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.

PTAC: An Interpretable Dual-Path Framework for Temperature-Adaptive Protein Discovery

Jiayu Wang
Beijing University of Chemical Technology
Junchao Shen
Beijing University of Chemical Technology
Ella A. Yi
American High School
Wenya Wang
Beijing University of Chemical Technology
Guimin Zhang
zhangguimin@mail.buct.edu.cn

Beijing University of Chemical Technology

Research Article

Keywords: Protein temperature adaptation, Interpretable model, Psychrophilic protein, Thermophilic proteins, Enzyme discovery

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11572915/v1

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

Additional Declarations: No competing interests reported.

PTAC: An Interpretable Dual-Path Framework for Temperature-Adaptive Protein Discovery

Jiayu Wang[a,b,†], Junchao Shen[a,†], Ella A. Yi[c], Wenya Wang[a], Guimin Zhang[a,b,*]

a. State Key Laboratory of Green Biomanufacturing, College of Life Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

b. Beijing Synthetic Biomanufacturing Innovation Center, Beijing 102209, China

c. American High School, 36300 Fremont Blvd, Fremont, CA 94536, USA

†. These authors contributed equally to this work.

* Corresponding Author: zhangguimin@mail.buct.edu.cn, zhangguimin6@hotmail.com.

Abstract

Reliable prediction of protein temperature adaptation can facilitate enzyme discovery and engineering. However, most existing AI-based methods rely on a single protein language model and inadequately capture sequence patterns across temperature-adaptation classes. Here, we developed PTAC, a dual-path feature-fusion framework that formulates temperature adaptation as a three-class classification task through integrating a general protein language model ProtT5 and a temperature-related encoder ProPrime. On the benchmark datasets, PTAC achieved an overall accuracy of 94.9% and substantially improved recognition of the under-represented psychrophilic class, boosting F1 score to 93.3%, significantly outperforming the reported methods (70.7%). PTAC exhibits strong generalization across remotely homologous and previously unseen sequences, as well as across distinct protein property prediction tasks, maintaining an accuracy of over 80% for halophilic and acidophilic proteins. Interpretability analyses identified sequence features associated with temperature adaptation and deepened our understanding of environmental adaptation mechanisms. Moreover, PTAC achieved 92.0% average accuracy in a practical secreted-protein screening scenario. These results establish PTAC as an effective computational framework for screening temperature-adaptive proteins and mining target enzymes.

Keywords: Protein temperature adaptation; Interpretable model; Psychrophilic protein; Thermophilic proteins; Enzyme discovery

1. Introduction

The temperature range over which enzymes remain active largely determines their value in environmental and industrial contexts [1]. For example, psychrophilic enzymes serve low-temperature bioprocesses including cold laundry detergents and cold food treatment [2]. Thermophilic enzymes maintain activity at high temperatures and are widely used in starch liquefaction, biorefinery and high-temperature biocatalysis to accelerate reactions and minimize microbial contamination[3], whereas mesophilic enzymes dominate conventional biomanufacturing for food additives and pharmaceutical biosynthesis under mild ambient conditions [4,5]. Therefore, rapid and accurate assessment of protein temperature adaptability is important for enzyme discovery and engineering. However, determining temperature adaptability still largely relies on labor-intensive and low-throughput experiments, which limits the large-scale identification of target enzymes from the rapidly expanding protein sequence databases [6].

These limitations have motivated the development of computational approaches for rapid prediction and screening of temperature-adaptive proteins. Early strategies relied mainly on handcrafted features. For example, Wang et al. combined these features with kernel principal component analysis (KPCA) with support vector machines (SVM) to predict thermophilic proteins [7]. Subsequently, deep learning methods were employed to automatically extract hidden features from protein sequences. For instance, DeepTP combined deep learning architectures with physicochemical descriptors to classify thermophilic proteins [8]. With the rapid development of pretrained protein language models (PLMs), Lv et al. proposed the PTSP-BERT, which uses BERT-bfd features to classify thermophilic, mesophilic, and psychrophilic proteins [9]. Peng et al. constructed ESM-PsyPred, which uses ESM-2 to extract deep semantic representations of protein sequences and uses a SVM for temperature classification [10]. These studies show that PLMs capture deep sequence information that traditional handcrafted features are unable to represent, thereby providing a more informative feature base for complex protein classification tasks.

Although significant advancements have been made in predicting protein temperature adaptation, existing methods still suffer from certain limitations. First, the accuracy of existing three-class models that simultaneously distinguish psychrophilic, mesophilic, and thermophilic proteins remain limited,

which are mostly maintained between 80%-90% [9,10], leaving room for improvement. Second, the ability of current models to identify psychrophilic proteins remains limited, with an F1 of only 70.68% [9]. Finally, most existing models provide limited interpretability for three-class temperature classification and fail to identify the sequence features underlying temperature adaptation.

To address these limitations, we developed a three-class Protein Temperature-Adaptation Classifier (PTAC) by integrating the sequence representation from ProtT5 and ProPrime, followed by a bidirectional gated recurrent unit (BiGRU) and MLP classifier for three-class prediction of thermophilic, mesophilic, and psychrophilic proteins. Notably, PTAC achieved an overall accuracy of 94.9% on the independent test set and substantially improved psychrophilic protein prediction, with the F1 increasing from 70.7% to 93.3%. Furthermore, interpretability analysis identified temperature-adaptation features in the three protein classes consistent with previous reports, while candidate protein mining simulations evaluated the model's applicability to large-scale screening. Overall, these results highlight PTAC as a robust and practical computational tool for large-scale screening of temperature-adaptive enzymes from genomic and metagenomic data.

2. Materials and methods

2.1 Dataset preparation

To develop a three-class temperature adaptation model for psychrophilic, mesophilic, and thermophilic proteins, we used a previously reported protein temperature adaptation dataset [11], following its original data curation and partitioning procedures. The final dataset contains 2,400 sequences, and labels were defined as 0, 1, and 2 for mesophilic, thermophilic, and psychrophilic, respectively. The dataset was divided into training (1,920 sequences) and test (480 sequences) sets in a 4:1 ratio. For model input, sequences were limited to a maximum of 1,024 residues, with longer sequences truncated, shorter sequences padded, and attention masks applied.

For transfer evaluation, the halophilic task employed the dataset from HPClas [12]. For pH adaptation classification, we constructed an internal pHmin[13] dataset using protein pH ranges annotations from BRENDA [14] and corresponding sequences from UniProt [15] and NCBI [16]. The lowest pH value was used as the sample label. A total of 77,671 sequences were collected. After filtering

for non-standard residues and sequences length (100 to 800 residues) followed by MMseqs2 [17] redundancy removal (80% coverage, 50% identity), 6,386 high-quality sequences were retained as the pHbre dataset (**Fig. S1a**). We found that extreme acidophilic proteins were scarce in this dataset, therefore, we try to supplement the insufficient data. We selected 11 acidophilic bacteria capable of growing at pH < 2 from BacDive [18] and obtained their complete NCBI proteomes (133,530 sequences). After SignalP 6.0 [19] and TMHMM 2.0 [20] filtering and subsequent sequence processing, 1,423 secreted proteins were retained as pHbac dataset (**Fig. S1a**). A pH threshold of 5 was used to define positive and negative samples; samples with pH values below 5 were considered positive, whereas those with pH values above 5 were considered negative. For independent testing, 200 positive and 200 negative samples were randomly selected from pHmin dataset.

