UNITACT: SEQUENCE-ONLY UNIFIED ANTIBODY–ANTIGEN INTERACTION MODELING FOR CONTACT AND INTERFACE PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Identifying antibody–antigen interfaces is essential for understanding immune recognition and the effects of antigen mutations. Given a binding antibody–antigen pair, however, their relative binding geometry is unknown from sequences or monomer structures alone, making the joint prediction of antibody-side paratopes, antigen-side epitopes, and residue-level contacts challenging. Existing approaches either rely on complex structures or focus on individual interface tasks, and residue-pair predictions can be dominated by marginal residue preferences rather than partner-specific interactions.

We introduce UniTact, a unified framework that predicts paratopes, epitopes, and residue-level contacts from a shared antibody–antigen contact surface. The model takes sequences as its only raw input and incorporates separately predicted monomer structures to construct interaction representations through intrachain encoding, cross-attention, and pair updates. We further introduce a double-centering strategy on pair scores to reduce additive row and column effects, making the readouts more sensitive to pair-specific variation.

On SAbDab2, our method achieves an epitope AUPRC of 0.2232, improving over the sequence-based baseline SPPIDER-seq (0.2034), while obtaining paratope AUROC of 0.9561 and contact AUROC of 0.9203. Further score decomposition, controlled experiments, and split-based analyses help characterize the sources of residue-level bias and clarify the remaining challenges of pair-specific contact prediction.

1 INTRODUCTION

Given an antibody–antigen pair known to bind, locating their interface is fundamental to understanding antibody recognition and how antigen mutations alter it. Experimental studies show that mutations in the same antigen can affect different antibodies differently (Greaney et al., 2021). The interface gives rise to three related prediction targets: the antibody-side paratope, the antigen-side epitope, and residue-level contacts between the two chains. The challenge is that the sequences or monomer structures of the binding partners do not reveal their relative orientation in the complex; different antibodies may also bind distinct regions of the same antigen. We therefore seek to predict the interface of a given antibody–antigen pair without access to an experimentally determined complex structure.

Existing methods differ in both their required inputs and their prediction targets. Structure-based approaches use surface geometry to estimate binding sites or residue contacts. DiscoTope-3.0 uses inverse-folding representations of antigen structures to predict epitopes without conditioning on a specific antibody (Høie et al., 2024). MIPE combines sequence and structural information from both partners and estimates an interchain interaction matrix (Wang et al., 2024). EpiFormer uses geometric encoding and cross-attention to exchange information between the two chains during encoding (Ahmed et al., 2026). Methods that require only sequences as raw input reduce dependence on structural inputs: Paraplume predicts antibody paratopes (Athènes et al., 2026), while SPPIDER-seq uses cross-attention to predict partner-conditioned binding sites (Porollo et al., 2026). These approaches have advanced binding-site prediction and interchain modeling, but differ in the tasks they support and the inputs they require.

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.

Cross-chain information exchange and pairwise scores alone do not establish that a model correctly identifies which residues contact each other. Epitopes depend on the specific antibody, while true contacts are sparse among candidate residue pairs. A strong antibody-side binding-site signal can therefore produce favorable contact rankings even when pair-specific interactions remain weak. In our unified evaluation, paratope prediction is substantially stronger than epitope prediction, and high contact AUROC does not necessarily imply comparably effective contact detection. Decomposing the contact scores further shows that our model's rankings are largely driven by antibody-side row effects. Dataset splits affect how progress is assessed, as does metric choice under severe class imbalance (Saito & Rehmsmeier, 2015). These observations motivate separate evaluation of residue-level predictions, pair-specific interactions, and contact detection.

In this work, we introduce UniTact to jointly predict all three tasks from a shared antibody–antigen contact surface. The model takes sequences as its raw input and incorporates predicted monomer geometry. It constructs an interaction representation through intrachain encoding, cross-attention, and axial pair updates, then applies double centering over valid residue pairs to attenuate additive row and column effects. Paratope and epitope predictions are read out from the pairwise component of the contact surface together with residue-level scores. This design shares an interaction representation across the three tasks while making residue-level readouts more sensitive to pair-specific variation. On SAbDab2, our model achieves an epitope AUPRC of 0.2232, compared with 0.2034 for the sequence-input baseline SPPIDER-seq; its paratope and contact AUROCs are 0.9561 and 0.9203, respectively. Our contributions are threefold. First, we formulate the three supervision targets from a common contact matrix and model them with separate readouts over a shared interaction representation. Second, we apply a parameter-free double-centering projection to isolate the nonadditive component of pair scores while retaining residue-level terms for contact prediction. Third, we use score decomposition and evaluation across SAbDab2 and two AsEP splits to distinguish interface-residue ranking from effective residue-pair contact detection and to examine how conclusions depend on the data split.

2 Related Work

2.1 Antibody–Antigen Interface Prediction

Antibody–antigen interface prediction is commonly studied at three levels: antibody-side *paratopes*, antigen-side *epitopes*, and residue-pair-level *contacts*. Among sequence-based methods, ParaDeep employs a BiLSTM–CNN for paratope prediction but does not model antigen information (Udomwong et al., 2025); Paraplume leverages protein language model representations for paratope prediction and is likewise restricted to the antibody side (Athènes et al., 2026); SPPIDER-seq combines ESM-2 representations with cross-attention to predict partner-aware interface residues, but targets general protein–protein interactions and does not explicitly model residue-pair contacts (Porollo et al., 2026).

Structure-aware methods further incorporate geometric information. DiscoTope-3.0 predicts epitopes using ESM-IF1 representations but does not condition its predictions on a specific antibody. MIPE integrates sequence and structural information and constructs antibody–antigen interaction matrices, but primarily targets residue-level paratope and epitope prediction. EpiFormer models antibody–antigen interactions using GNNs and interleaved cross-attention, but similarly focuses on epitope prediction rather than explicit residue-pair contacts. ImaPEp predicts binding compatibility between paratope and epitope patches, but relies on antibody–antigen complex structures to construct candidate interface patches, limiting its applicability when the bound complex structure is unavailable (Li et al., 2024).

