Improved compound–protein interaction site and binding affinity prediction using self-supervised protein embeddings

Jialin Wu, Zhe Liu, Xiaofeng Yang[*] and Zhanglin Lin[*]

*Correspondence:
biyangxf@scut.edu.cn;
zhanglinlin@scut.edu.cn

School of Biology and Biological
Engineering, South China
University of Technology, 382
East Outer Loop Road, University
Park, Guangzhou 510006,
Guangdong, China

Abstract

Background: Compound–protein interaction site and binding affinity predictions are crucial for drug discovery and drug design. In recent years, many deep learning-based methods have been proposed for predications related to compound–protein interaction. For protein inputs, how to make use of protein primary sequence and tertiary structure information has impact on prediction results.

Results: In this study, we propose a deep learning model based on a multi-objective neural network, which involves a multi-objective neural network for compound–protein interaction site and binding affinity prediction. We used several kinds of self-supervised protein embeddings to enrich our protein inputs and used convolutional neural networks to extract features from them. Our results demonstrate that our model had improvements in terms of interaction site prediction and affinity prediction compared to previous models. In a case study, our model could better predict binding sites, which also showed its effectiveness.

Conclusion: These results suggest that our model could be a helpful tool for compound–protein related predictions.

Keywords: Compound–protein interaction, Binding affinity, Deep learning, Self-supervised protein embedding

Background

In order to advance drug design, many compound–protein interaction prediction methods have been proposed [1]. As available data and computational methods continue to grow, this field has attracted a significant amount of attention. To date, several deep learning models [2–11] have been incorporated into compound–protein-related tasks. In terms of compound–protein interaction prediction, DeepConV-DTI [6] uses compound fingerprints and protein sequences as inputs, which are then processed by fully connected neural networks and convolutional neural networks, respectively. DeepConV-DTI yielded improved prediction accuracy compared with previous models such as MFDR [2] or DeepDTI [3]. DrugVQA [11] uses compound simplified molecular input line entry system (SMILES) [12] strings and protein distance matrices as inputs, which

Wu *et al. BMC Bioinformatics*   (2022) 23:543

Page 2 of 12

are then processed by bidirectional long short-term memory networks and convolutional neural networks, respectively. This model outperforms some previous methods [9, 13] on the area under receiver operator characteristic curve (AUC) scores and provides a way to show important sites on compounds and proteins by attention visualization. In terms of binding affinity prediction, DeepDTA [4] and GraphDTA [5] are representative models. DeepDTA uses compound SMILES strings and protein sequences to predict affinity. These compound SMILES strings and protein sequences are both processed by convolutional neural networks. Compared with DeepDTA, GraphDTA uses compound graphs and graph neural networks instead of SMILES strings and convolutional neural networks, which causes lower prediction errors. Recently, a multi-objective neural network (MONN) [7] was proposed, which combines interaction site prediction and binding affinity prediction. Compound graphs and protein sequences are used in this model, which are processed by graph convolutional networks and convolutional neural networks, respectively. Compared with previous models [8–10], the classification AUC of interaction site prediction was significantly improved. These interaction site prediction results were further utilized to benefit the prediction of binding affinities.

However, for most of the previously mentioned models [2–9], protein representations are simply encoded by protein primary sequences. Several self-supervised learning approaches have become available, such as UniRep [14] and TAPE-BERT [15], which learn from millions of protein sequences. These protein embeddings have shown good performance for protein stability prediction and green fluorescence protein (GFP) activity prediction [15]. Recently, we introduced PtsRep [16], a self-supervised learning method trained on 35,568 protein tertiary structures. PtsRep was shown to have comparable or better performance than UniRep and TAPE-BERT in terms of protein stability prediction and GFP activity prediction [15]. We reasoned that these protein embeddings would be useful for improving predictions related to compound–protein interaction. To this end, we used MONN [7] as a backbone, but incorporated the aforementioned self-supervised protein embeddings to better improve the protein process module (we have termed this model as SPE-MONN, Fig. 1). Our results indeed showed that these
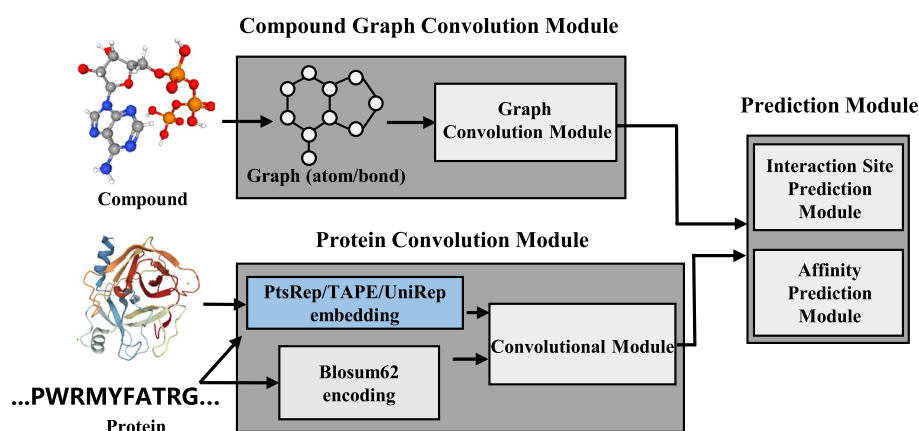

