PepPFN: protein-peptide binding residues prediction via pre-trained module-based Fourier Network

Xue Li
College of Computer Science and Technology
China University of Petroleum (East China)
Qingdao, China
Centre for Computational Biology
Duke-NUS Medical School
Singapore
xueleecs@gmail.com

Ben Cao
School of Computer Science and Technology
Dalian University of Technology
Dalian, China
*CFAR, A*STAR*
Singapore
bencaocs@gmail.com

Hongzhen Ding
College of Computer Science and Technology
China University of Petroleum (East China)
Qingdao, China
s22070050@s.upc.edu.cn

Na Kang*
The Ninth Department of Health Care Administration, the Second Medical Center
Chinese PLA General Hospital
Beijing, China
331839312@qq.com

Tao Song*
College of Computer Science and Technology
China University of Petroleum (East China)
Qingdao, China
tsong@upc.edu.cn

***Abstract*—Molecular activities in vivo, such as protein-peptide binding residues, contribute to understanding disease mechanisms and discovering targeted drugs, but are often limited by the cost of wet experiments. Therefore, this paper proposes a Pre-trained module-based Fourier Network inspired by Fourier Neural Operators for predicting protein-peptide binding residues (PepPFN). First, the model captures high-latent protein representation and implicit information through Bert and ESM, respectively. Subsequently, the contrastive learning module is used to optimize the representation of binding residues to mitigate the effects of data diversity. Furthermore, the order and frequency information of the residual is obtained by strengthening the model with Fourier network. Comprehensive analysis of multiple tasks and protein visualization results demonstrates that PepPFN outperforms state-of-the-art methods on multiple performance metrics. Datasets and the source code is available at https://github.com/xueleecs/PepPFN.git.**

Keywords—protein-peptide binding residues, Pre-trained module, Fourier network, deep learning

I. INTRODUCTION

Proteins, essential molecular entities within living organisms, intricately engage in a myriad of biological processes [1]. Peptides are the building blocks of proteins[2] and protein-peptide interactions[3] is one of the most important interactions [4], and the identification of protein-peptide binding residues is necessary to understand the mechanism of protein function and discover new drugs[5]. However, owing to the small size [6], weak affinity [7] and strong flexibility [8] of peptides, determining the structure of protein-peptide complexes becomes challenging. This complexity not only complicates the identification of protein-peptide binding residues through experimental methods but also escalates the time and cost of experiments [9]. Therefore, it is urgent and important to develop an effective method for predicting protein-peptide binding residues.

Currently, computational methods for predicting protein-peptide binding residues in prediction tasks are categorized into two groups, structure-based and sequence-based approaches. Structure-based methods like PepSite [10], SPRINT-Str [11], and Peptimap [12] typically characterize proteins based on structural properties such as Accessible Surface Area (ASA) and Secondary Structure (SS). On the other hand, sequence-based methods encompass SPRINT-Seq [13], PepBind [14], Visual [15], PepNN-Seq [16], and PepBCL [17]. Most of these methods tend to utilize amino acid composition information, physicochemical properties, and evolutionary information to describe proteins[18], and often perform feature extraction in a single dimension, which limits the accuracy of prediction. This is related to the characterization of proteins and the feature extraction network of models, as the singularity of protein embedding methods makes it difficult to fully reflect protein characteristics. Finally, the scarcity of positive samples in imbalanced datasets imposes stricter requirements on model performance.

To address these issues, we here propose a pre-trained module-based Fourier Network, called PepPFN, which makes full use of Fourier layers to extract multiple dimensions of information. Specially, this model utilizes pre-trained protein language models (Bert and ESM) to encode residues into high-latent representations, encompassing various aspects such as protein structure. Subsequently, the introduction of a contrastive learning module enhances adaptive learning of binding residue representations, ensuring high-quality representations while making optimal use of available samples. Following that, the Fourier network is employed to

* Corresponding authors: Na Kang; Tao Song
This work was supported by National Key Research and Development Project of China (2021YFA1000103，2021YFA1000102), National Natural Science Foundation of China (Grant Nos. 61873280, 61972416, 62272479, 62202498), Taishan Scholarship (tsqn201812029), Foundation of Science and Technology Development of Jinan (201907116), Shandong Provincial Natural Science Foundation(ZR2021QF023), Fundamental Research Funds for the Central Universities (21CX06018A), Spanish project PID2019-106960GB-I00, Juan de la Cierva IJC2018-038539-I.

