**	**1.545**	**0.796**	**2.709**

Table S15: Rosetta interface-energy metrics on the full SAbDab design test set before internal test-set deduplication.

Method	Paired antibodies		Nanobodies	
	IMP ↑	Mean dG ↓	IMP ↑	Mean dG ↓
DiffAb*	0.369	26.11	0.203	12.99
AbX	0.292	1.37	–	–
dyMEAN*	0.314	2630.74	–	–
IgGM	0.533	-19.31	0.311	4.86
RFAntibody	0.441	94.72	0.311	11.74
IgGM2-D	**0.595**	**-23.31**	**0.554**	**-24.11**

geometric inconsistencies and atomic clashes. Relaxation is thus necessary for a consistent comparison under the Rosetta energy function. This is consistent with the side-chain-generation limitation discussed in Appendix B.4: directly generated or packed structures are not yet suitable for Rosetta scoring without local refinement. Under the identical protocol, IgGM2-D still achieves the best post-relaxation IMP and mean $dG_{\text{separated}}$ for both receptor classes, suggesting that its generated interfaces more often sit in geometries that can be refined into favorable local conformations.

D.5 Full-Set Sequence-Design Results

We additionally report CDR co-design results on the full SAbDab design test set before internal test-set deduplication (269 complexes; Section D.1). Tables S13, S14, and S15 mirror the main sequence-design tables and report AAR, framework-aligned local Cα RMSD, and Rosetta interface-energy metrics on the full set. The overall trends are consistent with the deduplicated main results, showing that IgGM2-D maintains strong local structural quality and favorable interface-energy metrics even when all test-set sequences are included.

D.6 TCR Sequence–Structure Co-Design

VDJdb-2025-TCR50 construction. We construct an all-six-CDR test set from paired human, MHC-I-restricted TCR$\alpha\beta$ records in the 2025 VDJdb release. After filtering epitope and CDR3 overlap against the STCRDab design training and validation sets (Leem et al., 2018) (CDR3 identity < 0.80 for both chains), we sample one TCR per remaining epitope, yielding 60 candidates. We keep records in which both chains have complete V/J, CDR3, epitope, and MHC annotations. TCR sequences are reconstructed from VDJdb annotations with Thimble, IMGT variable domains are extracted with ANARCI, and the corresponding TCR–pMHC complexes are modeled with AlphaFold-3 (Abramson et al., 2024) as five-chain inputs ($B/A/M/N/P$ for β/α/MHC/β_2M/peptide). We retain 50 complexes with overall ipTM ≥ 0.8 and use the predicted structures as references; the retained models are converted to PDB and bioassembly files, indexed, and annotated with the six IMGT CDRs. Temporal splitting of STCRDab leaves five leak-free TCR–pMHC targets.

Table S16 compares amino-acid recovery on the five leak-free STCRDab TCR–pMHC targets using the same CDR masks across methods. The methods differ in their structural inputs: inverse-folding

Table S16: Mean all-six-CDR amino-acid recovery on five leak-free STCRDab TCR–pMHC targets.

Method	Designs/target	All CDR	H1	H2	H3	L1	L2	L3
ProteinMPNN	10	0.433	0.518	0.350	0.446	0.390	0.400	0.450
Rosetta FastDesign	10	0.397	0.529	0.264	0.391	0.328	0.345	**0.494**
ESM-IF1	10	0.349	0.143	0.306	0.373	0.443	0.326	0.394
IgGM2-D	1	**0.626**	**0.840**	**0.833**	**0.466**	**0.633**	**0.807**	0.455

Table S17: Effect of sequence-masking schedule on antibody design.

