TRUSTCDR: RELIABILITY-GUIDED CDR CO-DESIGN UNDER LOCAL ANTIGEN UNCERTAINTY

Anonymous authors
Paper under double-blind review

ABSTRACT

Target-specific antibody CDR sequence–structure design uses antigen geometry to guide binding-interface generation. Recent antigen-aware generators encode this geometry but generally do not estimate its local reliability and may therefore assign comparable trust to geometrically heterogeneous cues. Antigen structures used for design may be experimentally reconstructed, predicted, docked, or preprocessed and may also contain conformationally variable regions. Such structures can be globally plausible while their local geometric cues differ in reliability. Consequently, lower-reliability cues may exert excessive influence on CDR–antigen relation updates and degrade interface quality. We introduce TrustCDR, which estimates a reference-free, perturbation-based relative geometric reliability signal and converts it into relative trust weights that modulate residual CDR–antigen pair updates. Starting from a single observed antigen, TrustCDR constructs a reference-free perturbation ensemble and evaluates residue-level stability across coordinate, topology, orientation, and surface cues. It preserves the base antigen condition and uses down-only raw weights with mean-preserving normalization to reduce the relative influence of lower-reliability regions. On the RAbD test set, TrustCDR achieves 68.52% AAR, 0.57Å RMSD, and 0.76 DockQ under clean inputs, while maintaining 0.75 DockQ under a 1.5Å epitope perturbation. It also achieves IMP rates of 18.50% and 20.40% under clean and epitope-perturbed inputs, respectively.

1 INTRODUCTION

Antibody recognition is governed by highly variable complementarity-determining regions (CDRs). Their amino-acid identities define possible chemical interactions, while their conformations arrange these interactions at the binding interface. This sequence–structure coupling is well established by canonical-conformation and structural-clustering studies Al-Lazikani et al. (1997); North et al. (2011). Successful computational design must therefore generate plausible CDR sequence–structure pairs that adapt to a target antigen.

Framework-conditioned reconstruction can generate plausible CDRs, but target-specific design additionally requires the generated interface to respond to the supplied antigen. Antigen-aware generators therefore encode an epitope or an antibody–antigen complex, although nominal conditioning can still show limited sensitivity to the supplied antigen Olsen et al. (2024); Zhang et al. (2025); Kong et al. (2023b); Luo et al. (2022); Kong et al. (2023a); Wu et al. (2025b); Ahmed & Patterson (2026).

Design antigens may come from experiments, prediction, docking, or preprocessing and may also exhibit conformational heterogeneity Dunbar et al. (2014); Raybould et al. (2019); Jumper et al. (2021); Evans et al. (2022); Lin et al. (2023); Abramson et al. (2024). Although these sources differ, they can share a practical consequence: local coordinates, neighborhoods, orientations, and surface exposure may vary in stability even when the global fold remains plausible Frauenfelder et al. (1991); James & Tawfik (2003); Bowman & Geissler (2014); Feidakis et al. (2024). Existing antigen-aware generators generally do not estimate this residue-level stability explicitly before using antigen geometry in CDR–antigen relation updates. Their learned relations may capture binding relevance, but they do not ensure that less stable geometry exerts less influence. Paired native and perturbed antigen structures are rarely available for supervised training. TrustCDR therefore

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.

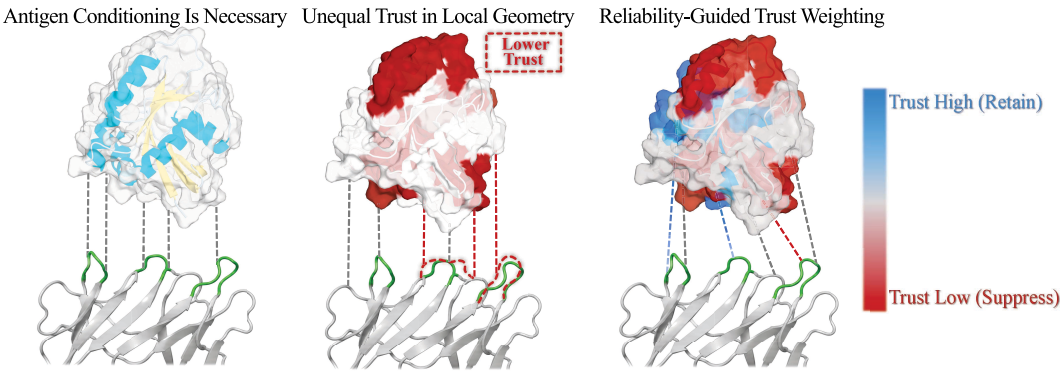

Figure 1: Motivation of TrustCDR: antigen conditioning is necessary for target-specific design, but local geometric cues should not be trusted uniformly.

estimates relative local stability from the supplied antigen itself, without requiring such paired data. Figure 1 summarizes this distinction and motivates reliability-guided relative trust weighting. We introduce **TrustCDR**, which estimates residue-level geometric reliability as relative local stability and converts it into relative trust weights for residual pair updates. Here, reliability is an estimated property of the supplied antigen geometry, whereas relative trust denotes the derived multiplicative weight applied to a residual pair update.

Starting from a single observed antigen structure, TrustCDR constructs a reference-free antigen-only perturbation ensemble and evaluates coordinate, topology, orientation, and surface stability. The resulting multi-channel descriptor guides a residual pair correction and its pair-specific applicability gate, while a scalar reliability score produces relative trust weights. A down-only rule attenuates updates supported by lower-reliability regions. Mean-weight-preserving normalization avoids globally weakening the residual path and leaves the base antigen-conditioned representation unchanged. Our contributions are summarized as follows:

- We identify the lack of explicit reliability-guided weighting for locally heterogeneous antigen geometry as an underexplored problem in antigen-aware CDR generation.
- We propose TrustCDR, which performs reference-free residue-level reliability estimation and reliability-guided relative trust weighting of residual CDR–antigen pair updates.
- We evaluate TrustCDR under clean and locally perturbed antigen inputs, together with affinity-oriented screening, mechanism ablations, and antigen-dependence diagnostics.

2 Related Work

Antigen-conditioned antibody CDR co-design. Computational antibody design spans energy-based search and learned sequence–structure generation. RosettaAntibodyDesign optimizes antibody sequences and conformations through statistical energies and Monte Carlo sampling Adolf-Bryfogle et al. (2018); Leman et al. (2020). Learned sequence priors, optimization, and retrieval systems further improve structure-conditioned candidate search Khan et al. (2023); Zhang et al. (2025). MEAN and dyMEAN use equivariant complex modeling for antigen-conditioned design, including end-to-end full-atom generation Kong et al. (2023b;a). DiffAb and AbX jointly model CDR identities and geometry through diffusion, with AbX incorporating evolutionary and physical constraints Luo et al. (2022); Zhu et al. (2024). FlowDesign introduces flow matching, while dyAb and RAAD address pre-binding flexibility and dynamic antibody–antigen relations Wu et al. (2025b); Tan et al. (2025); Wu et al. (2025c). Other approaches explore Poisson flows, score-based generation, sequence–structure codes, and full-atom modeling Huang et al. (2023); Xie et al. (2024); Zhu et al. (2025); Wu et al. (2025a); Yang et al. (2025). Concurrent anonymous work under review studies energy-tilted interface potentials for score-based antibody design, learning a label-free paratope–epitope potential from native and decoy complexes for training-time energy–geometry calibration Anonymous (2027). In contrast, TrustCDR estimates residue-level relative geometric stability from an antigen-only perturbation ensemble and uses the resulting reliability-derived rela-

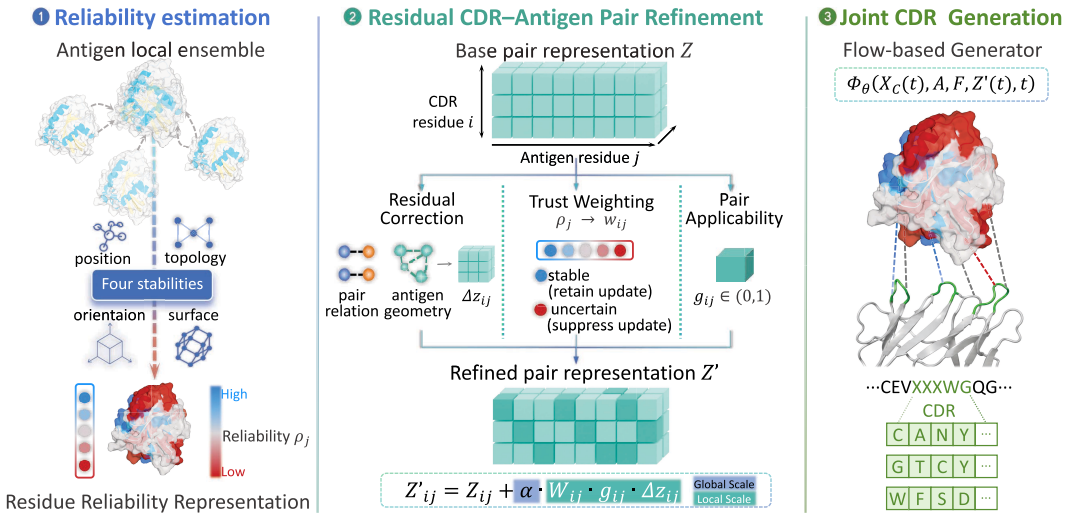

Figure 2: Framework of TrustCDR. A local antigen ensemble estimates residue-level reliability ρ_j from position, topology, orientation, and surface stability. The resulting trust weight w_{ij} modulates the residual correction together with the pair-applicability gate, while the base pair representation is retained for joint CDR generation.

tive trust weights to modulate residual CDR–antigen pair updates, without learning a native–decoy interface energy.

Antigen structural uncertainty and reliability-aware conditioning. Local reliability varies across protein structures, and variation in predicted geometry can affect docking and other structure-based tasks Jumper et al. (2021); Alkhouri et al. (2023); Scardino et al. (2023). Proteins also exhibit conformational heterogeneity and ligand-induced changes Frauenfelder et al. (1991); James & Tawfik (2003); Bowman & Geissler (2014); Feidakis et al. (2024). For antibodies, bound and unbound comparisons reveal loop-dependent CDR changes that can affect docking, while antigen recognition can involve conformational diversity Liu et al. (2024); Dreyer et al. (2026); Barozet et al. (2018); Lee (2021). These effects matter because coordinate, local-frame, and neighborhood changes directly alter CDR–antigen pair features and can bias relation updates. Additional conformations can help when reliable alternate states are available, but many design settings provide only one observed antigen structure. TrustCDR therefore estimates local reliability from that structure without requiring known alternate conformations or treating synthetic perturbations as biological ground truth.

