PatchET: Learning Enzyme Temperature Properties Through Patch-Based Neural Architectures

Ziqi Zhang[1] Runze Yang[2, 3], Longbing Cao[2], Zhaohong Deng[1*]

[1]School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, 214122, Jiangsu, China
[2]School of Computing, Macquarie University, Sydney, 2113, NSW, Australia
[3]The School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai, 200240, China
7233115021@stu.jiangnan.edu.cn, runze.yang@hdr.mq.edu.au, longbing.cao@mq.edu.au, dengzhaohong@jiangnan.edu.cn

Abstract

Understanding enzyme thermal properties is essential for biotechnology and protein engineering, yet experimental measurements of attributes such as temperature optimum, stability, and range remain labor-intensive and costly. Prior studies have shown that specific regions within enzyme sequences disproportionately influence thermal behavior—an aspect often overlooked by existing deep learning models. In this work, we introduce **PatchET**, a biologically inspired deep learning model that predicts enzyme thermal properties directly from amino acid sequences. PatchET employs a dual-stage, patch-based architecture that captures both **intra-patch** local features and **inter-patch** global dependencies, reflecting the hierarchical nature of protein thermal adaptation. Alongside the model, we curate a comprehensive benchmark, including a refined dataset for temperature optimum and the first publicly available dataset for temperature range prediction. PatchET achieves state-of-the-art performance across three key tasks—temperature optimum, stability, and range—and serves as the first dedicated model for temperature range prediction. Extensive ablation studies further validate the effectiveness of our architectural design. Together, PatchET and the accompanying benchmark provide a unified and generalizable framework for modeling enzyme thermal properties, offering new tools for the rational design of thermostable enzymes.

Code and Datasets — https://github.com/RIA-lab/PatchET

Introduction

An enzyme's thermal properties—namely, its temperature optimum, stability, and operational range—are key determinants of its suitability for industrial applications (Daniel, Danson, and Eisenthal 2001; Daniel et al. 2008; Daniel and Danson 2010; Lee, Monk, and Daniel 2013). The temperature optimum defines the point of maximum catalytic activity; temperature stability reflects the enzyme's ability to maintain structural integrity and function under heat stress; and the operational range characterizes the span of temperatures over which the enzyme remains catalytically active. To identify or engineer enzymes tailored to specific thermal environments, strategies such as directed evolution, rational design, and semi-rational design are commonly employed. However, these approaches rely on accurate determination of enzyme thermal properties (Xu et al. 2020), which remains a labor-intensive, time-consuming, and costly experimental task.

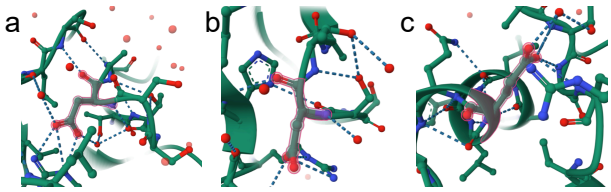

Figure 1: Structure of Aqualysin I (PDB ID: 4DZT), highlighting three key residues: Asp183 (a), Asp212 (b), and Glu237 (c).

Recent studies demonstrate that deep learning holds great potential for analyzing enzyme thermal properties. For example, TOMER (Li et al. 2019; Gado, Beckham, and Payne 2020), Preoptem (Zhang et al. 2022), DeepET (Li et al. 2022), and Seq2Topt (Qiu et al. 2025) have addressed the temperature optimum prediction task. These models typically rely on representations extracted directly from pretrained protein language models (Zhang et al. 2025c,b). However, such approaches treat the enzyme sequence as a homogeneous whole, overlooking the fact that different regions contribute unequally to thermal properties. As a result, they often lack the precision required for high-stakes industrial or therapeutic applications. Biological evidence (Sakaguchi et al. 2014; Chronopoulou et al. 2023; Lee et al. 2014; Kannan and Vishveshwara 2000) suggests that enzyme thermal behavior is shaped by both localized structural motifs and long-range cooperative interactions across the sequence. Figure 1 illustrates Aqualysin I as an example: a thermophilic subtilisin-like protease stabilized by a sparse but functionally critical network of salt bridges. Three residues—Asp183, Asp212, and Glu237—play pivotal roles in this stabilization. Asp183 forms a buried salt bridge with Arg12, and its mutation results in marked thermal instability and secondary structure disruption. Asp212 and Glu237 contribute to peripheral and C-terminal stabilization via electrostatic interactions. These residues often reside within the same local region or engage in long-range interactions with other residues, suggesting that models ca-

*Corresponding author

pable of learning both intra-region specificity and inter-region cooperativity can better capture such functional patterns.

Motivated by these biological insights, we introduce **PatchET**, a biologically inspired deep learning model for predicting enzyme thermal properties directly from amino acid sequences using a patching approach. PatchET employs a novel two-stage architecture that explicitly models both local and global contributions. In the first stage, the **Intra-patch Backbone** segments the sequence into local patches and applies self-attention independently within each patch, capturing fine-grained features such as structural motifs and stabilizing subregions. In the second stage, the **Inter-patch Backbone** models cross-patch interactions using convolution and attention mechanisms, enabling the model to learn long-range dependencies, such as domain–domain cooperativity and distributed stability networks. This patch-based paradigm enhances interpretability while significantly improving predictive performance.

Despite the promise of deep learning, progress has been hindered by data scarcity and the lack of efficient evaluation methods. There is a significant gap between the vast number of known enzyme sequences and the limited availability of annotated temperature data (Rigoldi et al. 2018). While the UniProt database contains over 230 million enzyme sequences, the BRENDA database includes fewer than 80,000 experimentally determined temperature values (Consortium 2024; Chang et al. 2020). This scarcity poses a substantial barrier to model training.

