Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

GL-E2EATP: improving protein-ATP binding residue prediction using global and local embedding of protein language model

Bing Rao
Hangzhou City University
Jie Bai
Hangzhou City University
Maha A. Thafar
Taif University
Somayah Albaradei
King Abdulaziz University
Kamran Arshad
University of Management and Technology
Apilak Worachartcheewanh
Mahidol University
Muhammad Arif
mfarif@hbku.edu.qa
Hamad bin Khalifa University

Article

Keywords: Protein-ATP binding residue prediction, Bioinformatics, Protein language model, Multi-head attention

Posted Date: March 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9088034/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

GL-E2EATP: improving protein-ATP binding residue prediction using global and local embedding of protein language model

Bing Rao[1], Jie Bai[1], Maha A. Thafar[2], Somayah Albaradei[3], Kamran Arshad[4, 5], Apilak Worachartcheewan[6, *] and Muhamad Arif[7, *]

[1] School of Information and Electrical Engineering, Hangzhou City University, Hangzhou, 310015, China.

[2] Department of Computer Science, College of Computers and Information Technology, Taif University, P.O.Box 110099, Taif, 21944, Saudi Arabia

[3] Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, 21589 Jeddah, Saudi Arabia.

[4]School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 200, China.

[5] Department of Software Engineering, School of System and Technology, University of Management and Technology C-II, Johar Town, Lahore, Pakistan.

[6] Department of Community Medical Technology, Faculty of Medical Technology, Mahidol University, Bangkok, 10700, Thailand.

[7] College of Science and Engineering, Hamad Bin Khalifa University, Doha 34110, Qatar.

* mfarif@hbku.edu.qa (M.A), apilak.woa@mahidol.edu (A.W) are the corresponding author of this work.

ABSTRACT

Identification of ATP-binding residues in proteins is of paramount importance for elucidating the mechanisms underlying protein functions and advancing drug discovery efforts. Several computational approaches have been developed for predicting ATP binding sites, yet their predictive performance remains suboptimal, primarily due to inadequate features descriptors and learning models. In this study, we developed a novel end-to-end deep learning (DL) model called GL-E2EATP, for predicting ATP-binding residues with improved accuracy. The proposed model use self-supervise learning strategy by extracting both global and local embeddings from protein language model generated by protein sequences. Specifically, we leverage a pre-trained DL-based biological language model, ESM2, to autonomously generate biologically relevant features. Building upon ESM2, two different neural network modules, i.e., convolutional layers and multi-head attention layers, are employed to separately extract global information for whole protein sequences and local information for the potential ATP-binding residues. Empirical evaluations conducted on two independent test datasets reveal that GL-E2EATP outperforms existing ATP-based prediction methods, achieving superior Matthews correlation coefficient (MCC), area under the ROC curve (AUC), and area under the Precision-Recall curve (AUCPR) metrics. Comprehensive analyses anticipate that GL-E2EATP will serve an efficient solution for characterizing large-scale prediction of ATP-binding sites from protein sequences. The standalone package for GL-E2EATP is downloadable at https://github.com/Robin8990/gl-e2eatp for academic use.

Keywords: Protein-ATP binding residue prediction, Bioinformatics, Protein language model, Multi-head attention

INTRODUCTION

Protein-adenosine triphosphate (ATP) interactions play a pivotal role in a diverse range of biological processes, including cellular signaling, metabolic pathways, and enzymatic regulation [1-3]. ATP-binding sites within proteins represent crucial targets for small-molecule drug development, rendering their precise identification indispensable for elucidating protein functions and propelling advancements in drug discovery [4]. Despite extensive experimental endeavors aimed at characterizing protein-ATP interactions, which have yielded thousands of protein-ATP complex structures archived in the Protein Data Bank (PDB) [5], conventional wet-lab methodologies remain constrained by their time-intensive, laborious, and costly nature. This has resulted in a pronounced disparity between the number of annotated and unannotated proteins, highlighting the pressing need for the development of efficient and high-accuracy computational approaches to predict ATP-binding sites, particularly in light of the exponential growth of protein sequence data in the postgenomic era [6, 7].

Computational methods for predicting ATP-binding sites from broad spectrum can be classified into ligand-specific based and general-purpose based approaches. General-purpose approaches, likewise LIGSITE [8], ConCavity [9], FINDSITE [10], 3DLigandSite [11], and COACH [12], are designed to identify ligand-binding sites without considering the ligand type, but they predominantly rely on experimentally determined protein structures. This structural dependency significantly limits their applicability, as a substantial proportion of proteins lack experimentally resolved structural data. Recent advancements in protein structure prediction, particularly with tools like MODELLER [13], I-TASSER [14], Rosetta [15], AlphaFold2 [16, 17], and AlphaFold3 [18], have markedly enhanced the accuracy of protein structure modeling, partially mitigating this limitation. Nevertheless, many general-purpose methods were developed during the early stages of computational biology and have not undergone consistent updates, leading to suboptimal predictive performance compared to more modern methodologies. In contrast, ligand-specific methods [6, 19-23], such as ATPint [24], ATPsite [25], DELIA [26], and LigBind [27], are tailored to predict binding sites for specific ligands, such as ATP. These approaches are typically divided into sequential-based structured-based computational schemes. Structure-based methods, exemplified by DELIA, utilize protein structural information, including 2D distance matrices, in conjunction with advanced DL-based structures for example convolutional neural networks (CNNs) and bidirectional long short-term memory (BiLSTM) networks. However, their effectiveness is constrained by the limited availability of high-resolution 3D protein structures. On the other hand, sequence-based methods, such as ATPseq [6] and DeepATPseq [23], rely on features derived from

position-specific scoring matrices (PSSMs) and other sequence-based descriptors, offering broader applicability due to the relative ease of acquiring sequence data. Despite these advantages, sequence-based methods often face challenges in capturing high-latent discriminative features from protein sequences, which can limit their predictive accuracy.

Recently, to address these limitations, several protein language model (PLM)-based approaches have been proposed for predicting protein-ATP binding residues. For instance, E2EATP [28], developed by our research group, leverages a pretrained protein language model, ESM2, with an embedding dimension of 1280. Similarly, ATP_mCNN [29] employs the ProtTrans [30] PLM in conjunction with multi-scale sliding windows to extract local contextual information surrounding target residues. Another notable method, ATP-Deep [31], utilizes the ProtT5-XL-UniRef50 PLM, pretrained on a comprehensive dataset of 2 billion primary proteins, to advance the efficacy and precision of ATP-binding residue prediction.. Furthermore, ATP-Deep incorporates a co-attention mechanism to model complex interdependencies among elements within and across sequences. Despite these advancements, most existing PLM-based methods predominantly focus on local residue context information, often neglecting the global structural and functional context of entire proteins. This oversight imposes a significant limitation on further improvements in the prediction and characterizing of ATP-binding proteins solely from sequence information.

In the present study, we present GL-E2EATP, an innovative large language-based deep learning framework specifically designed to enhance ATP-binding residue prediction through comprehensive extraction of both global and local features from protein sequences. Our proposed GL-E2EATP architecture leverages the protein language model (PLM) ESM2 [32], which generates high-dimensional feature embeddings (5120 dimensions) for protein sequence representation. Building upon the output of ESM2, we have developed an end-to-end neural network architecture comprising two specialized modules for global and local feature extraction, respectively, to construct an accurate sequence-driven predictor for ATP-binding proteins. The global embedding module incorporates a residual network architecture to capture discriminative global patterns across entire protein sequences. The local embedding module integrates two advanced attention mechanisms – multi-head attention and out-product-based attention sub-modules – to thoroughly exploit contextual information surrounding target residues. Extensive benchmarking on two independent test datasets demonstrates that GL-E2EATP achieves superior performance, with an average accuracy of 0.973 and coverage of 61.9% of all ATP-binding sites, while significantly outperforming existing state-of-the-art sequence-based methods in terms of Matthews correlation coefficient (MCC). To support the research community and enhance the promote reproducibility,

the standalone protocol of GL-E2EATP method is freely accessible at https://github.com/jun-csbio/gl-e2eatp/.

