ConTact: Contact-First Antibody CDR Design via Explicit Interface Reasoning

Anonymous Authors[1]

Abstract

Computational antibody CDR design methods condition on antigen structure to generate binding loops, yet existing architectures conflate two fundamentally distinct sub-problems: identifying which CDR positions will contact the antigen, and selecting amino acids at those positions. This conflation forces models to learn contact reasoning implicitly through uniform message passing, diluting antigen signal across all positions equally. We introduce CONTACT, a contact-then-act architecture that explicitly decomposes CDR design into three cascaded stages: learning surface complementarity fingerprints, predicting CDR-antigen contacts, and injecting contact-gated antigen features into the sequence head. A distance-biased cross-attention module encodes geometric priors favoring spatial neighbors, while a contact-weighted cross-entropy loss concentrates gradient signal on binding-critical positions. On CHIMERA-BENCH, CONTACT achieves the best structural quality (7% RMSD improvement over the next-best baseline), best epitope awareness (10% F1 score over GNN baselines), and competitive sequence recovery (AAR 0.38) among several CDR-H3 design baselines.

1. Introduction

Antibodies bind antigens through their complementarity-determining regions (CDRs), six hypervariable loops whose sequence and structure determine binding specificity (Chothia & Lesk, 1987). Computational CDR design methods condition on antigen structure to generate sequences and backbone conformations for these loops (Luo et al., 2022; Kong et al., 2023a;b; Wu et al., 2025b). Yet a growing body of evidence shows that existing methods largely fail to leverage antigen information. Predictions remain nearly unchanged when the antigen is removed (Li et al., 2025), and BLOSUM substitution matrices explain model outputs as well as learned likelihoods (Uçar & Sormanni, 2025; Chinery et al., 2024).

We argue that a fundamental cause is architectural: current methods conflate two distinct sub-problems into a single prediction head. The first sub-problem is *where* the CDR will contact the antigen, i.e., which CDR positions form binding interactions. The second is *what* amino acids to place at those positions, given the local chemistry of the binding partner. Equivariant GNNs such as MEAN (Kong et al., 2023a) and RAAD (Wu et al., 2025b) propagate antigen information through uniform message passing that treats all antigen residues equivalently. Diffusion-based methods like DiffAb (Luo et al., 2022) concatenate antibody and antigen residues into a flat graph with only a fragment-type embedding to distinguish them. Even dyMEAN (Kong et al., 2023b), which uses a shadow paratope mechanism and edge distance prediction for contact-aware graph construction, does not use predicted contacts to modulate sequence prediction. In all cases, the model must simultaneously discover which positions are binding-relevant and what residues belong there, with a uniform cross-entropy loss that allocates equal learning capacity to every position.

The CDR-antigen interface is inherently sparse. A CDR-H3 of length 10–25 typically forms only 5–15 contacts with the antigen, and the amino acid identity at contact positions is directly constrained by the chemistry of the binding partner: hydrophobic pockets select for complementary hydrophobic CDR residues, while charged patches favor oppositely charged side chains. Non-contact positions are primarily constrained by backbone geometry and loop stability. Treating these two classes of positions equally wastes learning capacity on the less informative non-contact positions.

We propose CONTACT, a contact-first architecture that decomposes CDR design into three explicit stages, addressing the *where* before the *what*. First, the model learns surface complementarity fingerprints that characterize the local binding environment at each CDR position, inspired by molecular surface fingerprints (Gainza et al., 2020; 2023). Second, it predicts which CDR positions will contact the antigen using a supervised contact predictor. Third, it selectively injects local antigen features into the CDR repre-

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Submitted to the 2026 Workshop on Generative and Agentic AI for Biology (ICML 2026). Do not distribute.

sentation, gated by the predicted contact confidence, so that antigen information flows preferentially to binding-critical positions. A distance-biased cross-attention module provides geometric inductive bias by favoring spatial neighbors, and a contact-weighted cross-entropy loss concentrates gradient signal on positions the model identifies as contacts.

Our contributions are:

1. We identify the conflation of contact identification and sequence prediction as a structural limitation of existing CDR design architectures, and propose the *contact-first* design paradigm that decomposes these sub-problems into an explicit three-stage cascade.

2. We introduce a contact-gated injection mechanism with double gating (learned gate $\times$ contact confidence) that selectively routes antigen information to binding-relevant CDR positions, preventing noise from distant antigen residues at non-contact positions.

3. We demonstrate on CHIMERA-BENCH that CONTACT achieves the best RMSD (1.63 Å, 7% over next-best), epitope F1 (0.79, 10% over GNN baselines), fnat (0.67), and AAR (0.38) among eleven baselines.

2. Related Work

Equivariant GNN methods. MEAN (Kong et al., 2023a) introduced multi-channel equivariant attention with alternating intra-segment and inter-segment layers for CDR design. dyMEAN (Kong et al., 2023b) extended this with a shadow paratope mechanism that predicts inter-chain edge distances for dynamic graph construction, making it the closest existing work to contact-aware CDR design. RAAD (Wu et al., 2025b) defines eight relation-aware edge types with Bernoulli edge sampling and a contrastive specificity loss applied at test-time optimization. CONTACT differs from all three in that it uses predicted contacts to directly modulate the sequence prediction head through gated injection and position-specific loss weighting, rather than using contact-related information solely for graph topology (dyMEAN) or test-time optimization (RAAD).

Diffusion and flow methods. DiffAb (Luo et al., 2022) models CDR generation as a joint diffusion process over coordinates, orientations, and amino acid types. AbFlowNet (Abir et al., 2025) extends this with flow matching and trajectory balance loss. AbMEGD (Chen et al., 2025) and RADAb (Wang et al., 2024) add retrieval-augmented and multi-expert components. dyAb (Tan et al., 2025) applies flow matching with structure relaxation. FlowDesign (Wu et al., 2025a) follows a diagnose-then-fix approach, identifying that standard Gaussian priors are poorly suited for CDR generation and replacing them with data-driven prior distributions. All these methods treat antigen conditioning as a flat concatenation of antibody and antigen residues with fragment-type embeddings, applying uniform attention without distinguishing contact from non-contact positions. CONTACT addresses a complementary limitation: not the prior distribution, but the conditioning mechanism itself.

