



Machine learning-guided multi-site combinatorial mutagenesis enhances the thermostability of pectin lyase

Zhihui Zhang^a, Zhixuan Li^a, Manli Yang^a, Fengguang Zhao^b, Shuangyan Han^{a,*}

^a Guangdong Key Laboratory of Fermentation and Enzyme Engineering, School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China

^b School of Light Industry and Engineering, South China University of Technology, Guangzhou 510006, China

ARTICLE INFO

Keywords:

Protein engineering
Machine learning
Thermostability

ABSTRACT

Enhancing the thermostability of enzymes is crucial for industrial applications. Methods such as directed evolution are often limited by the huge sequence space and combinatorial explosion, making it difficult to obtain optimal mutants. In recent years, machine learning (ML)-guided protein engineering has become an attractive tool because of its ability to comprehensively explore the sequence space of enzymes and discover superior mutants. This study employed ML to perform combinatorial mutation design on the pectin lyase PMGL-Ba from *Bacillus licheniformis*, aiming to improve its thermostability. First, 18 single-point mutants with enhanced thermostability were identified through semi-rational design. Subsequently, the initial library containing a small number of low-order mutants was utilized to construct an ML model to explore the combinatorial sequence space (theoretically 196,608 mutants) of single-point mutants. The results showed that the ML-predicted second library was successfully enriched with highly thermostable combinatorial mutants. After one iteration of learning, the best-performing combinatorial mutant in the third library, P36, showed a 67-fold and 39-fold increase in half-life at 75 °C and 80 °C, respectively, as well as a 2.1-fold increase in activity. Structural analysis and molecular dynamics simulations provided insights into the improved performance of the engineered enzyme.

1. Introduction

Enzymes serve as efficient and environmentally friendly biocatalysts, widely used in industries such as food, chemicals, pharmaceuticals, and environmental protection [1]. Their high efficiency and excellent selectivity make them indispensable in these fields. However, the inherent instability of natural enzymes severely limits their industrial applications, especially when elevated temperatures are required to reduce biomass viscosity, accelerate reactions, and lower production costs [2]. Therefore, enhancing enzyme stability is crucial for industrial applications.

Currently, widely used methods to enhance enzyme stability mainly include directed evolution, rational design, and semi-rational design. Some computational design approaches, such as FireProt [3], PROSS [4], and MutCompute [5], have successfully been used to enhance the thermostability of various enzyme proteins. However, due to the difficulty in exploring huge sequence spaces, these methods often fail to obtain the most desirable mutations [6]. To achieve superior performance, it is usually necessary to introduce multi-site combinatorial

mutations into protein sequences, but the “combinatorial explosion” issue presents significant challenges for rational design and experimental testing [7].

In recent years, ML-guided protein engineering is becoming an attractive methodology, providing a new strategy to enhance enzyme thermostability. By generating an initial library through saturation mutagenesis or random mutagenesis and using the data from this library to build ML models, which can predict the fitness of unknown sequences [8]. Compared to traditional methods like directed evolution, ML can explore huge sequence spaces and highly enrich for superior mutants with the desired functions. Currently, the application of ML in protein engineering mainly focuses on combinatorial saturation mutagenesis at a few sites in the active pocket, with less emphasis on multi-site combinatorial mutagenesis [9–11]. Nevertheless, some studies have shown that ML has significant potential in guiding multi-site combinatorial mutations in proteins. For instance, Cadet et al. successfully predicted the enantioselectivity of 512 mutants of *Aspergillus niger* epoxide hydrolase using the innovSAR method, discovering a combinatorial mutant with an *E*-value nearly 2.2 times higher than the best existing

* Corresponding author.

E-mail address: syhan@scut.edu.cn (S. Han).

<https://doi.org/10.1016/j.ijbiomac.2024.134530>

Received 17 June 2024; Received in revised form 25 July 2024; Accepted 4 August 2024

Available online 5 August 2024

0141-8130/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

mutant [12]. Similarly, Chen et al. demonstrated that ML models can predict the fitness of higher-order combinatorial mutants using data from lower-order mutants [13]. Therefore, studying ML-based multi-site combinatorial mutagenesis is of great significance for protein engineering.

In this study, we introduced ML-guided multi-site combinatorial mutagenesis to enhance the thermostability of pectin lyase (PNL, EC 4.2.2.10) (Fig. 1). PNL can degrade highly esterified pectin polymers via β -elimination reaction without generating harmful methanol [14], which plays a crucial role in industries such as food, textile, and paper [15]. However, its poor thermostability severely limits its industrial applications. We used a novel PNL, PMGL-Ba, previously discovered in our laboratory [16], as the starting enzyme. The ML model was trained to predict the thermostability of higher-order combinatorial mutants using a small amount of low-order mutant data based on obtaining single-point mutants with enhanced thermostability. Compared to the wild type (WT), the thermostability of the ML-predicted combinatorial mutant was significantly enhanced. Our study provides important insights and guidance for designing multi-site combinatorial mutations in protein engineering and developing high thermotolerance enzymes.

2. Material and methods

2.1. Genes, plasmids, and strains

The PMGL-Ba gene (NCBI: WP_009329358.1) was amplified from the pPICZ α A-PMGL-Ba plasmid vector constructed in our laboratory [16]. Then, it was ligated into the pET-28a(+) vector through homologous recombination, transformed into *E. coli* Top10, and the recombinant plasmid pET-28a(+)-PMGL-Ba was obtained after correct sequencing. The construction strain for PMGL-Ba and its mutants was *E. coli* Top10, and the protein expression strain was *E. coli* BL21 (DE3).

2.2. Site-directed mutagenesis

Using the nucleotide sequence of the recombinant plasmid pET-28a(+) - PMGL-Ba from the WT strain as a template, primers with mutation sites were designed (see Table S1). A whole plasmid polymerase chain reaction (PCR) was performed using KOD FX enzyme (TOYOBO, Japan). The PCR conditions were as follows: 94 °C, 5 min; 30 cycles (98 °C, 10 s; 55 °C, 30 s; 68 °C, 3 min 25 s); 68 °C, 10 min. The PCR products were digested with FastDigest *DpnI* endonuclease (Thermo Fisher, USA) at 37 °C for 30 min and then transformed into *E. coli* Top10. The transformants were selected for plasmid extraction and sequencing verification.

2.3. Enzyme expression and purification

The verified recombinant plasmid was transformed into *E. coli* BL21

(DE3) and cultured at 37 °C, 200 rpm. Upon reaching an OD₆₀₀ of 0.6–0.8, IPTG was added at a final concentration of 0.04 mmol/L to induce expression at 16 °C, 200 rpm. After incubation for 16 h, cells were collected by centrifugation (7000 rpm, 5 min) and then resuspended in Tris-HCl buffer (50 mM, pH 8.0). Cells were shattered by ultrasonication at 0 °C for 15 min, followed by centrifugation at 4 °C, 10,000 rpm for 30 min to remove cell debris, and the protein-containing supernatant was collected. WT and mutant proteins were purified using Ni-NTA column chromatography (Cytiva, USA), and the target proteins were eluted with 50 % imidazole buffer (0.3 M). The purified proteins were collected and dialyzed against Tris-HCl buffer (50 mM, pH 8.0) to remove the high concentration of imidazole. Protein detection was performed using SDS-PAGE gel electrophoresis, and protein concentration was determined using the Bradford assay kit (Generay, Shanghai, China).

2.4. Activity assay

The enzyme activity was determined using a 540 nm light absorption method, with pectin (Macklin, Shanghai, China) as the substrate. Take 10 μ L of appropriately diluted enzyme solution, mix it with 190 μ L of preheated 0.5 % pectin solution, and react at 60 °C for 10 min. Add 300 μ L of DNS solution, boil for 5 min, and measure the absorbance change at a wavelength of 540 nm to determine the enzyme activity. Relative activity was calculated by dividing the activity of the mutant by the activity of the WT and then multiplying by 100 %. Residual activity was calculated as a percentage of the initial activity of the corresponding enzyme.