METHODS

Benchmark Data Sets

In this study, we employ three training datasets (PATP-388, PATP-1930, and PATP-NW30-TRAIN) and two independent testing datasets (PATP-TEST and PATP-NW30-TEST) to rigorously examine the proposed GL-E2EATP method, against the existing approaches. The PATP-388 and PATP-TEST datasets were derived from an initial collection of 2,144 ATP-binding protein chains with clear annotations, sourced from the Protein Data Bank (PDB) prior to November 5, 2016 [6]. To ensure non-redundancy, the CD-HIT software [33] was utilized to remove protein pairs with sequence identities exceeding 40%. The resulting PATP-388 dataset comprises 388 protein sequences, containing 5,657 ATP-binding sites and 142,086 non-ATP-binding sites, conversely, the PATP-TEST dataset includes 41 protein sequences encompassing 674 ATP-binding residues and 14,159 non-ATP-binding residues. To further enhance the robustness and generalizability of our evaluation, we constructed an expanded training dataset, PATP-1930, by aggregating all available protein-ATP complex structures deposited in PDB repository up to 23rd of June, 2023. We applied the NW-align algorithm to ensure that all PATP-1930 sequences exhibit ≤40% sequence identity relative to those in PATP-TEST. The curated PATP-1930 dataset includes 1,930 proteins with 32,695 ATP-binding and 709,747 non-binding residues. Additionally, to assess the performance of our method under stricter sequence identity constraints, we utilized our previous datasets, i.e., PATP-NW30-TRAIN and PATP-NW30-TEST, which were constructed using a 30% sequence identity threshold. These datasets provide a more challenging evaluation scenario, ensuring that the performance of our method is not biased by high sequence similarity between training and testing data. The PATP-NW30-TRAIN dataset comprises 1,861 proteins annotated with 30,550 ATP-binding and 813,587 non-ATP-binding residues. The PATP-NW30-TEST dataset comprises 103 proteins with 1,552 annotated ATP-binding residues and their respective non-binding. The source code datasets are publically accessible at GitHub repository https://github.com/jun-csbio/gl-e2eatp/.

Feature Generation

In this study, we utilize ESM2, a pre-trained biological large language model developed by Facebook Research [34], to efficiently capture the hidden enriched information of protein. The feature embeddings generated by ESM2 have been successfully applied across various bioinformatics research domains, including B-cell epitope identification [35], the screening of

antihypertensive peptides [36], and discrimination of ATP binding sites[28]. ESM2 employs a deep transformer-based architecture trained on large-scale peptide sequences, to capture the local and global evolutionary properties of amino acid from protein sequence. This generates enrich and robust embeddings even effective on small dataset [37].

To identify the optimal model configuration, we systematically evaluated two pre-trained models, i.e., esm2_t33_650M_UR50D and esm2_t48_15B_UR50D, available on the ESM2 GitHub repository (https://github.com/facebookresearch/esm) on three training datasets, PATP-388, PATP-1930, and PATP-NE30-TRAIN, through comprehensive five-fold cross-validation tests. Based on performance evaluation, the pre-trained model 'esm2_t48_15B_UR50D' was selected for feature extraction. For a protein sequence of length L, ESM2 generates an embedding feature matrix of dimensions L×5120, where each residue is represented by a 5120-dimensional feature vector. This high-dimensional embedding captures intricate sequence-level information, facilitating robust functional predictions.

Global embedding module

In this study, global protein information is utilized to enhance the prediction accuracy of protein-ATP binding residues. As illustrated in Figure 1, leveraging the ESM2 output embeddings, where each outputted embedding with dimensions $L\times 5120$ for a protein with L amino acids, we developed a neural network module to integrate convolutional layers with attention pooling mechanisms for producing a fixed length embedding vector. This vector encapsulates the global embedding information, thereby improving the predictive performance for identifying protein-ATP binding residues. Concretely, the global embedding module first employ one convolutional neural network (CNN) layer with kernel size of 3 and padding size of 1 to extracting the neighboring information round each residue and rescaling the embedding dimension from 5120 to 128. Then, a deep residual network (ResNet) composed of 48 residual basic blocks is used to extract more discriminative information of potential ATP-binding residues. In ResNet, the output of i-th residual basic block admits a representation of the form as x_i=$x_{i-1} + f(x_{i-1}, w_i)$, where x_{i-1} is the output of the (i-1)-th residual basic block and w_i includes the weights of two CNN layers and one batch normalization (BN) layer. The function of f denotes the operation in the residual block. Gradients can flow smoothly from deeper to shallower layers by adding an identity shortcut, which makes the training of extremely deep residual neural networks possible. After ResNet, we employ an multi-head attention pooling layer following a linear layer to embed all global embedding information to

a fixed-length feature embedding with dimension of 17×128. The final dimension of global embedding of each protein is 17×128.

Local embedding module

In the current study, we aim to develop a local embedding neural network module (see Figure 1) with several core residual blocks (CRBs) to extract the local discriminative information for improving protein-ATP binding residue prediction performance. Each CRB mainly contains two parallel attention networks with multi-head attention neural network (MHA) and a newly designed out-product attention neural network (OPA), which can help us to extract the context information from a local sliding window centered as the *i*-th residue. Specifically, for each potential ATP-binding residue, we employ a sliding window with size of *wsize* (*wsize*=17 in this study) to extract the ESM2 embedding of the center residue (e.g., the *i*-th residue). The dimension of the ESM2 embedding of the *i*-th residue is 17×5120. Based on the local ESM2 embedding, we first employ one linear layer to squeeze the dimension of the embedding of each context residue from 5120 to 128. Then, the squeezed embedding (denoted as e_0) with dimension of 17×128 is separately inputted to two different attention neural networks (i.e., MHA and OPA) of the first CRB. In MHA, one multi-head attention layer with 8 heads is used to extract the discriminative information from the 17 context residues. Subsequently, the GeLU activation function and the layer normalization operation are employed to improve the non-linear learning and generalization abilities. The final outputted embedding of MHA is denoted as e_{MHA}. In OPA, two different linear layers are separately used to transform the inputted context embedding e_0 to e_1 and e_2 with the same dimensions of 17 ×128. Based e_1, the out-product operator is then employed to generate a new embedding e_3 with dimension of 17×17×128 to represent the relationship between each pair of the context residues. To fuse the information of e_2 and e_3, the matrix multiplication operator is utilized to generate the embedding e_4 with dimension of 17×128. The GeLU activation function and the layer normalization operation are also employed to generate the final embedding (i.e., e_{OPA}) of OPA. Finally, the outputted embedding of CRB is the embedding $e_{\mathrm{CRB}}= e_0+ e_{\mathrm{MHA}}+ e_{\mathrm{OPA}}$. This outputted embedding is subsequently inputted to the next CRB. The embedding of the last CRB is transformed by a linear layer finally.

Architecture of GL-E2EATP

Figure 1 illustrates the architecture of the proposed GL-E2EATP for protein-ATP binding residue prediction. As shown in Figure 1, in the model of GL-E2EATP, for a protein to be predicted, the model first extract feature embedding of each residue by calling the pretrained protein language

model ESM2 [32] and applying the local sliding window technique; then, the global and local embedding neural network modules are separately employed to extract the discriminative feature information from two different feature views; the outputted global and local embeddings are serially combined as the final feature embedding and fed to multilayer perceptron (MLP) neural network for predicting the probability value of a residue belonging to the class of ATP-binding residues. Finally, a fine-tuned threshold is utilized to determine whether the residue is an ATP-binding residue.