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00195

extract time-domain and spectral-domain features, capturing order and frequency information of residues. Finally, based on experimental results and protein visualization assessments, PepPFN demonstrates its performance advantage on the datasets.

II. Materials and Methods

A. Datasets

To evaluate the performance of PepPFN in protein-peptide binding residues prediction, we selected 2 benchmark datasets, which are the same as the other computational methods, namely Dataset 1 and Dataset 2 (TABLE I) [17]. For a fair comparison, we split Dataset 1 and Dataset 2 following the general method for model training and testing, respectively. In addition, considering the similarity of protein-protein binding sites to this task, it was also validated on this task (dataset 3) [19].

TABLE I. Summary of Datasets

Datasets	Dataset1		Dataset2		Dataset3	
	TR1154[a]	*TE125*[a]	*TR640*[a]	*TE639*[a]	*TR843*[a]	*TE164*[a]
No[b]. proteins	1154	125	640	639	843	164
No. residues	276 822	30 870	157 362	150 330	225 302	33 675
No. binding residues	15030	1719	8259	8490	32253	6096
No. non-binding residues	261 792	29 151	149 103	141 840	193 049	27 585

[a](TR: training set; TE: testing set)
[b](No.: number of)

B. Architecture of the proposed method

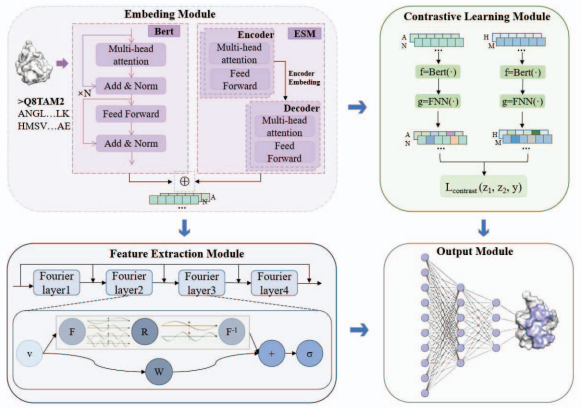

Fig. 1. Flowchart of the proposed PepPFN.

The model framework of PepPFN (Fig. 1) consists of four modules: the embedding module, feature extraction module, contrastive learning module, and output module. In the embedding module, the protein sequence employs Bert and ESM to obtain a high-latent representation vector, and then the obtained high-latent representation vector is input into the Fourier layer for feature extraction. At the same time, the initial residual connection is added to reduce information loss. Next, in the contrastive learning module, the model is optimized by calculating the contrastive loss between two training samples, so as to make the discrimination more accurate. Finally, the output module applies the fully connected layer to calculate whether each residue has peptide binding ability, so as to complete the prediction task. More details of the four modules are described below.

C. Embedding module

The Embedding module employs two pre-trained protein models, Bert and ESM, to capture structural and sequence information from residues, embedding biological sequences into high-latent vectors to enhance interpretability and richness [20]. The pre-trained model is trained based on a large number of protein sequences, and the original protein is first divided into multiple sequences according to a fixed maximum length. The input embeddings are then the sum of token embeddings and position embeddings with learnable weights. To explore token relationships, a percentage of input tokens are randomly masked, and the model is continuously trained to strengthen predictions for these masked tokens, accomplishing both masked language modeling and next sentence prediction tasks. Specifically, the pre-trained BERT model, ProtBert-Big Fantastic Database (BFD) [21,22], is trained on general domain Bert, utilizing a large amount of unlabeled protein data from BFD, encompassing 2.1 billion protein sequences for pre-training. ESM [23] is pre-trained on 250 million protein data, using distributed context vectors as protein representations. The resulting representation space demonstrates a multiscale organization, capturing long-range homology features extending from amino acid biochemical properties to proteins. Moreover, information about secondary and tertiary structure is embedded in the representation and can be identified through linear projection.

Since pre-trained models can effectively obtain interpretable data representations, this paper introduces Bert and ESM into PepPFN. The PepPFN performs fine-tuning on small datasets through pre-trained models to complete the adaptation of downstream tasks. Specifically, the PepPFN divides the protein sequence in the protein-peptide binding residues task into a fixed length, and then inputs it into a pre-trained model with parameters. After adjusting the parameters, it outputs an interpretable fusion representation at the biological word level.