Antigen conditioning failures. Multiple studies have documented that existing CDR design methods fail to effectively use antigen information. Li et al. (2025) showed that predictions remain nearly unchanged when the antigen is removed. Uçar & Sormanni (2025) demonstrated that BLOSUM substitution matrices explain model outputs as well as learned likelihoods. Chinery et al. (2024) found that simple computational methods can outperform deep learning in generating diverse, binder-enriched antibody libraries. RefineGNN (Jin et al., 2022b), which receives no antigen input, achieves the second-best binding metrics on CHIMERA-BENCH, further corroborating this failure. The contact-first decomposition in CONTACT directly targets this problem by providing an explicit, supervised pathway for antigen information to reach the sequence head.

Predict-then-design paradigms. The idea of predicting binding-relevant features before designing sequences has precedent in broader protein design. MaSIF-seed (Gainza et al., 2023) predicts favorable interaction sites on molecular surfaces using learned surface fingerprints, then designs binders targeting those sites. RFdiffusion (Watson et al., 2023) generates protein backbones first, then designs sequences with ProteinMPNN. CONTACT applies a similar predict-then-design strategy at the residue-contact level: predict which CDR positions will contact the antigen, then condition sequence design on those predictions. Unlike MaSIF-seed, which operates on molecular surfaces in a separate pipeline, CONTACT performs contact prediction and sequence design end-to-end within a single differentiable architecture.

3. Preliminaries

3.1. Task Definition

We adopt the formulation from CHIMERA-BENCH (Ahmed et al., 2026). Given an antigen structure $A = \{(s_j, \mathbf{x}_j) \mid j \in V_A\}$, an epitope specification $E \subseteq V_A$, and an antibody framework $F = \{(s_i, \mathbf{x}_i) \mid i \in V_{\text{FR}}\}$, the task is to design CDR residues $R = \{(s_k, \mathbf{x}_k) \mid k \in V_{\text{CDR}}\}$ that maximize the conditional likelihood subject to epitope contact constraints:

$$R^* = \arg\max_R \; p_\theta\big(R \mid A, E, F\big), \quad \text{s.t. } \mathcal{C}(R, A) \neq \emptyset \quad (1)$$

where each residue has amino acid type $s_k \in \{1, \ldots, 20\}$ and Cα coordinate $\mathbf{x}_k \in \mathbb{R}^3$. We denote by $\mathcal{C}(R, A) = \{j \in V_A \mid \exists\, k \in V_{\text{CDR}} : \|\mathbf{x}_k - \mathbf{x}_j\| < d_c\}$ the set of antigen residues contacted within cutoff d_c. We focus on CDR-H3, the most variable loop and primary determinant of antigen specificity (Chothia & Lesk, 1987).

3.2. Graph Construction

We represent the antibody-antigen complex as a heterogeneous graph $\mathcal{G} = (V, \mathcal{E})$. The node set $V = V_{\text{HC}} \cup V_{\text{LC}} \cup V_A \cup V_{\text{glob}} \cup V_{\text{vn}}$ contains residue nodes from the heavy chain (V_{HC}), light chain (V_{LC}), and antigen (V_A), three global delimiter tokens ($V_{\text{glob}} = \{\text{BOH}, \text{BOL}, \text{BOA}\}$), and N_{vn} virtual nodes (Sestak et al., 2026). Each residue node i carries amino acid type $s_i \in \{1, \ldots, 20\}$ and four backbone atom coordinates $\mathbf{X}_i = [\mathbf{x}_i^{\text{N}}, \mathbf{x}_i^{\text{C}\alpha}, \mathbf{x}_i^{\text{C}}, \mathbf{x}_i^{\text{O}}] \in \mathbb{R}^{4\times 3}$.

The edge set $\mathcal{E}$ is partitioned into 10 typed subsets that capture different structural relationships. Within each chain, we construct *radial edges* connecting all pairs within a Cα distance cutoff, *sequential edges* linking residues separated by one or two positions in primary sequence, and *KNN edges* connecting each residue to its nearest spatial neighbors. Across chains, we add *inter-chain radial edges* and *inter-chain KNN edges* that enable direct communication between antibody and antigen residues. Three *global-to-chain edges* connect the delimiter tokens to their respective chains. Two *virtual node edge types* connect each virtual node bidirectionally to all epitope and all CDR residues. This creates a two-hop shortcut between epitope and CDR, directly addressing the over-squashing problem (Alon & Yahav, 2021) where information from distant epitope residues dilutes through many layers of sequential message passing.

Each edge (i, j) carries a feature vector $\mathbf{e}_{ij}$ encoding edge type (one-hot), relative position in local coordinate frames, pairwise distance RBFs between backbone atom pairs, a quaternion encoding of relative backbone orientation, and local frame direction features. Virtual node edges use learnable feature vectors rather than geometric features.

3.3. Motivation: Contact-First Decomposition

Existing CDR design methods process antigen information through spatial message passing or cross-attention, but none separates the problem of *identifying contacts* from the problem of *designing residues at contacts*. MEAN (Kong et al., 2023a) alternates intra-segment and inter-segment equivariant attention layers, attending uniformly to all antigen residues. RAAD (Wu et al., 2025b) defines eight relation-aware edge types and uses Bernoulli edge sampling over antigen connections, but its contrastive specificity loss operates only at test-time optimization, not during training. dyMEAN (Kong et al., 2023b) introduces a shadow paratope that predicts inter-chain edge distances for graph construction, making it the closest to contact-aware, but these distances inform graph topology rather than the sequence prediction head. DiffAb (Luo et al., 2022) and AbFlowNet (Abir et al., 2025) concatenate all residues into a flat graph with a fragment-type embedding, applying uniform geometric attention. RefineGNN (Jin et al., 2022b) generates CDRs autoregressively without any antigen input, yet achieves surprisingly strong binding metrics, further highlighting the failure of existing conditioning approaches.