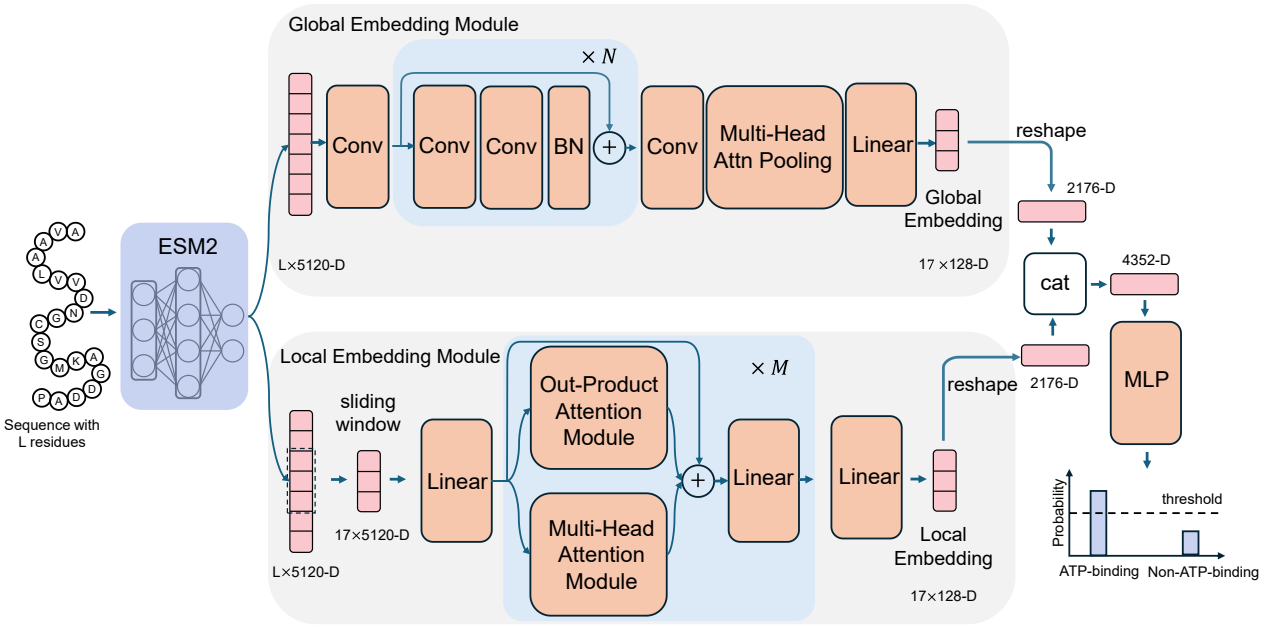


Figure 1. Architecture of GL-E2EATP.

Cross-Entropy Loss Function

In classification tasks, such as the prediction of ATP-binding proteins, the Cross-Entropy loss function is a fundamental concept in deep learning field. It quantifies the difference between two probability distributions: the ground-truth distribution and the prediction distribution generated by the deep learning model. Mathematically, for a classification task with C classes, e.g., ATP-binding residue prediction, the Cross-Entropy loss value L is defined as:

$$L = -\sum_{i=1}^{C} y_i \log(p_i) \quad (1)$$

where y_i represents the true label and p_i denotes the predicted probability for the i-th class. The loss function penalizes deviations between the predicted probabilities and the actual labels, encouraging the model to assign higher probabilities to the correct classes. In this study, we utilize the Cross-Entropy loss function to train the ATP-binding residue prediction models on different training data sets. To maximize the effectiveness of the Cross-Entropy loss during training, we ensure that each batch of data contains a balanced number of positive and negative samples, thereby

mitigating potential biases and enhancing model performance.

Implementation of GL-E2EATP

In this study, the proposed GL-E2EATP model for targeting ATP-binding residue was implemented using the PyTorch framework (version 1.13.1) and trained for 500 epochs on the PATP-388, PATP-1930, and PATP-NW30-TRAIN datasets. During the training stage, the values of the hyper-parameters were set to 10, 1e-5, and 0.3 for batch size, learning rate and dropout rate. The optimizer of Adam was employed to minimize the above cross-entropy loss. All the experiments were performed over a high-performance computing cluster equipped with four NVIDIA A100-SXM4-80GB GPUs.

Evaluation Indexes

Six threshold-dependent performance measures metrics, i.e., Matthew's correlation coefficient (*MCC*), Precision (*Pre*), F1-score (F_1), Sensitivity (*Sen*), Specificity (*Spe*), and Accuracy (*Acc*), are adopted to evaluate and compare the proposed GL-E2EATP method with existing ATP-binding residue prediction methods:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FN) \cdot (TP+FP) \cdot (TN+FN) \cdot (TN+FP)}} \quad (2)$$

$$Pre = \frac{TP}{TP+FP} \quad (3)$$

$$F_1 = \frac{2 \cdot Sen \cdot Pre}{Sen+Pre} \quad (4)$$

$$Sen = \frac{TP}{TP+FN} \quad (5)$$

$$Spe = \frac{TN}{TN+FP} \quad (6)$$

$$Acc = \frac{TN+TP}{TN+TP+FN+FP} \quad (7)$$

where in equation (2-7) TP denotes the true ATP-binding residues and FN denotes the incorrect ATP-binding residues,; TN denotes the correct non-ATP-binding residues and FP denotes the incorrect non-ATP-binding residues. In addition, two threshold-independent evaluation indexes, i.e., ROC curve and AUCPR, are employed to measure our proposed method.

EXPERIMENTAL RESULTS AND ANALYSIS

Performance comparison between different sliding window sizes in local embedding module

In this section, we compare the performance of different values of sliding window sizes (*wsize*) in the local embedding module on all three training data sets, i.e., PATP-388, PATP-1930, and PATP-NW30-TRAIN, over five-fold cross-validation tests. Concretely, we varied the value of *wsize* from 1 to 21 with a step size of 4. Table 1 demonstrates the performance comparison of *AUC, AUCPR, MCC, Pre, F1, Sen, Spe,* and *Acc* values of different *wsize* on the above three training data sets. Note that, in this section, when we train the prediction models with different *wsize*, the global embedding module will be ignored for training, resulting to the model only reflect the affection of different *wsize* in the local embedding module. To save the training time, the pretrained ESM2 model of 'esm2_t33_650M_UR50D' is employed to generate the protein embeddings and the model of each fold of cross-validations is trained with 100 epochs.

Table 1. Performance comparison between different sliding window sizes in local embedding module on PATP-388, PATP-1930, and PATP-NW30-TRAIN over five-fold cross-validation tests

Data set	*wsize*	AUC	AUCPR	*MCC*	*Pre*	F_1	*Sen*	*Spe*	*Acc*
PATP-388	1	0.923	0.538	0.540	0.569	0.557	0.545	0.984	0.967
	5	0.932	0.581	0.575	0.585	0.591	0.598	0.983	0.968
	9	0.932	0.590	0.583	0.607	0.599	0.591	0.985	0.970
	13	0.932	0.598	0.590	0.623	0.605	0.589	0.986	0.971
	17	0.934	0.602	0.592	0.625	0.607	0.590	0.986	0.971
	21	0.931	0.588	0.588	0.602	0.604	0.605	0.984	0.970
PATP-1930	1	0.989	0.883	0.834	0.802	0.841	0.884	0.990	0.985
	5	0.989	0.887	0.837	0.820	0.844	0.869	0.991	0.986
	9	0.989	0.887	0.839	0.814	0.846	0.880	0.991	0.986
	13	0.988	0.886	0.837	0.817	0.844	0.874	0.991	0.986
	17	0.989	0.888	0.841	0.816	0.847	0.881	0.991	0.986
	21	0.989	0.887	0.841	0.816	0.847	0.881	0.991	0.986
PATP-NW30-TRAIN	1	0.987	0.859	0.817	0.783	0.823	0.867	0.991	0.986
	5	0.988	0.861	0.822	0.797	0.828	0.862	0.992	0.987
	9	0.988	0.862	0.822	0.796	0.828	0.863	0.992	0.987
	13	0.988	0.862	0.822	0.797	0.828	0.862	0.992	0.987
	17	0.989	0.865	0.824	0.794	0.830	0.868	0.992	0.987
	21	0.988	0.860	0.822	0.799	0.828	0.859	0.992	0.987

