Enhancing Protein-Ligand Binding Affinity Prediction via Parameter-Efficient Fine-Tuning of Protein and Chemical Language Models

Ruikang Li[1], Jiaxian Yan[1], Kai Zhang[1(✉)], Yanjiang Chen[1], Qi Liu[1], Min Gao[2], and Enhong Chen[1]

[1] State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China, Hefei, China
[2] The First Affiliated Hospital of University of Science and Technology of China, Hefei, China
{liruikang,jiaxianyan}@mail.ustc.edu.cn, kkzhang08@ustc.edu.cn
yjchen@mail.ustc.edu.cn, qiliuql@ustc.edu.cn, gmbeauty@163.com
cheneh@ustc.edu.cn

Abstract. Predicting protein-ligand binding affinity (PLBA) is a crucial task in drug discovery. However, the performance and practicality of existing deep learning models are limited due to the scarcity of high-quality data. Protein language models (PLMs) and chemical language models (CLMs) offer a promising alternative for molecular representation, with the potential to significantly improve predictive accuracy. Nevertheless, the lack of effective integration frameworks has limited the full potential of PLMs and CLMs for PLBA and existing models often overlook the heterogeneity between language models and downstream tasks. To address these issues, we propose two frameworks utilizing parameter-efficient fine-tuning (PEFT) methods for PLBA prediction. The first framework **(KLG)** integrates **K**nowledge learned from **L**anguage models into advanced **G**eometric graph networks. The second framework **(LCB)** leverages **L**anguage models to generate representations of proteins and ligands, employing **C**ross-attention mechanisms and a four-**B**ranch neural network to process the features. To thoroughly explore the potential of PLMs and CLMs in the PLBA task, we fine-tune them in both frameworks using various PEFT methods, including Adapters, LoRA, BitFit, and QLoRA. Experimental results demonstrate the effectiveness and practicality of both frameworks, and we hope our work will inspire further applications of PLMs and CLMs in drug discovery.

Keywords: Protein-ligand binding affinity · Protein language models · Chemical language models · PEFT · Geometric graph neural networks

1 Introduction

In drug development, the primary focus is on proteins and ligand small molecules [32], where a critical step involves screening ligand molecules (small molecule

drugs) that bind most closely to proteins (targets) [26]. The protein-ligand binding affinity (PLBA) quantifies the strength of this interaction, with higher binding affinity indicating more favorable interactions [18]. Therefore, accurately and efficiently predicting PLBA is crucial for identifying promising drug candidates, significantly accelerating drug development and reducing associated costs [23]. Many advanced methods have been proposed to address PLBA prediction. Traditional empirical formula-based approaches were the first to emerge [8]. They established a foundation for computational tools in drug discovery, despite their limitations in accuracy. Building on these early methods, machine learning and deep learning methods were introduced [22].

Particularly with the rise of deep learning, models have enhanced their ability to autonomously learn complex features from data, leading to improved accuracy. Deep learning methods for predicting PLBA can be broadly categorized into structure-based methods and sequence-based methods. Structure-based methods, represented by geometric graph neural networks and 3D-CNNs (e.g., GIGN [27], OnionNet [33]), effectively capture 3D structural information and interactions, including both intramolecular and intermolecular interactions. In contrast, sequence-based methods (e.g., DeepDTAF [34]) utilize one-dimensional (1D) sequences, enabling the identification of long-range interactions.

Despite these advances, three key challenges remain. First, the severe scarcity of high-quality data, particularly costly three-dimensional structural data, limits the performance and practicality of existing deep learning models. The rise of large language models has introduced protein language models (PLMs) (e.g., ESM2 [15]) and chemical language models (CLMs) (e.g., ChemBERTa [3]), which are trained on large-scale datasets of protein sequences and ligand SMILES strings. These models capture valuable features, and when integrated with traditional deep learning methods, they enhance predictive performance, offering promising solutions to the data scarcity problem.

The second challenge is the absence of effective frameworks to effectively utilize PLMs and CLMs for the PLBA task. Several studies have applied PLMs and CLMs to PLBA task with promising results. Research [24] shows that integrating knowledge from PLMs significantly enhances the capabilities of geometric networks. Frameworks like PLAPT [19] use pre-trained transformers to process one-dimensional protein sequences and ligand SMILES strings, improving binding affinity predictions. However, these approaches leverage PLMs and CLMs in a relatively straightforward manner and may not fully exploit the rich feature representations and contextual information these models offer.

The third challenge lies in the significant heterogeneity between PLMs, CLMs and specific PLBA task, which can introduce a certain degree of uncertainty. Furthermore, the increasing number of model parameters and the length of protein sequences make fine-tuning computationally demanding, which is highly unfriendly for applications. Therefore, it is necessary to adopt parameter-efficient fine-tuning (PEFT) methods from the NLP domain to address this issue. PEFT models can achieve comparable or even better performance than traditional fine-

tuning while reducing memory requirements [21]. Nonetheless, no one has yet fine-tuning PLMs and CLMs within a unified framework simultaneously.

To address these challenges, we have specifically designed two novel frameworks for PLBA prediction that seamlessly integrate PLMs, CLMs, and PEFT methods, based on structure-based deep learning methods and sequence-based deep learning methods, respectively.

The first framework **(KLG)** is an enhancement of structure-based deep learning methods, integrating the knowledge of language models into state-of-the-art geometric graph neural network. It utilizes PLMs and CLMs to generate features for protein amino acids and ligand atoms, which are used as node embeddings and input into the advanced geometric graph neural network for processing. To enhance interpretability, we process proteins, ligands, and protein-ligand complexes and reflect binding affinity as energy changes.

The second framework **(LCB)** is an improvement upon sequence-based deep learning methods, generating two types of features for the sequence through protein and chemical language models: a feature matrix and overall sequence representations. We utilize a cross-attention mechanism to capture the interaction between the protein and ligand, ultimately predicting binding affinity through a four-branch neural network.

