# ChemGLaM: Chemical Genomics Language Models for Compound-Protein Interaction Prediction

**Takuto Koyama**
Graduate School of Medicine
Kyoto University
koyama.takuto.82j@st.kyoto-u.ac.jp

**Hayato Tsumura**
Graduate School of Medicine
Kyoto University
tsumura.hayato.22x@st.kyoto-u.ac.jp

**Shigeyuki Matsumoto**
Graduate School of Medicine
Kyoto University
matsumoto.shigeyuki.4z@kyoto-u.ac.jp

**Ryunosuke Okita**
Undergraduate School of Pharmaceutical Sciences
Kyoto University
okita.ryunosuke.75y@st.kyoto-u.ac.jp

**Ryosuke Kojima**
Graduate School of Medicine
Kyoto University
kojima.ryosuke.8e@kyoto-u.ac.jp

**Yasushi Okuno**
Graduate School of Medicine
Kyoto University
okuno.yasushi.4c@kyoto-u.ac.jp

## Abstract

Accurate prediction of compound-protein interaction (CPI) is of great importance for drug discovery. For creating generalizable CPI prediction deep learning (DL) models, the expansion of CPI data through experimental validation is crucial. However, the cost associated with these experimental validations is a bottleneck. Recently developed large language models (LLMs) such as chemical language models (CLMs) and protein language models (PLMs) have emerged as foundation models, demonstrating high generalization performance in various tasks involving compounds and proteins. Inspired by this, we propose a chemical genomics language model, ChemGLaM, for predicting compound-protein interactions. ChemGLaM is based on the 2 independent language models, MoLFormer for compounds and ESM-2 for proteins, and fine-tuned for the CPI datasets using an interaction block with a cross-attention mechanism. ChemGLaM is capable of predicting interactions between unknown compounds and proteins with higher accuracy than existing CPI prediction models, demonstrating that combining the independently pre-trained foundation models is effective for obtaining sophisticated representation of compound-protein interactions. Furthermore, visualizing the learned cross-attention map can offer explainable insights into the mechanism of compound-protein interaction. This study emphasizes the potential of integrating the independent foundation models for the tasks of multi-modality such as CPI prediction.

***K*eywords** Compound-Protein Interaction · Language Model

## 1 Introduction

Predicting compound-protein interactions (CPIs) is a critical part of drug discovery [1]. Since experimentally validating CPIs is time-consuming and costly, computational methods are promising approaches for predicting CPIs [2]. While structure-based approaches such as molecular docking and molecular dynamics (MD) simulations can effectively predict CPIs, the requirements of reliable three-dimensional (3D) protein structures and high computational costs significantly hamper the applicability of large-scale CPI prediction.

Deep learning (DL) has recently emerged as a potent CPI prediction approach. Many of these models belong to chemical genomics-based (structure-free) approaches that integrate chemical space and genomic space by using two-dimensional

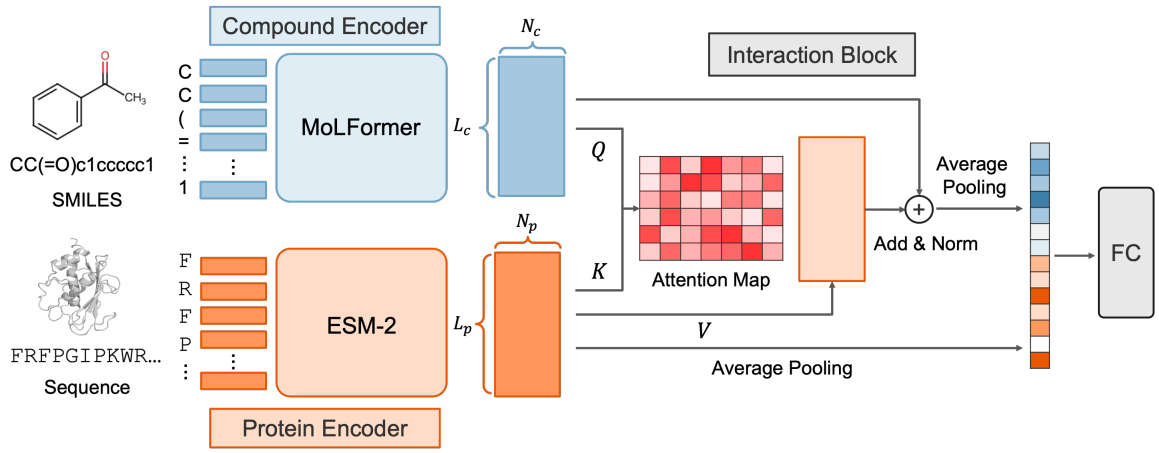


Figure 1: Schematic illustration of ChemGLaM framework

(2D) compound structures and protein sequences as inputs. Rapid growths of DL algorithms such as graph neural networks (GNNs)[3, 4], convolutional neural networks (CNNs)[5, 6], and transformer[7, 8] architectures have enabled DL models to obtain representations of 2D compound structures and protein sequences from publicly available or in-house CPI data.