D. Feature extraction module

The feature extraction layer mainly utilizes the Fourier layer, which is inspired by the Fourier neural operator [24]. The Fourier neural operator can break through the traditional neural network to learn the mapping between finite dimensional spaces, and can take care of learning the mapping relationship between function spaces. Operator learning can be viewed as an image-to-image problem, and Fourier layers can be viewed as an alternative to convolutional layers. The potential high-latent vector of each sequence obtained by the Embedding module can be regarded as a tractable vector, and the global information is obtained by virtue of the property of invariant resolution of the Fourier layer, rather than the local information limited to the CNN network. Specifically, as shown in Feature Extraction Module in Fig. 1, the process starts with the input vector v. On one hand, Fourier transformation F is applied to v. First, the Fourier transform F is applied to process the input vector. This transformation will decompose the input signal into components with different frequencies. Next, a linear transformation R is applied to the lower Fourier patterns and the higher patterns are filtered out, which helps to retain important information at lower frequencies. Then, the inverse Fourier transform F^{-1} is applied to restore the processed signal to a time-domain representation. On the other hand, a local linear

transformation *W* is applied in order to further extract features. Finally, the activation function is used to obtain a high-dimensional representation of the residues. The application of the Fourier network in the feature extraction module transforms the order of residues into the time domain, while the frequency is mapped into the spectral domain. The Fourier transform (Eq 1) and the inverse Fourier transform (Eq 2) can be converted to each other under appropriate conditions.

$$\hat{f}(\varphi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \varphi} dx \tag{1}$$

Where φ is any real number and the domain is the frequency domain, which we consider here to be the frequency of residues.

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\varphi) e^{-2\pi i \varphi x} d\varphi \tag{2}$$

Where *x* is any real number and the domain is the time domain, which we consider here to be the order of residues.

E. Contrastive learning module

Supported by data augmentation techniques, the acquisition of a large number of negative samples has become feasible, sparking widespread applications of contrastive learning in various domains [25]. In order to strengthen the constraint on negative samples, a contrastive learning module based on supervised learning is applied in this work, which can make the representations of the same class closer in the spatial mapping, while the representations of different classes are further away. The advantages of contrastive learning in this task have been demonstrated [17]. Then, a contrast loss $L_{contrast}$ is constructed to calculate the loss of imbalanced datasets to distinguish samples from the same class and different classes. For different sequence lengths, contrastive learning of residues is carried out by dividing the sequence into two parts in turn, and each residue can be more discriminative by minimizing the loss. For a pair of residues in a batch representing r_1, r_2, the loss function is defined as follows.

$$D(r_1, r_2) = 1 - cosine < r_1, r_2 > \tag{3}$$

$$L_{contrast}(r_1, r_2, y) = \frac{(1-y)D(r_1,r_2)^2}{2} + \frac{y\{D_{max} - \mathrm{D}(r_1,r_2)\}^3}{2} \tag{4}$$

And where y is calculated using the formula (5).

$$y = cat(\left(l_1(i) \oplus l_2(i)\right) for\ i\ in \left[0, \frac{len(r_1)}{2} - 1\right]) \tag{5}$$

Here, $D(r_1, r_2)$ represents the distance between two residues (r_1, r_2), and D_{Max} is obtained when two residues belong to different classes, which is equal to 2. In order to focus on minority categories, pairs of different categories are given a higher power of 3.

F. Output module

The output module is able to classify the latent vector after the Feature extraction module. The output module is composed of fully connected layers, and the number of nodes is 64, 32, 16, 2 respectively. Assuming that the vector output through the Fourier layers is *z*, and the calculation formula is as follows. At the same time, in order to improve the performance of the model, the loss function designed in this paper includes cross-entropy loss in addition to contrastive learning loss.