The main contributions of this work are summarized as follows:

- We propose two frameworks for PLBA prediction specifically tailored for structure-based methods and sequence-based deep learning methods that incorporate PLMs and CLMs. These frameworks address the challenges of data scarcity by utilizing the rich feature representations from PLMs and CLMs for enhanced prediction accuracy.
- To the best of our knowledge, this is the first work to simultaneously fine-tune both PLMs and CLMs using PEFT methods, and to conduct extensive ablation studies to systematically investigate the impact of PLMs, CLMs, and their PEFT methods on the PLBA task.
- Both frameworks strongly demonstrate effectiveness and significant practical potential, achieving state-of-the-art results on extensive datasets while exhibiting robust generalization capabilities, thereby laying a solid foundation for further applications of PLMs and CLMs in drug discovery.

2 Related Work

2.1 Advanced deep learning methods for PLBA prediction

Wu et al [24] were the first to integrate knowledge from protein language models (PLMs) into geometric graph networks. However, their approach used only frozen PLMs and truncated proteins based on binding pockets, limiting the full potential of PLMs. Furthermore, many current geometric GNNs heavily rely on optimized structural date. PAMNet [30] introduces a physics-informed bias to enhance structural awareness. Due to its ability to accurately and efficiently represent three-dimensional (3D) molecules of varying sizes and types, our first

framework **(KLG)** is built upon PAMNet. PLAPT [19] is a fully sequence-based method that uses frozen PLMs and CLMs for transfer learning. PLAPT performs well on larger datasets, but its model complexity is relatively simple. Our second framework **(CLB)** is an improvement over PLAPT.

2.2 Protein and Chemical language models

Notable protein language models include ESM-1, ESM-2, ESMFold [15], ProtT5 [7] and ProtBERT [2]. We utilize ESM-2 and ProtBERT to extract protein features. ESM-2 excels in capturing fine-grained amino acid-level information, while ProtBERT is well-suited for generating comprehensive embeddings of entire protein sequences. For ligands, the most prominent chemical language model is ChemBERTa [3], which is developed on the RoBERTa transformer architecture specifically designed for modeling ligand SMILES strings. We employ ChemBERTa to extract ligand features.

2.3 PEFT methods for protein language models

Parameter-efficient fine-tuning (PEFT) achieves comparable or even superior performance to traditional fine-tuning while significantly reducing the number of tunable parameters [21]. In protein-protein interaction (PPI) prediction, PEFT models have been shown to outperform traditional methods [1]. PLM-PLI [31] fine-tunes pre-trained protein language models, improving performance and reducing time costs across three PLI tasks.

Common PEFT methods include Adapters [9], LoRA [10], BitFit [28], and Qlora [5]. Among these, LoRA stands out as the top-performing PEFT method in natural language processing and is increasingly used for fine-tuning protein language models. Research has shown impressive results in protein property prediction tasks by fine-tuning ESM-2 using LoRA [20]. Exploring PEFT methods that are more suitable for PLMs and CLMs remains a worthwhile research topic.

3 Method

As illustrated in Figure 1 and Figure 2, we propose two advanced frameworks **KLG** and **CLB**, designed to enhance predictive performance in the PLBA task through protein and chemical language models. These frameworks integrate parameter-efficient fine-tuning (PEFT) methods for language models. Both proposed frameworks demonstrate effectiveness and practicality.

3.1 Problem formulation

Protein-ligand binding affinity prediction (PLBA) is a regression task. Given protein structure P, protein sequence S_p, ligand structure L and ligand SMILES string S_l, where the protein and ligand structures can form a complex (P, L). our goal is to train a model $f(P, L, S_p, S_l)$ to predict a scalar value that represents the binding affinity y.

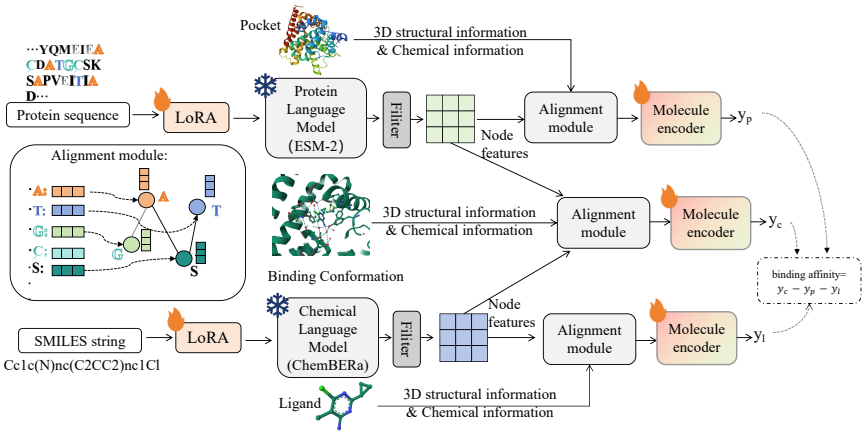

Fig. 1: Overview of KLG Framework. It consists of three main components: the Alignment Module, Molecule Encoder, and Prediction Module.

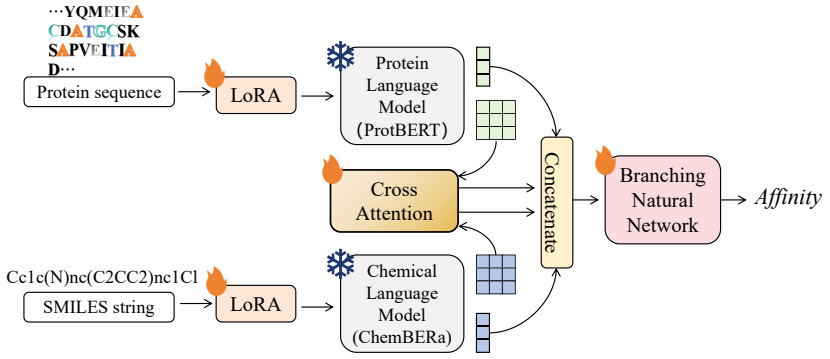

Fig. 2: Overview of CLB Framework. It is composed of three main components: Feature Extraction, Cross-Attention Mechanism, and Branching Neural Network.

3.2 KLG framework

Alignment module:

The Alignment Module integrates knowledge from protein and chemical language models fine-tuned with PEFT methods into the geometric network.