All these methods apply a uniform cross-entropy loss that treats every CDR position equally, whether it contacts the antigen or not. This is suboptimal because amino acid identity at contact positions is directly constrained by the chemistry of the binding partner, while non-contact positions are primarily constrained by backbone geometry and loop stability.

We formalize the alternative as a design principle: *contact prediction should precede sequence prediction*. If the model first identifies which CDR positions will contact the antigen, it can selectively route antigen information to those positions and concentrate learning capacity there. This decomposition draws on the broader paradigm of predict-then-design in structural biology (Gainza et al., 2023; Watson et al., 2023), but applies it at the residue-contact level within a single end-to-end architecture rather than as a separate pipeline.

3.4. Contact Definition

We define a CDR residue k as contacting the antigen if its Cα atom lies within 8 Å of any antigen Cα atom:

$$c_k = \mathbb{1}\left[\min_{j \in V_A} \|\mathbf{x}_k - \mathbf{x}_j\| < 8 \text{ Å}\right] \tag{2}$$

This threshold matches the symmetric contact definition used in the CHIMERA-BENCH evaluation metrics (fnat, iRMSD, DockQ). The binary labels $c_k \in \{0, 1\}$ serve as supervision for the contact prediction stage and as weights in the contact-weighted sequence loss.

4. Method

CONTACT consists of three components: (i) a VirtualNode-EGNN encoder that performs E(3)-equivariant message passing over the heterogeneous graph, (ii) a distance-biased cross-attention module that combines CDR and antigen representations with spatial priors, and (iii) a three-stage contact-first decoder that cascades complementarity fingerprinting, contact prediction, and contact-guided sequence generation. Figure 1 illustrates the full pipeline.

4.1. Feature Encoding

Each residue i in the antibody-antigen complex is represented by a feature vector $\mathbf{f}_i$ composed of five groups.

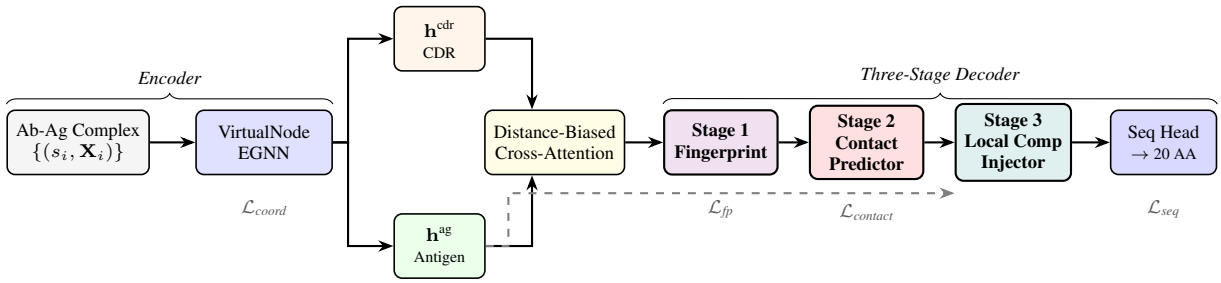

Figure 1. **ConTact architecture.** The encoder maps residue features through a VirtualNode-EGNN to produce per-residue embeddings and updated coordinates. CDR and antigen embeddings are combined via distance-biased cross-attention. The three-stage decoder cascades complementarity fingerprinting, contact prediction, and contact-guided local complementarity injection. Each stage conditions on the previous stage's output. The contact-weighted sequence loss $\mathcal{L}_{\text{seq}}$ up-weights positions identified as contacts by Stage 2. Dashed arrows indicate antigen features flowing directly to Stage 3 for local aggregation.

Amino acid identity: a one-hot encoding of the residue type, masked to the zero vector for all CDR positions during training to prevent trivial teacher-forcing. **Backbone distance RBFs**: intraresidue bond lengths (N–Cα, Cα–C, C–O) each expanded into Gaussian basis functions:

$$\phi_{\text{rbf}}(d)_m = \exp\left(-\frac{(d-\mu_m)^2}{2\varsigma^2}\right), \quad m = 1, \ldots, M \tag{3}$$

where μ_m are uniformly spaced centers and ς is the basis width. **Backbone angles**: bond angles and dihedral angles (ϕ, ψ, ω), each encoded as sine-cosine pairs. **Local frame directions**: unit vectors along the three local coordinate axes defined by the N-Cα-C backbone triangle. **Sinusoidal position embedding**: encoding of the residue index within its chain at multiple frequency scales.

A segment type indicator distinguishes heavy chain, light chain, and antigen residues. A dual-path MLP processes geometric and chemical features through separate pathways with SiLU activations, fuses the outputs, and projects to embedding dimension d:

$$\mathbf{h}_i^{(0)} = \text{MLP}_{\text{fuse}}\big([\text{MLP}_{\text{geom}}(\mathbf{f}_i^{\text{geom}}),\ \text{MLP}_{\text{chem}}(\mathbf{f}_i^{\text{chem}})]\big) \tag{4}$$

Epitope residues (those in E) receive an additional learnable embedding $\mathbf{e}_{\text{epi}}$ added to their representation, providing an explicit signal that these residues are part of the designated binding site.

4.2. VirtualNode-EGNN Encoder

The encoder applies multiple relation-aware E(3)-equivariant GNN layers (Satorras et al., 2021) on graph $\mathcal{G}$. Virtual nodes with learnable feature vectors and learnable coordinates participate in message passing through the VN-to-epitope and VN-to-CDR edge types. Because each virtual node connects to both all epitope residues and all CDR residues, information flows from any epitope residue to any CDR residue in exactly two message-passing steps. Without virtual nodes, this information must traverse the graph via sequential edges, suffering from over-squashing (Alon & Yahav, 2021) at bottleneck residues.

