



# Enhancing thermostability of a psychrophilic alpha-amylase by the structural energy optimization in the trajectories of molecular dynamics simulations

Qingbin Li<sup>a,b,1</sup>, Yaru Yan<sup>b,c,1</sup>, Xiaoqing Liu<sup>b</sup>, Ziding Zhang<sup>a,\*</sup>, Jian Tian<sup>b,\*</sup>, Ningfeng Wu<sup>b</sup>

<sup>a</sup> State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, China

<sup>b</sup> Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China

<sup>c</sup> College of Food Science and Technology, Agricultural University of Hebei, Baoding, Hebei Province 071001, China

## ARTICLE INFO

### Article history:

Received 22 July 2019

Received in revised form 18 September 2019

Accepted 1 October 2019

Available online 14 October 2019

### Keywords:

Cold-adapted alpha-amylase

Thermostability

Mutant

Molecular dynamics simulations

Energy calculation

## ABSTRACT

The cold-adapted alpha-amylase (PHA) from *Pseudoalteromonas haloplanktis* is a psychrophilic enzyme which demonstrates high activity at low temperatures, but poor thermostability. Most of the method only employed the crystal structure to design the target protein. However, the trajectory of protein molecular dynamics (MD) simulation contained clues about the protein stability. In this study, we combined MD simulation and energy optimization methods to design mutations located at non-conserved residues. Two single point mutants (S255K, S340P) and one integrated mutant (S255K/S340P) enhanced thermostability without affecting the optimal catalytic activity. After incubation at 40 °C for 80 min, the residual activities of mutants S255K, S340P and S255K/S340P were 1.6-, 2.4-, and 2.6-fold greater than that of the wild type (WT). Additionally, the catalytic efficiency values ( $k_{cat}/K_m$ ) of S255K, S340P, and S255K/S340P also increased 1.9-, 2.0-, and 2.7-fold when compared to WT.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

The alpha-amylase (EC 3.2.1.1) is one class of industrial enzymes that constitute approximately 30% of the world's enzyme market and are used in the industrial production of starches, detergents, textiles, and pharmaceuticals [1–3]. The cold-adapted alpha-amylase has emerged as a useful biocatalyst which has broad applications at low temperatures [4–6]. However, the enzyme is denatured after small temperature changes, which strongly limits its application in the biotechnology industry [7–9]. It is crucial to improve the thermostability of this biocatalyst without impacting its optimal temperature.

Various strategies have been suggested to enhance the enzyme thermostability, including: (a) a homologous alignment; (b) the optimization of protein surface charge; (c) a strategy based on disulfide bonds; (d) the strategy of proline effect; (e) a strategy based on temperature factor; and (f) the strategy of free energy based on protein folding [10–14]. The majority of these methods have attempted to increase the protein structure's rigidity by modifying the interaction of the residue with the covalent or non-

covalent bond that simulated with the crystal structure of proteins. However, the protein crystal structure provided by the Protein Data Bank (PDB) only represents a special conformation of the protein which is different from its natural structure at high temperatures. Therefore, mutations which are designed to enhance the thermostability *in silico* must consider various protein conformations, especially the conformation at high temperatures.

MD simulations have evolved as a well-developed technique to understand macromolecular conformations [15]. Protein can be simulated in the solution at different temperatures *in silico*. It was simple for us to generate many protein conformations, which are helpful further infer protein dynamic properties through structural analyses [16–20]. In this study, a strategy that combined MD simulation and energy optimization was used to find protein mutants with enhanced thermostability. MD simulation generated enough protein conformations. Position scanning was used to screen positive mutations with enhanced thermostability. Mutations that were positive on the majority of the protein simulated conformations were designed for our experimental evaluation.

Experimental results showed that two single-point mutants (S255K and S340P) yielded improved thermal thermostability. One double-point mutant (S255K/S340P) was constructed based on those two single point mutations. It was found that the  $T_m$  value of mutant (S255K/S340P) was 4.77 °C higher than that of the wild

\* Corresponding authors.

E-mail addresses: [zidingzhang@cau.edu.cn](mailto:zidingzhang@cau.edu.cn) (Z. Zhang), [tianjian@caas.cn](mailto:tianjian@caas.cn) (J. Tian).

<sup>1</sup> These authors have contributed equally to this work.

type (WT) enzyme. Since all designed mutations were located at non-conservative positions, the optimal temperature for each of the three mutants was also similar to that of the WT. The success of this rational design strategy for PHA indicates that it is an efficient method to design the protein, and that it can also be used for other proteins.

## 2. Materials and methods

### 2.1. Data set

Thermodynamic Database for Proteins and Mutants (Protherm) is a database that contains numerical thermodynamic parameters for proteins and their mutants [21]. From the Protherm dataset, we gathered 42 wild type proteins and these positive and negative mutants, for which  $\Delta T_m$  has been measured experimentally. The 3D structures of those proteins were downloaded from the Protein Data bank, <http://www.rcsb.org> [22]. Among them, the  $\Delta T_m$  values of positive mutants are greater 1 °C than that of the WT protein, and  $\Delta T_m$  values of negative mutants are less –1 °C than that of the WT protein.

### 2.2. MD simulation

The three-dimensional structures of four alpha-amylases and the 42 wild type proteins were downloaded from the PDB database. All non-protein and hydrogen atoms were removed and the protein structure was prepared using Discovery studio 2.5. Protein molecules were placed in cubic boxes at a minimum of 12 Å distance from the edge and were solvated using TIP3P [23] explicit water. The system was neutralized with 0.15 mol per liter of sodium chloride using VMD 1.9.2 [24], where approximate density was determined by the density of water at the corresponding temperatures.

The MD simulation was performed using NAMD 2.12 [25] with the CHARMM22 [26] force field. Simulations were carried out using periodic boundary conditions and a 12 Å cut-off for non-bonded interactions as well as a Particle Mesh Ewald for long-range electrostatics. A time step of 2 fs was used, and snapshots were saved every 1 ps. The protein was fully constrained, and the solvent minimized for 2000 steps using a conjugate gradient algorithm. The solvent was equilibrated for 100 ps under constant-temperature, constant-pressure NPT conditions. The solvent was then fully constrained and the protein minimized for 2000 steps. The entire system was then minimized for another 2000 steps and equilibrated for 100 ps under the same NPT conditions. Finally, a 20-ns MD simulation was performed at 300 K, 350 K, 400 K or 500 K in constant-energy, constant-volume ensemble NVE conditions.

### 2.3. Feature selection and Support Vector Machine (SVM) machine learning model

The structural conformations of simulated protein were extracted every 0.2 ns from the total 20 ns MD trajectory. Therefore, the 100 frames' structures were collected from each molecular simulation trajectory. As shown in Table S1, the wild type structure in each frame was mutated to the corresponding positive and negative mutants. Then the unfolding free energy change of each mutant was calculated with software FoldX 4.0 [27] and selected as the training features.

To build a suitable SVM machine learning model, the calculated  $\Delta\Delta G_u$  was selected as feature for SVM machine learning. The R package e1071 was used as the SVM machine learning package. The gaussian kernel was selected. The grid search was used for the parameter optimization of SVM machine learning, and the opti-

mal cost parameter and gamma parameter were used for the SVM machine learning model. On consideration of the small sample size, the 42-fold cross validation was conducted to evaluate the performance of the models.

### 2.4. Rational design of alpha-amylase mutations

In order to search for positive mutations, a position scanning protocol was carried out on protein structures that were extracted from the trajectory of the MD simulation. To this end, the unfolding free energy of each mutant was calculated as a criterion of thermostability. The scanning of mutations and the energy calculations were carried out using FoldX 4.0 [27].

