GLProtein: Global-and-Local Structure Aware Protein Representation Learning

Yunqing Liu[†] **Wenqi Fan**[*][†][‡] **Xiao-Yong Wei**[†][§] **Qing Li**[†]
[†]Department of Computing (COMP), The Hong Kong Polytechnic University
[‡]Department of Management and Marketing (MM), The Hong Kong Polytechnic University
[§]Department of Computer Science, Sichuan University
yunqing617.liu@connect.polyu.hk
wenqifan03@gmail.com, {cs007.wei, csqli}@comp.polyu.edu.hk

Abstract

Proteins are central to biological systems, participating as building blocks across all forms of life. Despite advancements in understanding protein functions through protein sequence analysis, there remains potential for further exploration in integrating protein structural information. We argue that the structural information of proteins is not only limited to their 3D information but also encompasses information from amino acid molecules (local information) to protein-protein structure similarity (global information). To address this, we propose **GLProtein**, the first framework in protein pre-training that incorporates both global structural similarity and local amino acid details to enhance prediction accuracy and functional insights. GLProtein innovatively combines protein-masked modelling with triplet structure similarity scoring, protein 3D distance encoding and substructure-based amino acid molecule encoding. Experimental results demonstrate that GLProtein outperforms previous methods in several bioinformatics tasks, including predicting protein-protein interactions, contact prediction, and so on.

1 Introduction

Proteins are fundamental to virtually every biological process, serving as the building blocks for cells and organs and acting as catalysts, messengers, and structural elements in all life forms. Understanding the structure and function of proteins is crucial for advances in health, agriculture, and environmental science, making protein research a cornerstone of biotechnology and medicinal science (Ding et al., 2019; Davis et al., 2024; Zhao et al., 2024). Recognizing the critical role of proteins in various scientific fields, many efforts have been made to design computational methods to further understand these crucial molecules (Weng et al., 2021; Zhao et al., 2020). Particularly, protein representation learning, as one significant part, involves capturing the complex features and relationships within proteins in a condensed form that can be utilized for various computational tasks and analyses. It is crucial for enhancing the understanding of protein structures and functions, improving predictive modelling in bioinformatics, facilitating the drug discovery process, and advancing our knowledge of biological systems through interpretable and efficient representations of proteins (Somnath et al., 2021; Liu et al., 2023; Gao et al., 2024).

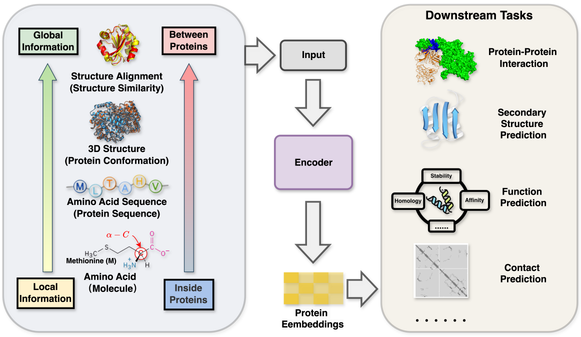


Figure 1: An illustration on protein representation learning flow. Protein information from local information (inside proteins) to global information (between proteins) can be used as input. This input undergoes encoding by a protein encoder to generate a protein representation across various downstream tasks.

In recent years, the success of language models in natural language processing (NLP) has paved the way for innovative approaches in bioinformatics areas, such as protein modeling (Xiao et al., 2021; Chowdhury et al., 2022), protein generation (Madani et al., 2020; Ferruz et al., 2022), and protein-protein interaction prediction (Wang et al., 2019; Ofer et al., 2021). To be specific, by treating protein sequences as linguistic strings, these models have demonstrated remarkable effectiveness in predicting protein function based on sequence data alone. Technically, as shown in Figure 1, protein sequences (e.g., the amino acid sequence '*MLTAHV...*') are treated as sentences in natural language and amino acids (e.g., 'M', 'L', and 'T')

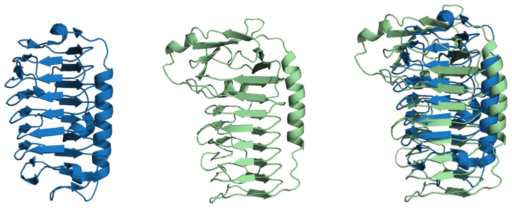
---

[*] Corresponding author

Findings of the Association for Computational Linguistics: EMNLP 2025, pages 4355–4372
November 4-9, 2025

resemble words. Thus, Leveraging the powerful BERT architecture originally developed for natural language, ProtBert (Elnaggar et al., 2021) adeptly adapts the BERT (Devlin et al., 2018) masked language modelling framework to the field of bioinformatics. This analogy allows ProtBert to employ the technique of predicting randomly masked elements in sequences, thereby learning to identify complex patterns and dependencies among amino acids. Similar to ProtBert, ESM (Rives et al., 2021; Verkuil et al., 2022; Hie et al., 2022) extends this paradigm by employing a more refined Transformer-based architecture, focusing on capturing the evolutionary relationships and functional dynamics within protein sequences. In other words, most existing protein modelling methods aim to perform protein representation learning by encoding the protein's sequence information for various downstream applications, such as amino acid contact prediction (Singh et al., 2022), protein homology detection (Kaminski et al., 2023), protein stability prediction (Chu et al., 2024), protein-protein interaction identification (Wang et al., 2019; Ofer et al., 2021), etc.

Despite the aforementioned successes, most existing protein language modelling methods suffer from intrinsic limitations. Specifically, most of their focuses have primarily been on the amino acid sequence, often neglecting the crucial aspects of protein structure. Proteins possess the ability to fold into diverse 3D shapes, interacting with various proteins and small molecules in biologically significant ways (Jumper et al., 2021; Mirdita et al., 2022; Tsaban et al., 2022). Since *protein's structure determines function* (Greslehner, 2018), utilizing protein 3D structure information effectively is crucial for protein language modelling, in which many studies have demonstrated the potential of pre-training on experimentally determined protein structures (Hermosilla and Ropinski, 2022; Su et al., 2023; Wang et al., 2022; Zhang et al., 2022). Nevertheless, these methods focus only on the structure within proteins and ignore the global similarities between proteins. We emphasize that the information on protein structure is not only limited to its structure (i.e., conformation) in 3D space but also includes information ranging from local amino acid molecules to the global structural similarity between proteins, as shown in Figure 1. Local information involves the detailed properties and orientations of individual amino acids, which can affect protein stability and biochemical activity (Renaud et al., 2021). These specifics are vital as they demonstrate how modifications or mutations at the amino acid level can alter the overall structure and functionality of the protein (Jumper et al., 2021). Furthermore, protein structure similarities provide information on evolutionary relationships and functional classes, which are crucial for understanding how structurally similar proteins of different species can perform similar or complementary functions within biological systems (Hamamsy et al., 2023). For example, as shown in Figure 2, the bacterial ice-binding protein FfIBP and the diatom adhesion protein CaTrailin_4 exhibit no detectable sequence similarity despite their functional similarities (Zackova Suchanova et al., 2023; Al-Fatlawi et al., 2023). Their predicted structures exhibit a remarkable similarity (TM-Score = 0.6), with both proteins adopting a beta-helical fold comprised of two units linked by an alpha helix. This structural topology is characteristic of ice-binding proteins. Such comparisons are key to predicting the functions of newly discovered proteins based on known structures, thereby enhancing our grasp of complex biological processes and interactions (Lipman and Pearson, 1985; Hamamsy et al., 2022, 2023). However, most existing approaches have ineffectively incorporated amino acid molecule information and protein structural similarities into protein representation learning.



(a) FfIBP (b) CaTrailin_4 (c) Alignment

Figure 2: An example of protein structure similarity. Given the predictive structures of a protein pair: (a) the bacterial ice-binding protein FfIBP and (b) the diatom adhesion protein CaTrailin_4 (Zackova Suchanova et al., 2023; Al-Fatlawi et al., 2023), (c) is FfIBP (blue) and CaTrailin_4 (green) structure alignment.

To eliminate these limitations, we propose a novel protein pre-training framework **GLProtein** with **G**lobal-and-**L**ocal **Protein** structure information for protein representation learning. Our major contributions are summarized as follows:

- We introduce a principled approach for capturing protein structural characteristics in a thorough and detailed manner. This approach incorporates a holistic view of protein structure data, encompassing global structural information, protein structure similarities, as well as local structure information such as protein 3D distance encoding and substructure-based molecular encoding. To the best of our knowledge, we are the first to investigate global and local protein structure information in protein language modelling.

- We propose a novel protein pre-training framework (GLProtein), where protein structure information is incorporated into protein language models for enhancing protein representation learning.
- The comprehensive experiments demonstrate the effectiveness of the proposed method on a wide range of downstream tasks, which verify the performance superiority of GLProtein.

2 Related Work

Protein Langauge Modelling. As an approach to protein representation learning, protein language modelling is a burgeoning field at the intersection of computational biology and natural language processing (NLP) (Hu et al., 2023; Li et al., 2024b,c,a). Inspired by the success of language models in NLP, researchers have adapted these techniques to analyse and predict the properties of protein sequences (Fan et al., 2025). Recent advancements have been dominated by the application of transformer-based models, which utilise self-attention mechanisms to capture relationships between amino acids in a sequence. ProtTrans (Elnaggar et al., 2021) and ESM (Beal, 2015; Verkuil et al., 2022; Hie et al., 2022), trained on large-scale protein databases, have shown remarkable ability in tasks such as protein classification and interaction prediction. Moreover, OntoProtein (Zhang et al.) and KeAP (Zhou et al., 2023a) incorporated external biological knowledge to enrich protein representations and enhance performance on various downstream tasks. However, most of these protein language models do not explicitly consider the spatial structure of proteins and structural similarities between proteins, like our proposed approach.

Protein Structure Modelling. The structure of a protein determines its functions. Thus, protein structure modelling has been treated as a reliable way to improve protein representation learning (Huang et al., 2024; AlQuraishi, 2021; Torrisi et al., 2020; Cheng et al., 2008). Some methods use Graph Neural Networks (GNNs) to handle the complex, non-linear relationships inherent in protein structure (Liao et al., 2023; Jha et al., 2022; Réau et al., 2023; Xu and Bonvin, 2024; Zhou et al., 2023b, 2024; Li et al., 2025). Moreover, RGN2 (Chowdhury et al., 2022) utilized a protein language model to learn structural information from unaligned protein sequences. GearNet (Zhang et al., 2022) focused on geometric pertaining and learned protein features by utilizing spatial relationships between amino acids. SaProt (Su et al., 2023) introduced the concept of a "structure-aware vocabulary" to integrate residue tokens with structure tokens. Similar to the knowledge hancing method, PST (Chen et al., 2024) enhances protein language models by integrating structural information through graph transformers to incorporate structural data. Unlike these models, we propose global structure learning and local structure learning methods, which could not only integrate protein structure information and amino acid information but also learn the structure similarity between different proteins by using TM-Score (Hamamsy et al., 2023).