3 Method

Overview. TrustCDR designs masked CDR residues conditioned on an antigen and the known antibody context. It preserves the base CDR–antigen pair representation and reweights only a residual branch. The method has three stages. First, a reference-free antigen-only perturbation ensemble estimates residue-level reliability from position, topology, orientation, and surface stability. Second, the resulting multi-channel reliability representation guides the residual correction $\Delta \mathbf{z}_{ij}$ and pair-applicability gate g_{ij}, while the scalar reliability score ρ_j is separately mapped to a relative trust weight w_{ij} through raw down-only weighting followed by mean-preserving normalization. Third, the refined pair representation conditions a joint generator for CDR amino-acid states, positions, and orientations. Figure 2 summarizes the overall framework.

3.1 Problem Formulation

Let $A = \{a_j\}_{j=1}^{L_A}$ denote the antigen. Let $F = \{f_i\}_{i=1}^{L_F}$ denote the known antibody context. A binary mask M_C identifies the CDR positions to be generated. The native sequence and structure of these masked positions are not used as inputs. This setting matches practical CDR redesign. The framework residues and antigen are observed. The CDR residues are treated as unknown variables.

For each masked CDR position $i \in \mathcal{I}_C$, the generated state is

$$x_i^C = (y_i, \mathbf{p}_i, R_i), \tag{1}$$

Here y_i is the amino-acid state. The vector $\mathbf{p}_i \in \mathbb{R}^3$ is the residue position. The matrix $R_i \in \mathrm{SO}(3)$ is the local orientation. The complete generated CDR state is $X_C = \{x_i^C\}_{i \in \mathcal{I}_C}$. The conditional design objective is

$$p_\theta(X_C \mid A, F, M_C). \tag{2}$$

During generation, pair relations are built from the current generated CDR state $X_C(t)$. They are not built from native CDR coordinates. This prevents leakage from the target CDR structure. It also makes the pair representation time dependent. We denote the resulting pair tensor by $Z(t)$ and omit t below when the dependence is clear. As the CDR state moves along the generative path, the CDR–antigen relation must be updated from the current state rather than from the final native answer.

3.2 Reference-Free Antigen Reliability Estimation

Given one observed antigen, TrustCDR estimates residue-wise relative geometric stability without a native reference, oracle epitope labels, or paired conformations. Reliability is derived from the stability of position, neighborhood topology, local orientation, and surface exposure under smooth local perturbations.

Starting from the observed antigen $A^{(0)}$, we construct an antigen-only ensemble

$$\mathcal{E}_A = \{A^{(k)}\}_{k=1}^K, \qquad K = 8. \tag{3}$$

Each $A^{(k)}$ preserves the antigen sequence and applies smoothed, locally correlated Cα offsets. The ensemble is used only to estimate reliability and is not supplied to the generator as multiple antigen conditions. For each supplied antigen structure, the resulting reliability field is computed once during preprocessing and cached for all generated candidates.

For antigen residue j, let $\mathbf{x}_j^{(k)}$ be its backbone position in conformer $A^{(k)}$. Let $N_j^{(k)}$ be its local antigen-residue neighborhood. Let $O_j^{(k)} \in \mathrm{SO}(3)$ be its local backbone frame. Let $\eta_j^{(k)}$ be its surface-exposure descriptor.

Definition 1 (Local perturbation response). TrustCDR defines four response scores:

$$\begin{aligned}
q_j^{\mathrm{coord}} &= \frac{1}{K}\sum_{k=1}^{K}\left\|\mathbf{x}_j^{(k)} - \mathbf{x}_j^{(0)}\right\|_2, \\
q_j^{\mathrm{topo}} &= \frac{1}{K}\sum_{k=1}^{K}\left(1 - \frac{|N_j^{(0)} \cap N_j^{(k)}|}{|N_j^{(0)} \cup N_j^{(k)}|}\right), \\
q_j^{\mathrm{orient}} &= \frac{1}{K}\sum_{k=1}^{K} d_{\mathrm{SO}(3)}\left(O_j^{(0)}, O_j^{(k)}\right), \\
q_j^{\mathrm{surf}} &= \mathrm{Std}_k\left[\eta_j^{(k)}\right].
\end{aligned} \tag{4}$$

The geodesic distance on $\mathrm{SO}(3)$ is denoted by $d_{\mathrm{SO}(3)}$. The four descriptors match geometric primitives used in CDR–antigen pair features. Coordinate deviation measures positional sensitivity, neighborhood Jaccard distance tests local connectivity, frame rotation measures directional stability, and exposure variation tests the local surface environment. Local-distance and structure-quality methods motivate positional and neighborhood stability Mariani et al. (2013); Hiranuma et al. (2021); orientation-aware protein networks motivate frame stability Li et al. (2022); and Cα-neighbor exposure measures support the surface descriptor Hamelryck (2005). All four scores are invariant to a common rigid transformation, so they characterize local structural response rather than the arbitrary global coordinate frame. Together they capture complementary geometric aspects without reducing reliability to a single distance fluctuation.

Definition 2 (Residue-level geometric reliability). Let $\mathcal{M} = \{\mathrm{coord}, \mathrm{topo}, \mathrm{orient}, \mathrm{surf}\}$. Because the descriptors have different units and scales, each channel is robustly standardized over valid residues of the same antigen and mapped to $[0, 1]$:

$$u_j^m = \mathcal{N}_A^m(q_j^m), \qquad u_j^m \in [0, 1]. \tag{5}$$

A larger u_j^m means lower stability along axis m. This produces a relative within-antigen perturbation score rather than an absolute confidence probability; precise normalization is given in the supplementary material. We also compute a monotonic reliability response:

$$r_j^m = \exp(-\lambda u_j^m). \tag{6}$$

We use the fixed response scale $\lambda = 1$. The multi-channel reliability descriptor is

$$U_{A,j} = \underset{m \in \mathcal{M}}{\mathrm{Concat}} \left[u_j^m, r_j^m\right]. \tag{7}$$

Because $|\mathcal{M}| = 4$, $U_{A,j}$ is an eight-dimensional descriptor that retains the source and response of each geometric channel. To derive relative trust weights, we summarize it into a scalar geometric reliability score:

$$\rho_j = \frac{1}{2|\mathcal{M}|} \sum_{m \in \mathcal{M}} \left[(1 - u_j^m) + r_j^m\right]. \tag{8}$$

The scalar score decreases as any normalized uncertainty increases, so larger ρ_j indicates more stable local antigen geometry. The descriptor $U_{A,j}$ supplies channel-specific responses to the residual network, whereas ρ_j provides the scalar used to derive relative trust weights. The analytical bounds and monotonicity proof are given in the supplementary material.

3.3 Residual CDR–Antigen Pair Refinement

The reliability estimates above describe which parts of the antigen provide stable geometric evidence, but they do not specify how a CDR–antigen relation should change. That change must also reflect the current generated CDR state and should not overwrite the original pair representation. TrustCDR therefore retains the base pair tensor as a skip path and learns a reliability-aware residual refinement. The refinement contains a vector proposal that specifies the feature update and a scalar gate that models its pair-specific applicability. At flow time t, both components are recomputed from $X_C(t)$ and $Z(t)$. Below, $\mathbf{z}_{ij}$ denotes the pair representation at the current flow step, while U_A and ρ remain fixed for the supplied antigen.

For CDR position i and antigen residue j, the base pair representation is

$$\mathbf{z}_{ij} = \phi_{\mathrm{pair}}\left(\mathbf{h}_i, \mathbf{h}_j, \mathbf{r}_{ij}^{\mathrm{seq}}, \mathbf{d}_{ij}, \boldsymbol{\omega}_{ij}\right), \tag{9}$$

where $\mathbf{h}_i$ and $\mathbf{h}_j$ are residue embeddings, $\mathbf{r}_{ij}^{\mathrm{seq}}$ encodes relative sequence and chain information, and $\mathbf{d}_{ij}$ and $\boldsymbol{\omega}_{ij}$ encode distance and orientation features.

Directly scaling $\mathbf{z}_{ij}$ can disturb the learned pair-feature distribution. It also does not define how the relation should be revised. TrustCDR therefore predicts a feature-space residual correction:

$$\Delta\mathbf{z}_{ij} = \Phi_\delta\left(\mathbf{z}_{ij}, \phi_{\mathrm{rel}}(U_{A,j}), d_{ij}, \mathbf{o}_{ij}\right), \tag{10}$$

where d_{ij} and $\mathbf{o}_{ij}$ are current distance and relative-direction features. The tensor $\Delta Z = \{\Delta\mathbf{z}_{ij}\}$ has the same feature dimension as the CDR–antigen slice of Z. It is a learned correction to the pair representation, not a displacement of atomic coordinates. Because Φ_δ receives the base pair feature, $U_{A,j}$, and the current pair geometry, the correction adapts to both the local reliability profile and the current CDR–antigen relation. The vector head proposes a pair-specific correction for every valid pair, but the residual branch need not be equally applicable to all pairs. We therefore predict a continuous applicability gate from the same reliability-aware geometric context:

$$g_{ij} = \sigma[\Phi_g(\phi_{\mathrm{rel}}(U_{A,j}), d_{ij}, \mathbf{o}_{ij})]. \tag{11}$$

Here $g_{ij} \in (0, 1)$ is a scalar soft gate learned from reliability-aware geometric context.

Thus, g_{ij} is predicted alongside, rather than from, $\Delta\mathbf{z}_{ij}$. g_{ij} controls the learned applicability of the proposal; it is not an explicit reliability score. Neither operation rescales the base representation, which remains available through the skip path.

3.4 Reliability-Guided Relative Trust Weighting

The residual proposal and applicability gate are learned from data, but neither provides an explicit monotonic weighting factor derived from antigen reliability. In particular, g_{ij} models whether a correction fits the current pair; it does not guarantee that lower-reliability antigen geometry receives a smaller multiplicative weight. We therefore convert ρ_j into an explicit relative trust weight. Lower-reliability geometry should attenuate a residual update, whereas high reliability alone does not justify raw amplification. TrustCDR implements this conservative behavior with a down-only raw-weighting rule.

