DeepProSite: structure-aware protein binding site prediction using ESMFold and pretrained language model

Yitian Fang, Yi Jiang, Leyi Wei, Qin Ma, Zhixiang Ren, Qianmu Yuan, Dong-Qing Wei

LifeLU reading group

presented by Özdeniz Dolu

26.06.2025

Introduction

- Understanding interactions between proteins and other biomolecules is crucial for understanding many biological processes.
- Traditional methods for binding site detection, such as X-ray crystallography, affinity purification-mass spectrometry etc. are expensive and time consuming.
- Beneficial to develop fast and accurate computational methods.

Introduction

There are generally 2 camps of computational methods in prediction of binding sites:

Sequence-based: DLPred, ProNA2020, DELPHI, SPRINT-Seq, PepBind, Visual, PepNN-Seq, pepBCL, and DNAPred.

Structure-based: SPPIDER, MaSIF-site, GraphPPIS, PepSite, Peptimap, SPRINT-Str, PepNNStruct, and GraphBind

Introduction

- Structure based methods require tertiary structure information which is much less abundant than sequence information.

- Sequence based methods often use evolutionary information. Extraction of evolutionary features (MSA, database search etc.) take time. Also have shortcomings for sequences with less similarity.

Introduction

Important developments:

Sequence to structure prediction: AlphaFold2, ESMFold and RoseTTAFold

ESMFold is noteworthy because it uses Protein Language Model(PLM) instead of MSA for extracting evolutionary features. Computationally lighter.

Transformer architecture in deep learning: Very significant performance increases throughout many deep learning tasks.

Graph neural networks: Use graph topology in deep learning

Introduction

Proposed model, DeepProSite, combines protein language model embeddings and ESMFold structure predictions for protein-protein/peptide binding site prediction.

Datasets

Datasets are compiled from previous studies.

Table 1. Statistics information of the benchmark datasets used in this study.

Type	Dataset	$N_{protein}$ [a]	N_{pos} [b]	N_{neg} [c]	PNratio [d]
Peptide	Pep_Train_1154	1154	15 030	261 792	0.057
	Pep_Test_125	125	1719	29 151	0.059
	Pep_Train_640	640	8259	149 103	0.055
	Pep_Test_639	639	8490	141 840	0.060
Protein	Pro_Train_335	335	10 374	55 992	0.185
	Pro_Test_60	60	2075	11 069	0.187
	Pro_Test_315	315	9355	55 976	0.167

[a] Number of proteins.
[b] Number of binding residues.
[c] Number of nonbinding residues.
[d] PNratio = N_{pos}/N_{neg}.

Protein graph construction

Prediction task is formulated as a “node classification task” in the protein graph.

If protein has n residues, for each residue (node);

- Sequence features
- Structural features
- 3D coordinates

$$H \in \mathbb{R}^{n \times d}$$

$$X \in \mathbb{R}^{n \times 3}$$

Predicted protein structures

Pretrained ESMFold model is used for structure information (esmfold_v1).

ESMFold is run on all benchmark datasets and structures of all proteins have been predicted.

Structural properties

For each residue in the ESMFold-predicted structures;

- Relative Solvent Accessibility (RSA)
- Secondary Structure Profile (8 Categories, one-hot encoded)
- Backbone torsion angles (phi and psy)

DSSP has been utilized for calculations.

In total, 14 features.

Language model embeddings

Protein language model ProtT5-XL-U50 (ProtT5 for short) is used to generate residue-level embeddings. Each embedding has 1024 features. Embeddings are normalized with min-max normalization.

$$x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$$

ProtT5 is a autoencoder utilizing Transformer architecture:

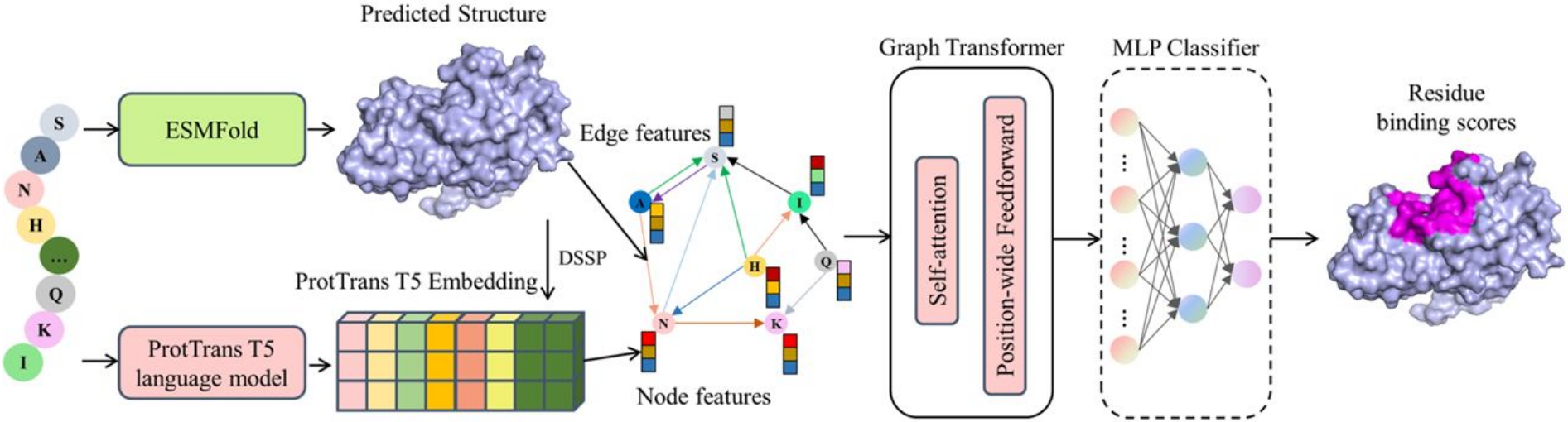
- Pre-trained on BFD (2.5B protein sequences)
- Fine-tuned on UniRef50 (0.5M protein sequences)