Fig. 1 The model architecture of SPE-MONN. This algorithm consists of three parts. (1) A protein convolution module. (2) A compound graph convolution module. (3) A prediction module for two tasks. This architecture was partially adapted from MONN [7]

Wu *et al. BMC Bioinformatics*   (2022) 23:543

Page 3 of 12

modifications were beneficial for predictions. Compared with previous advanced models [5, 7], SPE-MONN performed better for compound–protein interaction site and binding affinity prediction.

Methods

Prediction objective

Two prediction objective definitions for SPE-MONN were as described in MONN [7]. Briefly, one was the interaction site prediction between compound atoms and protein residues. The representation of interaction sites between a compound with N_m atoms and a protein with N_p residues was a pairwise interaction matrix $PIM \in R^{N_m \times N_p}$. Binding affinity predication was the second objective, which was a regression task.

Feature representation

The proposed model SPE-MONN utilized three types of embedding, PtsRep [16], UniRep [14], and TAPE-BERT [15]. In order to obtain the PtsRep embedding of a target protein, each residue of the protein sequence was represented by 10 properties (i.e., bulkiness, hydrophobicity, and relative spatial distance, among others) of its K nearest residues (*KNR*) in Euclidean space [17]. A bidirectional language model [18] taking *KNR* as an input was used to predict the two contiguous residues beside any given residues in both directions. This pre-trained model was used as a protein encoder. The PtsRep embedding of each protein had a shape like $N_p \times 768$. PtsRep was used by default for SPE-MONN below. UniRep [14], utilizing an mLSTM model pretrained on 24 million protein sequences, is one of the most effective self-supervised learning representations. TAPE-BERT [15] is another effective representation that was pre-trained on approximately 32 million protein sequences using the BERT [19] format. For simplicity, we will refer to TAPE-BERT as TAPE below. In this work, the 1900-dimensional UniRep embedding and the 768-dimensional TAPE embedding were used as a comparison. A compound graph used here was the same as in MONN [7]. Briefly, a compound can be described as a graph $G_m(V_m, E_m)$, where V_m consists of N_m atom nodes with a fixed dimension atom feature. E_m consists of edge information in a graph and the edge corresponds to those chemical bonds between the atoms in a given compound.

Model architecture of SPE-MONN

The architecture of SPE-MONN was a modification of that for MONN [7] in that the protein convolution module was expanded to contain two parts. The first one was a convolutional neural network (CNN) to extract intrinsic information from self-supervised protein embedding, while the second one was a CNN module for protein sequence evolutionary information extraction, similar to what was used in MONN [7].

For protein embedding extraction, a convolutional layer with Conv1D and Leaky Rectified Linear Unit (Leaky ReLU) were applied to reduce self-supervised embedding to *d1*-dimensions. In particular, a self-supervised embedding $P_0 \in R^{N_p \times d_0}$ was computed and represented by $P_1 \in R^{N_p \times d_1}$. P_1 further passed through multiple convolutional layers to obtain P_{embed}. The dimension of P_{embed} was consistent with P_1. The convolutional layer number N was 4 and the output dimension was 128.

$$P_K = LeakyReLU(Covn1d(P_{K-1})), K \in 1 \ldots N.$$

For evolutionary information extraction, each protein sequence was first encoded through the BLOSUM62 [20] matrix, and then processed via CNN to obtain the final sequence evolutionary representation P_{evo}. The outputs of the two CNN modules were combined for a combined protein representation, $P_{combine}$ as defined below.

$$P_{combine} = \alpha P_{embed} + (1 - \alpha)P_{evo}.$$

The compound graph convolution module and the downstream task prediction module were described in MONN [7]. Briefly, a graph convolution network was used here. A message passing unit [21] was used to aggregate information from neighbouring atoms and bonds and a graph warp unit [22] was used to aggregate information from super nodes, which represented compound global features.

For interaction site prediction, fully connected layers and a sigmoid function were applied to process compound representation and protein representation. A matrix with the shape $N_m \times N_p$ was the final interaction site prediction result. Based on a dual attention network [7, 23] and fully connected layers, the final affinity prediction results were obtained.

