

## 434 A Appendix

### 435 A.1 SE(3)

436 The collection of  $4 \times 4$  real matrices of the SE(3) is shown as:

$$\begin{bmatrix} R & \mathbf{t} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (11)$$

437 where  $R \in \text{SO}(3)$  and  $\mathbf{t} \in \mathbb{R}^3$ ,  $\text{SO}(3)$  is the 3D rotation group.  $R$  satisfying  $R^T R = I$  and  
 438  $\det(R) = 1$ .

### 439 A.2 Details of Model Architecture

440 As stated in Sec. 3.3 on sequence-structure graph convolution,  $l$  is set to be a constant number 11. We  
 441 increase the predefined radius  $r$  to  $2r$  after one pooling layer, and the number of feature channels for  
 442 node embeddings is also doubled. We use a Leaky ReLU function [13] as the activation  $\sigma(\cdot)$  in the  
 443 message passing layers.

444 We design the sequential and radius graph instead of the  $k$ -nearest neighbour graph because a constant  
 445  $k$  make some neighbor nodes far away from the center node. As shown in Figure 7, the distances  
 446 of a group of neighbor nodes ( $\|P_{i,C\alpha} - P_{j,C\alpha}\|$ ) are larger than  $20 \text{ \AA}$ , which cannot be seen as  
 447 contacts [9]. Therefore, the radius is initially set to 4, enlarging to 16 in deeper layers. There are  
 448 four message passing and pooling layers. In this condition, when the number of nodes decreases,  $l$  is  
 449 constant,  $r$  increases, neighbours of center nodes gradually cover more distant nodes.

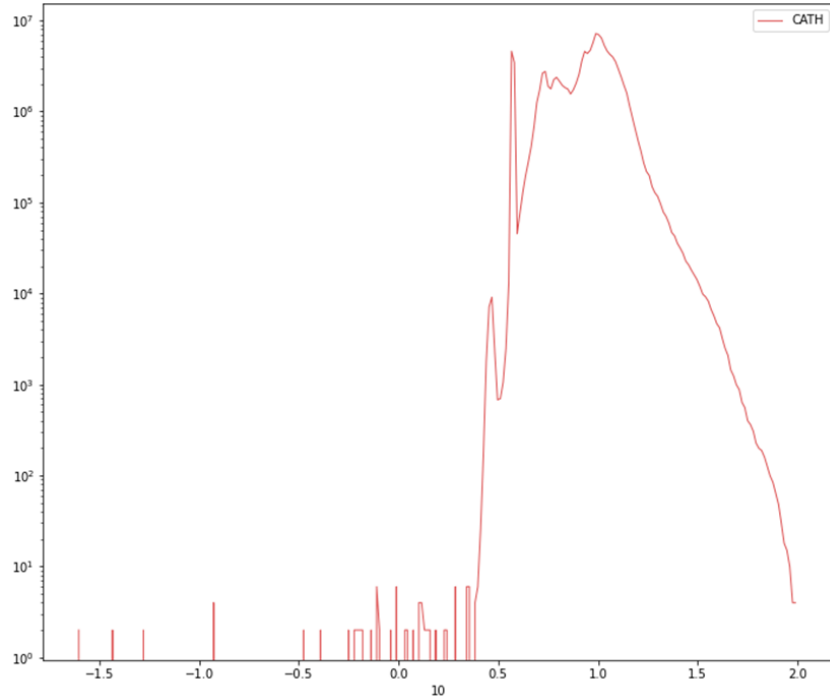


Figure 7: The histogram of distance statistics of  $k = 30$  nearest neighbor nodes of a protein dataset (CATH [29]). The horizontal axis denotes the distance in terms of exponents of 10, and the vertical axis represents the number of neighbor nodes with this distance.

Table 4: Dataset statistics. # X means the number of X.

Dataset	# Train	# Validation	# Test
Enzyme Commission	15, 550	1, 729	1, 919
Gene Ontology	29, 898	3, 322	3, 415
Fold Classification - Fold	12, 312	736	718
Fold Classification - Superfamily	12, 312	736	1, 254
Fold Classification - Family	12, 312	736	1, 272
Reaction Classification	29, 215	2, 562	5, 651

### A.3 Details of Datasets and Training Setup

For all datasets, we use a data augmentation strategy by adding noise for the training set to increase the variability of data. For example, we update the position of  $C_{\alpha i}$ ,

$$P_{i,C_\alpha} \leftarrow P_{i,C_\alpha} + N(\mu_N, \sigma_N^2) \quad (12)$$

where  $\mu_N, \sigma_N^2$  are the mean (expectation) and variance of the normal distribution  $N$ , which are set to 0 and 0.1 in experiments. Dataset statistics [53] of our four downstream tasks are summarized in Table 4.