2.2 Model architecture

PTAC consists of two pre-trained protein language models, ProtT5 [21] and ProPrime [22], followed by a feature fusion module, BiGRU [23], attention pooling, and an MLP classifier.

2.2.1 ProtT5 embedding features

ProtT5 [21] was used as a frozen feature extractor to generate residue-level representations from protein sequences. The model parameters were fixed during subsequent training. The hidden states of the last encoder layer were extracted to obtain residue-level features, yielding an (L×1,024) feature matrix for each protein sequence, where L denotes the sequence length.

2.2.2 ProPrime embedding features

ProPrime [22] is a protein language model incorporating a pre-training objective related to the optimal growth temperature of microorganisms, providing sequence features related to protein temperature adaptation. In this study, all parameters of ProPrime were frozen, and the model was used only to extract residue-level features. For a protein sequence length (L), ProPrime outputs an (L×1,280) feature matrix, where 1,280 denotes the feature dimension generated for each residue position.

To ensure consistent sequence positions for feature fusion, both ProtT5 and ProPrime used the same maximum sequence length (L) as input, with sequences truncated and padded accordingly. Their valid-position masks were aligned to the same sequence positions, and their intersection was used to construct the fusion mask. After linear projection, the ProtT5 and ProPrime features were concatenated along the

feature dimension. Padding positions and other invalid positions were masked using a fusion mask during the subsequent attention pooling process.

2.2.3 Dual-path token-level fusion and BiGRU encoding

The core module of this study is a dual-language model token-level fusion structure followed by a BiGRU [23]. ProtT5 and ProPrime generate residue-level features with dimensions of 1,024 and 1,280, respectively. Two independent linear projection layers were therefore applied to map the two feature sets to a common 512-dimensional space.

Then, the two projected feature sets are concatenated along the final feature dimension at the corresponding residue positions $\mathbf{X}=\mathrm{Concat}\left(\boldsymbol{Z}^{T5},\boldsymbol{Z}^{PP}\right)$. The fused sequence feature (**X**) was further input into a BiGRU to integrate forward and backward context information in the protein sequence. In this study, the number of hidden units in each unidirectional GRU was set to 256, and the forward and backward GRUs at the same residue position were concatenated along the feature dimension to obtain the BiGRU output $\boldsymbol{H}^{G}=\mathrm{Concat}\left(\overrightarrow{\mathbf{H}},\overleftarrow{\mathbf{H}}\right)$, where $\boldsymbol{H}^{G}\in\mathrm{R}^{B\times L\times 512}$. Thus, each residue position will eventually obtain a 512-dimensional bidirectional contextual representation that contains both upstream and downstream sequence information for that position.

2.2.4 Attention pooling and MLP classifier

The output of BiGRU remains a token-level representation. In this study, we calculated the mean activation value of each token vector as the attention score. The scores of all valid tokens were then normalized by softmax to obtain attention weights, with an attention mask was used to exclude padding positions from weight calculation. Finally, a global representation of the complete protein sequence was obtained by weighted summation of the BiGRU outputs for all valid tokens.

This protein-level discriminative representation was then input into MLP classifier [24], which generated three output logits corresponding to mesophilic, thermophilic, and psychrophilic proteins. Softmax was applied to obtain the class probabilities, and the class with the highest probability was assigned as the final prediction.

2.3 Model training

The model was trained using five-fold stratified cross-validation, with shuffling and a random seed

of 42. The ProtT5 and ProPrime encoders were frozen, and only the linear projection layers, BiGRU, and MLP classifier were updated. The maximum length of the protein sequence input was set to 1,024, the batch size to 8, and the dropout rate to 0.3. Cross-entropy loss and the AdamW optimizer [25] were used, with the initial learning rate of (1×10^{-4}) and weight-decay of (1×10^{-3}). The learning rate was decayed to 0.5 of its previous value after two consecutive epochs without improvement in validation loss, with a minimum learning rate of (1×10^{-6}). Early stopping was triggered after seven consecutive epochs without improvement, and the model with the lowest validation loss was saved. During the independent testing stage, the predicted probabilities of the five-fold models were averaged to obtain the final classification result.

All experiments were conducted on an NVIDIA GeForce RTX 5090 GPU using Python 3.12.3, PyTorch 2.8.0, Transformers 4.44.2, and CUDA 12.8. The trainable fusion module contained approximately 3.54 million parameters. With FP16 precision and a batch size of 8, the end-to-end inference throughput of the five-fold ensemble model was 15.605 +/- 0.028 sequences/s.

2.4 Baseline models and ablation experiments

The baseline models tested included RF [26], DT [27], GBDT [28], LightGBM [29], XGBoost [30], CatBoost [31], KNN [32], Bayes [33], and MLP [24]. All baseline models used the same fused feature inputs, the same data partitioning, and the same evaluation metrics. Ablation experiments included single-model versus dual-model features, different pooling methods, a linear classifier and an MLP classifier, the presence or absence of BiGRU, and different optimization settings. All ablation experiments were conducted with the same training workflow.

2.5 Evaluation metrics and visualization

Model performance was evaluated using accuracy (ACC), sensitivity (SN), specificity (SP), precision (PRE), F1 score, Matthews correlation coefficient (MCC), macro-averaged area under the receiver operating characteristic curve (AUROC), and macro-averaged area under the precision–recall curve (AUPRC). For the three-class task, a one-versus-rest (OvR) strategy was used to calculate true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each class, followed by the calculation of class-level metrics and macro-average metrics. In addition, t-SNE [34] was applied to visualize the feature embeddings from different stages of the model as two-dimensional

images. t-SNE was primarily used for exploratory visualization analysis and was not used as a quantitative basis for performance evaluation.

2.6 Calculation of amino acid composition and physicochemical properties

For each protein sequence, the frequency of 20 standard amino acids was calculated. According to their physicochemical properties, amino acids were divided into five classes: acidic (Asp, Glu), basic (Lys, Arg, His), polar (Ser, Thr, Asn, Gln), nonpolar (Ala, Val, Leu, Ile, Met, Phe, Trp, Pro, Gly), and aromatic (Phe, Trp, Tyr). The total proportion of each class was then calculated.

At pH 7.0, the net charge of each protein was calculated using the Henderson-Hasselbalch equation. The charge calculation accounted for nine ionizable groups and their corresponding typical pK_a values in the folded protein: basic groups—Lys (pK_a 10.5), Arg (pK_a 12.5), His (pK_a 6.0), and the N-terminus (pK_a 8.0)—and acidic groups—Asp (pK_a 3.9), Glu (pK_a 4.3), Cys (pK_a 8.3), Tyr (pK_a 10.1), and the C-terminus (pK_a 3.5).

The protonation fraction of basic groups was calculated as: $f = 1/(1+10^{\mathrm{pH-p}K_a})$.

The deprotonation fraction of acidic groups was calculated as: $f = 1/(1+10^{\mathrm{p}K_a-\mathrm{pH}})$.

The net charge was calculated by adding the positive and negative charges of all groups and then divided by the sequence length to obtain the normalized net charge ratio (%).

2.7 Identification of secreted proteins from three strains

The complete proteome sequences of all three strains were obtained from NCBI [16]. SignalP 6.0 [19] was used to predict signal peptides, retaining sequences with signal peptides and removing the signal peptide fragments. TMHMM [20] was then used to remove proteins with obvious transmembrane regions, followed by removal of sequences with abnormal lengths or containing non-canonical amino acids, forming the Final Samples.