Integration of language models into geometric network: We input the amino acid sequences $S_p = \{x1x2 \dots xn\}$ into the ESM-2 protein language model to extract the representations of each residue from the *last_hidden_state*, denoted as $\mathbf{h}'_p \in \mathbb{R}^{N \times \psi_{PLM}}$, where $\psi_{PLM} = 1280$ and N is the number of amino acids in the sequence. It is worth mentioning that we focus on the features of amino acids forming the protein pocket with the ligand. For example, if the amino acid sequence has 20 residues, we may only need the ones at positions $[5, 11, 12, 13, 19]$. We filter the representations to obtain the desired features, then concatenate

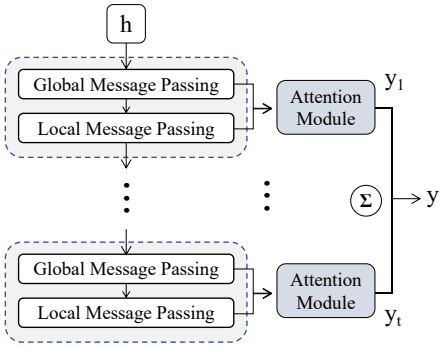

Fig. 3: Molecule encoder

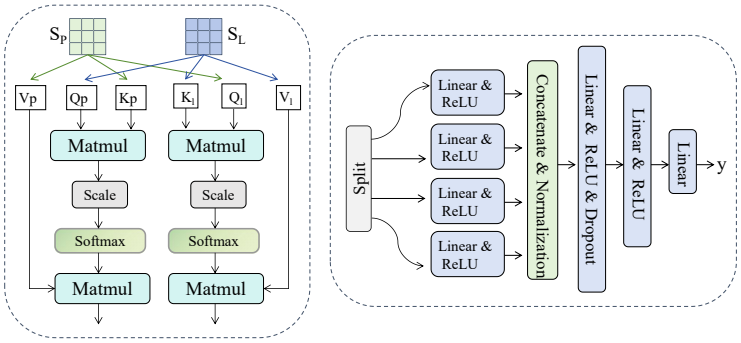

Fig. 4: Cross-attention mechanism (left) and Branching neural network (right)

them with the initial features $\mathbf{h}_p^0 \in \mathbb{R}^{n\times 21}$ (the one-hot encoding of the amino acid types) to produce new node input features $\mathbf{h}_p$ for the protein amino acid graph. Similarly, we input the SMILES strings $S_l = \{y1y2\ldots yn\}$ into the ChemBERTa2 chemical language model, extracting the representations of each token from the *last_hidden_state*, denoted as $\mathbf{h}_l' \in \mathbb{R}^{N\times\psi_{CLM}}$, where $\psi_{CLM} = 384$. Since ChemBERTa2 is based on the RoBERTa transformer architecture, it generates embeddings for each token. For example, with "NC(=N)N", the number of token is 7 (greater than the number of atoms, which is 4). Therefore, we need to filter the embeddings, retaining only those corresponding to heavy atoms. At this point, the representations of each atom are denoted as $\mathbf{h}_l' \in \mathbb{R}^{n\times\psi_{CLM}}$, where n equals the number of nodes in the ligand atom graph. We concatenate these representations with the initial features to obtain the input features $\mathbf{h}_l$ for the ligand atom graph. The key difference is that the initial features for each atom are obtained using the RDKit package [12], rather than simply one-hot encoding of the atom types.

Parameter-efficient Fine-tuning: We primarily focus on LoRA as our parameter-efficient fine-tuning method. LoRA introduces two low-rank matrices, A and B, into each adapted weight matrix. Given a weight matrix $\mathbf{W} \in \mathbb{R}^{d\times k}$,

LoRA adds new parameters, $\mathbf{A} \in \mathbb{R}^{r \times k}$ and $\mathbf{B} \in \mathbb{R}^{d \times r} (r << d, k)$. For input $\mathbf{x} \in \mathbb{R}^k$, LoRA modifies the stand forward pass of the layer from the original $h = Wx$ to the new form $h = Wx + BAx$. During back-propagation, only the weights of A and B are updated, while the weights of W remain unchanged. BAx is scaled by the quantity $\frac{\alpha}{r}$. Further parameter details will be introduced in subsequent sections. We also conducted experiments with Adapters [1], QLoRA [5] and BitFit [28]. Adapters insert small bottleneck layers into each layer of the model, Specifically, for a layer with input features$\mathbf{x} \in \mathbb{R}^k$, two projection matrices: $\mathbf{W}_{\text{down}} \in \mathbb{R}^{d \times r}$ and $\mathbf{W}_{\text{up}} \in \mathbb{R}^{r \times d}$ are introduced. The original forward pass $h = f(Wx)$ is modified by the Adapters to the following form:$h = W_{\text{up}} \sigma \left(W_{\text{down}} x\right) + x$. QLoRA builds upon LoRA by quantizating the base model's weights to a lower precision (typically 4 bits), reducing memory consumption. Specifically, QLoRA modifies the forward pass of the layer from $h = Wx$ to the new form $h = Q(W)x + BAx$, where $h = Q(W)x + BAx$ represents the quantized version of the weight matrix W. The BitFit method is relatively simple, it only updates the bias terms in the model. Given a layer with a weight matrix $\mathbf{W} \in \mathbb{R}^{d \times k}$ and a bias term $\mathbf{b} \in \mathbb{R}^d$, the forward pass $h = Wx + b$ remains unchanged. During back-propagation, only the bias terms b is updated, leaving the weights W are kept frozen.

Molecule encoder:

We adopt PaxNet (Figure 3) as our molecule encoder, as it effectively encodes both macromolecules and small molecules. Given a 3D molecular graph $\mathcal{G} = (\mathcal{H}, \mathcal{C}, \mathcal{E})$. Each node i is associated with an atom feature vector h_i and an atom coordinate c_i. The edge set $\mathcal{E}^L$ is constructed based on spatial distances between atoms. An edge is considered to exist between i-th node and j-th node if the distance $\left\|c_i^L - c_j^L\right\|_2$ is less than a predefined cutoff distance.

PaxNet uses global and local message passing to propagate information across the graph and process node embeddings. Each hidden layer performs both global and local message passing operations before moving to the next layer.