From Table 1, we can find that the average *AUC*, *AUCPR*, *MCC*, and F_1 values of *wisze*=17 are 0.971, 0.785, 0.752, and 0.761 on the three training data sets over five-fold cross-validation tests, which are consistently higher than those of other *wsize* values, with comparable average values of *Sen*, *Spe*, and *Acc*. Taking results on PATP-388 as one example, the *AUCPR*, *MCC*, *Pre*, and F_1

values of *wsize*=17 are 0.602, 0.592, 0.625, and 0.607, which are 11.90, 9.63, 9.84, and 8.98 percent higher than those of *wsize*=1 and 0.67, 0.34, 0.32, and 0.33 percent higher than those of the second-best wsize (i.e., *wsize*=13). In addition, the values of *AUCPR*, *MCC*, and F_1 tend to increase with the increase in *wsize* when *wsize*<17 and these values reach the highest value when *wsize*=17 on PATP-388 over five-fold cross-validation tests. The similar trends can be also found on PAT-1930 and PATP-NW30-TRAIN. The final value of *wsize* in the local embedding module is set to 17 in the following experiments.

Performance comparison between with- and without-global embedding module

To evaluate the usefulness of global embedding module (GEM), in this section, we compare the protein-ATP binding residue prediction performance between with- and without-GEM on three training data sets, i.e., PATP-388, PATP-1930, and PATP-NW30-TRAIN, over five-fold cross-validation tests. For the without-GEM configuration, prediction models are trained using the GL-E2EATP deep neural network architecture with all learnable parameters within the GEM component fixed at zero values. Conversely, the with-GEM models were trained utilizing the complete GL-E2EATP neural network architecture. Protein sequence embeddings were derived using the pretrained ESM2 model of 'esm2_t33_650M_UR50D', with each cross-validation fold undergoing 100 training epochs. Table 2 illustrates the comparison results between with- and without-GEM on three training data sets.

Table 2. Performance Comparison between with- and without-global embedding module on PATP-388, PATP-1930, and PATP-NW30-TRAIN over five-fold cross-validation tests

Data set	With-/without-GEM[a]	AUC	AUCPR	*MCC*	*Pre*	F_1	*Sen*	*Spe*	*Acc*
PATP-388	Without-	0.934	0.602	0.592	0.625	0.607	0.590	0.986	0.971
	With-	0.934	0.602	0.603	0.642	0.617	0.594	0.987	0.972
PATP-1930	Without-	0.989	0.888	0.841	0.816	0.847	0.881	0.991	0.986
	With-	0.991	0.887	0.843	0.826	0.850	0.874	0.992	0.986
PATP-NW30-TRAIN	Without-	0.989	0.865	0.824	0.794	0.830	0.868	0.992	0.987
	With-	0.990	0.867	0.828	0.807	0.834	0.862	0.992	0.988

[a] GEM denotes global embedding module

By observing Table 2, it is easy to find that with-GEM consistently outperforms without-GEM concerning the evaluation indexes of *MCC*, *Pre*, F_1 on all three training data sets, i.e., PATP-388, PATP-1930, and PATP-NW30-TRAIN, over five-fold cross-validation tests. Taking results on PATP-388 as an example, the values of *MCC*, *Pre*, F_1, *Sen*, *Spe*, and *Acc* of with-GEM are 0.603, 0.642, 0.617, 0.594, 0.987, and 0.972, which are approximately 1.86, 2.72, 1.65, 0.68, 0.10, and

1.03 percent higher than that of without-GEM, respectively. On PATP-1930 and PATP-NW30-TRAIN, we can also find that the models of with-GEM outperform those of without-GEM concerning the evaluation indexes of *AUC*, *MCC*, *Pre*, and F_1. Through an in-depth comparative analysis of the data presented in Table 2, we observed that the performance improvement achieved by the with-GEM model is significantly more pronounced on the PATP-388 dataset compared to the PATP-1930 and PATP-NW30-TRAIN datasets. This phenomenon can be potentially attributed to the relatively smaller protein size of PATP-388, wherein global information extracted by GEM plays a more crucial role in predictive performance. These findings suggest that the incorporation of global information through GEM can substantially enhance the predictive capability of protein-ATP binding residues, particularly in scenarios characterized by limited training data availability.

Performance comparison between two pretrained ESM2 models of esm2_t33_650M_UR50D and esm2_t48_15B_UR50D

In this section, we will compare the performance of ATP-binding residues prediction between two pretrained ESM2 models, i.e., 'esm2_t33_650M_UR50D' and 'esm2_t48_15B_UR50D', on three training data sets (PATP-388, PATP-1930, and PATP-NW30-TRAIN) over five-fold cross-validation tests. During the cross-validation tests, each model is trained with 100 epochs using the entire GL-E2EATP neural network architecture. The comparison results between two different pretrained ESM2 models are summarized in Table 3.

Table 3. Performance Comparison of different pretrained models of ESM2 on PATP-388, PATP-1930, and PATP-NW30-TRAIN over five-fold cross-validation tests

Data set	ESM2 Model	AUC	AUCPR	*MCC*	*Pre*	F_1	*Sen*	*Spe*	*Acc*
PATP-388	esm2_t33_650M_UR50D	0.934	0.602	0.603	0.642	0.617	0.594	0.987	0.972
	esm2_t48_15B_UR50D	0.935	0.609	0.607	0.686	0.617	0.561	0.990	0.973
PATP-1930	esm2_t33_650M_UR50D	0.991	0.887	0.843	0.826	0.850	0.874	0.992	0.986
	esm2_t48_15B_UR50D	0.991	0.888	0.847	0.830	0.854	0.878	0.992	0.987
PATP-NW30-TRAIN	esm2_t33_650M_UR50D	0.990	0.867	0.828	0.807	0.834	0.862	0.992	0.988
	esm2_t48_15B_UR50D	0.990	0.868	0.830	0.806	0.836	0.868	0.992	0.988

From Table 3, it is easy to observe that the *AUC*, *AUCPR*, *MCC*, *Pre*, F_1, *Sen*, *Spe*, and *Acc* values for 'esm2_t48_15B_UR50D' are 0.935, 0.609, 0.607, 0.686, 0.617, 0.561, 0.990, and 0.973 on PATP-388, 0.991, 0.888, 0.847, 0.830, 0.854, 0.878, 0.992, and 0.987 on PATP-1930, and 0.990, 0.868, 0.830, 0.806, 0.836, 0.868, 0.992, and 0.988 on PATP-NW-TRAIN, respectively. The *AUCPR* and *MCC* values of 'esm2_t48_15B_UR50D' are consistently higher than that of 'esm2_t33_650M_UR50D' on all three data sets. Taking PATP-388 as an example, the *AUCPR* and

MCC of ‘esm2_t48_15B_UR50D’ are 1.16 and 0.66 percent higher than that of ‘esm2_t33_650M_UR50D’, respectively. These data demonstrate the similar conclusion in the article of ESM2 [32], that is, the pretrained model of ‘esm2_t48_15B_UR50D’ with more parameters achieves higher performance than ‘esm2_t33_650M_UR50D’ on most of the downstream tasks, also holds true for the task of protein-ATP binding residue prediction in this study.