2.2 Protein Language Models

Protein language models learn general residue-level representations from large collections of unlabeled protein sequences. The ESM family demonstrated that masked language modeling can capture rich structural and functional information from protein sequences (Rives et al., 2021). ESM-2 further scaled model capacity and representation quality and served as the sequence representation backbone of ESMFold (Lin et al., 2023). More recently, ESM Cambrian (ESMC) further emphasizes

efficient and scalable protein representation learning, providing high-quality sequence representations for downstream protein modeling tasks (ESM Team, 2024).

2.3 Sequence-Derived Protein Structure Prediction

Recent advances in protein language models have reduced the dependence of protein structure prediction on conventional multiple sequence alignment pipelines. RaptorX-Single integrates pretrained protein language models with a structure generation module for single-sequence structure prediction (Jing et al., 2024). In the antibody domain, xTrimoABFold further exploits antibody-specific language representations for MSA-free antibody structure prediction (Wang et al., 2022). More recently, OpenDDE introduces an open-source all-atom biomolecular foundation model that uses co-folding as a general structural reasoning framework, extending sequence-driven structural modeling from individual proteins to diverse biomolecular complexes (Aureka AI OpenDDE Project, 2026).

3 Method

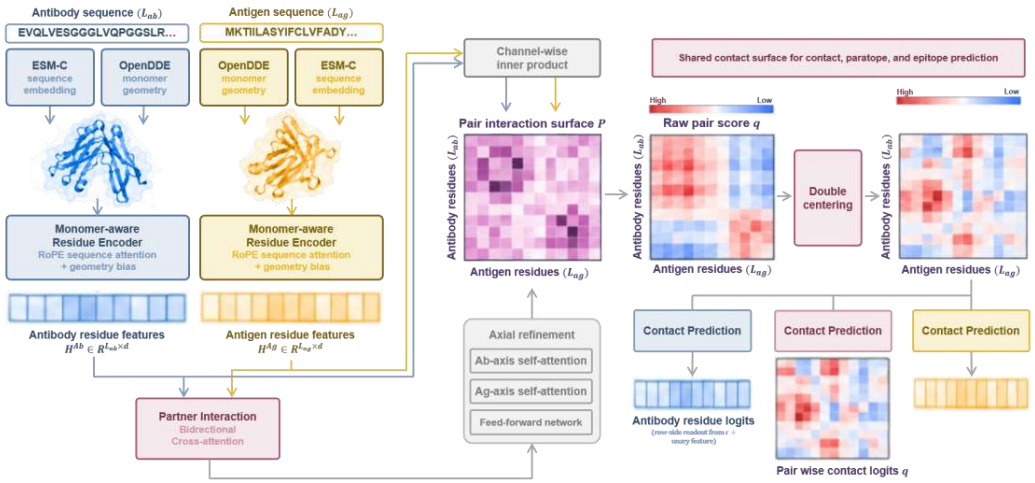

Figure 1: **Overview of UniTact.** Frozen ESMC embeddings and separately predicted OpenDDE monomer geometries are encoded for the antibody and antigen. Partner-conditioned residue features and a centered pair interaction map support the contact, paratope, and epitope heads. The sequence encoder uses rotary positional embeddings (RoPE) in self-attention, the geometry encoder adds intrachain distance biases, and the partner encoder exchanges information through bidirectional cross-attention. Channel-wise inner products construct the pair representation, followed by antibody- and antigen-axis self-attention and a feed-forward network (FFN).

3.1 Problem Definition

Given an antibody sequence s^{Ab} and an antigen sequence s^{Ag} of lengths L_a and L_g, respectively, UniTact predicts interchain contacts, antibody paratopes, and antigen epitopes. It uses no experimental complex structure or binding pose: all geometric information is derived from monomer structures predicted separately from the two sequences (Aureka AI OpenDDE Project, 2026). Let $Y \in \{0,1\}^{L_a \times L_g}$ be the native-complex contact labels constructed in Section 4.1, and $M_{ij} = m_i^{\mathrm{Ab}} m_j^{\mathrm{Ag}}$ the valid-pair mask. The residue labels are $y_i^{\mathrm{para}} = \bigvee_j Y_{ij}$ and $y_j^{\mathrm{epi}} = \bigvee_i Y_{ij}$. These exact relationships define the supervision labels; UniTact predicts the three tasks through separate readouts over a shared pair interaction representation.

3.2 Sequence and Monomer Geometry Encoding

Sequence representation. For each chain $c \in \{\mathrm{Ab}, \mathrm{Ag}\}$, frozen ESMC (ESM Team, 2024) produces $X_0^c \in \mathbb{R}^{L_c \times d_{\mathrm{ESMC}}}$. Independent linear projections give $X^c = X_0^c (W_{\mathrm{proj}}^c)^\top + \mathbf{1}(b_{\mathrm{proj}}^c)^\top \in \mathbb{R}^{L_c \times d}$, with $d = 128$. The two towers have the same architecture but separate trainable parameters.

Sequence encoder. A RoPE self-attention block (Su et al., 2021) and an FFN, each with pre-normalization and a residual connection, yield $H_{\mathrm{seq}}^c = \mathrm{SeqEncoder}^c(X^c)$ with sequence context and relative residue-position information.