Each layer l updates node features and coordinates simultaneously. For edge (i, j) of type t, the message function takes the concatenation of sender and receiver embeddings, an outer product geometry term, and the edge features:

$$\mathbf{m}_{ij}^{(l)} = \text{MLP}_{\text{msg}}^{(l)}\Big([\mathbf{h}_i^{(l)},\ \mathbf{h}_j^{(l)},\ \text{vec}\big(\Delta\mathbf{x}_{ij}(\Delta\mathbf{x}_{ij})^\top\big),\ \mathbf{e}_{ij}]\Big) \tag{5}$$

where $\Delta\mathbf{x}_{ij} = \mathbf{x}_i^{(l)} - \mathbf{x}_j^{(l)}$ and $\text{vec}(\cdot)$ flattens the 3×3 outer product matrix into a 9-dimensional vector. The entries of $\Delta\mathbf{x}_{ij}(\Delta\mathbf{x}_{ij})^\top$ are dot products of displacement components, which are invariant to rotations, translations, and reflections.

The model aggregates messages from all edge types with type-specific linear projections and updates node features via a residual connection:

$$\mathbf{h}_i^{(l+1)} = \mathbf{h}_i^{(l)} + \text{MLP}_{\text{node}}^{(l)}\left(\left[\mathbf{h}_i^{(l)},\ \sum_t \mathbf{W}_t^{(l)} \sum_{j \in \mathcal{N}_t(i)} \mathbf{m}_{ij}^{(l)}\right]\right) \tag{6}$$

where $\mathbf{W}_t^{(l)}$ is a type-specific projection matrix and $\mathcal{N}_t(i)$ denotes the neighbors of node i under edge type t. Coordinates are updated equivariantly by adding a weighted sum of displacement vectors:

$$\mathbf{x}_i^{(l+1)} = \mathbf{x}_i^{(l)} + \sum_t \frac{1}{|\mathcal{N}_t(i)|} \sum_{j \in \mathcal{N}_t(i)} \Delta\mathbf{x}_{ij} \cdot \text{MLP}_t^{\text{coord},(l)}(\mathbf{m}_{ij}^{(l)}) \tag{7}$$

where $\text{MLP}_t^{\text{coord},(l)}$ produces a scalar weight. The product of a displacement vector and a scalar computed from invariant inputs is equivariant by construction. After all layers, the encoder produces residue embeddings $\mathbf{h} \in \mathbb{R}^{N \times D}$ and updated backbone coordinates $\hat{\mathbf{X}}$.

4.3. Distance-Biased Cross-Attention

After encoding, we extract CDR node embeddings $\mathbf{H}_{\text{cdr}} \in \mathbb{R}^{L\times D}$ and antigen node embeddings $\mathbf{H}_{\text{ag}} \in \mathbb{R}^{M\times D}$. Standard cross-attention computes alignment scores purely from learned feature similarity, making no distinction between an antigen residue 5 Å from the CDR and one 30 Å away. CONTACT adds a Gaussian spatial bias that encodes a geometric inductive bias, since binding contacts are necessarily spatial neighbors.

We project queries $\mathbf{Q} = \mathbf{H}_{\text{cdr}}\mathbf{W}_Q$ and keys $\mathbf{K} = \mathbf{H}_{\text{ag}}\mathbf{W}_K$ into H attention heads. The attention score between CDR position i and antigen position j is:

$$\alpha_{ij}^{(h)} = \text{softmax}_j\left(\frac{(\mathbf{q}_i^{(h)})^\top \mathbf{k}_j^{(h)}}{\sqrt{D_h}} + \beta_{ij}\right) \tag{8}$$

where the distance bias β_{ij} decays as a Gaussian function of the Cα–Cα distance between the updated coordinates from the encoder:

$$\beta_{ij} = \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right), \quad d_{ij} = \|\hat{\mathbf{x}}_i^{\text{C}\alpha} - \hat{\mathbf{x}}_j^{\text{C}\alpha}\|_2 \tag{9}$$

with bandwidth σ. The Gaussian decay is near 1.0 for residues within van der Waals contact distance, drops to $e^{-2} \approx 0.14$ at 2σ (approximately the contact threshold), and becomes negligible beyond 3σ. The bias is shared across heads to provide a consistent spatial prior, while the learned query-key projections specialize to different aspects of the binding interaction.

The output for each CDR position concatenates the multi-head weighted sum with the original CDR embedding:

$$\mathbf{o}_i = \left[\mathbf{h}_i^{\text{cdr}},\ \Big\|_{h=1}^{H} \sum_{j=1}^{M} \alpha_{ij}^{(h)} \mathbf{v}_j^{(h)}\right] \tag{10}$$

where $\|$ denotes concatenation and $\mathbf{v}_j^{(h)} = \mathbf{h}_j^{\text{ag}}\mathbf{W}_V^{(h)}$ are value projections. The skip connection preserves the CDR embedding for downstream stages that primarily need structural context.

4.4. Decoder

STAGE 1: COMPLEMENTARITY FINGERPRINTING

The first decoder stage compresses the cross-attention output into a compact representation of the local surface complementarity at each CDR position. Binding interactions follow chemical complementarity patterns (hydrophobic-hydrophobic, charge-charge, donor-acceptor) that can be captured in a low-dimensional fingerprint, analogous to molecular fingerprints in cheminformatics. Given the cross-attention output $\mathbf{o}_i$ for CDR position i, an MLP produces a fingerprint vector:

$$\mathbf{f}_i = \text{MLP}_{\text{fp}}(\mathbf{o}_i) \tag{11}$$

We train the fingerprint with a contrastive loss that enforces structural consistency. CDR positions that face similar local antigen environments should have similar fingerprints, while positions facing dissimilar environments should have distinct fingerprints. We define ground-truth similarity between two CDR positions i and j (potentially from different complexes in the batch) based on the cosine similarity of their true local environment descriptors, computed from the 3D arrangement and amino acid composition of the nearest antigen residues around each position. Positive pairs $\mathcal{P}$ are those exceeding a similarity threshold. The loss follows an InfoNCE formulation:

$$\mathcal{L}_{\text{fp}} = -\frac{1}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} \log \frac{\exp(\mathbf{f}_i^\top \mathbf{f}_j/\tau_{\text{fp}})}{\sum_{k\in\mathcal{B}} \exp(\mathbf{f}_i^\top \mathbf{f}_k/\tau_{\text{fp}})} \tag{12}$$

where $\mathcal{B}$ is the set of all CDR positions in the batch and τ_{fp} is a temperature parameter. The contrastive objective ensures that the fingerprint captures *what kind of binding environment* a CDR position faces, conditioning the subsequent contact prediction stage.