K-Means clustering was applied to determine which clusters with lower interaction energy amino acid sites and two-clusters would be chosen for further sites selection (Fig. 1). In order to avoid mutations that had occurred in the conservative, binding and active sites, the NCBI analysis tool cd-search [28,29] was used to determine the location of those sites (Fig. S1). Finally, the locus of conservative residues (His<sup>7</sup>, Cys<sup>20</sup>, Gln<sup>33</sup>, Ser<sup>35</sup>, Asn<sup>38</sup>, His<sup>40</sup>, Gln<sup>51</sup>, Cys<sup>74</sup>, Asn<sup>88</sup>, Asp<sup>144</sup>, Asp<sup>174</sup>, Ala<sup>175</sup>, His<sup>178</sup>, Ala<sup>181</sup>, Gln<sup>199</sup>, Thr<sup>221</sup>, His<sup>263</sup>, Ala<sup>287</sup>, Ser<sup>304</sup> and Gln<sup>368</sup>) as well as the sites within five Å of active sites (Thr<sup>85</sup>, Asn<sup>88</sup>, Ala<sup>175</sup>, Ser<sup>176</sup>, Gln<sup>199</sup>, Thr<sup>221</sup>, His<sup>263</sup>) were avoided.

### 2.5. MD analysis methods

Analysis of the trajectory was performed with VMD 1.9.2. The protein's structural thermostability was further examined using the time evolution of the number of backbone H-bonds as well as the energy between one amino acid and the others. A hydrogen bond was considered to be formed if the distance between N and O was  $\leq 3.2$  Å and the angle of N–H...O was  $\geq 120^\circ$ . The energy between one amino acid and the others contains the energy of hydrogen bonds, ionic bonds and van der Waals forces. Images of hydrogen bond and energy analysis were generated by gnuplot 5.2, while the protein 3D structure diagram was generated using discovery studio 2.5.

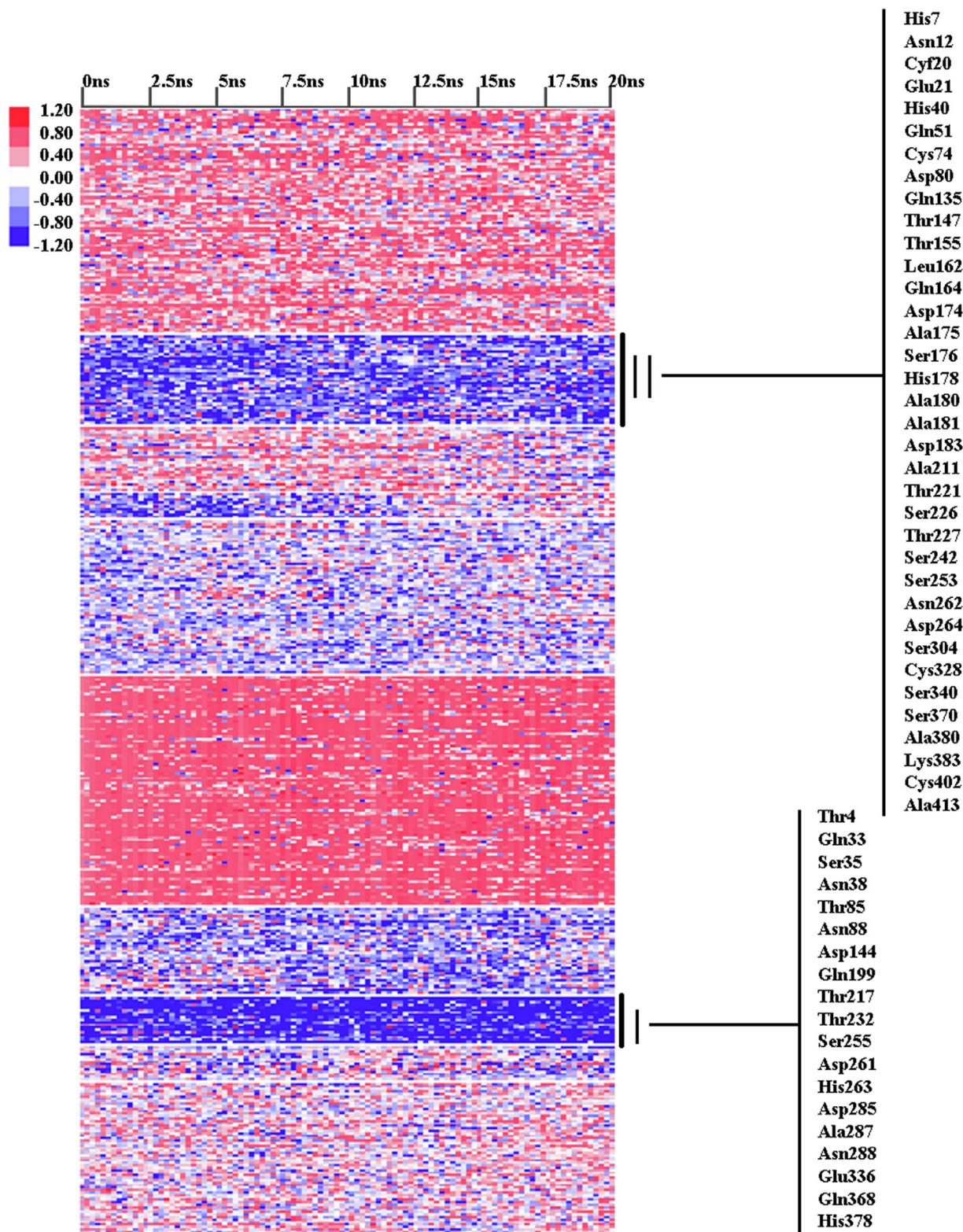
### 2.6. Bacterial strains, plasmids, culture conditions and chemicals

Purchased from Kangweishiji (Beijing, China), *E. coli* TOP10 and *E. coli* BL21 (DE3) were utilized as the hosts for cloning and gene expression, respectively. The plasmid pET-22b(+)-PHA harboring the gene PHA was synthesized by the GenScript Corporation (Nanjing, China), and the gene was optimized using Presyncodon [30]. The GenBank accession number of PHA gene and protein were MK627521 and CAA41481, respectively. All of the bacteria were cultivated at 37 °C or 16 °C in Luria Bertani (LB) medium. Both ampicillin (50 µg/mL) and IPTG (isopropyl β-D-1-thiogalactopyranoside) (1 mM/mL) were added to the growth medium when necessary. All chemicals used were analytical grade.

### 2.7. Site-directed mutagenesis

Mutations causing amino acid exchanges were performed using a two-step PCR method. Primers containing mutation sites are listed in Table S2. During the first phase, a DNA fragment which contained the mutation site was amplified using a typical PCR reaction: 1.5 µL Taqplus DNA polymerase (Tiangen, Cor. Beijing, China), 5 µL buffer, 200 mM dNTPs, 50 ng of pET22b(+)-PHA plasmid, 1 µL of each primer, followed by addition of ddH<sub>2</sub>O to 50 µL.

Thermal cycling was performed as follows: 95 °C for five minutes, followed by 30 cycles at 95 °C for 30 s, 55 °C for 30 s, 72 °C for one minute, and an elongation step at 72 °C for 10 min. DNA fragments were gel-purified using an AxyPrep DNA Gel Extraction



**Fig. 1.** The K-Means clustering of the unfolding free energy of mutants following MD simulation sampling. The x-coordinate is the time axis from 0 ns to 20 ns. The vertical axis represents different amino acid sites. Each square represents a mutant (saturation mutations were performed at each amino acid site, leaving the mutants with the lowest energy values): blue indicates that the energy of the mutant has been reduced relative to that of the wild type, while red indicates that the energy of the mutant has been increased relative to that of the wild type. The darker the color, the greater the variation.

Kit (Corning, Hangzhou, China). 500 ng of the fragments which were produced in the first PCR reaction were then used as megaprimers: 50 ng of pET22b(+)-PHA plasmid as template, 1.5  $\mu$ L FastPfu DNA polymerase (TransGen, Beijing, China), 10  $\mu$ L buffer, 200 mM dNTPs, followed by the addition of ddH<sub>2</sub>O to 50  $\mu$ L.