We first introduce Global Message Passing. In the global phase, we construct a molecular graph $G_{\boldsymbol{Global}}$ based on the cutoff distance $d_{\boldsymbol{Global}}$. For the PLBA task, $d_{\boldsymbol{Global}}$ is typically set to 12.0 Å. The message passing operations for neighboring nodes in $G_{\boldsymbol{Global}}$ are defined as follows:

$$\boldsymbol{m}_{ji}^{t-1} = \text{MLP}_m \left(\left[\boldsymbol{h}_j^{t-1} \left\| \boldsymbol{h}_i^{t-1} \right\| \boldsymbol{e}_{ji} \right] \right) \tag{1}$$

$$\boldsymbol{h}_i^t = \boldsymbol{h}_i^{t-1} + \sum_{j \in \mathcal{N}(i)} \boldsymbol{m}_{ji}^{t-1} \odot \phi_d \left(\boldsymbol{e}_{ji} \right) \tag{2}$$

Here, i and j are neighboring nodes in $G_{\boldsymbol{Global}}$, and e_{ji}, represents the edge embedding, encoding pairwise distance information using radial basis functions.

Next, we introduce Local Message Passing, which focuses on capturing finer 3D structural details. $G_{\boldsymbol{Local}}$ is defined by a cutoff distance $d_{\boldsymbol{Local}}$=2.0 Å. This phase incorporates additional embeddings to capture detailed 3D information,

including angular information. The message passing operations are as follows:

$$\boldsymbol{m}_{ji}^{\prime t-1} = \boldsymbol{m}_{ji}^{t-1} + \sum_{j' \in \mathcal{N}(i) \setminus \{j\}} \boldsymbol{m}_{j'i}^{t-1} \odot \phi_d\left(\boldsymbol{e}_{j'i}\right) \odot \phi_\theta\left(\boldsymbol{\theta}_{j'i,ji}\right) \tag{3}$$

$$+ \sum_{k \in \mathcal{N}(j) \setminus \{i\}} \boldsymbol{m}_{kj}^{t-1} \odot \phi_d\left(\boldsymbol{e}_{kj}\right) \odot \phi_\theta\left(\boldsymbol{\theta}_{kj,ji}\right) \tag{4}$$

$$\boldsymbol{h}_i^t = \boldsymbol{h}_i^{t-1} + \sum_{j \in \mathcal{N}(i)} \boldsymbol{m}_{ji}^{\prime t-1} \odot \phi_d\left(\boldsymbol{e}_{ji}\right) \tag{5}$$

Prediction Module:

Each hidden layer in PaxNet includes an attention module to generate node-level predictions. The node-level output is derived by taking the attention-weighted sum of the hidden states:

$$y_{\text{out},i}^t = \sum_m \alpha_{m,i}^t \left(\boldsymbol{W}_{\text{out}t_m}^t \boldsymbol{h}_{m,i}^t\right) \tag{6}$$

The final graph-level prediction is computed by summing the node-level outputs across layers and nodes:

$$y = \sum_{i=1}^{N} \sum_{t=1}^{T} y_{\text{out},i}^t \tag{7}$$

We maintain the original structure from the PaxNet paper. We create three weight-sharing replica networks, each intended to predict the target values for the protein-ligand complex, protein pocket, and ligand. The binding affinity is calculated by the following energy change formula: $y{=}y_{complex}{-}y_{pocket}{-}y_{ligand}$.

3.3 LCB framework

Feature extraction:

We use ProtBERT to extract features from protein sequences, where each sequence is represented as a string of amino acid residues. Since proteins often consist of multiple chains, but only one chain typically participates in ligand binding, we focus on the chain involved in binding. From ProtBERT, we obtain pooled features $\mathbf{s}_p^p \in \mathbb{R}^{1\times\psi_{PLM}}$ via the *pooler_output*, and feature matrix $\mathbf{S}_p^m \in \mathbb{R}^{N\times\psi_{PLM}}$ via the *last_hidden_state*. Similarly, ChemBERTa is used to extract both pooled features $\mathbf{s}_l^p \in \mathbb{R}^{1\times\psi_{CLM}}$ and feature matrix $\mathbf{S}_l^m \in \mathbb{R}^{M\times\psi_{CLM}}$ from ligands. A cross-attention mechanism is applied to process the protein feature matrix $\mathbf{S}_p^m \in \mathbb{R}^{N\times\psi_{PLM}}$ and the ligand feature matrix $\mathbf{S}_l^m \in \mathbb{R}^{M\times\psi_{CLM}}$. The resulting embeddings are concatenated with the pooled features $\mathbf{s}_p^p \in \mathbb{R}^{1\times\psi_{PLM}}$ and $\mathbf{s}_l^p \in \mathbb{R}^{1\times\psi_{CLM}}$ before being passed into the prediction module (Branching Neural Network). The parameter-efficient fine-tuning (PEFT) method used in LCB framework is consistent with KLG, as previously described.

Cross-attention mechanism:

Due to the ability of the cross-attention mechanism (Figure 4) to construct explicit interactions between two independent inputs and fully leverage their

correlations [11], we implement a cross-attention mechanism to establish interactions between protein sequences and ligand SMILES which can fully harness the potential of protein language models and chemical language models. Formally, the cross-attention mechanism is defined in the following way:

$$Q = S_p W^Q, K = S_l W^K, V = S_l W^V \tag{8}$$

$$S'_p = \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{9}$$

$$Q = S_l W^Q, K = S_p W^K, V = S_p W^V \tag{10}$$

$$S'_l = \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{11}$$

The resulting outputs, S'_p and S'_l, will undergo max pooling to form one-dimensional embeddings, then pass through a linear layer and ReLU activation function, and finally proceed to the prediction module for further processing.

Branching neural network:

The prediction module is a regression model based on BNN (Figure 4). The input one-dimensional feature vector is split into four parallel streams. Each stream is processed through its respective linear layer followed by a ReLU activation function. The outputs from these streams are then combined, normalized, and pass through three additional linear layers, each followed by a ReLU activation. To mitigate overfitting, a dropout layer is applied. The final output is a scalar value, y, representing the binding affinity metric.