Overall pipeline

Concatenation of structural and sequential features -> node features.

Each node (residue) is connected to 30 nearest neighbors based on coordinates.

Also, additional edge features are computed (explained later).



Geometric edge features

e_ij = edge feature between i'th and j'th residue. These are **translation** and **rotation invariant**.

First feature: radial distance (rbf) between backbone c_a atoms

Oi = local coordinate system of x_i (i'th residue)

Second feature: direction of j'th residue from i'th residues perspective

$$e_{ij}^{(s)} = \left(r(||x_j - x_i||),\ O_i^{\mathrm{T}} \frac{x_j - x_i}{||x_j - x_i||},\ q\left(O_i^{\mathrm{T}} O_j\right) \right)$$

$$v_i = \frac{x_i - x_{i-1}}{||x_i - x_{i-1}||},\ b_i = \frac{v_i - v_{i+1}}{||v_i - v_{i+1}||},\ n_i = \frac{v_i \times v_{i+1}}{||v_i \times v_{i+1}||}$$

$$O_i = [b_i\ n_i\ b_i \times n_i]$$

Geometric edge features

e_ij = edge feature between i’th and j’th residue. These are **translation** and **rotation invariant**.

Third feature: quaternion representation of spatial rotation matrix (Oi^TOj). Contains orientation information.

$$e_{ij}^{(s)} = \left(r(||x_j - x_i||),\ O_i^{\mathrm{T}} \frac{x_j - x_i}{||x_j - x_i||},\ q\left(O_i^{\mathrm{T}} O_j\right) \right)$$

$$v_i = \frac{x_i - x_{i-1}}{||x_i - x_{i-1}||},\ b_i = \frac{v_i - v_{i+1}}{||v_i - v_{i+1}||},\ n_i = \frac{v_i \times v_{i+1}}{||v_i \times v_{i+1}||}$$

$$O_i = [b_i\ n_i\ b_i \times n_i]$$

Graph Transformer

|| operation is vector concatenation.

Formulas describe how attention scores and “node feature updates” are calculated.

$$h_i' = h_i + \sum_{j \in N(i) \cup i} \alpha_{ij} W_V (h_j \parallel e_{ij})$$

$$\alpha_{ij} = \text{softmax}\left(\frac{(W_Q h_i)^T (W_K (h_j \parallel e_{ij}))}{\sqrt{d}} \right)$$

Multi-Layer Perceptron

Multilayer uses the representations in the last layer of Graph Transformer to make a binary prediction for each residue: binding site or not.

$$Y' = \text{Sigmoid}\left(H^{(L)} W + b\right)$$

Evaluation

Since it's a binary classification task, conventional evaluation metrics have been used. However, one important note here is that, residue-level binding site classification task is very imbalanced (many more non-binding sites than binding sites).

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

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

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{11}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FN} \times \text{FP}}{\sqrt{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}} \tag{13}$$

Results

Geometry feature improves the model performance

Transformer: A baseline model that is provided same node features but no edge features (no geometry information).