2.5. Thermostability analysis

An equal amount of enzyme solution (100 μ L) was added to the PCR tubes and incubated at different temperatures (65 °C, 70 °C, 75 °C, and 80 °C) for 30 min, then cooled on ice for 5 min. The residual activity was calculated using the method described above. The residual activity of WT and mutant enzymes was measured after incubation at 75 °C and 80 °C for different times using the same method. Relative thermostability was calculated by dividing the residual activity of the mutant by the residual activity of the WT and then multiplying by 100 %. The half-life ($t_{1/2}$) was calculated using the approach described by Zhang et al. [17].

2.6. Feature vector

We used two types of feature vector extraction methods based on various physicochemical properties of amino acids: AAindex [18] and protein descriptors [19]. AAindex has 566 different indices, representing various physicochemical properties of amino acid sequences. Protein descriptors include Amino Acid Composition (AAC), Dipeptide

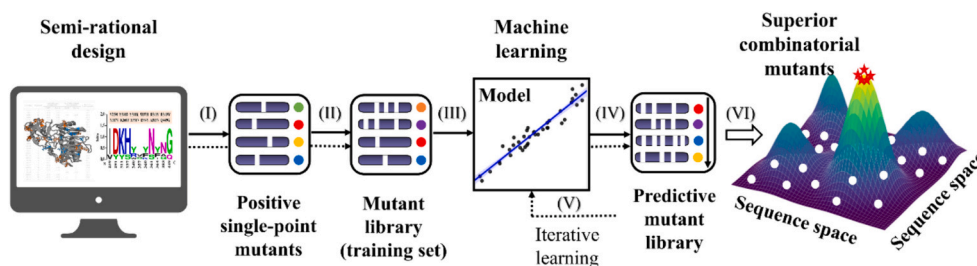


Fig. 1. Workflow of ML-guided multi-site combinatorial mutagenesis to enhance the thermostability of PMGL-Ba. The process includes six steps: (I) Screening thermostable single-point mutants using semi-rational design methods. (II) Constructing a mutant library by randomly combining these thermostable single-point mutants and combining them with the single-point mutants generated in step (I). (III) Building ML models using the data from the mutant library. (IV) Predicting and ranking all possible combinatorial mutants derived from the single-point mutants using the ML model. (V) Feeding the data of validated combinatorial mutants back into the ML model for iterative learning. (VI) Based on the above steps, a comprehensive exploration of the protein sequence space can be achieved, leading to the discovery of superior combinatorial mutants.

Composition (DC), Conjoint Triad (CT), Geary Autocorrelation (GA), Sequence-Order-Coupling Number (SOCN), and Quasi-Sequence-Order (QSO). Table S2 provides a concise overview of the specifics of each feature encoding method. In order to provide the model with richer sequence information, we combined AAindex with two protein descriptors.

2.7. Machine learning model

In this study, we tested six regression algorithms, namely Random Forest (RF) [20], Support Vector Regression (SVR) [21], Gradient Boosting Regression (GBR) [22], Extreme Gradient Boosting (XGB) [23], Adaptive Boosting (Ada) [24], and Bagging (Bag) [25]. The algorithms used in this study were sourced from the Scikit-Learn toolkit [26]. The parameters for all algorithms were used with the default settings from the sklearn models. Based on the combination of different feature vectors described above, a total of 8490 prediction models were established for each algorithm. We performed 10-fold cross-validation to evaluate the performance of the constructed model using the coefficient of determination R^2 and the root mean square error (RMSE) as evaluation metrics [27,28]. These metrics were calculated by the following formulas, respectively:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

where, y_i is the true value, $\hat{y}_i$ is the predicted value, $\bar{y}$ represents the mean of all actual observations, and n represents the size of the sample.

2.8. Enzymatic properties analysis

The optimal temperature was measured at pH 8.0 within a temperature range of 45–70 °C. The optimal pH was measured at 60 °C within a pH range of 6.5–10.0. The kinetic parameters of WT and mutant enzymes were determined at 60 °C, pH 8.0, and a pectin substrate concentration of 2–30 mg/mL. The V_{max} and K_m values were obtained by fitting the Michaelis-Menten equation using GraphPad Prism v9.5.0 software.

2.9. Bioinformatic analysis

The three-dimensional (3D) structure of WT enzymes was predicted using AlphaFold2 (<https://github.com/deepmind/alphafold>) and RoseTTAFold (<https://rosetta.bakerlab.org>). Mutant enzyme structures were predicted using AlphaFold2. Predicted structures were evaluated using a Ramachandran plot on the SAVES v6.0 (<https://saves.mbi.ucla.edu>) website. Intermolecular interactions, hydrophobicity, and surface charge distribution of WT and mutant enzymes were analyzed using LigPlot⁺ v2.2.8 and PyMol v2.5.4 software. Molecular docking was conducted using AutoDock Vina v1.1.2 software. The polygalacturonic acid molecular structure was from the TCMSP (<https://old.tcmssp-e.com/tcmssp.php>) website. The active pocket volume was calculated using POVME v2.2.0.

The molecular dynamics (MD) simulation was performed using the GROMACS 2022 software. The force field used was Charmm36 [29], and the water model used was TIP3P. Neutralize the system by adding Na⁺ and Cl⁻, perform energy minimization for 1000 ps using the steepest descent method, then slowly heat the system to 348 K, run 100 ps of NVT and NPT to equilibrate temperature and pressure. Finally, at 348 K, perform a 100 ns kinetic simulation. The root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg),

and solvent accessible surface area (SASA) of the protein were calculated using the GROMACS analysis module. The free energy landscape was plotted using principal component analysis.

2.10. Statistical analysis

The data are presented as the mean ± standard deviation (SD) of three independent experiments. Differences between groups were analyzed using one-way analysis of variance (ANOVA) with Tukey Honestly Significant Difference (HSD) test. A p -value of <0.05 was considered statistically significant.

3. Result

3.1. Screening for thermostable single-point mutants

In order to find the key residues that positively impact thermostability, we employed semi-rational design strategies, including computational design and consensus sequence design approaches. First, the 3D structure of PMGL-Ba was predicted using AlphaFold2 and RoseTTAFold. The results indicated that 99.5 % of the residues in the structure (AF2_M) predicted by AlphaFold2 and 97.8 % of the residues in the structure (RF_M) predicted by RoseTTAFold fell within the allowed region (Fig. S1). Both structures were considered appropriate for further analysis. Then, 54 hotspots were identified through the FireProt and BayeStab websites (Fig. 2A). Based on the principle of consensus mutagenesis, the amino acid sequence of PMGL-Ba was aligned with five known heat-resistant or highly active pectate lyases [30], which share similar 3D structures and catalytic mechanisms with PNLs [31]. This comparative analysis identified an additional 12 hotspots (Fig. 2B). Finally, 66 candidate hotspots were identified.

Subsequently, 63 mutants were successfully expressed in *E. coli* BL21 (DE3). PMGL-Ba loses more than half of its activity after being incubated at 65 °C for 30 min. Considering the trade-off between thermostability and activity [32], we evaluated the relative activity of each mutant and the residual activity after incubation for 30 min at 65 °C, as shown in Fig. 2C. Compared to the WT, fourteen mutants (R11H, T21V, Q35P, S115P, V195I, G200S, R310L, R340Y, R340W, Y357F, K360Y, S361N, T432P, and T451S) exhibited enhanced thermostability and either unchanged or improved activity. Five mutants (N255D, S264G, T305V, D308N, and W317Y) exhibited enhanced thermostability but reduced activity, particularly the D308N mutant, which lost nearly half of its activity. Mutants R340Y and S361N exhibited a significant enhancement in thermostability, with a >1.7-fold increase in residual activity after incubation for 30 min at 65 °C when compared to the WT. The remaining mutants exhibited a moderate increase in thermostability, with residual activities ranging from 1.1 to 1.5 times that of the WT. Therefore, it is necessary to combine these mutants to further improve the thermostability. Mutant D308N was not considered for the next step of combinatorial mutagenesis because of its more severe loss of activity. The other 18 mutations involve 17 amino acid residues. Theoretically, these mutations could produce $2^{16} \times 3 = 196,608$ combinatorial mutants, which is almost impossible to fully evaluate at the laboratory level. Therefore, to find the optimal combinatorial mutants, we plan to use ML to guide the design multi-site combinatorial mutations of these single-point mutants.