Within each supplied antigen, let μ_A and σ_A be the mean and standard deviation of ρ_j, and let $\mathcal{J}_i$ be the valid antigen partners of CDR position i. We compute

$$\begin{aligned} \widehat{\rho}_j &= \tanh\left(\frac{\rho_j - \mu_A}{\sigma_A + \varepsilon}\right), \qquad w_j^{\mathrm{raw}} = 1 - \beta \max(0, -\widehat{\rho}_j), \\ w_j^{\mathrm{clip}} &= \mathrm{clip}(w_j^{\mathrm{raw}}, 1-c, 1), \qquad \bar{w}_i = \frac{1}{|\mathcal{J}_i|} \sum_{j \in \mathcal{J}_i} w_j^{\mathrm{clip}}, \qquad w_{ij} = \frac{w_j^{\mathrm{clip}}}{\bar{w}_i + \varepsilon}. \end{aligned} \tag{12}$$

Here ε is a numerical stabilizer, while $\beta = 0.10$ and $c = 0.20$ are selected on the validation set. The max operation makes the weighting down-only; clipping limits attenuation, and normalization makes the mean relative weight around each CDR position approximately one.

For each valid pair $j \in \mathcal{J}_i$, the final refined representation is

$$\mathbf{z}'_{ij} = \mathbf{z}_{ij} + \alpha \left(g_{ij}\, w_{ij}\right) \Delta \mathbf{z}_{ij}, \tag{13}$$

where $\alpha \in \mathbb{R}$ is one global learnable residual scale shared by all CDR–antigen pairs. It contains no pair-specific reliability information. Initialized at zero, α gives $Z'(t) = Z(t)$ at the start of training, after which joint optimization can activate the residual branch. $\Delta \mathbf{z}_{ij}$ determines the feature correction, g_{ij} controls pair applicability, w_{ij} assigns reliability-derived relative trust, and α globally scales the residual branch. The base representation remains accessible through the skip path.

3.5 Joint Sequence–Structure Generation

The refined pair tensor $Z'(t) = \{\mathbf{z}'_{ij}(t)\}$ conditions joint CDR generation. We independently implement a flow-matching generator for joint CDR amino-acid states, positions, and orientations Lipman et al. (2023); Albergo & Vanden-Eijnden (2023). At flow time t, the velocity network predicts

$$\left(\mathbf{v}_\theta^{\mathrm{seq}}, \mathbf{v}_\theta^{\mathrm{pos}}, \mathbf{v}_\theta^{\mathrm{rot}}\right) = \Phi_\theta\left(X_C(t), A, F, Z'(t), t\right). \tag{14}$$

The generator is trained with equal-weight mean-squared flow-matching losses for amino-acid state, position, and orientation velocities. The full loss definition is provided in the supplementary material. The residual correction and the conditional generator are optimized jointly. For each supplied antigen, U_A is computed once and remains fixed during the flow trajectory. Sequence, position, and orientation share $Z'(t)$; after each flow update, $Z(t)$ and $Z'(t)$ are rebuilt from the current CDR state. Thus, the reliability field is fixed for a supplied antigen, while the CDR–antigen pair representation remains state dependent throughout generation.

4 Experiments

4.1 Experimental Setup

Dataset. All methods are trained and validated on the same 5,247 and 606 leakage-controlled SAbDab complexes, respectively, and are evaluated on the 58-complex RAbD test set Dunbar et al. (2014); Raybould et al. (2019); Adolf-Bryfogle et al. (2018). Potential train–test overlap is screened on both the antibody and antigen sides. Data curation, overlap screening, candidate aggregation, and preprocessing details are provided in the supplementary material.

Table 1: Overall performance under clean, moderate, and stress-test inputs (AAR: %; RMSD: Å). The final row reports TrustCDR's percentage improvement over the mean of the four baselines.

Method	Clean		Epitope 1.5Å		Epitope 3.0Å		Distant 1.5Å		Distant 3.0Å	
	AAR	RMSD	AAR	RMSD	AAR	RMSD	AAR	RMSD	AAR	RMSD
DiffAb (2022)	45.3	0.79	45.3	0.79	45.3	0.79	45.2	0.79	45.3	0.79
dyMEAN (2023)	<u>65.5</u>	<u>0.68</u>	<u>65.2</u>	**0.68**	<u>65.8</u>	**0.67**	<u>65.5</u>	**0.68**	**65.8**	**0.66**
dyAb (2025)	27.1	1.15	27.1	1.15	27.1	1.16	27.1	1.16	27.1	1.16
FlowDesign (2025)	48.8	2.17	48.8	2.27	49.0	2.26	48.7	2.28	49.2	2.27
TrustCDR	**68.5**	**0.57**	**68.8**	<u>0.73</u>	**66.1**	<u>0.73</u>	**68.5**	<u>0.73</u>	<u>64.1</u>	<u>0.73</u>
Δ_{avg} (%)	46.77	52.40	47.56	40.29	41.35	40.16	46.92	40.53	36.84	40.16

Perturbed-antigen robustness protocol. For each RAbD complex, we modify only the antigen coordinates and evaluate against the original native complex. These perturbed structures are synthetic inputs rather than alternative biological ground truths. Epitope and distant perturbations probe interface-local uncertainty and provide an off-interface control, respectively. The reported perturbation level denotes the RMSD after spatial smoothing, rather than the standard deviation of the initial noise. All methods receive identical perturbed antigen coordinates, and TrustCDR recomputes and caches reliability for each perturbed input. We use 1.5Å as moderate uncertainty and 3.0Å as a stress test; the mild 0.8Å results are provided in the supplementary material. The perturbed antigen is used only for conditioning during generation. After generation, each predicted antibody is paired with the original unperturbed native antigen, and DockQ is computed against the original native antibody–antigen complex.

Baselines and Metrics. We compare TrustCDR with DiffAb Luo et al. (2022), dyMEAN Kong et al. (2023a), dyAb Tan et al. (2025), and FlowDesign Wu et al. (2025b). All methods receive the complete antigen, native non-CDR framework, and native complex pose and jointly design all six masked CDRs; the baselines natively support this task without architectural adaptation. We report AAR, RMSD, DockQ Basu & Wallner (2016), and affinity improvement percentage (IMP) Luo et al. (2022). AAR measures native-sequence recovery, RMSD measures aligned CDR deviation Kabsch (1976), DockQ assesses interface quality, and IMP measures the fraction of candidates predicted to improve affinity. Higher AAR, DockQ, and IMP are better, whereas lower RMSD is better. Aggregate AAR and RMSD average all six CDRs; Clean denotes the unperturbed antigen. Bold and underlined values indicate the best and second-best results.

4.2 Overall and CDR-Level Performance

TrustCDR obtains the best aggregate clean AAR, RMSD, and DockQ among the compared methods. Relative to the mean baseline performance, TrustCDR improves AAR by 36.84–47.56%, reduces RMSD by 40.16–52.40%, and improves DockQ by 56.25–63.83% across the reported conditions. Under epitope 1.5Å perturbation, DockQ remains 0.75. At 3.0Å, TrustCDR achieves the highest AAR under epitope perturbation, reaching 66.1%. Under the 3.0Å distant perturbation, it retains the second-highest AAR of 64.1%, while dyMEAN gives the highest AAR and lowest RMSD. Complete DockQ values, including the mild 0.8Å conditions, are reported in Appendix C.1. Loop-level AAR and RMSD results for all perturbation levels are reported in Appendix C.2.

4.3 Ablation Study

We evaluate controlled variants of the reliability pathway and relative trust-weighting design. The first variant removes reliability estimation and all reliability-derived guidance from pair refinement. The second uses bidirectional trust weighting, allowing higher-reliability regions to be amplified and lower-reliability regions to be attenuated. The third retains raw down-only weighting but removes mean-preserving normalization. Four additional variants retain only the coordinate, topology, orientation, or surface reliability channel, while the final row gives the full TrustCDR setting.

Table 2: Mechanism ablation under clean, epitope 1.5Å, and distant 1.5Å inputs.

Variant	Clean			Epitope 1.5Å			Distant 1.5Å		
	AAR (%)	RMSD (Å)	DockQ	AAR (%)	RMSD (Å)	DockQ	AAR (%)	RMSD (Å)	DockQ
No reliability	64.3	0.75	0.76	64.2	0.76	0.71	64.4	0.75	0.76
Bidirectional	65.1	0.76	0.76	65.0	0.76	0.71	65.1	0.76	0.76
Down-only	65.5	0.76	0.76	65.3	0.76	0.71	65.5	0.75	0.76
Coordinate	65.3	0.70	**0.77**	66.3	0.71	0.71	65.2	**0.69**	0.75
Topology	65.3	0.71	0.76	66.3	0.71	0.71	65.2	0.71	0.75
Orientation	65.4	0.70	**0.77**	66.5	**0.70**	0.71	65.3	0.71	0.75
Surface	65.3	0.69	**0.77**	66.3	0.71	0.71	65.2	0.70	0.76
TrustCDR	**68.5**	**0.57**	0.76	**68.8**	0.73	**0.75**	**68.5**	0.73	**0.77**

Table 3: Interface quality and affinity improvement under clean, epitope 1.5Å, and distant 1.5Å inputs.

	DockQ			IMP (%)		
Method	Clean	Epitope 1.5Å	Distant 1.5Å	Clean	Epitope 1.5Å	Distant 1.5Å
DiffAb	0.72	0.71	0.71	11.10	7.40	11.10
dyMEAN	0.34	0.30	0.33	16.70	9.30	**14.80**
dyAb	0.26	0.26	0.25	9.60	15.40	13.50
FlowDesign	0.59	0.59	0.59	11.10	14.80	13.00
TrustCDR	**0.76**	**0.75**	**0.77**	**18.50**	**20.40**	**14.80**

Table 2 shows that removing reliability consistently reduces AAR and degrades clean RMSD and epitope-perturbed DockQ. Raw down-only weighting gives higher AAR than bidirectional weighting but remains below the full model, supporting both conservative attenuation and mean-preserving normalization. Individual channels occasionally improve isolated clean metrics but yield lower sequence recovery and perturbed-input DockQ. Combining all four channels with the complete weighting design gives the strongest overall balance across the reported AAR, RMSD, and DockQ results. Per-channel loop-level results and the sensitivity analysis for the attenuation strength are provided in Appendices C.4 and C.5, respectively.

4.4 Affinity-Oriented Evaluation

Sequence recovery, RMSD, and DockQ characterize native similarity and interface structure but not energetic favorability. IMP complements them by reporting the fraction of candidates predicted to improve affinity under a shared protocol. Table 3 shows that TrustCDR obtains the highest DockQ and IMP under the clean and epitope-perturbed inputs, and ties for the highest IMP under distant perturbation. Together with Table 1, this shows that strong sequence and structural metrics accompany a larger fraction of favorable candidates in the computational affinity screen. IMP remains a computational proxy rather than an experimental binding measurement. Complex-level paired bootstrap intervals for the DockQ differences across the four primary perturbed conditions are reported in Appendix C.7.