Despite this promising progress, the effectiveness of DL models heavily relies on the availability of a training dataset. In CPI prediction, their generalization performance is often limited by the lack of available CPI data. For instance, Davis[9] and PDBbind[10], which are often used for creating CPI prediction models, have only 30K and 20K samples, respectively. BindingDB[11], which is one of the largest CPI datasets, contains at most several million samples when curated as in other works [8, 12]. These numbers are small compared with those of other fields such as images and natural language processing (NLP).

Self-supervised learning is one of the powerful approaches to addressing this data scarcity problem. Self-supervised learning is an unsupervised method that learns an inherent representation from the input itself. The advantage of self-supervised learning is it can learn representation from large-scale unlabeled datasets, which enhances the generalization performance of models when fine-tuned for downstream tasks [13, 14]. Chemical language models (CLMs) such as ChemBERTa[15] and MoLFormer[16] are successful self-supervised approaches in molecular property prediction. Protein language models (PLMs) such as ESM-2[17], and ProtTrans[18] are effective approaches that can be applied to various tasks related to protein sequences.

Inspired by these language models, we propose a chemical genomics language model, ChemGLaM for compound-protein interaction prediction. ChemGLaM consists of three parts: (1) Compound Encoder, (2) Protein Encoder, (3) Interaction Block (Figure 1). The Compound Encoder and the Protein Encoder are derived from MoLFormer and ESM-2 with pre-trained model parameters, respectively, and their embeddings are integrated by the Interaction Block with cross-attention mechanisms. We fine-tuned the model except for the Protein Encoder with CPI data (a) BindingDB [11], (b) Davis [9], (c) PDBbind [10], and (d) Metz [19]. Our proposed model demonstrates that combining individually pre-trained chemical and protein language models is effective for the tasks of multi-modality such as CPI prediction. Furthermore, visualizing the learned cross-attention map can offer explainable insights into the mechanism of compound-protein interaction.

## 2 ChemGLaM architecture

ChemGLaM consists of three parts, (1) Compound Encoder, (2) Protein Encoder, and (3) Interaction Block. The following section presents a description of these model architectures. Figure 1 illustrates the overview of the ChemGLaM framework.

### 2.1 Compound Encoder: MoLFormer

MoLFormer[16] is a large-scale chemical language model that learns molecule representation by training on simplified molecular input line entry system (SMILES)[20] sequences of 1.1 billion unlabelled molecules from the PubChem [21] and ZINC[22] datasets. MoLFormer-XL, which is the largest variant of MoLFormer, outperforms the state-of-the-art

chemical foundation models in various prediction tasks. We adopted the MoLFormer pre-trained with 10% of the MoLFormer-XL training dataset since the full model isn't available in the public repository.

### 2.2 Protein Encoder: ESM-2

Evolutionary Scale Modeling version 2 (ESM-2)[17] is a state-of-the-art protein language model widely used for various tasks such as predicting the structure and function of proteins. We chose the pre-trained ESM-2-T36-3B-UR50D, which consists of 36 transformer blocks with 3B parameters trained from UniRef50[23], as a Protein Encoder of the ChemGLaM.

### 2.3 Interaction Block

Interaction Block learns compound-protein interactions using both of the encoders above. Interaction Block uses the two outputs of compound embeddings $H_c = (L_c, N_c)$ and protein embedding $H_p = (L_p, N_p)$ as input, where $L_c$ and $L_p$ are token sizes of a compound SMILES and a protein sequence, $N_c = 768$ and $N_p = 2560$ are the dimensions of the embeddings. After having these embeddings, Interaction Block updates the compound's representation with multi-head cross-attention mechanisms.