3. Results and discussion

3.1 Model architecture of PTAC

Here, we developed PTAC by fusing two protein language models (PLMs), comprising protein sequence input, dual-protein language model feature-extraction, BiGRU-based feature fusion, and MLP classification (**Fig. 1**). ProtT5 and ProPrime provide complementary sequence representations. ProtT5

can generate informative embeddings to describe general protein structure, function, and evolution, and also can effectively capture global evolutionary semantics and long-range dependencies [21,35], whereas ProPrime is a temperature-aware protein language model in which Optimal Growth Temperature (OGT) serves as one of the core pre-training tasks [22]. Specifically, residue-level representations of protein sequences were firstly extracted from ProtT5 and ProPrime, respectively. The two representations are separately projected into a common feature space and then fused at corresponding sequence positions. The resulting representations are subsequently integrated by a BiGRU and summarized through attention pooling before being fed into the MLP classifier, forming the final PTAC model.

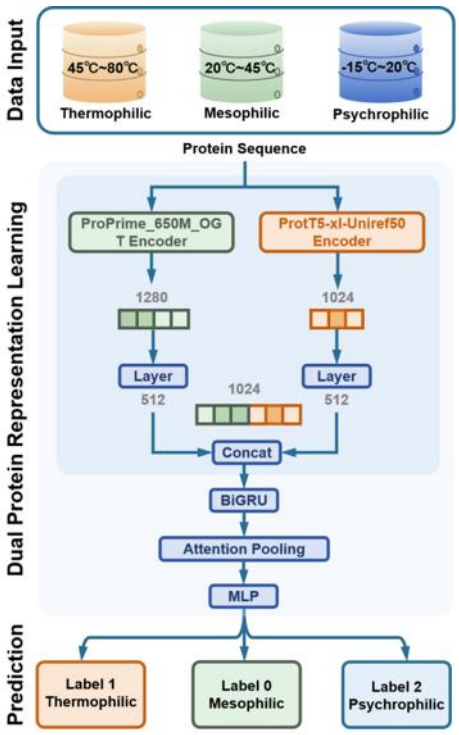


Fig. 1. Model architecture of PTAC

On the independent test set, PTAC achieved average performance values of 94.9% for accuracy (ACC), 94.8% for F1, and 92.5% for MCC. The F1 scores were 93.3%, 92.2%, and 98.9% for psychrophilic, mesophilic, and thermophilic proteins, respectively (**Table S1**). We rigorously compared PTAC with PTSP-BERT [9] under the same experimental setting. PTSP-BERT was retrained on the

training set used in this study, following the procedures described in the original publication, and evaluated on the same independent test set as PTAC (**Table 1**). PTAC improved ACC by 20.5% and MCC by 27.9% compared with PTSP-BERT. It also achieved higher F1, specificity (SP), AUROC, and AUPRC values. Notably, the model shows a more significant advantage in classifying psychrophilic proteins (**Fig. S2**), with precision increasing from 57.6% to 89.5% and F1 from 70.7% to 93.3%. This indicates that PTAC not only maintained strong performance in identifying thermophilic and mesophilic proteins, but also demonstrated superior discriminative ability for relatively less frequently mentioned psychrophilic protein class compared to existing methods.

Table 1. Comparison between PTSP-BERT and PTAC

Method	ACC	F1	MCC	SP	PRE	AUROC	AUPRC
PTSP-BERT	74.4%	74.2%	64.5%	87.4%	80.8%	93.2%	88.5%
PTAC	94.9%	94.8%	92.5%	95.0%	94.8%	99.2%	98.0%

The improved performance of PTAC, especially for psychrophilic protein classification, can be attributed to the fusion of ProtT5 and ProPrime, which preserves general protein features while incorporating complementary temperature-related information. Similar multi-model fusion strategies have been validated in other protein prediction tasks. For example, Xia et al. proposed DeepSaltPro by fusing Ankh and ESM-2 embeddings to improve the prediction performance of halophilic protein [36], while Pokharel et al. integrated embeddings from Ankh, ESM-2, and ProtT5 for protein O-GlcNAc site prediction [37]. More recently, systematic benchmarking has shown that protein language model embedding fusion can exploit complementary information across models and improve performance across multiple protein-related tasks [38]. These findings support the use of complementary protein language model representations.

3.2 Performance comparison of different language models and classifier

Insufficient sample size of psychrophilic proteins in dataset [11] is a typical class imbalance challenge in protein temperature adaptation classification. Therefore, MCC and F1 were used as the primary evaluation metrics, with accuracy and precision reported as complementary measures [39,40].

We first conducted a series of tests and comparisons on the different pre-trained PLMs. Using the VenusFactory [41] platform, we evaluated the feature extraction ability of representative PLMs. The comparative results of validation sets (**Table S2)** and independent test sets **(Fig. 2a, Table S1)** showed that the Prime [22], ProtTrans [21], and ESM [42] series models present relatively stable and excellent performance, maintaining high recognition accuracy across all three temperature classes, with precision and F1 scores exceeding 85% for each class (**Fig. 2b**). Among the evaluated models, our fused model PTAC outperform either model alone, with improvements of 7.5% or 3.0% in MCC and 5.0% or 2.0% in F1 compared with ProtT5 or ProPrime, respectively (**Table S1**). Notably, the fusion model PTAC also showed improved recognition of the under-represented psychrophilic class. Compared with ProtT5, PTAC improved the precision and F1 for psychrophilic proteins by 3.9% and 6.0%, respectively, while the corresponding improvements over ProPrime were 2.9% and 1.9% (**Table S1, Fig. 2b**). These results support the use of dual-PLM feature fusion to improve temperature-adaptation classification, particularly for the under-represented psychrophilic class.

We next evaluated the downstream classifier using the fixed ProtT5-ProPrime fusion features. MLP was compared with eight commonly used classifiers, including XGBoost, CatBoost, Random Forest (RF), Gradient Boosting Decision Tree (GBDT), LightGBM, k-Nearest Neighbors (KNN), Naive Bayes (Bay), and Decision Tree (DT) [24,26–33], on the validation set using five-fold cross-validation. As shown in **Fig. 2c**, MLP achieved the highest MCC (91.9%) and F1 (94.3%), together with the smallest five-fold standard deviations (1.1% for MCC and 0.7% for F1) (**Table S3**). CatBoost and XGBoost perform second best, while simpler models such as Decision Trees perform substantially worse because they could not effectively capture the complex nonlinear relationships in the high-dimensional fusion features. These results indicate that MLP can more fully learn the high-dimensional continuous features generated by protein language models, therefore achieving better predictive performance than traditional machine-learning classifiers.