3.2. Preparation of the initial library

Before proceeding with ML, we first prepared the initial library to create training data for the ML model. By performing random combinatorial mutations on the 18 single-point mutants, 88 combinatorial mutants were obtained (Fig. S2A). Together with the 18 single-point mutants, they formed the initial library (containing 106 mutants in total). The initial library was mainly dominated by low-order mutants (orders 1–5) (Fig. 3A). Under the additive assumption [33], almost all of

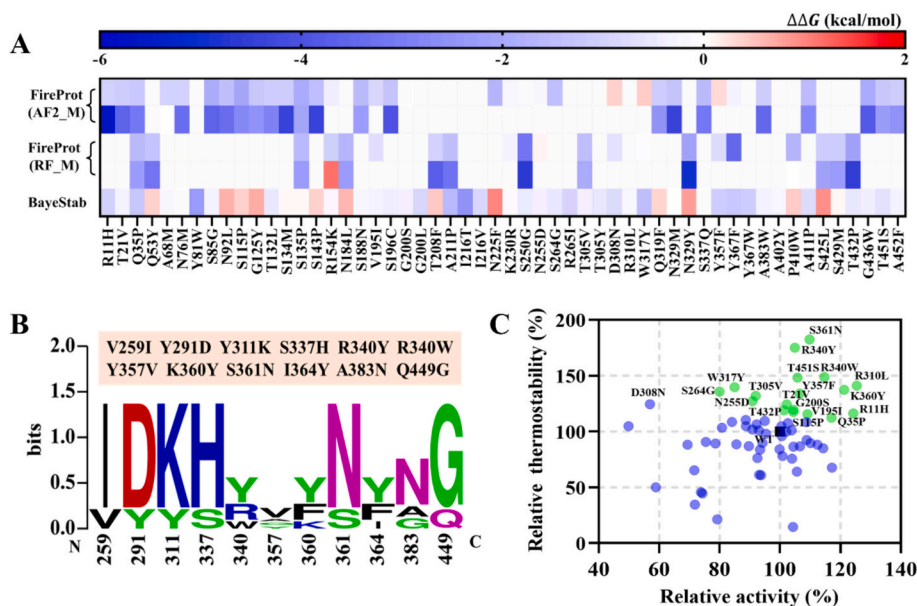


Fig. 2. Screening of PMGL-Ba thermostable single-point mutants. (A) $\Delta\Delta G$ heatmap of predicted candidate hotspots based on FireProt and Bayestab websites. (B) Candidate hotspots identified on the basis of the consensus mutation principle. (C) Relative activity and relative thermostability of single-point mutants obtained by the above methods. Relative activity is the ratio of the activity of the mutant to the WT before heat treatment. Relative thermostability is the ratio of the residual activity of the mutant to the WT after incubation at 65 °C for 30 min. The green represents the 18 thermostable mutants selected for further study.

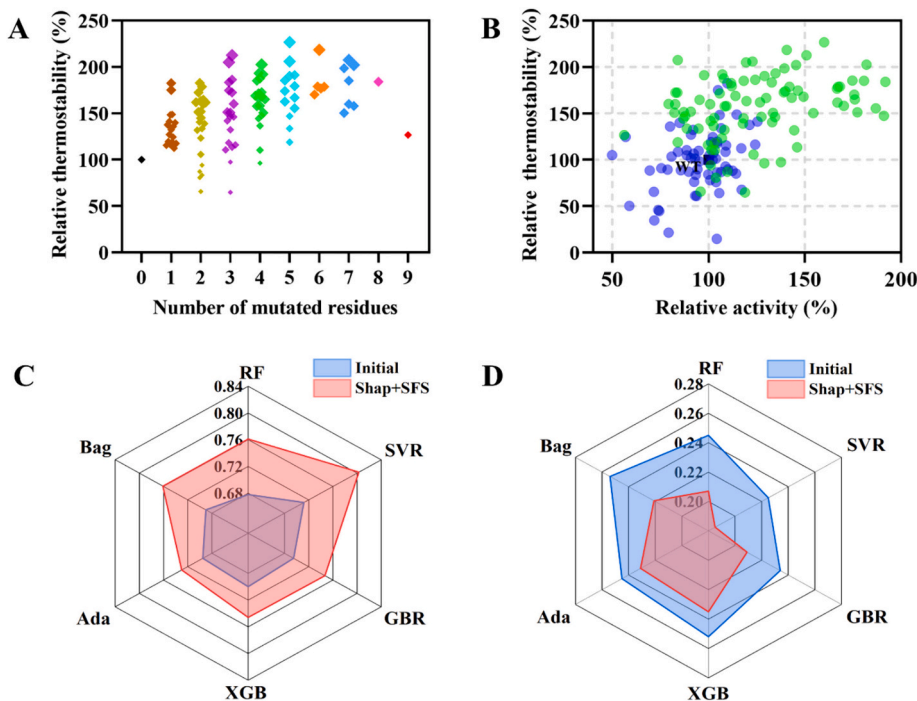


Fig. 3. Preparation of the ML training set and model construction. (A) Statistics of mutants with different mutation orders in the initial library. The initial library consisted of 88 combinatorial mutants and 18 thermostable single-point mutants. (B) The training data used for ML modeling. A total of 152 mutant data are included. The blue represents single-point mutants, while the green represents combinatorial mutants. R^2 scores (C) and RMSE scores (D) of different models before and after feature optimization.

the combinatorial mutants exhibited negative epistasis, and positive epistasis was rare (Fig. S2B). Moreover, the negative epistasis became stronger with an increasing number of mutations, which is consistent with previous studies [13]. This poses a significant challenge to traditional combinatorial mutation design methods, especially in the design of higher-order mutants, which requires more caution. This is also reflected in our experimental results (Fig. 3A). Hence, the adoption of ML

for multi-site combinatorial mutation design becomes crucial.

3.3. Machine learning modeling based on the initial library

We constructed ML prediction models using 106 mutants from the initial library and another 46 mutants (including WT) generated during the previous screening process (Fig. 3B). The model uses enzyme

sequences as input variants and generates predicted relative residual activities as output results. For the feature vectors used in the ML model, we considered various amino acid descriptors based on physicochemical properties (Table S2). To provide richer sequence information to the model and enable it to comprehensively capture complex molecular structures and properties for accurate prediction, we combined these descriptors (Fig. S3). Six regression algorithms (Fig. 3C,D), including RF, SVR, GBR, XGB, Ada, and Bag, were used. Due to the relatively small dataset, we evaluated the models using 10-fold cross-validation to mitigate the performance bias caused by data splitting. In previous studies, researchers have typically focused on improving model performance by combining various descriptors [8,11]. However, directly combining descriptors may lead to feature redundancy, increase the risk of overfitting, consume more computational resources, and complicate model inference and training. According to our related research, to avoid these problems, we adopted a “two-step strategy” to optimize the best model for each algorithm. First, the SHAP algorithm [34] was used to calculate the marginal contribution of each feature to the model and rank the features in terms of importance (Fig. S4). Subsequently, the Sequential Forward Search (SFS) algorithm [35] was used to search for the top-ranked features (top 300) one by one (Fig. S5). By employing this strategy, irrelevant features or those that have a negative effect on the model are removed, thereby enhancing the model's predictive performance. Finally, the SVR algorithm based on the AAindex_AAC_SOCN features performed the best, achieving the highest R^2 score of 0.803 with an RMSE of 0.185 based on 255 features (Fig. 3C,D). Thus, we chose this model as the final model to guide the directed evolution of PMGL-Ba. By using this model, we ranked all possible combinatorial mutants regarding the 17 mutated residues (i.e., $2^{16} \times 3 = 196,608$ mutants).

3.4. Iterative learning based on the second library

To assess the performance of the predicted combinatorial mutants, the second library was constructed by selecting 30 mutants with relatively high prediction rankings. Compared with the initial library, all combinatorial mutants in the second library showed higher thermostability than the WT (Fig. 4A). Twelve combinatorial mutants showed 100 % residual activity after incubation at 65 °C for 30 min, while the other mutants exhibited residual activities >1.7-fold to the WT. These results

demonstrate that ML successfully guided the multi-site combinatorial mutation of PMGL-Ba, significantly enriching combinatorial mutants with high thermostability. In addition, most of the combinatorial mutants in the second library maintained or even increased their activity while enhancing thermostability. For instance, combinatorial mutants P8, P11, and P13 showed more than a 1.5-fold increase in activity compared to the WT. However, there are still some combinatorial mutants experiencing a serious decrease in activity. This suggests that trade-off effects occurred in the sequence space explored in this round of ML. To avoid this problem, we try to make the ML model explore other sequence spaces by updating the training set, in hopes of avoiding this effect. To achieve this, we returned the thermostability data of 30 combinatorial mutants from the second library as an additional dataset to the ML model for iterative learning.