The following amplification program used was: 95 °C for five minutes, then 30 cycles at 95 °C for 30 s and 68 °C for five minutes, followed by an elongation step at 72 °C for 16 min. The original template was then digested using *DpnI* (NEB, USA) at 37 °C for three hours. Recycling purification of digestion products was undertaken

using a Universal DNA Purification Kit (Tiangen Cor. Beijing, China). To verify incorporation of the mutant genes, DNA sequencing was performed.

### 2.8. Protein expression, purification and analysis

Recombinant strains were grown overnight at 37 °C in Luria-Bertani (LB) medium supplemented with 50 µg/mL ampicillin. Following this, cultures were inoculated into 100 mL of LB medium containing 50 µg/mL ampicillin by 1% dilution and then grown at 37 °C with shaking at 200 rpm. When OD<sub>600</sub> levels reached 0.6–0.8, IPTG was added to a final concentration of 1.0 mM and then grown at 16 °C with shaking at 200 rpm for 16 h. Cells were harvested by centrifugation at 8000 rpm for five minutes, and then resuspended in 20 mM Tris-HCl buffer (pH 7.4) to be broken by an ultrasonicator. Supernatant was then collected from the cells which had been disrupted by being centrifuged at 12,000 rpm at 4 °C for 30 min. All the C-terminal His-tagged proteins were purified using an Ni-NTA agarose (Qiagen, USA) column. Purified proteins were stored in 20 mM Tris-HCl buffer (pH 7.4) at –20 °C. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis was undertaken using a 12% running gel. Protein concentration was measured with a BCA Protein Assay kit (CWBI, Beijing, China).

### 2.9. Measurement of enzymatic activity

The alpha-amylase enzyme activity was determined according to the modified method described previously [31]. One unit of enzyme was deemed to be the amount of amylase needed to release 1 mmol of reducing sugar per minute at 30 °C, pH 7.0, and 5 mM Ca<sup>2+</sup>. Reactions for both the WT and mutant enzymes were conducted using 450 µL 1% soluble starch (Sigma, USA) as a substrate and 50 µL enzyme for 10 min at 30 °C, pH 7.0, and 5 mM Ca<sup>2+</sup>. After color development, the reaction mixture was monitored at 550 nm using a SpectraMax M2 (Molecular Devices, USA) microplate reader.

### 2.10. Analysis of enzyme thermostability

In order to determine the thermostability of enzymes, they were diluted to the same concentration, and then incubated at 40 °C for 0, 20, 30, 40, 50, 60, 70 and 80 min respectively, and were then rapidly cooled. Finally, the residual activity was measured in the manner stated above. The activity of the enzyme without incubation was defined as 100%. Melting temperature ( $T_m$ ) was detected using a differential scanning calorimeter (DSC) (GE Healthcare, USA). Signal baseline drift was eliminated by allowing the DSC equipment to warm up for 10 min prior to use, with the temperature then increased from 10 °C to 70 °C at a rate of 10 °C/h.

### 2.11. Effect of temperature and pH on enzyme activity

The optimum temperatures of both the WT and mutants alpha-amylases were determined at temperatures ranging from 0 °C to 50 °C in sodium phosphate buffer (pH 7.0) and 5 mM Ca<sup>2+</sup>. The activity at optimum temperature was defined as 100%. The optimum pH of both the WT and the mutants were determined at pH values ranging from four to nine at 30 °C and 5 mM Ca<sup>2+</sup>. Activity of optimum pH was defined as 100%.

### 2.12. Analysis of kinetic parameters for WT and mutants

The kinetic parameters of the WT and the mutant enzymes were measured at the optimal temperature of 30 °C and pH 7.0, using 0, 1, 2, 3, 4, 5, 6, 8, 10, 12 and 14 mg/mL soluble starch as

a substrate. Each enzyme activity assay was undertaken with three replicates. The steady-state kinetic parameters  $K_m$  and  $k_{cat}$  of the WT and the mutants were determined through fitting the Michaelis–Menten equation to the initial velocity data using the GraphPad Prism (GraphPad Software Inc; La Jolla, CA, USA).

## 3. Results

### 3.1. MD simulations

In order to evaluate the performance of MD simulation on the alpha-amylase, four alpha-amylases were chosen and simulated with the MD tool NAMD at four temperatures (300 K, 350 K, 400 K and 500 K) and three replications, respectively. The four alpha-amylases were the psychrophilic alpha-amylase from *Pseudomonas haloplanktis* (PDB id: 1AQH [32], PHA), for which the optimal temperature ranges from 15 to 20 °C, the mesophilic alpha-amylase from *Bacillus amyloliquefaciens* (PDB id: 3BH4 [33], BAA) for which the optimal temperature is approximately 65 °C, the thermophilic alpha-amylase from *Bacillus licheniformis* (PDB id: 1BLI [34], BLA) for which the optimal temperature is approximately 90 °C, and the hyper-thermophilic alpha-amylase from *Pyrococcus woesei* (PDB id: 1MWO [35], PWA), for which the optimal temperature is approximately 98 °C.

The average RMSDs for all of the atoms for the three replicated MD simulations' trajectories at the four temperatures were calculated. As can be seen in Fig. S2, the simulated average RMSD at a high temperature was greater than that of a low temperature, suggesting that the average RMSD of the four alpha-amylases is sensitive to simulated temperature. In addition, the RMSDs of the PHA at all four temperatures were greater than that of the other three alpha-amylases (BAA, BLA and PWA). This result suggests that the thermostability of the PHA was lower than that of other enzymes. At each of the four simulated temperatures, the Pearson's correlation coefficient between simulated RMSD at 400 K and optimal temperature was –0.987, which was higher than the other three temperatures (Fig. S2). Additionally, the simulated temperatures were extracted from the simulation system in the NVE condition, which tended to be stable and only fluctuated between 398 K and 408 K (Fig. S3). Therefore, the simulated trajectories at 400 K reflected an appropriate conformation change and accordingly 400 K was set as the simulated temperature for the following simulations.

### 3.2. Correlation analysis between $\Delta\Delta G_u$ and $\Delta T_m$

The 42 proteins with positive and negative mutants were collected from the ProTherm database (Table S1). These 42 proteins were simulated with the software NAMD and the 100 frames' wild type structures were extracted from each molecular simulation trajectory. As shown in Table S1, each frame structure was mutated into the corresponding positive and negative mutants. Then the unfolding free energy change ( $\Delta\Delta G_u$ ) of each mutant was calculated with software FoldX. Finally, the Pearson correlation coefficient between the  $\Delta T_m$  values of the mutant and the calculated  $\Delta\Delta G_u$  values of the mutants were analyzed. As shown in Fig. 2, the red dot represents the Pearson correlation coefficient of the crystal structure and the other dots represent the Pearson correlation coefficients of the mutant structure obtained at different simulation time in the trajectory of the molecular simulation. As shown in Fig. 2, the Pearson correlation coefficients between the  $\Delta T_m$  values and calculated  $\Delta\Delta G_u$  values tend to be strengthened. This indicates that the structure simulated by molecular dynamics can provide more information about the protein stability.

### 3.3. Screening of stabilizing mutants in silico

To search for stabilizing (positive) mutations of PHA, 100 conformation structures of PHA were extracted in each trajectory of the MD simulation, which were then used to scan the free energy change of all possible single-point mutations. All free energy changes of the mutation were calculated in seven replications and clustered using the K-mean method. As shown in Fig. 1, different conformation structures obtained different screening results of the positive mutation, suggesting that the input structure was a key parameter for impacting the screen result. However, two regions (Fig. 1) showed consistently low values for the folding free energy changes ( $\Delta\Delta G^{\text{Fold}}$ ) [36]. Therefore, mutations in these two regions were selected as being potentially positive. Having eliminated positions located near the active and conservative sites, 10 mutants were selected to evaluate their effect on protein thermostability, measured by the following experiment.

### 3.4. Preliminary screening of mutants with normal activity

The genes of 10 mutants and WT were constructed in the expression plasmid pET-22b(+). The protein expression was induced overnight at 16 °C in *E. coli* strain BL21(DE3). The activities of the expressed WT and mutant proteins were measured in order to evaluate the effect of the mutations on protein function. As listed in Table 1, the enzyme activities of the five mutants (T4I, T147R, S226P, N288F and K383L) were lower than that of the WT. Considering that their potential industry applications were limited, they were not included in the following study. Thus, only five mutants (N63D, T217L, H378F, S255K and S340P) were retained for further evaluation.