GL-E2EATP outperforms the existing sequence-based methods on PATP-TEST

In this section, we present a comprehensive evaluation of the predictive performance of GL-E2EATP by conducting a comparative analysis against established state-of-the-art sequence-based methods for protein-ATP binding residue prediction. The benchmark methods include NsitePred [38], TargetS [20], ATPseq [6], DELIA(seq) [26], DeepATPseq [23], E2EATP [28], ATP_mCNN [29], and ATP-Deep [31]. The comparative assessment was performed using the independent testing dataset PATP-TEST to ensure unbiased evaluation. It should be noted that DELIA(seq) represents the sequence-specific variant of DELIA, exclusively trained on sequence-derived features [26]. Similar to our previous work E2EATP, we also train two models, i.e., GL-E2EATP(388) and GL-E2EATP(1930), on PATP-388 and PATP-1930, enabling a systematic comparison of model performance across different training conditions. Table 4 lists the performance comparison between GL-E2EATP with other state-of-the-art sequence-based protein-ATP binding residues prediction methods.

As presented in Table 4, GL-E2EATP(1930) demonstrates superior performance metrics, achieving MCC of 0.679, Pre of 0.727, F1 of 0.692, Sen of 0.660, Spe of 0.988, and Acc of 0.973. Comparative analysis reveals that GL-E2EATP(1930) outperforms the best-performing control method, E2EATP(1930), with relative improvements of 1.65%, 2.25%, and 1.47% in MCC, Pre, and F1, respectively. When compared to the second-best control method, ATP-Deep, GL-E2EATP(1930) shows enhancements of 3.03%, 5.81%, and 23.13% in MCC, F1, and Sen, respectively. Notably, while ATP-Deep achieves the highest *Pre*, this comes at the cost of reduced *Sen*, suggesting a substantial misclassification of positive samples (ATP-binding residues) as negative class (non-ATP-binding residues).

Furthermore, comparative evaluation between GL-E2EATP(388) and E2EATP(388) indicates consistent improvements across multiple evaluation metrics, particularly in *MCC*, F_1, and *Sen*. It is noteworthy that both GL-E2EATP(1930) and E2EATP(1930) demonstrate superior performance to

their PATP-388-trained counterparts (GL-E2EATP(388) and E2EATP(388)) in terms of the comprehensive evaluation metrics *MCC* and F_1. This performance enhancement persists despite the fact that neither PATP-1930 nor PATP-388 shares >40% sequence identity with any proteins in PATP-TEST. This observation strongly suggests that increased training data volume significantly enhances model predictive capability in protein-ATP binding residue prediction tasks.

Table 4. Performance Comparison between GL-E2EATP and other sequence-based prediction methods on PATP-TEST.

Method	*MCC*	*Pre*	F_1	*Sen*	*Spe*	*Acc*
NsitePred [a]	0.456	0.492	0.480	0.467	0.977	0.954
TargetS [a]	0.580	0.689	0.590	0.516	0.989	0.965
ATPseq [a]	0.639	0.781	0.642	0.545	0.993	0.972
DELIA(seq) [b]	0.612	0.715	0.622	0.550	N/A	N/A
DeepATPseq [c]	0.655	0.777	0.659	0.572	0.992	0.973
E2EATP(388) [d]	0.649	0.728	0.660	0.604	0.989	0.972
E2EATP(1930) [d]	0.668	0.711	0.682	0.656	0.987	0.972
ATP_mCNN [e]	0.523	N/A	N/A	0.831	0.931	0.927
ATP-Deep [f]	0.659	0.839	0.654	0.536	0.995	0.974
GL-E2EATP(388)	0.660	0.703	0.674	0.648	0.987	0.972
GL-E2EATP(1930)	0.679	0.727	0.692	0.660	0.988	0.973

[a] Data excerpted from Ref. [6]. [b] Data excerpted from Ref. [26]. [c] Data excerpted from Ref. [23]. [d] Data excerpted from Ref. [28]. [e] Data excerpted from Ref. [29]. [f] Resulting computed using the standalone program of ATP-Deep downloaded at https://github.com/jerry1984Y/ATP-Deep.

Figure 2 illustrates the ROC and Precision-Recall (PR) curves of nine computational models, i.e., GL-E2EATP(388), GL-E2EATP(1930), ATP-Deep, E2EATP(388), E2EATP(1930), DeepATPseq, ATPseq, TargetS, and NsitePred, evaluated on the independent testing dataset (PATP-TEST). It should be noted that the prediction results of DELIA(seq) were unavailable due to the limitation of its web server, which exclusively provides 3D-structure-based prediction functionality. Furthermore, technical issues encountered during the installation process prevented the successful implementation and testing of the ATP_mCNN software package.

The comparative analysis reveals that both GL-E2EATP(1930) and GL-E2EATP(388) demonstrate superior performance over the six existing sequence-based methods in terms of the evaluation index of AUC. Specifically, GL-E2EATP(1930) achieves an AUC value of 0.966, representing significant improvements of 1.79%, 5.69%, 5.81%, 10.40%, 10.02%, 10.78%, and 13.38% over ATP-Deep, E2EATP(1930), E2EATP(388), DeepATPseq, ATPseq, TargetS, and NsitePred, respectively. Regarding the Area Under the Precision-Recall Curve (AUCPR), GL-E2EATP(1930) attains the

second-highest value of 0.688, marginally lower than ATP-Deep (0.696). Detailed examination of Figure 2(B) reveals an interesting performance pattern: GL-E2EATP maintains higher precision values compared to ATP-Deep when Recall values are below 0.78, while exhibiting slightly lower precision than ATP-Deep when Recall exceeds 0.78. This observation suggests that GL-E2EATP demonstrates superior predictive accuracy in the lower recall range, whereas ATP-Deep performs better at higher recall levels.

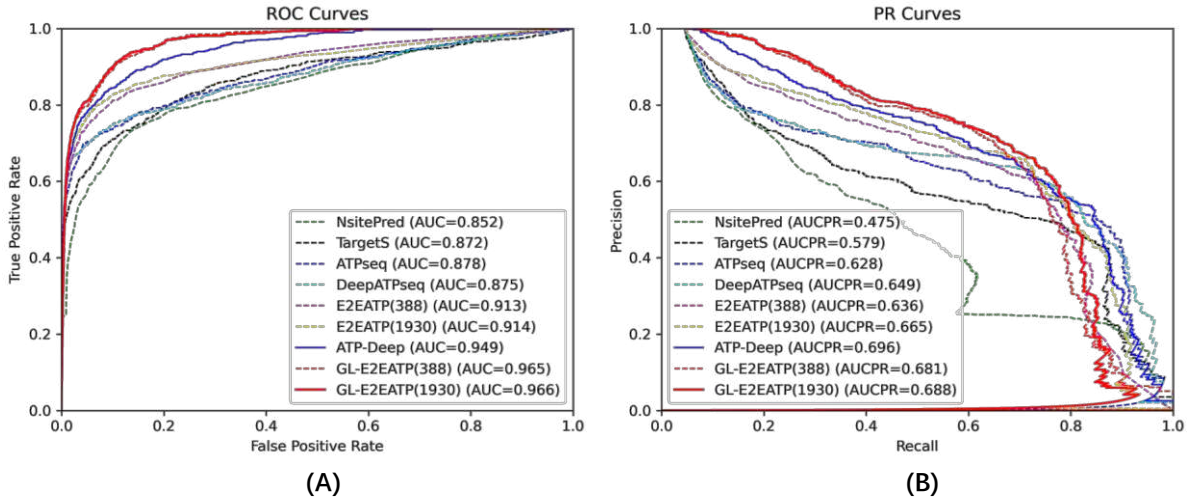


Fig. 2. ROC and Precision-Recall (PR) curves of GL-E2EATP (388 and 1930), E2EATP (388 and 1930), DeepATPseq, ATPseq, TargetS, and NsitePred on the independent testing data set (PATP-TEST). "388" and "1930" indicate the training data sets, i.e., PATP-388 and PATP-399, respectively. (A) ROC curves and (B) PR curves.