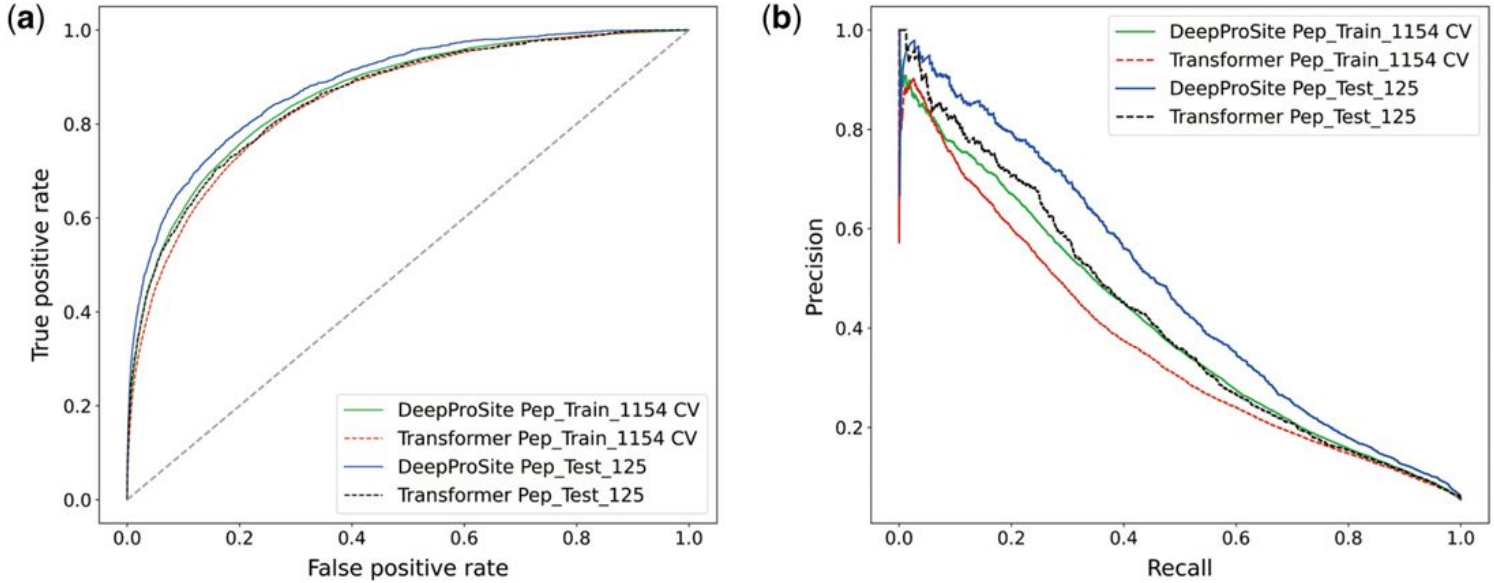
Table S5. Performance comparison with the geometric-agnostic baseline model Transformer on Pep_Test_125, Pep_Test_639, Pro_Test_60 and Pro_Test_315. The highest values are bolded.

Dataset	Method	Spe	Rec	Pre	F1	MCC	AUC	AUPRC	ACC
Pep_	Transformer	0.965	**0.441**	0.423	0.432	0.398	0.857	0.417	0.936
Test_125	DeepProSite	**0.983**	0.392	**0.578**	**0.467**	**0.451**	**0.883**	**0.480**	**0.950**
Pep_	Transformer	0.964	0.392	0.395	0.394	0.357	0.847	0.369	0.932
Test_639	DeepProSite	**0.972**	**0.400**	**0.460**	**0.428**	**0.397**	**0.861**	**0.411**	**0.940**
Pro_	Transformer	0.859	**0.550**	0.422	**0.478**	0.369	0.801	0.461	0.810
Test_60	DeepProSite	**0.917**	0.443	**0.501**	0.470	**0.379**	**0.813**	**0.490**	**0.842**
Pro_	Transformer	**0.854**	0.518	0.373	0.433	0.326	0.785	0.402	**0.806**
Test_315	DeepProSite	0.842	**0.576**	**0.378**	**0.457**	**0.355**	**0.805**	**0.432**	0.804

Results

Geometry feature improves the model performance

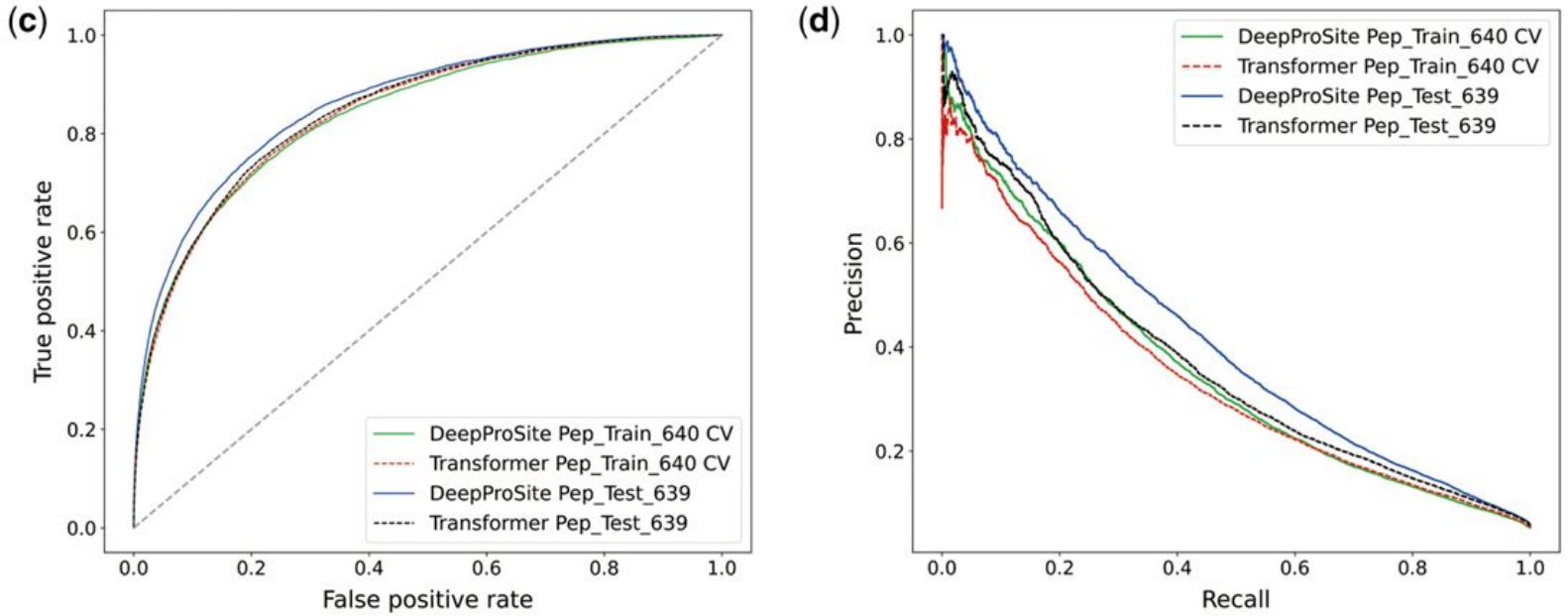
Transformer: A baseline model that is provided same node features but no edge features (no geometry information).



