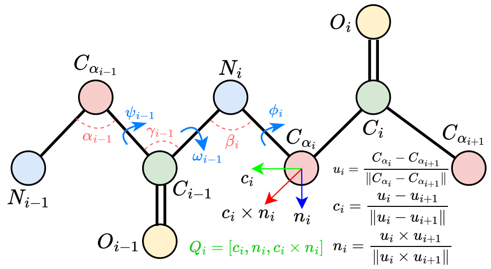
# Appendix

## A. Comparison of Method Characteristics

We compare various methods in Table. A1 from three key aspects: *(1) input contextual data*, including framework regions on the heavy chain, light chain, and antigen; *(2) architectural design*, the modeling levels of geometric information, and whether the modules are E(3)-equivariant; and *(3) optimization strategy*, whether it is autoregressive, whether it is iterative refinement-based or diffusion-based, and whether it evaluates the effects of model pre-training.

## B. Definitions of Node and Edge Features

The node and edge features are defined in terms of six key aspects: type, position, distance, direction, angle, and orientation. A summary of these features is available in Fig. A1.


$O_i$
$C_{\alpha_{i-1}}$
$N_i$
$\psi_{i-1}$
$\phi_i$
$\alpha_{i-1}$
$\gamma_{i-1}$
$\beta_i$
$\omega_{i-1}$
$C_{\alpha_i}$
$C_i$
$C_{\alpha_{i+1}}$
$N_{i-1}$
$C_{i-1}$
$c_i$
$c_i \times n_i$
$n_i$
$O_{i-1}$
$Q_i = [c_i, n_i, c_i \times n_i]$
$u_i = \frac{C_{\alpha_i} - C_{\alpha_{i+1}}}{\|C_{\alpha_i} - C_{\alpha_{i+1}}\|}$
$c_i = \frac{u_i - u_{i+1}}{\|u_i - u_{i+1}\|}$
$n_i = \frac{u_i \times u_{i+1}}{\|u_i \times u_{i+1}\|}$

| | Node $v_i$ | Edge $e_{i,j}$ |
|---|---|---|
| Type | $E_{\text{type}}(v_i)$ | $E_{\text{type}}(e_{i,j})$ |
| Position | $E_{\text{pos}}(i)$ | $E_{\text{pos}}(i-j)$ |
| Distance | $\|C_{\alpha_i} - C_i\|$ $\|C_{\alpha_i} - N_i\|$<br>$\|C_{\alpha_i} - O_i\|$ | $\|C_{\alpha_i} - C_j\|$ $\|C_{\alpha_i} - N_j\|$<br>$\|C_{\alpha_i} - O_j\|$ $\|C_{\alpha_i} - C_{\alpha_j}\|$ |
| Angle | Angle - $\alpha_i$, $\beta_i$, $\gamma_i$<br>Dihedral Angle - $\phi_i$, $\psi_i$, $\omega_i$ | — |
| Direction | $Q_i^\top \frac{C_i - C_{\alpha_i}}{\|C_i - C_{\alpha_i}\|}$ $Q_i^\top \frac{N_i - C_{\alpha_i}}{\|N_i - C_{\alpha_i}\|}$<br>$Q_i^\top \frac{O_i - C_{\alpha_i}}{\|O_i - C_{\alpha_i}\|}$ | $Q_i^\top \frac{C_j - C_{\alpha_i}}{\|C_j - C_{\alpha_i}\|}$ $Q_i^\top \frac{N_j - C_{\alpha_i}}{\|N_j - C_{\alpha_i}\|}$<br>$Q_i^\top \frac{O_j - C_{\alpha_i}}{\|O_j - C_{\alpha_i}\|}$ $Q_i^\top \frac{C_{\alpha_j} - C_{\alpha_i}}{\|C_{\alpha_j} - C_{\alpha_i}\|}$ |
| Orientation | — | $q\left(Q_i^\top Q_j\right)$ |

Figure A1: Node and edge features for single or pairwise residues, all invariant to rotation and translation. The definitions of bond angles ($\alpha_i$, $\beta_i$, $\gamma$), dihedral angles ($\psi_i$, $\phi_i$, $\omega_i$), and local coordinates frame $Q_i$ refer to (Gao et al. 2022).