To address this, we establish a robust benchmark for enzyme thermal property analysis by constructing and releasing two curated datasets derived from the BRENDA database. First, we introduce a significantly expanded and re-partitioned dataset for temperature optimum prediction. Unlike the existing dataset (n=2,917) used by TOMER and other models, which covers only EC classes 1 to 6, our new dataset includes 10,371 enzymes spanning all EC classes (1–7). It follows a rigorous partitioning strategy based on the protocol by Gado et al. (Gado et al. 2025), incorporating MMseqs2-based clustering (Steinegger and Söding 2017) to minimize sequence similarity between training and validation subsets. Our second curated dataset enables enzyme temperature range prediction, a previously neglected yet practically important task. These datasets, with high-quality annotations and strict evaluation protocols, provide a robust benchmark for reproducible comparisons of enzyme thermal properties, including temperature range and stability. Validated across all benchmark tasks using our newly constructed datasets, PatchET achieves state-of-the-art performance and establishes itself as a comprehensive and interpretable framework for enzyme thermal profiling and design.

Related Work

Computational Methods for Enzyme Thermal Property Prediction

Numerous computational methods have been developed to predict enzyme thermal properties. Some focus on narrow enzyme classes (Yan and Wu 2012, 2019; Zhang and Ge 2012; Chu et al. 2016; Foroozandeh Shahraki et al. 2021), achieving high performance within those domains but lacking generalizability across diverse enzyme families. Broader, class-agnostic predictors have also been proposed. However, many rely on auxiliary biological metadata, such as the optimal growth temperature (OGT) of the source organism (Li et al. 2019; Gado, Beckham, and Payne 2020; Wang et al. 2024), which limits their applicability when such metadata is unavailable.

To address these limitations, recent studies have developed OGT-independent, sequence-only models (Li et al. 2022; Zhang et al. 2022; Qiu et al. 2024). Representative tools include TOMER, Preoptem, DeepET, and Seq2Topt, which leverage embeddings from pre-trained protein language models to predict temperature optima. While these models represent significant progress, they often fail to capture the nuanced, region-specific contributions to thermal properties, resulting in limited predictive accuracy for industrial or therapeutic applications. Their core limitation lies in treating enzyme sequences as monolithic entities, applying global feature extraction without explicitly modeling local patterns or regional contributions (Zhang et al. 2025c,b).

In contrast, our proposed model, PatchET, diverges from this approach by incorporating domain knowledge into its architecture. It adopts a biologically informed, patch-aware design that captures both localized and distributed sequence effects, addressing a critical gap in prior sequence-based methods.

Benchmark Datasets for Thermal Property Prediction

Existing models for temperature optimum prediction rely on a single publicly available dataset (n=2,917), originally compiled from the BRENDA database for the TOMER study. This dataset has two major limitations: its limited size and scope (covering only EC classes 1–6) and a random partitioning strategy that may introduce sequence redundancy across training and test sets.

To enable more rigorous evaluation and broader enzyme coverage, we introduce a new benchmark dataset comprising 10,371 enzymes across all EC classes (1–7). This dataset employs a robust partitioning strategy using temperature stratification and MMseqs2-based clustering to minimize sequence similarity between subsets. Additionally, we propose the **first benchmark dataset for enzyme temperature range prediction**, addressing a critical gap in the literature where no prior dataset or method has tackled this task. These datasets, subjected to rigorous quality control and deduplication, provide a foundation for reproducible evaluation and future model development.

Patch-Based Modeling in Deep Learning

Patch-based learning has proven highly effective in various deep learning domains for capturing local semantics. In natural language processing, BERT leverages subword-based tokenization to focus on semantically meaningful units (Devlin et al. 2019; Schuster and Nakajima 2012). In computer vision, Vision Transformers (ViT) split images into

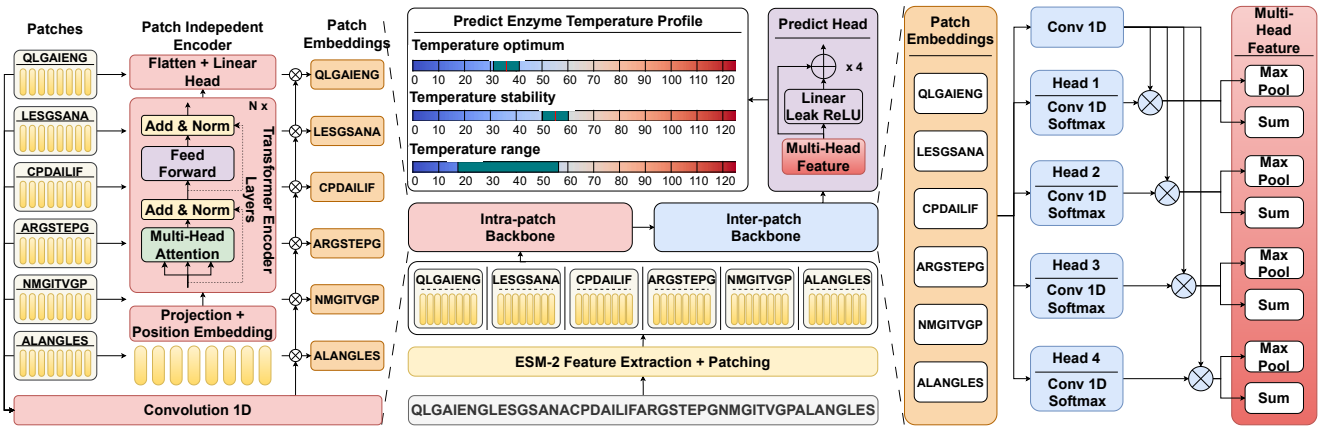

Figure 2: **The architecture of PatchET.** An ESM-2 pre-trained model is used for initial feature extraction. The extracted features are then segmented into patches. The Intra-patch Backbone models local interactions within each patch independently using a Transformer-based encoder. Subsequently, the Inter-patch Backbone captures inter-patch interactions through multi-head 1D convolution and attention mechanisms.