Results

Geometry feature improves the model performance

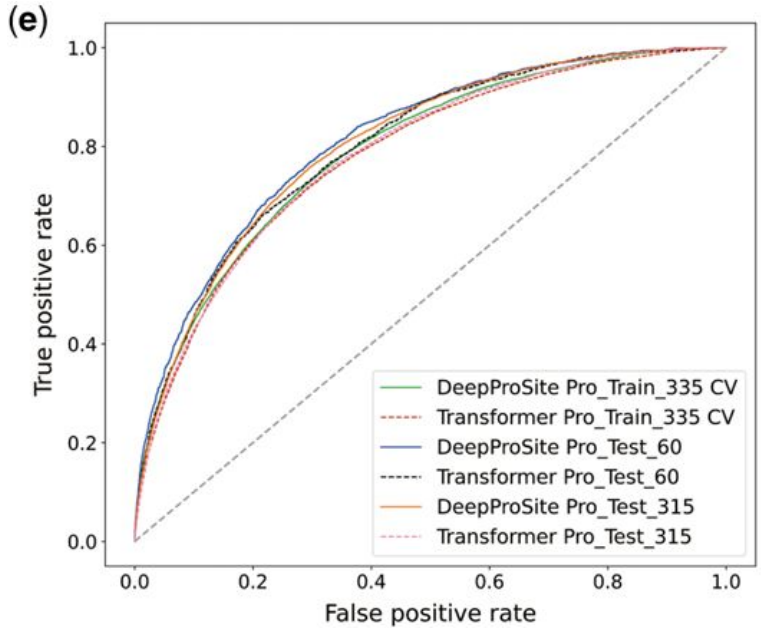
Transformer: A baseline model that is provided same node features but no edge features (no geometry information).

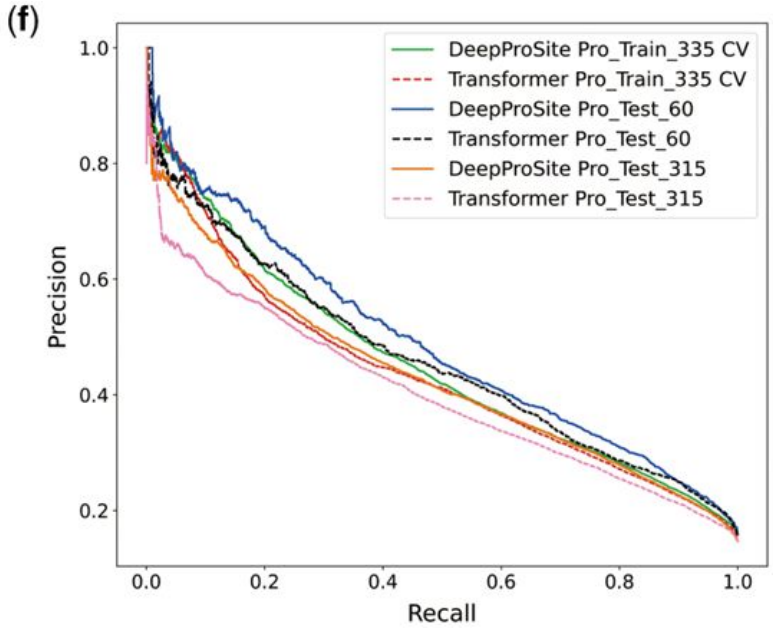


Results

Geometry feature improves the model performance

Transformer: A baseline model that is provided same node features but no edge features (no geometry information).