## C. Illustration of Edge Relations

We provide a schematic of edges relations in Fig. A2, where each edge $e_{i,j} \in \mathcal{E}$ is associated with a set of relations $R_{i,j} \in \mathcal{R}$. Besides, two relations $r_1$ (with sequence distance equal to 1) can derive a relation $r_2$ (with sequence distance equal to 2). In addition, an edge may connect two nodes (residues) due to both relations $r_3$ and $r_4$. We strongly encourage readers to combine Fig. A2 with the descriptions in the paper to understand the definition of edge relations.

## D. Feature/Coordinate Updating of Global Nodes

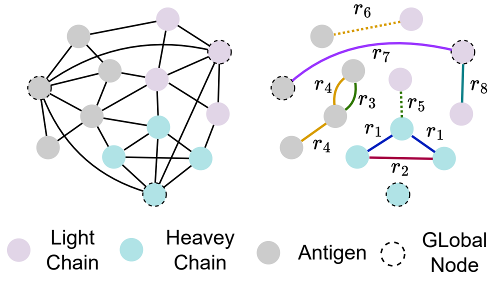


Figure A2: Illustration of edges relations. To avoid overly complexity, we visualize only some edges.

While node features and coordinates of global nodes can be updated as defined in Eq. (5)(7), we further provide their detailed derivations as below. The node feature of the global node $g_k$ $(1 \leq k \leq 3)$ of the $k$-th context component can be updated by local and global interactions, as follows:

$$\begin{aligned}\mathbf{h}_{g_k}^{(l+1)} = \phi_h\Big(\mathbf{h}_{g_k}^{(l)}, &\underbrace{\boldsymbol{W}_{r_7}^{(l)} \sum_{g_l \neq g_k} \alpha_{g_k,g_l}^{(r_7)} \mathbf{m}_{g_k,g_l,r_7}^{(l)}}_{\text{global interaction}} \\ &+ \underbrace{\boldsymbol{W}_{r_8}^{(l)} \sum_{\Phi(j)=g_k} \alpha_{g_k,j}^{(r_8)} \mathbf{m}_{g_k,j,r_8}^{(l)}}_{\text{local interaction}}\Big)\end{aligned} \tag{A.1}$$

where $\alpha_{g_k,g_l,r_7}^{(l)}$ and $\alpha_{g_k,j,r_8}^{(l)}$ are attention weights as calculated by Eq. (5). Besides, $\Phi(j) = g_k$ is the indicator function that determines whether node $v_j \in \mathcal{V}$ and $v_{g_k}$ belong to the same component. The node coordinate of the global node $g_k$ is updated by local and global interactions, as follows:

$$\begin{aligned}\mathbf{X}_{g_k}^{(l+1)} = \mathbf{X}_{g_k}^{(l)} + \frac{1}{2}\Big( &\underbrace{\sum_{g_l \neq g_k} \alpha_{g_k,g_l}^{(r_7)} \mathbf{X}_{g_k,g_l}^{(l)} \phi_z^{(r_7)}(\mathbf{m}_{g_k,g_l,r_7}^{(l)})}_{\text{global interaction}} \\ &+ \underbrace{\sum_{\Phi(j)=g_k} \alpha_{g_k,j}^{(r_8)} \mathbf{X}_{g_k,j}^{(l)} \phi_z^{(r_8)}(\mathbf{m}_{g_k,j,r_8}^{(l)})}_{\text{local interaction}}\Big)\end{aligned} \tag{A.2}$$