STAGE 2: CONTACT PREDICTION

The second stage predicts which CDR positions will form contacts with the antigen. This is the central component of the contact-first decomposition. Existing methods leave contact identification as an implicit byproduct of message passing (Kong et al., 2023a; Wu et al., 2025b) or graph construction (Kong et al., 2023b). CONTACT supervises contact prediction explicitly and uses the predictions to gate downstream information flow.

For each CDR position i, we aggregate features from its K nearest antigen neighbors (by Cα distance) using distance-weighted pooling:

$$\mathbf{a}_i = \frac{1}{K} \sum_{j\in\text{KNN}_K(i)} \phi_{\text{rbf}}(d_{ij}) \odot \mathbf{h}_j^{\text{ag}} \tag{13}$$

where $\phi_{\text{rbf}}(d_{ij})$ is a learned linear projection of the RBF encoding broadcast to match the antigen embedding dimension, and $\odot$ denotes elementwise multiplication. The contact predictor takes a concatenation of four inputs: the CDR embedding, the KNN-aggregated antigen features, an RBF encoding of the minimum distance to any antigen residue, and the complementarity fingerprint from Stage 1:

$$\hat{c}_i = \sigma\big(\text{MLP}_{\text{ct}}\big([\mathbf{h}_i^{\text{cdr}},\ \mathbf{a}_i,\ \phi_{\text{rbf}}(d_i^{\text{min}}),\ \mathbf{f}_i]\big)\big) \tag{14}$$

where $d_i^{\text{min}} = \min_{j\in V_A} \|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j\|$ is the minimum Cα distance to any antigen residue and σ denotes the sigmoid

function. Including the fingerprint $\mathbf{f}_i$ from Stage 1 creates a cascaded dependency, so that the quality of contact prediction depends on the learned complementarity representation.

The contact predictor outputs a soft probability $\hat{c}_i \in [0, 1]$ rather than a hard binary decision. The sigmoid output already provides smooth gradients, and hard thresholding introduced training instability in preliminary experiments. We supervise the contact predictor with a focal binary cross-entropy loss (Lin et al., 2017) that addresses the inherent class imbalance, where non-contact positions typically outnumber contacts by 3–5×:

$$\mathcal{L}_{\text{contact}} = -\frac{1}{L}\sum_{i=1}^{L}(1-\hat{p}_i)^{\gamma}\left[c_i \log \hat{c}_i + (1-c_i)\log(1-\hat{c}_i)\right] \tag{15}$$

where $c_i \in \{0, 1\}$ is the ground-truth contact label (Equation (2)), $\hat{p}_i$ denotes the predicted probability of the correct class, and γ is the focusing parameter. The factor $(1-\hat{p}_i)^{\gamma}$ down-weights well-classified examples, concentrating the learning signal on hard, ambiguous positions near the contact boundary.

Stage 3: Contact-Guided Local Complementarity Injection

The third stage uses the predicted contact confidence $\hat{c}_i$ from Stage 2 to selectively inject local antigen information into the CDR embeddings. Antigen features should influence CDR representations primarily at positions the model predicts will form contacts, while non-contact positions should rely mainly on their backbone geometry context.

For each CDR position i, we aggregate features from K-nearest antigen neighbors:

$$\mathbf{h}_i^{\text{local}} = \frac{1}{K}\sum_{j \in \text{KNN}_K(i)} \mathbf{h}_j^{\text{ag}} \tag{16}$$

A learned gate modulates the injection magnitude based on the CDR embedding and the contact prediction:

$$g_i = \sigma\left(\mathbf{w}_g^{\top}[\mathbf{h}_i^{\text{cdr}},\ \hat{c}_i] + b_g\right) \tag{17}$$

The enriched CDR embedding combines the original representation with the gated antigen information:

$$\mathbf{h}_i^{\text{enriched}} = \mathbf{h}_i^{\text{cdr}} + g_i \cdot \hat{c}_i \cdot \text{MLP}_{\text{proj}}(\mathbf{h}_i^{\text{local}}) \tag{18}$$

The product $g_i \cdot \hat{c}_i$ creates a double gating mechanism. The contact confidence $\hat{c}_i$ from Stage 2 provides a data-driven prior: at non-contact positions ($\hat{c}_i \approx 0$), the antigen contribution is effectively zeroed out regardless of the learned gate, preventing noise from distant antigen residues. The learned gate g_i provides fine-grained control, allowing the model to modulate injection magnitude even at contact positions based on local structural context.

The final representation for the sequence head concatenates the enriched embedding with the contact-masked cross-attention output:

$$\mathbf{z}_i = [\mathbf{h}_i^{\text{enriched}},\ \hat{c}_i \cdot \mathbf{o}_i^{\text{attn}}] \tag{19}$$

where $\mathbf{o}_i^{\text{attn}}$ is the attention output from Section 4.3. Multiplying by $\hat{c}_i$ further suppresses the attention-derived antigen information at non-contact positions.