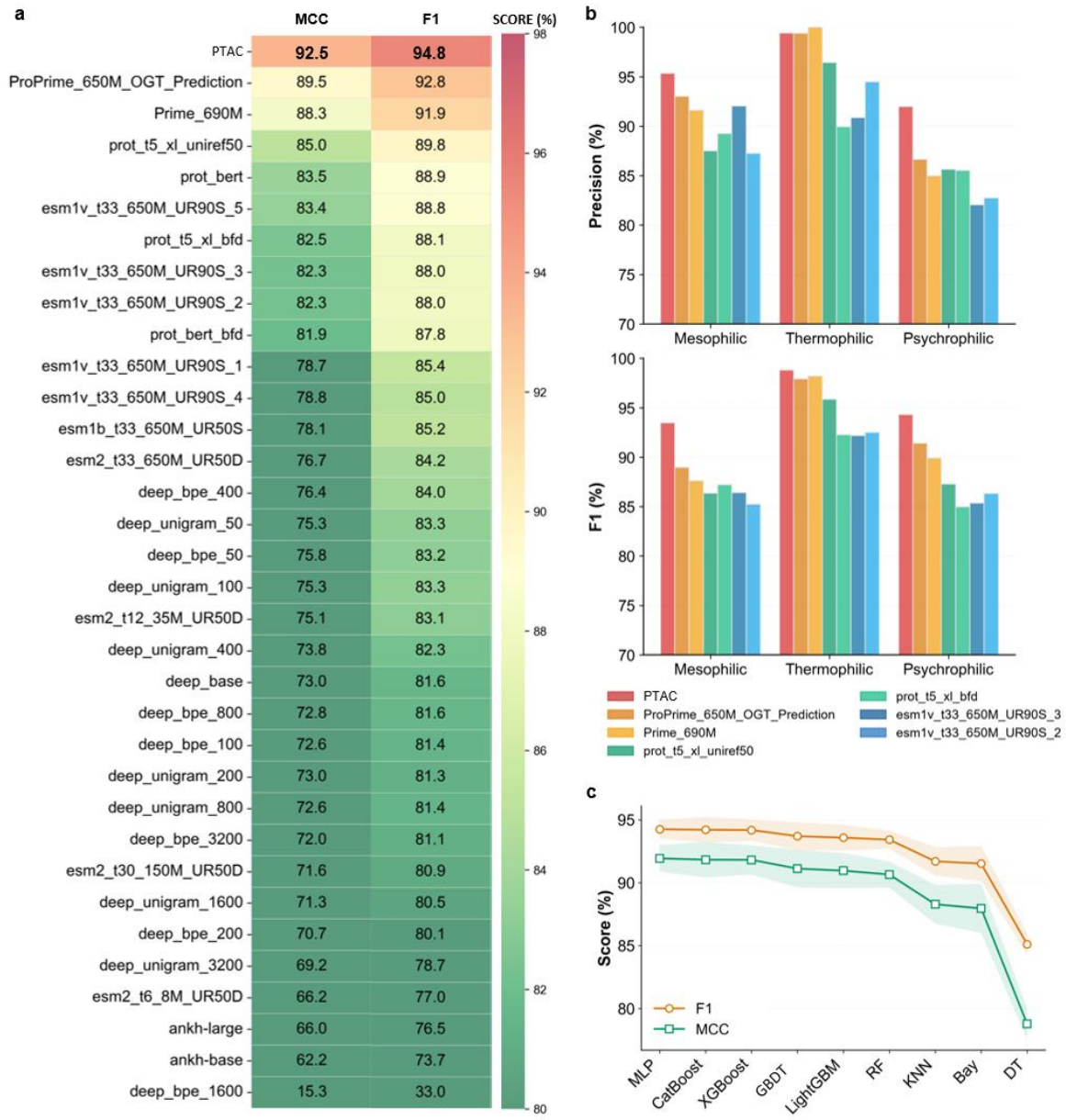


Fig. 2. Performance comparison of different protein language models and their downstream classifiers. (a) Overall performance heatmap of different protein language models on the independent test set. **(b)** Class-specific precision and F1 scores on the independent test set. **(c)** Performance comparison of different downstream classifiers based on the same fusion features (ProtT5 + ProPrime). The colored intervals above and below the lines represent the standard deviations of the five-fold cross-validation.

3.3 Model generalization of PTAC

The generalization ability of PTAC was evaluated from two aspects: performance across different

sequence similarity and transferability to other protein adaptation prediction tasks. First, we divided the independent test set into seven subsets according to the sequence similarity between the training set and the independent test set. The seven subsets were separately input into PTAC to evaluate its performance (**Fig. 3a-c, Table S4**). The results showed that even in the low homology subset with only 20% homology, the model maintained excellent predictive performance, with an MCC of 89.7%, an F1 of 93.1%, and an accuracy of 93.3%. As the sequence similarity decreased, the model's performance did not decline markedly; in the high homology subset with 80% similarity, the accuracy reached 100%. These results indicate that PTAC retained predictive performance across a wide range of sequence similarity, including distantly related sequences.

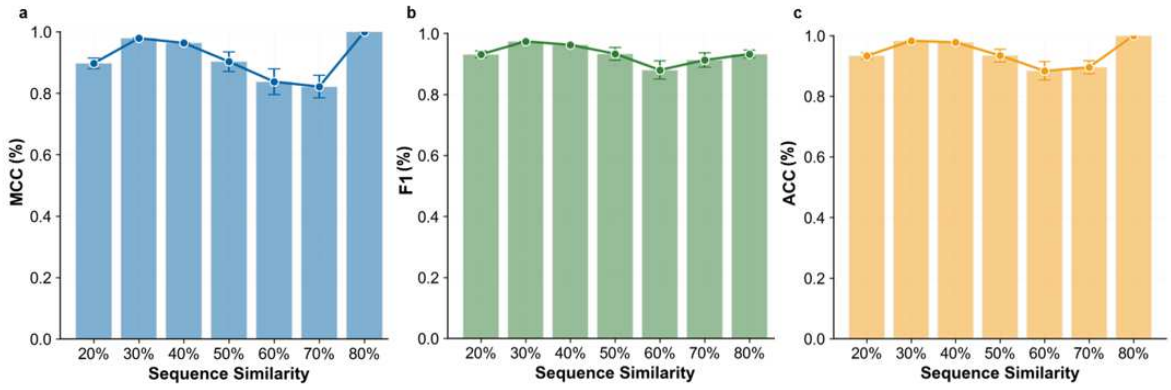


Fig. 3. Generalization performance across sequence similarity intervals. The independent test set was divided into seven threshold intervals with sequence identity ranging from 20% to 80%, corresponding to sample sizes of 193, 58, 89, 76, 36, 21, and 7, respectively. The MCC **(a)**, F1 **(b)**, and accuracy **(c)** of the complete model were then evaluated. The error bars indicate the standard deviation of prediction performance on the independent test set of the five-fold cross-validation models.

Then, the generalization ability of PTAC was further evaluated by predicting halophilic and acidophilic proteins. The core model architecture and pre-trained parameters were kept unchanged, with only the output classification head replaced by a binary classifier. The resulting models were fine-tuned on the corresponding task-specific datasets and evaluated on independent test sets. As shown in **Table 2**, after transfer training on test set, PTAC achieved an ACC of 88.3% and an F1 of 88.3% on the

halophilic protein, and an ACC of 80.6% and an F1 of 80.3% on the acidophilic protein. These results demonstrate that PTAC retains strong predictive performance when transferred to other protein property prediction tasks beyond temperature adaptation.

Table 2. Performance on prediction tasks of halophilic and acidophilic protein properties

Function	ACC	F1	MCC	SP	PRE	AUROC	AUPRC
Halophilic	88.3%	88.3%	76.7%	88.3%	88.4%	96.2%	96.4%
Acidophilic	80.6%	80.3%	63.0%	80.6%	82.5%	90.4%	90.2%

PTAC demonstrates strong generalization across remotely homologous and previously unseen sequences, as well as across distinct protein property prediction tasks, maintaining an accuracy of over 80% for halophilic and acidophilic proteins. This finding is also consistent with the complementary roles of the two feature paths. ProPrime appears to capture more temperature-related information [22], contributing to its stronger performance on the main task. In contrast, ProtT5 provides more general protein sequence representations [21,35], which may complement ProPrime and support stable performance in transfer tasks such as halophilic and acidophilic protein classification.