## E. On the proof of RA-EGN E(3)-Equivariance

**Theorem 1.** *We denote the $l$-th RA-EGN layer $g^{(l)}(\cdot)$ in our framework as $(\mathbf{H}^{(l+1)}, \mathbf{X}^{(l+1)}, \mathbf{E}^{(l+1)}, R^{(l+1)}) = g^{(l)}(\mathbf{H}^{(l)}, \mathbf{X}^{(l)}, \mathbf{E}^{(l)}, R^{(l)})$. For any given orthogonal matrix $\mathbf{O} \in \mathbb{R}^{3\times 3}$ and translation matrix $\mathbf{t} \in \mathbb{R}^3$, $g^{(l)}(\cdot)$ is E(3)-equivariant, i.e., we have $(\mathbf{H}^{(l+1)}, \mathbf{O}\mathbf{X}^{(l+1)} + \mathbf{t}, \mathbf{E}^{(l+1)}, R^{(l+1)}) = g^{(l)}(\mathbf{H}^{(l)}, \mathbf{O}\mathbf{X}^{(l)} + \mathbf{t}, \mathbf{E}^{(l)}, R^{(l)})$.*

*Proof.* For any orthogonal matrix $\mathbf{O} \in \mathbb{R}^{3\times 3}$, translation matrix $\mathbf{t} \in \mathbb{R}^3$, and $\mathbf{X}_{i,j}^{(l)} = \mathbf{X}_i^{(l)} - \mathbf{X}_j^{(l)}$, We can derive the relative coordinates between $\mathbf{O}\mathbf{X}_i^{(l)} + \mathbf{t}$ and $\mathbf{O}\mathbf{X}_j^{(l)} + \mathbf{t}$ as

$$(\mathbf{O}\mathbf{X}_i^{(l)} + \mathbf{t}) - (\mathbf{O}\mathbf{X}_j^{(l)} + \mathbf{t}) = \mathbf{O}\mathbf{X}_i^{(l)} - \mathbf{X}_j^{(l)}) = \mathbf{O}\mathbf{X}_{i,j}^{(l)} \tag{A.3}$$

Table A1: A comparison of various antibody design methods in three aspects: (1) Contextual data, where *"FR"* refers to *"Framework Region"*; (2) Architecture design, where *"Geometry"* refers to the scale of the involved geometric information; and (3) Optimization strategy, where *"Decoding"* refers to whether CDRs are generated in an autoregressive or full-shot fashion, *"No Iterative"* refers to whether iterative refinement or diffusion is used, and *"Pre-train"* refers to the effects of pre-training.

| **Method** | **Context** | | | **Architecture** | | **Optimization** | | |
|---|---|---|---|---|---|---|---|---|
| | **FR** | **Light Chain** | **Antigen** | **Geometry** | **Equivariant** | **Decoding** | **No Iterative** | **Pre-train** |
| RefineGNN (ICLR'22) | ✓ | ✗ | ✗ | Backbone | ✗ | Autoregressive | ✗ | ✗ |
| HSRN (ICML'22) | ✗ | ✗ | Epitope | Full-atom | ✓ | Autoregressive | ✗ | ✗ |
| DiffAb (NIPS'22) | ✓ | ✗ | Full | Residue (C-$\alpha$) | ✓ | Full-shot | ✗ | ✗ |
| MEAN (ICLR'23) | ✓ | ✓ | Epitope | Backbone | ✓ | Full-shot | ✗ | ✗ |
| dyMEAN (ICML'23) | ✓ | ✓ | Epitope | Full-atom | ✓ | Full-shot | ✗ | ✗ |
| ABGNN (KDD'23) | ✓ | ✗ | Epitope | Full-atom | ✓ | Full-shot | ✗ | ✓ |
| HTP (NeurIPS'23) | ✓ | ✓ | Epitope | Residue (C-$\alpha$) | ✓ | Full-shot | ✓ | ✓ |
| ADsigner (AAAI'24) | ✓ | ✓ | Epitope | Backbone | ✗ | Full-shot | ✓ | ✗ |
| RAAD (ours) | ✓ | ✓ | Full | Backbone | ✓ | Full-shot | ✓ | ✓ |