For compound–protein pairs, the loss function is defined as follows:

$$L_{total} = \lambda L_p + L_a.$$

L_p and L_a represent the loss function on interaction site prediction and affinity prediction, respectively. L_p can be defined as follows:

$$L_p = -\frac{1}{N}\sum\nolimits_{n=1}^{N}\sum\nolimits_{i=1}^{N_m^{(n)}}\sum\nolimits_{j=1}^{N_p^{(n)}} (PIM_{ij}^{(n)}\log\left(y_{ij}^{(n)}\right) + (1 - PIM_{ij}^{(n)})\log(1 - y_{ij}^{(n)})).$$

$PIM_{ij}^{(n)}$ is the label for interaction between the i-th compound atom and the j-th protein residue of the n-th pair. $y_{ij}^{(n)}$ is the prediction probability. L_a can be defined as follows:

$$L_a = \frac{1}{N}\sum\nolimits_{n=1}^{N}(y_a^{(n)} - \widehat{y}_a^{(n)})^2.$$

$\widehat{y}_a^{(n)}$ is the binding affinity label of the n-th pair, $y_a^{(n)}$ is the prediction score and λ is set to 0.1.

The SPE-MONN model was implemented using PyTorch and run on a Nvidia GeForce 2080 Ti. The Adam optimizer was used, and the learning rate was set to 0.0005 at first, followed by a change to a step size of 20.

Datasets

PDBbind [24, 25] provides binding affinity data for biomolecular complexes stored in the Protein Data Bank (PDB) [26]. The interaction sites in each complex were extracted using PLIP [27]. In order to compare performance with MONN, we used the same PDB-bind version, PDBBind v2018, and processed the data using the same processing methods as described for MONN [7]. The difference was that we used the PDB sequence instead of the UniProt sequence because the PtsRep construction required PDB data.

Wu *et al. BMC Bioinformatics*   (2022) 23:543

Page 5 of 12

After processing, 23,985 pairs were obtained for interaction site prediction and 14,402 of them with K_i or K_d affinity labels were used for binding affinity prediction. Referring to MONN [7], a clustering based cross-validation was used here. Single-linkage clustering [28] was used to gather similar data for drug compounds and proteins by distance measurements. According to the number of clusters and the number of elements in the cluster (Additional file 1: Tables S1, S2), the range of distance threshold was set from 0.3 to 0.6. For the new-compound setting and the new-protein setting, data splitting was based on compound clusters and protein clusters, respectively. Five-fold cross-validation was performed for these two settings. For the both-new setting, the protein and compound clusters were both considered and nine-fold cross-validation was performed. We used AUC to evaluate the performance of interaction site prediction, while we used the Pearson correlation coefficient and root mean square error (RMSE) to assess binding affinity prediction [7].

Results

Performance on the PDBbind v2018 dataset

The results from interaction site predications are shown in Fig. 2A. SPE-MONN-PtsRep (0.811, 0.849) outperformed MONN (0.729, 0.811) with an increase of 11.2%,
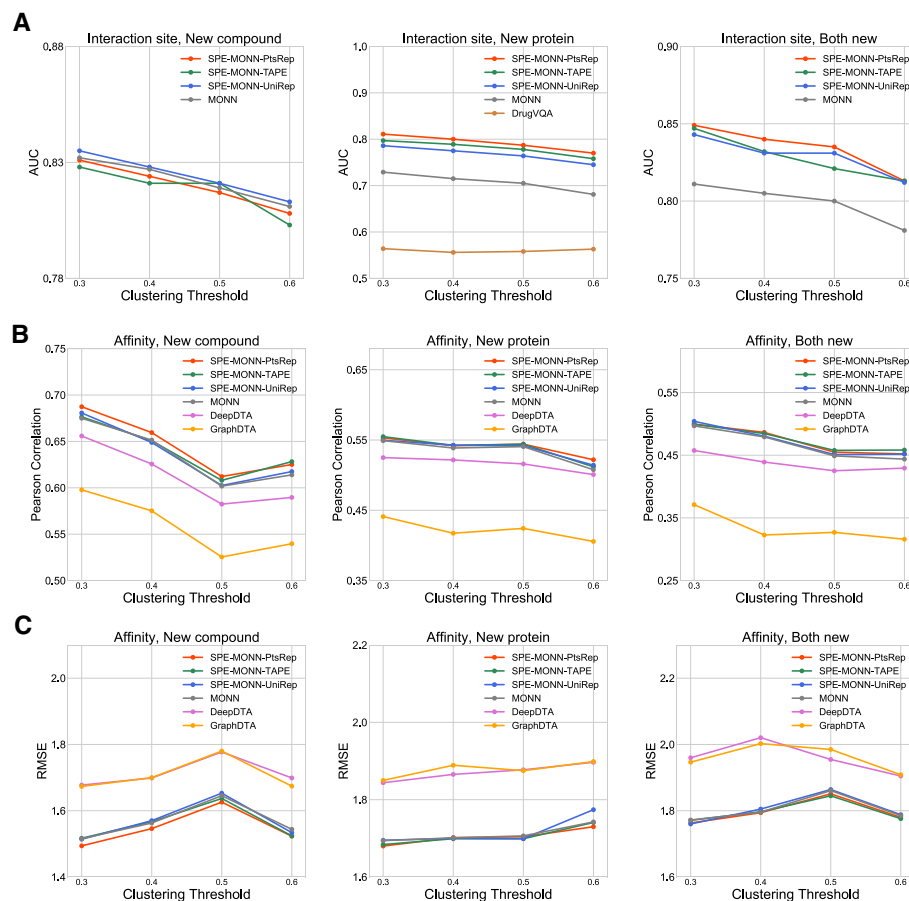