The third library was constructed using the same method as the second library. To our delight, almost all combinatorial mutants in the third library showed dual improvements in thermostability and activity (Fig. 4B). In particular, the combinatorial mutants P31, P36, P48, P54, and P55 not only showed 100 % residual activity after 30 min at 65 °C incubation but also exhibited more than a 1.8-fold higher activity compared to the WT. We found that the R310, R340, Y357, K360, and S361 residues were almost all mutated in the combinatorial mutants recommended by the two rounds of ML predictions (Fig. 4C,D). This suggests that these residues may have an important impact on enhancing the thermostability of the enzyme. Compared to the second library, the combinatorial mutants in the third library exhibited a higher frequency of mutations at residues R11, T21, Q35, S115, and V195, which are distant from the enzyme's active site (Fig. 4C,D). These sites were rarely observed in the combinatorial mutants with reduced activity (Table S3). Thus, this may be a reason for the reduced activity of some combinatorial mutants in the second library. We speculate that these residues may indirectly affect enzyme activity through interactions with other residues [36]. In addition, unlike the initial library, almost all of the combinatorial mutants in the second and third libraries are higher-order mutants (orders 6–11) (Fig. S6). These results show that our model can explore different sequence spaces based on different training data and predict the combined effects of higher-order mutations using a small amount of low-order mutation data.

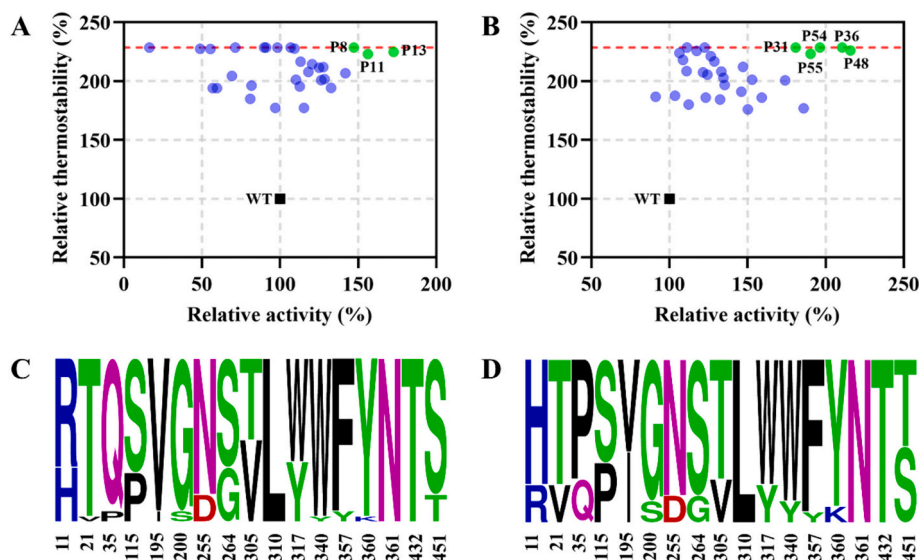


Fig. 4. Constructions of the second and third libraries and sequence logos of the combinatorial mutants predicted by ML. (A) (C) Second library. The more highly ranked 30 combinatorial mutants from the first round of ML prediction results. (B) (D) Third library. The more highly ranked 30 combinatorial mutants from the second round of ML prediction results. The red dashed positions represent 100 % residual activity. The green represents combinatorial mutants selected for further analysis. The columns in Fig. 4C and D represent mutation positions, namely, R11, T21, Q35, S115, V195, G200, N255, S264, T305, R310, W317, R340, Y357, K360, S361, T432, and T451.

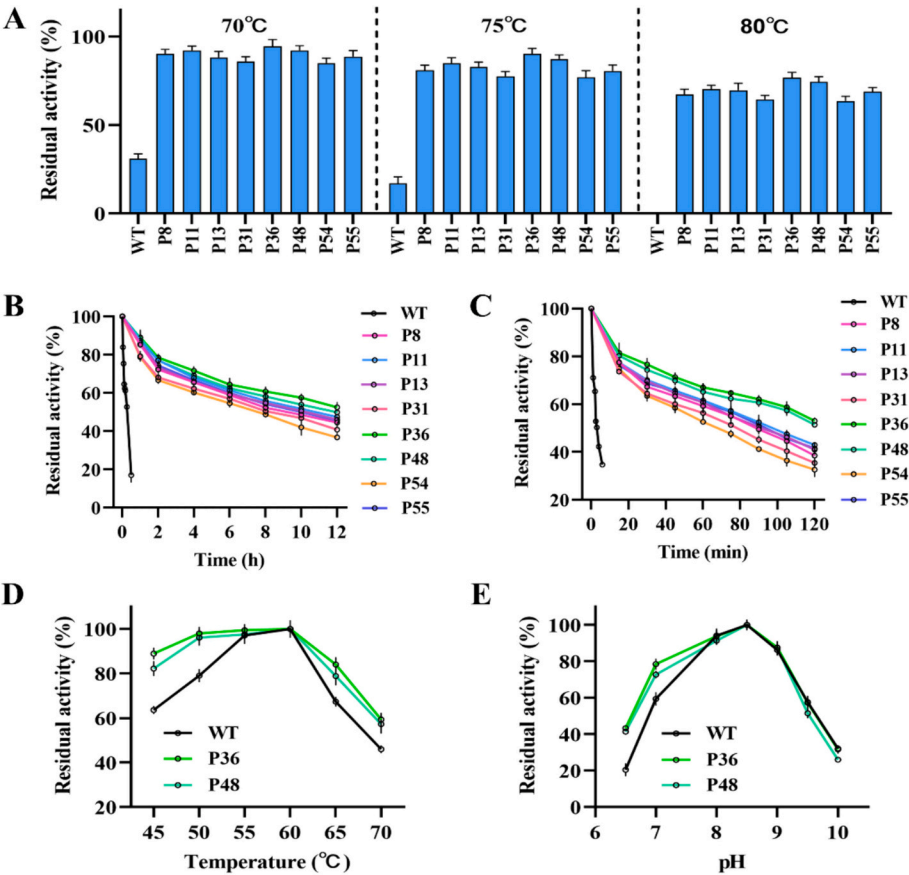


Fig. 5. Characterization of WT and predicted combinatorial mutants. (A) Residual activity after 30 min incubation at 70 °C, 75 °C, and 80 °C. (B) Half-life at 75 °C. (C) Half-life at 80 °C. (D) Optimum temperature. (E) Optimum pH.

3.5. Enzymatic properties of WT and predicted combinatorial mutants

To further evaluate the performance of combinatorial mutants predicted by ML, we selected the eight best-performance combinatorial mutants from the second and third libraries for comparison with the WT. As shown in Fig. 5A, the thermostability of all combinatorial mutants was significantly higher than that of the WT at 70 °C, 75 °C, and 80 °C, particularly at 75 °C and 80 °C. The WT enzyme showed only about 17 % residual activity after 30 min at 75 °C and a complete loss of activity after the same duration of incubation at 80 °C. In contrast, ML-predicted combinatorial mutants exhibited >75 % and 60 % residual activity after incubation at 75 °C and 80 °C for the same duration, respectively. Notably, P36 and P48 displayed residual activities of 90 % and 87 %, respectively, after 30 min at 75 °C, representing 5.3- and 5.1-fold increases compared to the WT. In addition, the half-life of P36 and P48 at

75 °C was extended from the original 12.45 min to 836.36 min and 762.43 min, respectively, representing a 67- and 61-fold increase (Fig. 5B). The half-life at 80 °C was prolonged from the original 3.98 min to 153.33 min and 146.81 min, respectively, showing a 39- and 37-fold increase (Fig. 5C). Similarly, other combinatorial mutants showed similar thermostability (Table 1). These results once again demonstrate the remarkable success of ML in guiding multi-site combinatorial mutations.