$$\hat{y} = Linear(relu(z)) \tag{6}$$

$$L_{loss} = L_{CE} + L_{contrast} \tag{7}$$

G. Evaluation metrics

In addressing the significant imbalance between positive and negative samples in the dataset, five indicators commonly used in imbalanced classification tasks are selected, which are Sensitivity, Specificity, Precision, Matthews Correlation Coefficient (MCC), and AUC. The formulas are as follows. Where TP (True Positive) is the correct prediction of binding residues in the sample, TN (True Negative) is the correct prediction of non-binding residues, FP (False Positive) is the prediction of non-binding residues in the sample as binding residues, and TP (true positive) is the prediction of binding residues. False Negative (FN) is the prediction of binding residues as non-binding residues. The larger the value of the above evaluation index, the better the performance of the method.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{8}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{9}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{10}$$

$$\text{MCC} = \frac{\text{TP}\times\text{TN} - \text{FP}\times\text{FN}}{\sqrt{(\text{TP+FP})(\text{TP+FN})(\text{TN+FP})(\text{TN+FN})}} \tag{11}$$

III. RESULT

A. Comparison with existing methods

To evaluate the performance of PepPFN, it is compared with nine existing methods, including PepSite [10], Peptimap [12], SPRINT-Seq [13], SPRINT-Str [11], PepBind [14], Visual [15], PepNN-Seq [26], PepBCL [17], SPPPred [27]. Out of the nine methods evaluated, six (SPRINT-Seq, Visual, PepNN-Seq, PepBCL, SPPPred) are sequence-based models, while the remaining three (PepSite, Peptimap, and SPRINT-Str) are structure-based. We conduct comparative experiments on two benchmark datasets TE125 and TE639, respectively.

TABLE II. PERFORMANCES OF THE PROPOSED PEPPFN MODEL AND THE PREVIOUS METHODS ON THE TE125 TEST SET

Methods	Sensitivity	Specificity	Precision	MCC	AUC
PepSite	0.180	0.970	-	0.200	0.610
Peptimap	0.320	0.950	-	0.270	0.630
SPRINT-Seq	0.210	0.960	-	0.200	0.680
SPRINT-Str	0.240	0.980	-	0.290	0.780
PepBind	0.344	-	0.469	0.372	0.793
Visual	**0.670**	0.680	-	0.170	0.730
PepNN-Seq	-	-	-	0.278	0.805
PepBCL	0.315	0.984	0.540	**0.385**	**0.815**
SPPPred	0.315	0.959	-	0.230	0.710
PepPFN	0.195	**0.992**	**0.600**	0.322	0.813

The comparison results are shown in TABLE II and III, respectively. At the same time, considering the similarity between protein-protein binding sites and this task, the performance evaluation is also carried out on the TE164 dataset, and the results are shown in Table IV. In the tables, the boldface is the highest value of the evaluation metric in the table. Since the source code of most methods is not directly

available, in order to ensure the fairness of the results, the results of the methods compared on the three datasets are directly from their research.

Table II shows the evaluation results of PepPFN in TE125. PepPFN achieves 0.992 specificity, 0.600 precision, and 0.813 AUC. Compared with other methods, PepPFN shows advantages in specificity, and precision. We observe similar results on the TE639 test set to the TE125 test set, with improved specificity, precision, and AUC compared to previous methods. The results of TE639 test set are shown in TABLE III, and the sensitivity, specificity, precision, MCC and AUC values are 0.127, 0.996, 0.680, 0.307 and 0.813, respectively. Although the PepPFN method does not perform the best in all metrics in both test sets, the overall performance is better than the other methods, and these advantages prove the good performance of the PepPFN model. This can be attributed to the embedding method being pretrained on a comprehensive set of protein sequences, covering information such as protein structure and function. Furthermore, the inclusion of the contrast module enhances the model's ability to address imbalanced data.

TABLE III. PERFORMANCES OF THE PROPOSED PEPPFN MODEL AND THE PREVIOUS METHODS ON THE TE639 TEST SET

Methods	Sensitivity	Specificity	Precision	MCC	AUC
PepBind	**0.317**	-	0.450	**0.348**	0.767
PepNN-Seq	-	-	-	0.251	0.792
PepBCL	0.252	0.983	0.470	0.312	0.804
PepPFN	0.127	**0.996**	**0.680**	0.307	**0.813**

To further illustrate the robustness of the PepPFN model and its performance advantage on imbalanced data, we added the task of protein-protein binding sites prediction. TABLE IV shows the prediction results of PepPFN in protein-protein binding sites. Compared with other methods [28], PepPFN achieves the Precision of 0.383 and AUC of 0.689, which reaches the maximum value. The value of MCC is 0.193, which is the second highest value. Since the comparison method did not calculate Sensitivity and Specificity, we only compared the same three indicators of Precision, MCC and AUC.