3 Methodology

In this section, we aim to introduce our proposed framework (GLProtein) as a novel solution to learn global and local protein structure information for protein representation learning. We develop GLProtein that incorporates both global and local protein structure information into protein representation learning. The framework of GLProtein, shown in Figure 3, consists of three components: *protein language modelling* (Section 3.1), *global structure information modelling* (Section 3.2), and *local structure information modelling* (Section 3.3).

3.1 Protein Language Modelling

As shown in the center part of Figure 3, protein language modelling forms the backbone of our proposed framework, which aims to learn protein representation. We adopt a masking strategy that each masked amino acid has an 80% probability of being masked for prediction, a 10% chance of being replaced with a random amino acid, and a 10% chance of remaining unchanged. We then integrate protein 3D distance encoding and substructure-based molecular encoding into a protein decoder, in which we will detail in the local structure information modelling component. Suppose that the number of masked amino acids is M and x_i denotes the i-th amino acid. $x_{\sim i}$ denotes the sequence of amino

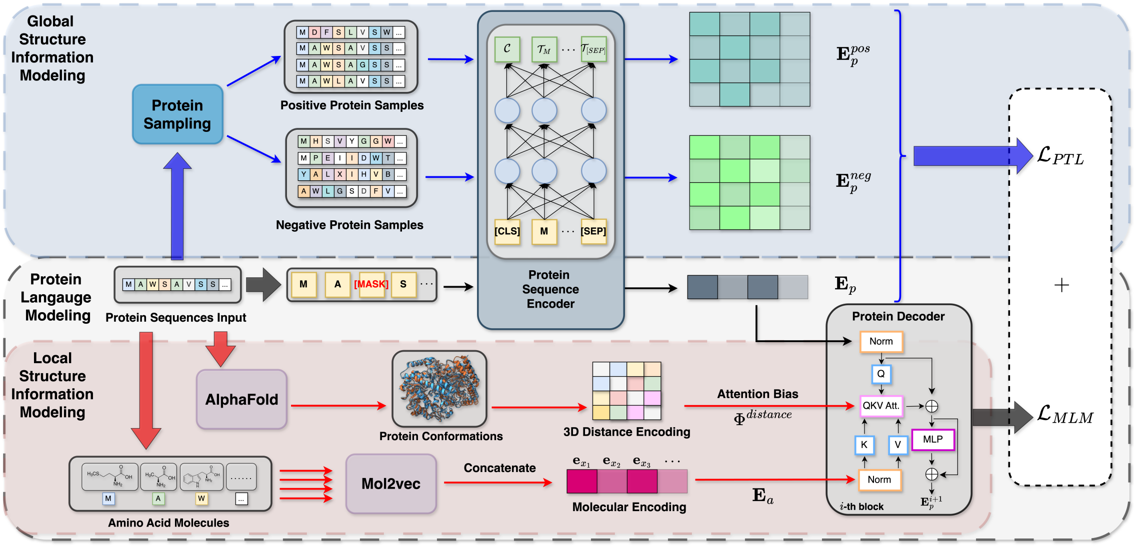


Figure 3: Overview of our proposed model, which jointly optimises global protein similarities and masked protein model with local structure information.

acids excluding the masked amino acid at position i. We leverage a cross-entropy loss $\mathcal{L}_{MLM}$ to estimate masked amino acids. Formally, the masked protein modelling objective can be defined as:

$$\mathcal{L}_{MLM} = -\log \sum_{i \in M} P(x_i | x_{\sim i}; \theta_E, \theta_D), \qquad (1)$$

where θ_E and θ_D denote the parameters of the protein sequence encoder and decoder, respectively. We initialise with a pre-trained BERT-like encoder: ProtBert (Elnaggar et al., 2021).

3.2 Global Structure Information Modelling

Protein structures encompass more than mere 3D spatial configurations; they also include global structural information that reflects similarities among proteins. To address this, we introduce the concept of global structure information, which contains the structure similarities between proteins, by leveraging the huge amount of self-supervised signals in protein sequences, as shown at the top of Figure 3. To be specific, given each input protein sequence, positive and negative protein sampling is designed to get the triplet (P, P_{pos}, P_{neg}) for capturing protein structure similarity features. Then, the protein triplets are encoded to protein representation for the calculation of the contrastive learning loss. This optimises the protein sequence encoder by bringing the representation of the input protein P and its positive samples P_{pos} closer together while pushing the representation of P and its negative samples P_{neg} further apart in the representation space.

Positive and Negative Protein Sampling. TM-score (Template Modeling Score) (Zhang and Skolnick, 2004; Xu and Zhang, 2010) is a widely used metric in structural biology for assessing the structural similarity between two protein structures. We utilize the TM-score to measure structural similarity between proteins, focusing on their overall global structure rather than mere sequence identity. Mathematically, the TM-score can be expressed as:

$$\text{TM-score} = max[\frac{1}{L_N} \sum_{i=1}^{L_r} \frac{1}{1 + (\frac{d_i}{d_0})^2}], \qquad (2)$$

where L_N is the length of the native structure, L_r is the length of the aligned residues to the template structures, d_i is the distance between the i-th pair of residues, and d_0 is a scaling factor.

We employ a two-pronged approach that utilizes the TM-Vec model (Hamamsy et al., 2023) to construct a robust set of positive and negative samples for our protein structure similarity analysis. For positive sample selection, we utilize the TM-Vec model to identify the top-K protein sequences that exhibit the highest TM-score values in relation to the template proteins.

In contrast, our negative sampling strategy employs a stochastic selection process followed by structural dissimilarity confirmation. Initially, we randomly select n proteins from our dataset. Subsequently, we employ the TM-Vec model to compute the TM-score between each selected protein and the template protein. Proteins with a TM-score $<$ 0.2 are classified as negative samples, as this threshold indicates a high degree of structural dissimilarity (Xu and Zhang, 2010).

Protein Triplet Modelling. After positive and negative protein sampling, we obtain the triplet

(P, P_{pos}, P_{neg}). Each protein in the triplet is passed to the protein sequence encoder, resulting in the protein representation, i.e., $\mathbf{E}_p \in \mathbb{R}^{L_p \times D}$, $\mathbf{E}_p^{pos} \in \mathbb{R}^{L_p \times D}$ and $\mathbf{E}_p^{neg} \in \mathbb{R}^{L_p \times D}$. L_p denotes the length of amino acid sequence and D stands for the feature dimension.

Since the task we focus on in this part is contrastive learning, the protein triplet loss is designed. This loss function operates by comparing three entities: anchor protein P, positive protein P_{pos} and negative protein P_{neg}. Thus, given protein representation triplet (P, P_{pos}, P_{neg}), the protein triplet loss $\mathcal{L}_{PTL}$ can be defined as:

$$\begin{aligned}&\mathcal{L}_{PTL}(P, P_{pos}, P_{neg}) = \\ &max(||\mathbf{E}_p - \mathbf{E}_p^{pos}||_2 - ||\mathbf{E}_p - \mathbf{E}_p^{neg}||_2 + \epsilon, 0),\end{aligned} \tag{3}$$

where $\mathbf{E}_p$, $\mathbf{E}_p^{pos}$, $\mathbf{E}_p^{neg} \in \mathbb{R}^{L_p \times D}$ are protein representation of the triplet (P, P_{pos}, P_{neg}). ϵ is a margin between positive and negative pairs.

3.3 Local Structure Information Modelling

While the global structure information modelling component is designed to identify structural similarities across different proteins, the local structure information modelling component zooms in on the specific, intricate features of a protein's internal structure, providing a more nuanced understanding. More specifically, in this part, we leverage the local structural details of proteins, including protein 3D distance encoding and substructure-based molecular encoding, to enhance the model's ability to learn and interpret this local configuration effectively, as shown at the bottom of Figure 3.

Protein 3D Distance Encoding. The 3D coordinates provide critical insights into how proteins fold and interact in three-dimensional space, influencing their stability, activity, and specificity (Liu et al., 2022; Peng et al., 2022; Su et al., 2023). We use AlphaFoldDB[1] as the 3D protein database and integrate the protein 3D distance encoding (Ying et al., 2021) to represent protein 3D structural information to ensure rotational and translational invariance. We propose to take advantage of the **alpha-carbon (α-C) coordinates** rather than the entire protein coordinates in protein representation learning. By capturing the backbone conformation, α-C coordinates effectively convey the protein's overall shape and folding pattern, which are critical for understanding its function. Moreover, leveraging α-C coordinates balances capturing essential structural information and maintaining computational efficiency.

Specifically, the coordinates of each α-C are processed to represent the current position of the amino acid in 3D space. Then, we encode the Euclidean distance metric to reflect the spatial relation between any pair of amino acids in the 3D space. Mathematically, given each amino acid pair (i, j), we first process their Euclidean distance with the Gaussian Basis Kernel function (Scholkopf et al., 1997), $\phi_{(i,j)}^k = -\frac{1}{\sqrt{2\pi}|\sigma^k|} \exp(-\frac{1}{2}(\frac{\gamma_{(i,j)}||\mathbf{r_i}-\mathbf{r_j}||+\beta_{(i,j)}-\mu^k}{|\sigma^k|})^2)$, where $k = 1, \ldots, K$. K is the number of Gaussian Basis kernels. Then, the 3D distance encoding can be calculated as follows:

$$\Phi_{(i,j)}^{distance} = GELU(\boldsymbol{\phi}_{(i,j)} \boldsymbol{W}_D^1) \boldsymbol{W}_D^2, \tag{4}$$

where $\boldsymbol{\phi}_{(i,j)} = [\phi_{(i,j)}^1; \ldots; \phi_{(i,j)}^K]^\top$. $\boldsymbol{W}_D^1 \in \mathbb{R}^{K \times K}$, $\boldsymbol{W}_D^2 \in \mathbb{R}^{K \times 1}$ are learnable parameters. $\gamma_{(i,j)}, \beta_{(i,j)}$ are learnable scalars indexed by the pair of amino acid types, and μ^k, σ^k are learnable kernel center and learnable scaling factor of the k-th Gaussian Basis Kernel. Denote $\Phi^{distance}$ as the matrix form of the 3D distance encoding, whose shape is $n \times n$.