4.5 External Study on a Neutralizing Antibody

To examine transfer beyond the benchmark split, we evaluate PDB 7V0P, which is excluded from the SAbDab-derived training and validation sets and from the RAbD test set. This cryo-EM complex contains SINV/EEEV bound to the Fab fragment of the potently neutralizing human antibody IgG-106, providing a biologically meaningful viral neutralization interface RCSB Protein Data Bank (2023); Williamson et al. (2023). Figure 3 compares the generated CDR structures with the experimental reference. TrustCDR gives the lowest mean RMSD of 0.99Å and the highest DockQ of 0.79, whereas dyMEAN gives the highest mean AAR of 65.5%. Per-loop values for this external case are

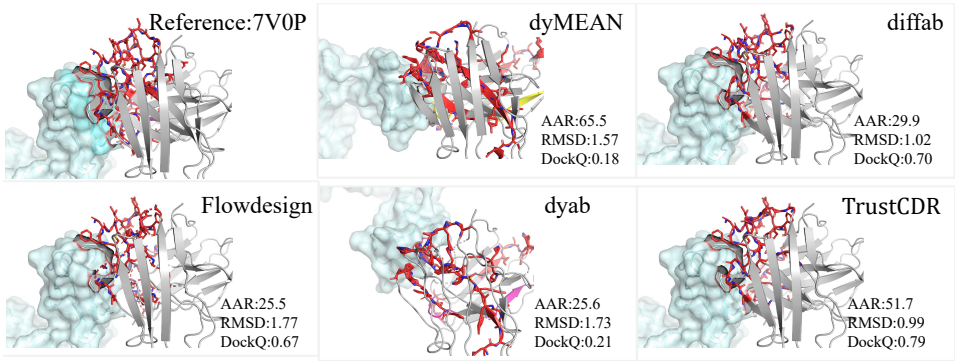

Figure 3: CDR-level results for the external 7V0P.

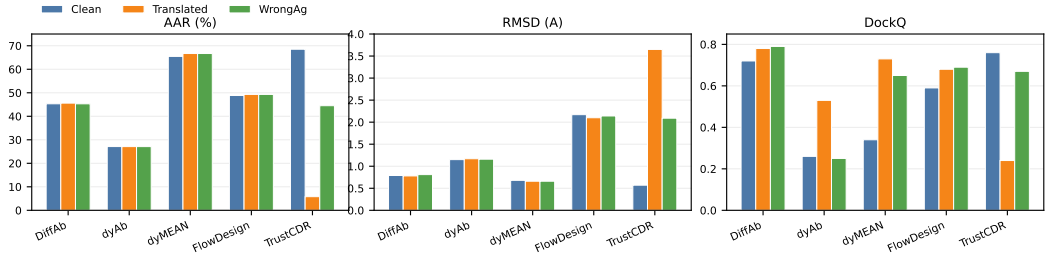

Figure 4: Antigen-condition diagnostic with Clean, Translated 1000Å, and WrongAg inputs.

reported in Appendix C.3. For this complex, TrustCDR better preserves loop geometry and interface arrangement even though another method recovers more native residues.

4.6 Antigen-Condition Diagnostics

We compare Clean, a 1000Å antigen translation that breaks the coordinate frame, and WrongAg, which substitutes another antigen. These corruptions diagnose condition utilization rather than realistic robustness, and TrustCDR recomputes reliability for each input. Figure 4 reports CDR AAR, RMSD, and DockQ; complete numerical values are reported in Appendix C.6. Across all four baselines, AAR and RMSD are nearly invariant to translation and replacement. Combined with their framework-conditioned pathways, this provides evidence that their sequence and loop-geometry outputs are driven mainly by antibody context rather than target-specific antigen information, the characteristic antigen-blindness pattern. TrustCDR instead degrades under both corruptions, showing dependence on a coherent, target-matched antigen. Table 1 separately tests robustness to local structural uncertainty.

5 Conclusion

We studied antigen-conditioned CDR co-design when local antigen geometry lacks an explicit residue-level reliability estimate. TrustCDR combines reference-free reliability estimation with reliability-guided relative trust weighting of residual pair corrections. It uses raw down-only weights followed by mean-preserving normalization to reduce the relative influence of lower-reliability regions while preserving the base representation. Experiments show strong clean performance, robust interface quality under perturbed inputs, and favorable affinity improvement rates.

TrustCDR does not require paired antigen conformations or oracle epitope reliability labels. The present study does not include experimentally paired antigen perturbation trajectories for supervised reliability calibration. Synthetic perturbations therefore provide controlled geometric stress tests, while evaluation against experimentally observed antigen variability remains future work.

Reproducibility Statement

The main paper and its appendix specify the reliability estimator, residual pair update, weighting rule, data split, evaluation metrics, perturbation protocol, implementation details, derivations, complete per-condition results, ablations, and parameter settings. The anonymous submission package contains the source files, figures, bibliography, configurations, and scripts needed to reproduce the paper.

AI Use Statement

The authors declare that no AI tools were used in the conduct of this research or in the preparation of this manuscript. All content, including the text, figures, analyses, and conclusions, was produced and reviewed solely by the authors.

References

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630:493–500, 2024. doi: 10.1038/s41586-024-07487-w.

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

Mansoor Ahmed and Murray Patterson. Agforce enables antigen-conditioned generative antibody design, 2026. URL https://arxiv.org/abs/2605.21610.

B. Al-Lazikani, A. M. Lesk, and C. Chothia. Standard conformations for the canonical structures of immunoglobulins. *Journal of Molecular Biology*, 273(4):927–948, 1997. doi: 10.1006/jmbi.1997.1354.

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=li7qeBbCR1t.

Ismail Alkhouri et al. On the robustness of alphafold: A covid-19 case study. *arXiv preprint arXiv:2301.04093*, 2023. URL https://arxiv.org/abs/2301.04093.

Anonymous. Energy-tilted interface potentials for antibody design, 2027. Manuscript under review.

Amélie Barozet, Marc Bianciotto, Thierry Siméon, Hervé Minoux, and Juan Cortés. Conformational changes in antibody fab fragments upon binding and their consequences on the performance of docking algorithms. *Immunology Letters*, 200:5–15, 2018. doi: 10.1016/j.imlet.2018.05.004.

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

Gregory R Bowman and Phillip L Geissler. Extensive conformational heterogeneity within protein cores. *The Journal of Physical Chemistry B*, 118(24):6417–6423, 2014.

Henri Dreyer, Henry Kenlay, and Charlotte M Deane. Conformation-aware structure prediction of antigen-recognizing immune proteins. *mAbs*, 18(1):2602217, 2026.

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.

Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Zidek, Russ Bates, Sam Blackwell, Jason Yim, et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2022. doi: 10.1101/2021.10.04.463034.

Christos P Feidakis, Radoslav Krivak, David Hoksza, and Marian Novotny. Ahoj-db: A pdb-wide assignment of apo & holo relationships based on individual protein–ligand interactions. *Journal of Molecular Biology*, 436(17):168545, 2024.

Hans Frauenfelder, Stephen G Sligar, and Peter G Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, 1991.

Thomas Hamelryck. An amino acid has two sides: A new 2d measure provides a different view of solvent exposure. *Proteins: Structure, Function, and Bioinformatics*, 59(1):38–48, 2005. doi: 10.1002/prot.20379. URL https://doi.org/10.1002/prot.20379.

Nao Hiranuma, Hahnbeom Park, Minkyung Baek, Ivan Anishchenko, Justas Dauparas, and David Baker. Improved protein structure refinement guided by deep learning based accuracy estimation. *Nature Communications*, 12:1340, 2021. doi: 10.1038/s41467-021-21511-x. URL https://doi.org/10.1038/s41467-021-21511-x.

Chutian Huang, Zijing Liu, Shengyuan Bai, Linwei Zhang, Chencheng Xu, Zhe Wang, Yang Xiang, and Yuanpeng Xiong. Pf-abgen: A reliable and efficient antibody generator via poisson flow. In *ICLR 2023-Machine Learning for Drug Discovery workshop*, 2023. URL https://hdl.handle.net/1783.1/126118.

Leo C James and Dan S Tawfik. Conformational diversity and protein evolution–a 60-year-old hypothesis revisited. *Trends in biochemical sciences*, 28(7):361–368, 2003.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

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

Asif Khan, Alexander I Cowen-Rivers, Antoine Grosnit, Derrick-Goh-Xin Deik, Philippe A Robert, Victor Greiff, Eva Smorodina, Puneet Rawat, Rahmad Akbar, Kamil Dreczkowski, et al. Toward real-world automated antibody design with combinatorial bayesian optimization. *Cell Reports Methods*, 3(1), 2023. URL https://doi.org/10.5281/zenodo.7344859.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. End-to-end full-atom antibody design. *ICML'23*, 2023a. URL https://doi.org/10.48550/arXiv.2302.00203.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. Conditional antibody design as 3d equivariant graph translation, 2023b. URL https://arxiv.org/abs/2208.06073.

Kelly K Lee. How a broadly neutralizing antibody grapples with antigenic and conformational diversity in dengue virus. *Cell*, 184(25):6015–6016, 2021.

Julia Koehler Leman, Brian D Weitzner, Steven M Lewis, Jared Adolf-Bryfogle, Nawsad Alam, Rebecca F Alford, Melanie Aprahamian, David Baker, Kyle A Barlow, Patrick Barth, et al. Macromolecular modeling and design in rosetta: recent methods and frameworks. *Nature methods*, 17 (7):665–680, 2020.

Jiahan Li, Shitong Luo, Congyue Deng, Chaoran Cheng, Jiaqi Guan, Leonidas J. Guibas, Jian Peng, and Jianzhu Ma. Orientation-aware graph neural networks for protein structure representation learning. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=WcTLZrpzfe.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *Proceedings of the Eleventh International Conference on Learning Representations (ICLR)*, 2023.

Chu'nan Liu, Lilian Denzler, Oliver EC Hood, and Andrew CR Martin. Do antibody cdr loops change conformation upon binding? *mAbs*, 16(1):2322533, 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. URL https://doi.org/10.1101/2022.07.10.499510.

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. URL https://doi.org/10.1093/bioinformatics/btt473.

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.