GL-E2EATP outperforms the existing structure-based methods on PATP-TEST

To comprehensively assess the predictive capability of GL-E2EATP, we conducted a comparative performance analysis between GL-E2EATP(1930) and four state-of-the-art structure-based methods for protein-ATP binding residue prediction: COACH, ATPbind, DELIA, and LigBind. This evaluation is performed using the independent PATP-TEST dataset. To ensure a fair comparison between these structure-based approaches and our sequence-based GL-E2EATP method, we implemented a standardized protocol wherein three-dimensional structural information is generated for each input protein using AlphaFold2 [16], rather than utilizing experimentally determined structures from the wet-lab experiments. The comparative performance metrics of GL-E2EATP, ATPbind, DELIA, and LigBind on the PATP-TEST dataset are systematically presented in Table 5.

Table 5. Performance Comparison between GL-E2EATP and four structure-based prediction methods, COACH, ATPbind, DELIA, and LigBind, on the independent testing data set PATP-TEST.

Method	*MCC*	*Pre*	F_1	*Sen*	*Spe*	*Acc*

COACH [a]	0.639	0.627	0.655	0.687	0.981	0.967
ATPbind [b]	0.664	0.776	0.671	0.591	0.992	0.974
DELIA [c]	0.654	0.716	0.667	0.625	0.988	0.972
LigBind [d]	0.667	0.694	0.682	0.670	0.986	0.971
GL-E2EATP	0.679	0.727	0.692	0.660	0.988	0.973

[a] Resulting computed using the standalone program of COACH downloaded at https://zhanggroup.org/COACH/. [b] Resulting computed using the standalone program of ATPbind downloaded at https://github.com/jun-csbio/ATPbind/. [c] Resulting computed using the webserver of DELIA at http://www.csbio.sjtu.edu.cn/bioinf/delia/. [d] Data excerpted from Ref [31].

Based on the performance metrics presented in Table 5, GL-E2EATP demonstrates superior performance compared to the four structure-based prediction methods, i.e., COACH, ATPbind, DELIA, and LigBind, on the independent testing dataset PATP-TEST. Specifically, GL-E2EATP achieves the highest MCC of 0.679, indicating a better overall balance between true positives, true negatives, false positives, and false negatives. In terms of Pre, GL-E2EATP (0.727) outperforms COACH (0.639) and LigBind (0.667). Although ATPbind shows a higher *Pre* of 0.776. However, GL-E2EATP compensates with a higher F_1 score of 0.692, which reflects a better harmonic mean of precision and recall, suggesting a more robust performance in identifying true positives while minimizing false negatives. Additionally, GL-E2EATP exhibits competitive Sen at 0.660, which is higher than ATPbind and DELIA, indicating a better ability to correctly identify positive instances. The Spe of GL-E2EATP is 0.988, comparable to ATPbind and DELIA, and slightly higher than COACH and LigBind. Finally, GL-E2EATP achieves an accuracy Acc of 0.973, which is marginally higher than COACH, DELIA, and LigBind, and very close to ATPbind (0.974). These findings demonstrate that GL-E2EATP could be one of the robust alternatives to the existing structure-based methods.

Comparisons of GL-E2EATP with existing methods on PATP-NW30-TEST

Table 6 lists the comparison results between GL-E2EATP and the state-of-the-art existing protein-ATP binding residue prediction methods, i.e., DeepATPseq, ATPseq, COACH, ATPbind, DELIA, E2EATP, ATP_mCNN, ATP-Deep, and LigBind, on the new independent testing data set, i.e., PATP-NW30-TEST. To make a fair comparison, we employ the training set of PATP-NW30-TRAIN, which have no protein with >40% sequence identity against any proteins in PATP-NW30-TEST, to train the prediction model of GL-E2EATP in this section. The performance of the control structure-based methods are gained using the AlphFold2-predicted structure as input.

Table 6. Performance Comparison between GL-E2EATP and the existing methods, i.e., DeepATPseq, ATPseq, COACH, ATPbind, DELIA, ATP_mCNN, and ATP-Deep, on the independent testing data set PATP-NW30-TEST.

Method	*MCC*	*Pre*	F_1	*Sen*	*Spe*	*Acc*
ATPseq [a]	0.425	0.581	0.424	0.334	0.991	0.968
DELIA [b]	0.501	0.598	0.511	0.446	0.989	0.970
DeepATPseq [c]	0.510	0.758	0.486	0.358	0.996	0.973
COACH [d]	0.525	0.529	0.542	0.557	0.982	0.967
ATPbind [a]	0.553	0.739	0.546	0.432	0.994	0.974
ATP_mCNN [e]	0.562	N/A	N/A	0.465	0.993	0.974
E2EATP [f]	0.566	0.662	0.574	0.506	0.991	0.973
ATP-Deep [g]	0.583	0.621	0.597	0.575	0.987	0.972
GL-E2EATP	0.588	0.628	0.602	0.577	0.987	0.973

[a] Resulting computed using the standalone program of ATPbind downloaded at https://github.com/jun-csbio/ATPbind/. [b] Resulting computed using the webserver of DELIA at http://www.csbio.sjtu.edu.cn/bioinf/delia/. [c] Resulting computed using the standalone program of DeepATPseq downloaded at https://github.com/Zlinlin211/DeepATPseq. [d] Resulting computed using the standalone program of COACH downloaded at https://zhanggroup.org/COACH/. [e] Data excerpted from Ref. [29]. [f] Data excerpted from Ref. [28]. [g] Resulting computed using the standalone package downloaded at https://github.com/jerry1984Y/ATP-Deep.

From Table 6, it is easy to find that GL-E2EATP could achieve the highest *MCC*, F_1, and *Sen* values among the nine comparison prediction methods. Specifically, the *MCC*, F_1, and *Sen* values of GL-E2EATP are 0.588, 0.602, and 0.577, which are 0.86%, 0.84%, and 0.35% higher than that of ATP-Deep, 3.89%, 4.88%, and 14.03% higher than that of E2EATP, 17.37%, 5.02%, and 29.37% higher than that of DELIA, 6.33%, 10.26%, and 33.56% higher than that of ATPbind, 12.00%, 11.07%, and 3.59% higher than that of COACH, 28.35%, 41.98%, and 72.75% higher than that of ATPseq, and 15.29%, 23.87%, and 61.17% higher than that of DeepATPseq, respectively. Compared to ATP-mCNN, GL-E2EATP achieves 4.63% and 24.09% improvements of *MCC* and *Sen*. These findings demonstrate that GL-E2EATP an higher overall performance on PATP-NW30-TRAIN.

Figure 3 demonstrates the ROC and Precision-Recall curves of GL-E2EATP, ATP-Deep, E2EATP, DeepATPseq, ATPseq, COACH, ATPbind, and DELIA on the independent testing data set PATP-NW30-TEST. From Figure 3, we can observe that GL-E2EATP could achieve the highest AUC value (0.938) and the second-highest AUCRP value (0.597). Concretely, the AUC and AUCPR values of GL-E2EATP are 3.42% and 1.53% higher than that of E2EATP, 6.96% and 15.70% higher than that of ATPbind, 17.69% and 6.04% higher than that of COACH, 16.96% and 28.11% higher than that of DeepATPseq, 3.30% and 19.88% higher than that of DELIA, and 13.15% and 55.06% higher than that of ATPseq, respectively. Compared to ATP-Deep, GL-E2EATP achieves an improvement of AUC of 0.86%, although it obtains a slightly lower AUCPR value.