Substructure-based Molecular Encoding. As more detailed information about protein localisation, amino acid molecules play a crucial role in protein representation learning, as they form the essential building blocks of proteins and provide the foundational data for understanding protein structure and function (Lieu et al., 2020; Lopez and Mohiuddin, 2024). To learn the **fine-grained amino acid structure** information, we introduce substructure-based molecular encoding to leverage the inherent relationships between molecule motifs and substructural features in amino acid molecules. In practice, we utilize the mol2vec (Jaeger et al., 2018) to process and derive representations for all amino acid molecules to obtain fine-grained molecular structure information. For protein P, we have

$$\mathbf{E}_a(P) = \text{Concat}(\mathbf{e}_{x_1}, \mathbf{e}_{x_2}, \ldots, \mathbf{e}_{x_i}, \ldots, \mathbf{e}_{x_L}),$$

where $\mathbf{e}_{x_i} \in \mathbb{R}^{1 \times d}$, L is the length of the protein sequence, $\mathbf{e}_{x_i}$ is the i-th amino acid molecule embedding, and d stands for the feature dimension of the amino acid molecule.

3.4 Model Training

In this part, we will first detail the protein decoder process, which combines protein language mod-

---

[1] https://alphafold.ebi.ac.uk/

elling and local structure information modelling components. Finally, the pre-training objective of the whole framework will be stated.

Protein Decoder. As shown in Figure 3, the decoder treats protein representation $\mathbf{E}_p$ as a query, while the substructure-based molecular encodings $\mathbf{E}_a$ are attended to as keys and values and protein 3D distance encoding $\Phi^{distance}$ is attended to as attention bias. Taking the i-th layer as an example, the inputs to the protein decoder include $\mathbf{E}_p^i$, $\Phi^{distance}$ and $\mathbf{E}_a$. The substructure-based molecular encoding $\mathbf{E}_a$ is firstly queries by $\mathbf{E}_p^i$ as the key and value:

$$\begin{aligned}Q_p^i &= \text{Norm}(\mathbf{E}_p^i)W_Q^i,\\ K_a^i &= \text{Norm}(\mathbf{E}_a)W_K^i,\\ V_a^i &= \text{Norm}(\mathbf{E}_a)W_V^i,\end{aligned}$$

where W_Q^i,W_K^i,W_V^i are learnable matrices. Norm stands for the layer normalization (Ba et al., 2016).

Then, Attention (Vaswani et al., 2017) is applied to $\{Q_p^i, K_a^i, V_a^i\}$, where the representation of protein sequence extracts helpful, relevant information from the substructure-based molecular encoding. The obtained representation o_p^i stores the helpful structure information for restoring missing amino aids. We then add up o_p^i and $\mathbf{E}_p^i$ to integrate information, resulting in the representation $\hat{\mathbf{E}}_p^i$ as follows:

$$\begin{aligned}o_p^i &= \text{Attention}(Q_p^i, K_a^i, V_a^i, \Phi^{distance}),\\ \hat{\mathbf{E}}_p^i &= \text{Norm}(\mathbf{E}_p^i) + o_p^i.\end{aligned}$$

The resulting representation $\hat{\mathbf{E}}_p^i$ integrates the helpful, relevant structure information that benefits the restoration of missing amino acids. We finally forward $\hat{\mathbf{E}}_p^i$ through a residual multi-layer perceptron to obtain the output representation of the i-th block, which also serves as the input to the $(i+1)$-th block.

Pre-training Objective. To estimate the model parameters of GLProtein, we adopt the masked protein modelling object and global protein triplet objective to construct the overall model. We jointly optimize the overall objective as follows:

$$\mathcal{L} = \mathcal{L}_{MLM} + \alpha\mathcal{L}_{PTL}, \tag{5}$$

where α is the hyper-parameter.

4 Experiments

In this section, we evaluate the generalization ability of the learned protein representation by fine-tuning the pre-trained model across a diverse array of downstream applications, including amino acid contact prediction, protein homology detection, protein stability prediction, protein-protein interaction identification, protein-protein binding affinity prediction and semantic similarity inference.

Pretraining Datasets. Swiss-Prot (Boeckmann et al., 2003) offers a comprehensive and manually curated protein sequence database that includes nearly 600k protein sequences. We use it as pertaining dataset. Additionally, we use AlphaFoldDB to obtain the protein 3D coordinate datasets.

Implementation Details. We conducted some experiments and compared GLProtein with baselines regarding pre-training and inference time in contact prediction tasks, as shown in Appendix Table 5. GLProtein outperforms baselines in multiple downstream tasks with similar parameters. During pre-training, GLProtein is trained for 300k steps using a learning rate of 1e-5, weight decay of 0.01 over four GPUs (NVIDIA A6000, 48G Memory each). We uniformly fine-tuned all downstream tasks without structural information to ensure fair and unbiased comparisons. For the amino acid contact prediction and protein-protein interaction task, we randomly selected five random seeds to fine-tune our model and the baseline model separately and report the results.

4.1 Downstream Tasks

4.1.1 Amino Acid Contact Prediction

Overview. Amino acid contact prediction is a critical task in computational biology, aiming to identify pairs of amino acids within a protein that are in close spatial proximity. Given an input protein sequence, our model predicts whether pairs of amino acids from the same sequence are in contact. The model accomplishes this by generating a probability contact matrix for each input protein. We tested the model on the dataset collected and organized by ProteinNet (AlQuraishi, 2019) and TAPE (Rao et al., 2019).

Baselines. We evaluate our model compared with ten baselines. Specifically, we employed variations of LSTM (Hochreiter and Schmidhuber, 1997), ResNet (He et al., 2016) and Transformer (Vaswani et al., 2017) proposed by the TAPE benchmark (Rao et al., 2019). ProtBert (Elnaggar et al., 2021) is a BERT-like model pretrained on UniRef100 (Suzek et al., 2007, 2015). ESM-2 (Rives et al., 2021; Verkuil et al., 2022;

	$6 \leq seq < 12$			$12 \leq seq < 24$			$24 \leq seq$		
	P@L	P@L/2	P@L/5	P@L	P@L/2	P@L/5	P@L	P@L/2	P@L/5
LSTM	$0.26_{(\pm 0.02)}$	$0.36_{(\pm 0.01)}$	$0.49_{(\pm 0.03)}$	$0.20_{(\pm 0.02)}$	$0.26_{(\pm 0.02)}$	$0.34_{(\pm 0.03)}$	$0.20_{(\pm 0.01)}$	$0.23_{(\pm 0.02)}$	$0.27_{(\pm 0.02)}$
ResNet	$0.25_{(\pm 0.02)}$	$0.34_{(\pm 0.02)}$	$0.46_{(\pm 0.02)}$	$0.28_{(\pm 0.01)}$	$0.25_{(\pm 0.01)}$	$0.35_{(\pm 0.03)}$	$0.10_{(\pm 0.03)}$	$0.13_{(\pm 0.02)}$	$0.17_{(\pm 0.03)}$
Transformer	$0.28_{(\pm 0.03)}$	$0.35_{(\pm 0.01)}$	$0.46_{(\pm 0.02)}$	$0.19_{(\pm 0.02)}$	$0.25_{(\pm 0.02)}$	$0.33_{(\pm 0.01)}$	$0.17_{(\pm 0.02)}$	$0.20_{(\pm 0.02)}$	$0.24_{(\pm 0.02)}$
ProtBert	$0.30_{(\pm 0.03)}$	$0.40_{(\pm 0.02)}$	$0.52_{(\pm 0.02)}$	$0.27_{(\pm 0.03)}$	$0.35_{(\pm 0.02)}$	$0.47_{(\pm 0.01)}$	$0.20_{(\pm 0.01)}$	$0.26_{(\pm 0.02)}$	$0.34_{(\pm 0.01)}$
OntoProtein	$0.37_{(\pm 0.02)}$	$0.46_{(\pm 0.01)}$	$0.57_{(\pm 0.03)}$	$0.32_{(\pm 0.01)}$	$0.40_{(\pm 0.02)}$	$0.50_{(\pm 0.02)}$	$0.24_{(\pm 0.03)}$	$0.31_{(\pm 0.01)}$	$0.39_{(\pm 0.03)}$
LM-GVP	$0.35_{(\pm 0.02)}$	$0.42_{(\pm 0.02)}$	$0.49_{(\pm 0.02)}$	$0.33_{(\pm 0.03)}$	$0.43_{(\pm 0.02)}$	$0.51_{(\pm 0.03)}$	$0.26_{(\pm 0.02)}$	$0.37_{(\pm 0.02)}$	$0.43_{(\pm 0.03)}$
GearNet	$0.39_{(\pm 0.02)}$	$0.46_{(\pm 0.02)}$	$0.57_{(\pm 0.02)}$	$0.36_{(\pm 0.03)}$	$0.44_{(\pm 0.02)}$	$0.55_{(\pm 0.03)}$	$0.29_{(\pm 0.02)}$	$0.37_{(\pm 0.01)}$	$0.45_{(\pm 0.02)}$
SaProt	$0.41_{(\pm 0.02)}$	$0.39_{(\pm 0.03)}$	$0.42_{(\pm 0.02)}$	$0.38_{(\pm 0.01)}$	$0.37_{(\pm 0.01)}$	$0.41_{(\pm 0.01)}$	$0.24_{(\pm 0.02)}$	$0.27_{(\pm 0.03)}$	$0.37_{(\pm 0.02)}$
KeAP	$0.41_{(\pm 0.04)}$	$0.52_{(\pm 0.02)}$	$0.62_{(\pm 0.03)}$	$0.36_{(\pm 0.01)}$	$0.45_{(\pm 0.01)}$	$0.57_{(\pm 0.01)}$	$0.29_{(\pm 0.02)}$	$0.37_{(\pm 0.03)}$	$0.46_{(\pm 0.02)}$
ESM-2	$0.42_{(\pm 0.02)}$	$0.49_{(\pm 0.03)}$	$0.63_{(\pm 0.01)}$	$0.37_{(\pm 0.01)}$	$0.43_{(\pm 0.01)}$	$0.57_{(\pm 0.02)}$	$0.30_{(\pm 0.02)}$	$0.38_{(\pm 0.03)}$	$0.46_{(\pm 0.02)}$
GLProtein	$\mathbf{0.45}_{(\pm 0.02}$	$\mathbf{0.55}_{(\pm 0.02}$	$\mathbf{0.66}_{(\pm 0.01}$	$\mathbf{0.39}_{(\pm 0.03}$	$\mathbf{0.48}_{(\pm 0.01}$	$\mathbf{0.58}_{(\pm 0.02}$	$\mathbf{0.31}_{(\pm 0.02}$	$\mathbf{0.40}_{(\pm 0.01}$	$\mathbf{0.47}_{(\pm 0.02)}$

Table 1: Comparisons on amino acid contact prediction. **seq** signifies the distance, measured in terms of amino acid units, between two selected amino acids. **P@L, P@L/2, P@L/5** denote the precision scores calculated upon top L (i.e., L most likely contacts), top L/2, and top L/5 predictions, respectively. The best results are bolded, and the second-best results are underlined.