TABLE IV. PERFORMANCES OF THE PROPOSED PEPPFN MODEL AND THE PREVIOUS METHODS ON THE TE164 TEST SET

Methods	Precision	MCC	AUC
SPPIDER	0.253	0.090	0.528
PSIVER	0.216	0.043	0.554
CRFPPI	0.280	0.121	0.608
SSWRF	0.266	0.103	0.606
SCRIBER	0.327	0.179	0.657
DLPred	0.338	0.192	0.672
DELPHI	0.352	**0.209**	0.685
PepPFN	**0.383**	0.193	**0.689**

B. *Ablation experiments of the model*

To investigate the effectiveness of each component in the PepPFN model, a series of ablation experiments are conducted based on TE125 (TABLE V). Due to the leading advantage of the pre-trained model in the field of protein representation, this part conducts the ablation experiment based on the pre-trained model. The methods for ablation experiments are Bert and methods with contrastive learning methods, including Bert (Bert+CL), ESM (ESM+CL), Bert combined with ESM (Bert+ESM+CL), and models with different Fourier layers added (Bert+ESM+CL+FNX, X refers to 3, 4, and 5). The two methods of Bert and Bert+CL are carried out under the same hyperparameters, and the values in the table show that the effect of Bert without contrastive learning is not significant. This suggests that the loss has a substantial impact on the model, further confirming the effectiveness of the loss function. The results of Bert+CL, ESM+CL, and Bert+ESM+CL demonstrate the necessity of Bert and ESM, affirming the complementary of the two pre-training models with high-latent features. The results of exploring the influence of different Fourier layers on the model (Bert+ESM+FNX) show that the PepPFN (Bert+ESM+FN4) model achieves the best performance.

TABLE V. PERFORMANCES OF THE DIFFERENT COMPONENT IN PEPPFN

Methods	Sensitivity	Specificity	Precision	MCC	AUC
Bert	0.176	0.877	0.078	0.037	0.550
Bert+CL	0.606	0.776	0.137	0.230	0.764
ESM+CL	**0.639**	0.799	0.158	0.240	0.785
Bert+ESM +CL	0.240	0.980	0.417	0.287	0.766
Bert+ESM +CL+FN3	0.327	0.977	0.458	**0.357**	0.811
PepPFN	0.200	**0.992**	**0.600**	0.322	**0.813**
Bert+ESM +CL+FN5	0.204	0.990	0.550	0.313	0.804

C. *Case study*

In order to visually show the ability of the model to predict protein-peptide binding residues, we randomly selected three protein sequences (pdbID: 1dpuA, 1jmqA and 1kugA) from TE125 for prediction. The visualized predictions are shown in Fig. 2, where gray means non-binding residues and blue means binding residues.

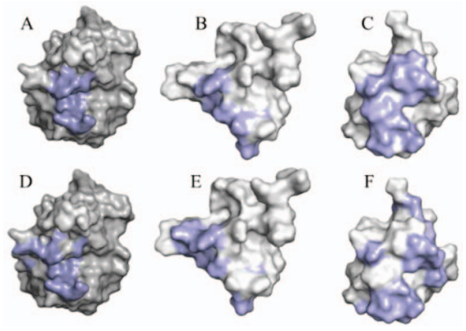

Fig. 2. 3D structure visualization of three proteins (pdbID: 1dpuA, 1jmqA and 1kugA)

Fig. 3. Comparing the experimental data of residues with the prediction of PepPFN (with pdbID 1dpuA, 1jmqA and 1kugA)

Fig. 2A-C shows the actual binding residue obtained from biological experiments and Fig. 2D-F shows the residues predicted by PepPFN. We also present from the unfolded sequence(Fig. 3). The orange residues indicate the binding residues obtained in the experiments, while the green residues indicate the binding residues predicted by PepPFN. As can be seen from the protein visualization results, the binding residues predicted by the proposed method are very similar to the actual binding residues. In particular, compared with discrete regions, the model has more advantages in continuous regions of binding residues, which is also more consistent with the law that protein binding regions tend to be continuous and concentrated.

