Improved thermostability of creatinase from Alcaligenes Faecalis through non-biased phylogenetic consensus-guided mutagenesis

Xue Bai
Institute of Biothermal Science and Technology, University of Shanghai for Science and Technology

Daixi Li
Institute of Biothermal Science and Technology, University of Shanghai for Science and Technology

Fuqiang Ma
CAS Key Lab of Bio-Medical Diagnostics, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences

Xi Deng
State Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University School of Life Sciences and Biotechnology

Manjie Luo
Wuhan Hzymes Biotechnology Co., Ltd

Yan Feng
State Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University School of Life Sciences and Biotechnology

Guangyu Yang (yanggy@sjtu.edu.cn)
Shanghai Jiao Tong University School of Life Sciences and Biotechnology

Research

Keywords: creatinase, thermostability, consensus approach, multiple sequence alignment, phylogenetic analysis

DOI: https://doi.org/10.21203/rs.3.rs-49252/v2

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

Background: Enzymatic quantification of creatinine has become an essential method for clinical evaluation of renal function. Although creatinase (CR) is frequently used for this purpose, its poor thermostability severely limits industrial applications. Herein, we report a novel creatinase from Alcaligenes faecalis (afCR) with higher catalytic activity and lower $K_M$ value, than currently used creatinases. Furthermore, we developed a non-biased phylogenetic consensus method to improve the thermostability of afCR.

Results: We applied a non-biased phylogenetic consensus method to identify 59 candidate consensus residues from 24 creatinase family homologs for screening afCR mutants with improved thermostability. Twenty-one amino acids of afCR were selected to mutagenesis and 11 of them exhibited improved thermostability compared to the parent enzyme (afCR-M0). Combination of single-site mutations in sequential screens resulted in a quadruple mutant D17V/T199S/L6P/T251C (M4-2) which showed ~1700-fold enhanced half-life at 57 ºC and a 4.2 ºC higher $T_{50}^{15}$ than that of afCR-M0. The mutant retained catalytic activity equivalent to afCR-M0, and thus showed strong promise for application in creatinine detection. Structural homology modeling revealed a wide range of potential molecular interactions associated with individual mutations that contributed to improving afCR thermostability.

Conclusions: Results of this study clearly demonstrated that the non-biased-phylogenetic consensus design for improvement of thermostability in afCR is effective and promising in improving the thermostability of more enzymes.

1. Background

Creatinine is the final product of phosphocreatine metabolism in humans [1], and has been established as a reliable clinical marker for the determination of renal and muscular dysfunction. One of the most commonly used methods for the detection of creatinine is based on an enzymatic cascade [2] that includes creatininase (E.C.3.5.2.10), creatinase (E.C.3.5.3.3), and sarcosine oxidase (E.C.1.5.3.1) (Fig.1). In this way, creatinine is eventually converted into $\text{H}_2\text{O}_2$, and thus the concentration of creatinine can be determined by converting $\text{H}_2\text{O}_2$ into detectable signal by horseradish peroxidase (Fig.1). In this system, creatinase is the rate-limiting enzyme and several reported creatinases exhibit low catalytic activity and poor thermostability [3-6]. Consequently, genome mining for new homologs of this enzyme, with subsequent protein engineering to improve its catalytic properties and thermostability have recently become the focus of increasing research attention.

Over the years, many different approaches have been used to modify the stability of proteins. Classic protein stability design strategies are usually based on disulfide bonds design [7], optimization of protein surface charges [8], B-factor design [9], Proline effect design [10]. Recently, design tools rely on the energy function or the machine-learning algorithm have been developed to predict changes in protein stability, such as CC/PBSA [11], I-Mutant 2.0 [12], Fold-X [13], MUpro [14], PopMUSIC [15], Rosetta [16]. However,
these approaches usually require a crystal structure of target protein and in-depth understanding of the structure-function relationship, which limited their widely applications.

Consensus design for protein engineering is a method for identifying conserved amino acid residues across a set of homologous sequences to find sites that can serve as strong candidates for mutagenesis to improve thermostability [17-19], and which could make accurate predictions independent of structural information. In 1994, Steipe et al. first proposed the consensus concept and successfully applied it to design stable immunoglobulin variable domains [20]. Subsequently, Lehmann et al. created consensus sequences from a set of homologous phytases to achieve an astonishingly ~26 °C higher melting temperature than the wild-type phytase [21, 22]. Thus far, this method has also been successfully applied to improve the thermostability of sucrose phosphorylase, β-lactamase, and fibronectin type III (FN3) domain [23-25].