Sequence-masking schedule	All-CDR AAR ↑	Mean DockQ ↑	SR (DockQ ≥ 0.23) ↑	iRMSD (Å) ↓	LRMSD (Å) ↓	IMP ↑
Noise-aligned masking (ours)	0.531	**0.285**	**0.443**	**6.34**	**16.74**	**0.591**
MFDesign-D-style linear masking	**0.533**	0.232	0.355	8.85	22.03	0.502

Table S18: Sampling ablation on the full 269-complex design test set at a fixed $30/70$ sequence/structure training ratio. The backbone-noise scale and phase-transition step vary; the selected setting is bolded.

Noise	T_{switch}	All AAR ↑	Paired AAR ↑	Nano AAR ↑	HCDR3 RMSD ↓	IMP ↑
0.1	25	0.523	0.579	0.373	3.846	0.595
0.1	50	0.528	0.589	0.369	3.830	0.561
0.1	**75**	**0.530**	**0.587**	**0.379**	**3.702**	**0.584**
0.1	200	0.522	0.582	0.362	3.618	0.543
0.5	25	0.522	0.580	0.371	3.993	0.561
0.5	50	0.528	0.584	0.380	4.052	0.602
0.5	75	0.527	0.583	0.382	4.051	0.550
0.5	200	0.525	0.583	0.373	3.897	0.580

baselines receive reference receptor structures, whereas IgGM2-D generates receptor coordinates from sequence and fixed target context.

E Supplementary Ablations

E.1 Sequence-Masking Schedule

IgGM2 aligns sequence masking with the continuous structural noise level σ through Eq. (7), rather than tying it to a discrete sequence timestep as in the MFDesign-style linear schedule (Yang et al., 2026). To isolate this choice, we trained a baseline that forces the continuous structural noise to align with a discrete sequence timestep. As shown in Table S17, the two schedules yield nearly identical sequence recovery (0.531 versus 0.533), but linear masking substantially degrades structural and interface quality: DockQ drops from 0.285 to 0.232, success rate from 0.443 to 0.355, and IMP from 0.591 to 0.502, while iRMSD and LRMSD both worsen. This supports aligning sequence masking with the native continuous diffusion process rather than imposing a discrete sequence-noise schedule on structural diffusion.

E.2 Backbone Noise and Sampling-Stage Ablation

Table S18 reports sampling-stage and backbone-noise ablations on the full 269-complex design test set before test-internal deduplication. The sequence/structure training ratio is fixed at $30/70$ throughout. At noise scale 0.1, T_{switch}=75 gives the highest All AAR (0.530). At noise scale 0.5, no transition step exceeds 0.528 All AAR, although step 50 reaches a higher IMP (0.602). We use noise scale 0.1 and T_{switch}=75 for the reported configuration.

F VISUALIZATION

F.1 TCR–PMHC COMPLEXES

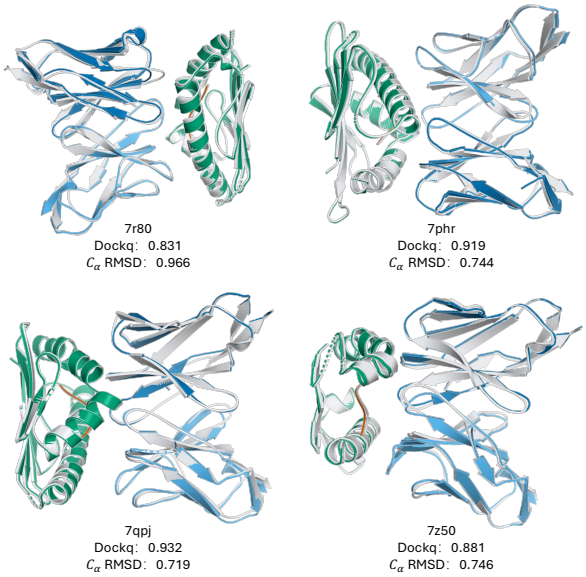


Figure S1: Representative visualization for STCRDab-22-TCR_pMHC complex prediction.