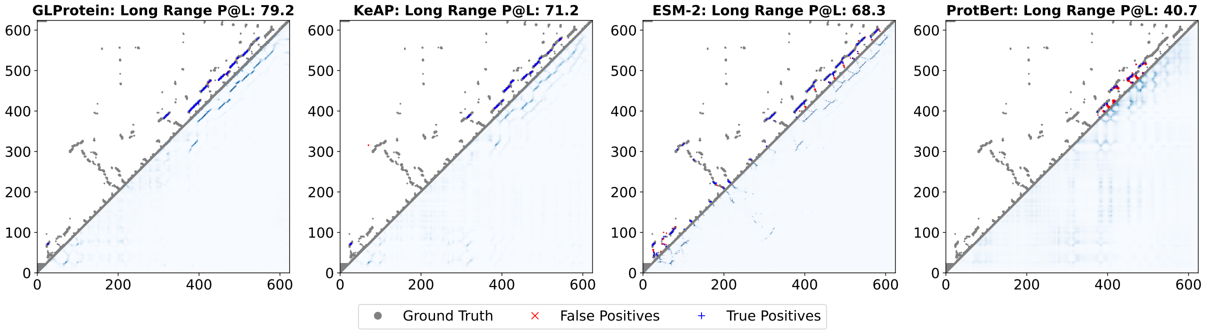


Figure 4: An example of amino acid contacts (top-L predictions for ProteinNet (AlQuraishi, 2019) test example TBM-hard#T0912). Raw contact probabilities are shown below the diagonal, top L contacts are shown above the diagonal (blue: true positives, red: false positives, grey: ground-truth contacts).

Hie et al., 2022) feature a transformer architecture pre-trained on the representative sequences from UniRef50 (Suzek et al., 2007, 2015). OntoProtein (Zhang et al.) and KeAP (Zhou et al., 2023a) are the most recent knowledge-based pre-training methodologies. SaProt (Su et al., 2023) is the most recent structure-based protein language model. LM-GVP (Wang et al., 2022) and GearNet (Zhang et al., 2022) are famous geometric methods for protein representation learning. We uniformly fine-tuned all downstream tasks without structural information to ensure fair and unbiased comparisons. Since all structural tokens are masked, residual information still exists; we substitute Foldseek structure tokens with "#" when fine-tuning SaProt.

Results. Table 1 shows the experimental results of amino acid contact prediction. Specifically, we notice that our model GLProtein consistently outperforms other models in short- ($6 \leq seq < 12$), medium- ($12 \leq seq < 24$) and long-range ($seq > 24$) contact predictions. Notably, our model demonstrates better performance compared to SaProt, which is also a structure-based language model. We also randomly selected a protein from the contact test dataset for visual analysis. As shown in Figure 4, the left is our GLProtein's result of amino acid contacts. The right three are the contact maps of three baseline models, including KeAP, ESM-2 and ProtBERT. Figure 4 shows more visually that GLProtein's prediction on the task of contact prediction is closer to labels, i.e., better performance on long-range contact prediction. We attribute the enhancements in performance achieved by GLProtein to its innovative integration of global and local structural information, which allows the pre-trained model to gain a deeper understanding of protein structure. More results can be found in the Appendix A.1.

4.1.2 Protein-Protein Interaction

Overview. Protein-protein interaction (PPI) is fundamental to virtually all biological processes and pathways in living organisms. It refers to the physical contact between two or more amino acid sequences. In this paper, we only focus on two-protein cases where a pair of protein sequences

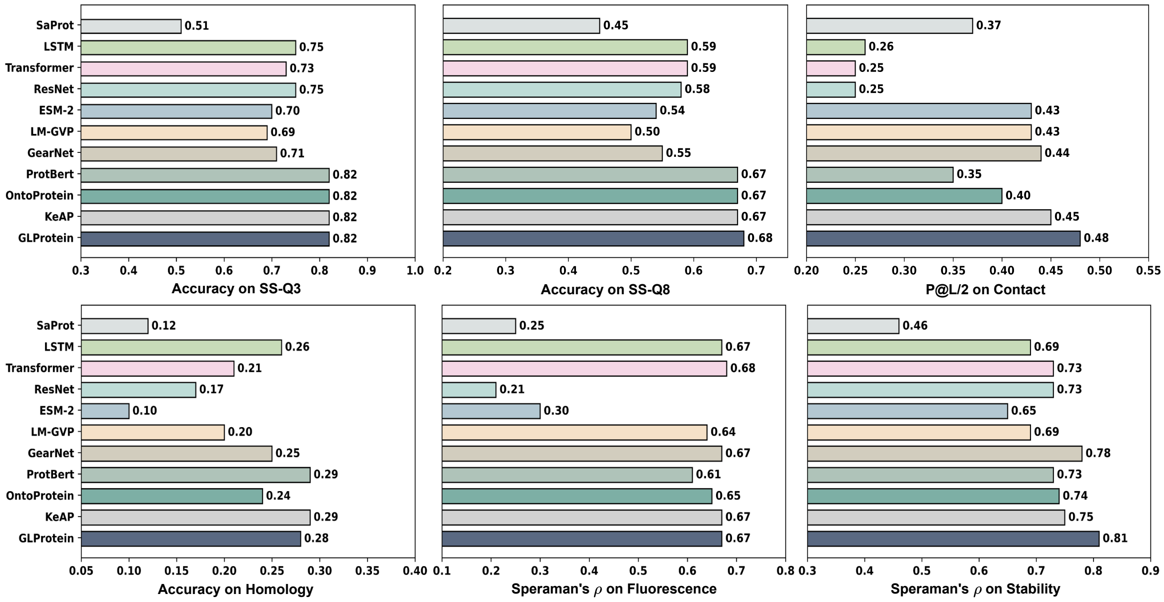


Figure 5: Results on TAPE Benchmark encompass various evaluations. SS is a secondary structure task that is evaluated in CB315. We report medium- and long-range results using P@L/2 metrics in contact prediction task. In fluorescence and stability prediction tasks, we use Spearman's ρ metric for evaluation. We also provide a related table in Appendix A.1.

serve as the inputs. The objective is to accurately predict the specific types of interactions that occur between each pair of proteins.

In our experiments, we focus on predicting 7 interaction types between protein pairs, namely reaction, binding, post-translational modifications, activation, inhibition, catalysis, and expression. The challenge of PPI prediction is approached as a multi-label classification problem. We conducted our experiments using three datasets: SHS27K (Chen et al., 2019), SHS148K (Chen et al., 2019) and STRING (Lv et al., 2021). Both SHS27K and SHS148K are considered subsets of STRING, with proteins excluded if they have fewer than 50 amino acids or exhibit 40% or higher sequence identity. We followed OntoProtein's setting to generate test sets and employed Breadth-First Search (BFS) and Depth-First Search (DFS) techniques across these datasets. The F1 score is utilized as the primary metric for evaluating performance.

Baselines. Following OntoProtein (Zhang et al.) and KeAP(Zhou et al., 2023a), we have expanded our baseline models to include four additional methods: DPPI (Hashemifar et al., 2018), DNN-PPI (Li et al., 2018), PIPR (Chen et al., 2019) and GNN-PPI (Lv et al., 2021). These are incorporated alongside existing baselines such as ProtBert (Elnaggar et al., 2021), ESM-2 (Beal, 2015), OntoProtein (Zhang et al.), KeAP (Zhou et al., 2023a), SaProt (Su et al., 2023), LM-GVP (Wang et al., 2022), GearNet (Zhang et al., 2022), DeepInter (Lin et al., 2023), MAPE-PPI (Wu et al., 2024), ProLLM (Jin et al., 2024) and ESM-C (Hayes et al., 2025), providing a comprehensive set of comparisons in our analysis.

	SHS27k		**SHS148k**		**STRING**	
Methods	BFS	DFS	BFS	DFS	BFS	DFS
DPPI	40.27$_{(\pm0.74)}$	44.86$_{(\pm0.87)}$	51.26$_{(\pm0.66)}$	51.43$_{(\pm0.94)}$	55.79$_{(\pm0.81)}$	64.72$_{(\pm0.94)}$
DNN-PPI	47.97$_{(\pm0.94)}$	52.85$_{(\pm0.91)}$	55.90$_{(\pm0.67)}$	57.82$_{(\pm0.78)}$	52.74$_{(\pm0.89)}$	62.99$_{(\pm0.93)}$
PIPR	43.67$_{(\pm0.99)}$	56.76$_{(\pm0.82)}$	60.10$_{(\pm0.85)}$	61.83$_{(\pm0.94)}$	53.65$_{(\pm0.88)}$	66.46$_{(\pm0.92)}$
GNN-PPI	62.47$_{(\pm0.65)}$	73.19$_{(\pm0.89)}$	71.01$_{(\pm0.92)}$	81.54$_{(\pm0.87)}$	75.34$_{(\pm0.82)}$	90.01$_{(\pm0.78)}$
ProtBert	68.44$_{(\pm0.78)}$	72.36$_{(\pm0.85)}$	70.06$_{(\pm0.88)}$	77.46$_{(\pm0.62)}$	66.08$_{(\pm0.91)}$	86.45$_{(\pm0.82)}$
DeepInter	77.31$_{(\pm1.14)}$	77.18$_{(\pm0.84)}$	74.52$_{(\pm0.99)}$	76.60$_{(\pm0.49)}$	77.82$_{(\pm0.98)}$	80.04$_{(\pm1.18)}$
OntoProtein	71.37$_{(\pm0.84)}$	76.28$_{(\pm0.77)}$	74.60$_{(\pm0.56)}$	76.33$_{(\pm0.69)}$	75.64$_{(\pm0.91)}$	90.23$_{(\pm0.79)}$
KeAP	78.51$_{(\pm0.95)}$	78.84$_{(\pm0.85)}$	74.26$_{(\pm0.89)}$	81.99$_{(\pm0.92)}$	80.08$_{(\pm0.79)}$	88.47$_{(\pm0.71)}$
LM-GVP	80.25$_{(\pm1.24)}$	79.42$_{(\pm0.83)}$	77.6$_{(\pm0.76)}$	80.36$_{(\pm0.97)}$	81.17$_{(\pm0.58)}$	85.67$_{(\pm0.74)}$
MAPE-PPI	83.63$_{(\pm0.76)}$	81.01$_{(\pm0.58)}$	84.57$_{(\pm0.91)}$	83.62$_{(\pm0.69)}$	87.18$_{(\pm0.82)}$	87.46$_{(\pm0.59)}$
GearNet	85.46$_{(\pm0.61)}$	82.73$_{(\pm0.69)}$	80.02$_{(\pm1.26)}$	82.28$_{(\pm0.93)}$	85.55$_{(\pm0.74)}$	88.03$_{(\pm0.51)}$
ESM-2	94.01$_{(\pm0.77)}$	87.32$_{(\pm0.97)}$	91.46$_{(\pm0.63)}$	85.24$_{(\pm0.46)}$	88.13$_{(\pm0.71)}$	85.53$_{(\pm0.55)}$
SaProt	91.18$_{(\pm0.73)}$	88.85$_{(\pm1.04)}$	90.75$_{(\pm0.91)}$	80.67$_{(\pm0.90)}$	88.23$_{(\pm0.81)}$	88.90$_{(\pm0.74)}$
ProLLM	91.49$_{(\pm0.91)}$	88.38$_{(\pm0.78)}$	90.90$_{(\pm1.03)}$	85.34$_{(\pm0.61)}$	87.38$_{(\pm0.77)}$	86.99$_{(\pm0.57)}$
ESM-C	92.46$_{(\pm1.02)}$	88.14$_{(\pm0.75)}$	91.86$_{(\pm0.51)}$	86.11$_{(\pm0.78)}$	87.41$_{(\pm0.95)}$	86.78$_{(\pm0.66)}$
GLProtein	**96.32**$_{(\pm0.86)}$	**91.23**$_{(\pm0.92)}$	**93.78**$_{(\pm0.77)}$	**86.14**$_{(\pm0.69)}$	**89.41**$_{(\pm0.66)}$	**91.35**$_{(\pm0.89)}$