$$Q = H_c, K = H_p, V = H_p \tag{1}$$

$$\mathrm{MultiHeadAttention}(Q, K, V) = \mathrm{Concat}(Z_1, Z_2, ..., Z_h)W^O \tag{2}$$

$$Z_i = \mathrm{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{3}$$

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

where $W^O$, $W_i^Q$, $W_i^K$, $W_i^V$ denote a linear layer. We set the number of heads as $h = 8$ and the number of dimensions $d_k$ for each head is $768/8 = 96$. Cross-attention enables the model to dynamically focus on different parts of the input data. This study applies cross-attention to the compound and protein embeddings to learn what and where to focus. The outputs of cross-attention are added by compound embeddings, followed by layer normalization[24].

$$X_c = \mathrm{LayerNormalization}\left(\mathrm{MultiHeadAtention}(H_c, H_p, H_p) + H_c\right) \tag{5}$$

Finally, the average pooled compound feature and the protein embedding are concatenated and passed through fully connected layers with skip connections[25].

## 3 Dataset

To evaluate the effectiveness of ChemGLaM, we prepared the following four datasets: (1) BindingDB, (2) Davis, (3) PDBbind, and (4) Metz. Table 1 shows the statistics on these CPI datasets.

Table 1: Summary of CPI datasets.

| Name | Task | Compounds | Proteins | Interactions | Positives | Negatives |
|---|---|---|---|---|---|---|
| BindingDB | Classification | 49752 | 813 | 61270 | 33777 | 27493 |
| Davis | Classification | 68 | 442 | 30056 | 9125 | 20931 |
| PDBbind | Regression | 14694 | 3705 | 18512 | - | - |
| Metz | Regression | 1423 | 170 | 35259 | - | - |

### 3.1 BindingDB

BindingDB [11] is a public, web-accessible database of measured binding affinities focusing chiefly on the interactions of small molecules and proteins. We utilized a binary classification dataset preprocessed by Gao et al. [26] with 33,777 positive examples and 27,493 examples.

### 3.2 Davis

Davis[9] dataset contains data on the binding affinities of a large number of kinase inhibitors against a significant portion of the human kinome. It includes experimentally measured binding constants (Kd values) for kinase-inhibitor pairs. We prepared the same versions of the Davis dataset from DeepDTA [6]. An example is annotated as positive if pKd is more than 5, and negative if pKd is 5 (indicating inactive in this dataset), resulting in 9,125 positive examples and 20,931 negative examples.

### 3.3 PDBbind

PDBbind[10] is a comprehensive collection of the binding affinities of proteins and nucleic acids in the Protein Data Bank (PDB). This database is specially curated for researchers involved in the development of molecular docking algorithms and for the study of protein-ligand interactions. We created a regression dataset from PDBbind version 2020 general set with 18,512 examples.

### 3.4 Metz

The Metz[19] is a dataset for the compound-protein interaction research, comprising 1,421 compounds and 156 proteins. About 42% of its compound-protein pairs include binding affinity data as pKi values, representing log-transformed kinase inhibition constants. This dataset's consistent experimental conditions and comparable binding affinities make it a valuable tool for evaluating predictive models in pharmacological studies. We created a regression dataset from Metz dataset that is the same version as Zhao et al[27].

## 4 Training and Evaluation

### 4.1 Training Details

Training ChemGLaM involves the fine-tuning of the Compound Encoder, which is the pre-trained MoLFormer-XL, along with Interaction Blocks. For the Protein Encoder, we decided to keep it fixed, considering that the ESM-2, which has 3 billion parameters, is excessively large for the CPI datasets. We trained ChemGLaM using Adam[28] optimizer with learning rate= $3 \times 10^{-5}$ over 10 epochs. For regression tasks, we additionally implemented *ExponentialLR* scheduler[29] where the learning rate is multiplied by 0.95 for each epoch. The batch size was set to 8 for all datasets. This training was processed in a single NVIDIA A100 40GB.

### 4.2 Model Evaluation: Zero-shot prediction

To verify the generalization performance for unknown CPIs, this study conducted model evaluations akin to zero-shot prediction, unless otherwise stated. More specifically, we implemented 5-fold cross-validation on test data where both compounds and proteins are unknown, thereby preventing the overestimation of model performance commonly observed in random splits.

### 4.3 Baseline Models

Hereafter are baseline models for the model evaluation. To evaluate the effectiveness of self-supervised pre-training in CPI prediction, we mainly focused on supervised learning approaches as baseline models.

**GraphDTA**[4] is one of the most frequently used graph-based method that uses a 2D graph for a compound and a sequence for a protein as input. The 2D graph is processed through Graph Isomorphism Network (GIN)[30] blocks and the sequence is passed through CNN blocks.