Next, we selected the combinatorial mutants P36 and P48 for further research, as they also showed the highest relative activity (2.1- and 2.2-fold higher than the WT, respectively) in the case of having the highest thermostability (Table 1). The optimal temperature and pH of the P36 and P48 did not change compared to WT, but they showed higher sustained activity over a wider temperature range (45 °C–65 °C) and pH range (7.0–9.0) (Fig. 5D,E). Furthermore, the P36 and P48 showed

Table 1
Enzymatic properties of WT and predicted combinatorial mutants.

Enzyme	Relative activity (%)	t _{1/2} (75 °C) (min)	t _{1/2} (80 °C) (min)
WT (PMGL-Ba)	100 ± 5.26	12.45	3.98
P8 (T305V-R310L-W317Y-R340W-Y357F-K360Y-S361N-T451S)	147.15 ± 2.34	663.42	100.00
P11 (R11H-S115P-T305V-R310L-W317Y-R340W-Y357F-K360Y-S361N-T451S)	156.27 ± 3.94	707.69	111.29
P13 (R11H-Q35P-R310L-R340W-Y357F-K360Y-S361N)	172.75 ± 2.70	698.14	104.55
P31 (R11H-Q35P-S115P-V195I-G200S-R310L-R340Y-Y357F-K360Y-S361N)	180.78 ± 4.23	642.85	90.79
P36 (R11H-Q35P-S115P-V195I-R310L-R340W-Y357F-K360Y-S361N)	210.53 ± 3.21	836.36	153.33
P48 (R11H-R310L-R340W-Y357F-K360Y-S361N)	215.75 ± 5.22	762.43	146.81
P54 (R11H-Q35P-S264G-R310L-R340W-Y357F-K360Y-S361N-T451S)	196.26 ± 2.34	565.57	80.23
P55 (R11H-Q35P-V195I-G200S-R310L-R340Y-Y357F-K360Y-S361N)	190.29 ± 4.11	679.80	106.15

Table 2

Kinetic parameters of WT and predicted combinatorial mutants.

Enzyme	K_m ($\text{g} \cdot \text{L}^{-1}$)	K_{cat} (s^{-1})	K_{cat}/K_m ($\text{L} \cdot \text{g}^{-1} \cdot \text{s}^{-1}$)
WT	10.21 ± 0.21	1224.96 ± 32.33	119.98
P36	10.39 ± 0.11	3090.56 ± 29.82	297.46
P48	10.32 ± 0.23	3198.80 ± 34.25	309.96

higher reaction rates than WT at all given substrate concentrations (2–30 mg/mL) (Fig. S7). Notably, the K_m values of P36 and P48 remained relatively unchanged, but the K_{cat}/K_m values increased by 2.48- and 2.58-fold, respectively (Table 2). This indicates a significant improvement in the catalytic efficiency of P36 and P48 towards the substrate, leading to increased activity.

3.6. Structural analysis and MD simulation

To provide insight into how these mutations improve thermostability, we selected the combination mutant P36 with the highest enhancement in thermostability for analysis. The structure of P36 was determined by AlphaFold2 (Fig. S8). The structural comparison between the WT and P36 indicates that five aspects of changes have improved

thermostability as follows: rearrangement and optimization of the protein surface charges induced by the R11H, Q35P, R310L, and R340W mutations (Fig. 6B), enhancement of hydrophobic interactions in the protein interior (R11H, S115P, and V195I) (Fig. S9), enhancement of hydrogen-bonding interactions resulting from the V195I, Y357F, and S361N mutations, introduction of new $\pi - \pi$ interactions between residues H338, H358, and Y430 and the mutant residue K360Y (Fig. 6C), and enhancement of structural stability by introducing proline at residue 35 (Q35P) and 115 (S115P) in the loop region (Fig. S10) [37–39].

To further understand the reason for the enhanced thermostability of the mutants, we conducted 100 ns MD simulations at 348 K for WT and P36. The results showed that the combinatorial mutant P36 overall exhibited lower RMSF, particularly in the 30–40, 99–135, 250–255, 345–375, and C-terminal amino acid regions, indicating increased structural rigidity in these regions [40] (Fig. 6D). Meanwhile, P36 exhibited a lower RMSD values throughout the simulations (average RMSD for WT: 2.4779 Å; average RMSD for P36: 1.5923 Å) as well as a lower Rg values (average Rg for WT: 2.2850 nm; average Rg for P36: 2.2728 nm) (Figs. 6E and S11A). This indicates that P36 is more compact and stable during the simulation [40,41]. The SASA of P36 did not change significantly compared to WT (Fig. S11B). In addition, the

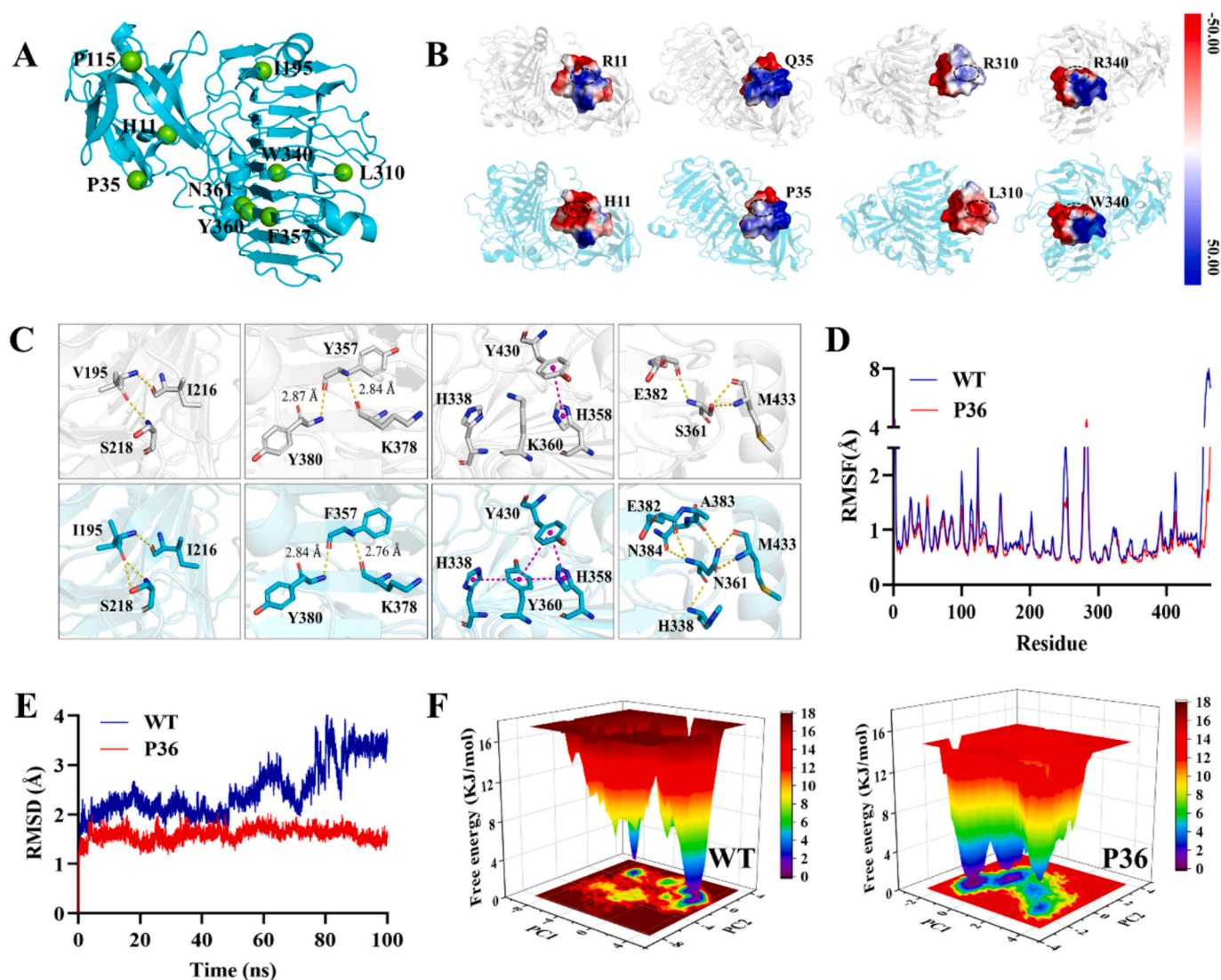


Fig. 6. Structural analysis and MD simulation of WT and combinatorial mutant P36. (A) Distribution of mutation sites in the combinatorial mutant P36. (B) Analysis of protein surface electrostatic charge. (C) Structural differences between WT and P36. The yellow dotted lines represent hydrogen bonds. The purple dotted lines represent $\pi - \pi$ interactions. (D) RMSF. (E) RMSD. (F) Free energy landscape.

overall energy of the free energy landscape of P36 is lower than that of the WT and it exhibits multiple minima, indicating that P36 is more stable in the conformational space [42] (Fig. 6F). In summary, the enhanced thermostability of the mutants can be attributed to the optimization of surface charge distribution and the increase in intermolecular interactions. These changes lead to an overall increase in protein rigidity and more stable conformations.