F.2 ANTIBODY

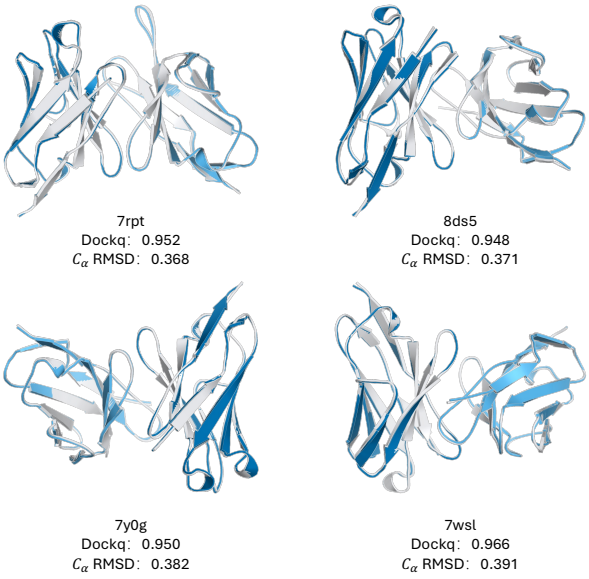


Figure S2: Representative visualization for SAbDab-22H2-Ab antibody prediction.

F.3 NANOBODY

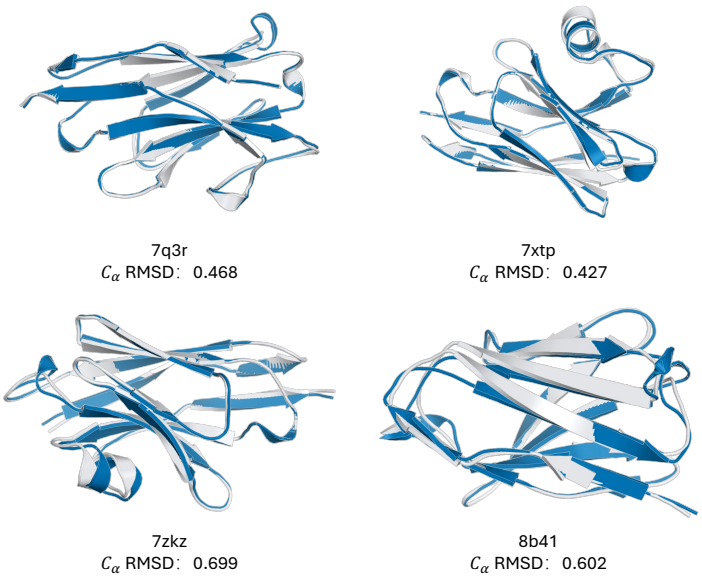


Figure S3: Representative visualization for SAbDab-22H2-Nano nanobody prediction.

F.4 UNLIGANDED TCR

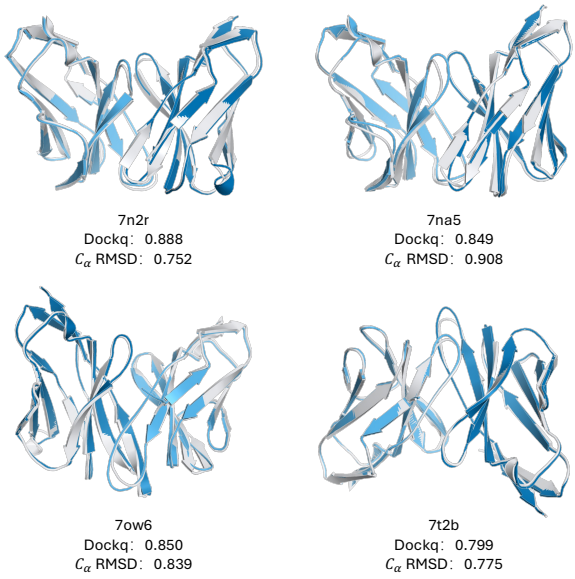


Figure S4: Representative visualization for STCRDab-22 unliganded TCR prediction.