fixed-size patches to model local spatial patterns effectively (Dosovitskiy et al. 2021), inspiring extensions such as BEiT and Masked Autoencoders (MAE) (Bao et al. 2022). Patch-based modeling is also prevalent in speech and time-series domains; for example, Wav2Vec 2.0 employs temporal convolutions to extract local acoustic features (Baevski et al. 2020), while PatchTST utilizes patch-wise self-attention to capture short-range dynamics in time-series data (Nie et al. 2022).

These successes highlight the value of patching for extracting interpretable, localized features, motivating its application to protein sequence modeling, where thermal properties are often governed by specific subregions.

Methods

Biological Motivation and Model Overview

Enzyme thermal behavior is shaped by both localized structural motifs and long-range cooperative interactions across the protein sequence. Specific subregions—such as hydrophobic cores, salt bridges, and active sites—disproportionately influence thermostability and optimal activity. However, existing deep learning models often treat enzyme sequences as holistic inputs, failing to account for the unequal contributions of different regions.

To better reflect biological reality, we designed **PatchET**, a patch-aware architecture that explicitly models local and global contributions to enzyme thermal properties. This design draws inspiration from the success of patch-based models in domains such as computer vision and time-series forecasting, tailored to the unique biological context of enzyme function.

As illustrated in Figure 2, PatchET consists of two main stages:

- **Intra-patch Backbone:** The sequence embedding, extracted from a pre-trained protein language model, is divided into non-overlapping patches. Each patch is processed independently via a Transformer-based encoder to capture localized interactions and motif-level features.
- **Inter-patch Backbone:** Patch-level representations are processed through convolutional and attention layers to model global dependencies and distributed thermal determinants across the entire sequence.

The model outputs a scalar value representing the predicted thermal property (temperature optimum, stability, or range), depending on the task. This two-stage structure enables PatchET to capture the heterogeneous contributions of different sequence regions and achieve superior generalization across diverse enzyme families.

Initial Sequence Feature Extraction and Patching

Given an amino acid sequence, we obtain its contextual embeddings using a pre-trained language model (e.g., ESM-2) as follows:

$$\mathbf{H} = \mathrm{ESM}(\mathbf{x}) \in \mathbb{R}^{L \times d}, \tag{1}$$

where L is the sequence length and d is the hidden dimensionality (e.g., 640).

We then divide $\mathbf{H}$ into non-overlapping patches of length p, resulting in:

$$\mathbf{H}_{\text{patch}} = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N\}, \quad \mathbf{h}_i \in \mathbb{R}^{p \times d}, \tag{2}$$

where $N = \left\lfloor \frac{L}{p} \right\rfloor$ is the number of patches.

Intra-patch Backbone: Modeling Local Dependencies

To model local interactions within each patch, we apply an intra-patch attention encoder. Each patch $\mathbf{h}_i$ is processed through a Transformer-based encoder layer that includes multi-head self-attention and feed-forward sublayers:

$$\mathbf{z}_i = \mathrm{IntraEncoder}(\mathbf{h}_i) \in \mathbb{R}^{p \times d'}, \tag{3}$$

where $\mathbf{z}_i$ denotes the refined representation of patch i. Positional encoding is added before attention to preserve sequence order.

After encoding, each patch representation is aggregated and passed through a linear projection to obtain the intra-patch attention score:

$$\boldsymbol{\alpha}_i = \mathrm{softmax}(\mathrm{PatchLinear}(\mathbf{z}_i)) \in \mathbb{R}^{d'}, \tag{4}$$

which is used to reweight patch features from a parallel convolutional projection $\mathbf{v}_i$:

$$\tilde{\mathbf{z}}_i = \boldsymbol{\alpha}_i \odot \mathbf{v}_i, \tag{5}$$

where $\odot$ denotes element-wise multiplication.

Inter-patch Backbone: Modeling Global Interactions

To capture inter-patch interactions, the sequence of patch-level features $\{\tilde{\mathbf{z}}_1, \ldots, \tilde{\mathbf{z}}_N\}$ is processed using a 1D convolution:

$$\mathbf{Z}_{\text{inter}} = \mathrm{Conv1D}([\tilde{\mathbf{z}}_1, \ldots, \tilde{\mathbf{z}}_N]) \in \mathbb{R}^{N \times d'}, \tag{6}$$

followed by attention-weighted aggregations. For each attention head h, attention weights are computed as:

$$\mathbf{w}^{(h)} = \mathrm{softmax}(\mathrm{Conv1D}^{(h)}(\mathbf{Z}_{\text{inter}})), \tag{7}$$

and used to produce weighted sum and max pooling:

$$\mathbf{s}^{(h)} = \sum_{i=1}^{N} \mathbf{w}_i^{(h)} \cdot \mathbf{Z}_{\text{inter},i}, \tag{8}$$

$$\mathbf{m}^{(h)} = \max_i \left(\mathbf{w}_i^{(h)} \cdot \mathbf{Z}_{\text{inter},i} \right). \tag{9}$$

The final inter-patch feature vector is obtained by concatenating all heads:

$$\mathbf{z}_{\text{final}} = [\mathbf{s}^{(1)}; \mathbf{m}^{(1)}; \ldots; \mathbf{s}^{(H)}; \mathbf{m}^{(H)}], \tag{10}$$

where H is the number of inter-patch attention heads.