**TransformerCPI**[8] is a representative transformer-based method that utilizes atom sequence embeddings from 2D graphs of compounds and sequence embeddings from protein sequences. These embeddings are processed through a Transformer Encoder for the protein sequence and a Transformer Decoder for the atom sequence.

**HyperAttentionDTI**[31] is one of the successful CNN-based methods that encodes compound SMILES and protein sequences with 1D-CNNs and learn CPIs using attention blocks.

**PerceiverCPI**[32] is a state of the art model based on the nested cross-attention network that combines the representation of 2D graphs, molecule fingerprints, and protein sequences.

**AttentionDTA**[27] is also based on 1D-CNNs and attention mechanisms for learning the representations of compound SMILES and protein sequences, but it was implemented for predicting binding affinity and has state of the art performance.

**Autodock Vina**[33] is a software for molecular docking. We utilized the Vina scoring function for rescoring the affinity of the crystal structures from the PDBbind dataset.

For the models implemented only for classification or regression tasks, we additionally implemented the training scripts with binary cross-entropy and mean squared error as the loss function for the classification and regression tasks, respectively.

## 5 Results and Discussion

### 5.1 Performance evaluation: Classification

We evaluated the ChemGLaM model's performance on the BindingDB dataset in zero-shot predictions, comparing it with other models (Table 2). The evaluation metrics included Area Under the Receiver Operating Characteristic Curve (AUROC), Area Under the Precision-Recall Curve (AUPRC), Accuracy, and F1 score. ChemGLaM demonstrated outstanding performance across all metrics. It achieved the highest AUROC score of 0.828 (±0.018), AUPRC of 0.827 (±0.049), and Accuracy of 0.755 (±0.019) significantly surpassing other models in the evaluation. This high performance indicates the effectiveness of large-scale pre-training in enhancing the generalization capabilities of ChemGLaM in predicting truly unknown CPIs. The model also excelled in F1 score of 0.756 (±0.036), suggesting a high success rate of positive classifications, further underscoring its overall efficacy in the dataset. Similarly, ChemGLaM outperforms these baseline models on the Davis dataset (Table 3), confirming the versatility of the ChemGLaM.

Table 2: Performance evaluation for the BindingDB dataset.

| Model | AUROC (Std) | AUPRC (Std) | Accuracy (Std) | F1 (Std) |
|---|---|---|---|---|
| GraphDTA | 0.689(0.063) | 0.715 (0.107) | 0.572 (0.084) | 0.649 (0.073) |
| TransformerCPI | 0.682(0.037) | 0.718 (0.060) | 0.627 (0.053) | 0.586 (0.032) |
| HyperAttentionDTI | 0.792(0.041) | 0.803 (0.055) | 0.695 (0.060) | 0.663 (0.092) |
| PerceiverCPI | 0.531(0.120) | 0.578(0.142) | 0.565 (0.098) | 0.407 (0.086) |
| AttentionDTA | 0.779(0.014) | 0.788 (0.063) | 0.710 (0.025) | 0.708 (0.033) |
| ChemGLaM | **0.828 (0.018)** | **0.827 (0.049)** | **0.755 (0.019)** | **0.756 (0.036)** |

Std: Standard deviation

Table 3: Performance evaluation for the Davis dataset.

| Model | AUROC (Std) | AUPRC (Std) | Accuracy (Std) | F1 (std) |
|---|---|---|---|---|
| GraphDTA | 0.661(0.072) | 0.470(0.131) | 0.694 (0.049) | 0.382 (0.150) |
| TransformerCPI | 0.664(0.072) | 0.460 (0.165) | 0.705 (0.104) | 0.426 (0.174) |
| HyperAttentionDTI | 0.671(0.045) | 0.499 (0.098) | 0.614 (0.067) | 0.486 (0.084) |
| PerceiverCPI | 0.671(0.074) | 0.464 (0.111) | 0.707 (0.029) | 0.410 (0.141) |
| AttentionDTA | 0.679(0.039) | 0.511 (0.076) | 0.596 (0.072) | 0.484 (0.075) |
| ChemGLaM | **0.737 (0.054)** | **0.555 (0.133)** | **0.730 (0.039)** | **0.501 (0.153)** |