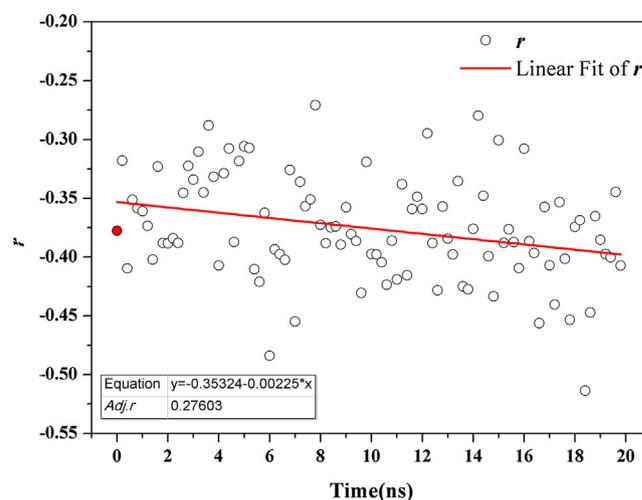
### 3.5. Thermostability assays of WT and mutants

In order to determine the thermostability of the five mutants, we purified the proteins of the five mutants and WT with a Ni<sup>2+</sup>-chelating affinity column (Fig. S4) and measured the  $T_m$  values of those proteins with DSC.  $T_m$  values of N63D, T217L and H378F reduced by 0.77 °C, 0.50 °C and 3.70 °C, while  $T_m$  values of S255K and S340P increased by 1.52 °C and 3.38 °C respectively (Table 2) when compared to the WT protein. This indicates that two stabilizing mutations (S255K and S340P) were identified.

A mutant with double mutations (S255K and S340P) was constructed. The protein of this double-point mutant was then purified to evaluate the cumulative effect of the two mutations. The  $T_m$  value of double-point mutant (S255K/S340P) was 43.47 °C, which was 4.77 °C higher (Table 2) than that of the WT. Following the incubation at 40 °C for 80 min, the residual activities of S255K, S340P and S255K/S340P retained 46%, 69% and 76% of the maximal activity and were found to be 1.6-, 2.4-, and 2.6-fold higher than that of the WT (Fig. 3). These results suggest that the two substitutions (S255K and S340P) improved the thermostability of WT, and the cumulative effect of S255K and S340P existed.

### 3.6. Temperature and pH dependency of WT and mutants

The pH dependence of WT and mutant activities was investigated in various buffers at pH values ranging from four to nine, under standard assay conditions. Results show that the optimal pH of all variants was 7.0, similar to that of the WT (Fig. 4A). The effect of temperature on WT and mutant activities was measured at various temperatures ranging from 0 to 50 °C, under standard assay conditions. Results show that the optimal temperature of variants was 30 °C, again similar to that of the WT (Fig. 4B). Therefore, the substitution of amino acid at 255 and 340 had no effect on either optimal pH or temperature. Additionally, proteins WT,



**Fig. 2.** The correlation coefficient between  $\Delta T_m$  and  $\Delta\Delta G_d$  varies with simulation time. The red dot represents the Pearson correlation coefficient corresponding to the mutant structure without molecular simulation and the other dots represent the Pearson correlation coefficients of the mutant structure obtained at different moments of molecular simulation.

**Table 1**

The difference in enzyme activity and thermostability between selected mutants and wild type.

Selected sites	Wild	Mutation	Result
4	THR	ILE	Enzyme activity <sup>↓↓</sup>
63	ASN	ASP	Thermostability <sup>↓</sup>
147	THR	ARG	Enzyme activity <sup>↓↓</sup>
217	THR	LEU	Thermostability <sup>↓</sup>
226	SER	PRO	Enzyme activity <sup>↓↓</sup>
255	SER	LYS	Thermostability <sup>↑</sup>
288	ASN	PHE	Enzyme activity <sup>↓↓</sup>
340	SER	PRO	Thermostability <sup>↑↑</sup>
378	HIS	PHE	Thermostability <sup>↓</sup>
383	LYS	LEU	Enzyme activity <sup>↓↓</sup>

“↓”: Decreased activity or decreased thermostability.

“↑”: Increased activity or increased thermostability.

S255K, S340P and S255K/S340P had approximately 15, 18, 20 and 24% activity relative to the corresponding maximal value at 0 °C and less than 10% residual activity at 50 °C, suggesting that both the WT and variants had the typical traits of cold-adapted enzymes.

### 3.7. Kinetic parameters of WT and mutants

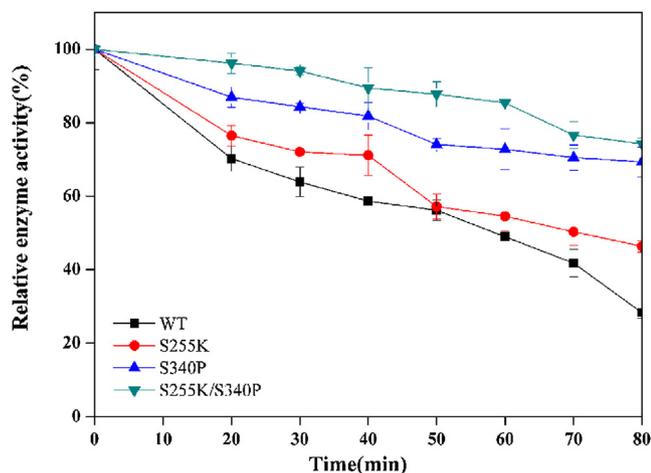
The kinetic parameters ( $K_m$  and  $k_{\text{cat}}$ ) of both the WT and mutants were measured at pH 7.0 and 30 °C with 11 different substrate concentrations of soluble starch, from zero to 14 mg/mL. Compared with the WT, the catalytic efficiency ( $k_{\text{cat}}/K_m$ ) of S255K, S340P and S255K/S340P increased by 88%, 105% and 173% (Table 3) respectively, suggesting that the three mutants not only enhanced their thermostability, but also improved catalytic efficiency. Further, the cumulative effect of S255K and S340P improved the catalytic efficiency of the S255K/S340P mutant.

### 3.8. Mutations on residues S255 and S340

In order to test whether or not the mutation S255K and S340P was best suited to be in those positions, five typical amino acids (Alanine, Glutamine acid, Leucine, Arginine and Tyrosine) were

**Table 2**The  $T_m$  values of the WT and mutants.

	WT	S255K	S340P	S255K/S340P	N63D	T217L	H378F
$T_m$ (°C)	38.7 ± 0.3	40.22 ± 0.1	42.08 ± 0.4	43.47 ± 0.3	37.93 ± 0.5	38.23 ± 0.3	35 ± 0.5