3.4 Model visualization and interpretability analysis

To visually demonstrate the feature learning process of the model, we used t-SNE dimensionality reduction method [34] to sequentially visualize the features of ProtT5 [21,35], ProPrime [22], dual-path fusion, and MLP output [24] into two-dimensional images. In the features of a single PLM, the original ProtT5 embeddings (**Fig. 4a**) show significant overlap among the three protein classes. The original ProPrime embeddings (**Fig. 4b**) show modest improvement, with thermophilic proteins aggregating into a relatively separate cluster; however mesophilic and psychrophilic proteins continued to overlap extensively. By using complementary features from ProtT5 and ProPrime through a dual-path strategy, the separation between the three protein classes is significantly improved (**Fig. 4c**). Thermophilic proteins clustered completely independently, whereas mesophilic and psychrophilic proteins formed distinct core regions with minimal boundary overlap. This shows that multi-model fusion successfully combines the strengths of different PLMs. After the final nonlinear transformation of the MLP, the three

classes of temperature-adaptive protein samples formed three independent clusters with clear boundaries and compact internal structures (**Fig. 4d**). The visualization results confirm that the model progressively refines features at each stage, ultimately achieving effective feature discrimination for temperature adaptive proteins.

To explore whether the model has learned a biologically meaningful discriminative pattern, we selected three proteins with the highest prediction confidence and three proteins with the lowest prediction confidence in the independent test set from each class (18 proteins in total) (**Table S5**) and systematically compared them in sequence composition, physicochemical properties, and structural features. Protein length analysis (**Fig. 4e**) showed that low-confidence psychrophilic and thermophilic proteins generally had shorter sequences and simpler secondary structure compositions (**Fig. S3a**), suggesting that insufficient sequence information may be a key factor influencing prediction bias. Analysis of the acidic and basic amino acid composition (**Fig. 4f-g**) showed that thermophilic proteins had a higher proportion of acidic and basic residues, consistent with an extensive salt-bridge network [43]. Notably, low-confidence psychrophilic proteins exhibited an abnormally high proportion of acidic and basic residues, a pattern contrary to the characteristic reliance of psychrophilic proteins on flexible structures for low-temperature catalysis [44] which may therefore represent a key factor contributing to model misclassification. Polar/nonpolar amino acids analysis (**Fig. 4h-i**) revealed confidence-related differences. High-confidence psychrophilic proteins had the highest polar residue proportion, consistent with enhanced surface hydrophilicity and structural flexibility [45–47]. Low-confidence psychrophilic proteins exhibited reduced polar residues, whereas low-confidence mesophilic proteins had an abnormally high proportion of polar residues, resulting in overlapping characteristics and potential misclassification. Furthermore, low-confidence thermophilic proteins contained fewer nonpolar residues than high-confidence thermophilic proteins, suggesting weakened hydrophobic cores [46,48] and deviation from canonical thermophilic features. Net-charge analysis (**Fig. 4j**) further supported these results. In the high confidence samples, the net charge of mesophilic proteins was close to zero, the thermophilic proteins were positively charged, and psychrophilic proteins were negatively charged. The surface electrostatic potential distribution was also consistent with the overall net-charge features (**Fig. S3b**). The positive surface charges of thermophilic proteins can reduce nonspecific protein aggregation

at high temperatures; the negative surface charge of psychrophilic proteins can attract more solvent water molecules and form a thicker, more fluid hydration layer, thereby maintaining conformational flexibility and catalytic activity at low temperatures [47]. The dispersed net charge distribution may interfere with model predictions, particularly for samples near the class boundaries.

Residue frequency analysis showed that confidence differences were associated with residue preferences. High-confidence samples exhibited typical features consistent with their temperature-adaptation classes [43–45,47–50], whereas low-confidence samples generally showed mixed features or deviated from typical patterns. In mesophilic proteins, low-confidence samples showed lower levels of hydrophobic residues (Ala and Leu) and structurally rigid residues (Pro), but higher levels of polar (Asn) and positively charged (Lys) residues (**Fig. 4k**), resulting in mixed sequence features of mesophilic and psychrophilic proteins. In thermophilic proteins, low-confidence samples showed lower levels of flexibility-related (Gly) and hydrophobic (Val) residues, but higher levels of negatively charged residues (Asp) (**Fig. 4l**), resulting in mixed thermophilic- and mesophilic-like features. Similarly, in psychrophilic proteins, low-confidence samples showed lower levels of hydrophobic (Ile) and hydroxy (Ser) residues, but higher levels of negatively charged residues (Glu) (**Fig. 4m**), resulting in mixed psychrophilic- and mesophilic-like features.These specific differences in residues composition provide more detailed clues to understanding the basis of model discrimination.