Contact-Weighted Sequence Head

The sequence head maps the final representation $\mathbf{z}_i$ to amino acid logits $\boldsymbol{\ell}_i \in \mathbb{R}^{20}$ via an MLP. Rather than standard per-position cross-entropy, we apply a contact-weighted variant that allocates more learning capacity to positions predicted to form binding contacts:

$$\mathcal{L}_{\text{seq}} = -\frac{1}{L}\sum_{i=1}^{L} w_i \log \frac{\exp(\ell_i^{y_i})}{\sum_{a=1}^{20}\exp(\ell_i^{a})} \tag{20}$$

where y_i is the ground-truth amino acid at position i and the position-specific weight is:

$$w_i = 1 + \alpha \cdot \hat{c}_i \tag{21}$$

The hyperparameter α controls the relative up-weighting of contact positions. This reweighting follows from the observation that standard cross-entropy distributes learning capacity uniformly, treating a non-contact glycine at the loop apex the same as a contact-forming tryptophan buried in an antigen pocket. By up-weighting contacts, the model receives stronger gradient signal at precisely the positions where amino acid identity is most constrained by the antigen.

At inference, the predicted amino acid at each position is $\hat{s}_i = \arg\max_a \ell_i^a$. The contact predictions $\hat{c}_i$ can also be inspected to verify which positions the model believes form contacts.

4.5. Training Objective

The full training objective combines seven loss terms:

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\text{seq}} + \lambda_{\text{coord}}\mathcal{L}_{\text{coord}} + \lambda_{\text{contact}}\mathcal{L}_{\text{contact}} + \lambda_{\text{fp}}\mathcal{L}_{\text{fp}} \\ + \lambda_{\text{pair}}\mathcal{L}_{\text{pair}} + \lambda_{\text{dock}}\mathcal{L}_{\text{dock}} + \lambda_{\text{aux}}\mathcal{L}_{\text{aux}}\end{aligned} \tag{22}$$

The coordinate loss $\mathcal{L}_{\text{coord}}$ is a smooth-ℓ_1 (Huber) loss on predicted versus true Cα coordinates for CDR positions:

$$\mathcal{L}_{\text{coord}} = \frac{1}{L}\sum_{k \in V_{\text{CDR}}} \text{smooth}_{\ell_1}\left(\hat{\mathbf{x}}_k^{\text{C}\alpha} - \mathbf{x}_k^{\text{C}\alpha,\text{true}}\right) \tag{23}$$

The pairing loss $\mathcal{L}_{\text{pair}}$ is an InfoNCE contrastive loss that matches mean-pooled CDR and antigen embeddings within

the batch, treating cognate pairs as positives:

$$\mathcal{L}_{\text{pair}} = -\frac{1}{B}\sum_{i=1}^{B}\log\frac{\exp(\bar{\mathbf{h}}_i^{\text{cdr}}\cdot\bar{\mathbf{h}}_i^{\text{ag}}/\tau_p)}{\sum_{k=1}^{B}\exp(\bar{\mathbf{h}}_i^{\text{cdr}}\cdot\bar{\mathbf{h}}_k^{\text{ag}}/\tau_p)} \tag{24}$$

The docking loss $\mathcal{L}_{\text{dock}}$ penalizes the minimum predicted Cα distance from each CDR residue to epitope atoms when it exceeds a cutoff, encouraging the predicted backbone to dock near the epitope. The auxiliary loss $\mathcal{L}_{\text{aux}}$ is a CDR feature reconstruction regularizer that prevents representation collapse. All loss weights λ are determined by hyperparameter sweeps using Weights & Biases (W&B).

5. Experiments

5.1. Setup

Dataset and metrics. We evaluate on CHIMERA-BENCH (Ahmed et al., 2026), comprising 2,922 antibody-antigen complexes with the epitope-group split (2,338/292/292 train/val/test). We report CDR-H3 results across eight metrics. Sequence quality is measured by amino acid recovery (AAR), contact AAR (CAAR, restricted to positions within 8 Å of the antigen), and perplexity (PPL). Structural quality is measured by Cα RMSD. Binding quality is measured by fraction of native contacts (fnat), interface RMSD (iRMSD), DockQ (Basu & Wallner, 2016), and epitope F1. All interface metrics use symmetric Cα–Cα contacts at 8 Å restricted to CDR residues.

Baselines. We compare against eleven baselines spanning four architectural families: equivariant GNNs (RAAD (Wu et al., 2025b), MEAN (Kong et al., 2023a), dyMEAN (Kong et al., 2023b)), diffusion and flow models (DiffAb (Luo et al., 2022), AbFlowNet (Abir et al., 2025), AbMEGD (Chen et al., 2025), RADAb (Wang et al., 2024), dyAb (Tan et al., 2025)), ODE (AbODE (Verma et al., 2023)), and autoregressive (RefineGNN (Jin et al., 2022b), AbDockGen (Jin et al., 2022a)). All models are retrained on CHIMERA-BENCH with their original hyperparameters.

Implementation details. CONTACT has 9.68M trainable parameters. The feature encoder projects 108D input features to 32D embeddings. The VN-EGNN encoder uses 4 layers with 256D hidden features and 3 virtual nodes. The cross-attention module uses 2 heads with bandwidth $\sigma = 4$ Å. The complementarity fingerprint is 32-dimensional. The contact predictor MLP has two hidden layers of 128 units. The contact weight $\alpha = 4.47$ and focal loss $\gamma = 2$. Loss weights are $\lambda_{\text{coord}} = 0.598$, $\lambda_{\text{contact}} = 1.763$, $\lambda_{\text{fp}} = 0.020$, $\lambda_{\text{pair}} = 0.103$, $\lambda_{\text{dock}} = 0.233$, $\lambda_{\text{aux}} = 0.200$, determined by Weights & Biases sweep. The notably large weight on λ_{contact} reflects the importance of accurate contact prediction for downstream sequence quality. We train with Adam (lr $= 6.31 \times 10^{-4}$, exponential decay $\gamma_{\text{lr}} = 0.944$ per epoch), gradient clipping at 0.5, batch size 8, dropout 0.1, and early stopping with patience 10 on validation loss. Training takes approximately 1.6 hours on a single NVIDIA H100 80GB GPU.