Tobias H Olsen, Iain H Moal, and Charlotte M Deane. Addressing the antibody germline bias and its effect on language models for improved antibody design. *Bioinformatics*, 40(11):btae618, 2024. URL https://doi.org/10.1093/bioinformatics/btae618.

Matthew I J Raybould, Claire Marks, Alan P Lewis, Jiye Shi, Alexander Bujotzek, Bruck Taddese, and Charlotte M Deane. Thera-sabdab: the therapeutic structural antibody database. *Nucleic Acids Research*, 48(D1):D383–D388, 09 2019. ISSN 0305-1048. doi: 10.1093/nar/gkz827. URL https://doi.org/10.1093/nar/gkz827.

RCSB Protein Data Bank. 7V0P: Cryo-em structure of sinv/eeev in complex with the fab fragment of the potently neutralizing human antibody igg-106, 2023. URL https://www.rcsb.org/structure/7V0P.

Valerio Scardino, Juan I Di Filippo, and Claudio N Cavasotto. How good are alphafold models for docking-based virtual screening? *iScience*, 26(1), 2023. URL https://www.cell.com/iscience/fulltext/S2589-0042(22)02193-9.

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.

Lauren E. Williamson, Anumita Bandyopadhyay, Kelsey Bailey, Deepti Sirohi, Thomas Klose, Justin G. Julander, Richard J. Kuhn, and James E. Jr. Crowe. Structural constraints link differences in neutralization potency of human anti-eastern equine encephalitis virus monoclonal antibodies. *Proceedings of the National Academy of Sciences*, 120:e2213690120, 2023. doi: 10.1073/pnas.2213690120.

Jiayang Wu, Xingyi Zhang, Xiangyu Dong, Kun Xie, Ziqi Liu, Wensheng Gan, Sibo Wang, and Le Song. Towards more accurate full-atom antibody co-design, 2025a. URL https://arxiv.org/abs/2502.19391.

Jun Wu, Xiangzhe Kong, Ningguan Sun, Jing Wei, Sisi Shan, Fuli Feng, Feng Wu, Jian Peng, Linqi Zhang, Yang Liu, and Jianzhu Ma. Flowdesign: Improved design of antibody cdrs through flow matching and better prior distributions. *Cell Systems*, 16(6):101270, 2025b. ISSN 2405-4712. doi: https://doi.org/10.1016/j.cels.2025.101270. URL https://www.sciencedirect.com/science/article/pii/S2405471225001036.

Lirong Wu, Haitao Lin, Yufei Huang, Zhangyang Gao, Cheng Tan, Yunfan Liu, Tailin Wu, and Stan Z. Li. Relation-aware equivariant graph networks for epitope-unknown antibody design and specificity optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 895–904, 2025c. doi: 10.1609/aaai.v39i1.32074.

Xuezhi Xie, Pedro A Valiente, Jin Sub Lee, Jisun Kim, and Philip M Kim. Antibody-sgm, a score-based generative model for antibody heavy-chain design. *Journal of Chemical Information and Modeling*, 64(17):6745–6757, 2024. URL https://doi.org/10.1021/acs.jcim.4c00711.

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. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Xingyi Zhang, Kun XIE, Ningqiao Huang, Peilin Zhao, Wei Liu, Sibo Wang, Kangfei Zhao, and Biaobin Jiang. Igseek: Fast and accurate antibody sequence design via structure retrieval, 2025. URL https://openreview.net/forum?id=KgiMUvJcwm.

Tian Zhu, Milong Ren, and Haicang Zhang. Antibody design using a score-based diffusion model guided by evolutionary, physical and geometric constraints. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 62531–62548. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/zhu24j.html.

Yibo Zhu, Xiumin Shi, Jingjuan Zhang, Weizhong Sun, and Lu Wang. Abegdiffuser: Antibody sequence-structure codesign with equivariant graph neural networks and diffusion models. *Journal of Chemical Theory and Computation*, 21(21):11307–11317, 2025. URL https://doi.org/10.1021/acs.jctc.5c00990.

A Implementation and Evaluation Details

A.1 Data and Evaluation Protocol

All antibody–antigen complexes are collected from SAbDab. Antibody sequences are clustered with MMseqs2 easy-cluster at 50% sequence identity and 80% coverage, and clusters overlapping the RAbD test set are excluded from training and validation. Potential overlap is additionally screened on the antigen side using sequence- and structure-based criteria; the corresponding thresholds and scripts are provided in the anonymous code package. The resulting split contains 5,247 training and 606 validation complexes, while the 58-complex RAbD test set is reserved exclusively for evaluation. Preprocessing retains standard amino acids, removes non-protein entries and hydrogen atoms, validates antibody numbering, and masks missing backbone coordinates.

For every method, the complete antigen, native non-CDR antibody framework, and native antibody–antigen relative pose are provided, while the identities and coordinates of all six CDRs are simultaneously masked. The compared baselines are evaluated in their original antigen-conditioned sequence–structure design formulations without task-specific architectural modification. Each method is trained with three fixed random seeds: 2025, 2026, and 2027. Each method generates ten candidates per test complex, and metrics are averaged first over candidates and then over the test set. RMSD is computed on generated Cα coordinates after alignment to the native antibody context.

A.2 Reliability Estimation

For each supplied antigen, TrustCDR constructs $K = 8$ sequence-preserving conformers using small, locally smoothed Cα perturbations. These conformers are used only to estimate reliability and are not supplied as multiple generator conditions. The resulting field is cached and shared across candidates and flow steps for that antigen. The four uncertainty channels measure coordinate displacement, nearest-neighbor overlap, local-frame change, and variation in surface exposure. For each channel, raw values are standardized over valid residues using the median and interquartile range, clipped to $[-3, 3]$, and linearly mapped to $[0, 1]$; a constant channel maps to zero. This robust within-antigen normalization makes heterogeneous descriptors comparable while preserving their relative spatial variation. It represents relative geometric reliability within the supplied antigen rather than a calibrated physical probability. The implementation computes $r_j^m = \mathrm{clip}(\exp(-u_j^m), 0.01, 0.99)$ for numerical stability. The main paper states the unclipped analytical response; the next section derives its properties and the slightly tighter bounds induced by endpoint clipping. The full field $U_{A,j}$ contains the four normalized uncertainties and their monotonic responses, while ρ_j aggregates them into a scalar geometric reliability score used to derive down-only relative trust weights. Each contribution to ρ_j decreases strictly as its normalized uncertainty increases.

A.3 Leakage Control and Perturbation Evaluation

The antigen and antibody framework are observed, but the native identities and coordinates of all six masked CDRs are excluded from the generation input. CDR–antigen relations are constructed from the current generated CDR state, and the reliability estimator uses only the supplied antigen without native contact or reliability labels.

For evaluation-condition construction only, the core epitope contains antigen residues with at least one heavy atom within 6.0Å of a CDR heavy atom. This set is expanded to antigen residues whose Cα atom lies within 8.0Å of a core-epitope Cα atom, and the resulting expanded epitope is used for interface-local perturbation. The distant control selects antigen residues whose Cα atoms are more than 20.0Å from the epitope.

Initial random displacements are spatially smoothed with a Gaussian kernel of width 4.0Å and rescaled to the specified target Cα RMSD. Thus, 0.8Å, 1.5Å, and 3.0Å denote target perturbation RMSDs rather than Gaussian noise standard deviations. Only antigen coordinates are modified, and all methods receive identical perturbed structures under each condition. The native antibody and clean complex remain the evaluation reference. The perturbed antigen is used only as the conditioning structure. For DockQ evaluation, each generated antibody is combined with the original unperturbed native antigen and compared with the original native antibody–antigen complex. TrustCDR recomputes and caches the reliability field for each perturbed antigen before generation.

Table 4: Reliability estimation settings.

Item	Value	Description
Ensemble size	$K = 8$	Antigen-only local perturbations
Raw noise scale	0.30Å	Cα Gaussian offsets
Smoothing window	3 residues	Sequence-local moving average
Topology neighbors	16	Antigen Cα nearest neighbors
Surface cutoff	10Å	Neighbor-count exposure proxy
Axes	4	Coordinate, topology, orientation, surface
u_j^m	$[0, 1]$	Normalized uncertainty on axis m
r_j^m	Clipped $\exp(-u_j^m)$	Monotonic reliability response
ρ_j	Bounded scalar	Residue-level geometric reliability

Table 5: Training time for 200,000 iterations.

Method	Training time (h)
DiffAb	22.15
TrustCDR	30.58
FlowDesign	34.75
dyMEAN	43.73
dyAb	46.95

A.4 Training Details

The residual correction ΔZ matches the CDR–antigen pair-feature dimension and is not a coordinate displacement. The global scale α is initialized to zero, while a small nonzero residual-head initialization allows it to receive a gradient before the residual branch becomes active. The sequence, position, and orientation flow-matching losses have equal weight:

$$\mathcal{L} = \mathcal{L}_{\text{seq}} + \mathcal{L}_{\text{pos}} + \mathcal{L}_{\text{rot}}. \tag{15}$$

The default trust-weighting configuration uses $\beta = 0.10$, selected on the validation set, and a fixed safety cap $c = 0.20$ before valid-pair mean normalization.

All methods are trained for exactly 200,000 iterations, and the checkpoint at iteration 200,000 is used for evaluation without validation- or test-based checkpoint selection. TrustCDR uses Adam with a learning rate of 10^{-4} and a batch size of 6. The remaining method-specific optimization settings, loss configurations, and data-augmentation parameters are reported in the supplementary material or provided in the anonymous configuration files. Table 5 reports wall-clock training time on the shared hardware environment.

TrustCDR trains faster than FlowDesign, dyMEAN, and dyAb, while DiffAb has the lowest training time.

A.5 Flow Backbone and Modular Transfer

The flow-matching backbone receives the current CDR state, the antigen, the known antibody framework, the refined pair representation $Z'(t)$, and the flow time t. It predicts sequence, position, and orientation velocities from shared conditional features. The implementation uses 128-dimensional residue features and 64-dimensional pair features in a six-layer geometric-attention flow backbone with 12 attention heads. The reliability encoder and pair-modulation heads use 128-dimensional

hidden states; the residual and applicability heads contain two and three layers, respectively. The refined pair representation is rebuilt after every flow update from the current CDR state, while the antigen reliability field remains cached for the supplied antigen. The reliability encoder, residual correction, and pair-applicability gate act before this shared generator input; they do not replace the generator backbone.