**Settings** The proposed models are conducted on a single NVIDIA-SMI A100 GPU, through PyTorch 1.13+cu117 and PyTorch Geometric 2.3.1 with CUDA 11.2. The number of the initial feature channels is 256. The learning rate is set to 0.001. More details about implementation is shown in Table 5.

Table 5: More details of training setup

Hyper-parameter	Fold	Enzyme Reaction	GO	EC
Batch size	4	4	24	64
Epoch	400	400	500	500

### A.4 Evaluation Metric $F_{\max}$

$F_{\max}$  is calculated by first determining the precision and recall for each protein, then averaging these results over all proteins [53, 15, 19].  $p_i^j$  is the prediction probability for the  $j$ -th class of the  $i$ -th protein, given the decision threshold  $t \in [0, 1]$ , the precision and call are give as:

$$\text{precision}_i(t) = \frac{\sum_j \mathbb{I}[(p_i^j \geq t) \cap b_i^j]}{\sum_j \mathbb{I}[p_i^j \geq t]}, \quad \text{recall}_i(t) = \frac{\sum_j \mathbb{I}[(p_i^j \geq t) \cap b_i^j]}{\sum_j b_i^j}$$

where  $b_i^j \in \{0, 1\}$  is the corresponding binary class label, and  $\mathbb{I} \in \{0, 1\}$  is an indicator function. If there are  $N$  proteins in total, then the average precision and recall are defined as:

$$\text{precision}(t) = \frac{\sum_i^N \text{precision}_i(t)}{\sum_i^N \left( \left( \sum_j (p_i^j \geq t) \right) \geq 1 \right)}, \quad \text{recall}(t) = \frac{\sum_i^N \text{recall}_i(t)}{N}$$

Finally,  $F_{\max}$  is defined as the maximum value of F-score over all thresholds,

$$F_{\max} = \max_t \left\{ \frac{2 \cdot \text{precision}(t) \cdot \text{recall}(t)}{\text{precision}(t) + \text{recall}(t)} \right\} \quad (13)$$

### A.5 More Results of GO Term Prediction

For GO term prediction, we also apply different cutoff splits. Proteins in the test set are categorized into five groups based on their similarity to the training set ( 30%, 40%, 50%, 70%, and 95%). As

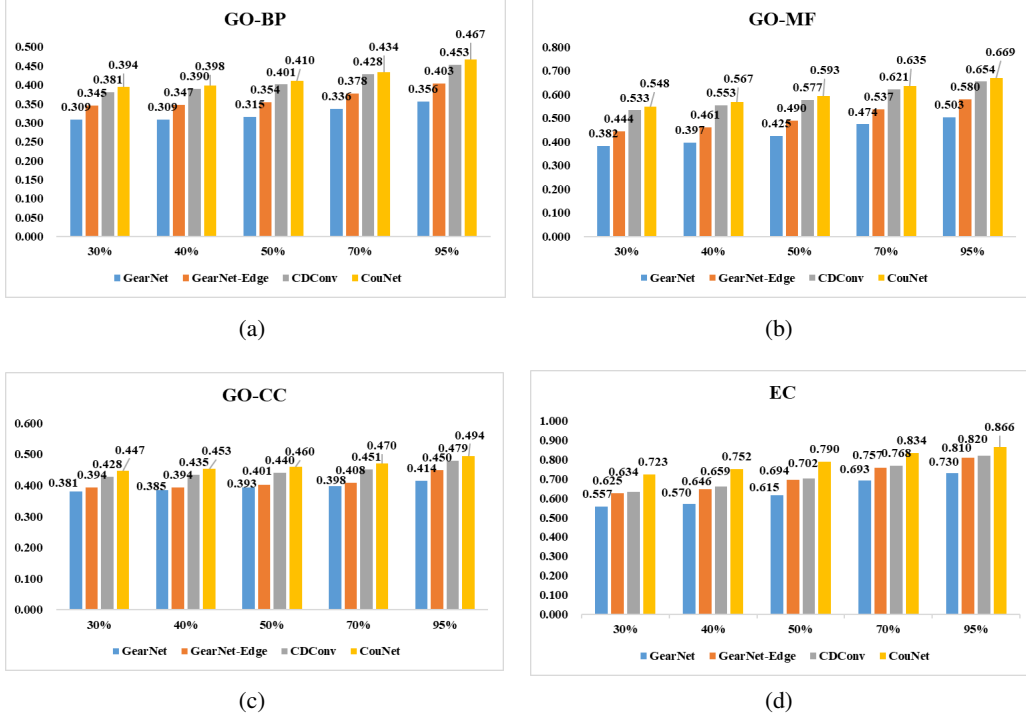


Figure 8:  $F_{\max}$  on GO term and EC number prediction under different cutoffs.