The message aggregation of Eq. (3) can be re-writed as

$$\begin{aligned}
&\phi_m\Big(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \frac{(\mathbf{OX}_{i,j}^{(l)})^\top \mathbf{OX}_{i,j}^{(l)}}{\|(\mathbf{OX}_{i,j}^{(l)})^\top \mathbf{OX}_{i,j}^{(l)}\|_F}, \mathbf{E}_{i,j}^{(l)}\Big) \\
&= \phi_m\Big(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \frac{(\mathbf{X}_{i,j}^{(l)})^\top \mathbf{O}^\top \mathbf{OX}_{i,j}^{(l)}}{\|(\mathbf{X}_{i,j}^{(l)})^\top \mathbf{O}^\top \mathbf{OX}_{i,j}^{(l)}\|_F}, \mathbf{E}_{i,j}^{(l)}\Big) \qquad \text{(A.4)} \\
&= \phi_m\Big(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \frac{(\mathbf{X}_{i,j}^{(l)})^\top \mathbf{X}_{i,j}^{(l)}}{\|(\mathbf{X}_{i,j}^{(l)})^\top \mathbf{X}_{i,j}^{(l)}\|_F}, \mathbf{E}_{i,j}^{(l)}\Big) = \mathbf{m}_{i,j,r}^{(l)}
\end{aligned}$$

Therefore, the message aggregation is $E(3)$-invariant. Since the input node feature $\mathbf{h}_i^{(l)}$, edge feature $\mathbf{E}_{i,j}^{(l)}$, and message $\mathbf{m}_{i,j,r}^{(l)}$ are all $E(3)$-invariant, it is easy to derive that updated node features, edge features and edge relations in the $l$-th RA-EDGN layer are also $E(3)$-invariant.

Following that, we can prove that the node coordinates are also $E(3)$- equivariant, as follows:

$$\begin{aligned}
&(\mathbf{OX}_i^{(l)} + \mathbf{t}) + \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^{(r)}} \alpha_{i,j,r}^{(l)} \, (\mathbf{OX}_{i,j}^{(l)}) \, \phi_z^{(r)}(\mathbf{m}_{i,j,r}^{(l)}) \\
&= \mathbf{O}\Big(\mathbf{X}_i^{(l)} + \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^{(r)}} \alpha_{i,j,r}^{(l)} \, \mathbf{X}_{i,j}^{(l)} \, \phi_z^{(r)}(\mathbf{m}_{i,j,r}^{(l)})\Big) + \mathbf{t} \\
&= \mathbf{OX}_i^{(l+1)} + \mathbf{t}
\end{aligned} \qquad \text{(A.5)}$$

With the above derivations, it is easy to derive that

$$\begin{aligned}
&(\mathbf{H}^{(l+1)}, \mathbf{OX}^{(l+1)} + \mathbf{t}, \mathbf{E}^{(l+1)}, R^{(l+1)}) \\
&= g^{(l)}(\mathbf{H}^{(l)}, \mathbf{OX}^{(l)} + \mathbf{t}, \mathbf{E}^{(l)}, R^{(l)})
\end{aligned} \qquad \text{(A.6)}$$

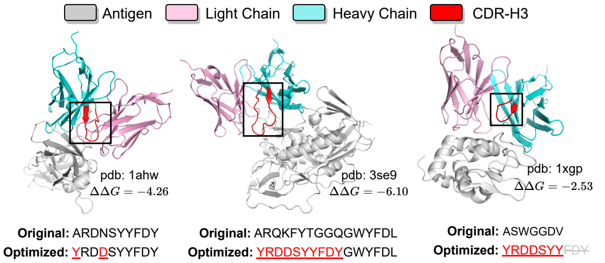


Figure A3: A case study of "universal" antibodies, in which optimized CDR sequences tailored for one antigen (pdb: 1ahw) can be used for other antigens (pdb: 3se9 and 1xgp) with some minor length adaptations (e.g., replacements and cropping) to improve their binding affinity, i.e., lower $\Delta\Delta G$.