Std: Standard deviation

### 5.2 Performance evaluation: Regression

We also evaluated the performance of binding affinity prediction for the PDBbind dataset, focusing on their performance in terms of Root Mean Square Error (RMSE), Pearson correlation coefficient, Spearman correlation coefficient, and Mean Absolute Error (MAE). ChemGLaM achieved the lowest RMSE of 1.530 (±0.069), the highest Pearson correlation of 0.598 (±0.039), and the highest Spearman correlation of 0.598 (±0.042), along with the lowest MAE of 1.201 (±0.045) (Table 4). Conventional chemical genomics-based methods such as GraphDTA and TransformerCPI displayed inferior generalization performance compared with the Vina scoring function. This indicates the limitations of supervised methods in the context of truly unseen CPIs. In contrast, our ChemGLaM is capable of predicting binding affinities more accurately even in truly unknown pairs, highlighting the generalizability of ChemGLaM in affinity prediction. We also

evaluated the performance of the binding affinity prediction for the Metz dataset and confirmed that our ChemGLaM outperforms the baseline models in RMSE (0.851 (±0.043)), Pearson correlation coefficient (0.549 (±0.035)), and Spearman correlation coefficient (0.491 (±0.044)).

Table 4: Performance evaluation for the PDBbind dataset.

| Model | RMSE (Std) | Pearson (std) | Spearman (Std) | MAE (Std) |
|---|---|---|---|---|
| Vina | – | – | 0.511(0.037) | – |
| GraphDTA | 1.755(0.137) | 0.444(0.043) | 0.449(0.037) | 1.386(0.095) |
| TransformerCPI | 1.733(0.050) | 0.474(0.032) | 0.468(0.024) | 1.369(0.041) |
| HyperAttentionDTI | 1.580(0.039) | 0.542 (0.026) | 0.539 (0.021) | 1.237 (0.035 |
| PerceiverCPI | 1.687(0.054) | 0.462 (0.047) | 0.450 (0.040) | 1.338 (0.039) |
| AttentionDTA | 1.615(0.075) | 0.523 (0.047) | 0.509 (0.044) | 1.283 (0.057) |
| ChemGLaM | **1.530 (0.069)** | **0.598 (0.039)** | **0.598 (0.042)** | **1.201 (0.045)** |

Std: Standard deviation

Table 5: Performance evaluation for the Metz dataset.

| Model | RMSE (Std) | Pearson (Std) | Spearman (Std) | MAE (Std) |
|---|---|---|---|---|
| GraphDTA | 1.314(0.108) | 0.363(0.054) | 0.330(0.056) | 1.034(0.111) |
| TransformerCPI | 0.879(0.045) | 0.442(0.048) | 0.400 (0.052) | 0.705 (0.035) |
| HyperAttentionDTI | 0.907(0.082) | 0.419 (0.052) | 0.368 (0.046) | 0.693 (0.048) |
| PerceiverCPI | 0.913(0.029) | 0.399(0.064) | 0.399(0.064) | 0.704(0.024) |
| AttentionDTA | 0.858(0.046) | 0.482 (0.023) | 0.419 (0.033) | **0.655 (0.033)** |
| ChemGLaM | **0.851 (0.043)** | **0.549 (0.035)** | **0.491 (0.044)** | 0.689 (0.040) |

Std: Standard deviation

### 5.3 Relation between the performance of protein language models and their CPI prediction

We investigated the relationship between the efficacy of protein language models (PLMs) and their ability to predict compound-protein interactions (CPIs). We drew upon the performance metrics of PLMs from the study of Lin et al. [17], and trained a ChemGLaM model using a Protein Encoder based on these PLMs, evaluating CPI prediction performance on the PDBbind dataset. Our analysis shows that PLMs with superior performance, such as lower validation perplexity and higher scores in key benchmarks (LR P@L, CASP14[34], and CAMEO[35]), tend to yield more accurate CPI predictions (Table 6). This is evidenced by improved metrics like lower root mean square error (RMSE) and higher Pearson correlation coefficients in our CPI results. In particular, the ESM-2 (3B) model, which exhibited the best performance in the study by Lin et al.[17], also showed the best CPI prediction results on our PDBbind dataset, marked by the lowest RMSE and the highest Pearson coefficient. This observation suggests that a deep understanding of protein properties by advanced PLMs is essential for enhancing the performance of CPI predictions. The same would be true for the chemical language models.