Fig. 2 The performances on the PDBbind dataset. The AUC results of interaction site prediction (**A**). The Pearson correlation coefficient results (**B**) and RMSE results (**C**) of binding affinity prediction

Wu *et al. BMC Bioinformatics*   (2022) 23:543

Page 6 of 12

4.7% for the AUC in the new-protein and the both-new settings, respectively, when the clustering threshold was set to 0.3. In addition, the model using TAPE or UniRep showed the same trend as the model using PtsRep in these two settings. The AUC of the TAPE model in the new-protein setting was 0.797 and the UniRep model was 0.786. In the both-new setting, their AUCs were 0.847 and 0.843, respectively. In the new-compound setting, the AUC of SPE-MONN-UniRep was 0.835, which was better than MONN (0.832). The AUCs of SPE-MONN-PtsRep (0.829) and SPE-MONN-TAPE (0.828) were slightly lower than MONN. The performance rankings of the four models in the clustering thresholds 0.4, 0.5, and 0.6 were similar to that in 0.3. We also ran DrugVQA under the new-protein setting, and the AUCs were found to be lower than those of both SPE-MONN and MONN in all the four clustering thresholds. Taken together, using self-supervised protein embeddings and a modification of the protein module in MONN made the site predictions on the protein and the overall site predictions more accurate, and SPE-MONN-PtsRep had the best performance using this data set.

The results in terms of binding affinity prediction are shown in Fig. 2B, C. We added DeepDTA and GraphDTA for comparison. They are representative models on binding affinity prediciions. GraphDTA had four variants of graph neural networks, we tested on all of them, and the data presented here was from the GCN-GAT variant, which had the best performance. As shown in Fig. 2B, C, in the new-compound setting SPE-MONN-PtsRep generally outperformed MONN in terms of both Pearson correlation coefficient and the RMSE. The Pearson correlation coefficient for SPE-MONN-PtsRep in this setting when the clustering threshold set to 0.3 was 0.687, and it was 0.675 for MONN, 0.656 for DeepDTA, and 0.598 for GraphDTA, while the RMSE of SPE-MONN-PtsRep, MONN, DeepDTA and GraphDTA were 1.491, 1.516, 1.677 and 1.673, respectively. SPE-MONN-PtsRep returned the lowest RMSE among all of these models. In the new-protein and the both-new settings, the average increases of the Pearson correlation coefficient of SPE-MONN-PtsRep compared to MONN at 4 clustering thresholds were both 0.006, when compared to DeepDTA, the average increases were 0.024 and 0.035, and when compared to GraphDTA, the corresponding increases were 0.118 and 0.139. The average decreases of the RMSE of SPE-MONN-PtsRep compared to MONN at 4 clustering thresholds were 0.007 and 0.006, when compared to DeepDTA, the average decreases were 0.167 and 0.163, and when compared to GraphDTA, the average decreases were 0.174 and 0.164. Thus, compared to DeepDTA and GraphDTA, SPE-MONN showed significant improvement both in the new-protein and the both-new settings. The results of SPE-MONN and MONN were close, but the former still performed slightly better overall.

In addition, we have been compared our models with the HPC/HWPC models [29], which were also trained and tested on PDBbind dataset. We followed HPC/HWPC models, and conducted experiments on PDB-2016, PDB-2013 and PDB-2007. As shown on Table 1, on PDB-2016, our models were comparable or slightly better than the HPC model on both Pearson correlation coefficient and RMSE; while the HPC-HWPCs model performed better or slightly better than our models overall, except for partial results on RMSE, for which our SPE-MONN-TAPE/UniRep model performed better. According to this literature, we also conducted experiments to compare these

Table 1 The Pearson correlation coefficient results and RMSE results (in parentheses) for our models and HPC/HWPC models in PDBbind-v2016, PDBbind-v2013 and PDBbind v2007

Dataset	SPE-MONN-PtsRep	SPE-MONN-TAPE	SPE-MONN-UniRep	HPC	HPC-HWPCs (η1 & η2)
PDB-2016	0.810 (1.321)	0.816 (1.301)	0.823 (1.272)	0.810 (1.359)	0.831 (1.307)
PDB-2013	0.791 (1.431)	0.796 (1.420)	0.797 (1.412)	0.770 (1.508)	0.784 (1.483)
PDB-2007	0.835 (1.389)	0.831 (1.399)	0.822 (1.429)	0.813 (1.423)	0.829 (1.403)

models on PDB-2013 and PDB-2007. Our three models performed slightly better on both Pearson correlation coefficient and RMSE compared with HPC and HPC-HWPCs models.