Table 2: Protein-Protein Interaction Prediction Results. Breath-First Search (BFS) and Depth-First Search (DFS) are strategies that split the training and testing PPI datasets. The best results are bolded, and the second-best results are underlined.

Results. As shown in Table 2, the results clearly indicate that our method consistently outperforms all other methods, including the structure-based protein language model such as SaProt and multimodal protein language model such as ESM-C, across all datasets and both BFS and DFS evaluation metrics. The observed decline in performance can be linked to the growing amount of fine-tuning data, transitioning from SHS27k to STRING, which diminished the influence of pre-training. We believe that the structural similarities between proteins identified during the pre-training step enable GLProtein to excel in the PPI task, resulting in its outstanding performance.

5 Conclusion and Future Work

In this work, we propose GLProtein, a general protein language model with global and local protein structure information. GLProtien outperforms the previous protein representation learning model on most downstream applications, demonstrating the performance superiority of GLProtein. In the future, we aim to further enhance GLProtein's capabilities by exploring novel avenues for incorporating multi-modal data sources, refining the model's interpretability, and extending its applicability to a wider array of biological contexts.

6 Limitations

We have observed that GLProtein underperforms on certain individual tasks. For instance, in the protein-protein binding affinity prediction task, ESM-2 surpasses GLProtein. This task focuses on predicting changes in binding affinity resulting from protein mutations. We believe that GLProtein's limited performance is attributed to its lack of mutation information, whereas ESM-2 incorporates multiple sequence alignment (MSA) data during training, which includes mutation insights. Similarly, in the Fluorescence task, GLProtein does not demonstrate significant improvement when tasked with distinguishing closely related proteins. We hypothesize that while GLProtein effectively learns structural similarities among different proteins during pre-training, it excels at identifying differences between dissimilar structures but struggles to differentiate between similar ones. We plan to further investigate these issues in our future research.

7 Acknowledgements

The research described in this paper has been partially supported by the National Natural Science Foundation of China (project no. 62102335 and 62372314), General Research Funds from the Hong Kong Research Grants Council (project no. PolyU 15207322, 15200023, 15206024, and 15224524), internal research funds from Hong Kong Polytechnic University (project no. P0042693, P0048625, and P0051361). This work was supported by computational resources provided by The Centre for Large AI Models (CLAIM) of The Hong Kong Polytechnic University.

References

Ali Al-Fatlawi, Martin Menzel, and Michael Schroeder. 2023. Is protein blast a thing of the past? *nature communications*, 14(1):8195.

Mohammed AlQuraishi. 2019. Proteinnet: a standardized data set for machine learning of protein structure. *BMC bioinformatics*, 20:1–10.

Mohammed AlQuraishi. 2021. Machine learning in protein structure prediction. *Current opinion in chemical biology*, 65:1–8.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. 2016. Layer normalization. *arXiv preprint arXiv:1607.06450*.

Daniel J Beal. 2015. Esm 2.0: State of the art and future potential of experience sampling methods in organizational research. *Annu. Rev. Organ. Psychol. Organ. Behav.*, 2(1):383–407.

Brigitte Boeckmann, Amos Bairoch, Rolf Apweiler, Marie-Claude Blatter, Anne Estreicher, Elisabeth Gasteiger, Maria J Martin, Karine Michoud, Claire O'Donovan, Isabelle Phan, and 1 others. 2003. The swiss-prot protein knowledgebase and its supplement trembl in 2003. *Nucleic acids research*, 31(1):365–370.

John-Marc Chandonia, Naomi K Fox, and Steven E Brenner. 2019. Scope: classification of large macromolecular structures in the structural classification of proteins—extended database. *Nucleic acids research*, 47(D1):D475–D481.

Dexiong Chen, Philip Hartout, Paolo Pellizzoni, Carlos Oliver, and Karsten Borgwardt. 2024. Endowing protein language models with structural knowledge. *arXiv preprint arXiv:2401.14819*.

Muhao Chen, Chelsea J-T Ju, Guangyu Zhou, Xuelu Chen, Tianran Zhang, Kai-Wei Chang, Carlo Zaniolo, and Wei Wang. 2019. Multifaceted protein–protein interaction prediction based on siamese residual rcnn. *Bioinformatics*, 35(14):i305–i314.

Jianlin Cheng, Allison N Tegge, and Pierre Baldi. 2008. Machine learning methods for protein structure prediction. *IEEE reviews in biomedical engineering*, 1:41–49.

Ratul Chowdhury, Nazim Bouatta, Surojit Biswas, Christina Floristean, Anant Kharkar, Koushik Roy, Charlotte Rochereau, Gustaf Ahdritz, Joanna Zhang, George M Church, and 1 others. 2022. Single-sequence protein structure prediction using a language model and deep learning. *Nature Biotechnology*, 40(11):1617–1623.

Simon KS Chu, Kush Narang, and Justin B Siegel. 2024. Protein stability prediction by fine-tuning a protein language model on a mega-scale dataset. *PLOS Computational Biology*, 20(7):e1012248.

James A Cuff and Geoffrey J Barton. 1999. Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins: Structure, Function, and Bioinformatics*, 34(4):508–519.

Steven J Davis, Kathleen Alexander, Juan Moreno-Cruz, Chaopeng Hong, Matthew Shaner, Ken Caldeira, and Ian McKay. 2024. Food without agriculture. *Nature Sustainability*, 7(1):90–95.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Bo Ding, Yue Weng, Yunqing Liu, Chunlan Song, Le Yin, Jiafan Yuan, Yanrui Ren, Aiwen Lei, and Chien-Wei Chiang. 2019. Selective photoredox trifluoromethylation of tryptophan-containing peptides. *European Journal of Organic Chemistry*, 2019(46):7596–7605.

Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rehawi, Yu Wang, Llion Jones, Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, and 1 others. 2021. Prottrans: Toward understanding the language of life through self-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(10):7112–7127.

Wenqi Fan, Yi Zhou, Shijie Wang, Yuyao Yan, Hui Liu, Qian Zhao, Le Song, and Qing Li. 2025. Computational protein science in the era of large language models (llms). *arXiv preprint arXiv:2501.10282*.

Noelia Ferruz, Steffen Schmidt, and Birte Höcker. 2022. Protgpt2 is a deep unsupervised language model for protein design. *Nature communications*, 13(1):4348.

Bowen Gao, Bo Qiang, Haichuan Tan, Yinjun Jia, Minsi Ren, Minsi Lu, Jingjing Liu, Wei-Ying Ma, and Yanyan Lan. 2024. Drugclip: Contrasive protein-molecule representation learning for virtual screening. *Advances in Neural Information Processing Systems*, 36.

Gregor P Greslehner. 2018. What do molecular biologists mean when they say'structure determines function'?

Tymor Hamamsy, James T Morton, Daniel Berenberg, Nicholas Carriero, Vladimir Gligorijevic, Robert Blackwell, Charlie EM Strauss, Julia Koehler Leman, Kyunghyun Cho, and Richard Bonneau. 2022. Tm-vec: template modeling vectors for fast homology detection and alignment. *bioRxiv*, pages 2022–07.

Tymor Hamamsy, James T Morton, Robert Blackwell, Daniel Berenberg, Nicholas Carriero, Vladimir Gligorijevic, Charlie EM Strauss, Julia Koehler Leman, Kyunghyun Cho, and Richard Bonneau. 2023. Protein remote homology detection and structural alignment using deep learning. *Nature biotechnology*, pages 1–11.

Somaye Hashemifar, Behnam Neyshabur, Aly A Khan, and Jinbo Xu. 2018. Predicting protein–protein interactions through sequence-based deep learning. *Bioinformatics*, 34(17):i802–i810.

Thomas Hayes, Roshan Rao, Halil Akin, Nicholas J Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, Vincent Q Tran, Jonathan Deaton, Marius Wiggert, and 1 others. 2025. Simulating 500 million years of evolution with a language model. *Science*, 387(6736):850–858.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Pedro Hermosilla and Timo Ropinski. 2022. Contrastive representation learning for 3d protein structures. *arXiv preprint arXiv:2205.15675*.

Brian Hie, Salvatore Candido, Zeming Lin, Ori Kabeli, Roshan Rao, Nikita Smetanin, Tom Sercu, and Alexander Rives. 2022. A high-level programming language for generative protein design. *bioRxiv*, pages 2022–12.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Jie Hou, Badri Adhikari, and Jianlin Cheng. 2018. Deepsf: deep convolutional neural network for mapping protein sequences to folds. *Bioinformatics*, 34(8):1295–1303.

Hanxu Hu, Yunqing Liu, Zhongyi Yu, and Laura Perez-Beltrachini. 2023. Improving user controlled table-to-text generation robustness. *arXiv preprint arXiv:2302.09820*.

Yufei Huang, Siyuan Li, Lirong Wu, Jin Su, Haitao Lin, Odin Zhang, Zihan Liu, Zhangyang Gao, Jiangbin Zheng, and Stan Z Li. 2024. Protein 3d graph structure learning for robust structure-based protein property prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 12662–12670.

Sabrina Jaeger, Simone Fulle, and Samo Turk. 2018. Mol2vec: unsupervised machine learning approach with chemical intuition. *Journal of chemical information and modeling*, 58(1):27–35.