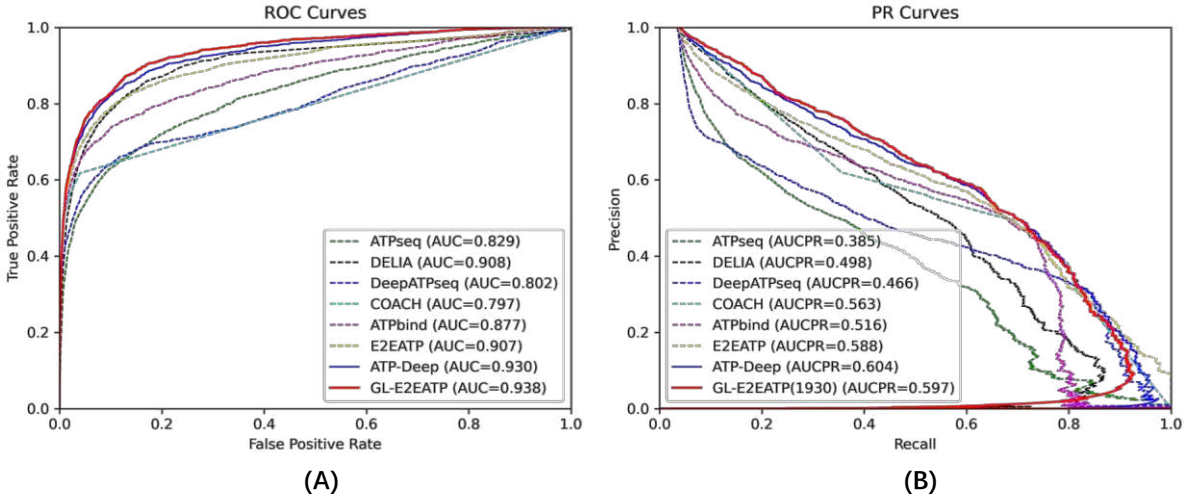


Fig.3. ROC and Precision-Recall (PR) curves of GL-E2EATP, ATP-Deep, E2EATP, ATPbind, COACH, DeepATPseq, DELIA and ATPseq on the independent testing data set PATP-NW30-TEST. (A) ROC curves and (B) PR curves.

CONCLUSIONS

Accurate identification of ATP-binding sites represents a fundamental step in elucidating the molecular mechanisms underlying protein-ATP interactions, which is essential for advancing our understanding of diverse biological processes and cellular functions. In this study, we present GL-E2EATP, a novel deep learning framework designed to enhance the prediction accuracy of protein-ATP binding residues. The proposed method integrates a pretrained protein language model, ESM2, to generate comprehensive protein sequence embeddings, coupled with a sophisticated neural network architecture that combines convolutional layers and multi-head attention mechanisms for feature extraction and model optimization. Extensive evaluations on independent benchmark datasets demonstrate that GL-E2EATP achieves superior performance compared to existing state-of-the-art methods, including both sequence-based and predicted structure-based approaches for ATP-binding residue prediction. To facilitate further research in this field, we have made the standalone implementation of GL-E2EATP freely available for academic use at https://github.com/Robin8990/gl-e2eatp/.

In our future research, to further advance the performance of ATP-binding residue prediction, we intend to pursue the following strategic directions: (1) the development of an advanced graph neural network architecture to extract more discriminative structural features from protein structures, thereby enhancing the precision of structure-based predictive modeling; and (2) the implementation of a deep transfer learning framework to facilitate knowledge transfer from protein residues interacting with other nucleotides, including but not limited to ADP and GTP. While GL-E2EATP

demonstrates significant predictive capabilities, there remains potential for further refinement. Nevertheless, it currently represents one of the most accurate computational frameworks for sequence-based protein-ATP residue prediction, leveraging state-of-the-art protein language models and deep learning methodologies.

DATA AND SOFTWARE AVAILABILITY

The data and code are freely accessible at https://github.com/Robin8990/gl-e2eatp for academic use.

AUTHOR INFORMATION

Corresponding Author

* Email: mfarif@hbku.edu.qa (M.A), apilak.woa@mahidol.edu(A.W),

Author Contributions

B.R., J.B., conceived the experiment(s) and manuscript writing. M.A.T., S.A., performed data collection and formal analysis and validation, K.A, conducted Initial review and editing the draft, statistical analysis and figure generation, M.A, conceptualization, project management, supervision, writing review and editing, and A.W formal analysis, funding.

ORCID

Apilak Worachartcheewan: 0000-0003-3021-3632

Muhammad Arif: 0000-0003-3950-6618

Declaration

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

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

ACKNOWLEDGEMENTS

This work was supported by the National Key R&D Program of China (No. 2023YFB3209803), the National Natural Science Foundation of China (No. 61902352), Natural Science Foundation of Zhejiang (No. LY21F020025) and the Research Foundation of Hangzhou City University (No. X-202302), Mahidol University (Fundamental Fund: fiscal year 2025 by National Science Research and Innovation Fund (NSRF)).The researchers would like to acknowledge Deanship of Scientific

Research, Taif University for funding this work.

Funding

This research was funded by Mahidol University (Fundamental Fund: fiscal year 2026 by National Science Research and Innovation Fund (NSRF)).

REFERENCES

[1] M. J. Haas, and A. D. Mooradian, "Potential Therapeutic Agents That Target ATP Binding Cassette A1 (ABCA1) Gene Expression," *Drugs,* vol. 82, no. 10, pp. 1055-1075, 2022/07/01, 2022.

[2] B. Sun, T. Liu, M. Zhang, and S. Li, "Synergistic Binding of ATP and Nucleic Acids Necessitates UPF1's ATPase Functional Cycle," *Journal of Chemical Information and Modeling,* vol. 63, no. 11, pp. 3474-3485, 2023/06/12, 2023.

[3] G. A. Wyant, W. Yu, I. P. Doulamis, R. S. Nomoto, M. Y. Saeed, T. Duignan, J. D. McCully, and W. G. Kaelin, "Mitochondrial remodeling and ischemic protection by G protein-coupled receptor 35 agonists," *Science,* vol. 377, no. 6606, pp. 621-629, 2022.

[4] R. Mo, T. Jiang, R. DiSanto, W. Tai, and Z. Gu, "ATP-triggered anticancer drug delivery," *Nature Communications,* vol. 5, no. 1, pp. 3364, 2014/03/11, 2014.

[5] S. K. Burley, C. Bhikadiya, C. Bi, S. Bittrich, H. Chao, L. Chen, P. A. Craig, G. V. Crichlow, K. Dalenberg, J. M. Duarte, S. Dutta, M. Fayazi, Z. Feng, J. W. Flatt, S. Ganesan, S. Ghosh, D. S. Goodsell, R. K. Green, V. Guranovic, J. Henry, B. P. Hudson, I. Khokhriakov, C. L. Lawson, Y. Liang, R. Lowe, E. Peisach, I. Persikova, D. W. Piehl, Y. Rose, A. Sali, J. Segura, M. Sekharan, C. Shao, B. Vallat, M. Voigt, B. Webb, J. D. Westbrook, S. Whetstone, J. Y. Young, A. Zalevsky, and C. Zardecki, "RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning," *Nucleic Acids Research,* vol. 51, no. D1, pp. D488-D508, 2022.

[6] J. Hu, Y. Li, Y. Zhang, and D.-J. Yu, "ATPbind: Accurate Protein–ATP Binding Site Prediction by Combining Sequence-Profiling and Structure-Based Comparisons," *Journal of chemical information and modeling,* vol. 58, no. 2, pp. 501-510, 2018.

[7] L. Rao, N.-X. Jia, J. Hu, D.-J. Yu, and G.-J. Zhang, "ATPdock: a template-based method for ATP-specific protein–ligand docking," *Bioinformatics,* vol. 38, no. 2, pp. 556-558, 2021.