Table 6: Relation between the performance of protein language models and that of CPI prediction for PDBbind dataset.

| | Performance of protein language models [17] | | | | PDBbind | |
|---|---|---|---|---|---|---|
| Protein Language Model | Validation Perplexity | LR P@L | CASP14 | CAMEO | RMSE (Std) | Pearson (Std) |
| ESM-2 (150M) | 7.75 | 0.44 | 0.49 | 0.65 | 1.560(0.047) | 0.587(0.012) |
| ESM-2 (650M) | 6.95 | 0.48 | 0.51 | 0.70 | 1.577(0.043) | 0.578(0.016) |
| Prot-T5-XL-UR50 (3B) | – | 0.52 | 0.50 | 0.69 | 1.577(0.051) | 0.588(0.126) |
| ESM-2 (3B) | **6.49** | **0.54** | **0.52** | **0.72** | **1.530 (0.069)** | **0.598 (0.039)** |

### 5.4 Visualizing interactions with attention maps

The advantage of our ChemGLaM model lies in its ability to investigate relationships between compound SMILES and protein sequences through the attention mechanism. This suggests that the visualizations of attention weights

could offer insights into the mechanisms of CPIs. To visualize CPIs using attention weights, we reconstructed the ChemGLaM model for the PDBbind dataset, employing a random split for test data. We extracted attention weights from a representative example, glutathione transporter (GshT) in complex with glutathione disulfide (GSSG), the experimental structure of which (PDB entry=4C0R) demonstrates that GSSG binds to the hydrophobic pocket formed between N- and C-terminal lobes in GshT, with the complex stabilized by multiple hydrogen bonds, hydrophobic, and van der Waals interactions [36]. Intriguingly, mapping the pooled scores onto the chemical structure of GSSG and the tertiary structure of GshT revealed that the highlighted regions were spatially adjacent and agreed well with the observed interaction pattern (Figure 3A, B, and C). Notably, the highlighted carbonyl groups in GSSH and Thr12 and Thr119 in GshT play a pivotal role in the complex formation by making hydrogen bonding interactions in the experimental structure (Figure 3C). Together with them, several residues consisting of the hydrophobic pocket, including Val115, were highlighted in the visualization. These observations suggest the possibility that ChemGLaM learned the protein folding and the binding event in the three-dimensional space solely from compound SMILES and protein sequences. Similar observations highlighting the binding pocket in ChemGLaM's attention visualizations were obtained for other proteins (Figure 3D, PDB entries=5YQW, 4F8J, and 4WMU), emphasizing the effectiveness of learning CPIs using attention blocks for providing insights into interaction mechanisms.