Intrachain geometry encoder. The separately predicted monomer of chain c yields a $\mathrm{C}\alpha$ intrachain distance matrix $D^c \in \mathbb{R}^{L_c \times L_c}$; it supplies no relative antibody–antigen pose. Each distance is expanded into radial basis function (RBF) features $g_{ik}^c = \mathrm{RBF}(D_{ik}^c) \in \mathbb{R}^{n_r}$, then converted into a head-specific bias $\beta_{ik}^{c,h} = [\mathrm{MLP}_{\mathrm{geo}}^c(g_{ik}^c)]_h$. Let $Q_i^{c,h}$ and $K_k^{c,h}$ be the RoPE-rotated query and key projections of $\mathrm{LN}(H_{\mathrm{seq}}^c)$ at positions i and k. The geometry-aware attention weights are

$$a_{ik}^{c,h} = \underset{k:m_k^c=1}{\mathrm{softmax}} \left(\frac{\langle Q_i^{c,h}, K_k^{c,h} \rangle}{\sqrt{d_h}} + \beta_{ik}^{c,h} \right). \tag{1}$$

Here d_h is the per-head width. RoPE encodes relative sequence position, while the bias encodes intrachain spatial proximity in the predicted fold. Standard residual attention and FFN updates yield $H_{\mathrm{geo}}^c = \mathrm{GeometryEncoder}^c(H_{\mathrm{seq}}^c, D^c)$.

3.3 Cross-Chain Interaction Encoding

The partner encoder exchanges information in both directions. Write $(c, d) \in \{(\mathrm{Ab}, \mathrm{Ag}), (\mathrm{Ag}, \mathrm{Ab})\}$ for the query and partner chains. With direction-specific projections $Q_i^{c \leftarrow d,h} = W_q^{c \leftarrow d,h} \mathrm{LN}(H_{\mathrm{geo},i}^c)$ and $K_j^{c \leftarrow d,h} = W_k^{c \leftarrow d,h} \mathrm{LN}(H_{\mathrm{geo},j}^d)$, cross-attention (Vaswani et al., 2017) is

$$A_{ij}^{c \leftarrow d,h} = \underset{j:m_j^d=1}{\mathrm{softmax}} \left(\frac{\langle Q_i^{c \leftarrow d,h}, K_j^{c \leftarrow d,h} \rangle}{\sqrt{d_h}} \right). \tag{2}$$

The reverse direction has independent parameters; both updates use the same pre-update states. No cross-chain RoPE or geometry bias is used because the chains have neither a shared sequence coordinate system nor a known binding pose. Value aggregation and residual updates give, for both (c, d) choices,

$$H^c = \mathrm{Partner}_{c \leftarrow d}(H_{\mathrm{geo}}^c, H_{\mathrm{geo}}^d). \tag{3}$$

Thus H^{Ab} and H^{Ag} are partner-conditioned residue representations.

3.4 Pair Interaction Encoding and Decoding

Pair encoder. To represent individual residue pairs, each tower projects into C interaction channels: $f_{ir}^c = [W_c \mathrm{LN}(H_i^c)]_r \in \mathbb{R}^{d_o}$ for $c \in \{\mathrm{Ab}, \mathrm{Ag}\}$. The channel-wise compatibility defines a multichannel pair surface,

$$P_{ijr}^{(0)} = \frac{\langle f_{ir}^{\mathrm{Ab}}, f_{jr}^{\mathrm{Ag}} \rangle}{\sqrt{d_o}}, \qquad P^{(0)} \in \mathbb{R}^{L_a \times L_g \times C}. \tag{4}$$

Axial pair decoder. Attention along the antibody axis exchanges pair features that share an antigen residue; attention along the antigen axis exchanges features that share an antibody residue (Ho et al., 2019). This avoids attention over all $L_a L_g$ pairs at once:

$$\begin{aligned} P^{(1)} &= P^{(0)} + \mathrm{Attn}_{\mathrm{Ab}}(\mathrm{LN}(P^{(0)})), \\ P^{(2)} &= P^{(1)} + \mathrm{Attn}_{\mathrm{Ag}}(\mathrm{LN}(P^{(1)})), \\ P &= P^{(2)} + \mathrm{FFN}_{\mathrm{pair}}(\mathrm{LN}(P^{(2)})). \end{aligned} \tag{5}$$

The refined surface $P \in \mathbb{R}^{L_a \times L_g \times C}$ retains pair-level context for all three tasks.

3.5 Centered Interaction Readout

A shared projection produces the scalar map $q_{ij} = w_{\text{int}}^{\top} \text{LN}(P_{ij:})$. Raw pair scores can contain additive row and column effects, $q_{ij} \approx r_i + c_j + \text{const}$, which do not identify specific residue-pair coupling. We remove these effects using a mask-aware double centering. Let $\mathcal{J}_i = \{j : M_{ij} = 1\}$ and $\mathcal{I}_j = \{i : M_{ij} = 1\}$. The row, column, and global averages are, in one notation,

$$\bar{q}_{i\cdot} = \langle q_{ij} \rangle_{j \in \mathcal{J}_i}, \quad \bar{q}_{\cdot j} = \langle q_{ij} \rangle_{i \in \mathcal{I}_j}, \quad \bar{q} = \langle q_{ij} \rangle_{M_{ij}=1}. \tag{6}$$

Here $\langle \cdot \rangle$ denotes the arithmetic mean over the indicated valid indices. The resulting pair-specific signal is

$$E_{ij} = [\Pi q]_{ij} = M_{ij}(q_{ij} - \bar{q}_{i\cdot} - \bar{q}_{\cdot j} + \bar{q}). \tag{7}$$

The fixed operator Π removes purely antibody-row and antigen-column additive components; $E \in \mathbb{R}^{L_a \times L_g}$ is used by every task head.