While our models do not comprehensively outperform the HPC-HWPCs model, but our models can also perform interaction site prediction in addition to affinity prediction. Moreover, compared with the machine learning method (hypergraph-based persistent cohomology) used in the HPC-HWPCs model, our models require less time. We surmise that it may be possible to improve the overall performance of the HPC-HWPCs model using the protein embeddings from TAPE/UniRep/PtsRep.

In general, SPE-MONN produced improved results both in terms of interaction site prediction and binding affinity prediction tasks, and SPE-MONN-PtsRep had the best performance. The above results were obtained when the α for protein representation combination was 0.5.

Identifying how to make better use of protein representations

By processing protein inputs through the protein convolution module, three different kinds of protein representations $\mathbf{P_{embed}}$, $\mathbf{P_{evo}}$ and $\mathbf{P_{combine}}$ were obtained. In order to explore which protein representation was more suitable for interaction site prediction and binding affinity prediction, we applied the three types of protein representations mentioned above to two prediction tasks, respectively.

The results in terms of interaction site prediction are shown in Fig. 3A. In the new-compound setting, the three models using $\mathbf{P_{embed}}$, $\mathbf{P_{evo}}$ and $\mathbf{P_{combine}}$ performed closely. In the new-protein setting, the results of the $\mathbf{P_{evo}}$ model were the worst. Compared with the $\mathbf{P_{evo}}$ model, the $\mathbf{P_{embed}}$ model improved significantly, while the $\mathbf{P_{combine}}$ model had the best performance. In the both-new setting, the performance rankings of the three models was consistent with that in the new-protein setting. The results on binding affinity prediction are shown in Fig. 3B and C. In the new-compound setting, the $\mathbf{P_{evo}}$ model had the lowest Pearson correlation coefficient and the highest RMSE. The other two models had better performance than the $\mathbf{P_{evo}}$ model. In contrast, in the new-protein and the both-new settings, the results were different from the new-compound setting. The $\mathbf{P_{evo}}$ model had the best performance, while the $\mathbf{P_{embed}}$ model had the worst performance. The $\mathbf{P_{combine}}$ model had improved results compared with the $\mathbf{P_{embed}}$ model, but this was still inferior to the $\mathbf{P_{evo}}$ model.

On interaction site prediction, the protein embedding representation was beneficial for the improvement of model performance. Compared with the protein evolutionary representation, the protein embedding representation had been pre-trained on a large number of protein sequences or structures, so it contained more diversified biological
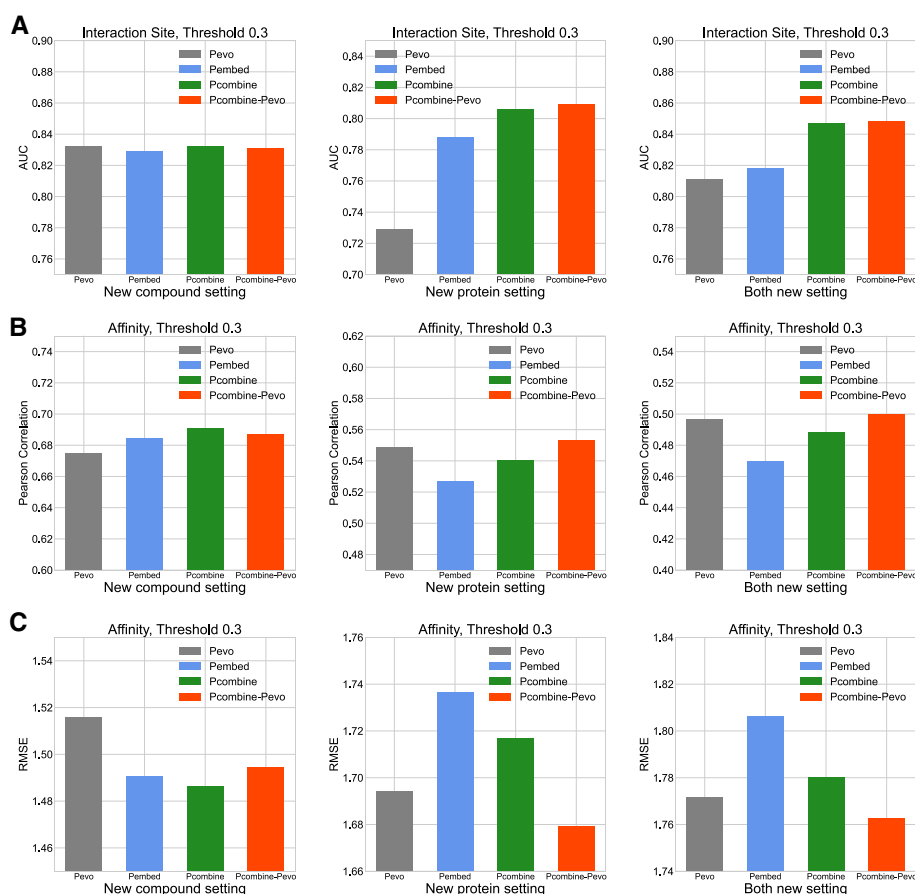
Fig. 3 Impacts on different types of protein representations.The AUC results of interaction site prediction (**A**). The Pearson correlation coefficient results (**B**) and RMSE results (**C**) of binding affinity prediction