The module can be transferred to an antigen-conditioned generator that exposes a compatible CDR–antigen pair representation. The pair-feature dimensions and update interface must match, and the added module must be trained or fine-tuned with the target backbone. This interface-level compatibility does not constitute a cross-backbone experimental evaluation.

A.6 Computing Environment

Experiments were conducted on a Linux workstation with two Intel Xeon E5-2690 v3 CPUs (24 cores and 48 threads in total), 128 GB DDR4 memory, and four NVIDIA GeForce RTX 4090 GPUs with 24 GB memory each. The system used Ubuntu 22.04.5 LTS, Linux kernel 6.8.0-124-generic, NVIDIA driver 570.86.10, Python 3.8.20, PyTorch 2.4.1, and CUDA 12.1. The anonymous code package contains the configurations and scripts used for preprocessing, training, generation, evaluation, and analysis.

B Theoretical Properties of Reliability-Guided Relative Trust Weighting

This section establishes properties of the reliability construction and its use in residual pair refinement. The results concern the weighting mechanism itself; they do not constitute a guarantee on the complete generative trajectory or experimental binding.

B.1 Robust Channel Normalization

Let $\mathcal{V}_A$ denote the valid residues of antigen A. For uncertainty channel m, the implementation first computes

$$z_j^m = \mathrm{clip}\left(\frac{q_j^m - \mathrm{med}_{\ell\in\mathcal{V}_A}\, q_\ell^m}{\max(\mathrm{IQR}_{\ell\in\mathcal{V}_A}\, q_\ell^m, \varepsilon_q)}, -3, 3\right). \tag{16}$$

It then maps the clipped values to a common interval:

$$u_j^m = \begin{cases} \dfrac{z_j^m - z_{\min}^m}{z_{\max}^m - z_{\min}^m}, & z_{\max}^m - z_{\min}^m > \varepsilon_q, \\ 0, & \text{otherwise}, \end{cases} \tag{17}$$

where the extrema are taken over $\mathcal{V}_A$. Thus $u_j^m \in [0,1]$, and a constant channel contributes no artificial spatial variation. For a positive affine rescaling $q_j^{m\prime} = aq_j^m + b$ with $a > 0$, the median and interquartile range transform as $a\,\mathrm{med}(q^m) + b$ and $a\,\mathrm{IQR}(q^m)$. Away from the numerical safeguard, Equations equation 16–equation 17 therefore produce the same u_j^m. This explains why heterogeneous channels can be combined as relative within-antigen responses, but it does not make u_j^m a calibrated probability.

B.2 Rigid-Transform Invariance of Local Responses

Consider a common rigid transformation $T(\mathbf{x}) = R\mathbf{x} + \mathbf{t}$ with $R \in \mathrm{SO}(3)$, applied jointly to the observed antigen and every member of its perturbation ensemble. The coordinate response is unchanged because

$$\left\|T(\mathbf{x}_j^{(k)}) - T(\mathbf{x}_j^{(0)})\right\|_2 = \left\|R(\mathbf{x}_j^{(k)} - \mathbf{x}_j^{(0)})\right\|_2 = \left\|\mathbf{x}_j^{(k)} - \mathbf{x}_j^{(0)}\right\|_2. \tag{18}$$

Pairwise distances are also preserved, so the nearest-neighbor sets and distance-based surface-exposure descriptors do not change. Consequently, the topology Jaccard response and surface-exposure variation are invariant. For local frames, $O_j^{(k)}$ transforms to $RO_j^{(k)}$, and

$$(RO_j^{(0)})^\mathsf{T}(RO_j^{(k)}) = (O_j^{(0)})^\mathsf{T} O_j^{(k)}. \tag{19}$$

The relative rotation, and hence its geodesic distance on $\mathrm{SO}(3)$, is unchanged. The four response scores therefore describe local ensemble variation rather than the arbitrary global coordinate frame.

B.3 Reliability Bounds and Monotonicity

For the analytical response $r_j^m = \exp(-\lambda u_j^m)$ with $u_j^m \in [0,1]$, each channel contributes

$$h(u) = \frac{(1-u) + \exp(-\lambda u)}{2}. \tag{20}$$

Its endpoint values and derivative are

$$h(0) = 1, \qquad h(1) = \frac{e^{-\lambda}}{2}, \qquad h'(u) = -\frac{1 + \lambda e^{-\lambda u}}{2} < 0. \tag{21}$$

Since ρ_j is the mean of $h(u_j^m)$ over the four channels,

$$\frac{e^{-\lambda}}{2} \le \rho_j \le 1, \qquad \frac{\partial \rho_j}{\partial u_j^m} = -\frac{1 + \lambda e^{-\lambda u_j^m}}{2|\mathcal{M}|} < 0. \tag{22}$$

Thus increasing any normalized uncertainty strictly lowers the scalar reliability while the other channels are fixed.

The implementation uses $\lambda = 1$ and clips the response to $[0.01, 0.99]$. Because $e^{-u} \ge e^{-1} > 0.01$ on $[0,1]$, only the upper endpoint is affected:

$$r_{\mathrm{impl}}(u) = \min(0.99, e^{-u}). \tag{23}$$

Let $u_0 = -\log(0.99)$. The implemented scalar therefore satisfies

$$\frac{e^{-1}}{2} \le \rho_j^{\mathrm{impl}} \le 0.995, \tag{24}$$

and its channel-wise derivative is, except at the clipping boundary,

$$\frac{\partial \rho_j^{\mathrm{impl}}}{\partial u_j^m} = \begin{cases} -\dfrac{1}{2|\mathcal{M}|}, & 0 \le u_j^m < u_0, \\ -\dfrac{1 + e^{-u_j^m}}{2|\mathcal{M}|}, & u_0 < u_j^m \le 1. \end{cases} \tag{25}$$

Endpoint clipping therefore tightens the upper bound but preserves strict monotonicity of ρ_j because the $(1 - u_j^m)$ term remains active.

B.4 Mean Preservation and Relative Trust

For completeness, the main-text description is implemented as

$$\widehat{\rho}_j = \tanh\left(\frac{\rho_j - \mu_A}{\sigma_A + \varepsilon}\right), \qquad w_j^{\mathrm{raw}} = 1 - \beta \max(0, -\widehat{\rho}_j), \qquad w_j^{\mathrm{clip}} = \mathrm{clip}(w_j^{\mathrm{raw}}, 1-c, 1). \tag{26}$$

The clipped weights are then normalized over the valid antigen partners of each CDR position as follows.

Since $\widehat{\rho}_j \in [-1,1]$, the down-only rule gives $w_j^{\mathrm{raw}} \in [1-\beta, 1]$. After clipping to $[1-c, 1]$, define

$$\ell = 1 - \min(\beta, c), \qquad \ell \le w_j^{\mathrm{clip}} \le 1. \tag{27}$$

For valid pair set $\mathcal{J}_i$, let

$$\bar{w}_i = \frac{1}{|\mathcal{J}_i|} \sum_{j \in \mathcal{J}_i} w_j^{\mathrm{clip}}, \qquad w_{ij} = \frac{w_j^{\mathrm{clip}}}{\bar{w}_i + \varepsilon}. \tag{28}$$

The row-wise mean is

$$\frac{1}{|\mathcal{J}_i|} \sum_{j \in \mathcal{J}_i} w_{ij} = \frac{\bar{w}_i}{\bar{w}_i + \varepsilon} \approx 1. \tag{29}$$

Table 6: DockQ under all antigen perturbation conditions.

Method	Clean	Epitope 1.5Å	Epitope 3.0Å	Distant 1.5Å	Distant 3.0Å
DiffAb	<u>0.72</u>	<u>0.71</u>	<u>0.72</u>	<u>0.71</u>	<u>0.72</u>
dyMEAN	0.34	0.30	0.34	0.33	0.35
dyAb	0.26	0.26	0.26	0.25	0.26
FlowDesign	0.59	0.59	0.59	0.59	0.59
TrustCDR	**0.76**	**0.75**	**0.76**	**0.77**	**0.75**

The approximation is exact in the limit $\varepsilon \to 0$. Because every weight in one row has the same positive denominator, normalization preserves ordering:

$$w_j^{\text{clip}} \le w_k^{\text{clip}} \quad \Longrightarrow \quad w_{ij} \le w_{ik}. \tag{30}$$

Moreover,

$$\frac{\ell}{1+\varepsilon} \le w_{ij} \le \frac{1}{\ell+\varepsilon}. \tag{31}$$

With $\beta = 0.10$ and $c = 0.20$, $\ell = 0.90$, giving the approximate interval $[0.90, 1.11]$. Values above one arise only from relative mean preservation; the raw down-only transformation never amplifies a residue.

B.5 Base-Path Preservation and Residual Influence

For a valid CDR–antigen pair, the reliability-weighted residual update is

$$\mathbf{z}'_{ij} - \mathbf{z}_{ij} = \alpha g_{ij} w_{ij} \Delta \mathbf{z}_{ij}. \tag{32}$$

Because $g_{ij} \in (0, 1)$ and Equation equation 31 bounds w_{ij},

$$\left\|\mathbf{z}'_{ij} - \mathbf{z}_{ij}\right\|_2 \le \frac{|\alpha|}{\ell+\varepsilon} \left\|\Delta \mathbf{z}_{ij}\right\|_2. \tag{33}$$

Under the default settings, reliability-guided weighting can scale the proposal by at most approximately 1.11 before the global coefficient α. At initialization, $\alpha = 0$ gives $\mathbf{z}'_{ij} = \mathbf{z}_{ij}$ exactly, so the residual branch starts from the base pair representation. The learned vector $\Delta \mathbf{z}_{ij}$ determines the correction, g_{ij} determines pair applicability, and w_{ij} imposes an explicit reliability-derived relative trust scale.

C Additional Experimental Results

Bold numbers in benchmark, case-study, and sensitivity tables mark the best value in each metric column.

C.1 Additional Mild-Perturbation Results

Table 7 reports the 0.8Å conditions omitted from the main comparison. TrustCDR obtains the highest AAR and DockQ under both perturbation regions, while dyMEAN has the lowest RMSD. The mild tests therefore preserve the interface-quality advantage observed under the stronger perturbations in the main text.

C.2 CDR-Level Performance under Antigen Perturbations

The main paper reports clean CDR-level results. Tables 9 and 10 provide the corresponding raw AAR and RMSD values for every perturbed condition. DockQ is defined for the complete antibody–antigen interface and therefore has no CDR-level decomposition. Under the 0.8Å and 1.5Å perturbations, TrustCDR retains the highest AAR on most loops and consistently gives the lowest H3 RMSD. At 3.0Å, TrustCDR and dyMEAN each achieve the highest AAR on three CDRs. TrustCDR gives the higher aggregate AAR under epitope perturbation (66.14% versus 65.76%), whereas dyMEAN remains higher under distant perturbation (65.79% versus 64.10%).