Molecular docking simulation analysis reveals the reason for the improved catalytic efficiency of the combinatorial mutant P36 (Fig. 7). Hydrogen bonding between P36 and substrate molecules was enhanced, and pocket hydrophobicity was increased compared to the WT (Fig. 7A). This helps to enhance the binding between the enzyme and the substrate, facilitating the enzyme's capture of the substrate [43]. Notably, the steric hindrance at one end of the P36 active pocket decreased, making it more open, and the volume increased from 745.0 Å to 946.0 Å (Figs. 7B and S12). These changes benefit substrate entry and product release [43,44]. Residue R347 has been identified as the catalytic base for initiating the β -elimination reaction by pectin lyase, extracting protons from C⁵ for the catalytic reaction [45]. The reduction in steric hindrance optimized the spatial conformation of the substrate in the pocket, leading to tighter binding of the substrate to the enzyme and shortening the distance of proton extraction, thereby enhancing the catalytic efficiency (Fig. 7C).

4. Discussion

Thermostability is a crucial indicator for evaluating the feasibility of enzymes in industrial applications [46]. Enhancing enzyme thermostability can broaden the application range of enzymes, improve industrial efficiency, and reduce production costs, which is significant for industries such as chemicals, pharmaceuticals, and biofuels [2]. PMGL-Ba, an alkaline PNL from *Bacillus licheniformis*, has great potential for

application in the textile and paper industries [16]. However, its poor thermostability severely limits its industrial applications. In this study, ML-guided multi-site combinatorial mutagenesis was employed to enhance the thermostability of PMGL-Ba. Compared to the WT, the best-performing combinatorial mutant, P36, exhibited a 67-fold and 39-fold increase in half-life at 75 °C and 80 °C, respectively, as well as a 2.1-fold increase in activity (Table 1). Furthermore, P36 demonstrated good performance over a broad range of temperatures and pH levels (Fig. 5D, E). Therefore, mutant P36 can serve as a powerful candidate enzyme for applications such as ramie degumming in the textile industry, pulp bleaching in the paper industry, and the treatment of pectin-containing industrial wastewater.

Previous studies have shown that the SVR algorithm performs well in predicting enzyme thermostability [47]. Among the six regression algorithms tested in this study, SVR achieved the highest performance scores, regardless of feature optimization (Fig. 3C,D). Therefore, for predicting the thermostability of enzymes, the SVR algorithm can be preferred. During the ML phase, we initially used 152 mutants as the training set to construct the ML model and ranked all possible combinatorial mutants of the 18 PMGL-Ba mutants. Since these 18 mutants involved 17 residues, the size of combinatorial sequence space was $2^{16} \times 3 = 196,608$. Although the dataset of 152 mutants is relatively small, the ML-generated libraries were enriched with high-performance mutants (Fig. 4A,B). Similarly, in several previous studies, when using ML to explore the sequence space of all four-point mutants ($20^4 = 160,000$), the initial training set size was usually around 100–150, and high-performance mutants were successfully identified [9,11,48]. Therefore, this training set size should be sufficient for discovering high-performance mutants from a sequence space containing approximately 200,000 mutants. However, using more data might yield better predictions, but this must be weighed against the experimental burden and time costs. With limited experimental data, iterative learning can

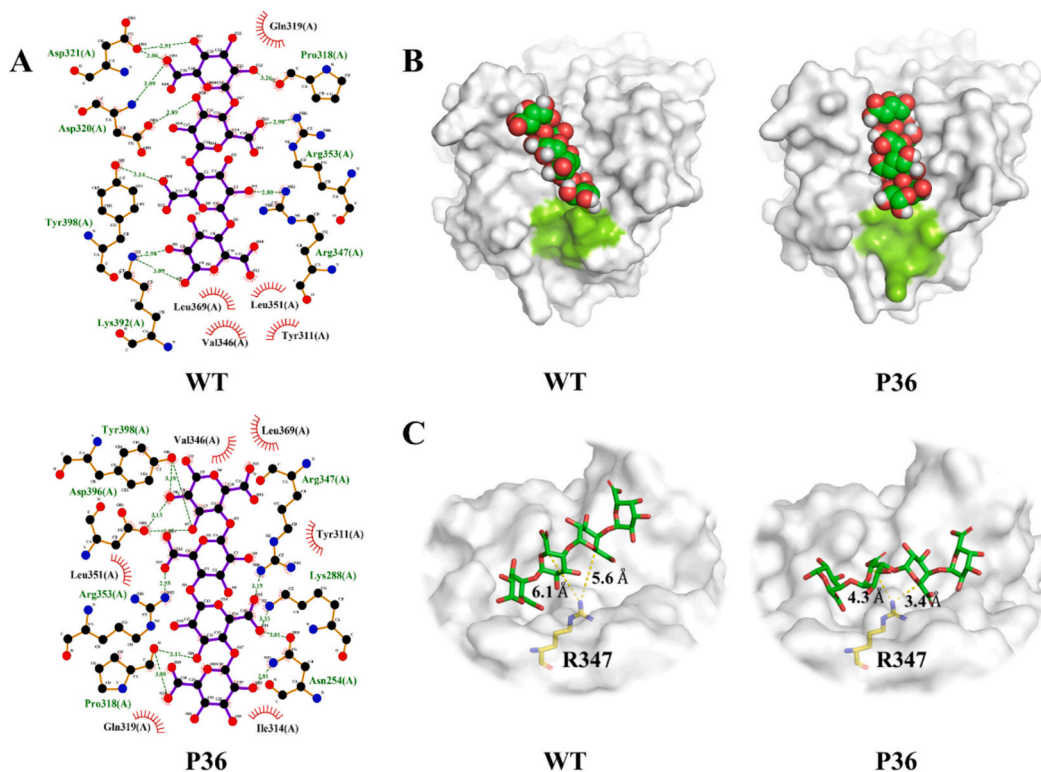


Fig. 7. Docking simulation analysis between polygalacturonic acid substrate and WT and mutant P36. (A) Interaction analysis between the WT and P36 and the substrate molecules. The purple structural formula in the center represents the substrate molecule. The red eyelash shape represents hydrophobic interactions, and the green dotted lines represents hydrogen bonds. (B) Structural comparison of the active pockets in WT and P36. (C) Distance analysis of the N atom in R347 to the C⁵ atom of the monosaccharide unit of the substrate molecule.

gradually improve model performance [8]. In this study, after one round of iterative learning, the combinatorial mutants in the third library were overall superior to those in the second library (Table 1), and the best-performing combinatorial mutant, P36, also came from the third library (Fig. 4B), indicating that iterative learning helps in identifying better mutants. Compared to the second library, most combinatorial mutants in the third library showed improvements in thermostability and activity, with significant sequence differences from the second library mutants (Fig. 4). Furthermore, the ML model's training data primarily consisted of low-order mutants (orders 1–5), but the model-recommended combinatorial mutants were all high-order mutants (orders 6–11) (Fig. S6). These results suggest that ML can explore different sequence spaces based on varying training data and predict high-order mutants from low-order mutants.