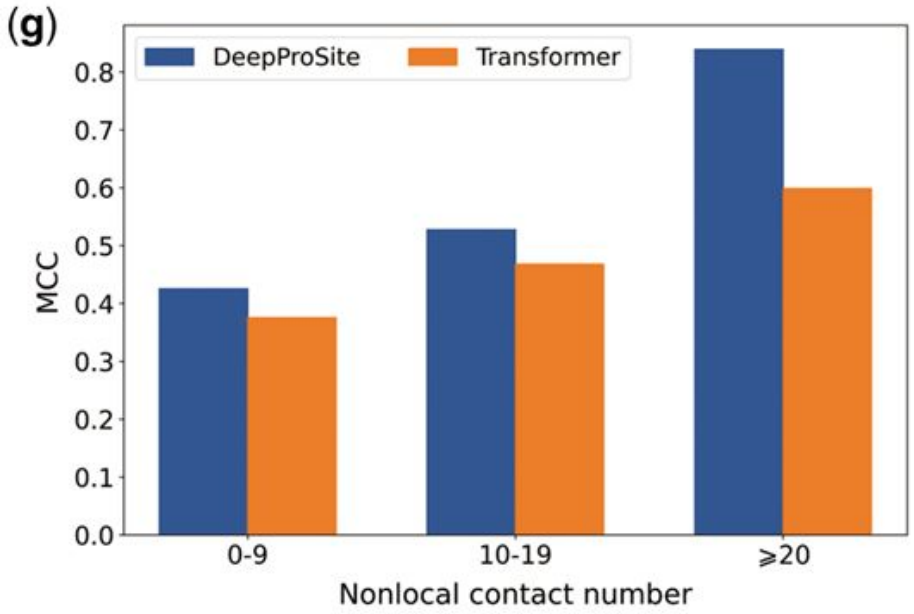


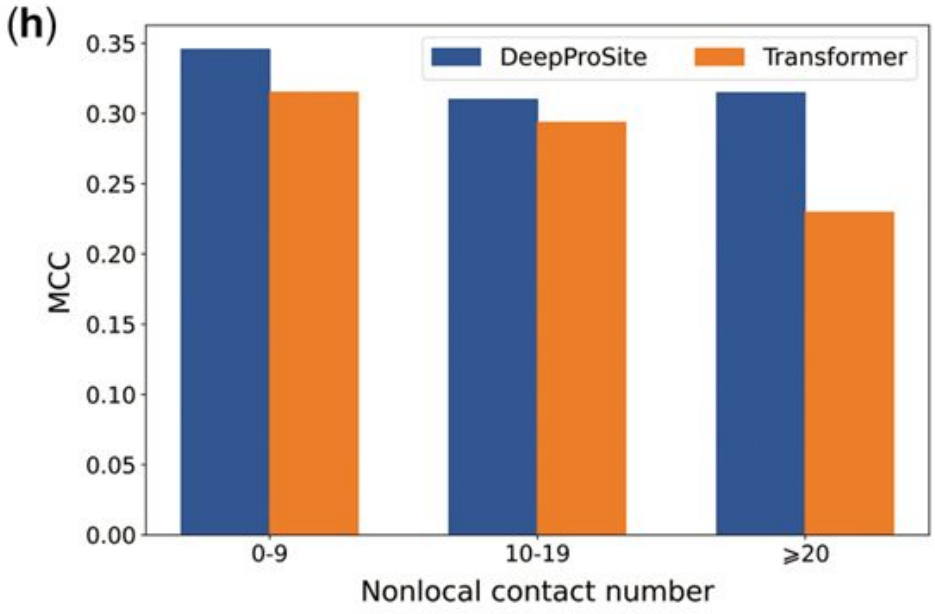


Results

Geometry feature improves the model performance

Here we see that as the non-local contact numbers increase, difference between DeepProSite from Transformer starts to become more visible.





Results

Feature importance and model ablation

EVO refers to a process of extracting evolutionary information based on conventional methods (position-specific scoring matrix (PSSM) and hidden Markov model (HMM) profile). (This is to replace PLM embeddings).

Table 2. Comparison of feature performance in predicting PBPs on Pep_Train_1154 and Pep_Test_125.[a]

Feature	Pep_Train_1154 CV		Pep_Test_125	
	AUC	AUPRC	AUC	AUPRC
EVO	0.831	0.323	0.855	0.426
DSSP	0.778	0.273	0.810	0.337
ProtT5	0.858	0.386	0.880	0.467
ProtT5+EVO+DSSP	0.861	0.402	0.875	0.470
EVO+DSSP	0.835	0.342	0.852	0.427
ProtT5+EVO	0.862	0.386	**0.883**	0.477
ProtT5+DSSP (DeepProSite)	**0.864**	**0.404**	**0.883**	**0.480**

Results

Feature importance and model ablation

Using AlphaFold2 improves the model but it's too costly for small improvement.

Table S6. The predictive performance of DeepProSite on Pep_Test_125 when using different predicted structures.

Structural information	Spe	Rec	Pre	F1	MCC	AUC	AUPRC	ACC
AlphaFold2 predicted structures	**0.985**	0.383	**0.603**	**0.468**	**0.457**	**0.888**	**0.482**	**0.952**
ESMFold predicted structures	0.983	**0.392**	0.578	0.467	0.451	0.883	0.480	0.950

Results

Feature importance and model ablation

30-nearest neighbors work better than fixed threshold.

Attention Matters.

Table S7. Performance comparison of fixed distance and fixed number on Pep_Test_125.