IV. Conclusion

In this study, we construct a pre-trained module-based Fourier Network, PepPFN, which can effectively predict binding and non-binding residues in protein sequences. The model employs two high-performing pre-training models for protein language to capture the essence of proteins, producing interpretable vectors with high-latent representation vectors of structural, functional, and biological characteristics. Additionally, it is demonstrated that different pre-training models emphasize distinct aspects and the Bert and ESM models with embedded modules can complement high-latent features. The added contrastive learning can alleviate the problem of data imbalance, so that samples from the same class have more similar representations, while samples from different classes have more different representations, making the model discrimination more biased. And the Fourier layer network can complete the multi-dimensional extraction of residue features to reduce the loss of information.

The reliability of PepPFN model has been confirmed through visual comparisons of randomly selected proteins. Also same time, in order to show the generality of the model, we also predict the protein-protein binding site task and achieve good results. Compared with other methods, although PepPFN is not the best in all indicators, the comprehensive evaluation performance is still in a better position. This may be attributed to the fact that the pre-trained model is built on the protein language models, and the data used by the pre-trained model is more biased towards protein sequences, which does not fully reflect the advantage of short peptides. At the same time, the prediction of protein-peptide binding residues involves an imbalanced sample problem, and a large number of negative samples can easily lead to false positive prediction results. In the future, employing fine-tuning methods can align the generated representations with downstream tasks more effectively, thereby improving the accuracy of those tasks.

Acknowledgment

We thank our partners who provided all the help during the research process and the team for their great support.

References

[1] J. Wang, Y. Chu, J. Mao, H.-N. Jeon, H. Jin, A. Zeb, Y. Jang, K.-H. Cho, T. Song, and K. T. No, "De novo molecular design with deep molecular generative models for PPI inhibitors," Brief. Bioinformatics, vol. 23, no. 4, pp. bbac285, July 2022.

[2] X. Li, P. Han, W. Chen, C. Gao, S. Wang, T. Song, M. Niu, and A. Rodriguez-Patón, "MARPPI: boosting prediction of protein–protein interactions with multi-scale architecture residual network," Brief. Bioinformatics, vol. bbac524, January 2023.

[3] V. Neduva, R. Linding, I. Su-Angrand, A. Stark, F. d. Masi, T. J. Gibson, J. Lewis, L. Serrano, and R. B. Russell, "Systematic discovery of new recognition peptides mediating protein interaction networks," PLoS biol., vol. 3, no. 12, pp. e405, November 2005.

[4] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, and R. D. Schaeffer, "Accurate prediction of protein structures and interactions using a three-track neural network," Science, vol. 373, no. 6557, pp. 871-876, Augest 2021.

[5] A. Chandra, A. Sharma, I. Dehzangi, T. Tsunoda, and A. Sattar, "PepCNN deep learning tool for predicting peptide binding residues in proteins using sequence, structural, and language model features," Sci. Rep., vol. 13, no. 1, pp. 20882, November, 2023.

[6] Y. Lei, S. Li, Z. Liu, F. Wan, T. Tian, S. Li, D. Zhao, and J. Zeng, "A deep-learning framework for multi-level peptide–protein interaction prediction," Nat. Commun., vol. 12, no. 1, pp. 5465, September 2021.

[7] P. Vlieghe, V. Lisowski, J. Martinez, and M. Khrestchatisky, "Synthetic therapeutic peptides: science and market," Drug Discov. Today, vol. 15, no. 1-2, pp. 40-56, 2010.

[8] H. J. Dyson, and P. E. Wright, "Intrinsically unstructured proteins and their functions," Nat. Rev. Mol. Cell Biol., vol. 6, no. 3, pp. 197-208, March 2005.

[9] A. C.-L. Lee, J. L. Harris, K. K. Khanna, and J.-H. Hong, "A comprehensive review on current advances in peptide drug development and design," Int. J. Mol. Sci., vol. 20, no. 10, pp. 2383, May 2019.

[10] E. Petsalaki, A. Stark, E. García-Urdiales, and R. B. Russell, "Accurate prediction of peptide binding sites on protein surfaces," PLoS Comput. Biol., vol. 5, no. 3, pp. e1000335, March 2009.