3.6 Unified Contact and Interface Readouts

Contact prediction. The contact logit combines the centered interaction with unary interface tendencies on both chains:

$$z_{ij}^{\text{contact}} = E_{ij} + w_u^{\top} H_i^{\text{Ab}} + w_v^{\top} H_j^{\text{Ag}} + b_c. \tag{8}$$

Interface prediction. The two residue-level logits combine unary scores with aggregations of E along opposite axes:

$$\begin{aligned} s_i^{\text{para}} &= g_{\text{para}}(H_i^{\text{Ab}}) + \text{Agg}_i^{\text{para}}(E), \\ s_j^{\text{epi}} &= g_{\text{epi}}(H_j^{\text{Ag}}) + \text{Agg}_j^{\text{epi}}(E). \end{aligned} \tag{9}$$

For paratope, $\text{Agg}_i^{\text{para}}(E) = \log\big(|\mathcal{J}_i|^{-1} \sum_{j \in \mathcal{J}_i} e^{E_{ij}}\big)$ is a length-normalized log-mean-exp. For epitope, let $\mathcal{T}_j \subseteq \mathcal{I}_j$ contain the $k_j = \min(2, |\mathcal{I}_j|)$ largest interactions, set $\mu_j = k_j^{-1} \sum_{i \in \mathcal{T}_j} \sigma(E_{ij})$, and use $\text{Agg}_j^{\text{epi}}(E) = \text{logit}[\text{clip}(\mu_j, \epsilon, 1-\epsilon)]$. The corresponding probabilities are obtained by applying the sigmoid function to these logits.

3.7 Training Objective

Let $\mathcal{S} = \{\text{para}, \text{epi}, \text{contact}\}$. For each task $t \in \mathcal{S}$, write a_u^t for its logit (s_i^{para}, s_j^{epi}, or z_{ij}^{contact}), $p_u^t = \sigma(a_u^t)$, and y_u^t for the corresponding label. All losses below are evaluated on valid prediction units.

Balanced BCE–Tversky objective. For each task, the positive and negative BCE terms are normalized separately; a Tversky term (Salehi et al., 2017) uses class-normalized soft true positives, false positives, and false negatives. The task loss is

$$\mathcal{L}_{\text{fused}}^t = 1 - \text{Tv}^t + \eta \mathcal{L}_{\text{bBCE}}^t, \qquad \eta = 0.1. \tag{10}$$

We use Tversky weights $(\alpha, \beta) = (1, 2)$ for contact and epitope and $(1, 1)$ for paratope; the exact normalization and component formulas are given in Appendix A.

Contact marginal constraints. To constrain the amount of contact around each residue, let $\mathcal{I} = \{i : m_i^{\text{Ab}} = 1\}$ and $\mathcal{J} = \{j : m_j^{\text{Ag}} = 1\}$. We match observed and predicted contact counts along both axes:

$$\begin{aligned} \mathcal{L}_{\text{marg}}^{\text{row}} &= \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \left| \sum_{j \in \mathcal{J}_i} (p_{ij}^{\text{contact}} - Y_{ij}) \right|, \\ \mathcal{L}_{\text{marg}}^{\text{col}} &= \frac{1}{|\mathcal{J}|} \sum_{j \in \mathcal{J}} \left| \sum_{i \in \mathcal{I}_j} (p_{ij}^{\text{contact}} - Y_{ij}) \right|. \end{aligned} \tag{11}$$

Prevalence constraint. For task t, let Ω_t contain its valid prediction units. The predicted and observed positive fractions are matched by

$$\mathcal{L}_{\text{prev}}^{t} = \left| \frac{1}{|\Omega_t|} \sum_{u \in \Omega_t} (p_u^t - y_u^t) \right|. \tag{12}$$

Total objective. The resulting objective is

$$\begin{aligned} \mathcal{L} = \lambda_p \mathcal{L}_{\text{fused}}^{\text{para}} + \lambda_e \mathcal{L}_{\text{fused}}^{\text{epi}} + \lambda_c \mathcal{L}_{\text{fused}}^{\text{contact}} \\ + \lambda_{\text{marg}} (\mathcal{L}_{\text{marg}}^{\text{row}} + \mathcal{L}_{\text{marg}}^{\text{col}}) + \lambda_{\text{prev}} \sum_{t \in \mathcal{S}} \mathcal{L}_{\text{prev}}^{t}. \end{aligned} \tag{13}$$

4 Experiments

This section describes the datasets, training and evaluation protocols, and controlled experiments for analyzing pair-specific interactions and the epitope prediction bottleneck.

4.1 Data and Preprocessing

Datasets and splits. We evaluate on SAbDab2 (Capel et al., 2026) and AsEP (Liu et al., 2024). For SAbDab2, we retain complexes containing Fab, Fv, or single-domain heavy-chain (SD-H) antibodies and adopt the joint antibody–antigen sequence similarity-based splitting strategy for antigen-aware tasks to reduce information leakage between training and test sets. For AsEP, we use its two proposed splitting strategies: the epitope-ratio split stratifies examples according to the proportion of epitope residues; the epitope-group split assigns different epitope groups to separate subsets. The resulting dataset sizes are summarized in Table 1. The two AsEP splits share the same sequences, structures, and labels, but differ in the assignment of samples to training, validation, and test sets.

Table 1: Dataset splits used for evaluation. SAbDab2 uses joint antibody–antigen sequence clustering; AsEP[1] is stratified by epitope ratio; AsEP[2] is based on epitope groups with disjoint antigens across splits.

Dataset and split	Train	Validation	Test
SAbDab2	2,887	321	455
AsEP[1]	1,175	138	149
AsEP[2]	1,157	158	147

Model inputs. We encode antibody and antigen sequences using frozen ESMC-300M (ESM Team, 2024) and use its 960-dimensional residue-level representations. Frozen OpenDDE (Aureka AI OpenDDE Project, 2026) independently predicts the monomer structure of each chain, from which we compute intrachain residue distances. Structure prediction does not use multiple sequence alignments (MSAs) or templates. The model receives neither complex structures nor interchain geometric information.