shown in Figure 6, the results of GO term prediction are presented in Figure 8(a)-(c). The proposed model CoupleNet achieves the highest  $F_{\max}$  scores across all cutoffs on these tasks. Even when there is a low similarity between the training and test sets, our model also has higher scores, which demonstrates the superiority and robustness of the proposed model.

## A.6 Completeness Analysis

Given a protein 3D graph  $G = (\mathcal{V}, \mathcal{E}, \mathcal{P})$ , we capture the geometric representations based on the atoms' 3D positions and use sequential and structural representations as the node and edge features. For a 3D structure, based on the definition of completeness in Sec. 3.1 and the rigorously demonstrated method to show the calculated geometries can achieve completeness for structures [47], we guarantee the completeness of the selected geometric representations at the base and backbone levels of structures.

The geometric representations are SE(3) invariant (distances, angles) and SE(3) equivariant (directions, orientations). Therefore, it is natural for Eq. 3 to hold from right to left. To demonstrate Eq. 3 holding from left to right, we need to show  $\mathcal{F}(G) \Rightarrow T_g(\mathcal{P})$ , where  $T_g$  does not change the 3D conformation of a 3D graph. Thus we need to show positions can be determined by  $\mathcal{F}(G)$ .

The base approach CoupleNet<sub>aa</sub> only considers the  $C_{\alpha}$  coordinates and constructs LCS for each residue.  $\mathcal{F}(G)_{aa}$  provides complete representations. First, when  $n = 1$ , it holds. Assume the case  $n = k$  holds such that  $\mathcal{F}(G)_{aa}$  is complete. Then we need to prove the case  $n = k + 1$  still holds. This is obvious because if  $v_j$  is the  $(k + 1)$ -th node connected to node  $v_i$  among the existing  $k$  nodes, the LCS  $Q_j$  can be easily obtained from  $Q_i$  and  $\mathcal{F}(G)_{aa}$ .

When considering the backbone atoms  $C_{\alpha}, C, N, O$ ,  $\mathcal{F}(G)_{aa}$  is complete. As shown in Figure 3, the remaining degree of freedom at the backbone level is the rotation angles  $\Phi, \Psi, \Omega$  based on the rigid bond lengths and angles. Such backbone torsion angles are calculated and concatenated with  $x_{i,aa}$  into  $x_i$ . Besides, for any residues  $i$  and  $j$ , the calculated six inter-residue geometries fully define the relative locations of backbone atoms. Therefore, there are no other remaining degrees of freedom. Consequently, the obtained geometric representations at the backbone level are complete.

Table 6: More Ablation of our proposed method

Method	Fold Classification			Enzyme Reaction	GO			EC
	Fold	Superfamily	Family		BP	MF	CC	
CoupleNet	60.6	82.1	99.7	89.0	0.467	0.669	0.494	0.866
w/o sequence	60.0	81.6	99.6	88.4	0.441	0.650	0.456	0.700
w/o structure	26.1	36.4	92.9	81.3	0.406	0.586	0.427	0.625

## A.7 More Results of Ablation Study

Table 3 presents an ablation study of the proposed CoupleNet model. Apart from removing  $\Phi, \Psi, \Omega$  or  $d, \omega, \theta, \varphi$  and using the base model CoupleNet<sub>aa</sub>, we conduct more ablation experiments on the four tasks. The results are shown in Table 6.

Compared with the full model, we consider removing either the sequence or structure information to analyze their importance. Removing the sequence information means removing the encoding of amino acid types for each node. Removing the structure information means removing features related to protein geometry ( $\mathcal{F}(G)_{aa}, \Phi, \Psi, \Omega, d, \omega, \theta, \varphi$ , and we omit related subscripts for brevity).

As shown in Table 6, removing either sequence or structure causes a performance drop on all tasks, demonstrating that both types of information are critical for the proposed method. When removing the structure, the performance decreases more significantly, suggesting that structural information provides more important and comprehensive clues compared with sequence information alone. Combining these diverse data sources leads to optimal predictive performance.