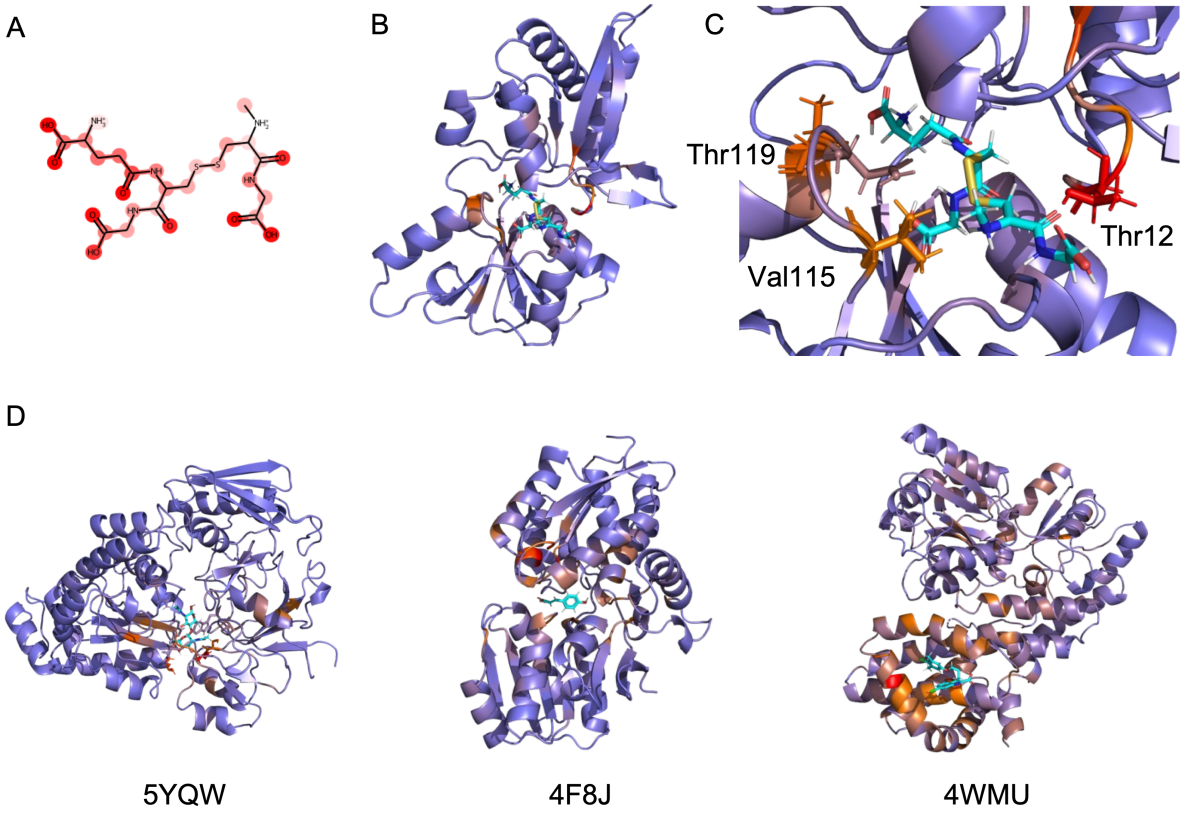


Figure 2: Visualization of interactions on 4C0R with attention map. Atoms of glutathione disulfide (GSSG) (A) and residues of glutathione transporter (GshT) (B) with high attention scores are highlighted with red. (C) an enlarged view of (B). (D) Attention visualization for 5YQW, 4F8J, and 4WMU.

## 6 Conclusion

This research presents ChemGLaM, a novel chemical genomics language model for predicting CPIs. By leveraging the capabilities of pretrained chemical (MoLFormer) and protein (ESM-2) language models, ChemGLaM integrates the representations of compounds and proteins through a multi-head cross-attention mechanism. Our method demonstrates significant effectiveness over other DL-based approaches in CPI prediction. In our evaluations, ChemGLaM exhibited superior performance across various datasets, including BindingDB, Davis, PDBbind, and Metz. This success is attributed to the model's ability to leverage large-scale molecule and protein datasets for individual pre-training, thereby enhancing their efficacy in tasks involving multi-modality such as CPI prediction. Furthermore, the use of visualized

cross-attention maps in our model provides explainable insights into compound-protein interactions. The results from ChemGLaM pave the way for further exploration and refinement in this domain. Exploring more sophisticated architectures and training techniques could further enhance the model's generalizability and explainability. Additionally, the advancement of foundation models related to compounds and proteins will elevate CPI prediction capabilities. ChemGLaM has a substantial potential to offer a scalable and efficient identification of novel compound-protein interactions.

## 7 Use of AI Software

ChatGPT (GPT-4) was used to improve the wording of some paragraphs, but not to generate new content.

## 8 Acknowledgments

This research was supported by JSPS KAKENHI grant number 22K06112 to S.M.

## References

[1] Michael J Keiser, Vincent Setola, John J Irwin, Christian Laggner, Atheir I Abbas, Sandra J Hufeisen, Niels H Jensen, Michael B Kuijer, Roberto C Matos, Thuy B Tran, et al. Predicting new molecular targets for known drugs. *Nature*, 462(7270):175–181, 2009.

[2] Ricardo Macarron, Martyn N Banks, Dejan Bojanic, David J Burns, Dragan A Cirovic, Tina Garyantes, Darren VS Green, Robert P Hertzberg, William P Janzen, Jeff W Paslay, et al. Impact of high-throughput screening in biomedical research. *Nature reviews Drug discovery*, 10(3):188–195, 2011.

[3] Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design*, 30:595–608, 2016.

[4] Thin Nguyen, Hang Le, Thomas P Quinn, Tri Nguyen, Thuc Duy Le, and Svetha Venkatesh. GraphDTA: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, 37(8):1140–1147, 10 2020.

[5] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[6] Hakime Öztürk, Arzucan Özgür, and Elif Ozkirimli. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics*, 34(17):i821–i829, 09 2018.

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

[8] Lifan Chen, Xiaoqin Tan, Dingyan Wang, Feisheng Zhong, Xiaohong Liu, Tianbiao Yang, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, and Mingyue Zheng. TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics*, 36(16):4406–4414, 05 2020.

[9] Mindy I. Davis, Jeremy P. Hunt, Sanna Herrgard, Pietro Ciceri, Lisa M. Wodicka, Gabriel Pallares, Michael Hocker, Daniel K. Treiber, and Patrick P. Zarrinkar. Comprehensive analysis of kinase inhibitor selectivity. *Nature Biotechnology*, 29(11):1046–1051, Nov 2011.

[10] Zhihai Liu, Yan Li, Li Han, Jie Li, Jie Liu, Zhixiong Zhao, Wei Nie, Yuchen Liu, and Renxiao Wang. PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics*, 31(3):405–412, 10 2014.