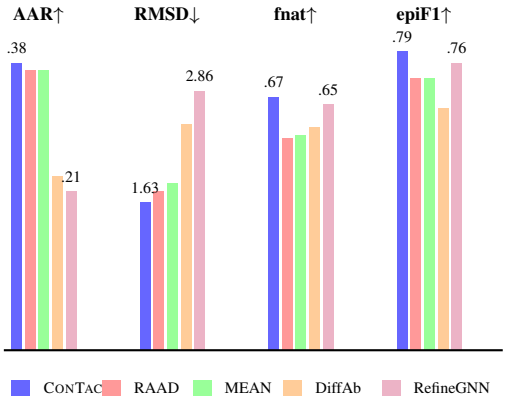

Figure 2. Comparison of CONTACT against representative baselines on four key metrics. CONTACT achieves the best AAR, lowest RMSD, highest fnat, and highest epiF1. For RMSD (↓), shorter bars are better.

5.2. Main Results

5.3. Results and Discussion

Table 1 presents the full comparison. CONTACT achieves the best performance across structure, binding, and epitope metrics simultaneously. Figure 2 highlights three key metrics against representative baselines.

Structural quality. CONTACT achieves the lowest RMSD (1.63 Å) among all methods, improving over RAAD by 7% (1.63 vs 1.75 Å) and over MEAN by 11% (1.63 vs 1.84 Å). The contact-guided injection produces CDR backbones that are both geometrically accurate and properly positioned relative to the epitope. The docking loss contributes to this, but the contact prediction stage is the key differentiator: by identifying which positions will interact with the antigen, the model generates backbone conformations that better accommodate the binding geometry.

Binding quality. CONTACT achieves the highest fnat (0.67), lowest iRMSD (1.35 Å), and highest DockQ (0.73). RefineGNN achieves strong binding metrics (fnat 0.65, DockQ 0.73) despite receiving no antigen input, confirming that backbone geometry alone carries substantial information about interface contacts (Li et al., 2025). CONTACT matches or surpasses RefineGNN on all binding metrics while additionally conditioning on the antigen, as reflected in its substantially higher epitope F1.

Table 1. CDR-H3 design on CHIMERA-BENCH. Best in **bold**, second-best underlined.

	Method	AAR↑	CAAR↑	PPL↓	RMSD↓	fnat↑	iRMSD↓	DockQ↑	epiF1↑
GNN	RAAD	0.37±.12	0.21±.22	3.27±.48	1.75±.77	0.56±.30	1.48±.69	0.70±.15	0.72±.25
	MEAN	0.37±.13	**0.24**±.23	3.10±.47	1.84±.75	0.57±.31	1.53±.72	0.69±.15	0.72±.25
	dyMEAN	0.37±.13	0.22±.23	3.29±.40	2.22±.97	0.53±.31	1.95±.90	0.65±.15	0.64±.28
Diff./Flow	DiffAb	0.23±.12	0.14±.19	–	2.49±1.28	0.59±.31	2.15±.96	0.65±.16	0.64±.25
	AbFlowNet	0.23±.11	0.14±.18	–	2.38±1.22	0.60±.31	2.04±.95	0.66±.16	0.65±.25
	AbMEGD	0.21±.12	0.12±.16	–	2.44±1.29	0.56±.29	2.12±.95	0.64±.14	0.64±.25
	RADAb	0.20±.12	0.11±.17	–	5.33±17.2	0.49±.32	3.65±11.4	0.60±.17	0.60±.27
	dyAb	0.19±.08	0.09±.14	–	2.34±.87	0.14±.21	2.94±.95	0.45±.09	0.24±.31
ODE	AbODE	0.26±.12	0.20±.22	11.70±4.34	14.64±3.21	0.11±.21	4.37±2.18	0.37±.15	0.27±.25
AR	RefineGNN	0.21±.11	0.10±.14	8.46±3.28	2.86±.87	0.65±.28	1.42±.71	0.73±.14	0.76±.22
	AbDockGen	0.24±.12	0.12±.18	8.04±2.65	4.67±1.32	0.41±.26	3.02±1.22	0.55±.14	0.62±.22
	CONTACT (ours)	**0.38**±.13	0.20±.21	**3.28**±.47	**1.63**±.74	**0.67**±.28	**1.35**±.71	**0.73**±.19	**0.79**±.20

Epitope awareness. CONTACT achieves the best epitope F1 at 0.79, surpassing RefineGNN (0.76) by 4% and RAAD/MEAN (both 0.72) by 10%. This is the most direct validation of the contact-first approach. The explicit contact prediction stage forces the model to identify which antigen residues are binding-relevant, and the contact-gated injection transfers this awareness to the sequence head. Methods without explicit contact reasoning (RAAD, MEAN, dyMEAN) achieve substantially lower epitope F1, confirming that uniform message passing fails to capture epitope specificity.

Sequence recovery. CONTACT achieves an AAR of 0.38, the highest among all baselines. The three GNN baselines (RAAD, MEAN, dyMEAN) reach 0.37. Contact AAR (CAAR 0.20) remains comparable to most baselines (RAAD 0.21, MEAN 0.24), indicating that predicting the exact amino acid at contact positions remains fundamentally challenging across all current methods. The contact-weighted CE concentrates gradient signal on contact positions, but the underlying difficulty of predicting antigen-specific amino acid identity persists. This gap (AAR 0.38 vs CAAR 0.20) suggests the bottleneck is not learning capacity allocation but the information content of Cα-level antigen representations, which may not fully capture the chemical constraints imposed by antigen binding pockets.

The contact-first hypothesis. The results collectively validate the contact-first decomposition. CONTACT dominates on structural and binding metrics because the explicit contact prediction stage provides a supervised bridge between the antigen representation and the sequence head. Without this bridge, methods route antigen information through uniform message passing (MEAN, RAAD), fragment-type embeddings (DiffAb, AbFlowNet), or graph topology (dyMEAN), all of which provide coordinate-level guidance but fail to translate into position-specific sequence preferences. The RefineGNN comparison is particularly instructive: it achieves strong binding metrics from backbone geometry alone, demonstrating the baseline signal available without any antigen conditioning. CONTACT leverages this geometric signal through the VN-EGNN encoder while adding antigen conditioning through the three-stage decoder, achieving both the structural quality of geometry-only methods and the epitope awareness that requires explicit antigen reasoning.