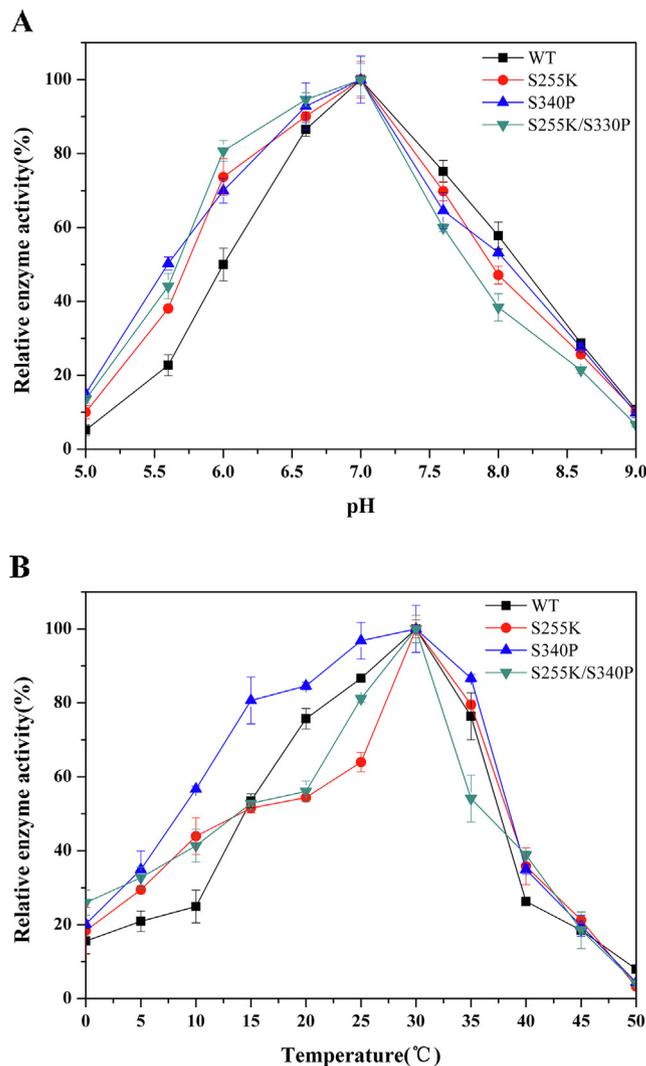
**Fig. 3.** The thermostability assays of the WT and mutants. Enzymes were incubated at 40 °C and pH 7.0 for different periods, and the residual activity of WT and mutants was measured. Activity of the enzyme without incubation was defined as 100%. Data points correspond to mean values of three independent experiments.

mutated in the residues S255 and S340 respectively. All 10 mutations were constructed, and the corresponding proteins were expressed and purified. As shown in Table S3, the  $T_m$  values of all 10 mutants were lower than that of the WT, indicating that Lysine and Proline were probably the most suitable options of residues S255 and S340, and that our rational design method is a powerful method for selecting the best suited mutations for improving thermostability.

### 3.9. Interaction energy analysis

In order to investigate what structural factors of S255K and S340P contributed to the protein thermostability, we analyzed interaction energy between amino acids and other residues of the simulated conformations from MD. As shown in Fig. S5, the amino acid interaction energy map of the WT protein revealed a high interaction energy area (Glu<sup>118</sup>, Glu<sup>138</sup>, Asp<sup>144</sup> and Asp<sup>146</sup>), suggesting that these residues were negative on protein thermostability.

Structure analysis of PHA indicated that the four amino acids (Glu<sup>118</sup>, Glu<sup>138</sup>, Asp<sup>144</sup> and Asp<sup>146</sup>) were located at Domain B, which is one of the three structural domains of alpha amylase (Fig. 5A) and is responsible for maintaining protein thermostability [37,38]. Moreover, this region is located on the surface of the protein and close to three functionally important areas. These three areas contain two substrate binding areas (Trp<sup>48</sup> to Gln<sup>51</sup> and Lys<sup>177</sup> to His<sup>178</sup>) and one catalytic activity area (Asp<sup>174</sup>, Glu<sup>200</sup>, Asp<sup>264</sup>) (Fig. 5B). The cumulative energy among the different simulated conformations reflected the protein interaction energy map and found which residues related to the protein thermostability. As shown in Fig. 6, the mutations S255K and S340P optimized the amino acid interaction energies, and found that the amino acid interaction energies in residues Glu<sup>118</sup>, Glu<sup>138</sup>, Asp<sup>144</sup> and Asp<sup>146</sup> were lower than that of the WT protein. Additionally, mutation



**Fig. 4.** The effect of temperature and pH on the activity of the WT and mutants. For each enzyme, total activity at the optimal temperature and pH were set at 100%. A = optimal pH, while B = optimal temperature. Data points correspond to the mean values of all three independent experiments.

**Table 3**

The kinetic parameters of the WT and mutants.

	$K_m$ (mg/mL)	$k_{cat}$ (min <sup>-1</sup> )	$k_{cat}/K_m$ (mL/mg·min)
WT	10.4 ± 0.10	7.67 ± 0.05	0.74 ± 0.01
S255K	26.84 ± 0.84	37.26 ± 0.92	1.39 ± 0.05
S340P	14.73 ± 0.55	22.35 ± 0.75	1.52 ± 0.03
S255K/S340P	14.24 ± 0.85	28.72 ± 1.01	2.02 ± 0.07

S255K optimized the interaction energy between the residues Ser<sup>255</sup> and the other residues (Fig. 6). Analysis of interaction energy showed that the double mutation (S255K/S340P) combined the energy contribution of S255K and S340P, which in turn optimized

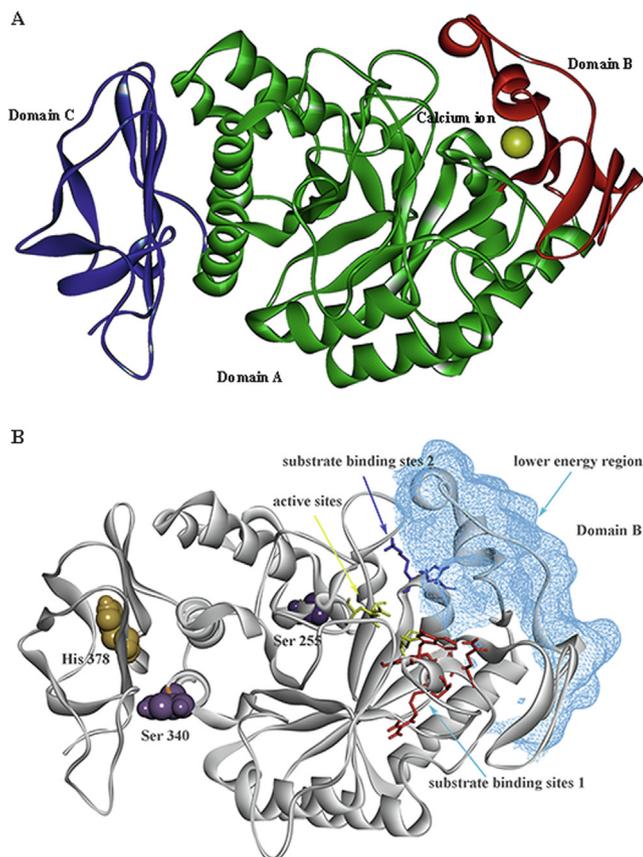


Fig. 5. (A) Distribution of three domains of alpha-amylase. (B) Location of the lower energy region and its neighboring functionally important sites.