semantics. For binding affinity prediction, the P_{evo} model relative to the P_{embed} model showed better performance. These results suggested that, although the protein embedding already contained abundant information, in terms of sequence evolutionary information, it was still lacking, and this term was an important factor that affected affinity prediction. To some extent, the results of the $P_{combine}$ model showed the benefit of P_{evo}. Compared with the P_{embed} model, the $P_{combine}$ model's performance on interaction site prediction was further improved. In terms of binding affinity prediction, its performance was also improved but lower than the P_{evo} model.

In order to achieve the optimal performance for both predictions, $P_{combine}$ was used for interaction site prediction and P_{evo} was used for binding affinity prediction. The results are shown in Fig. 3. In both tasks, this model achieved the best performance. Therefore, this usage of protein representation was adopted.

We then compared the performance of five different combination ratios, namely, 0, 0.2, 0.5, 0.8, and 1, when the clustering threshold was 0.3. The results on interaction site prediction are shown in Additional file 1: Table S3. In the new-compound setting, the results of the five combination ratios were equivalent. In both the new-protein and the both-new settings, the model's performance increased gradually when α increased from 0 to 0.5, and decreased gradually when α exceeded 0.5. When $\alpha = 0.5$, the model

Wu *et al. BMC Bioinformatics*   (2022) 23:543

Page 9 of 12

obtained the highest AUC value. The results on binding affinity prediction are shown in Additional file 1: Table S3. Among these three settings, when $\alpha = 0.5$, the model's Pearson correlation coefficient was the highest and its RMSE value was the lowest. Combining the results from two predictions, it could be concluded that $\alpha = 0.5$ was the best combination hyperparameter, and decreasing or increasing this value led to the degradation of model performance.

Case study

The SARS-CoV-2 main protease is considered as a drug promising target [30]. Some inhibitors interacting with it had been selected that are thought to function against this virus. There was a study [31] that reported 8 inhibitors interacting with the SARS-CoV-2 main protease. Of these inhibitors, 4 of them interacting with the protease (PDB id: 6W63) were identified from FDA-approved drugs, including Dobutamine, Apicidin, Nelfinavir, and Teniposide. The others interacting with the protease (PDB id: 6Y2F) were collated from CHEMBL, namely, CHEMBL206650 (C1), CHEMBL303543 (C2), CHEMBL127888 (C3), and CHEMBL573507 (C4). We applied our trained models to predict the interaction sites. A total of 25 sites were identified [31] and Table 2 lists the rankings of interaction sites

Table 2 Results of interaction site prediction ranking on SARS-CoV-2 related proteins and compounds

Compound	Site	Ranking			
		SPE-MONN-PtsRep	SPE-MONN-TAPE	SPE-MONN-UniRep	MONN
Dobutamine	41	**1**	**1**	**1**	**1**
	49	106	41	**4**	50
	166	**4**	**2**	**8**	**3**
Apicidin	166	**2**	**2**	11	**2**
	189	**7**	**4**	40	16
Nelfinavir	141	39	26	44	17
	143	**5**	**6**	58	**4**
	166	**3**	**2**	11	**2**
	189	**8**	**4**	32	15
Teniposide	141	**4**	**6**	46	**4**
	166	**7**	**3**	24	12
C1	140	**6**	16	17	**2**
	143	**3**	**5**	25	**3**
	144	**9**	**8**	16	16
	167	26	57	135	116
	189	**10**	**4**	15	12
C2	166	**1**	**1**	**2**	**1**
	167	26	52	131	119
	189	**8**	**4**	14	12
C3	25	12	12	**6**	20
	142	**5**	**9**	29	**7**
	189	**9**	**4**	15	**10**
C4	141	13	29	36	14
	142	**5**	**9**	26	**6**
	166	**1**	**1**	**2**	**1**

Items in bold were predicted to be the top 10 sites

predicted by SPE-MONN and MONN. Items in bold were predicted to be the top 10 sites. We counted the top 10 sites on the protein side of each compound–protein pair. Among these 25 sites, SPE-MONN-PtsRep revealed 19 of the top 10 sites, while MONN returned 13. The results from SPE-MONN-TAPE and SPE-MONN-UniRep were 18 and 6, respectively. Of the average results from site prediction ranking, SPE-MONN-PtsRep suggests a real site on the 13th out of all sites and MONN suggests a site on the 19th. The ranking results of SPE-MONN-TAPE and SPE-MONN-UniRep were 13th and 25th, respectively. SPE-MONN-PtsRep and SPE-MONN-TAPE thus could more accurately find possible compound binding sites on proteins. These results suggested that the SPE-MONN model was a helpful tool for analyzing the interaction between compounds and proteins.