Prediction

The combined feature vector is passed through a sequence of residual dense blocks (RDBlocks) and a final regression head to predict the temperature optimum:

$$\mathbf{z}_{\text{pred}} = \mathrm{RDBlock}_n(\ldots \mathrm{RDBlock}_1(\mathbf{z}_{\text{final}})), \tag{11}$$

$$\hat{y} = \mathbf{w}^\top \mathbf{z}_{\text{pred}} + b. \tag{12}$$

Benchmark

Dataset Construction

The datasets for temperature optimum and temperature range prediction were sourced from the BRENDA enzyme functional database (accessed in December 2024) (Consortium 2024; Chang et al. 2020), initially comprising 40,732 and 6,632 entries, respectively. To ensure data quality and consistency, the following preprocessing steps were applied:

1. Samples lacking UniProt accession numbers or valid numerical annotations were excluded.
2. Entries associated with multiple accession numbers were split, with each instance corresponding to a single accession number and one thermal property.
3. Duplicate entries with the same accession number were removed to eliminate redundancy.

After preprocessing, the resulting datasets contained 10,371 samples for temperature optimum prediction and 1,818 samples for temperature range prediction.

Dataset	Partition			Total
	Train	Validation	Test	
Temperature optimum	8,347	987	1,037	10,371
Temperature range	1,464	172	182	1,818
Temperature stability	2,798	311	345	3,454

Table 1: Dataset statistics for each task, including partition sizes.

Dataset partitioning. 10% of the data were randomly held out as an independent test set. The remaining 90% were used for training and validation, structured through clustering and stratification procedures to minimize sequence similarity between splits. For the temperature optimum dataset, sequences were first grouped by temperature intervals. Within each group, MMseqs2 was used to cluster sequences based on similarity. Clusters were then randomly split, with 90% allocated to training and 10% to validation. This approach reduces sequence similarity between training and evaluation sets, promoting robust generalization. For the temperature range dataset, clustering was applied directly without temperature-based grouping due to its smaller size. The temperature stability dataset was sourced from previous work by Zhang et al. (Zhang et al. 2025a). A summary of dataset sizes and partitions is provided in Table 1.

Loss Function

We employed two loss functions for training PatchET, each tailored to the prediction task. For scalar temperature prediction tasks, such as temperature optimum and stability, we used a **weighted root mean squared error** loss. This formulation emphasizes specific temperature intervals by assigning custom weights. The loss is defined as:

$$\mathcal{L} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} w_i (\hat{y}_i - y_i)^2}, \tag{13}$$

where w_i is the weight assigned to sample i.

For temperature range prediction, we designed a custom **temperature range loss**. This loss penalizes deviations from the true interval boundaries and enforces the constraint that the lower bound must not exceed the upper bound. The loss is computed as:

$$\begin{aligned} \mathcal{L}_{\text{range}} = \mathrm{MSE}(\hat{y}_{\text{low}}, y_{\text{low}}) + \mathrm{MSE}(\hat{y}_{\text{high}}, y_{\text{high}}) \\ + \lambda \cdot \max(0, \hat{y}_{\text{low}} - \hat{y}_{\text{high}}), \end{aligned} \tag{14}$$

where $\hat{y}_i = [\hat{y}_{i,\text{low}}, \hat{y}_{i,\text{high}}]$ and $y_i = [y_{i,\text{low}}, y_{i,\text{high}}]$ denote the predicted and true temperature ranges for the i-th sample, and λ is a penalty coefficient enforcing $\hat{y}_{\text{low}} \leq \hat{y}_{\text{high}}$. In our experiments, we set $\lambda = 1.0$.

Model	Low Boundary				High Boundary				Mean Overlap Ratio
	Pearson	Spearman	MAE	RMSE	Pearson	Spearman	MAE	RMSE	
BiLSTM	0.470	**0.471**	11.53	14.96	0.402	0.437	15.50	19.41	0.565
CNN	0.469	0.409	11.72	15.26	0.478	0.431	14.00	17.07	0.650
Transformer	0.472	0.427	13.28	16.85	0.360	0.306	16.49	20.24	0.495
Light Attention	0.041	0.065	12.96	17.16	0.108	0.109	16.42	20.06	0.598
RNN	0.468	0.446	12.29	16.02	0.512	**0.520**	15.57	19.86	0.578
DeepET	0.471	0.468	18.19	23.39	0.433	0.403	40.67	44.81	0.059
Seq2Topt	0.455	0.396	11.64	15.60	0.484	0.474	13.83	17.70	0.683
TemStaPro	0.354	0.300	12.43	15.81	0.343	0.291	15.53	18.60	0.559
PatchET	**0.473**	0.418	**11.28**	**14.87**	**0.514**	0.470	**13.53**	**17.05**	**0.707**

Table 2: Performance of all models on the temperature range dataset, evaluated for both range boundaries and mean overlap ratio.