4 Experiments

4.1 Experimental setup

Datasets We utilized two well-established and authoritative datasets to comprehensively evaluate the model's performance on PLBA task.

PDBbind v2016 is a well-known benchmark for PLBA task [25], consisting of three subsets: the general set (13,283 complexes), the refined set (4,057 high-quality complexes selected from the general set), and the core set (290 highest-quality complexes). To ensure fairness and robust evaluation, we used the core set as the test set. While the refined set, after removing overlapping parts, was randomly shuffled and split into training and validation sets with a 9:1 ratio.

In this paper, to ensure data accuracy and consistency, we employed the ProDy [29] and OpenBabel [16] packages to extract protein sequences, validating the results against those obtained using PyMOL [4]. For efficiency, we truncated the protein sequences. As shown in the Figure 5 and Figure 6 below, we analyzed the lengths of protein sequences in the PDBbind refined and core sets.

CSAR-HiQ is commonly used to evaluate model generalization after training on the PDBbind refined set [13]. This publicly available dataset contains 3D protein-ligand complexes with experimental affinity labels [6] and serves as an important supplementary dataset for PLBA task. Since it overlaps with the

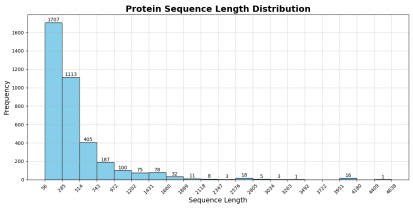

Fig. 5: Protein sequence length distribution (refined set)

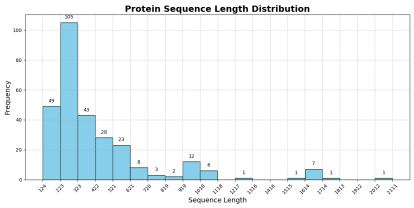

Fig. 6: Protein sequence length distribution (core set)

PDBbind dataset, we created an independent test set of 135 samples from CSAR-HiQ, excluding those already present in the PDBbind v2016 refined set.

Evaluation Metrics We used the MSE (Mean Squared Error) as the loss function. During testing, prediction performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Standard Deviation (SD), and Pearson's correlation coefficient (R) [25].

Implementation Details In the KLG framework, we set the pocket truncation distance to 8 Å, and the molecule encoder has 3 layers. We fine-tuned the PLMs and CLMs with the LoRA configuration (r=8, lora_alpha=32, lora_dropout=0.1). QLoRA applied 4-bit quantization, and the Adapter's reduction factor was set to 16. The KLG model was trained with Adam optimizer, learning rate of 5×10^{-4}, batch size of 8, and hidden layer dimension of 128. For the CLB framework, the configurations and PEFT settings were similar to KLG, with all branches unified to 256, 512 dimensions. Training ran for 800 epochs with early stopping after 80 epochs if no improvement was observed. Both frameworks were implemented in PyTorch [17] and trained on four Tesla V100 GPUs. Details of the PLMs and CLMs used are in Table 1. Code is released at: `https://github.com/kronos7777777/KLG-CLB`.

Table 1: Utilized protein and chemical language models

Model	Architecture (Pretraining)	Number of Parameters (Encoder)	Encoder Layers	Emb Size	Huggingface Model Checkpoint
ESM2 35M	Enc	35 M	12	480	esm2_t12_35M_UR50D
ESM2 150M	Enc	150 M	30	640	esm2_t30_150M_UR50D
ESMFold	Enc	690 M(+3.0 B)	48(+36)	-	esmfold_v1
ProtBert	Enc	420M	30	1024	prot_bert
ChemBERTa	Enc	442M	6	768	ChemBERTa-zinc-base-v1
ChemBERTa2	Enc	77 M	3	384	ChemBERTa-77M-MTR

Table 2: Test Performance on PDBbind Core Set and CSAR-HiQ Dataset. The best result is in bold.

Method		PDBbind core set				CSAR-HiQ dataset			
		RMSE ↓	MAE ↓	SD ↓	R ↑	RMSE ↓	MAE ↓	SD ↓	R ↑
ML-based	LR	1.675 (0.000)	1.358 (0.000)	1.612 (0.000)	0.671 (0.000)	2.071 (0.000)	1.622 (0.000)	1.973 (0.000)	0.652 (0.000)
	SVR	1.555 (0.000)	1.264 (0.000)	1.493 (0.000)	0.727 (0.000)	1.995 (0.000)	1.553 (0.000)	1.911 (0.000)	0.679 (0.000)
	RF-Score	1.446 (0.008)	1.161 (0.007)	1.335 (0.010)	0.789 (0.003)	1.947 (0.012)	1.466 (0.009)	1.796 (0.020)	0.723 (0.007)
CNN-based	Pafnucy	1.585 (0.013)	1.284 (0.021)	1.563 (0.022)	0.695 (0.011)	1.939 (0.103)	1.562 (0.094)	1.885 (0.071)	0.686 (0.027)
	OnionNet	1.407 (0.034)	1.078 (0.025)	1.391 (0.038)	0.768 (0.014)	1.927 (0.071)	1.471 (0.093)	1.839 (0.071)	0.690 (0.044)
Sequence-based	DeepDTA	1.443 (0.030)	1.148(0.028)	1.445(0.028)	0.749 (0.018)	-	-	-	-
	DeepDTAF	1.355(0.030)	1.073(0.029)	1.337(0.028)	0.789(0.018)	2.765(0.050)	2.318(0.045)	1.679(0.045)	0.543(0.035)
GNN-based	MAT	1.457 (0.037)	1.154 (0.037)	1.445 (0.033)	0.747 (0.013)	1.879 (0.065)	1.435 (0.058)	1.821 (0.089)	0.715 (0.030)
	DimeNet	1.453 (0.027)	1.138 (0.026)	1.434 (0.023)	0.752 (0.019)	1.805 (0.036)	1.338 (0.026)	1.764 (0.032)	0.723 (0.020)
	CMPNN	1.408 (0.028)	1.117 (0.031)	1.399 (0.025)	0.765 (0.009)	1.839 (0.096)	1.411 (0.064)	1.767 (0.103)	0.720 (0.051)
	IGN	1.519 (0.055)	1.187 (0.042)	1.513 (0.052)	0.718 (0.023)	2.051 (0.077)	1.604 (0.074)	1.944 (0.095)	0.687 (0.046)
	SIGN	1.316 (0.031)	1.027 (0.025)	1.312 (0.035)	0.797 (0.012)	1.753 (0.031)	1.327 (0.020)	1.709 (0.044)	0.754 (0.014)
	GIANT	1.269 (0.020)	0.999 (0.018)	1.265 (0.024)	0.814 (0.008)	1.666 (0.024)	1.242 (0.030)	1.633 (0.034)	0.779 (0.011)
Ours	KLG	**1.255 (0.023)**	**0.983 (0.018)**	**1.243 (0.023)**	**0.824 (0.008)**	**1.609 (0.036)**	**1.210 (0.032)**	**1.542 (0.038)**	**0.805 (0.020)**
	LCB	1.329 (0.024)	1.023(0.020)	1.301 (0.030)	0.801 (0.010)	1.712 (0.040)	1.329 (0.031)	1.661 (0.049)	0.758 (0.025)