Supervision. Experimental complex structures are used only for constructing supervision labels. Following a commonly used distance-based definition for antibody–antigen interfaces (Gordon et al., 2023), a residue pair is defined as a contact if the distance between any pair of atoms from the two residues is no greater than 4.5 Å. Paratope and epitope labels are obtained by applying logical OR operations over the antigen and antibody axes of the contact matrix, respectively. Residues without resolved coordinates are excluded from label construction, loss computation, and evaluation. Sequence representations, predicted structures, and experimental labels are mapped to a common residue index through sequence alignment.

4.2 Implementation and Training

All experiments use the same model configuration: a residue hidden width of 128, one interaction block, one dense geometric attention stream with distance bias, and one axial pair-update layer. The

pair representation uses 32 channels. Intrachain distances are encoded with 16 radial basis functions and truncated at 20 Å. The epitope top-k readout uses $k = 2$. Ablation experiments modify only the corresponding component while keeping other settings unchanged.

The model is trained with AdamW using an initial learning rate of 3×10^{-4}, one-epoch linear warmup, and cosine decay. Training lasts at most 40 epochs with an early-stopping patience of 8 epochs. The loss weights are: $\lambda_c = 1$, $\lambda_p = \lambda_e = 0.25$, $\lambda_{\text{marg}} = 0.1$, and $\lambda_{\text{prev}} = 0.2$. Batches are constructed with a budget of 5.5×10^5 padded residue pairs, and gradient checkpointing is applied to the interaction block.

Main experiments use 10 random seeds, while ablations use 5 of them. For each run, a single checkpoint is selected according to validation loss and used for all three tasks. The test set is not used for hyperparameter or checkpoint selection.

4.3 Evaluation Protocol

Contact, paratope, and epitope predictions are evaluated using complex-level macro averaging. We report AUROC, AUPRC, precision, recall, F1, and MCC, where AUPRC is computed using non-interpolated average precision. For threshold-dependent metrics, our model uses a threshold of 0.5; undefined ratios caused by zero denominators are assigned a value of zero and retained in the macro average. Cross-method comparisons mainly rely on threshold-independent AUROC and AUPRC.

The main tables report means over 10 runs; ablations use five matched seeds for paired comparisons.

4.4 Analysis of Pair-Specific Interaction

We ask whether contact scores reflect specific residue pairs or mostly individual residues. For both the raw pair scores q and final contact logits z^{contact}, we write $S_{ij} = \bar{S} + r_i + c_j + \varepsilon_{ij}$ on valid pairs, where r_i and c_j are row and column effects and $\varepsilon = \Pi S$ is the pair-specific residual (Section 3.5). We measure variance contributions in both maps and compare contact rankings from the full logits, their additive component, and the residual. Comparing q with z^{contact} shows the effect of the unary scores added by the contact head.

Across five matched seeds, we remove double centering and, separately, feed full contact logits instead of E to the paratope and epitope readouts. We also vary the epitope top-k width while keeping the learned pair representation fixed.

For antigens bound by different antibodies, we align residues and measure the overlap between the corresponding epitope label sets using Jaccard similarity. Frozen-input probes compare sequence features, predicted monomer geometry, and a global antibody summary. Further structural, antigen-feature, and pooling controls are described in the Appendix; none uses a complex pose or native interface geometry.

5 Results

5.1 Main Results on Three Tasks

Table 2 compares the three prediction tasks on the SAbDab2 test set, grouping methods by the raw inputs required at inference. Among sequence-only methods, our model achieves an epitope AUPRC of 0.2232, compared with 0.2034 for SPPIDER-seq, and a paratope AUROC of 0.9561, compared with 0.9514 for Paraplume. Its paratope recall reaches 0.8759, although its AUPRC of 0.7278 is slightly below Paraplume's 0.7319. Methods using experimental structures and sequences achieve higher epitope AUPRC; EpiFormer reaches 0.3480. For contact prediction, our model attains the highest AUROC in the table (0.9203), but its AUPRC of 0.0208 is below that of the structure-based ProTact-best-antibody (Rao et al., 2026) (0.0302). Its contact recall is 0.0058, indicating that stronger residue-pair ranking has yet to translate into effective contact detection. Section 5.3 examines the source of the ranking signal.

Table 2: Performance on the SAbDab2 test set, grouped by the raw inputs required at inference. A dash denotes an unsupported task. Our monomer structures are predicted from sequences, so our model is categorized as sequence-only, and bold marks the column-wise best value within that group only. Contact columns carry no bold mark, as our model is the only sequence-only method supporting that task.

Input type	Model	Contact			Epitope			Paratope		
		AUROC	AUPRC	Recall	AUROC	AUPRC	Recall	AUROC	AUPRC	Recall
Structure-only	ProTact-best-antibody	0.8610	0.0302	0.0066	—	—	—	—	—	—
	ProTact-best-dips	0.5652	0.0034	0.0016	—	—	—	—	—	—
Structure + sequence	DiscoTope-3.0	—	—	—	0.6862	0.2577	0.5346	—	—	—
	EpiFormer	—	—	—	0.7677	0.3480	0.4785	—	—	—
	MIPE	0.4992	0.0015	0.0001	0.7918	0.2514	0.1064	0.8433	0.6214	0.4029
Sequence-only	SPPIDER-seq	—	—	—	0.6337	0.2034	**0.5222**	0.9481	0.6981	0.7330
	ParaDeep	—	—	—	—	—	—	0.9089	0.5833	0.7349
	Paraplume	—	—	—	—	—	—	0.9514	**0.7319**	0.7485
	Ours	0.9203	0.0208	0.0058	**0.6371**	**0.2232**	0.3167	**0.9561**	0.7278	**0.8759**

5.2 Results on the Two AsEP Splits

Table 3 reports our model's results on SAbDab2 and the two AsEP splits. The epitope-group split serves as our primary AsEP evaluation: its epitope AUROC/AUPRC are 0.6556/0.2639, compared with 0.7602/0.4485 on the epitope-ratio split. Contact AUPRC likewise rises from 0.0565 to 0.1804. The AsEP splits share the same processed sequences, geometries, and labels but assign complexes differently to training, validation, and test sets.