Kanchan Jha, Sriparna Saha, and Hiteshi Singh. 2022. Prediction of protein–protein interaction using graph neural networks. *Scientific Reports*, 12(1):8360.

Mingyu Jin, Haochen Xue, Zhenting Wang, Boming Kang, Ruosong Ye, Kaixiong Zhou, Mengnan Du, and Yongfeng Zhang. 2024. Prollm: protein chain-of-thoughts enhanced llm for protein-protein interaction prediction. *arXiv preprint arXiv:2405.06649*.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, and 1 others. 2021. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589.

Kamil Kaminski, Jan Ludwiczak, Kamil Pawlicki, Vikram Alva, and Stanislaw Dunin-Horkawicz. 2023. plm-blast: distant homology detection based on direct comparison of sequence representations from protein language models. *Bioinformatics*, 39(10):btad579.

Hang Li, Xiu-Jun Gong, Hua Yu, and Chang Zhou. 2018. Deep neural network based predictions of protein interactions using primary sequences. *Molecules*, 23(8):1923.

Jiatong Li, Junxian Li, Yunqing Liu, Dongzhan Zhou, and Qing Li. 2024a. Tomg-bench: Evaluating llms on text-based open molecule generation. *arXiv preprint arXiv:2412.14642*.

Jiatong Li, Yunqing Liu, Wenqi Fan, Xiao-Yong Wei, Hui Liu, Jiliang Tang, and Qing Li. 2024b. Empowering molecule discovery for molecule-caption translation with large language models: A chatgpt perspective. *IEEE transactions on knowledge and data engineering*, 36(11):6071–6083.

Jiatong Li, Yunqing Liu, Wei Liu, Jingdi Le, Di Zhang, Wenqi Fan, Dongzhan Zhou, Yuqiang Li, and Qing Li. 2024c. Molreflect: Towards in-context fine-grained alignments between molecules and texts. *arXiv preprint arXiv:2411.14721*.

Yiran Li, Gongyao Guo, Jieming Shi, Sibo Wang, and Qing Li. 2025. Efficient integration of multi-view attributed graphs for clustering and embedding. In *2025 IEEE 41st International Conference on Data Engineering (ICDE)*, pages 3863–3875. IEEE.

Ningyi Liao, Siqiang Luo, Xiang Li, and Jieming Shi. 2023. Ld2: Scalable heterophilous graph neural network with decoupled embeddings. *Advances in neural information processing systems*, 36:10197–10209.

Elizabeth L Lieu, Tu Nguyen, Shawn Rhyne, and Jiyeon Kim. 2020. Amino acids in cancer. *Experimental & molecular medicine*, 52(1):15–30.

Peicong Lin, Huanyu Tao, Hao Li, and Sheng-You Huang. 2023. Protein–protein contact prediction by geometric triangle-aware protein language models. *Nature Machine Intelligence*, 5(11):1275–1284.

David J Lipman and William R Pearson. 1985. Rapid and sensitive protein similarity searches. *Science*, 227(4693):1435–1441.

Chengyi Liu, Wenqi Fan, Yunqing Liu, Jiatong Li, Hang Li, Hui Liu, Jiliang Tang, and Qing Li. 2023. Generative diffusion models on graphs: methods and applications. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 6702–6711.

Meng Liu, Youzhi Luo, Kanji Uchino, Koji Maruhashi, and Shuiwang Ji. 2022. Generating 3d molecules for target protein binding. *arXiv preprint arXiv:2204.09410*.

Michael J Lopez and Shamim S Mohiuddin. 2024. Biochemistry, essential amino acids. In *StatPearls [Internet]*. StatPearls Publishing.

Guofeng Lv, Zhiqiang Hu, Yanguang Bi, and Shaoting Zhang. 2021. Learning unknown from correlations: Graph neural network for inter-novel-protein interaction prediction. *arXiv preprint arXiv:2105.06709*.

Ali Madani, Bryan McCann, Nikhil Naik, Nitish Shirish Keskar, Namrata Anand, Raphael R Eguchi, Po-Ssu Huang, and Richard Socher. 2020. Progen: Language modeling for protein generation. *arXiv preprint arXiv:2004.03497*.

Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. 2022. Colabfold: making protein folding accessible to all. *Nature methods*, 19(6):679–682.

Iain H Moal and Juan Fernández-Recio. 2012. Skempi: a structural kinetic and energetic database of mutant protein interactions and its use in empirical models. *Bioinformatics*, 28(20):2600–2607.

Dan Ofer, Nadav Brandes, and Michal Linial. 2021. The language of proteins: Nlp, machine learning & protein sequences. *Computational and Structural Biotechnology Journal*, 19:1750–1758.

Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. 2022. Pocket2mol: Efficient molecular sampling based on 3d protein pockets. In *International Conference on Machine Learning*, pages 17644–17655. PMLR.

Roshan Rao, Nicholas Bhattacharya, Neil Thomas, Yan Duan, Peter Chen, John Canny, Pieter Abbeel, and Yun Song. 2019. Evaluating protein transfer learning with tape. *Advances in neural information processing systems*, 32.

Manon Réau, Nicolas Renaud, Li C Xue, and Alexandre MJJ Bonvin. 2023. Deeprank-gnn: a graph neural network framework to learn patterns in protein–protein interfaces. *Bioinformatics*, 39(1):btac759.

Nicolas Renaud, Cunliang Geng, Sonja Georgievska, Francesco Ambrosetti, Lars Ridder, Dario F Marzella, Manon F Réau, Alexandre MJJ Bonvin, and Li C Xue. 2021. Deeprank: a deep learning framework for data mining 3d protein-protein interfaces. *Nature communications*, 12(1):7068.

Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C Lawrence Zitnick, Jerry Ma, and 1 others. 2021. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15):e2016239118.

Gabriel J Rocklin, Tamuka M Chidyausiku, Inna Goreshnik, Alex Ford, Scott Houliston, Alexander Lemak, Lauren Carter, Rashmi Ravichandran, Vikram K Mulligan, Aaron Chevalier, and 1 others. 2017. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science*, 357(6347):168–175.

Bernhard Scholkopf, Kah-Kay Sung, Christopher JC Burges, Federico Girosi, Partha Niyogi, Tomaso Poggio, and Vladimir Vapnik. 1997. Comparing support vector machines with gaussian kernels to radial basis function classifiers. *IEEE transactions on Signal Processing*, 45(11):2758–2765.

Jaspreet Singh, Thomas Litfin, Jaswinder Singh, Kuldip Paliwal, and Yaoqi Zhou. 2022. Spot-contact-lm: improving single-sequence-based prediction of protein contact map using a transformer language model. *Bioinformatics*, 38(7):1888–1894.

Vignesh Ram Somnath, Charlotte Bunne, and Andreas Krause. 2021. Multi-scale representation learning on proteins. *Advances in Neural Information Processing Systems*, 34:25244–25255.

Jin Su, Chenchen Han, Yuyang Zhou, Junjie Shan, Xibin Zhou, and Fajie Yuan. 2023. Saprot: Protein language modeling with structure-aware vocabulary. biorxiv.

Baris E Suzek, Hongzhan Huang, Peter McGarvey, Raja Mazumder, and Cathy H Wu. 2007. Uniref: comprehensive and non-redundant uniprot reference clusters. *Bioinformatics*, 23(10):1282–1288.

Baris E Suzek, Yuqi Wang, Hongzhan Huang, Peter B McGarvey, Cathy H Wu, and UniProt Consortium. 2015. Uniref clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, 31(6):926–932.

Mirko Torrisi, Gianluca Pollastri, and Quan Le. 2020. Deep learning methods in protein structure prediction. *Computational and Structural Biotechnology Journal*, 18:1301–1310.

Tomer Tsaban, Julia K Varga, Orly Avraham, Ziv Ben-Aharon, Alisa Khramushin, and Ora Schueler-Furman. 2022. Harnessing protein folding neural networks for peptide–protein docking. *Nature communications*, 13(1):176.

Serbulent Unsal, Heval Atas, Muammer Albayrak, Kemal Turhan, Aybar C Acar, and Tunca Doğan. 2022. Learning functional properties of proteins with language models. *Nature Machine Intelligence*, 4(3):227–245.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Robert Verkuil, Ori Kabeli, Yilun Du, Basile IM Wicky, Lukas F Milles, Justas Dauparas, David Baker, Sergey Ovchinnikov, Tom Sercu, and Alexander Rives. 2022. Language models generalize beyond natural proteins. *BioRxiv*, pages 2022–12.

Yanbin Wang, Zhu-Hong You, Shan Yang, Xiao Li, Tong-Hai Jiang, and Xi Zhou. 2019. A high efficient biological language model for predicting protein–protein interactions. *Cells*, 8(2):122.

Zichen Wang, Steven A Combs, Ryan Brand, Miguel Romero Calvo, Panpan Xu, George Price, Nataliya Golovach, Emmanuel O Salawu, Colby J Wise, Sri Priya Ponnapalli, and 1 others. 2022. Lm-gvp: an extensible sequence and structure informed deep learning framework for protein property prediction. *Scientific reports*, 12(1):6832.

Yue Weng, Bo Ding, Yunqing Liu, Chunlan Song, Lo-Ying Chan, and Chien-Wei Chiang. 2021. Late-stage photoredox c–h amidation of n-unprotected indole derivatives: Access to n-(indol-2-yl) amides. *Organic Letters*, 23(7):2710–2714.

Lirong Wu, Yijun Tian, Yufei Huang, Siyuan Li, Haitao Lin, Nitesh V Chawla, and Stan Z Li. 2024. Mape-ppi: Towards effective and efficient protein-protein interaction prediction via microenvironment-aware protein embedding. *arXiv preprint arXiv:2402.14391*.

Yijia Xiao, Jiezhong Qiu, Ziang Li, Chang-Yu Hsieh, and Jie Tang. 2021. Modeling protein using large-scale pretrain language model. *arXiv preprint arXiv:2108.07435*.

Jinrui Xu and Yang Zhang. 2010. How significant is a protein structure similarity with tm-score= 0.5? *Bioinformatics*, 26(7):889–895.

Xiaotong Xu and Alexandre MJJ Bonvin. 2024. Deeprank-gnn-esm: a graph neural network for scoring protein–protein models using protein language model. *Bioinformatics advances*, 4(1):vbad191.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. 2021. Do transformers really perform badly for graph representation? *Advances in neural information processing systems*, 34:28877–28888.