Evaluation Metrics

Scalar Temperature Prediction. We evaluated scalar predictions (e.g., temperature optimum and stability) using four standard metrics: Pearson and Spearman correlation coefficients, mean absolute error (MAE), and root mean squared error (RMSE). Given N samples with predictions $\hat{y}_i$ and ground truth y_i, the metrics are:

$$\text{Pearson} = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}, \quad (15)$$

$$\text{Spearman} = \text{Pearson}(\text{rank}(\hat{y}_i), \text{rank}(y_i)), \quad (16)$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|, \quad (17)$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}. \quad (18)$$

Interval Prediction. For temperature range prediction, we computed the *mean overlap ratio* between predicted $[T^{\text{pred}}_{\text{low},i}, T^{\text{pred}}_{\text{high},i}]$ and true $[T^{\text{true}}_{\text{low},i}, T^{\text{true}}_{\text{high},i}]$ intervals:

$$\text{Intersection}_i = \max\Big(0,\ \min(T^{\text{pred}}_{\text{high},i},\ T^{\text{true}}_{\text{high},i}) - \max(T^{\text{pred}}_{\text{low},i},\ T^{\text{true}}_{\text{low},i})\Big) \quad (19)$$

$$\text{Length}^{\text{true}}_i = T^{\text{true}}_{\text{high},i} - T^{\text{true}}_{\text{low},i} \quad (20)$$

$$\text{OverlapRatio}_i = \begin{cases} \frac{\text{Intersection}_i}{\text{Length}^{\text{true}}_i}, & \text{if Length}^{\text{true}}_i > 0, \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

$$\text{MeanOverlapRatio} = \frac{1}{N}\sum_{i=1}^{N}\text{OverlapRatio}_i \quad (22)$$

We also evaluated the accuracy of interval boundaries using MAE, RMSE, Pearson, and Spearman metrics for the predicted lower and upper bounds separately.

Experiments

We evaluate the performance of PatchET against five standard deep learning models (BiLSTM, CNN, Transformer, Light Attention, RNN) and three state-of-the-art enzyme temperature prediction methods (TemStaPro, Seq2Topt, and DeepET) on three benchmark datasets (Schuster and Paliwal 1997; Collobert et al. 2011; Vaswani et al. 2017; Stärk et al. 2021; Elman 1990; Pudžiuvelytė et al. 2024; Qiu et al. 2025; Li et al. 2022). All experiments were conducted on a machine running Ubuntu 22.04.5 LTS with Python 3.12.7 and PyTorch 2.5.1. The hardware environment included an NVIDIA H100 GPU with CUDA version 12.2.

Results on Temperature Range Dataset

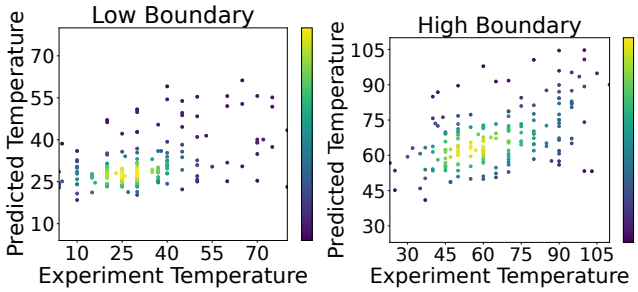

Figure 3: **Scatter plots of predicted versus actual values** for PatchET on low and high boundaries.

Table 2 presents the performance of all models on the temperature range dataset, evaluating predictions for both low and high temperature boundaries and the mean overlap ratio between predicted and true intervals. PatchET achieves the best overall performance, with the lowest MAE and RMSE for both boundaries (11.28 and 14.87 for low; 13.53 and 17.05 for high) and the highest mean overlap ratio (0.707), demonstrating superior accuracy in interval prediction.

Seq2Topt ranks second, with strong results, particularly for high-boundary prediction (MAE = 13.83) and a mean overlap ratio of 0.683. Traditional models, such as BiLSTM, CNN, and RNN, show moderate accuracy. CNN performs

slightly better on the high boundary, while RNN exhibits stronger correlation metrics. The Transformer model performs relatively well on the low boundary but poorly on the high boundary, resulting in a low mean overlap ratio (0.495). The scatter plots in Figure 3 confirm PatchET's superior accuracy and consistent coverage in predicting both boundaries, establishing it as the most precise model for temperature range estimation.

Results on Temperature Optimum Dataset

Model	Overall			
	Pearson	Spearman	MAE	RMSE
BiLSTM	0.449	0.347	12.18	17.32
CNN	0.511	0.418	11.68	15.51
Transformer	0.521	0.416	12.38	16.72
Light Attention	0.515	0.413	11.59	15.45
RNN	0.459	0.358	13.51	18.20
DeepET	0.450	0.382	11.79	14.96
Seq2Topt	0.526	0.423	11.59	15.67
TemStaPro	0.501	0.414	11.79	15.54
PatchET	**0.551**	**0.458**	**9.69**	**13.88**

Table 3: Performance on the temperature optimum dataset. Best results are bolded.

As shown in Table 3, PatchET achieves the best performance across all evaluation metrics on the temperature optimum dataset, with the highest Pearson (0.551) and Spearman (0.458) correlations and the lowest RMSE (13.88) and MAE (9.69), outperforming all baseline models.

Compared to the previous best method, Seq2Topt, PatchET reduces RMSE by over 2 units and MAE by nearly 2 points, indicating significant improvements in predictive accuracy. Conventional architectures like BiLSTM and RNN exhibit higher errors, while advanced attention-based models, such as Transformer and Light Attention, are surpassed by PatchET in both correlation and error metrics.