4.2 Experimental Results

We trained our models on the PDBbind refined set and evaluated their performance on the PDBbind core set and CSAR-HiQ set, as detailed in Table 2. We compare KLG and LCB with four families of methods. The results clearly demonstrate that KLG outperformed all existing state-of-the-art methods across all metrics on both public datasets [14]. These results indicate that integrating PLMs and CLMs into geometric graph neural networks has led to state-of-the-art performance. In the context of the critical scarcity of high-quality 3D data, the KLG framework leverages PLMs, CLMs, and PEFT to further enhance model performance. Additionally, the sequence-based method CLB demonstrated comparable performance to 3D geometric structure-based methods across all metrics in both public datasets, despite the limited training data. Given that one-dimensional data is relatively easier to obtain and trains faster, CLB holds significant practical value for future research and applications.

4.3 Ablation Studies

In Table 3, we perform an ablation study on the KLG framework, exploring the effects of PLMs, CLMs, and various PEFT methods. Key findings are as follows:

Without fine-tuning with PEFT methods, the RMSE values of KLG on the two public datasets are lower when using information from either protein language models (PLMs) or chemical language models (CLMs) individually, compared to when both are integrated. This highlights the importance of combining both PLM and CLM information for accurate binding affinity predictions within a well-structured framework, despite their differing informational content. Furthermore, as the model size increases, the performance of both PLMs and CLMs improves, thereby enhancing the KLG framework's effectiveness. Thus, integrating insights from large models into the geometric deep learning framework shows substantial promise.

Table 3: Ablation Study on the Impact of PLMs, CLMs, and PEFT Methods in the KLG Framework. The best result is in bold.

Ablation Study				PDBbind core set				CSAR-HiQ dataset			
level[1]	PLMS	CLMs	PEFT	RMSE↓	MAE↓	R↑	SD↓	RMSE↓	MAE↓	R↑	SD↓
a	-	-	-	1.263	0.987	0.815	1.261	-	-	-	-
r	-	-	-	1.314	1.027	0.798	1.318	-	-	-	-
r	ESM2-150	-	-	1.310	1.025	0.799	1.298	1.731	1.325	0.735	1.687
r	-	CB2	-	1.322	1.023	0.798	1.315	1.744	1.338	0.729	1.701
r	ESM2-35	CB2[p]	-	1.308	1.022	0.801	1.295	1.683	1.267	0.763	1.649
r	ESM2-150	CB2	-	1.301	1.020	0.805	1.290	1.668	1.252	0.759	1.589
r	ESM2Fold	CB2	-	1.282	1.017	0.819	1.259	1.643	1.245	0.781	1.580
r	ESM2-150	CB2	Bitft	1.279	1.009	0.811	1.261	1.629	1.227	0.769	1.577
r	ESM2-150	CB2	Adapter	1.265	1.001	0.810	1.253	1.624	1.220	0.775	1.575
r	ESM2-150	CB2	Qlora	1.262	0.993	0.815	1.251	1.618	1.217	0.776	1.563
r	ESM2-150	CB2	LoRA	**1.255**	**0.983**	**0.824**	**1.243**	**1.609**	**1.210**	**0.805**	**1.542**

[1]: The protein processed as a residue graph is denoted as "r", and the protein processed as an atom graph is denoted as "a". [p]: ChemBERTa2; ChemBERTa-77M-MTR. [q]: ChemBERTa; ChemBERTa-zinc-base-v1.

When employing PEFT methods, the RMSE values of KLG on the two public datasets are significantly higher than those without PEFT. Specifically, on the PDBbind dataset, compared to the previous method (level=r) that did not utilize PLMs, CLMs and PEFT, KLG achieved improvements of 4.5%, 4.3%, 3.3%, and 5.7% in RMSE, MAE, SD, and R, respectively. Furthermore, compared to the method that used PLMs and CLMs but did not use PEFT methods to finetuen, KLG demonstrated improvements of 3.5%, 3.6%, 2.4%, and 3.6% in RMSE, MAE, SD, and R, respectively. Among various PEFT methods, LoRA delivered the best performance, while BitFit, despite requiring the fewest parameters, performed the worst. Overall, employing PEFT methods for both PLMs and CLMs is essential, as it dramatically reduces the computational resources required for fine-tuning while allowing these models to fully adapt to downstream tasks.

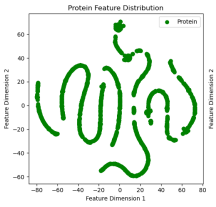

Fig. 7: Features visualization before fine-tuning

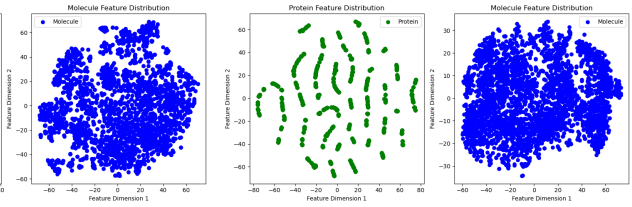

Fig. 8: Features visualization after 40 epochs of fine-tuning

In Table 4, we present the results of an ablation study on the CLB framework, exploring the impact of PLMs, CLMs, and PEFT various methods. Our key findings are summarized as follows:

Table 4: Ablation Study on the Impact of PLMs, CLMs, and PEFT Methods in the CLB Framework, The best result is in bold.