Discussion

Accurate prediction of compound–protein interaction-related tasks could facilitate improved drug discovery and drug design. In this work, we present SPE-MONN, a compound–protein interaction site and binding affinity predicting method based on MONN, and this model utilizes self-supervised protein embedding. Compared with previous models [4, 5, 7, 11], the results demonstrated that our model outperformed them in both prediction tasks. PtsRep protein embedding likely enriches potential protein information due to the use of protein structure and protein properties. TAPE and UniRep protein embeddings learn the intrinsic information from proteins from a large number of protein sequences. They are thus helpful for predictions because more useful information is included in protein embeddings. Evolutionary information from protein sequences was also shown to be a factor influencing prediction [7], so we used both protein embedding and protein sequence evolutionary information together to improve the overall performance of our model. The results from our case study showed that our model could also make predictions that were closer to the correct results. Thus, our model is a useful tool for compound–protein-related predictions.

We also attempted to explore a graph neural network more suitable for two prediction tasks, including three graph neural network modules [5], GAT [32], GIN [33] and GCN-GAT [5], which were used to replace the compound convolution module in SPE-MONN. The results (Additional file 1: Fig. S1) showed that the current graph convolution module [7] was the best choice.

The advent of AlphaFold2 [34] greatly reduces the difficulty of obtaining protein structures. We expect more advanced protein embedding methods will emerge along with the increasing entries for protein structures. Recently, there was a method demonstrated in the literature [35] that uses 3D-CNN to process protein tertiary structures and it has shown the ability to sense interactions within proteins and succeed in mutation guidance. This and others protein processing methods [36–38] will have implications for predictive accuracy on compound–protein interaction, and accelerate the process of drug discovery and design.

Conclusion

In this paper, a model called SPE-MONN is proposed to predict the compound–protein interactions. It is based on the published model MONN and we utilize the protein embedding gained from self–supervised learning and modify the related protein convolution module for improvement. The experimental results show that the performance of

Wu *et al. BMC Bioinformatics*   (2022) 23:543

Page 11 of 12

SPE-MONN is improved both in interaction site prediction and binding affinity prediction. The case study also show that our model can better predict the interaction sites. In general, SPE-MONN could be a helpful tool for compound–protein interaction study.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-022-05107-w.

Additional file 1. Supplementary Figures S1–S2 and Tables S1-S3. Fig. S1. Impacts on different graph neural networks. The AUC results of interaction site prediction (**A**). The Pearson correlation coefficient results (**B**) and RMSE results (**C**) of binding affinity prediction. **Fig. S2.** Impacts of different lambda values for SPE-MONN-PtsRep. The AUC results of interaction site prediction (**A**). The Pearson correlation coefficient results (**B**) and RMSE results (**C**) of binding affinity prediction. **Table S1.** Clustering results of compounds at different clustering thresholds. **Table S2.** Clustering results for proteins at different clustering thresholds. **Table S3.** Impacts of the hyperparameter α on protein representation combination. The area under receiver operator characteristic curve (AUC) is the result of interaction site prediction. The Pearson correlation coefficient and root mean square error (RMSE) are results of binding affinity prediction. NC, NP, and NN are new-compound, new-protein, both-new settings, respectively.

Author contributions

JW designed and implemented the algorithm, carried out the experiments, and drafted the manuscript; ZL analysed the data; XY and ZL supervised the study, and revised the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

The dataset and code used in this article can be found in: https://github.com/Jwoods14/SPE-MONN.git.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References

1. Mathur A, Loskill P, Shao K, et al. Human iPSC-based cardiac microphysiological system for drug screening applications. Sci Rep. 2015;5:8883.
2. Hu P-W, Chan KC, You Z-H. Large-scale prediction of drug-target interactions from deep representations. Int Jt Conf Neural Netw. 2016;1236–43.
3. Wen M, Zhang Z, Niu S, et al. Deep-learning-based drug–target interaction prediction. J Proteome Res. 2017;16(4):1401–9.
4. Ozturk H, Ozgur A, Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction. Bioinformatics. 2018;34(17):i821–9.
5. Nguyen T, Le H, Quinn TP, et al. GraphDTA: predicting drug–target binding affinity with graph neural networks. Bioinformatics. 2021;37(8):1140–7.
6. Lee I, Keum J, Nam H. DeepConv-DTI: prediction of drug-target interactions via deep learning with convolution on protein sequences. PLoS Comput Biol. 2019;15(6):e1007129.
7. Li S, Wan F, Shu H, et al. MONN: a multi-objective neural network for Predicting compound-protein interactions and affinities. Cell Syst. 2020;10(4):308–22.e311.
8. Gao KY, Fokoue A, Luo H, et al. Interpretable drug target prediction using deep neural representation. Int Jt Conf Artif Intell. 2018;3371–7.
9. Tsubaki M, Tomii K, Sese J. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. Bioinformatics. 2019;35(2):309–18.
10. Karimi M, Wu D, Wang Z, et al. DeepAffinity: interpretable deep learning of compound-protein affinity through unified recurrent and convolutional neural networks. Bioinformatics. 2019;35(18):3329–38.
11. Zheng S, Li Y, Chen S, et al. Predicting drug–protein interaction using quasi-visual question answering system. Nat Mach Intell. 2020;2(2):134–40.

12. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Model. 1988;28(1):31–6.
13. Ballester PJ, Mitchell JB. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics. 2010;26(9):1169–75.
14. Alley EC, Khimulya G, Biswas S, et al. Unified rational protein engineering with sequence-based deep representation learning. Nat Methods. 2019;16(12):1315–22.
15. Rao R, Bhattacharya N, Thomas N, et al. Evaluating protein transfer learning with TAPE. Adv Neural Inf Process Syst. 2019;32:9689.
16. Luo J, Cai Y, Wu J, et al. Self-supervised representation learning of protein tertiary structures (PtsRep) and its implications for protein engineering. *bioRxiv* 2021:2020.2012. 2022.423916.
17. Wang J, Cao H, Zhang JZ, et al. Computational protein design with deep learning neural networks. Sci Rep. 2018;8(1):1–9.
18. Peters ME, Neumann M, Iyyer M, et al. Deep contextualized word representations. 2018, arXiv http://arxiv.org/abs/1802.05365.
19. Devlin J, Chang M-W, Lee K, et al: Bert: pre-training of deep bidirectional transformers for language understanding. 2018 arXiv http://arxiv.org/abs/1810.04805.
20. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci. 1992;89(22):10915–9.
21. Lei T, Jin W, Barzilay R, et al: Deriving neural architectures from sequence and graph kernels. 2017 arXiv http://arxiv.org/abs/1705.09037.
22. Ishiguro K, Maeda S-i, Koyama M. Graph warp module: an auxiliary module for boosting the power of graph neural networks. 2019 arXiv http://arxiv.org/abs/1902.01020.
23. Nam H, Ha J-W, Kim J. Dual attention networks for multimodal reasoning and matching. IEEE Int Conf Comput Vis Pattern Recognit 2017:299–307.
24. Wang R, Fang X, Lu Y, et al. The PDBbind database: collection of binding affinities for protein–ligand complexes with known three-dimensional structures. J Med Chem. 2004;47(12):2977–80.
25. Wang R, Fang X, Lu Y, et al. The PDBbind database: methodologies and updates. J Med Chem. 2005;48(12):4111–9.
26. Berman HM, Kleywegt GJ, Nakamura H, et al. The Protein Data Bank archive as an open data resource. J Comput-Aided Mol Des. 2014;28(10):1009–14.
27. Salentin S, Schreiber S, Haupt VJ, et al. PLIP: fully automated protein-ligand interaction profiler. Nucleic Acids Res. 2015;43(W1):W443–7.
28. Gower JC, Ross GJ. Minimum spanning trees and single linkage cluster analysis. J R Stat Soc C-Appl. 1969;18(1):54–64.
29. Liu X, Wang X, Wu J, et al. Hypergraph-based persistent cohomology (HPC) for molecular representations in drug design. Brief Bioinf. 2021;22(5):1–12.
30. Ullrich S, Nitsche C. The SARS-CoV-2 main protease as drug target. Bioorg Med Chem Lett. 2020;30(17):127377.
31. Pant S, Singh M, Ravichandiran V, et al. Peptide-like and small-molecule inhibitors against Covid-19. J Biomol Struct Dyn. 2021;39(8):2904–13.
32. Velickovic P, Cucurull G, Casanova A, et al. Graph attention networks. Stat. 2017;1050:20.
33. Xu K, Hu W, Leskovec J, et al. How powerful are graph neural networks? arXiv 2018.
34. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–9.
35. Shroff R, Cole AW, Diaz DJ, et al. Discovery of novel gain-of-function mutations guided by structure-based deep learning. ACS Synth Biol. 2020;9(11):2927–35.
36. Hu L, Pan X, Tang Z, et al. A fast fuzzy clustering algorithm for complex networks via a generalized momentum method. IEEE Trans Fuzzy Syst. 2021;30(9):3473–85.
37. Pan X, Hu L, Hu P, et al. Identifying protein complexes from protein-protein interaction networks based on fuzzy clustering and GO semantic information. IEEE/ACM Trans Comput Biol Bioinf 2021:1–13.
38. Wang X, Yang W, Yang Y, et al. PPISB: a novel network-based algorithm of predicting protein–protein interactions with mixed membership stochastic blockmodel. IEEE/ACM Trans Comput Biol Bioinf 2022:1–8.