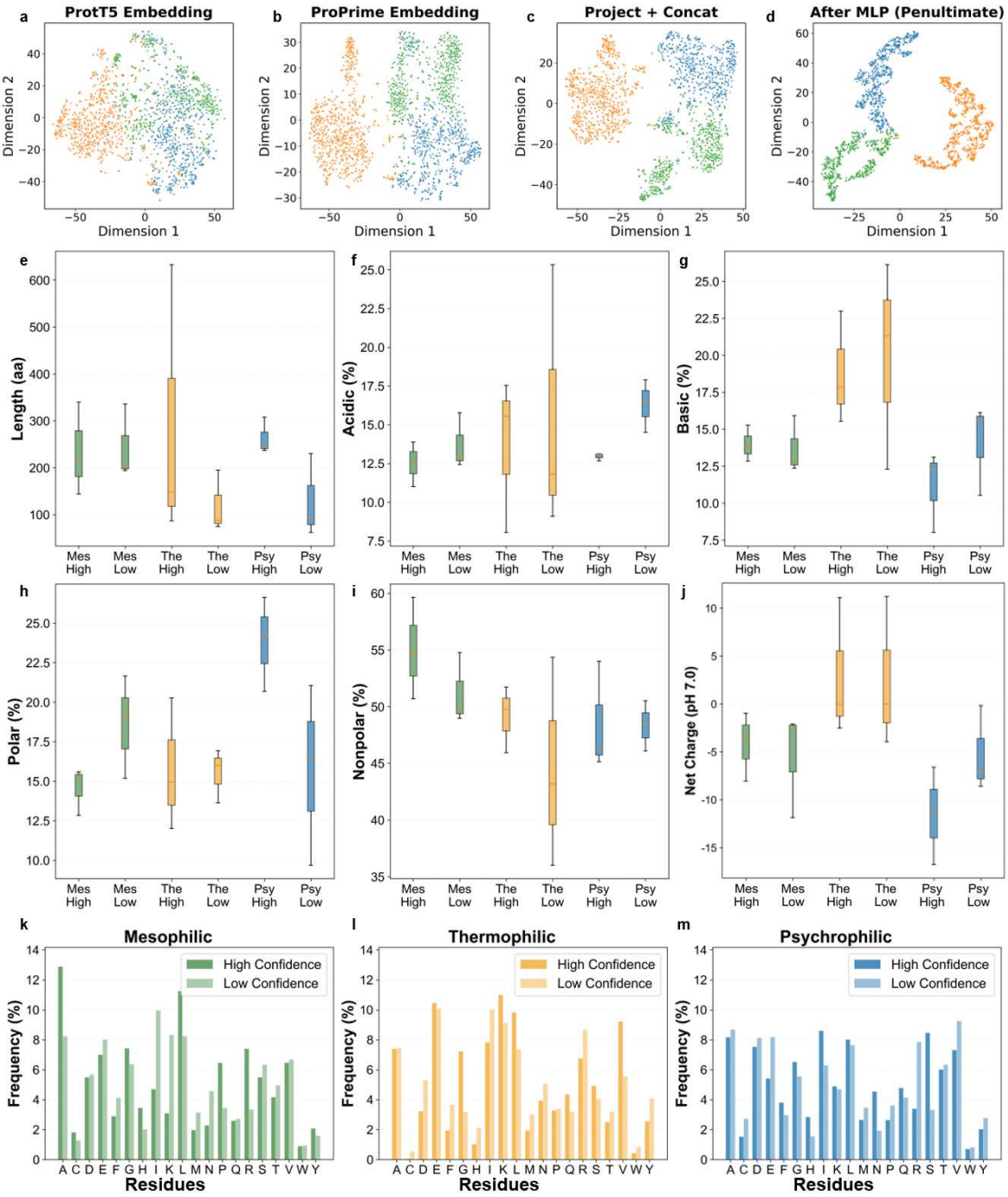


Fig. 4. Dimensionality reduction visualization and comparative analysis of sequence features. Green represents mesophilic proteins, yellow represents thermophilic proteins, and blue represents psychrophilic proteins. **(a)** Original ProtT5 embeddings. **(b)** Original ProPrime embeddings. **(c)** Projected features after concatenation. **(d)** MLP hidden-layer features. **(e)** Protein length distribution. **(f)** Acidic amino acid content. **(g)** Basic amino acid content. **(h)** Polar amino acid content. **(i)** Nonpolar amino acid content. **(j)** Protein net-charge distribution at pH 7.0. **(k-m)** Amino acid frequencies of

mesophilic **(k)**, thermophilic **(l)**, and psychrophilic **(m)** proteins. The group abbreviations are: Mes-High (high-confidence mesophilic proteins), Mes-Low (low-confidence mesophilic proteins), The-High (high-confidence thermophilic proteins), The-Low (low-confidence thermophilic proteins), Psy-High (high-confidence psychrophilic proteins), and Psy-Low (low-confidence psychrophilic proteins). Data are shown as mean +/- standard deviation.

3.5 The application of PTAC

The PTAC exhibited robust classification performance on the independent test set. To further evaluate the applicability of PTAC in practical scenarios, we selected secreted proteins from representative microorganisms with different temperature-related ecological backgrounds as external validation objects. The optimal growth temperature of a host strain can macroscopically characterize the overall temperature adaptation tendency of its secretome [51,52]. Therefore, we used the optimal growth temperature of each strain as a reference label for the corresponding secretory protein prediction results, rather than treating it as equivalent to the true optimal temperature of each protein.

For external validation, we selected three representative microorganisms with distinct temperature-adaptation profiles. *Atrimonas thermophila* M15^T is a thermophilic bacterium with an optimal growth temperature of 70-75 °C [53]. *Methyloraptor flagellatus* S20^T is a mesophilic bacterium with an optimal growth temperature of 30-35 °C [54]. *M. burtonii* DSM 6242 grows optimally at approximately 23.4 °C and can grow at temperatures near the freezing point, with a theoretical minimum growth temperature as low as -2.5 °C [55]. All secretory proteins from the three strains were data cleaned and input into PTAC. The prediction results were shown in **Table 3**.

Table 3. Prediction results for secretory proteins from three representative strains

Strain	Growth Temperature	Final Samples	Thermo-philic	Meso-philic	Psychro-philic	Accuracy
A. thermophila M15^T	70-75 ℃	23	22	0	1	95.6%
M. flagellatus S20^T	30-35 ℃	317	0	317	0	100.0%

Strain	Growth Temperature	Final Samples	Thermo-philic	Meso-philic	Psychro-philic	Accuracy
M. burtonii DSM 6242	23.4 ℃ (-2.5 ℃)	91	6	12	73	80.2%

Final Samples: Secretory protein samples after data cleaning

For the thermophilic bacterium *A. thermophila* M15^T, 22 out of the 23 proteins were predicted as thermophilic, achieving an accuracy of 95.6%. The only misclassified protein, BER92295.1, was annotated as a putative thioredoxin/peroxiredoxin-associated oxidoreductase. The conformational flexibility of its catalytic motifs may resemble that of cold-adapted proteins [56], contributing to its misclassification. For the mesophilic bacterium *M. flagellatus* S20^T, all 317 predicted secreted protein were classified as mesophilic, achieving 100.0% accuracy. This result indicates that PTAC has a relatively stable ability to distinguish proteins from both thermophilic and mesophilic backgrounds. For the low-temperature-adapted archaeon *M. burtonii* DSM 6242, PTAC prediction was more challenging: 73 of the 91 proteins were predicted to be psychrophilic, corresponding to an overall proportion of 80.2%, while the remaining 18 were classified as mesophilic (12) or thermophilic (6). However, the lower psychrophilic prediction rate does not necessarily indicate a deficiency in PTAC, as proteins from low-temperature-adapted microorganisms do not necessarily exhibit typical psychrophilic characteristics. For example, among the 18 proteins not classified as psychrophilic, the c-type cytochrome WP_011498485.1 was classified as mesophilic by PTAC with Tm value of 48.3 ℃, measured by wet experiment [57]. In addition, during long-term evolution, secretory proteins from thermophiles are under strong selective pressure to maintain stability and activity at high temperatures. In contrast, psychrophilic proteins do not face an equivalent selective pressure for low-temperature adaptation, and therefore may not undergo the same degree of evolutionary optimization. Consequently, proteins from psychrophilic microorganisms are not necessarily characterized by typical psychrophilic temperature properties. This is also supported by previous studies showing that the alkaline phosphatase SIB1 APase from the psychrophilic bacterium *Shewanella sp*. SIB1 has an optimal reaction temperature as high as 50 ℃ [58].

4. Conclusion

In all, we developed PTAC, a dual-path framework for three-class prediction of protein temperature adaptation. PTAC achieved an overall accuracy of 94.9% and a psychrophilic F1 of 93.3%, demonstrating improved discrimination across temperature-adaptation classes, particularly for the under-represented psychrophilic class. Beyond the benchmark dataset, PTAC maintained robust performance on low-homology and previously unseen sequences and achieved accuracies above 80% in halophilic and acidophilic protein classification, supporting the transferability of its learned representations to broader protein adaptation tasks. Interpretability analysis further identified sequence features associated with temperature adaptation. In a practical secretory protein screening scenario, PTAC achieved 92.0% average accuracy, highlighting its potential for protein and enzyme discovery. Overall, PTAC provides an effective and interpretable framework that extends temperature-adaptation prediction beyond conventional sequence classification toward broader protein screening applications.

Data availability: The data and code for this study can be found in a GitHub repository accompanying this manuscript: https://github.com/sjc88661/PTAC

Contributions

J.W.: Writing- original draft, Conceptualization, Investigation, Formal analysis, Visualization. **J. S.:** Writing-original draft, Investigation, Formal analysis. **E. A. Y.:** Data curation, Validation. **W. W.:** Writing-review & editing, Supervision. **G. Z.:** Writing-review & editing, Supervision, Project administration, Funding acquisition. All authors reviewed the manuscript.

Funding

This work was supported by National Key Research and Development Program of China (2022YFC2105003) and Beijing Municipal Science and Technology Program (Z251100000125007), China.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Dumorné K, Córdova DC, Astorga-Eló M, Renganathan P. Extremozymes: a potential source for industrial applications. J Microbiol Biotechnol. 2017;27:649–59. https://doi.org/10.4014/jmb.1611.11006.