Although often effective, the consensus approach suffers from a major flaw, that is, the overrepresentation of one or a few sub-families of homologous sequences in the sequence space, which results in bias in the final consensus sequence. This bias may obscure conserved residues, thereby limiting access to the potentially high thermostable properties offered by less-characterized families and evolutionary lineages [26]. Consequently, a purely statistical approach of simply replacing all non-consensus residues in conserved positions of a sequence or motif with the consensus residue may fail to result in a more stable protein [17]. In a previous study, Bloom et al. demonstrated that utilizing a likelihood-based method to account for bias in a phylogeny can relatively reduce database bias for homologous sequences [27].

In this work, we cloned and characterized a novel creatinase from *Alcaligenes faecalis* (afCR). The enzyme showed relatively high catalytic efficiency and a low $K_M$ value compared with previous reported CR from *Flavobacterium*, *Pseudomonas*, which make it a promising candidate for the applications in the enzymatic determination of creatinine [28, 29]. However, the thermostability of afCR is poor, thus limiting its use with many reagents. To address this issue, we used a non-biased consensus method based on phylogenetic analyses to identify conserved residues to target for improving thermostability of afCR. We determined the optimal conditions for wild-type afCR activity, and generated combinatorial mutants to increase afCR thermostability. In addition, we used structural homology modeling to examine the interactions potentially underlying the improvements conferred by individual mutations, and found a wide range of interactions that increase the thermostability of the creatinase enzyme without reducing its activity. This work modifies our conventional understanding of the effects of some amino acid substitutions on thermostability, and provides a reliable system for similar engineering of other proteins for industrial applications.

2. Results And Discussion

2.1 Cloning, expression, and characterization of wild-type afCR gene
Creatinase is used as a key enzyme for enzymatic measurement of creatinine which catalyzes the hydrolysis of creatine to sarcosine and urea. It has been found in various kinds of bacteria such as *Flavobacterium*, *Pseudomonas*, *Arthrobacter* and *Bacillus* [6, 28, 30-32]. In this work, we analyzed the creatinase family sequences base on phylogenetic tree (Fig.S1) and found a creatinase gene from *Alcaligenes faecalis* in the NCBI database. Its amino acid sequence comparison between creatinase from *Pseudomonas*, *Flavobacterium*, and *Arthrobacter* creatinase showed about 60% homology. We amplified *af*CR gene and cloned it into pANY1 expression vector. The recombinant DNA was expressed in *E.coli* BL21 (DE3), then purified by affinity chromatography and its activity was determined. The results of *af*CR enzymatic characterization showed maximum activity of 14 U/mg (activity was determined by using the 1mg/ml enzyme at 37 °C) and a lower $K_M$ value (23.6 mM), providing a new source of creatinase for the enzymatic determination of creatinine.

2.2 Effects of temperature, pH, and metal ions on wild-type *af*CR activity

In order to provide an initial characterization of the optimal temperature for wild-type *af*CR (WT), we first examined its activity across temperatures ranging 25-55 °C. We found that WT activity gradually increased from 30 °C to 37 °C but was rapidly inactivated at higher temperatures, thus indicating that 37 °C was the optimum temperature for *af*CR reactions (Fig.2A). Therefore, we examined WT activity over a range of reaction pH values (4.5-10.0) to determine at which values it functioned most efficiently (Fig.2B). The results revealed a marked change in activity over this range, steadily increasing up to pH 8.0, after which the rate of *af*CR relative activity declined, with more than 50% of maximum activity observed at a broad pH plateau between 7.5 and 9.0. These results thus demonstrated that pH 8.0 and 37 °C were optimal conditions for *af*CR reactions.

In order to investigate the effects of metal ions on WT *af*CR enzyme stability, the WT was individually incubated with equimolar concentrations of Co$^{2+}$, Fe$^{2+}$, Fe$^{3+}$, Mn$^{2+}$, Zn$^{2+}$, Ca$^{2+}$, Mg$^{2+}$, Cu$^{2+}$, and Na$^+$. The WT exhibited differences in sensitivity to the metal ions (Fig.2C). Specifically, Mn$^{2+}$ and Mg$^{2+}$ ions enhanced WT *af*CR activity. In the presence of EDTA, Fe$^{3+}$, Fe$^{2+}$, Ca$^{2+}$ WT retained >80% of its initial activity, which has little effect on *af*CR activity; in contrast, Cu$^{2+}$, Co$^{2+}$, Zn$^{2+}$ and Na$^+$ ions inhibited WT activity to varying degrees.