For an $L$-layer RA-EGN encoder, we have

$$\begin{aligned}
&(\mathbf{H}^{(L)}, \mathbf{OX}^{(L)} + \mathbf{t}, \mathbf{E}^{(L)}, R^{(L)}) \\
&= g^{(L-1)}(\mathbf{H}^{(L-1)}, \mathbf{OX}^{(L-1)} + \mathbf{t}, \mathbf{E}^{(L-1)}, R^{(L-1)}) \\
&= g^{(L-2)}(\mathbf{H}^{(L-2)}, \mathbf{OX}^{(L-2)} + \mathbf{t}, \mathbf{E}^{(L-2)}, R^{(L-2)}) \\
&= \cdots \\
&= g^{(0)}(\mathbf{H}^{(0)}, \mathbf{OX}^{(0)} + \mathbf{t}, \mathbf{E}^{(0)}, R^{(0)})
\end{aligned} \qquad \text{(A.7)}$$

Since $\mathbf{H}^{(0)}$, $\mathbf{E}^{(0)}$, and $R^{(0)}$ are constructed based on the backbone coordinates $\mathbf{X}^{(0)}$ and keep $E(3)$-invariant, the resulting encoder is also $E(3)$-equivariant as a whole.

## F. Huber Loss Function

The huber loss (Huber 1992) helps to lead to a stable training procedure (Kong, Huang, and Liu 2022), defined as follows:

$$l(x, y) = \begin{cases} 0.5(x-y)^2, \text{ if } |x-y| < \delta \\ \delta \cdot (|x-y| - 0.5 \cdot \delta), \text{ else} \end{cases} \qquad \text{(A.8)}$$

where we set $\delta = 1$ in this paper.

Table A2: Performance comparison of models pre-trained with different data, including 1D protein sequences (UniRef), 1D antibody sequences (OAS), 3D protein structures (AlphaFoldDB), as well as hierarchical pre-training.

| **Pre-training** | | **Modeling** | | **Generation** | | | **Optimization** |
|---|---|---|---|---|---|---|---|
| Encoder | Pre-training Data | AAR (↑) | RMSD (↓) | AAR (↑) | TM-score (↑) | RMSD (↓) | SP-score (↑) |
| no pre-training | | 67.64 | 0.772 | 41.26 | 0.9874 | 1.463 | 11.55 |
| ESM-1b | UniRef | 68.10 | 0.768 | 41.40 | 0.9866 | 1.475 | 10.94 |
| AbBERT | OAS | 68.32 | **0.738** | 41.78 | 0.9876 | 1.437 | 11.37 |
| GearNet-Edge | AlphaFoldDB | 66.38 | 0.761 | 40.19 | 0.9883 | 1.413 | **11.87** |
| ESM-1b | UniRef & OAS | **68.74** | 0.745 | **41.94** | **0.9890** | **1.394** | 11.17 |

## G. Iterative Target Augmentation (ITA)

The pseudo-code of ITA-based antibody specificity optimization is placed in Algorithm. 1.

Algorithm 1: Algorithm for ITA-based Optimization

```
Input: A set of antibodies $\mathcal{D}$ to be optimized, a antibody generator $\mathcal{F}_\Theta$, and a $\Delta\Delta G$ predictor $f$.
for each iteration do
    for each antibody $s$ in $\mathcal{D}$ do
        Remove CDRs from antibody $s$ and generate 50 candidate antibodies by generator $\mathcal{F}_\Theta$.
        Sort candidates by predictor $f$, update the candidate queue, and update the training set $\mathcal{Q}$.
    end for
    Sample a batch of high-affinity antibodies from $\mathcal{Q}$.
    Fine-tune the generator $\mathcal{F}_\Theta$ by minimizing the total loss $\mathcal{L}_{\text{opt}}$ in Eq. (12) by back-propagation.
end for
  Generate high-affinity antibodies with generator $\mathcal{F}_{\Theta_*}$.
return Fine-tuned antibody generator $\mathcal{F}_{\Theta_*}$.
```