[11] Michael K. Gilson, Tiqing Liu, Michael Baitaluk, George Nicola, Linda Hwang, and Jenny Chong. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Research*, 44(D1):D1045–D1053, 10 2015.

[12] Mostafa Karimi, Di Wu, Zhangyang Wang, and Yang Shen. DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics*, 35(18):3329–3338, 02 2019.

[13] Jun Xia, Yanqiao Zhu, Yuanqi Du, and Stan Z. Li. A systematic survey of chemical pre-trained models. In Edith Elkind, editor, *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pages 6787–6795. International Joint Conferences on Artificial Intelligence Organization, 8 2023. Survey Track.

[14] Bozhen Hu, Jun Xia, Jiangbin Zheng, Cheng Tan, Yufei Huang, Yongjie Xu, and Stan Z. Li. Protein language models and structure prediction: Connection and progression, 2022.

[15] Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. Chemberta: large-scale self-supervised pretraining for molecular property prediction. *arXiv preprint arXiv:2010.09885*, 2020.

[16] Jerret Ross, Brian Belgodere, Vijil Chenthamarakshan, Inkit Padhi, Youssef Mroueh, and Payel Das. Large-scale chemical language representations capture molecular structure and properties. *Nature Machine Intelligence*, 4(12):1256–1264, 2022.

[17] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

[18] Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rehawi, Yu Wang, Llion Jones, Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, Debsindhu Bhowmik, and Burkhard Rost. Prottrans: Toward understanding the language of life through self-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7112–7127, 2022.

[19] James T Metz, Eric F Johnson, Niru B Soni, Philip J Merta, Lemma Kifle, and Philip J Hajduk. Navigating the kinome. *Nature chemical biology*, 7(4):200–202, 2011.

[20] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.

[21] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, and Evan E Bolton. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Research*, 47(D1):D1102–D1109, 10 2018.

[22] John J Irwin and Brian K Shoichet. Zinc- a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, 45(1):177–182, 2005.

[23] Baris E. Suzek, Yuqi Wang, Hongzhan Huang, Peter B. McGarvey, Cathy H. Wu, and the UniProt Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, 31(6):926–932, 11 2014.

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

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

[26] Kyle Yingkai Gao, Achille Fokoue, Heng Luo, Arun Iyengar, Sanjoy Dey, and Ping Zhang. Interpretable drug target prediction using deep neural representation. In *IJCAI*, volume 2018, pages 3371–3377.

[27] Qichang Zhao, Guihua Duan, Mengyun Yang, Zhongjian Cheng, Yaohang Li, and Jianxin Wang. Attentiondta: Drug–target binding affinity prediction by sequence-based deep learning with attention mechanism. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 20(2):852–863, 2023.

[28] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[29] Zhiyuan Li and Sanjeev Arora. An exponential learning rate schedule for deep learning. *arXiv preprint arXiv:1910.07454*, 2019.

[30] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2018.

[31] Qichang Zhao, Haochen Zhao, Kai Zheng, and Jianxin Wang. HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics*, 38(3):655–662, 10 2021.

[32] Ngoc-Quang Nguyen, Gwanghoon Jang, Hajung Kim, and Jaewoo Kang. Perceiver CPI: a nested cross-attention network for compound–protein interaction prediction. *Bioinformatics*, 39(1):btac731, 11 2022.

[33] Oleg Trott and Arthur J Olson. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.

[34] Andriy Kryshtafovych, Torsten Schwede, Maya Topf, Krzysztof Fidelis, and John Moult. Critical assessment of methods of protein structure prediction (casp)—round xiv. *Proteins: Structure, Function, and Bioinformatics*, 89(12):1607–1617, 2021.

[35] Jürgen Haas, Alessandro Barbato, Dario Behringer, Gabriel Studer, Steven Roth, Martino Bertoni, Khaled Mostaguir, Rafal Gumienny, and Torsten Schwede. Continuous automated model evaluation (cameo) complementing the critical assessment of structure prediction in casp12. *Proteins: Structure, Function, and Bioinformatics*, 86:387–398, 2018.

[36] Bjorn Vergauwen, Kenneth Verstraete, Dilani B Senadheera, Ann Dansercoer, Dennis G Cvitkovitch, Eric Guédon, and Savvas N Savvides. Molecular and structural basis of glutathione import in gram-positive bacteria via gsht and the cystine abc importer tcybc of s treptococcus mutans. *Molecular microbiology*, 89(2):288–303, 2013.