Table 3: Performance of our model on SAbDab2 and the two AsEP splits. Metrics are macro-averaged over complexes.

Dataset and split	Task	AUROC	AUPRC	Recall
SAbDab2	Epitope	0.6371	0.2232	0.3167
	Paratope	0.9561	0.7278	0.8759
	Contact	0.9203	0.0208	0.0058
AsEP: epitope group	Epitope	0.6556	0.2639	0.3607
	Paratope	0.9569	0.7320	0.8711
	Contact	0.9394	0.0565	0.1510
AsEP: epitope ratio	Epitope	0.7602	0.4485	0.6068
	Paratope	0.9691	0.7815	0.9026
	Contact	0.9620	0.1804	0.3588

5.3 Additive Effects in Pair Scores and Contact Logits

The decomposition in Section 4.4 separates two questions: whether the pre-centering pair scores q are additive, and how the unary terms restored in the final contact logits affect ranking. In the final logits, antibody-side row effects outweigh the pair-specific residual. This helps interpret the gap between contact AUROC (0.9203), AUPRC (0.0208), and recall (0.0058): favorable overall rankings do not by themselves establish effective contact detection. The variance decomposition and ranking-control procedures are described in Appendix C; the variance table there concerns an earlier, uncentered configuration.

5.4 Centering Ablations and Partner Dependence

Matched-seed ablations remove double centering, replace E-based interface readouts with pooling of full contact logits, and vary the epitope top-k width. The control settings are described in Appendix C.

Epitope sets for the same antigen paired with different antibodies have limited overlap, supporting the need for partner-conditioned prediction. Changes to pooling and antigen-side inputs yield limited gains in our controls. These observations are consistent with insufficient partner-specific information in the learned interaction scores; they do not isolate a single cause. The input and pooling control designs are described in Appendix C.

6 Conclusion

We formulate contact, paratope, and epitope prediction using labels derived from a common antibody–antigen contact matrix. Under a shared contact criterion, the paratope and epitope labels are exact logical-OR projections along the antigen and antibody axes, respectively. The three predictions use separate task readouts over a shared pair interaction representation. To reduce the influence of additive row and column effects on residue-level predictions, we apply parameter-free double centering over the rectangular domain of valid residue pairs and use the centered interaction scores for residue-level readouts.

On SAbDab2, our method achieves higher epitope AUPRC and paratope AUROC than the other sequence-only methods, as well as the highest contact AUROC among the compared methods. Nevertheless, epitope performance remains below that of methods using experimental structures, while contact precision and recall remain low. Score decomposition shows that antibody-side row effects strongly influence contact rankings; a high AUROC alone therefore does not establish effective contact detection. Results on the two AsEP splits further show that performance comparisons must account for the data split.

Double centering changes how the readouts use learned interaction scores, but cannot create residue-pair information the model has not learned. Predicted monomer structures provide no complex pose, and transferring antibody information more effectively to antigen-side representations remains important for epitope and contact prediction. Future work could provide more direct supervision for the pairwise interaction component and test whether predicted contact surfaces can serve as useful constraints for downstream co-folding.

AI Use Statement

Generative AI tools were used to assist with manuscript writing, including language polishing, ensuring consistent terminology, and editing text drafted by the authors in the introduction, methodology, experiments, appendix, and other sections. The authors take full responsibility for the final content of the paper.

References

Mansoor Ahmed, Huirong Chai, Haoxin Wang, Hemanth Venkateswara, and Murray Patterson. EpiFormer: Learning antigen-antibody interactions for epitope prediction via geometric deep learning, 2026.

Gabriel Athènes, Adam Woolfe, Thierry Mora, and Aleksandra M. Walczak. Paraplume: A fast and accurate antibody paratope prediction method provides insights into repertoire-scale binding dynamics. *PLOS Computational Biology*, 22(2):e1013981, 2026. doi: 10.1371/journal.pcbi.1013981.

Aureka AI OpenDDE Project. Folding, reasoning, and scaling with open-source drug discovery engine, 2026.

Henriette L. Capel, Odysseas Vavourakis, Benjamin H. Williams, Christopher R. Taylor, and Charlotte M. Deane. SAbDab2: The structural antibody database in the age of machine learning. *bioRxiv*, 2026. doi: 10.64898/2026.06.16.732554.

ESM Team. ESM Cambrian: Revealing the mysteries of proteins with unsupervised learning, 2024. URL https://www.evolutionaryscale.ai/blog/esm-cambrian.

Gemma L. Gordon, Henriette L. Capel, Bora Guloglu, Eve Richardson, Ryan L. Stafford, and Charlotte M. Deane. A comparison of the binding sites of antibodies and single-domain antibodies. *Frontiers in Immunology*, 14:1231623, 2023. doi: 10.3389/fimmu.2023.1231623.

Allison J. Greaney, Tyler N. Starr, Pavlo Gilchuk, Seth J. Zost, Elad Binshtein, Andrea N. Loes, Sarah K. Hilton, John Huddleston, Rachel Eguia, Katharine H. D. Crawford, Adam S. Dingens, Rachel S. Nargi, Rachel E. Sutton, Naveenchandra Suryadevara, Paul W. Rothlauf, Zhuoming Liu, Sean P. J. Whelan, Robert H. Carnahan, James E. Jr. Crowe, and Jesse D. Bloom. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host & Microbe*, 29(1):44–57.e9, 2021. doi: 10.1016/j.chom.2020.11.007.

Jonathan Ho, Nal Kalchbrenner, Dirk Weissenborn, and Tim Salimans. Axial attention in multidimensional transformers, 2019.