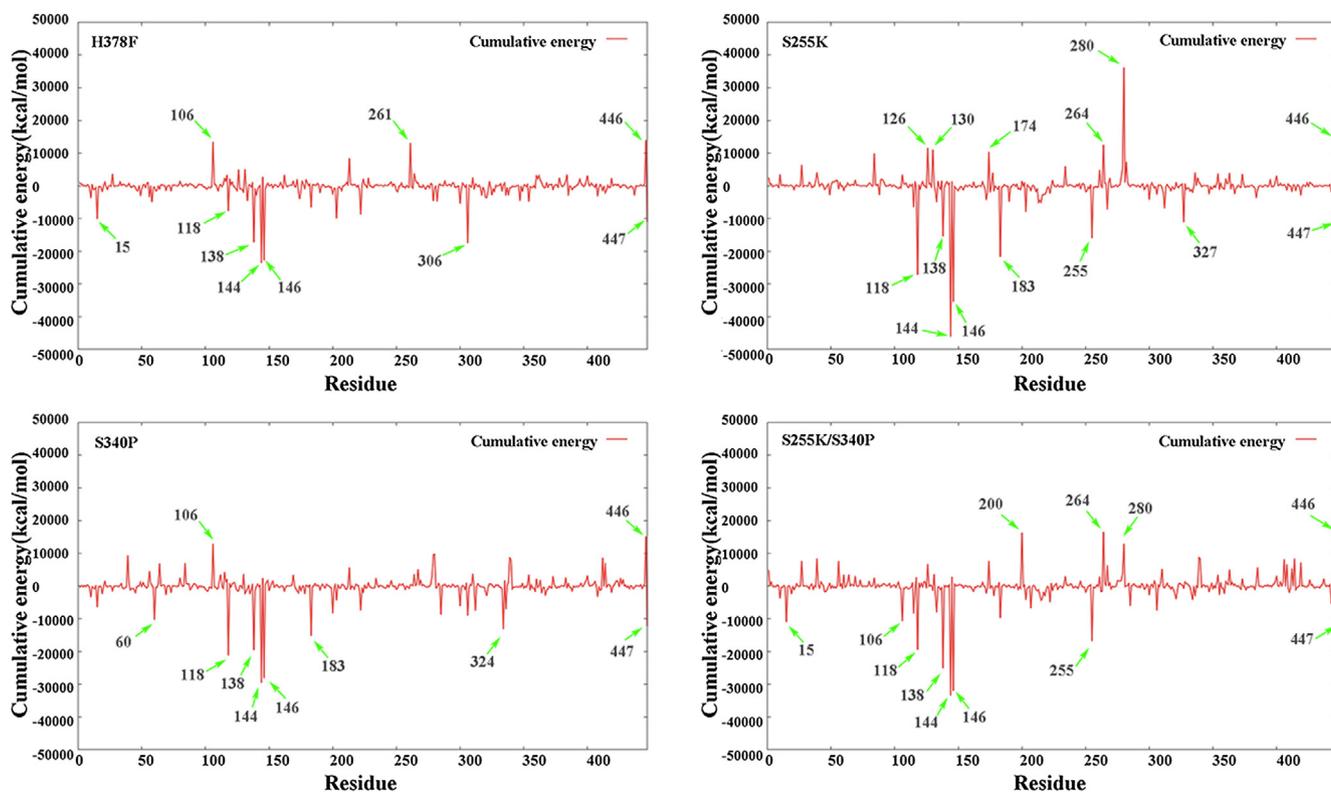


Fig. 6. The cumulative change of energy at the amino acid sites in molecular simulation. The x-coordinate is the amino acid site. The ordinate is the energy accumulation, which represents a cumulative change of energy at each amino acid site in the molecular simulation process, relative to the previous time. The amino acid sites which have accumulative energy values greater than 10,000 kcal/mol or less than -10,000 kcal/mol are marked with green arrows to indicate that they are sites with significant changes.

interaction energies in the residues (Glu<sup>118</sup>, Glu<sup>138</sup>, Asp<sup>144</sup>, Asp<sup>146</sup> and Ser<sup>255</sup>). Therefore, the mutations S255K and S340P give a positive effect on protein thermostability by optimizing the structural energy.

#### 4. Discussion

Protein engineering strategies have been frequently used to optimize enzyme traits. Both rational design and directed evolution have also been used to tailor enzyme properties [39,40]. Rational design has often increased protein stabilities through the rigidification of flexible sites, improving the hydrophobic packing of the protein core [41], creating a favorable network of both positive and negative charges at the protein surface [42] and engineering stabilizing disulfide bridges into proteins [43].

In the process of our previous method exploration, the classification and regression results of SVM were used as a support for our method and also provided us with a new direction for engineering enzyme stability. As shown in Fig. 7, the AUC value of the ROC curve and predicted accuracy of the SVM model reached 0.73, and 0.70, respectively. The regression prediction presented that the correlation coefficient between  $\Delta T_m$  measured in the experiment and  $\Delta T_m$  predicted by SVM was 0.42 (Fig. 8). Considering that our sample size is relatively small, with only 42 pairs of positive and negative mutants, the result of SVM machine learning is pretty good. If we increase the number of samples and try more machine learning methods, we may get a more ideal result. Therefore, the method of combining MD with machine learning can also be regarded as an engineering method to enhanced thermostability and these contents and results will be reflected in our next work.

And in the current study, we proposed a rational design method which combined MD simulations and structural energy optimization to design site-directed mutations of PHA with improved ther-

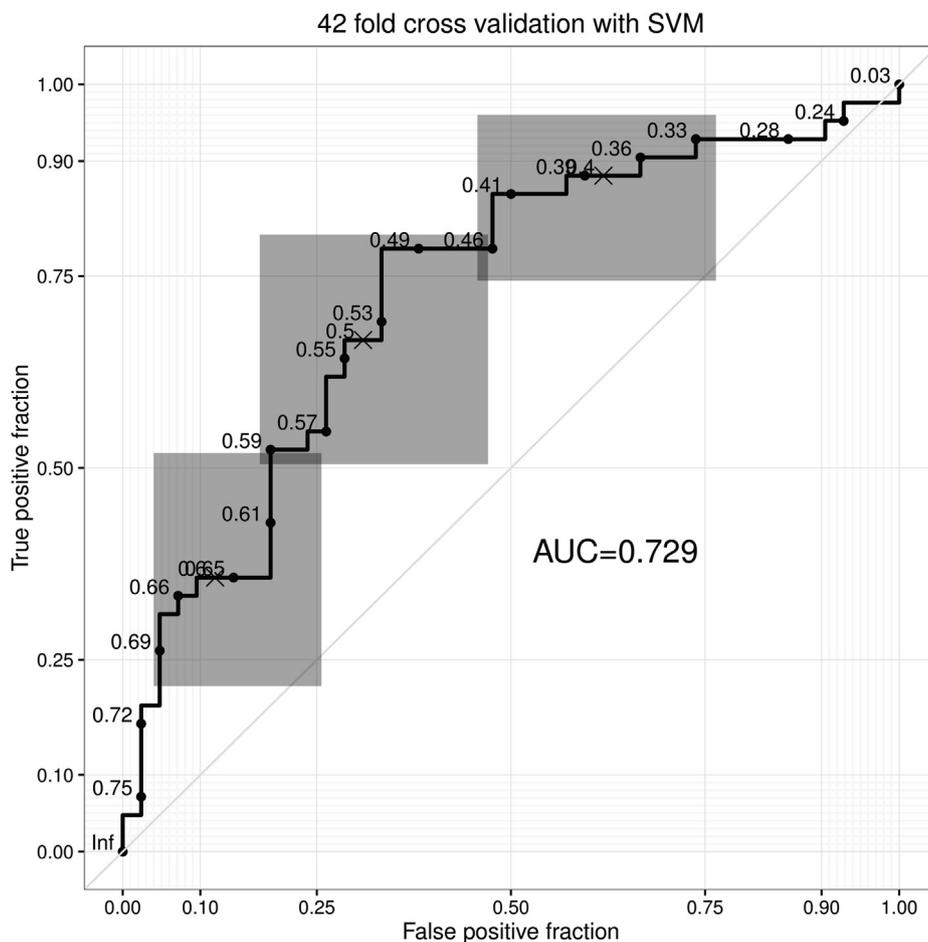


Fig. 7. The ROC curve for SVM classification with 42 proteins with positive and negative mutants and the AUC value is 0.73.

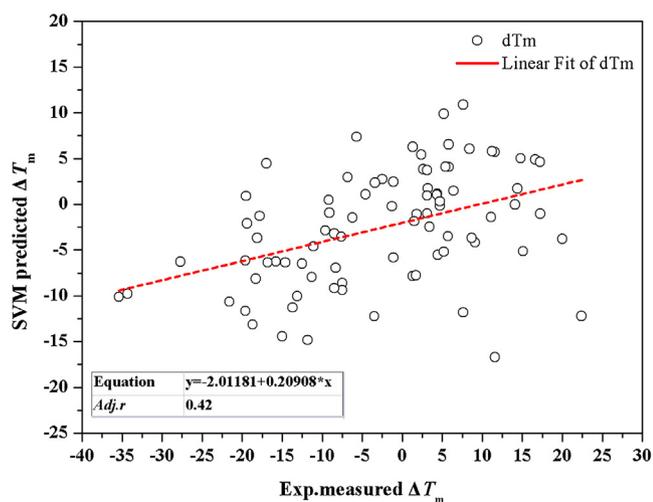


Fig. 8. Linear fitting between  $\Delta T_m$  predicted by SVM and  $\Delta T_m$  measured by experiment. The correlation coefficient between  $\Delta T_m$  predicted by SVM and  $\Delta T_m$  measured by experiment reached 0.42.

mostability. This design strategy was based on energy calculations and identified 10 single-point mutations showing predicted stabilizing effects. When compared to directed evolution, in which  $>10^4$  variants were needed to be screened, our strategy greatly reduced the amount of experimental screening. When compared with other

rational design methods, our strategy demonstrated an obvious improvement. The majority of existing methods have attempted to increase the protein structure's thermostability through crystal structure of proteins [12–14]. The protein crystal structure deposited in the PDB database is simply a special conformation of the protein, which is different to the real structure at high temperatures. Therefore, the design of mutations which enhances the thermostability *in silico* must consider various protein conformations, especially the conformation at high temperatures.