Neighbors selection	Spe	Rec	Pre	F1	MCC	AUC	AUPRC	ACC
C_α atoms <10 Å	0.979	0.408	0.539	0.464	0.442	0.879	0.464	0.948
k-nearest neighbors	**0.983**	**0.392**	**0.578**	**0.467**	**0.451**	**0.883**	**0.480**	**0.950**

Table S8. The ablation study on attention mechanism on Pep_Test_125.

Structural information	Spe	Rec	Pre	F1	MCC	AUC	AUPRC	ACC
DeepProSite w/o attention	**0.985**	0.353	**0.586**	0.440	0.430	0.870	0.457	**0.950**
DeepProSite	0.983	**0.392**	0.578	**0.467**	**0.451**	**0.883**	**0.480**	**0.950**

Results

Comparison with state-of-the-art methods on peptide datasets

Table 3. Performance comparison of DeepProSite with state-of-the-art methods on Pep_Test_125 dataset.[a]

Method	Spe	Rec	Pre	MCC	AUC
Pepsite	0.970	0.180		0.200	0.610
Peptimap	0.950	0.320		0.270	0.630
SPRINT-Seq	0.960	0.210		0.200	0.680
SPRINT-Str	0.980	0.240		0.290	0.780
Visual	0.680	0.670		0.170	0.730
PepBind		0.344	0.469	0.372	0.793
PepNN-Seq				0.278	0.805
PepNN-Struct				0.321	0.841
PepBCL	**0.984**	0.315	0.540	0.385	0.815
DeepProSite	0.983	**0.392**	**0.578**	**0.451**	**0.883**

Table 4. Performance comparison of DeepProSite with state-of-the-art methods on Pep_Test_639 dataset.

Method	Spe	Rec	Pre	MCC	AUC
PepBind		0.317	0.450	0.348	0.767
PepNN-Seq				0.251	0.792
PepNN-Struct				0.301	0.838
PepBCL	**0.983**	0.252	**0.470**	0.312	0.804
DeepProSite	0.972	**0.400**	0.460	**0.397**	**0.861**

Results

Comparison with state-of-the-art methods on protein datasets

Table 5. Performance comparison of DeepProSite with state-of-the-art methods on Pro_Test_60 dataset.

Method	ACC	Rec	Pre	F1	MCC	AUC	AUPRC
PSIVER	0.561	0.534	0.188	0.278	0.074	0.573	0.190
ProNA2020	0.738	0.402	0.275	0.326	0.176	N/A	N/A
SCRIBER	0.667	0.568	0.253	0.350	0.193	0.665	0.278
DLPred	0.682	0.565	0.264	0.360	0.208	0.677	0.294
DELPHI	0.697	0.568	0.276	0.372	0.225	0.699	0.319
DeepPPISP	0.657	0.539	0.243	0.335	0.167	0.653	0.276
SPPIDER	0.752	0.557	0.331	0.415	0.285	0.755	0.373
MaSIF-site	0.780	0.561	0.370	0.446	0.326	0.775	0.439
GraphPPIS	0.776	0.584	0.368	0.451	0.333	0.786	0.429
RGN	0.785	**0.587**	0.382	0.463	0.349	0.791	0.441
DeepProSite	**0.842**	0.443	**0.501**	**0.470**	**0.379**	**0.813**	**0.490**

Table 6. Performance comparison of DeepProSite with state-of-the-art methods on Pro_Test_315 dataset.

Method	ACC	Rec	Pre	F1	MCC	AUC	AUPRC
DeepPPISP	0.603	0.622	0.206	0.310	0.157	0.660	0.256
SPPIDER	0.744	0.613	0.305	0.407	0.294	0.783	0.376
MaSIF-site	0.764	0.589	0.322	0.417	0.304	0.778	0.372
GraphPPIS	0.739	**0.689**	0.313	0.430	0.329	0.798	0.423
DeepProSite	**0.804**	0.576	**0.378**	**0.457**	**0.355**	**0.805**	**0.432**

Results

Comparison with state-of-the-art methods on protein datasets

All four structure-based algorithms were trained with bound structures.

Therefore, they had a significant performance drop with unbound structures.