Jirina Zackova Suchanova, Gust Bilcke, Beata Romanowska, Ali Fatlawi, Martin Pippel, Alastair Skeffington, Michael Schroeder, Wim Vyverman, Klaas Vandepoele, Nils Kröger, and 1 others. 2023. Diatom adhesive trail proteins acquired by horizontal gene transfer from bacteria serve as primers for marine biofilm formation. *New Phytologist*, 240(2):770–783.

Ningyu Zhang, Zhen Bi, Xiaozhuan Liang, Siyuan Cheng, Haosen Hong, Shumin Deng, Qiang Zhang, Jiazhang Lian, and Huajun Chen. Ontoprotein: Protein pretraining with gene ontology embedding. In *International Conference on Learning Representations*.

Yang Zhang and Jeffrey Skolnick. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4):702–710.

Zuobai Zhang, Minghao Xu, Arian Jamasb, Vijil Chenthamarakshan, Aurelie Lozano, Payel Das, and Jian Tang. 2022. Protein representation learning by geometric structure pretraining. *arXiv preprint arXiv:2203.06125*.

Hao Zhao, Xiangwen Fan, Zhaohai Bai, Lin Ma, Chao Wang, Petr Havlík, Zhenling Cui, Juraj Balkovic, Mario Herrero, Zhou Shi, and 1 others. 2024. Holistic food system innovation strategies can close up to 80% of china's domestic protein gaps while reducing global environmental impacts. *Nature Food*, pages 1–11.

Jingtian Zhao, Yang Cao, and Le Zhang. 2020. Exploring the computational methods for protein-ligand binding site prediction. *Computational and structural biotechnology journal*, 18:417–426.

Hong-Yu Zhou, Yunxiang Fu, Zhicheng Zhang, Bian Cheng, and Yizhou Yu. 2023a. Protein representation learning via knowledge enhanced primary structure reasoning. In *The Eleventh International Conference on Learning Representations*.

Ziang Zhou, Zhihao Ding, Jieming Shi, Qing Li, and Shiqi Shen. 2024. Tined: Gnns-to-mlps by teacher injection and dirichlet energy distillation. *arXiv preprint arXiv:2412.11180*.

Ziang Zhou, Jieming Shi, Renchi Yang, Yuanhang Zou, and Qing Li. 2023b. Slotgat: slot-based message passing for heterogeneous graphs. In *International Conference on Machine Learning*, pages 42644–42657. PMLR.

A Appendix

Table of Content:

A.1 Results on TAPE Benchmark

In addition to the Figure version, we also provide results on TAPE benchmark in a tabular version. As shown in Table 6, our model GLProtein performs competitive performance on many tasks, especially on the contact prediction and stability prediction tasks.

A.2 Secondary structure prediction

Overview. Secondary structure is a fundamental aspect of computational biology, aiming to determine the local structures of protein segments. this task is a sequence-to-sequence task where each input protein is mapped to a type of local structure. We report accuracy on a per-amino acid basis on the CB513 dataset (Cuff and Barton, 1999). **Baselines.** We evaluate our model compared with ten baselines. Specifically, we employed variations of LSTM (Hochreiter and Schmidhuber, 1997), ResNet (He et al., 2016) and Transformer (Vaswani et al., 2017) proposed by the TAPE benchmark (Rao et al., 2019). ProtBert (Elnaggar et al., 2021) is a BERT-like model pre-trained on UniRef100 (Suzek et al., 2007, 2015). ESM-2 (Rives et al., 2021; Verkuil et al., 2022; Hie et al., 2022) feature a transformer architecture pre-trained on the representative sequences from UniRef50 (Suzek et al., 2007, 2015). OntoProtein (Zhang et al.) and KeAP (Zhou et al., 2023a) are the most recent knowledge-based pre-training methodologies. SaProt (Su et al., 2023) is the most recent structure-based protein language model. LM-GVP (Wang et al., 2022) and GearNet (Zhang et al., 2022) are famous geometric methods for protein representation learning.

Results. For the secondary structure (SS-Q3 and SS-Q8), as shown in Figure 5, GLProtein outperforms other baselines in SS-Q8 task and shows competitive performance with ProtBERT, OntoProtein and KeAP in SS-Q3 task. Considering the approaches taken by Saprot, LM-GVP, and GearNet, which also incorporate protein structural information, the evident performance superiority of GLProtein over these methods indicates that it offers a more effective option for structure-based protein representation learning. We attribute the enhancements in performance achieved by GLProtein to its innovative integration of global and local structural information, which allows the pre-trained model to gain a deeper understanding of protein structure.

A.3 Homology Detection, Fluorescence and Stability Prediction

Overview of homology detection. The task of predicting remote homology in proteins can be viewed as a classification problem at the molecular level. The objective is to input a protein sequence into the homology detection model, which then identifies the correct types of protein fold. In our paper, this presents a significant challenge with 1,195 distinct protein folds to classify. We utilize data sources from (Hou et al., 2018) and present the average accuracy achieved on the fold-level heldout set.

Overview of fluorescence prediction. In the realm of protein science, fluorescence prediction is a vital task that involves estimating the fluorescence properties of proteins. This is a regression task where each input protein is mapped to a label measuring the most extreme circumstances in which the protein maintains its fold above a concentration threshold. We use the data from (Rocklin et al., 2017) and use Spearman's rank correlation coefficient as the metric.

Overview of stability prediction. Stability prediction involves estimating the resilience of a protein's structure under various environmental conditions, a critical factor in understanding its functional efficacy and therapeutic potential. This regression task focuses on predicting the intrinsic stability of proteins, which is essential for assessing their capacity to preserve their structural integrity under severe conditions. To assess the effectiveness of our model, we measure its performance using Spearman's rank correlation coefficient across the entire test set (Rocklin et al., 2017).

Baselines. As shown in Figure 5, we included ten protein model as baselines.

Results. As for fluorescence prediction, Figure 5 shows our model has the most competitive performance compared to Transformer and KeAP. In the domain of stability prediction, our model again shows the highest performance with a score of 0.81. This is significantly higher compared to other models, indicating its potential utility in applications like drug design and protein engineering, where stability is paramount.

A.4 Protein Function Prediction

Overview. Protein function prediction aims to assign biological or biochemical roles to proteins, and we also regard this task as a sequence classification task. Following KeAP (Zhou et al., 2023a), we divide protein attributes into three groups: molecular function (MF), biological process (BP) and cellular component (CC), and report the Spearman's rank correlation scores for each group.

Baselines. We evaluate our model compared with five baselines, including ESM-2, ProtBERT, OntoProtein, SaProt and KeAP.

Methods	MF	BP	CC	Avg
ESM-2	0.31	**0.42**	0.28	0.34
ProtBert	**0.41**	0.35	0.36	0.37
OntoProtien	**0.41**	0.36	0.36	0.38
SaProt	0.40	0.40	0.39	**0.40**
KeAP	0.40	0.40	**0.40**	**0.40**
GLProtein	**0.41**	0.40	0.39	**0.40**

Table 3: Comparisons on semantic similarity inference. The best results are bolded, and the second-best results are underlined.

Results. Table 3 assesses the performance of various computational models in prediction protein functions in three categories: MF, BP and CC. Additionally, an average score (Avg) is calculated for each method to provide a holistic view of performance across all categories. These models all show a balanced performance in three groups. It is worth noting that our model does not use any protein attribute-related knowledge and is comparable to OntoProtein and KeAP, which do. It also demonstrates the superiority of our approach.

A.5 Protein-Protein Binding Affinity Estimation

Overview. In this task, we focus on assessing how well protein representations can predict changes in binding affinity caused by protein mutations. This regression task involves associating each protein pair with a numerical value. Following the methodology described in (Unsal et al., 2022), we employ Bayesian ridge regression on the outcomes of element-wise multiplication of representations derived from pre-trained protein models. This approach is designed to enhance the accuracy of binding affinity predictions. We used the SKEMPI dataset from (Moal and Fernández-Recio, 2012) and reported the mean square error of 10-fold cross-validation.

Baselines. We evaluate our model compared with six baselines. Specifically, we employed PIPR (Chen et al., 2019), ProtBert (Elnaggar et al., 2021), ESM-2 (Rives et al., 2021; Verkuil et al., 2022; Hie et al., 2022), SaProt (Su et al., 2023), OntoProtein (Zhang et al.) and KeAP (Zhou et al., 2023a). PIPR is a siamese-residual-RCNN-based model for multifaceted protein–protein interaction prediction. ProtBert is a BERT-like model pre-trained on UniRef100 (Suzek et al., 2007, 2015). ESM-2 feature a transformer architecture pre-trained on the representative sequences from UniRef50 (Suzek et al., 2007, 2015). SaProt is the most recent structure-based protein language model. OntoProtein and KeAP are the most recent knowledge-based pre-training methodologies.

Methods	Affinity($\downarrow$)
PIPR	0.63
ProtBert	0.58
ESM-2	**0.48**
SaProt	0.58
OntoProtien	0.59
KeAP	0.52
GLProtein	0.52

Table 4: Comparisons on protein-protein binding affinity prediction, with the best result bolded and the second best underlined. The notion ↓ signifies a preference for lower values, reflecting a superior predictive performance in this context.

Results. Table 4 compares several methods of predicting the binding affinity of protein interactions, where a lower score indicates superior performance. GLProtein outperforms PIPR, ProtBert, SaProt and OntoProtein. It also shows the competitive perfor-

	Parameters	Resouces	Pre-training	Inference (40 examples)
ProtBert	400M	A single TPU Pod V3-512	400k steps	2.02s
OntoProtein	400M	4 NVIDIA 48G A6000 GPUs	3 Days (continue pertaining on ProtBert)	1.91s
KeAP	400M	4 NVIDIA 48G A6000 GPUs	3 Days (continue pertaining on ProtBert)	1.94s
SaProt	650M	64 NVIDIA 80G A100 GPUs	3 Months	3.02s
ESM-2	650M	-	-	2.45s
GLProtein	400M	4 NVIDIA 48G A6000 GPUs	3 Days (continue pertaining on ProtBert)	1.93s

Table 5: Comparison of the number of parameters, resources, pre-training time, and inference time for GLProtein and baselines.

mance of KeAP and ESM-2.

A.6 Ablation Study

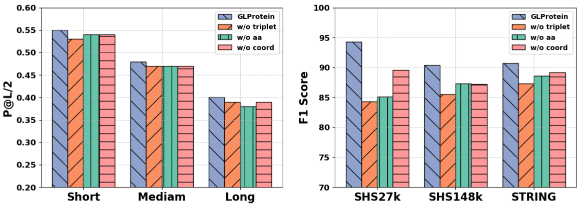


Figure 6: **Left**: Ablations of three proposed approaches. Long-range P@L/2 results are reported for contact prediction. Right: Ablations of three proposed approaches. F1 scores are reported for protein-protein interaction tasks.