Table 7: Additional results under mild antigen perturbations.

Method	Epitope 0.8Å			Distant 0.8Å		
	AAR (%)	RMSD (Å)	DockQ	AAR	RMSD	DockQ
DiffAb	45.29	0.79	0.71	45.21	0.79	0.72
dyMEAN	65.25	**0.69**	0.31	65.43	**0.68**	0.33
dyAb	27.12	1.15	0.26	27.12	1.16	0.26
FlowDesign	48.83	2.28	0.59	48.71	2.28	0.59
TrustCDR	**68.71**	0.73	**0.75**	**66.86**	0.73	**0.76**

Table 8: Clean CDR-level performance.

Method	AAR (%)						RMSD (Å)					
	H1	H2	H3	L1	L2	L3	H1	H2	H3	L1	L2	L3
DiffAb	64.09	35.20	26.56	52.60	49.77	43.66	0.59	0.46	1.86	0.71	0.33	0.77
dyMEAN	75.30	67.04	**40.27**	73.56	78.31	58.33	0.60	0.48	1.70	0.46	**0.19**	**0.66**
dyAb	53.09	43.57	19.24	9.48	7.27	30.05	1.02	1.15	1.85	1.39	0.41	1.11
FlowDesign	63.16	40.17	21.13	62.12	60.87	45.60	1.90	1.74	3.05	2.23	1.81	2.27
TrustCDR	**80.17**	**69.83**	37.23	**74.65**	**83.37**	**65.85**	**0.45**	**0.32**	**1.12**	**0.44**	0.36	0.74

Clean CDR-level performance. Because aggregate metrics can hide loop-specific behavior, this table reports clean CDR-level AAR and RMSD. TrustCDR gives the highest AAR on five loops and the lowest RMSD on four; dyMEAN retains the best H3 AAR and the lowest L2 and L3 RMSD.

C.3 External Case Study

We additionally evaluate the 7V0P_H_L_A complex under the same candidate-aggregation protocol. This external complex is excluded from the SAbDab-derived training and validation sets and from the RAbD test set. PDB 7V0P is a cryo-EM structure of SINV/EEEV bound to the Fab fragment of the potently neutralizing human antibody IgG-106 RCSB Protein Data Bank (2023); Williamson et al. (2023). Table 11 uses the same display precision as the main-text visualization: one decimal place for AAR and two decimal places for RMSD and DockQ.

dyMEAN gives the highest mean AAR in this case, while TrustCDR gives the lowest mean RMSD and the highest DockQ. TrustCDR also attains the highest L2 recovery. The case therefore illustrates a sequence–geometry trade-off: the strongest native-sequence recovery does not necessarily yield the most accurate mean loop geometry or interface quality.

C.4 Detailed Ablation Results

Figure 5 resolves the aggregate ablation results by loop. The down-only variant gives the strongest recovery on L2 and L3 across the three conditions, and gives the lowest H3 RMSD among the incomplete variants. Bidirectional trust weighting improves some H1 and H2 sequence scores but does not yield a consistent structural advantage. The loop-wise patterns therefore agree with the main ablation: raw down-only weighting without mean preservation is a useful partial design, while the complete relative trust-weighting design gives the strongest aggregate behavior.

Table 12 provides the complete CDR-level breakdowns for the four single-channel variants. Individual channels occasionally improve isolated geometric metrics, including clean DockQ values of up to 0.77 and loop-specific RMSD reductions, but none matches the sequence recovery of the full multi-channel model. Under epitope perturbation, all single-channel variants obtain 0.71 DockQ, compared with 0.75 for TrustCDR. Under distant perturbation, their DockQ values range from 0.75 to 0.76, whereas the full model reaches 0.77. These results indicate that individual geometric cues can benefit particular metrics, while their combination provides a stronger balance of sequence recovery and perturbed-interface quality.

Table 9: CDR-level AAR (%) under all antigen perturbation conditions.

Condition	Method	H1	H2	H3	L1	L2	L3
Epitope 0.8Å	DiffAb	64.16	35.72	26.18	52.85	49.46	43.35
	dyMEAN	74.90	67.04	**39.11**	73.27	78.31	58.90
	dyAb	53.09	43.57	19.24	9.48	7.27	30.05
	FlowDesign	62.60	40.29	19.70	62.00	60.80	47.59
	TrustCDR	**78.94**	**71.84**	38.11	**75.77**	**82.60**	**65.00**
Epitope 1.5Å	DiffAb	63.89	35.56	26.48	52.73	49.47	43.54
	dyMEAN	74.90	67.07	**38.65**	73.56	78.31	58.91
	dyAb	53.09	43.57	19.24	9.48	7.27	30.05
	FlowDesign	62.60	39.95	20.60	62.00	60.00	47.75
	TrustCDR	**78.50**	**71.84**	37.79	**76.10**	**82.85**	**65.60**
Epitope 3.0Å	DiffAb	64.10	35.92	26.30	52.77	49.17	43.29
	dyMEAN	75.03	67.13	**41.27**	**74.16**	78.31	**58.64**
	dyAb	53.09	43.57	19.24	9.48	7.27	30.05
	FlowDesign	62.69	39.69	21.00	62.01	60.90	47.92
	TrustCDR	**78.44**	**71.31**	37.17	70.73	**82.11**	57.08
Distant 0.8Å	DiffAb	63.93	35.33	26.35	52.58	49.87	43.20
	dyMEAN	75.10	67.04	**40.44**	**73.56**	78.31	58.15
	dyAb	53.09	43.57	19.24	9.48	7.27	30.05
	FlowDesign	61.80	40.10	19.80	62.10	60.84	47.60
	TrustCDR	**79.18**	**67.24**	36.96	73.14	**81.16**	**63.50**
Distant 1.5Å	DiffAb	63.90	35.25	26.38	52.56	49.86	43.16
	dyMEAN	75.30	67.04	**40.50**	73.56	78.31	58.42
	dyAb	53.09	43.57	19.24	9.48	7.27	30.05
	FlowDesign	61.80	40.00	19.80	62.11	60.83	47.60
	TrustCDR	**79.92**	**70.80**	37.04	**74.65**	**83.62**	**65.03**
Distant 3.0Å	DiffAb	64.05	35.24	26.43	52.59	49.84	43.59
	dyMEAN	75.01	**67.04**	**41.72**	74.16	78.31	**58.52**
	dyAb	53.09	43.57	19.24	9.48	7.27	30.05
	FlowDesign	63.21	40.12	21.11	62.13	60.87	47.60
	TrustCDR	**79.24**	63.71	27.15	**74.68**	**82.63**	57.18

C.5 Parameter Sensitivity

We study sensitivity to the attenuation-strength parameter β. After selecting $\beta = 0.10$ on the validation set, we vary $\beta \in \{0.05, 0.10, 0.20, 0.30\}$ while fixing the safety cap at $c = 0.20$ and keeping all other trust-weighting settings unchanged.

With $c = 0.20$ fixed, the within-condition DockQ range is at most 0.04 across the tested β values, while $\beta = 0.10$ gives the lowest clean RMSD. It also gives the highest AAR under clean input, both epitope perturbations, and distant 1.5Å, but not under distant 0.8Å. These results support the selected attenuation strength while showing that performance is not restricted to a single β value.

C.6 Antigen-Condition Diagnostics

The diagnostic compares the original antigen with a 1000Å translation and an antigen taken from another sample.

The diagnostic inputs probe condition utilization under gross corruption rather than ordinary local robustness. For every evaluated baseline, AAR and RMSD are almost invariant to both translation and antigen replacement. Together with the framework-conditioned pathways in these generators, this pattern suggests limited sensitivity to the supplied antigen condition in this evaluation protocol. TrustCDR responds differently: translation lowers AAR to 5.80 and raises RMSD to 3.65Å, while WrongAg also degrades all three metrics. DockQ is not used alone to infer condition utilization because translation or replacement directly changes the evaluated interface. The controlled local

Table 10: CDR-level RMSD (Å) under all antigen perturbation conditions.

Condition	Method	H1	H2	H3	L1	L2	L3
Epitope 0.8Å	DiffAb	**0.60**	**0.46**	1.87	0.72	0.34	0.78
	dyMEAN	**0.60**	0.49	1.72	**0.46**	**0.19**	**0.67**
	dyAb	1.02	1.15	1.85	1.39	0.41	1.11
	FlowDesign	2.25	1.82	3.15	2.27	1.91	2.25
	TrustCDR	0.62	0.58	**1.35**	0.57	0.42	0.82
Epitope 1.5Å	DiffAb	**0.59**	**0.45**	1.88	0.72	0.34	0.78
	dyMEAN	0.60	0.49	1.68	**0.46**	**0.19**	**0.66**
	dyAb	1.02	1.15	1.85	1.39	0.41	1.11
	FlowDesign	2.25	1.80	3.15	2.27	1.91	2.24
	TrustCDR	0.62	0.57	**1.36**	0.56	0.43	0.82
Epitope 3.0Å	DiffAb	0.60	**0.46**	1.86	0.71	0.33	0.77
	dyMEAN	**0.59**	0.48	1.64	**0.45**	**0.19**	**0.66**
	dyAb	1.03	1.15	1.86	1.39	0.41	1.11
	FlowDesign	2.24	1.78	3.13	2.26	1.90	2.23
	TrustCDR	0.63	0.58	**1.37**	0.58	0.43	0.82
Distant 0.8Å	DiffAb	**0.59**	**0.46**	1.86	0.71	0.33	0.77
	dyMEAN	0.60	0.48	1.69	**0.46**	**0.19**	**0.66**
	dyAb	1.02	1.15	1.86	1.39	0.41	1.12
	FlowDesign	2.26	1.82	3.13	2.28	1.94	2.26
	TrustCDR	0.62	0.57	**1.36**	0.58	0.42	0.82
Distant 1.5Å	DiffAb	**0.59**	**0.46**	1.86	0.71	0.33	0.78
	dyMEAN	0.60	0.49	1.69	**0.46**	**0.19**	**0.66**
	dyAb	1.02	1.15	1.86	1.39	0.41	1.13
	FlowDesign	2.26	1.82	3.13	2.28	1.94	2.27
	TrustCDR	0.62	0.57	**1.36**	0.58	0.42	0.82
Distant 3.0Å	DiffAb	**0.59**	**0.46**	1.86	0.72	0.33	0.77
	dyMEAN	**0.59**	0.48	1.61	**0.45**	**0.19**	**0.66**
	dyAb	1.03	1.15	1.85	1.39	0.41	1.12
	FlowDesign	2.26	1.79	3.13	2.28	1.91	2.26
	TrustCDR	0.62	0.57	**1.36**	0.58	0.43	0.82

Table 11: CDR-level sequence and structural results for the external 7V0P_H_L_A case.