6. Conclusion

In this paper, we propose CONTACT, which decomposes antibody CDR design into explicit contact prediction followed by contact-guided sequence generation. The three-stage cascade (complementarity fingerprinting, contact prediction, contact-gated injection) provides a supervised pathway for antigen information to reach the sequence head at binding-relevant positions. Experiments on CHIMERA-BENCH demonstrate that this contact-first decomposition achieves the best structural quality, epitope awareness, and composite binding scores among eleven baselines.

Contact AAR remains comparable to baselines, indicating that predicting antigen-specific amino acid identity at binding positions is a fundamental bottleneck not resolved by contact-aware conditioning alone. Future work should explore richer antigen representations (side-chain geometry, surface electrostatics) and multi-modal sequence heads that capture the combinatorial nature of contact residue selection.

Impact Statement

This paper presents work whose goal is to advance computational antibody design. Designed sequences require extensive experimental validation before any therapeutic application. We see no specific negative societal consequences that must be highlighted.

References

Abir, A. R., Shahgir, H. S., Ratul, M. R. Z., Tahmid, M. T., Steeg, G. V., and Dong, Y. Abflownet: Optimizing antibody-antigen binding energy via diffusion-gflownet fusion. *arXiv preprint arXiv:2505.12358*, 2025.

Ahmed, M., Taj, N., Khan, I. U., Venkateswara, H., and Patterson, M. CHIMERA-bench: A benchmark dataset for epitope-specific antibody design. In *ICLR 2026 Workshop on Generative and Experimental Perspectives for Biomolecular Design*, 2026. URL https://openreview.net/forum?id=PyZvVIJbSy.

Alon, U. and Yahav, E. On the bottleneck of graph neural networks and its practical implications. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=i80OPhOCVH2.

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

Chen, J., Cai, X., Wu, J., and Hu, W. Antibody design and optimization with multi-scale equivariant graph diffusion models for accurate complex antigen binding. *arXiv preprint arXiv:2506.20957*, 2025.

Chinery, L., Hummer, A. M., Mehta, B. B., Akbar, R., Rawat, P., Slabodkin, A., Le Quy, K., Lund-Johansen, F., Greiff, V., Jeliazkov, J. R., and Deane, C. M. Simple computational methods can outperform deep learning in designing diverse, binder-enriched antibody libraries. *bioRxiv*, 2024. doi: 10.1101/2024.03.26.586756.

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

Gainza, P., Sverrisson, F., Monti, F., Rodola, E., Boscaini, D., Bronstein, M., and Correia, B. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, 17(2):184–192, 2020.

Gainza, P., Wehrle, S., Van Hall-Beauvais, A., Marchand, A., Scheck, A., Harteveld, Z., Buckley, S., Ni, D., Tan, S., Sverrisson, F., et al. De novo design of protein interactions with learned surface fingerprints. *Nature*, 617 (7959):176–184, 2023.

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

Jin, W., Wohlwend, J., Barzilay, R., and Jaakkola, T. Iterative refinement graph neural network for antibody sequence-structure co-design. In *International Conference on Learning Representations*, 2022b.

Kong, X., Huang, W., and Liu, Y. Conditional antibody design as 3D equivariant graph translation. In *International Conference on Learning Representations*, 2023a.

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

Li, Y., Lang, Y., Xu, C., Zhou, Y., Pang, Z., and Greisen, P. J. Benchmarking inverse folding models for antibody CDR sequence design. *PLOS ONE*, 20(6):e0324566, 2025. doi: 10.1371/journal.pone.0324566.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. Focal loss for dense object detection. In *IEEE International Conference on Computer Vision*, pp. 2980–2988, 2017.

Luo, S., Su, Y., Peng, X., Wang, S., Peng, J., and Ma, J. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. *Advances in Neural Information Processing Systems*, 35: 9754–9767, 2022.

Satorras, V. G., Hoogeboom, E., and Welling, M. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Sestak, F., Schneckenreiter, L., Brandstetter, J., Hochreiter, S., Mayr, A., and Klambauer, G. VN-EGNN: E(3)-equivariant graph neural networks with virtual nodes enhance protein binding site identification. *Journal of Cheminformatics*, 18:11, 2026. doi: 10.1186/s13321-025-01127-9.

Tan, C., Zhang, Y., Gao, Z., Huang, Y., Lin, H., Wu, L., Wu, F., Blanchette, M., and Li, S. Z. dyab: Flow matching for flexible antibody design with alphafold-driven prebinding antigen. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 782–790, 2025.

Uçar, T. and Sormanni, P. Blosum is all you learn—generative antibody models reflect evolutionary priors. *bioRxiv*, pp. 2025–10, 2025.

Verma, Y., Heinonen, M., and Garg, V. Abode: Ab initio antibody design using conjoined odes. In *International Conference on Machine Learning*, pp. 35037–35050. PMLR, 2023.

Wang, Z., Ji, Y., Tian, J., and Zheng, S. Retrieval augmented diffusion model for structure-informed antibody design and optimization. *arXiv preprint arXiv:2410.15040*, 2024.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borber, A. J., Ragotte, R. J., et al. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976): 1089–1100, 2023.

Wu, J., Kong, X., Sun, N., Wei, J., Shan, S., Feng, F., Wu, F., Peng, J., Zhang, L., Liu, Y., and Ma, J. Flowdesign: Improved design of antibody cdrs through flow matching and better prior distributions. *Cell Systems*, 2025a. doi: 10.1016/j.cels.2025.101270.

Wu, L., Lin, H., Huang, Y., Gao, Z., Tan, C., Liu, Y., Wu, T., and Li, S. Z. 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, 2025b.