[11] G. Taherzadeh, Y. Zhou, A. W.-C. Liew, and Y. Yang, "Structure-based prediction of protein–peptide binding regions using Random Forest," Bioinformatics, vol. 34, no. 3, pp. 477-484, February 2018.

[12] A. Lavi, C. H. Ngan, D. Movshovitz - Attias, T. Bohnuud, C. Yueh, D. Beglov, O. Schueler - Furman, and D. Kozakov, "Detection of peptide - binding sites on protein surfaces: The first step toward the modeling and targeting of peptide - mediated interactions," Proteins, vol. 81, no. 12, pp. 2096-2105, September 2013.

[13] G. Taherzadeh, Y. Yang, T. Zhang, A. W. C. Liew, and Y. Zhou, "Sequence-based prediction of protein‐peptide binding sites using support vector machine," J. Comput. Chem., vol. 37, no. 13, pp. 1223-1229, February 2016.

[14] Z. Zhao, Z. Peng, and J. Yang, "Improving sequence-based prediction of protein–peptide binding residues by introducing intrinsic disorder and a consensus method," J. Chem. Inf. Model., vol. 58, no. 7, pp. 1459-1468, June 2018.

[15] W. Wardah, A. Dehzangi, G. Taherzadeh, M. A. Rashid, M. G. Khan, T. Tsunoda, and A. Sharma, "Predicting protein-peptide binding sites with a deep convolutional neural network," J. Theor. Biol., vol. 496, pp. 110278, 2020.

[16] O. Abdin, H. Wen, and P. M. Kim, "Sequence and structure based deep learning models for the identification of peptide binding sites," NeurIPS, vol. 33, February 2020.

[17] R. Wang, J. Jin, Q. Zou, K. Nakai, and L. Wei, "Predicting protein–peptide binding residues via interpretable deep learning," Bioinformatics, vol. 38, no. 13, pp. 3351-3360, July 2022.

[18] P. B. Moore, W. A. Hendrickson, R. Henderson, and A. T. Brunger, "The protein-folding problem: Not yet solved," Science, vol. 375, no. 6580, pp. 507-507, February 2022.

[19] Z. Hou, Y. Yang, Z. Ma, K.-c. Wong, and X. Li, "Learning the protein language of proteome-wide protein-protein binding sites via explainable ensemble deep learning," Commun. Biol., vol. 6, no. 1, pp. 73, January 2023.

[20] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, "Pre-trained models for natural language processing: A survey," Sci. China Technol. Sci., vol. 63, no. 10, pp. 1872-1897, September 2020.

[21] A. Rives, J. Meier, T. Sercu, S. Goyal, Z. Lin, J. Liu, D. Guo, M. Ott, C. L. Zitnick, and J. Ma, "Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences," PNAS, vol. 118, no. 15, pp. e2016239118, December 2021.

[22] M. Steinegger, M. Mirdita, and J. Söding, "Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold," Nat. methods, vol. 16, no. 7, pp. 603-606, June 2019.

[23] W. Yuan, G. Chen, and C. Y.-C. Chen, “FusionDTA: attention-based feature polymerizer and knowledge distillation for drug-target binding affinity prediction,” Brief. Bioinformatics, vol. 23, no. 1, pp. bbab506, January 2022.

[24] Y. Lee, “Autoregressive Renaissance in Neural PDE Solvers,” ICLR Blogposts 2023, 2023.

[25] S. Ge, S. Mishra, S. Kornblith, C.-L. Li, and D. Jacobs, "Hyperbolic contrastive learning for visual representations beyond objects." CVPR, pp. 6840-6849, 2023.

[26] O. Abdin, S. Nim, H. Wen, and P. M. Kim, “PepNN: a deep attention model for the identification of peptide binding sites,” Commun. Biol., vol. 5, no. 1, pp. 503, May 2022.

[27] S. Shafiee, A. Fathi, and G. Taherzadeh, “SPPPred: Sequence-based Protein-Peptide binding residue Prediction using genetic programming and ensemble learning,” IEEE/ACM Trans. Comput. Biol. Bioinform., December 2022.

[28] Y. Li, G. B. Golding, and L. Ilie, “DELPHI: accurate deep ensemble model for protein interaction sites prediction,” Bioinformatics, vol. 37, no. 7, pp. 896-904, March 2021.