Method	AAR (%)						
	H1	H2	H3	L1	L2	L3	Mean
TrustCDR	71.4	40.0	25.0	54.5	**85.7**	33.3	51.7
dyMEAN	**80.0**	**57.1**	**44.4**	**66.7**	66.7	**77.8**	**65.5**
DiffAb	28.6	20.0	0.0	54.5	42.9	33.3	29.9
FlowDesign	42.9	20.0	0.0	36.4	42.9	11.1	25.5
dyAb	30.0	**57.1**	27.8	16.7	0.0	22.2	25.6

Method	RMSD (Å)							DockQ
	H1	H2	H3	L1	L2	L3	Mean	
TrustCDR	**2.17**	0.85	1.41	0.29	0.80	0.42	**0.99**	**0.79**
DiffAb	2.97	**0.84**	**1.12**	0.42	0.22	0.57	1.02	0.70
dyAb	3.24	1.40	3.64	1.00	0.30	0.79	1.73	0.21
dyMEAN	3.61	1.58	3.58	**0.28**	**0.07**	**0.28**	1.57	0.18
FlowDesign	2.92	1.94	1.68	1.44	1.18	1.43	1.77	0.67

perturbation experiments separately assess robustness when the antigen condition remains biologically matched.

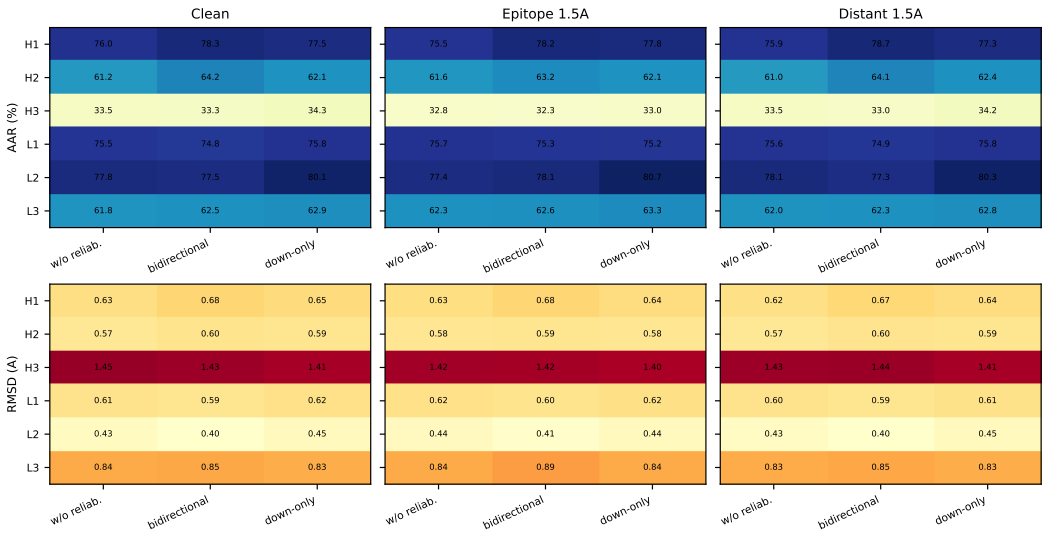

Figure 5: CDR-level ablation details. The three columns correspond to clean, epitope 1.5Å, and distant 1.5Å inputs. Variant names follow the main ablation table.

C.7 Complex-Level Paired Bootstrap Analysis

To quantify uncertainty in the interface-quality comparison, we perform a complex-level paired bootstrap over the four primary perturbed conditions (epitope 1.5Å, epitope 3.0Å, distant 1.5Å, and distant 3.0Å). For each complex, we first average DockQ over its generated candidates and then average the paired differences across the four conditions. We resample the common test complexes 10,000 times and report percentile 95% confidence intervals for TrustCDR minus each baseline.

All intervals are above zero, indicating a stable absolute DockQ advantage under the evaluated perturbation conditions. This analysis measures the uncertainty of the absolute performance difference; it does not establish that TrustCDR has a smaller perturbation-induced degradation than every baseline.

Table 12: Detailed single-channel ablation results under clean, epitope 1.5Å, and distant 1.5Å conditions. All values, including overall means, are truncated to two decimal places to match the main-text formatting.

Condition	Variant	AAR (%)							DockQ
		H1	H2	H3	L1	L2	L3	Mean	
Clean	Coordinate-only reliability	80.16	60.57	33.45	75.05	79.13	63.36	65.28	0.77
	Topology-only reliability	80.38	59.71	34.49	75.33	79.09	62.87	65.31	0.76
	Orientation-only reliability	80.32	60.11	35.03	75.25	78.60	63.22	65.42	0.77
	Surface-only reliability	80.87	59.94	35.42	75.01	78.11	62.40	65.29	0.77
Epitope 1.5Å	Coordinate-only reliability	78.17	62.64	35.61	77.35	79.59	64.43	66.29	0.71
	Topology-only reliability	78.25	62.13	35.31	77.19	80.57	64.63	66.34	0.71
	Orientation-only reliability	78.17	63.16	35.52	76.95	79.59	65.64	66.50	0.71
	Surface-only reliability	78.20	61.21	35.88	76.70	81.31	64.63	66.32	0.71
Distant 1.5Å	Coordinate-only reliability	79.45	60.00	33.51	75.32	79.34	63.83	65.24	0.75
	Topology-only reliability	80.57	60.69	34.27	75.17	78.36	62.40	65.24	0.75
	Orientation-only reliability	80.32	59.77	34.12	75.67	78.85	63.03	65.29	0.75
	Surface-only reliability	80.38	60.23	34.34	75.09	79.34	61.86	65.20	0.76

Condition	Variant	RMSD (Å)						
		H1	H2	H3	L1	L2	L3	Mean
Clean	Coordinate-only reliability	0.58	0.54	1.30	0.55	0.38	0.82	0.70
	Topology-only reliability	0.61	0.52	1.39	0.55	0.39	0.79	0.71
	Orientation-only reliability	0.60	0.51	1.36	0.55	0.39	0.80	0.70
	Surface-only reliability	0.60	0.52	1.31	0.55	0.39	0.79	0.69
Epitope 1.5Å	Coordinate-only reliability	0.57	0.57	1.35	0.53	0.40	0.80	0.71
	Topology-only reliability	0.57	0.58	1.35	0.53	0.40	0.79	0.71
	Orientation-only reliability	0.55	0.57	1.29	0.53	0.41	0.83	0.70
	Surface-only reliability	0.56	0.58	1.34	0.53	0.40	0.80	0.71
Distant 1.5Å	Coordinate-only reliability	0.59	0.54	1.29	0.55	0.38	0.81	0.69
	Topology-only reliability	0.61	0.53	1.41	0.55	0.39	0.80	0.71
	Orientation-only reliability	0.60	0.52	1.39	0.55	0.38	0.80	0.71
	Surface-only reliability	0.59	0.53	1.37	0.55	0.39	0.79	0.70

Table 13: Sensitivity to attenuation strength β with $c = 0.20$ fixed under clean, 0.8Å, and 1.5Å conditions.

β	Condition	AAR (%)	RMSD (Å)	DockQ
0.05	clean	68.27	0.73	**0.77**
0.05	epitope 0.8Å	68.21	**0.72**	**0.75**
0.05	epitope 1.5Å	68.15	**0.72**	0.71
0.05	distant 0.8Å	68.27	**0.73**	**0.76**
0.05	distant 1.5Å	68.30	**0.73**	0.76
0.10	clean	**68.52**	**0.57**	0.76
0.10	epitope 0.8Å	**68.71**	0.73	**0.75**
0.10	epitope 1.5Å	**68.78**	0.73	**0.75**
0.10	distant 0.8Å	66.86	**0.73**	**0.76**
0.10	distant 1.5Å	**68.51**	**0.73**	**0.77**
0.20	clean	68.29	0.73	**0.77**
0.20	epitope 0.8Å	68.25	0.73	0.74
0.20	epitope 1.5Å	68.20	0.73	0.71
0.20	distant 0.8Å	68.30	**0.73**	**0.76**
0.20	distant 1.5Å	68.33	**0.73**	0.76
0.30	clean	68.29	0.73	**0.77**
0.30	epitope 0.8Å	68.20	0.73	0.74
0.30	epitope 1.5Å	68.17	0.74	0.71
0.30	distant 0.8Å	**68.32**	**0.73**	**0.76**
0.30	distant 1.5Å	68.31	**0.73**	0.76

Table 14: Antigen-condition diagnostic. AAR and RMSD are mean CDR metrics, and DockQ is Total DockQ. This diagnostic has no best-value marking.

Method	Metric	Clean	Translated	WrongAg
DiffAb	AAR (%)	45.31	45.60	45.30
DiffAb	RMSD (Å)	0.79	0.78	0.81
DiffAb	DockQ	0.72	0.78	0.79
dyAb	AAR (%)	27.12	27.10	27.10
dyAb	RMSD (Å)	1.15	1.17	1.16
dyAb	DockQ	0.26	0.53	0.25
dyMEAN	AAR (%)	65.47	66.70	66.70
dyMEAN	RMSD (Å)	0.68	0.66	0.66
dyMEAN	DockQ	0.34	0.73	0.65
FlowDesign	AAR (%)	48.84	49.30	49.30
FlowDesign	RMSD (Å)	2.17	2.10	2.14
FlowDesign	DockQ	0.59	0.68	0.69
TrustCDR	AAR (%)	68.52	5.80	44.50
TrustCDR	RMSD (Å)	0.57	3.65	2.09
TrustCDR	DockQ	0.76	0.24	0.67

Table 15: Complex-level paired bootstrap for DockQ differences across the four primary perturbed conditions.

Baseline	n	ΔDockQ (95% CI)
DiffAb	58	$+0.039$ $[+0.032, +0.046]$
FlowDesign	57	$+0.163$ $[+0.155, +0.171]$
dyMEAN	58	$+0.409$ $[+0.383, +0.433]$
dyAb	55	$+0.493$ $[+0.479, +0.505]$