[2] Kuddus M, Roohi, Bano N, Sheik GB, Joseph B, Hamid B, Sindhu R, Madhavan A. Cold-active microbial enzymes and their biotechnological applications. Microb Biotechnol. 2024;17:e14467. https://doi.org/10.1111/1751-7915.14467.

[3] Li J, Sun L, Huo Y-X. High-temperature catalytic platform powered by thermophilic microorganisms and thermozymes. Synth Biol Eng. 2025;3:10001. https://doi.org/10.70322/sbe.2025.10001.

[4] De La Fuente M, Lombardero L, Gómez-González A, Solari C, Angulo-Barturen I, Acera A, Vecino E, Astigarraga E, Barreda-Gómez G. Enzyme therapy: current challenges and future perspectives. Int J Mol Sci. 2021;22:9181. https://doi.org/10.3390/ijms22179181.

[5] Pingoud A, Wilson GG, Wende W. Type II restriction endonucleases—a historical perspective and more. Nucleic Acids Res. 2016;44:8011. https://doi.org/10.1093/nar/gkw513.

[6] Robinson SL, Piel J, Sunagawa S. A roadmap for metagenomic enzyme discovery. Nat Prod Rep. 2021;38:1994–2023. https://doi.org/10.1039/D1NP00006C.

[7] Wang X-F, Gao P, Liu Y-F, Li H-F, Lu F. Predicting thermophilic proteins by machine learning. Curr Bioinform. 2020;15:493–502. https://doi.org/10.2174/1574893615666200207094357.

[8] Zhao J, Yan W, Yang Y. DeepTP: a deep learning model for thermophilic protein prediction. Int J Mol Sci. 2023;24:2217. https://doi.org/10.3390/ijms24032217.

[9] Lv Z, Wei M, Pei H, Peng S, Li M, Jiang L. PTSP-BERT: predict the thermal stability of proteins using sequence-based bidirectional representations from transformer-embedded features. Comput Biol Med. 2025;185:109598. https://doi.org/10.1016/j.compbiomed.2024.109598.

[10] Peng C, Bian Y, Yuan C, Chen Y, Liu D, Lu F, Liu F, Liu Y. ESM-PsyPred: leveraging protein

language models for accurate prediction of psychrophilic proteins. Interdiscip Sci Comput Life Sci. 2026. https://doi.org/10.1007/s12539-025-00810-7.

[11] Huang A, Lu F, Liu F. Discrimination of psychrophilic enzymes using machine learning algorithms with amino acid composition descriptor. Front Microbiol. 2023;14:1130594. https://doi.org/10.3389/fmicb.2023.1130594.

[12] Hu S, Wang X, Wang Z, Jiang M, Wang S, Wang W, Song J, Zhang G. HPClas: a data-driven approach for identifying halophilic proteins based on catBoost. mLife. 2024;3:mlf2.12125. https://doi.org/10.1002/mlf2.12125.

[13] Hu S, Pan T, Wang J, Koh HY, Wang Z, Zhang Y, Zhu L, Li Z, Webb GI, Song J, Zhang G. ACENet: a graph neural network for predicting enzyme pHmin using structure and sequence insights. bioRxiv. 2026. https://doi.org/10.64898/2026.03.11.711222.

[14] Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, Neumann-Schaal M, Jahn D, Schomburg D. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. Nucleic Acids Res. 2021;49:D498–D508. https://doi.org/10.1093/nar/gkaa1025.

[15] UniProt Consortium. UniProt: the universal protein knowledgebase in 2025. Nucleic Acids Res. 2025;53:D609–D617. https://doi.org/10.1093/nar/gkae1010.

[16] Sayers EW, Beck J, Bolton EE, Brister JR, Chan J, Connor R, Feldgarden M, Fine AM, Funk K, Hoffman J, Kannan S, Kelly C, Klimke W, Kim S, Lathrop A, Marchler-Bauer A, Murphy TD, O'Sullivan C, Schmieder E, Skripchenko Y, Stine A, Thibaud-Nissen F, Wang J, Ye J, Zellers E, Schneider VA, Pruitt KD. Database resources of the National Center for Biotechnology Information in 2025. Nucleic Acids Res. 2025;53:D20–D29. https://doi.org/10.1093/nar/gkae979.

[17] Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35:1026–8. https://doi.org/10.1038/nbt.3988.

[18] Reimer LC, Sardà Carbasse J, Koblitz J, Ebeling C, Podstawka A, Overmann J. Bac Dive in 2022: the knowledge base for standardized bacterial and archaeal data. Nucleic Acids Res. 2022;50:D741–D746. https://doi.org/10.1093/nar/gkab961.

[19] Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, Von Heijne G, Nielsen H. SignalP 6.0 predicts all five types of signal peptides using protein

language models. Nat Biotechnol. 2022;40:1023–5. https://doi.org/10.1038/s41587-021-01156-3.

[20] Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567–80. https://doi.org/10.1006/jmbi.2000.4315.

[21] Elnaggar A, Heinzinger M, Dallago C, Rehawi G, Wang Y, Jones L, Gibbs T, Feher T, Angerer C, Steinegger M, Bhowmik D, Rost B. ProtTrans: toward understanding the language of life through self-supervised learning. IEEE Trans Pattern Anal Mach Intell. 2022;44:7112–27. https://doi.org/10.1109/TPAMI.2021.3095381.

[22] Jiang F, Li M, Dong J, Yu Y, Sun X, Wu B, Huang J, Kang L, Pei Y, Zhang L, Wang S, Xu W, Xin J, Ouyang W, Fan G, Zheng L, Tan Y, Hu Z, Xiong Y, Feng Y, Yang G, Liu Q, Song J, Liu J, Hong L, Tan P. A general temperature-guided language model to design proteins of enhanced stability and activity. Sci Adv. 2024;10:eadr2641. https://doi.org/10.1126/sciadv.adr2641.

[23] Li Z, Yu Y. Protein secondary structure prediction using cascaded convolutional and recurrent neural networks. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, NY: AAAI Press; 2016. pp. 2560–7. https://www.ijcai.org/Proceedings/16/Papers/364.pdf.

[24] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323:533–6. https://doi.org/10.1038/323533a0.

[25] Loshchilov I, Hutter F. Decoupled weight decay regularization. In: International Conference on Learning Representations. 2017. https://openreview.net/forum?id=rk6qdCgC7.Accessed 11 Oct 2026.

[26] Breiman L. Random forests. Mach Learn. 2001;45:5–32. https://doi.org/10.1023/A:1010933404324.

[27] Quinlan JR. Induction of decision trees. Mach Learn. 1986;1:81–106. https://doi.org/10.1007/BF00116251.

[28] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29:1189–232. https://www.jstor.org/stable/2699986.Accessed 30 May 2026.

[29] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y. LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems. 2017. https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-

Abstract.html.Accessed 30 May 2026.

[30] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: Association for Computing Machinery; 2016. pp. 785–94. https://doi.org/10.1145/2939672.2939785.

[31] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: Advances in Neural Information Processing Systems. 2018. https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html.Accessed 30 May 2026.

[32] Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans Inf Theory. 1967;13:21–7. https://doi.org/10.1109/TIT.1967.1053964.

[33] Domingos P, Pazzani M. On the optimality of the simple Bayesian classifier under zero-one loss. Mach Learn. 1997;29:103–30. https://doi.org/10.1023/A:1007413511361.

[34] Husnain M, Missen MMS, Mumtaz S, Luqman MM, Coustaty M, Ogier J-M. Visualization of high-dimensional data by pairwise fusion matrices using t-SNE. Symmetry. 2019;11:107–22. https://doi.org/10.3390/sym11010107.