Following *in silico* design and evaluation, two useful mutations (S255K and S340P) and a double mutation (S255K/S340P) were identified. All three of these mutations were located in the non-conservative region, far away the active center, and enhanced thermostability without affecting its activity and optimal catalyzing condition. High levels of hydrolytic activity at low temperatures are a characteristic of this enzyme. However, the increase of thermostability caused by mutations is often accompanied by a change of optimal temperature [5,43], which will in turn lead to energy consumption in industrial applications. Mutations may also lead to changes in optimal pH values [43]. However, the optimal pH value of our mutants is consistent with that of the WT protein, and to some extent expands the optimal pH range of the enzyme. Therefore, when compared with previous studies, the mutations we obtained are better suitable for industrial applications.

It is interesting that the mutation H378F of PHA resulted in instability, although could optimize the structure interaction energy (Fig. S6). Indeed, there exist two stable hydrogen bonds between His<sup>378</sup> and its neighboring residues (Ser<sup>370</sup> and Thr<sup>394</sup>) (Fig. S7). These two hydrogen bonds are located at the cavity of

the C-domain and link the two folded sheets (Fig. S7B). While the mutation H378F may optimize residue interaction energy, it also has the potential to destroy these two hydrogen bonds between His<sup>378</sup> and Ser<sup>370</sup>/Thr<sup>394</sup>, and thus impair the protein's thermostability (Fig. S7C). Although our method considered the amino acid interaction energy, it ignored the key hydrogen bond in the structure, which could lead to some false positive mutations. Therefore, future development of our method should integrate the key structural bond analysis and other parameters in order to design the positive mutation with higher accuracy.

## Funding

This work was supported by the National Natural Science Foundation of China (NSFC, Grant 31371748) and Opening Project of the State Key Laboratory of Agrobiotechnology (2018SKLAB6-21).

## Ethical statement

This article does not contain any studies with human participants or animals performed by any of the authors.

## Declaration of Competing Interest

None.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijbiomac.2019.10.004>.

## References

- [1] S. Janecek, B. Svensson, E.A. MacGregor, Alpha-amylase: an enzyme specificity found in various families of glycoside hydrolases, *Cell. Mol. Life Sci.* 71 (7) (2014) 1149–1170, <https://doi.org/10.1007/s00018-013-1388-z>.
- [2] V. Lombard, H.G. Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, *Nucleic Acids Res.* 42 (D1) (2014) D490–D495, <https://doi.org/10.1093/nar/gkt1178>.
- [3] B. Khemakhem, M. Ben Ali, N. Aghajari, M. Juy, R. Haser, S. Bejar, Engineering of the alpha-amylase from *Geobacillus stearothermophilus* US100 for detergent incorporation, *Biotechnol. Bioeng.* 102 (2) (2009) 380–389, <https://doi.org/10.1002/bit.22083>.
- [4] R. Cavicchioli, T. Charlton, H. Ertan, S.M. Omar, K.S. Siddiqui, T.J. Williams, Biotechnological uses of enzymes from psychrophiles, *Microb. Biotechnol.* 4 (4) (2011) 449–460, <https://doi.org/10.1111/j.1751-7915.2011.00258.x>.
- [5] G. Yang, H. Yao, M. Mozzicafreddo, P. Ballarini, S. Pucciarelli, C. Miceli, Rational engineering of a cold-adapted alpha-amylase from the Antarctic ciliate *Euplotes focardii* for simultaneous improvement of thermostability and catalytic activity, *Appl. Environ. Microb.* 83 (13) (2017), <https://doi.org/10.1128/AEM.00449-17>.
- [6] K.S. Siddiqui, R. Cavicchioli, T. Thomas, Thermodynamic activation properties of elongation factor 2 (EF-2) proteins from psychrotolerant and thermophilic Archaea, *Extremophiles* 6 (2) (2002) 143–150, <https://doi.org/10.1007/s007920100237>.
- [7] R. Cavicchioli, K.S. Siddiqui, D. Andrews, K.R. Sowers, Low-temperature extremophiles and their applications, *Curr. Opin. Biotechnol.* 13 (3) (2002) 253–261, [https://doi.org/10.1016/S0958-1669\(02\)00317-8](https://doi.org/10.1016/S0958-1669(02)00317-8).
- [8] C. Gerday, M. Aittaleb, M. Bentahir, J.P. Chessa, P. Claverie, T. Collins, S. D'Amico, J. Dumont, G. Garsoux, D. Georgette, A. Hoyoux, T. Lonhienne, M.A. Meuwis, G. Feller, Cold-adapted enzymes: from fundamentals to biotechnology, *Trends Biotechnol.* 18 (3) (2000) 103–107, [https://doi.org/10.1016/S0167-7799\(99\)01413-4](https://doi.org/10.1016/S0167-7799(99)01413-4).
- [9] F. Sarmiento, R. Peralta, J.M. Blamey, Cold and hot extremozymes: industrial relevance and current trends, *Front. Bioeng. Biotechnol.* 3 (2015) 148, <https://doi.org/10.3389/fbioe.2015.00148>.
- [10] Y.H. Liu, H. Liu, L. Huang, S. Gui, D. Zheng, L.B. Jia, Y. Fu, F.P. Lu, Improvement in thermostability of an alkaline lipase I from *Penicillium cyclopium* by directed evolution, *RSC Adv.* 7 (61) (2017) 38538–38548, <https://doi.org/10.1039/C7RA06307E>.
- [11] W.H. Yang, Y.Z. Yang, L.D. Zhang, H. Xu, X.J. Guo, X. Yang, B. Dong, Y.H. Cao, Improved thermostability of an acidic xylanase from *Aspergillus sulphureus* by combined disulphide bridge introduction and proline residue substitution, *Sci. Rep.* 7 (2017), <https://doi.org/10.1038/s41598-017-01758-5>.
- [12] M.H. Chang, X.Y. Chu, J.Z. Lv, Q.B. Li, J. Tian, N.F. Wu, Improving the thermostability of acidic pullulanase from *Bacillus naganensis* by rational design, *PLoS One* 11 (10) (2016), <https://doi.org/10.1371/journal.pone.0165006>.
- [13] J. Tian, J.C. Woodard, A. Whitney, E.I. Shakhnovich, Thermal stabilization of dihydrofolate reductase using Monte Carlo unfolding simulations and functional consequences, *PLoS Comput. Biol.* 11 (4) (2015), <https://doi.org/10.1371/journal.pcbi.1004207>.
- [14] B. Jaouadi, N. Aghajari, R. Haser, S. Bejar, Enhancement of the thermostability and the catalytic efficiency of *Bacillus pumilus* CBS protease by site-directed mutagenesis, *Biochimie* 92 (4) (2010) 360–369, <https://doi.org/10.1016/j.biochi.2010.01.008>.
- [15] A. Hospital, J.R. Goni, M. Orozco, J.L. Gelpi, Molecular dynamics simulations: advances and applications, *Adv. Appl. Bioinf. Chem.* 8 (2015) 37–47, <https://doi.org/10.2147/AABC.S70333>.
- [16] M.C. Childers, V. Daggett, Validating molecular dynamics simulations against experimental observables in light of underlying conformational ensembles, *J. Phys. Chem. B* 122 (26) (2018) 6673–6689, <https://doi.org/10.1021/acs.jpcc.8b02144>.
- [17] S. Leone, D. Picone, Molecular dynamics driven design of pH-stabilized mutants of MNEI, a sweet protein, *PLoS One* 11 (6) (2016), <https://doi.org/10.1371/journal.pone.0158372>.
- [18] N.L. Salimi, B. Ho, D.A. Agard, Unfolding simulations reveal the mechanism of extreme unfolding cooperativity in the kinetically stable alpha-lytic protease, *PLoS Comput. Biol.* 6 (2) (2010), <https://doi.org/10.1371/journal.pcbi.1000689>.
- [19] A. Tse, G.M. Verkhrivker, Molecular dynamics simulations and structural network analysis of c-Abl and c-Src kinase core proteins: capturing allosteric mechanisms and communication pathways from residue centrality, *J. Chem. Inf. Model.* 55 (8) (2015) 1645–1662, <https://doi.org/10.1021/acs.jcim.5b00240>.
- [20] Y. Xie, J. An, G.Y. Yang, G. Wu, Y. Zhang, L. Cui, Y. Feng, Enhanced enzyme kinetic stability by increasing rigidity within the active site, *J. Biol. Chem.* 289 (11) (2014) 7994–8006, <https://doi.org/10.1074/jbc.M113.536045>.
- [21] M.D. Kumar, K.A. Bava, M.M. Gromiha, P. Prabakaran, K. Kitajima, H. Uedaira, A. Sarai, ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions, *Nucleic Acids Res.* 34 (Database issue) (2006) D204–D206, <https://doi.org/10.1093/nar/gkj103>.
- [22] S.K. Burley, H.M. Berman, G.J. Kleywegt, J.L. Markley, H. Nakamura, S. Velankar, Protein Data Bank (PDB): the single global macromolecular structure archive, *Methods Mol. Biol.* 1607 (2017) 627–641, [https://doi.org/10.1007/978-1-4939-7000-1\\_26](https://doi.org/10.1007/978-1-4939-7000-1_26).
- [23] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (2) (1983) 926–935, <https://doi.org/10.1063/1.445869>.
- [24] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph. Model.* 14 (1) (1996) 33–38, [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [25] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, Scalable molecular dynamics with NAMD, *J. Comput. Chem.* 26 (16) (2005) 1781–1802, <https://doi.org/10.1002/jcc.20289>.
- [26] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B* 102 (18) (1998) 3586–3616, <https://doi.org/10.1021/jp973084f>.
- [27] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, The FoldX web server: an online force field, *Nucleic Acids Res.* 33 (Web Server issue) (2005) W382–W388, <https://doi.org/10.1093/nar/gki387>.
- [28] A. Marchler-Bauer, M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, R. C. Geer, J. He, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, S.H. Bryant, CDD: NCBI's conserved domain database, *Nucleic Acids Res.* 43 (Database issue) (2015) D222–D226, <https://doi.org/10.1093/nar/gku1221>.
- [29] A. Marchler-Bauer, S.H. Bryant, CD-Search: protein domain annotations on the fly, *Nucleic Acids Res.* 32 (Web Server issue) (2004) W327–31, <https://doi.org/10.1093/nar/gkh454>.
- [30] J. Tian, Y.R. Yan, Q.X. Yue, X.Q. Liu, X.Y. Chu, N.F. Wu, Y.L. Fan, Predicting synonymous codon usage and optimizing the heterologous gene for expression in *E. coli*, *Sci. Rep.* 7 (2017), <https://doi.org/10.1038/s41598-017-10546-0>.
- [31] T. Huma, M.H. Rashid, M.R. Javed, A. Ashraf, Gamma ray mediated mutagenesis of *Phialocephala humicola*: effect on kinetics and thermodynamics of alpha-amylase production, *Afr. J. Microbiol. Res.* 6 (22) (2012) 4639–4646, <https://doi.org/10.5897/AJMR11.754>.
- [32] N. Aghajari, G. Feller, C. Gerday, R. Haser, Crystal structures of the psychrophilic alpha-amylase from *Alteromonas haloplacis* in its native form and complexed with an inhibitor, *Protein Sci.* 7 (3) (1998) 564–572, <https://doi.org/10.1002/pro.5560070304>.
- [33] J. Alikhajeh, K. Khajeh, B. Ranjbar, H. Naderi-Manesh, Y.H. Lin, E.H. Liu, H.H. Guan, Y.C. Hsieh, P. Chuankhayan, Y.C. Huang, J. Jeyaraman, M.Y. Liu, C.J. Chen, Structure of *Bacillus amyloliquefaciens* alpha-amylase at high resolution: implications for thermal stability, *Acta Crystallogr. F* 66 (2010) 121–129, <https://doi.org/10.1107/S1744309109051938>.
- [34] M. Machius, N. Declerck, R. Huber, G. Wiegand, Activation of *Bacillus licheniformis* alpha-amylase through a disorder → order transition of the