We investigated the effects of employing diverse protein structure information fusion strategies. First of all, the exclusion of the global structure information modelling component (representation as "*w/o triplet*" in Figure 6) resulted in varying degrees of performance deterioration across contact prediction and protein-protein interaction prediction tasks. This observation suggests that our global structure similarities through protein triplet contrastive learning stand out as a more efficacious choice. Subsequently, upon removing the proposed substructure-based molecular encoding from the local protein structure information component (denoted as "*w/o aa*" in Figure 6), we noted a decline in performance by approximately 2.5% and 8% for contact prediction and protein-protein interaction tasks, respectively. This underscores the essential role of substructure-based molecular encoding within our proposed methodologies. Finally when the protein 3D distance encoding was omitted from the local structure information modelling component (indicated as "*w/o coord*" in Figure 6), a similar trend of performance degradation was observed, further emphasizing the indispensability of this strategy within our architectural framework.

A.7 Parameter Sensitivity Study

In this section, we explore the impact of the parameters in the model on the final performance of our protein model. We experimented with the number of protein samples in the protein local structure information modelling component and the coefficient of contrastive learning loss for the protein triplet, respectively.

As shown in Figure 7, we test the number of protein samples from 1 to 4 on the contact prediction task. We observe that as the number of protein samples increases, the performance of our model improves to varying degrees in short-, medium- and long-range contact prediction. This also shows that our proposed protein triplet approach indeed enables the protein language model to capture the structural similarity features among proteins. Due to computational and memory cost considerations, we ended up constructing 4 protein positive samples and 4 protein negative samples for each protein.

As shown in Figure 8, we test the value of the coefficient α of contrastive learning loss for the protein triplet. We divided the experiment into 6 groups and set the values of α to 0.1, 0.3, 0.5, 1, 3, and 5. Then, we evaluated them using the protein-protein interaction prediction task on SHS27k, SHS148k, and STRING datasets, respectively. We observe that the model achieves the best performance when the value of α is set to 1. Thus, we choose $\alpha = 1$ in this paper as our model's setting.

A.8 Visualization of Protein Representations

To facilitate a more intuitive comparison, we utilize t-SNE to visualize the protein representations produced by GLProtein, ESM2, KeAP, and ProtBert. The visualization results, based on the non-redundant subset ($PIDE \leq 40\%$) of the SCOPe database (Chandonia et al., 2019), are illustrated in Figure 9. As depicted in this figure, the representations for alpha and beta proteins generated by GLProtein are distinctly separated, whereas those produced by ESM-2, KeAP, and ProtBert are more closely intertwined.

	Structure			Evolutionary	Engineering	
	SS-Q3	SS-Q8	Contact	Homology	Fluorescence	Stability
SaProt	0.51	0.45	0.37	0.12	0.25	0.46
LSTM	0.75	0.59	0.26	0.26	0.67	0.69
Transformer	0.73	0.59	0.25	0.21	**0.68**	0.73
ResNet	0.75	0.58	0.25	0.17	0.21	0.73
ESM-2	0.70	0.54	0.43	0.10	0.30	0.65
LM-GVP	0.69	0.50	0.43	0.20	0.64	0.69
GearNet	0.71	0.55	0.44	0.25	0.67	0.78
ProtBert	**0.82**	0.67	0.35	**0.29**	0.61	0.73
OntoProtein	**0.82**	0.67	0.40	0.24	0.65	0.74
KeAP	**0.82**	0.67	0.45	**0.29**	0.67	0.75
GLProtein	**0.82**	**0.68**	**0.48**	0.28	0.67	**0.81**

Table 6: Results on TAPE Benchmark. SS is a secondary structure task that is evaluated in CB315. In contact prediction, we test medium- and long-range using P@L/2 metrics. In protein engineering tasks, we test fluorescence and stability prediction using Spearman's ρ metric.

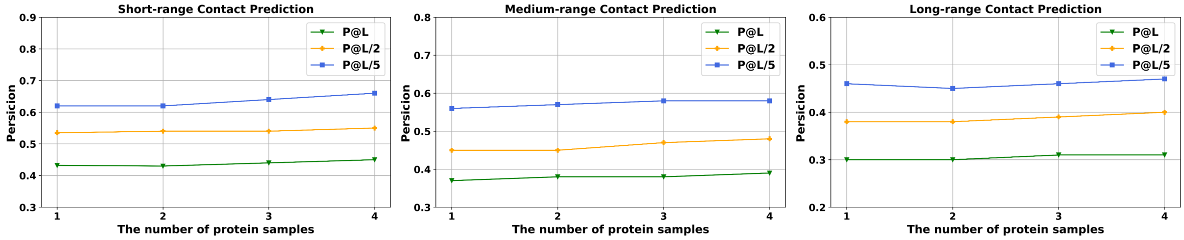


Figure 7: Parameter sensitivity study on the number of protein samples in the local structure information component.

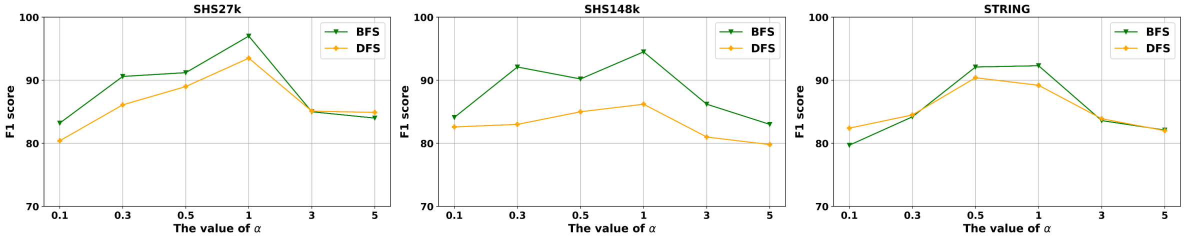
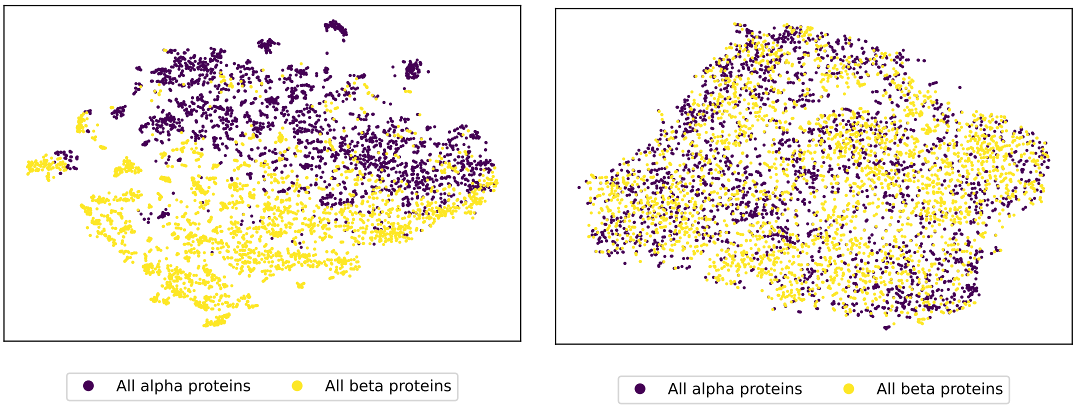


Figure 8: Parameter sensitivity study on the value of the coefficient α of contrastive learning loss for the protein triplet in the local structure information component.

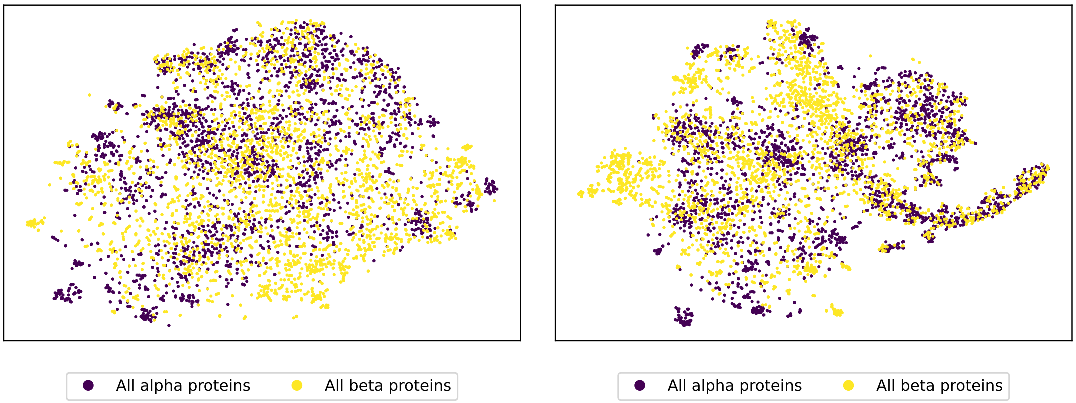
Task	Epoch	Batch Size	Warmup Ratio	Learning Rate	Freeze Bert	Optimizer
Contact	5	8	0.08	3e-5	False	AdamW
Homology	10	32	0.08	4e-5	False	AdamW
Stability	5	64	0.08	1e-5	False	AdamW
SS-Q3	5	32	0.08	3e-5	False	AdamW
SS-Q8	5	32	0.08	3e-5	False	AdamW
Fluorescence	15	64	0.10	1e-3	True	Adam

Table 7: Hyper-parameters for fine-tuning.



Our Proposed GLProtein

ESM-2



KeAP

ProtBert

Figure 9: Embedding visualizations of GLProtein, ESM-2, KeAP and ProtBert on SCOPe database.

A.9 Time Complexity Analysis

We provide a more specific complexity analysis as follows: protein encoder operates at approximately $O(L^2 d)$, where L is the length of protein sequence and d is the embedding dimension. Triplet protein sampling operates at approximately $O(L^3)$, reducing the complexity to $O(L^2)$ by TM-Vec. Triplet loss operates at approximately $O(3Ld) \to O(Ld)$. Protein 3D distance encoding operated at approximately $O(KL^2 d)$, where K is the number of Gaussian Basis kernels. Substructure-based molecular encoding operates at approximately $O(Ld)$. Protein decoder operates at approximately $O(L^2 d)$. Total computation cost operated at $O_{total} = O_{encoder} + O_{global} + O_{local} + O_{decoder} = O(L^2 d) + O(L^2) + O(Ld) + O(KL^2 d) + O(L^2 d) = O((K+1)L^2 d)$.

A.10 Hyper-parameters for Fine-tuning

The hyper-parameters for fine-tuning are provided in the Table 7. Specifically, we follow the hyper-parameter settings in GNN-PPI (Lv et al., 2021) for PPI prediction. For protein binding affinity prediction and semantic similarity inference, we follow the fine-tuning configuration in PROBE (Unsal et al., 2022).