This study further investigates the mechanisms underlying the enhancement of enzyme thermostability and activity. It has been reported that optimizing the distribution of surface charges can improve protein thermostability [37]. In this study, multiple positions in the combinatorial mutant P36 exhibited significant changes in surface electrostatic potential (Fig. 6B), potentially optimizing the enzyme's surface charge distribution and thereby enhancing thermostability. Compared to the WT, P36 showed increased hydrogen bonds, $\pi-\pi$ interactions, and hydrophobic interactions (Figs. 6C and S9). These intermolecular interactions play crucial roles in maintaining enzyme thermostability. Hydrogen bonds and $\pi-\pi$ interactions strengthen the connections between amino acids within the protein, making the structure more stable [38]; increased hydrophobic interactions can significantly reduce conformational entropy, making the protein structure more compact, thereby improving conformational stability [39]. Additionally, the introduction of proline at positions Q35 and S115 in the loop regions of P36 increased the structural rigidity of these areas (Fig. S10). Compared to the WT, P36 exhibited lower RMSF, RMSD, and Rg values, as well as lower energy within the specified conformational space (Figs. 6D,E,F and S11), consistent with the structural analysis results. In summary, the optimization of surface charge distribution, enhanced intermolecular interactions, and increased protein structural rigidity are partial reasons for the improved thermostability of the mutant.

Interestingly, although increasing enzyme thermostability often comes at the cost of reduced activity, the catalytic activity and efficiency of the combinatorial mutant P36 were significantly enhanced (Tables 1,2). Compared to the WT, P36 exhibited reduced steric hindrance at one end of the active pocket, with the pocket volume increasing and becoming more open (Figs. 7B and S12). Studies have shown that reduced steric hindrance and increased pocket volume facilitate substrate entry and product release, thereby enhancing the enzyme's catalytic efficiency [44]. Furthermore, the reduction in steric hindrance leads to tighter substrate binding to the enzyme and a shorter distance for proton extraction (Fig. 7C), accelerating the reaction process. Overall, the reduction in pocket steric hindrance and the optimization of substrate conformation are primary reasons for the improved catalytic activity and efficiency of the mutant.

Despite the significant achievements of our study, there are some limitations. For instance, the number of experimentally validated mutants is limited, so the globally optimal mutant may not have been discovered. This modeling approach requires substantial computational resources and is at risk of overfitting. We did not assess the long-term stability of these mutants under industrial conditions, which is crucial for practical applications. Additionally, there may be biases in the selection of initial single-point mutants, potentially affecting the performance of the final combinatorial mutants. The generalizability of this method to other enzymes and properties has yet to be verified. Future research should strengthen experimental validation, optimize the model to reduce computational resource demands, and assess the applicability of this approach to different enzymes and properties to enhance its generalizability and predictive accuracy.

5. Conclusion

In this study, ML was proven to be effective in guiding multi-site combinatorial mutagenesis of enzymes, which is beneficial for obtaining stable enzyme mutants. This approach enriched stable mutants using only 152 data points, which represents about 0.08 % of the sequence space. The best-performing combinatorial mutant, P36, exhibited a 67-fold and 39-fold increase in half-life at 75 °C and 80 °C, respectively, as well as a 2.1-fold increase in activity. Additionally, executing ML with different training data may reveal diverse enhanced mutants. The ML model can effectively address the challenges posed by epistasis effects, having the ability to predict from low-order mutants to higher-order combinatorial mutants. Lastly, structural analyses and MD simulations provided insights into the improved performance of P36. In conclusion, this study provides a rational method and guidance for designing multi-site combinatorial mutants in protein engineering and developing highly thermostable enzymes, as well as offering a powerful candidate biocatalyst for industries such as textiles and paper manufacturing.

CRediT authorship contribution statement

Zhihui Zhang: Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Zhixuan Li:** Validation, Methodology, Investigation. **Manli Yang:** Supervision, Project administration. **Fengguang Zhao:** Supervision, Project administration. **Shuangyan Han:** Writing – review & editing, Resources, Funding acquisition, Conceptualization.

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.

Data availability

The thermostability data (training set) of PMGL-Ba mutants are available in [Data Set S1 \(XLSX\)](#). The code used in this study is based on our another research and has been made publicly available at https://github.com/Zhang66-star/scut_ProFP.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (NO. 2022YFC2105500).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijbiomac.2024.134530>.

References

- [1] J. Chapman, A.E. Ismail, C.Z. Dinu, Industrial applications of enzymes: recent advances, techniques, and outlooks, *Catalysts* 8 (2018) 238, <https://doi.org/10.3390/catal8060238>.
- [2] L. Zhao, Z. Ma, Q. Wang, Engineering the thermostability of sucrose synthase by reshaping the subunit interaction contributes to efficient UDP-glucose production, *J. Agr. Food Chem.* 71 (2023) 3832–3841, <https://doi.org/10.1021/acs.jafc.2c08642>.
- [3] M. Musil, J. Stourac, J. Bendl, FireProt: web server for automated design of thermostable proteins, *Nucleic Acids Res.* 45 (2017) W393–W399, <https://doi.org/10.1093/nar/gkx285>.
- [4] J.J. Weinstein, A. Goldenzweig, S. Hoch, PROSS 2: a new server for the design of stable and highly expressed protein variants, *Bioinformatics* 37 (2021) 123–125, <https://doi.org/10.1093/bioinformatics/btaa1071>.
- [5] R. Shroff, A.W. Cole, D.J. Diaz, Discovery of novel gain-of-function mutations guided by structure-based deep learning, *ACS Synth. Biol.* 9 (2020) 2927–2935, <https://doi.org/10.1021/acssynbio.0c00345>.