[35] Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, Guo D, Ott M, Zitnick CL, Ma J, Fergus R. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci U S A. 2021;118:e2016239118. https://doi.org/10.1073/pnas.2016239118.

[36] Xia Y, Guo Q, Liu T. DeepSaltPro: enhancing halophilic protein prediction accuracy and efficiency via multi-protein language model integration. Int J Biol Macromol. 2025;333:148947. https://doi.org/10.1016/j.ijbiomac.2025.148947.

[37] Pokharel S, Pratyush P, Ismail HD, Ma J, Kc DB. Integrating embeddings from multiple protein language models to improve protein O-GlcNAc site prediction. Int J Mol Sci. 2023;24:16000. https://doi.org/10.3390/ijms242116000.

[38] Ko YS, Parkinson J, Wang W. Scalable embedding fusion with protein language models: insights from benchmarking text-integrated representations. Brief Bioinform. 2026;27:bbag014. https://doi.org/10.1093/bib/bbag014.

[39] Jiang J, Zhang C, Ke L, Hayes N, Zhu Y, Qiu H, Zhang B, Zhou T, Wei G-W. A review of machine learning methods for imbalanced data challenges in chemistry. Chem Sci. 2025;16:7637–58. https://doi.org/10.1039/d5sc00270b.

[40] Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020;21:6. https://doi.org/10.1186/s12864-019-6413-7.

[41] Tan Y, Liu C, Gao J, Wu B, Li M, Wang R, Zhang L, Yu H, Fan G, Hong L, Zhou B. VenusFactory: an integrated system for protein engineering with data retrieval and language model fine-tuning. In: Mishra P, Muresan S, Yu T, editors. Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations). Vienna, Austria: Association for Computational Linguistics; 2025. pp. 230–41. https://doi.org/10.18653/v1/2025.acl-demo.23.

[42] Meier J, Rao R, Verkuil R, Liu J, Sercu T, Rives A. Language models enable zero-shot prediction of the effects of mutations on protein function. In: Advances in Neural Information Processing Systems. 2021. pp. 29287–303. https://proceedings.neurips.cc/paper/2021/hash/f51338d736f95dd42427296047067694-Abstract.html.Accessed 3 Oct 2026.

[43] Modarres HP, Mofrad MR, Sanati-Nezhad A. Protein thermostability engineering. RSC Adv. 2016;6:115252–70. https://doi.org/10.1039/C6RA16992A.

[44] Mandelman D, Ballut L, Wolff DA, Feller G, Gerday C, Haser R, Aghajari N. Structural determinants increasing flexibility confer cold adaptation in psychrophilic phosphoglycerate kinase. Extremophiles. 2019;23:495–506. https://doi.org/10.1007/s00792-019-01102-x.

[45] Feller G, Gerday C. Psychrophilic enzymes: hot topics in cold adaptation. Nat Rev Microbiol. 2003;1:200–8. https://doi.org/10.1038/nrmicro773.

[46] Wilson MC. Comparing the thermal, chemical and mechanical stabilities of extremophilic cold shock protein variants. PhD thesis. Leeds: University of Leeds; 2016. https://etheses.whiterose.ac.uk/15982/.

[47] Feller G. Psychrophilic enzymes: from folding to function and biotechnology. Scientifica. 2013;2013:512840. https://doi.org/10.1155/2013/512840.

586 [48] Haney PJ, Badger JH, Buldak GL, Reich CI, Woese CR, Olsen GJ. Thermal adaptation analyzed
587 by comparison of protein sequences from mesophilic and extremely thermophilic Methanococcus
588 species. Proc Natl Acad Sci U S A. 1999;96:3578–83. https://doi.org/10.1073/pnas.96.7.3578.
589 [49] Zeldovich KB, Berezovsky IN, Shakhnovich EI. Protein and DNA sequence determinants of
590 thermophilic adaptation. PLoS Comput Biol. 2007;3:e5. https://doi.org/10.1371/journal.pcbi.0030005.
591 [50] Závodszky P, Kardos J, Svingor Á, Petsko GA. Adjustment of conformational flexibility is a key
592 event in the thermal adaptation of proteins. Proc Natl Acad Sci U S A. 1998;95:7406–11.
593 https://doi.org/10.1073/pnas.95.13.7406.
594 [51] Hobbs JK, Jiao W, Easter AD, Parker EJ, Schipper LA, Arcus VL. Change in heat capacity for
595 enzyme catalysis determines temperature dependence of enzyme catalyzed rates. ACS Chem Biol.
596 2013;8:2388–93. https://doi.org/10.1021/cb4005029.
597 [52] Prentice EJ, Hicks J, Ballerstedt H, Blank LM, Liang L, Schipper LA, Arcus VL. The inflection
598 point hypothesis: the relationship between the temperature dependence of enzyme-catalyzed reaction
599 rates and microbial growth rates. Biochemistry. 2020;59:3562–9.
600 https://doi.org/10.1021/acs.biochem.0c00530.
601 [53] Kawamoto H, Watanabe M, Mochimaru H, Nakahara N, Meng X-Y, Sakamoto S, Morinaga K,
602 Katayama T, Yoshioka H, Nomura N, Tamaki H. Atrimonas thermophila gen. nov., sp. nov., a novel
603 anaerobic thermophilic bacterium of the phylum Atribacterota isolated from deep subsurface gas field
604 and proposal of Atrimonadaceae fam. nov. within the class Atribacteria in the phylum Atribacterota. Syst
605 Appl Microbiol. 2024;47:126515. https://doi.org/10.1016/j.syapm.2024.126515.
606 [54] Saltykova VA, Danilova OV, Oshkin IY, Belova SE, Suzina NE, Pimenov NV, Dedysh SN.
607 Methyloraptor flagellatus gen. nov., sp. nov., novel Ancalomicrobiaceae-affiliated facultatively
608 methylotrophic bacteria that feed on methanotrophs of the genus Methylococcus. Syst Appl Microbiol.
609 2025;48:126565. https://doi.org/10.1016/j.syapm.2024.126565.
610 [55] Franzmann PD, Springer N, Ludwig W, Conway De Macario E, Rohde M. A methanogenic
611 archaeon from Ace Lake, Antarctica: Methanococcoides burtonii sp. nov. Syst Appl Microbiol.
612 1992;15:573–81. https://doi.org/10.1016/S0723-2020(11)80117-7.
613 [56] Copley SD, Novak WRP, Babbitt PC. Divergence of function in the thioredoxin fold suprafamily:

614 evidence for evolution of peroxiredoxins from a thioredoxin-like ancestor. Biochemistry.
615 2004;43:13981–95. https://doi.org/10.1021/bi048947r.
616 [57] Sakaguchi R, Fujiyoshi S, Wakai S, Yamanaka M, Sambongi Y. Thermal destabilization mechanism
617 of cytochrome c′ from psychrophilic Shewanella violacea. Biosci Biotechnol Biochem. 2021;85:1121–
618 7. https://doi.org/10.1093/bbb/zbab007.
619 [58] Suzuki Y, Mizutani Y, Tsuji T, Ohtani N, Takano K, Haruki M, Morikawa M, Kanaya S. Gene
620 cloning, overproduction, and characterization of thermolabile alkaline phosphatase from a
621 psychrotrophic bacterium. Biosci Biotechnol Biochem. 2005;69:364–73.
622 https://doi.org/10.1271/bbb.69.364.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PTACSupplementaryMaterials.docx