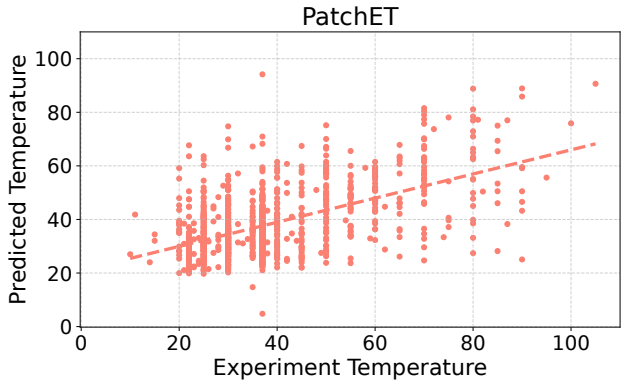

Figure 4: **Scatter plots of predicted versus actual values** for PatchET.

These results highlight the effectiveness of PatchET's patch-based design, which enhances its ability to capture complex sequence–temperature relationships. The consistent improvements over diverse baselines confirm PatchET's robustness and suitability for temperature optimum prediction. Additional breakdowns by enzyme class are provided in the Appendix. In addition, Figure 4 shows scatter plots comparing predicted versus actual temperature values for PatchET. Full scatter plots for all models are available in the Appendix.

Results on Temperature Stability Dataset

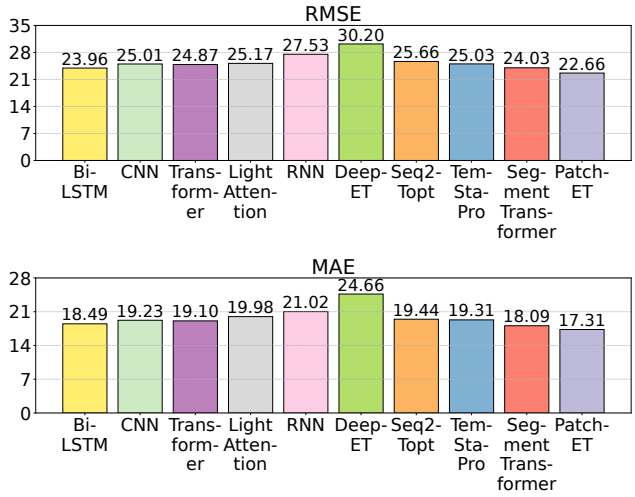

Figure 5: RMSE and MAE of PatchET and baseline models on the temperature stability dataset.

Figure 5 reports the RMSE and MAE for all models on the temperature stability dataset. PatchET achieves the lowest RMSE (22.66) and MAE (17.31), clearly outperforming all baselines. Segment-Transformer ranks second (RMSE = 24.03, MAE = 18.09), followed by BiLSTM (RMSE = 23.96, MAE = 18.49), both showing reasonable generalization. Other models, including CNN, Transformer, Seq2Topt, and TemStaPro, exhibit comparable but less competitive performance (RMSE around 25, MAE around 19). Light Attention and RNN yield slightly higher errors. DeepET performs the worst (RMSE = 30.20, MAE = 24.66), indicating limited capability for modeling temperature stability. Overall, PatchET demonstrates robust generalization and superior predictive power across all metrics for the temperature stability task.

Ablation Study

To evaluate the contributions of PatchET's core components, we conducted an ablation study on the temperature optimum prediction task (Table 4). Specifically, we varied the intra-patch length, which defines the local region processed independently, and the inter-patch kernel size, which controls the convolutional receptive field across patches.

The results demonstrate that both parameters significantly influence performance. Increasing the intra-patch length from 5 to 20 leads to consistent improvements, with the best

performance achieved at a patch length of 25 and a kernel size of 5 (RMSE = 13.884, MAE = 9.693). Further increasing the patch length to 25 or 50 yields marginal or slightly degraded results, suggesting diminishing returns from excessive patch aggregation.

Intra	Inter	Metric			
Length	Size	RMSE	MAE	Pearson	Spearman
5	3	15.339	11.298	0.511	0.401
	5	15.278	10.562	0.526	0.402
	7	14.450	10.449	0.533	0.440
10	3	14.478	10.465	0.517	0.435
	5	14.201	10.230	0.554	0.450
	7	14.917	10.706	0.534	0.401
20	3	13.925	9.975	0.562	0.442
	5	14.694	10.847	0.535	0.428
	7	14.143	10.151	0.546	0.447
25	3	13.991	9.888	0.556	0.456
	5	13.884	9.693	0.551	0.458
	7	13.813	9.902	0.550	0.442
50	3	14.052	10.150	0.533	0.424
	5	13.646	9.817	0.558	0.449
	7	13.873	9.757	0.553	0.462
w/o Intra-patch		17.139	13.025	0.493	0.397
w/o Inter-patch		15.508	11.111	0.407	0.337

Table 4: Ablation study on intra-patch length and inter-patch kernel size for temperature optimum prediction.

A similar trend is observed for the inter-patch kernel size. A kernel size of 5 consistently outperforms smaller (3) and larger (7) sizes across most patch lengths, indicating the importance of a balanced receptive field for inter-patch communication.

To further isolate the effects of each module, we examined two model variants: one without the intra-patch module and one without the inter-patch module. Removing intra-patch modeling causes a significant performance drop (RMSE = 17.139), while removing inter-patch modeling also leads to substantial degradation (RMSE = 15.508, Pearson = 0.407). These results underscore the necessity of both local and global sequence modeling for accurate thermal prediction.