Magnus Haraldson Høie, Frederik Steensgaard Gade, Julie Maria Johansen, Charlotte Würtzen, Ole Winther, Morten Nielsen, and Paolo Marcatili. DiscoTope-3.0: improved b-cell epitope prediction using inverse folding latent representations. *Frontiers in Immunology*, 15:1322712, 2024. doi: 10.3389/fimmu.2024.1322712.

Xiaoyang Jing, Fandi Wu, Xiao Luo, and Jinbo Xu. Single-sequence protein structure prediction by integrating protein language models. *Proceedings of the National Academy of Sciences*, 121(13): e2308788121, 2024. doi: 10.1073/pnas.2308788121.

Dong Li, Fabrizio Pucci, and Marianne Rooman. Prediction of paratope–epitope pairs using convolutional neural networks. *International Journal of Molecular Sciences*, 25(10):5434, 2024. doi: 10.3390/ijms25105434.

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

Chu'nan Liu, Lilian Denzler, Yihong Chen, Andrew C. R. Martin, and Brooks Paige. AsEP: Benchmarking deep learning methods for antibody-specific epitope prediction. In *Advances in Neural Information Processing Systems*, volume 37, 2024. Datasets and Benchmarks Track.

Aleksey Porollo, Om Jadhav, Aaron Alvarez, and Jichao Chen. SPPIDER-seq: sequence-based partner-aware predictor of protein-protein interaction sites. *Bioinformatics*, 42(7):btag388, 2026. doi: 10.1093/bioinformatics/btag388.

Jiahua Rao, Deqin Liu, Xiaolong Zhou, Qianmu Yuan, Wentao Wei, Wei Lu, Jixian Zhang, Yu Rong, Yuedong Yang, and Shuangjia Zheng. Accurate protein-protein interactions modeling through physics-informed geometric invariant learning. *Communications Biology*, 9:685, 2026. doi: 10.1038/s42003-026-09809-2.

Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C. Lawrence Zitnick, Jerry Ma, and Rob Fergus. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15):e2016239118, 2021. doi: 10.1073/pnas.2016239118.

Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3):e0118432, 2015. doi: 10.1371/journal.pone.0118432.

Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Tversky loss function for image segmentation using 3d fully convolutional deep networks, 2017.

Jianlin Su, Yu Lu, Shengfeng Pan, Ahmed Murtadha, Bo Wen, and Yunfeng Liu. RoFormer: Enhanced transformer with rotary position embedding, 2021.

Piyachat Udomwong, Thanathat Pamonsupornwichit, Kanchanok Kodchakorn, and Chatchai Tayapiwatana. Paradeep: sequence-based deep learning for residue-level paratope prediction using chain-aware bilstm-cnn models. *Frontiers in Bioinformatics*, 5:1684042, 2025. doi: 10.3389/fbinf.2025.1684042.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Yining Wang, Xumeng Gong, Shaochuan Li, Bing Yang, Yiwu Sun, Chuan Shi, Yangang Wang, Cheng Yang, Hui Li, and Le Song. xTrimoABFold: De novo antibody structure prediction without msa, 2022.

Zhiwei Wang, Yongkang Wang, and Wen Zhang. Improving paratope and epitope prediction by multi-modal contrastive learning and interaction informativeness estimation. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 6053–6061, 2024. doi: 10.24963/ijcai.2024/669.

A Loss Function Details

For one task and one complex, let Ω^+ and Ω^- be the valid positive and negative prediction units. Write $p_u = \sigma(a_u)$, $\ell_u^+ = -\log p_u$, and $\ell_u^- = -\log(1-p_u)$. Define $N_\pm = \max(|\Omega^\pm|, 1)$, $\chi_\pm = \mathbb{1}[|\Omega^\pm| > 0]$, and $n_{\text{obs}} = \chi_+ + \chi_-$. Averaging the classes present in the complex gives

$$\mathcal{L}_{\text{bBCE}} = \frac{1}{n_{\text{obs}}} \left(\chi_+ \frac{\sum_{u\in\Omega^+} \ell_u^+}{N_+} + \chi_- \frac{\sum_{u\in\Omega^-} \ell_u^-}{N_-} \right). \tag{14}$$

Thus a complex containing only one observed class is normalized by that class alone. The normalized contributions to the Tversky score are

$$\begin{aligned} \text{TP} &= \frac{\sum_{u\in\Omega^+} p_u}{N_+}, & \text{FP} &= \frac{\sum_{u\in\Omega^-} p_u \ell_u^-}{N_-}, \\ \text{FN} &= \frac{\sum_{u\in\Omega^+} (1-p_u)\ell_u^+}{N_+}. \end{aligned} \tag{15}$$

We use

$$\text{Tv} = \frac{\text{TP} + \epsilon}{\text{TP} + \alpha\,\text{FP} + \beta\,\text{FN} + \epsilon}, \qquad \mathcal{L}_{\text{fused}} = 1 - \text{Tv} + \eta \mathcal{L}_{\text{bBCE}}, \tag{16}$$

with $\eta = 0.1$. The contact and epitope tasks use $(\alpha, \beta) = (1, 2)$, and the paratope task uses $(1, 1)$. The small $\epsilon > 0$ stabilizes the ratio. These quantities are computed on valid units for each complex before averaging across complexes. The marginal-count and prevalence penalties are defined in Eqs. (11) and (12).