Ablation Study				PDBbind core set				CSAR-HiQ dataset			
CA[t]	PLMs	CLMs	PEFT	RMSE↓	MAE↓	R↑	SD↓	RMSE↓	MAE↓	R↑	SD↓
-	ProtBert	CB	-	1.705	1.468	0.619	1.699	-	-	-	-
✓	ESMFold	CB[q]	-	1.480	1.157	0.749	1.476	1.906	1.479	0.707	1.861
✓	ProtBert	CB	-	1.507	1.185	0.694	1.495	1.952	1.519	0.697	1.881
✓	✓[w]	CB	LoRA	1.875	1.504	0.470	1.821	-	-	-	-
✓	ProtBert	CB	BitFit	1.404	1.099	0.792	1.378	1.741	1.381	0.728	1.762
✓	ProtBert	CB	Adapter	1.346	1.047	0.784	1.328	1.776	1.369	0.736	1.725
✓	ProtBert	CB	Qlora	1.335	1.038	0.793	1.312	1.734	1.355	0.746	1.695
✓	ProtBert	CB	LoRA	**1.329**	**1.023**	**0.801**	**1.301**	**1.712**	**1.329**	**0.758**	**1.661**

[t]: Whether to use the cross-attention mechanism. [q]: ChemBERTa: ChemBERTa-zinc-base-v1. [w]: Fine-tuning only the chemical language model.

In the absence of fine-tuning with PEFT methods and without a cross-attention mechanism, models relying solely on one-dimensional data exhibit poor performance, particularly when trained on limited datasets. This suggests that the cross-attention mechanism significantly enhances model performance by fully leveraging the information within PLMs and CLMs. Moreover, as the parameter count of PLMs and CLMs increases, the performance of CLB improves substantially, as it relies entirely on the features provided by these models.

After fine-tuning with PEFT techniques, the RMSE values of CLB on the two public datasets show significant improvement (an average of 11%), indicating that the CLB framework greatly benefits from PEFT methods. As illustrated in Figure 7 and Figure 8, PEFT improves the feature distribution of proteins and ligands, allowing the model to learn richer and more detailed content and produce features better suited for binding affinity prediction task.

The ranking of PEFT methods in the CLB framework aligns with the experimental results observed in the KLG framework. Notably, fine-tuning either PLMs or CLMs in isolation can lead to decreased performance, likely due to data limitations and the cross-attention mechanism's inability to capture meaningful interaction information. CLB leverages the advantages provided by PLMs and CLMs more directly, making these findings particularly significant.

4.4 Parameters Analysis

We conducted hyperparameter experiments on $d_{\boldsymbol{Local}}$ in the KLG framework (Figure 9), varying its value from 1 to over 5Å. In these experiments, $d_{\boldsymbol{Global}}$ was set to 12Å, and no fine-tuning was applied. The model achieved optimal performance with $d_{\boldsymbol{Local}} = 2$Å, effectively capturing both short- and long-range interactions within the 3D structure.

As shown in Figure 10 and 11. We tested various LoRA ranks and found that both KLG and CLB performed best with ranks of 4 and 8, while performance declined at ranks 1 and 2. These findings highlight the importance of fine-tuning

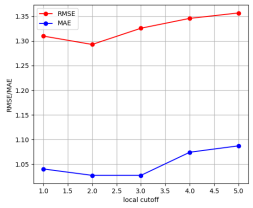

Fig. 9: The RMSE and MAE of KLG with different d_{Local}

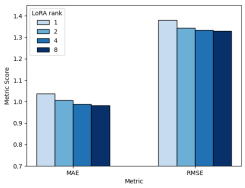

Fig. 10: Impact of LoRA rank (KLG)

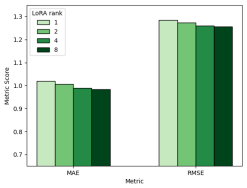

Fig. 11: Impact of LoRA rank (CLB)

PLMs and CLMs with PEFT methods, demonstrating that fully fine-tuned large language models can enhance performance on the downstream PLBA task.

5 Conclusion

In this paper, we thoroughly explore the potential of PLMs and CLMs to enhance PLBA prediction. To address the heterogeneity between language models and downstream tasks, as well as overcome computational resource limitations, we introduce PEFT methods for effectively fine-tuning. Furthermore, to tackle the challenge of lacking suitable frameworks, we propose two novel frameworks, KLG and CLB, based on structure-based and sequence-based deep learning methods, respectively. Extensive experiments conducted on two widely recognized benchmark datasets strongly demonstrate the practicality and effectiveness of both frameworks. Ablation studies highlight the critical role of PLMs, CLMs, and PEFT methods in improving the performance of both frameworks. We hope that our research establishes a solid foundation for the broader applications of PLMs and CLMs in drug discovery and other specific scientific tasks.

6 Acknowledgment

This research was partially supported by the National Natural Science Foundation of China (Grants No.62406303), Anhui Provincial Natural Science Foundation (No. 2308085QF229), Anhui Science and Technology Innovation Plan (No.202423k09020010) and the Fundamental Research Funds for the Central Universities (No. WK2150110034).