[8] M. Hendlich, F. Rippmann, and G. Barnickel, "LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins," *Journal of Molecular Graphics and Modelling,* vol. 15, no. 6, pp. 359-363, 1997.

[9] J. A. Capra, R. A. Laskowski, J. M. Thornton, M. Singh, and T. A. Funkhouser, "Predicting Protein Ligand Binding Sites by Combining Evolutionary Sequence Conservation and 3D Structure," *Plos Computational Biology,* vol. 5, no. 12, pp. e1000585, Dec, 2009.

[10] M. Brylinski, and J. Skolnick, "A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation," *Proceedings of the National Academy of Sciences,* vol. 105, no. 1, pp. 129-134, 2008.

[11] M. N. Wass, L. A. Kelley, and M. J. E. Sternberg, "3DLigandSite: predicting ligand-binding sites using similar structures," *Nucleic Acids Research,* vol. 38, no. suppl_2, pp. W469-W473, 2010.

[12] J. Yang, A. Roy, and Y. Zhang, "Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment," *Bioinformatics,* vol. 29, no. 20, pp. 2588-2595, 2013.

[13] N. Eswar, D. Eramian, B. Webb, M.-Y. Shen, and A. Sali, "Protein Structure Modeling with MODELLER," *Structural Proteomics: High-Throughput Methods*, B. Kobe, M. Guss and T. Huber, eds., pp. 145-159, Totowa, NJ: Humana Press, 2008.

[14] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang, "The I-TASSER Suite: protein structure and function prediction," *Nature methods,* vol. 12, no. 1, pp. 7-8, 2015.

[15] C. A. Rohl, C. E. M. Strauss, K. M. S. Misura, and D. Baker, "Protein Structure Prediction Using Rosetta," *Methods in Enzymology*, pp. 66-93: Academic Press, 2004.

[16] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek

dek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, "Highly accurate protein structure prediction with AlphaFold," *Nature,* vol. 596, no. 7873, pp. 583-589, 2021.

[17] K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Žídek, A. Bridgland, A. Cowie, C. Meyer, and A. Laydon, "Highly accurate protein structure prediction for the human proteome," *Nature,* vol. 596, no. 7873, pp. 590-596, 2021.

[18] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, and J. Bambrick, "Accurate structure prediction of biomolecular interactions with AlphaFold 3," *Nature,* vol. 630, no. 8016, pp. 493-500, 2024.

[19] D.-J. Yu, J. Hu, Z.-M. Tang, H.-B. Shen, J. Yang, and J.-Y. Yang, "Improving protein-ATP binding residues prediction by boosting SVMs with random under-sampling," *Neurocomputing,* vol. 104, pp. 180-190, 2013/03/15/, 2013.

[20] D. J. Yu, J. Hu, J. Yang, H. B. Shen, J. Tang, and J. Y. Yang, "Designing Template-Free Predictor for Targeting Protein-Ligand Binding Sites with Classifier Ensemble and Spatial Clustering," *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* vol. 10, no. 4, pp. 994-1008, 2013.

[21] D.-J. Yu, J. Hu, Y. Huang, H.-B. Shen, Y. Qi, Z.-M. Tang, and J.-Y. Yang, "TargetATPsite: A template-free method for ATP-binding sites prediction with residue evolution image sparse representation and classifier ensemble," *Journal of Computational Chemistry,* vol. 34, no. 11, pp. 974-985, 2013.

[22] J. Hu, Y. Li, W.-X. Yan, J.-Y. Yang, H.-B. Shen, and D.-J. Yu, "KNN-based dynamic query-driven sample rescaling strategy for class imbalance learning," *Neurocomputing,* vol. 191, pp. 363-373, 2016/05/26/, 2016.

[23] J. Hu, L.-L. Zheng, Y.-S. Bai, K.-W. Zhang, D.-J. Yu, and G.-J. Zhang, "Accurate prediction of protein-ATP binding residues using position-specific frequency matrix," *Analytical Biochemistry,* vol. 626, pp. 114241, 2021/08/01/, 2021.

[24] J. S. Chauhan, N. K. Mishra, and G. P. S. Raghava, "Identification of ATP binding residues of a protein from its primary sequence," *BMC Bioinformatics,* vol. 10, no. 1, pp. 434, 2009/12/19, 2009.

[25] K. Chen, M. J. Mizianty, and L. Kurgan, "ATPsite: sequence-based prediction of ATP-binding residues," *Proteome Science,* vol. 9, no. 1, pp. S4, 2011/10/14, 2011.

[26] C.-Q. Xia, X. Pan, and H.-B. Shen, "Protein–ligand binding residue prediction enhancement through hybrid deep heterogeneous learning of sequence and structure data," *Bioinformatics,* vol. 36, no. 10, pp. 3018-3027, 2020.

[27] Y. Xia, X. Pan, and H.-B. Shen, "LigBind: identifying binding residues for over 1000 ligands with relation-aware graph neural networks," *Journal of Molecular Biology,* vol. 435, no. 13, pp. 168091, 2023.

[28] B. Rao, X. Yu, J. Bai, and J. Hu, "E2EATP: fast and high-accuracy protein–ATP binding residue prediction via protein language model embedding," *Journal of Chemical Information and Modeling,* vol. 64, no. 1, pp. 289-300, 2023.

[29] V.-T. Le, M.-S. Malik, Y.-J. Lin, Y.-C. Liu, Y.-Y. Chang, and Y.-Y. Ou, "ATP_mCNN: Predicting ATP binding sites through pretrained language models and multi-window neural networks," *Computers in Biology and Medicine,* vol. 185, pp. 109541, 2025.

[30] A. Elnaggar, M. Heinzinger, C. Dallago, G. Rehawi, Y. Wang, L. Jones, T. Gibbs, T. Feher, C. Angerer, and M. Steinegger, "Prottrans: Toward understanding the language of life through self-supervised learning," *IEEE transactions on pattern analysis and machine intelligence,* vol. 44, no. 10, pp. 7112-7127, 2021.

[31] J.-S. Wu, Y. Liu, F. Ge, and D.-J. Yu, "Prediction of protein-ATP binding residues using multi-view feature learning via contextual-based co-attention network," *Computers in Biology and Medicine,* vol. 172, pp. 108227, 2024.

[32] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, and S. Candido, "Language models of protein sequences at the scale of evolution enable accurate structure prediction," *BioRxiv,* vol. 2022, pp. 500902, 2022.

[33] W. Li, and A. Godzik, "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences," *Bioinformatics,* vol. 22, no. 13, pp. 1658-1659, 2006.

[34] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, and A. Rives, "Evolutionary-scale prediction of atomic-level protein structure with a language model," *Science,* vol. 379, no. 6637, pp. 1123-1130, 2023.

[35] Y. Zeng, Z. Wei, Q. Yuan, S. Chen, W. Yu, Y. Lu, J. Gao, and Y. Yang, "Identifying B-cell epitopes using AlphaFold2 predicted structures and pretrained language model," *Bioinformatics,* vol. 39, no. 4, 2023.

[36] Z. Du, X. Ding, W. Hsu, A. Munir, Y. Xu, and Y. Li, "pLM4ACE: A protein language model based predictor for antihypertensive peptide screening," *Food Chemistry,* vol. 431, pp. 137162, 2024/01/15/, 2024.

[37] K. Palepu, M. Ponnapati, S. Bhat, E. Tysinger, T. Stan, G. Brixi, S. R. T. Koseki, and P. Chatterjee, "Design of Peptide-Based Protein Degraders via Contrastive Deep Learning," *bioRxiv*, pp. 2022.05.23.493169, 2022.

[38] K. Chen, M. J. Mizianty, and L. Kurgan, "Prediction and analysis of nucleotide-binding residues using sequence and sequence-derived structural descriptors," *Bioinformatics,* vol. 28, no. 3, pp. 331-341, 2012.