- [6] M.J. Menke, Y.-F. Ao, U.T. Bornscheuer, Practical machine learning-assisted design protocol for protein engineering: transaminase engineering for the conversion of bulky substrates, *ACS Catal.* 14 (2024) 6462–6469, <https://doi.org/10.1021/acscatal.4c00987>.
- [7] V. Sayous, P. Lubrano, Y. Li, Unbiased libraries in protein directed evolution, *BBA-Proteins Proteom.* 1868 (2020) 140321, <https://doi.org/10.1016/j.bbapap.2019.140321>.
- [8] Y. Saito, M. Oikawa, T. Sato, Machine-learning-guided library design cycle for directed evolution of enzymes: the effects of training data composition on sequence space exploration, *ACS Catal.* 11 (2021) 14615–14624, <https://doi.org/10.1021/acscatal.1c03753>.
- [9] Z. Wu, S.J. Kan, R.D. Lewis, Machine learning-assisted directed protein evolution with combinatorial libraries, *Proc. Natl. Acad. Sci. U. S. A.* 116 (2019) 8852–8858, <https://doi.org/10.1073/pnas.1901979116>.
- [10] R. Hu, L. Fu, Y. Chen, Protein engineering via Bayesian optimization-guided evolutionary algorithm and robotic experiments, *Brief. Bioinform.* 24 (2023) bbac570, <https://doi.org/10.1093/bib/bbac570>.
- [11] T. Matsushita, S. Kishimoto, K. Hara, Functional enhancement of flavin-containing monooxygenase through machine learning methodology, *ACS Catal.* 14 (2024) 6945–6951, <https://doi.org/10.1021/acscatal.4c00826>.
- [12] F. Cadet, N. Fontaine, G. Li, A machine learning approach for reliable prediction of amino acid interactions and its application in the directed evolution of enantioselective enzymes, *Sci. Rep.* 8 (2018) 16757, <https://doi.org/10.1038/s41598-018-35033-y>.
- [13] Y. Chen, R. Hu, K. Li, Deep mutational scanning of an oxygen-independent fluorescent protein CreiLOV for comprehensive profiling of mutational and epistatic effects, *ACS Synth. Biol.* 12 (2023) 1461–1473, <https://doi.org/10.1021/acssynbio.2c00662>.
- [14] K. Yadav, S. Dwivedi, S. Gupta, Recent insights into microbial pectin lyases: a review, *Process Biochem.* (2023) 1359–1513, <https://doi.org/10.1016/j.procbio.2023.10.008>.
- [15] S. Samanta, Microbial pectinases: a review on molecular and biotechnological perspectives, *J. Microb. Biotech. Food* 9 (2019) 248, <https://doi.org/10.15414/jmbfs.2019.9.2.248-266>.
- [16] J. Li, M. Yang, F. Zhao, Efficient expression of an alkaline pectin lyase from *Bacillus licheniformis* in *Pichia pastoris*, *Bioresour. Bioprocess.* 11 (2024) 37, <https://doi.org/10.1186/s40643-024-00752-w>.
- [17] T. Zhang, P. Liu, H. Wei, Protein engineering of glucosylglycerol phosphorylase facilitating efficient and highly regio- and stereoselective glycosylation of polyols in a synthetic system, *ACS Catal.* 12 (2022) 15715–15727, <https://doi.org/10.1021/acscatal.2c05232>.
- [18] S. Kawashima, M. Kanehisa, AAindex: amino acid index database, *Nucleic Acids Res.* 28 (2000) 374, <https://doi.org/10.1093/nar/28.1.374>.
- [19] J. Dong, Z.-J. Yao, L. Zhang, PyBioMed: a python library for various molecular representations of chemicals, proteins and DNAs and their interactions, *J. Chem.* 10 (2018) 1–11, <https://doi.org/10.1186/s13321-018-0270-2>.
- [20] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32, <https://doi.org/10.1023/A:1010933404324>.
- [21] M. Awad, R. Khanna, M. Awad, Support vector regression, in: *Efficient Learning Machines*, Apress, Berkeley, CA, 2015, pp. 67–80, https://doi.org/10.1007/978-1-4302-5990-9_4.
- [22] J.H. Friedman, Stochastic gradient boosting, *Computational statistics & data analysis* 38 (2002) 367–378, [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2).
- [23] Z.A. Ali, Z.H. Abduljabbar, H.A. Taher, Exploring the power of eXtreme gradient boosting algorithm in machine learning: a review, *Academic Journal of Nawroz University* 12 (2023) 320–334, <https://doi.org/10.25007/ajnu.v12n2a1612>.
- [24] A.J. Ferreira, M.A. Figueiredo, Boosting algorithms: a review of methods, theory, and applications, *Ensemble machine learning* (2012) 35–85, https://doi.org/10.1007/978-1-4419-9326-7_2.
- [25] L. Breiman, Bagging predictors, *Mach. Learn.* 24 (1996) 123–140, <https://doi.org/10.1007/BF00058655>.
- [26] F. Pedregosa, G. Varoquaux, A. Gramfort, Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830, <https://doi.org/10.5555/1953048.2078195>.
- [27] O. Renaud, M.-P. Victoria-Feser, A robust coefficient of determination for regression, *J. Stat. Plan. Inference* 140 (2010) 1852–1862, <https://doi.org/10.1016/j.jspi.2010.01.008>.
- [28] T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature, *Geosci. Model Dev.* 7 (2014) 1247–1250, <https://doi.org/10.5194/gmd-7-1247-2014>.
- [29] J. Huang, S. Rauscher, G. Nawrocki, CHARM36: an improved force field for folded and intrinsically disordered proteins, *BpJ* 112 (2017) 175a–176a, <https://doi.org/10.1016/j.bpj.2016.11.971>.
- [30] X. Wang, Z. Lu, T. Xu, Improving the specific activity and thermo-stability of alkaline pectate lyase from *Bacillus subtilis* 168 for bioscouring, *Biochem. Eng. J.* 129 (2018) 74–83, <https://doi.org/10.1016/j.bej.2017.11.001>.
- [31] L. Zheng, Y. Xu, Q. Li, Pectinolytic lyases: a comprehensive review of sources, category, property, structure, and catalytic mechanism of pectate lyases and pectin lyases, *Bioresour. Bioprocess.* 8 (2021) 1–13, <https://doi.org/10.1186/s40643-021-00432-z>.
- [32] Q. Hou, M. Rooman, F. Pucci, Enzyme stability-activity trade-off: new insights from protein stability weaknesses and evolutionary conservation, *J. Chem. Theory Comput.* 19 (2023) 3664–3671, <https://doi.org/10.1021/acs.jctc.3c00036>.
- [33] K.S. Sarkisyan, D.A. Bolotin, M.V. Meer, Local fitness landscape of the green fluorescent protein, *Nature* 533 (2016) 397–401, <https://doi.org/10.1038/nature17995>.
- [34] Y. Bi, D. Xiang, Z. Ge, An interpretable prediction model for identifying N7-methylguanosine sites based on XGBoost and SHAP, *Mol Ther-Nucl Acids* 22 (2020) 362–372, <https://doi.org/10.1016/j.omtn.2020.08.022>.
- [35] P. Pudil, J. Novotná, J. Kittler, Floating search methods in feature selection, *PaRel* 15 (1994) 1119–1125, [https://doi.org/10.1016/0167-8655\(94\)90127-9](https://doi.org/10.1016/0167-8655(94)90127-9).
- [36] H. Yu, P.A. Dalby, Coupled molecular dynamics mediate long- and short-range epistasis between mutations that affect stability and aggregation kinetics, *Proc. Natl. Acad. Sci. U. S. A.* 115 (2018) E11043–E11052, <https://doi.org/10.1073/pnas.1810324115>.
- [37] K.L. Schweiker, A. Zarrine-Afsar, A.R. Davidson, Computational design of the Fyn SH3 domain with increased stability through optimization of surface charge-charge interactions, *Protein Sci.* 16 (2007) 2694–2702, <https://doi.org/10.1110/ps.073091607>.
- [38] H. Wu, Q. Chen, W. Zhang, Overview of strategies for developing high thermostability industrial enzymes: discovery, mechanism, modification and challenges, *Crit. Rev. Food Sci.* 63 (2023) 2057–2073, <https://doi.org/10.1080/10408398.2021.1970508>.
- [39] Y. Li, C. Li, H. Huang, Significantly enhanced thermostability of *Aspergillus niger* xylanase by modifying its highly flexible regions, *J. Agric. Food Chem.* 70 (2022) 4620–4630, <https://doi.org/10.1021/acs.jafc.2c01343>.
- [40] R. Wang, S. Wang, Y. Xu, Enhancing the thermostability of *Rhizopus chinensis* lipase by rational design and MD simulations, *Int. J. Biol. Macromol.* 160 (2020) 1189–1200, <https://doi.org/10.1016/j.ijbiomac.2020.05.243>.
- [41] X.-D. Pei, H.-L. Fan, D.-Q. Jiao, Rational engineering S1' substrate binding pocket to enhance substrate specificity and catalytic activity of thermal-stable keratinase for efficient keratin degradation, *Int. J. Biol. Macromol.* 263 (2024) 130688, <https://doi.org/10.1016/j.ijbiomac.2024.130688>.
- [42] J. Bi, S. Chen, X. Zhao, Computation-aided engineering of starch-debranching pullulanase from *Bacillus thermovorans* for enhanced thermostability, *Appl. Microbiol. Biot.* 104 (2020) 7551–7562, <https://doi.org/10.1007/s00253-020-10764-z>.
- [43] L. Cui, A. Cui, Q. Li, Molecular Evolution of an Aminotransferase Based on Substrate–Enzyme Binding Energy Analysis for Efficient Valienamine Synthesis, *ACS Catal.* 12 (2022) 13703–13714, <https://doi.org/10.1021/acscatal.2c03784>.
- [44] T. Wu, Y. Wang, N. Zhang, Reshaping substrate-binding pocket of leucine dehydrogenase for bidirectionally accessing structurally diverse substrates, *ACS Catal.* 13 (2022) 158–168, <https://doi.org/10.1021/acscatal.2c04735>.
- [45] P. Sánchez-Torres, J. Visser, J.A. Benen, Identification of amino acid residues critical for catalysis and stability in *Aspergillus niger* family 1 pectin lyase A, *Biochem. J.* 370 (2003) 331–337, <https://doi.org/10.1042/bj20021071>.
- [46] Z. Xu, Y.-K. Cen, S.-P. Zou, Recent advances in the improvement of enzyme thermostability by structure modification, *Crit. Rev. Biotechnol.* 40 (2020) 83–98, <https://doi.org/10.1080/07388551.2019.1682963>.
- [47] C.W. Chen, J. Lin, Y.W. Chu, iStable: off-the-shelf predictor integration for predicting protein stability changes, *BMC Bioinformatics* 14 (2013) 1–14, <https://doi.org/10.1186/1471-2105-14-S2-S5>.
- [48] Y. Saito, M. Oikawa, H. Nakazawa, Machine-learning-guided mutagenesis for directed evolution of fluorescent proteins, *ACS Synth. Biol.* 7 (2018) 2014–2022, <https://doi.org/10.1021/acssynbio.8b00155>.