- substrate-binding site mediated by a calcium-sodium-calcium metal triad, *Structure* 6 (3) (1998) 281–292, [https://doi.org/10.1016/S0969-2126\(98\)00032-X](https://doi.org/10.1016/S0969-2126(98)00032-X).
- [35] A. Linden, O. Mayans, W. Meyer-Klaucke, G. Antranikian, M. Wilmanns, Differential regulation of a hyperthermophilic alpha-amylase with a novel (Ca, Zn) two-metal center by zinc, *J. Biol. Chem.* 278 (11) (2003) 9875–9884, <https://doi.org/10.1074/jbc.M211339200>.
- [36] Y.F. Bu, Y.L. Cui, Y. Peng, M.R. Hu, Y.E. Tian, Y. Tao, B. Wu, Engineering improved thermostability of the GH11 xylanase from *Neocallimastix patriciarum* via computational library design, *Appl. Microbiol. Biotechnol.* 102 (8) (2018) 3675–3685, <https://doi.org/10.1007/s00253-018-8872-1>.
- [37] O. Prakash, N. Jaiswal, Alpha-amylase: an ideal representative of thermostable enzymes, *Appl. Biochem. Biotechnol.* 160 (8) (2010) 2401–2414, <https://doi.org/10.1007/s12010-009-8735-4>.
- [38] J.E. Nielsen, T.V. Borchert, Protein engineering of bacterial alpha-amylases, *BBA-Protein Struct. Mol.* 1543 (2) (2000) 253–274, [https://doi.org/10.1016/S0167-4838\(00\)00240-5](https://doi.org/10.1016/S0167-4838(00)00240-5).
- [39] M.D. Toscano, K.J. Woycechowsky, D. Hilvert, Minimalist active-site redesign: teaching old enzymes new tricks, *Angew. Chem. Int. Ed.* 46 (18) (2007) 3212–3236, <https://doi.org/10.1002/anie.200604205>.
- [40] T.M. Penning, J.M. Jez, Enzyme redesign, *Chem. Rev.* 101 (10) (2001) 3027–3046, <https://doi.org/10.1021/cr000049n>.
- [41] B. Borgo, J.J. Havranek, Automated selection of stabilizing mutations in designed and natural proteins, *Proc. Natl. Acad. Sci. USA* 109 (5) (2012) 1494–1499, <https://doi.org/10.1073/pnas.1115172109>.
- [42] A.V. Gribenko, M.M. Patel, J. Liu, S.A. McCallum, C.Y. Wang, G.I. Makhatadze, Rational stabilization of enzymes by computational redesign of surface charge-charge interactions, *Proc. Natl. Acad. Sci. USA* 106 (8) (2009) 2601–2606, <https://doi.org/10.1073/pnas.0808220106>.
- [43] L. Liu, Z.M. Deng, H.Q. Yang, J.H. Li, H.D. Shin, R.R. Chen, G.C. Du, J. Chen, In silico rational design and systems engineering of disulfide bridges in the catalytic domain of an alkaline alpha-amylase from *Alkalimonas amylolytica* to improve thermostability, *Appl. Environ. Microb.* 80 (3) (2014) 798–807, <https://doi.org/10.1128/AEM.03045-13>.