## H. Time Complexity Analysis

The time complexity of the relation-aware encoder (consisting of multiple layers of RA-EGNs) comes from five main components: (1) message aggregation $\mathcal{O}\big(|\mathcal{E}|(d_nF + d_eF)\big)$ , where $|\mathcal{E}|$ is the number of edges, $d_n$ and $d_e$ are feature dimension of nodes and edges, and $F$ is the hidden dimension; (2) node relation updating $\mathcal{O}(|\mathcal{E}|F)$; (3) node feature updating $\mathcal{O}\big(|\mathcal{R}| \cdot |\mathcal{V}|F^2 + |\mathcal{R}| \cdot |\mathcal{E}|(d_n + F)\big)$, where $|\mathcal{V}|$ and $|\mathcal{R}|$ is the number of nodes and relations; (4) edge feature updating $\mathcal{O}\big(|\mathcal{E}|(d_nF + d_eF)\big)$; (5) node coordinate updating $\mathcal{O}(|\mathcal{R}| \cdot |\mathcal{E}|F)$. The total time complexity is $\mathcal{O}\Big(|\mathcal{E}|(d_n + d_e)F + |\mathcal{R}| \cdot \big(|\mathcal{V}|F^2 + |\mathcal{E}|(d_n + F)\big)\Big)$ is linear w.r.t the number of nodes $|\mathcal{V}|$ and edges $|\mathcal{E}|$. In this paper, we define a total of 8 edge relations, i.e., $|\mathcal{R}| = 8 \ll |\mathcal{V}|$, which is almost negligible in the time complexity analysis.

## I. Implementation Details and Hyperparameters

The following hyperparameters are determined by an AutoML toolkit NNI with the hyperparameter search spaces as: Adam optimizer (Kingma and Ba 2014) with $lr = \{0.0005, 0.001\}$, batch size $B = \{4, 8, 16\}$, training epoch $E_{\text{train}} = 20$, fine-tuning epoch $E_{\text{fine}} = 20$, number of nearest neighbors $K = 8$ for graph construction, embedding size $F_e = 32$ for $E_{\text{type}}(v_i)$, hidden dimension $F = \{128, 256\}$, layer number $L = \{3, 4\}$, and loss weight $\eta = \{0.8, 1.0\}$ for $\mathcal{L}_{\text{struct}}$. In addition, $\phi_m(\cdot)$, $\phi_z^{(r)}(\cdot)$, $\phi_\omega^{(r)}(\cdot)$, $\phi_h(\cdot)$, and $\phi_e(\cdot)$ are all implemented as one- or two-layer MLPs with $\mathrm{SiLU}(\cdot)$ (Elfwing, Uchibe, and Doya 2018) as the activation function in this paper. For a fairer comparison, we select the mode checkpoint with the lowest loss on the validation set for testing. Moreover, the experiments on both baselines and our approach are implemented based on the standard implementation using the PyTorch 1.8.0 with Intel(R) Xeon(R) Gold 6240R @ 2.40GHz CPU and 8 NVIDIA A100 GPUs. For complex initialization, we remove the CDR $\mathcal{C}$ from the antibody, and then we use HDOCK (Yan et al. 2017) to dock it to the target antigen to obtain an antibody-antigen complex. Since the CDR $\mathcal{C}$ is unknown at first, we initialize its 3D coordinates according to the even distribution between the two residues right before and after the CDR.

## J. Effects of Pre-training

To explore how pre-training influences RAAD in antibody design, we consider four pre-training schemes: (1) ESM-1b (Rives et al. 2021) encoder pre-trained on the UniRef dataset (Suzek et al. 2007); (2) AbBERT (Gao et al. 2023) encoder pre-trained on the OAS dataset (Olsen, Boyles, and Deane 2022); (3) GearNet-edge (Zhang et al. 2022) pre-trained on the AlphaFoldDB dataset (Varadi et al. 2022); (4) ESM-1b encoder hierarchically pre-trained on the UniRef and OAS datasets (Wu and Li 2023). Following (Gao et al. 2023), we exploit the pre-trained encoders by concatenating the pre-trained and original node features, forwarding with a linear transformation, and taking the fused features as inputs. It can be observed from Table. A2 that (1) all kinds of pre-training strategies help to improve performance; (2) sequence pre-training helps more for sequence generation, while structure pre-training is more beneficial for structure prediction; (3) pre-training with antibody sequences helps much more than general protein sequences, but combining the two for hierarchical pre-training outperforms the both.