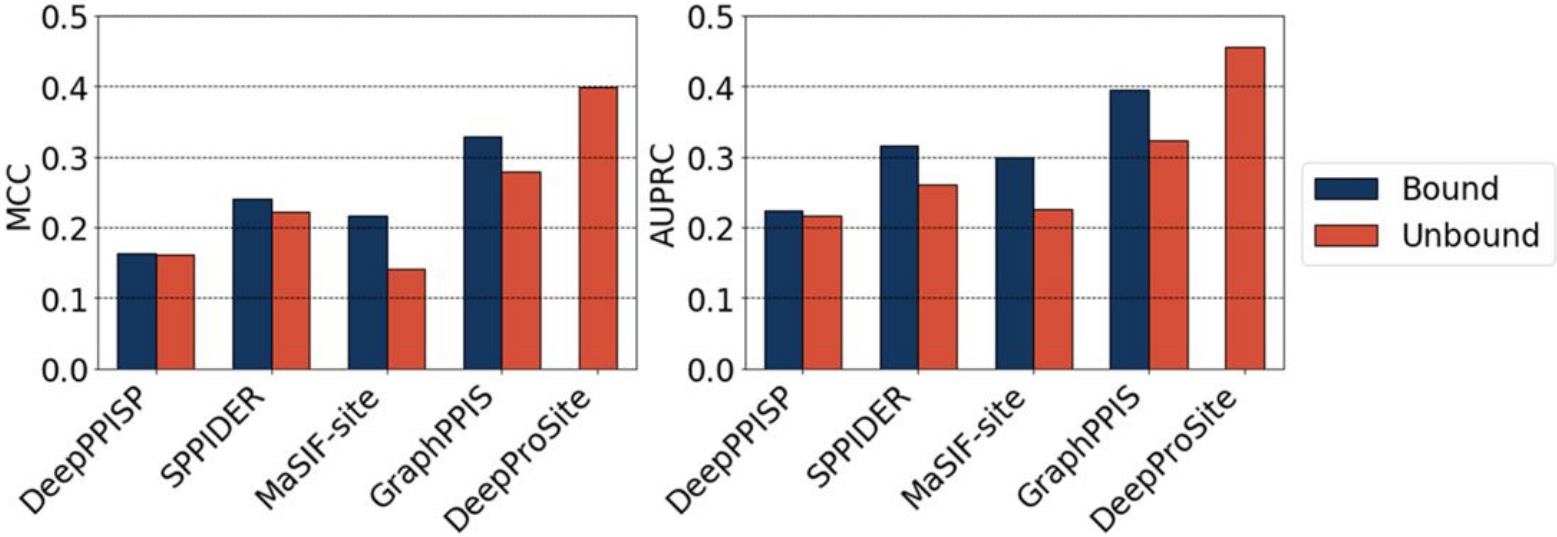


Figure 3. Performance comparison of DeepProSite with structure-based methods on 31 proteins with bound and unbound structures.

Results

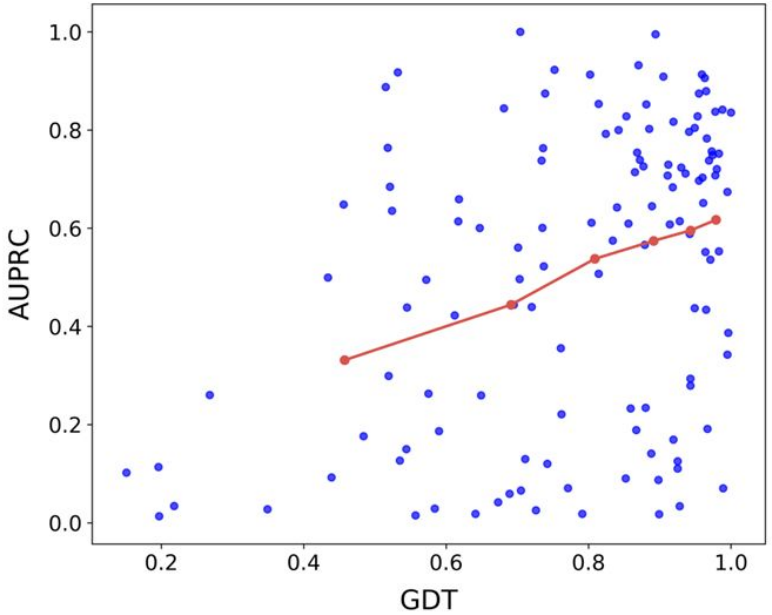
Influence of predicted protein structure quality

The average global distance test (GDT): GDT is calculated between predicted and native structures (based on SPalign).

Based on GDT scores, test data is put into

6 different bins. For each bin, AUPRC

is calculated (red dots)



Results

Case Study

4L3O (chain A), 4BVX (chain A), were randomly selected from the Pep_Test_125 and Pro_Test_315 datasets for illustrative purposes.