B Additive Pair Scores and Double Centering

Where additive terms arise. For interaction channel r, write each projected residue vector as its chain mean plus a deviation: $f_{ir}^{\text{Ab}} = \bar{f}_r^{\text{Ab}} + \delta_{ir}^{\text{Ab}}$ and $f_{jr}^{\text{Ag}} = \bar{f}_r^{\text{Ag}} + \delta_{jr}^{\text{Ag}}$. Expanding Eq. (4) gives

$$\begin{aligned} \sqrt{d_o}\, P_{ijr}^{(0)} = \langle \bar{f}_r^{\text{Ab}}, \bar{f}_r^{\text{Ag}} \rangle + \langle \delta_{ir}^{\text{Ab}}, \bar{f}_r^{\text{Ag}} \rangle \\ + \langle \bar{f}_r^{\text{Ab}}, \delta_{jr}^{\text{Ag}} \rangle + \langle \delta_{ir}^{\text{Ab}}, \delta_{jr}^{\text{Ag}} \rangle. \end{aligned} \tag{17}$$

The four terms are constant, antibody-specific, antigen-specific, and pair-specific, respectively. This identity explains how additive structure can enter a pair representation; it does not determine their relative magnitudes after the axial decoder.

Effect on residue-level pooling. Suppose a scalar pair map is strictly additive, $S_{ij} = r_i + c_j$, and a permutation-invariant pooling function F obeys $F(x + t\mathbf{1}) = F(x) + t$. Then

$$F(\{S_{ij}\}_{i\in\mathcal{I}}) = c_j + F(\{r_i\}_{i\in\mathcal{I}}). \tag{18}$$

Mean, max, and length-normalized log-mean-exp pooling satisfy this identity. The antibody-side terms contribute the same offset to every antigen residue, so these poolers cannot obtain residue-pair discrimination from an additive map alone. The epitope pooling used in our model averages the top-k sigmoid values before returning to logit space; it is not translation equivariant, but its ranking is monotone in c_j for a strictly additive map, apart from ties introduced by clipping. This observation concerns the pair-map readout only: the separate unary epitope term in Eq. (9) can also carry partner information.

Projection on the valid domain. The valid residue pairs form a rectangle $\Omega = \mathcal{I} \times \mathcal{J}$ because $M_{ij} = m_i^{\mathrm{Ab}} m_j^{\mathrm{Ag}}$. Subtracting the row and column means and adding the global mean, as in Eq. (7), yields an interaction map $E = \Pi q$ whose valid row and column means are zero. On this domain, Π is an idempotent orthogonal projection; its kernel consists of maps of the form $r_i + c_j$. In particular, $\Pi E = E$. Centering does not create new pairwise information: it removes additive components from the scalar interaction readout while the contact head retains its own unary terms.

C Supplementary Interaction Analyses

Final-model score decomposition and ranking controls. For each complex, we decompose the pre-centering pair score q and the final contact logit z^{contact} on valid residue pairs. For either map $S \in \{q, z^{\mathrm{contact}}\}$, write $S_{ij} = \bar{S} + r_i^S + c_j^S + \varepsilon_{ij}^S$, where $r_i^S = \bar{S}_{i\cdot} - \bar{S}$, $c_j^S = \bar{S}_{\cdot j} - \bar{S}$, and $\varepsilon^S = \Pi S$. On the rectangular valid domain, these centered row, column, and residual terms are orthogonal. We compare their squared norms as fractions of the centered map's squared norm for each nonconstant map, then average the fractions across complexes. To inspect contact ranking, we compare the full logits, their additive component $\bar{z} + r_i^z + c_j^z$, and their residual ε_{ij}^z against the same contact labels using the protocol in Section 4.3. For the final model, $\varepsilon^z = E$ on valid pairs by Eq. (8); its additive component contains the antibody- and antigen-side unary scores.

Centering and interface-readout controls. We compare the full model with three targeted changes. First, removing double centering passes the raw pair score q in place of E to the contact and interface readouts. Second, we retain the centered contact head but feed the full contact logits z^{contact} in place of E to the paratope and epitope aggregators. Third, we vary the epitope top-k width while keeping the learned pair representation fixed. The model variants use the same data splits, inputs, training schedule, and evaluation metrics; comparisons that require training use matched random seeds and checkpoints selected by validation loss. No variant receives a native complex pose.

Frozen-input probes and feature controls. We compare fixed sequence representations, predicted monomer geometry, and a global summary of the paired antibody to examine how antigen-side and partner information affect epitope prediction. Structural controls vary only the available monomer-derived information, antigen-feature controls vary the antigen-side representation, and pooling controls change the reduction from pair interactions to residue scores. These controls use the same splits and epitope evaluation protocol as the main experiments. Geometry is derived from the separately predicted monomers; no control supplies native interchain distances or a known binding pose.

Score decomposition in an earlier configuration. To inspect the relative scale of pair and residue effects, we decomposed scores from an earlier, larger three-block configuration *without* double centering. Denote its raw scalar pair score by q and its contact logit by $z_{\mathrm{unc},ij} = q_{ij} + u_i + v_j + b$. For each complex, we separated the row, column, and residual components on valid pairs and then averaged their variance fractions across complexes.

Table 4: Variance fractions (%) in an exploratory, uncentered configuration. These values do not estimate the variance fractions of the final model.

Score map	Ab row	Ag column	Pair residual
Uncentered contact logit z_{unc}	93.50	5.00	1.50
Pre-centering pair score q	82.97	7.83	9.20

The larger residual fraction in q shows that this earlier model produced pair-specific variation whose share was reduced after adding the unary contact terms. It motivates inspecting additive effects before the final readout, without making a quantitative claim about the final model's logits. For the final configuration, $E = \Pi q$ by construction; decomposing E on the same valid rectangle would give zero row and column components.

Variation across antibody partners. To measure whether the same antigen presents the same labeled interface to different antibodies, we align residues within antigen sequence groups and compare the resulting epitope sets. The analysis contains 1,467 antibody-pair comparisons across 533 antigens with multiple antibody partners. The median Jaccard similarity is zero, and 84% of comparisons are below 0.3. These label differences motivate partner-conditioned prediction. Comparisons within a shared antigen group are related observations, so these counts are descriptive rather than independent replicates.