Number of Heads	RMSE	MAE	Pearson	Spearman
1	13.922	10.104	0.545	0.433
2	13.921	9.948	0.541	0.438
3	13.896	9.895	0.551	0.428
4	13.884	9.693	0.551	0.458
5	13.896	9.734	0.550	0.451

Table 5: Effect of varying the number of attention heads in the inter-patch backbone. Arrows indicate direction of improvement.

We also assessed the impact of varying the number of attention heads in the inter-patch backbone (Table 5). Multi-head attention enables the model to learn diverse relational patterns among patches by attending to multiple subspaces.

Increasing the number of heads from 1 to 4 leads to progressive improvements across all metrics, with the best performance achieved at 4 heads (RMSE = 13.884, MAE = 9.696, Pearson = 0.551, Spearman = 0.458). Adding a fifth head slightly reduces performance, suggesting potential over-fragmentation and diminished expressiveness. These results highlight the importance of tuning the number of attention heads to optimize the balance between model capacity and stability.

Overall, our ablation studies confirm the critical roles of intra- and inter-patch modeling in PatchET and demonstrate the performance benefits of carefully selecting architectural hyperparameters.

Conclusion

In this study, we introduced **PatchET**, a biologically inspired deep learning model for predicting enzyme thermal properties directly from amino acid sequences. Motivated by evidence that enzyme thermal behavior is governed by both localized structural motifs and long-range interactions, PatchET employs a novel patch-based architecture that explicitly captures **intra-patch** local patterns and **inter-patch** global dependencies. This design reflects the hierarchical and region-specific nature of protein thermal adaptation, enabling the model to identify critical sequence features often overlooked by prior approaches.

Beyond architectural innovation, we addressed a critical need for standardized evaluation in enzyme thermal property prediction. We curated and released a comprehensive benchmark suite, including a refined dataset for temperature optimum prediction and the first publicly available dataset for temperature range prediction—an underexplored yet biologically significant property.

Extensive experiments across three tasks—temperature optimum, stability, and range—demonstrate that PatchET achieves state-of-the-art performance across all evaluation metrics. Ablation studies validate the critical contributions of both intra- and inter-patch components, confirming their synergistic roles in enhancing predictive accuracy. Notably, PatchET is the first model specifically designed for temperature range prediction, significantly expanding the scope of enzyme thermal modeling and addressing a previously neglected aspect of enzyme functionality.

In summary, PatchET advances the field by integrating biological insight with computational innovation. Its patch-based design and accompanying benchmark resources provide a robust and generalizable framework for enzyme thermal property prediction, supporting the rational design of thermostable enzymes for industrial and biomedical applications.

Acknowledgments

We acknowledge financial support from the National Natural Science Foundation of China (62176105), the China Scholarship Council (202406790100).

References

Baevski, A.; Zhou, Y.; Mohamed, A.; and Auli, M. 2020. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, 33: 12449–12460.

Bao, H.; Dong, L.; Piao, S.; and Wei, F. 2022. BEiT: BERT Pre-Training of Image Transformers. In *International Conference on Learning Representations*.

Chang, A.; Jeske, L.; Ulbrich, S.; Hofmann, J.; Koblitz, J.; Schomburg, I.; Neumann-Schaal, M.; Jahn, D.; and Schomburg, D. 2020. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Research*, 49(D1): D498–D508.

Chronopoulou, E. G.; Mutabdzija, L.; Poudel, N.; Papageorgiou, A. C.; and Labrou, N. E. 2023. A Key Role in Catalysis and Enzyme Thermostability of a Conserved Helix H5 Motif of Human Glutathione Transferase A1-1. *International journal of molecular sciences*, 24(4): 3700.

Chu, Y.; Yi, Z.; Zeng, R.; and Zhang, G. 2016. Predicting the optimum temperature of β-agarase based on the relative solvent accessibility of amino acids. *Journal of Molecular Catalysis B: Enzymatic*, 129: 47–53.

Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; and Kuksa, P. 2011. Natural language processing (almost) from scratch. *Journal of machine learning research*, 12(7).

Consortium, T. U. 2024. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Research*, 53(D1): D609–D617.

Daniel, R. M.; and Danson, M. J. 2010. A new understanding of how temperature affects the catalytic activity of enzymes. *Trends in biochemical sciences*, 35(10): 584–591.

Daniel, R. M.; Danson, M. J.; and Eisenthal, R. 2001. The temperature optima of enzymes: a new perspective on an old phenomenon. *Trends in biochemical sciences*, 26(4): 223–225.

Daniel, R. M.; Danson, M. J.; Eisenthal, R.; Lee, C. K.; and Peterson, M. E. 2008. The effect of temperature on enzyme activity: new insights and their implications. *Extremophiles*, 12: 51–59.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 4171–4186.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*.

Elman, J. L. 1990. Finding structure in time. *Cognitive science*, 14(2): 179–211.

Foroozandeh Shahraki, M.; Farhadyar, K.; Kavousi, K.; Azarabad, M. H.; Boroomand, A.; Ariaeenejad, S.; and Hosseini Salekdeh, G. 2021. A generalized machine-learning aided method for targeted identification of industrial enzymes from metagenome: A xylanase temperature dependence case study. *Biotechnology and Bioengineering*, 118(2): 759–769.

Gado, J. E.; Beckham, G. T.; and Payne, C. M. 2020. Improving enzyme optimum temperature prediction with resampling strategies and ensemble learning. *Journal of Chemical Information and Modeling*, 60(8): 4098–4107.

Gado, J. E.; Knotts, M.; Shaw, A. Y.; Marks, D.; Gauthier, N. P.; Sander, C.; and Beckham, G. T. 2025. Machine learning prediction of enzyme optimum pH. *Nature Machine Intelligence*, 1–14.