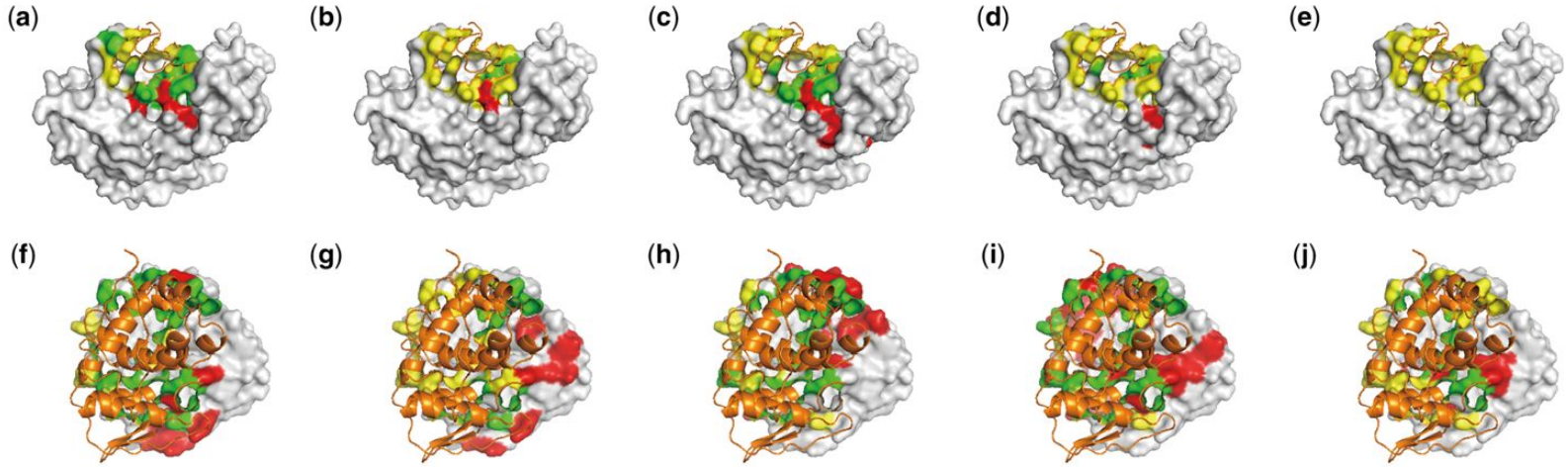


Figure 5. Visualization of predicted binding residues for two cases predicted by DeepProSite and other methods. The results predicted by DeepProSite (a), the geometric diagnostic baseline method Transformer (b), PepNN-Struct (c), PepBCL (d), and PepBind (e) are shown for the first protein (PDB ID: 4L3O, chain A) from Pep_Test_125. The results predicted by DeepProSite (f), Transformer (g), GraphPPIS (h), SPPIDER (i) and ProNA2020 (j) are shown for the second protein (PDB ID: 4BVX, chain A) from Pro_Test_315. The TP, FP, and FN are colored in green, red, and yellow, respectively.

Results

Extending DeepProSite to other types of ligands

Performance of DeepProSite upon being trained on other type of binding site prediction datasets is evaluated and compared with methods in their respective cases.

DeepProSite have shown good generalizability and performance in most cases.

Ligands Tested: DNA, RNA, ATP, HEME, carbohydrate, Mg2, Ca2, and Mn2

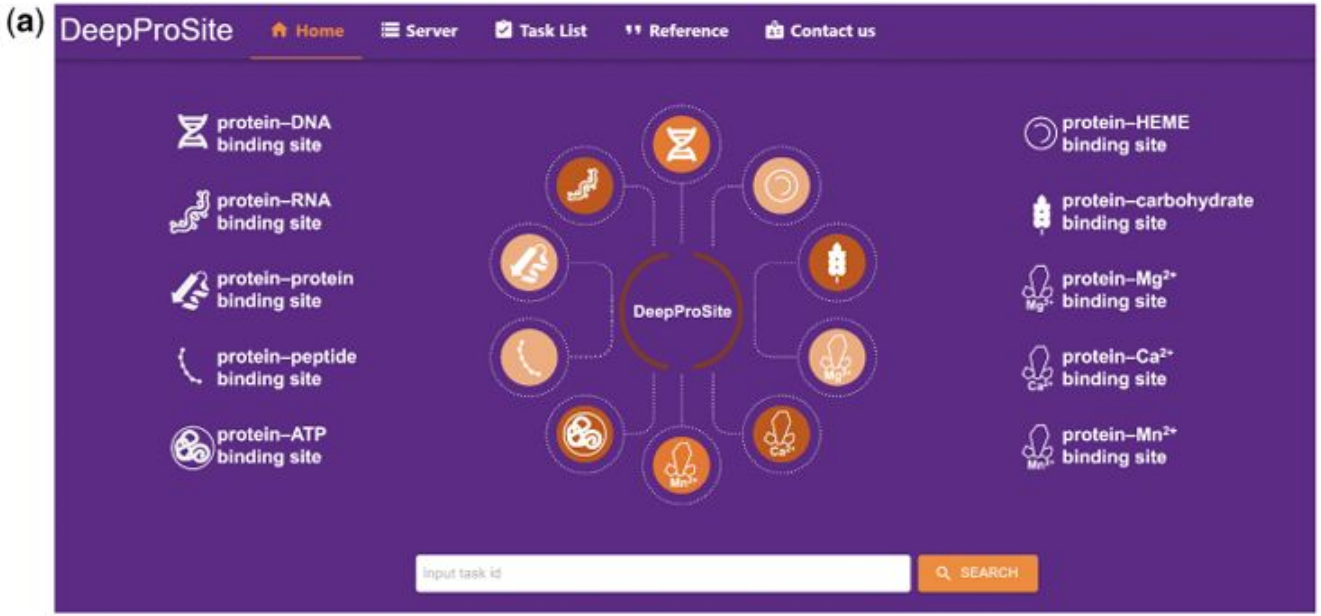
Tables are omitted in this slide because of their shapes.

(Supplementary Table S16, S17) (Table 7)

Results

Establishment of a webserver to facilitate the prediction of multiple types of binding sites

DeepProSite is open as a web service.



Conclusion

Proposed model, DeepProSite, have shown generalizability, robustness and performance throughout the experiments that have been conducted.

Importance of geometrical information and structure prediction quality was outlined.

It only requires sequence information.

References

Please refer to the article for the full list of references.

Thanks for listening