References

1. Balne, C.C.S., Bhaduri, S., Roy, T., Jain, V., Chadha, A.: Parameter efficient fine tuning: A comprehensive analysis across applications. arXiv preprint arXiv:2404.13506 (2024)
2. Brandes, N., Ofer, D., Peleg, Y., Rappoport, N., Linial, M.: Proteinbert: a universal deep-learning model of protein sequence and function. Bioinformatics **38**(8), 2102–2110 (2022)

3. Chithrananda, S., Grand, G., Ramsundar, B.: Chemberta: large-scale self-supervised pretraining for molecular property prediction. arXiv preprint arXiv:2010.09885 (2020)
4. DeLano, W.L., et al.: Pymol: An open-source molecular graphics tool. CCP4 Newsl. Protein Crystallogr **40**(1), 82–92 (2002)
5. Dettmers, T., Pagnoni, A., Holtzman, A., Zettlemoyer, L.: Qlora: Efficient finetuning of quantized llms. Advances in Neural Information Processing Systems **36** (2024)
6. Dunbar Jr, J.B., Smith, R.D., Damm-Ganamet, K.L., Ahmed, A., Esposito, E.X., Delproposto, J., Chinnaswamy, K., Kang, Y.N., Kubish, G., Gestwicki, J.E., et al.: Csar data set release 2012: ligands, affinities, complexes, and docking decoys. Journal of chemical information and modeling **53**(8), 1842–1852 (2013)
7. Elnaggar, A., Heinzinger, M., Dallago, C., Rehawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., Steinegger, M., et al.: Prottrans: Toward understanding the language of life through self-supervised learning. IEEE transactions on pattern analysis and machine intelligence **44**(10), 7112–7127 (2021)
8. Guedes, I.A., Pereira, F.S., Dardenne, L.E.: Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges. Frontiers in pharmacology **9**, 411637 (2018)
9. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International conference on machine learning. pp. 2790–2799. PMLR (2019)
10. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685 (2021)
11. Jin, Z., Wu, T., Chen, T., Pan, D., Wang, X., Xie, J., Quan, L., Lyu, Q.: Capla: improved prediction of protein–ligand binding affinity by a deep learning approach based on a cross-attention mechanism. Bioinformatics **39**(2), btad049 (2023)
12. Landrum, G.: Rdkit documentation. Release **1**(1-79), 4 (2013)
13. Li, R., Yan, J., Zhang, K.: Msp: A multi-level structures-based framework with multi-level pre-training for protein-ligand binding affinity prediction. In: 2024 International Conference on Computational Linguistics and Natural Language Processing (CLNLP). pp. 57–63. IEEE (2024)
14. Li, S., Zhou, J., Xu, T., Huang, L., Wang, F., Xiong, H., Huang, W., Dou, D., Xiong, H.: Giant: Protein-ligand binding affinity prediction via geometry-aware interactive graph neural network. IEEE Transactions on Knowledge and Data Engineering (2023)
15. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., et al.: Language models of protein sequences at the scale of evolution enable accurate structure prediction. bioRxiv (2022)
16. O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R.: Open babel: An open chemical toolbox. Journal of cheminformatics **3**, 1–14 (2011)
17. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
18. Rezaei, M.A., Li, Y., Wu, D., Li, X., Li, C.: Deep learning in drug design: protein-ligand binding affinity prediction. IEEE/ACM transactions on computational biology and bioinformatics **19**(1), 407–417 (2020)

19. Rose, T., Monti, N., Anand, N., Shen, T.: Plapt: Protein-ligand binding affinity prediction using pretrained transformers. bioRxiv pp. 2024–02 (2024)
20. Schmirler, R., Heinzinger, M., Rost, B.: Fine-tuning protein language models boosts predictions across diverse tasks. Nature Communications **15**(1), 7407 (2024)
21. Sledzieski, S., Kshirsagar, M., Baek, M., Dodhia, R., Lavista Ferres, J., Berger, B.: Democratizing protein language models with parameter-efficient fine-tuning. Proceedings of the National Academy of Sciences **121**(26), e2405840121 (2024)
22. Stepniewska-Dziubinska, M.M., Zielenkiewicz, P., Siedlecki, P.: Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. Bioinformatics **34**(21), 3666–3674 (2018)
23. Wang, X., Nie, F., Gao, Z., Li, G., Zhang, D., Zhang, J., Zhang, P., Wang, Z., Qu, A.: Studies on qsar models for the anti-virus effect of oseltamivir derivatives targeting h5n1 based on mix-kernel support vector machine. Chemometrics and Intelligent Laboratory Systems p. 105273 (2024)
24. Wu, F., Wu, L., Radev, D., Xu, J., Li, S.Z.: Integration of pre-trained protein language models into geometric deep learning networks. Communications Biology **6**(1), 876 (2023)
25. Yan, J., Ye, Z., Yang, Z., Lu, C., Zhang, S., Liu, Q., Qiu, J.: Multi-task bioassay pre-training for protein-ligand binding affinity prediction. Briefings in Bioinformatics **25**(1), bbad451 (2024)
26. Yan, J., Zhang, Z., Zhu, J., Zhang, K., Pei, J., Liu, Q.: Deltadock: A unified framework for accurate, efficient, and physically reliable molecular docking. arXiv preprint arXiv:2410.11224 (2024)
27. Yang, Z., Zhong, W., Lv, Q., Dong, T., Yu-Chian Chen, C.: Geometric interaction graph neural network for predicting protein–ligand binding affinities from 3d structures (gign). The journal of physical chemistry letters **14**(8), 2020–2033 (2023)
28. Zaken, E.B., Ravfogel, S., Goldberg, Y.: Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. arXiv preprint arXiv:2106.10199 (2021)
29. Zhang, S., Krieger, J.M., Zhang, Y., Kaya, C., Kaynak, B., Mikulska-Ruminska, K., Doruker, P., Li, H., Bahar, I.: Prody 2.0: increased scale and scope after 10 years of protein dynamics modelling with python. Bioinformatics **37**(20), 3657–3659 (2021)
30. Zhang, S., Liu, Y., Xie, L.: A universal framework for accurate and efficient geometric deep learning of molecular systems. Scientific Reports **13**(1), 19171 (2023)
31. Zhang, W., Hu, F., Li, W., Yin, P.: Does protein pretrained language model facilitate the prediction of protein-ligand interaction? Methods (2023)
32. Zhang, Z., Shen, W.X., Liu, Q., Zitnik, M.: Efficient generation of protein pockets with pocketgen. Nature Machine Intelligence pp. 1–14 (2024)
33. Zheng, L., Fan, J., Mu, Y.: Onionnet: a multiple-layer intermolecular-contact-based convolutional neural network for protein–ligand binding affinity prediction. ACS omega **4**(14), 15956–15965 (2019)
34. Zhu, Y., Zhao, L., Wen, N., Wang, J., Wang, C.: Datadta: a multi-feature and dual-interaction aggregation framework for drug–target binding affinity prediction. Bioinformatics **39**(9), btad560 (2023)