Kannan, N.; and Vishveshwara, S. 2000. Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein engineering*, 13(11): 753–761.

Lee, C. K.; Monk, C. R.; and Daniel, R. M. 2013. Determination of enzyme thermal parameters for rational enzyme engineering and environmental/evolutionary studies. *Protein Nanotechnology: Protocols, Instrumentation, and Applications, Second Edition*, 219–230.

Lee, C.-W.; Wang, H.-J.; Hwang, J.-K.; and Tseng, C.-P. 2014. Protein thermal stability enhancement by designing salt bridges: a combined computational and experimental study. *PloS one*, 9(11): e112751.

Li, G.; Buric, F.; Zrimec, J.; Viknander, S.; Nielsen, J.; Zelezniak, A.; and Engqvist, M. K. 2022. Learning deep representations of enzyme thermal adaptation. *Protein Science*, 31(12): e4480.

Li, G.; Rabe, K. S.; Nielsen, J.; and Engqvist, M. K. 2019. Machine learning applied to predicting microorganism growth temperatures and enzyme catalytic optima. *ACS synthetic biology*, 8(6): 1411–1420.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2022. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *The Eleventh International Conference on Learning Representations*.

Pudžiuvelytė, I.; Olechnovič, K.; Godliauskaite, E.; Sermokas, K.; Urbaitis, T.; Gasiunas, G.; and Kazlauskas, D. 2024. TemStaPro: protein thermostability prediction using sequence representations from protein language models. *Bioinformatics*, 40(4): btae157.

Qiu, S.; Hu, B.; Zhao, J.; Xu, W.; and Yang, A. 2024. Seq2Topt: a sequence-based deep learning predictor of enzyme optimal temperature. *bioRxiv*, 2024–08.

Qiu, S.; Hu, B.; Zhao, J.; Xu, W.; and Yang, A. 2025. Seq2Topt: a sequence-based deep learning predictor of enzyme optimal temperature. *Briefings in Bioinformatics*, 26(2): bbaf114.

Rigoldi, F.; Donini, S.; Redaelli, A.; Parisini, E.; and Gautieri, A. 2018. Engineering of thermostable enzymes for industrial applications. *APL bioengineering*, 2(1).

Sakaguchi, M.; Osaku, K.; Maejima, S.; Ohno, N.; Sugahara, Y.; Oyama, F.; and Kawakita, M. 2014. Highly conserved salt bridge stabilizes a proteinase K subfamily enzyme, Aqualysin I, from Thermus aquaticus YT-1. *AMB Express*, 4(1): 59.

Schuster, M.; and Nakajima, K. 2012. Japanese and korean voice search. In *2012 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 5149–5152. IEEE.

Schuster, M.; and Paliwal, K. K. 1997. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11): 2673–2681.

Stärk, H.; Dallago, C.; Heinzinger, M.; and Rost, B. 2021. Light attention predicts protein location from the language of life. *Bioinformatics Advances*, 1(1): vbab035.

Steinegger, M.; and Söding, J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology*, 35(11): 1026–1028.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, X.; Zong, Y.; Zhou, X.; Xu, L.; He, W.; and Quan, S. 2024. Artificial Intelligence-Powered Construction of a Microbial Optimal Growth Temperature Database and Its Impact on Enzyme Optimal Temperature Prediction. *The Journal of Physical Chemistry B*, 128(10): 2281–2292.

Xu, Z.; Cen, Y.-K.; Zou, S.-P.; Xue, Y.-P.; and Zheng, Y.-G. 2020. Recent advances in the improvement of enzyme thermostability by structure modification. *Critical reviews in biotechnology*, 40(1): 83–98.

Yan, S.; and Wu, G. 2019. Predictors for predicting temperature optimum in beta-glucosidases. *Journal of Biomedical Science and Engineering*, 12(8): 414–426.

Yan, S.-M.; and Wu, G. 2012. Prediction of optimal pH and temperature of cellulases using neural network. *Protein and Peptide Letters*, 19(1): 29–39.

Zhang, G.; and Ge, H. 2012. Prediction of xylanase optimal temperature by support vector regression. *Electronic Journal of Biotechnology*, 15(1): 7–7.

Zhang, Y.; Guan, F.; Xu, G.; Liu, X.; Zhang, Y.; Sun, J.; Yao, B.; Huang, H.; Wu, N.; and Tian, J. 2022. A novel thermophilic chitinase directly mined from the marine metagenome using the deep learning tool Preoptem. *Bioresources and Bioprocessing*, 9(1): 54.

Zhang, Z.; Chen, S.; Yang, R.; Wei, Z.; Zhang, W.; Wang, L.; Liu, Z.; Zhang, F.; Wu, J.; Pan, X.; Shen, H.; Cao, L.; and Deng, Z. 2025a. Modeling enzyme temperature stability from sequence segment perspective. arXiv:2507.19755.

Zhang, Z.; Wei, Z.; Qin, Z.; Wang, L.; Gong, J.; Shi, J.; Wu, J.; and Deng, Z. 2025b. Advancing Enzyme Optimal pH Prediction via Retrieved Embedding Data Augmentation. *Journal of Chemical Information and Modeling*.

Zhang, Z.; Yu, G.; Deng, Z.; Luo, C.; Cai, C.; Zhang, W.; Hu, F.; Choi, K.-S.; Wei, Z.; Wang, L.; et al. 2025c. SEFP: Structure-based Enzyme Function Prediction. *IEEE Transactions on Computational Biology and Bioinformatics*.