In particular, we have observed that the presence of Cu$^{2+}$ have a significant inhibition effect on the activity of *af*CR, which is similar to other previous reports on creatinase [2, 33]. It has been reported in previous studies that the metal-ions-induced enzyme inactivation may be due to protein aggregation [34]. However, the aggregation of *af*CR was not observed in our experiments. We speculated that Cu$^{2+}$ may interact with some key amino acid residues and lead to the loss of catalytic activity of *af*CR. However, the detailed mechanism of this effect requires further study.

2.3 Non-biased phylogenetic consensus method reveals 21 target consensus residues for mutagenesis
In order to employ a non-biased phylogenetic consensus method for identifying specific residues that will most effectively improve thermostability through mutation, we used wild-type afCR as a query for a blastP search of proteins in the NCBI database. This search yielded 45 CR sequences with a sequence similarity greater than 50% (Table S1) compared with wild-type afCR. We then aligned these sequences and deleted duplicates, as well as sequences that were too long or too short. A final set of 24 CR homologous sequences were selected for the construction of phylogenetic tree. In order to reduce the branch bias that may be introduced by the overrepresentation family in the database, we developed a non-biased phylogenetic consensus method based on phylogenetic relationships. To this end, we first constructed a neighbor-joining phylogenetic tree and determined the branch lengths for each sequence, representing the evolutionary distance between CR homologous sequences as a ratio of the number of non-identical residue pairs to the minimum length of the sequence (Fig.3 and Fig.S4).

Thus, the phylogenetic tree serves as a model for CR divergence caused by evolutionary pressures, with branch lengths used to calculate the weight of each sequence. The use of branch weights thus provided statistical independence among the different protein sequences, enabling the identification of conserved target residues by optimizing the occurrence frequency of amino acid residues (Table 1). We used a proprietary consensus wi.py script to introduce branch weights to calculating the consensus sequence (Fig.S2 and Fig.S4). Compared with the wild-type query sequence, 59 of the 404 amino acid positions were selected as candidate consensus residues if they appeared at a given position with >40% frequency among aligned sequences. This consensus cut-off can be adjusted according to the screening method of consensus mutants and the accuracy of other criteria.
In order to further narrow the pool of candidate residues for mutagenesis, we then considered the impacts on protein structure associated with each position of the 59 residues by applying the following criteria: 1) the substitution should be farther than 6 Å away from the active site to avoid the disruption of catalysis; and 2) we excluded amino acids with side chains that directly formed hydrogen bonds or salt bridges with other residues to avoid decreasing protein structure stability. Based on these screening criteria, we selected 21 mutations for further protein engineering by mutagenesis, including E349V, W59F, C331S,
K351E, T251C, F108Y, Y109F, D73T, L162A, V340L, P20T, G58D, Q165I, L6P, V362I, T117P, D17V, T199S, V33L, C52N, and K166A (See Table 1 for residue positions and distance from the active site).

Previously, there are many studies have reported that the distances in the range of 5-10 Å from active sites could be selected to design the mutagenesis library [35-37]. For different enzymes, the optimal distances may depend on their different structure and mechanism, which may not be consistent. Herein, we chose the criteria of > 6 Å based on the structural analysis and substrate docking mode of our creatinase. In fact, our successful design of the enzyme also proved that it is a reasonable consideration.

2.4 Construction and characterization of the afCR mutants

Based on the report of a I304L/F395V double mutant variant of creatinase from Erwinia with lower $K_M$ than its wild type enzyme [38], we first examined the effects of introducing two non-synonymous mutations (I304L/F395V) into WT afCR. We then examined the $K_M$ value for the double mutant (15.3 mM) and found that it was lower than that of the WT afCR (23.6 mM), while retaining similar $K_{cat}$ value. Accordingly, the I304L/F395V afCR variant (afCR-M0) was subsequently used as a parent template to construct more variants. To this end, 21 consensus variants were constructed based on the candidate residues identified above and expressed in E.coli. Eleven of the 21 variants showed improved thermostability compared to afCR-M0 at 55 ºC with 80% or higher activity compared to afCR-M0 (Table 1). The best variant D17V (afCR-M1) exhibited the highest half-life at 55 ºC (150 min), which was 12.9-fold higher than that of afCR-M0 (Fig.2D). The thermostability of the other positive mutants was increased by 1-2-fold compared with that of afCR-M0.

In fact, in previous studies, some researchers have applied consensus approach to design thermostable enzymes, which shown a relatively low design success rate of 20-38% [39-41]. Compared with the other studies that used traditional consensus design method, the non-biased-phylogenetic consensus design method was performed in this work with a higher design success rate of 52%. Analysis of these results also further demonstrated that the non-biased-phylogenetic consensus design is an effective approach and promising in improving the thermostability of more enzymes.

2.5 Combination of beneficial mutations and thermostability of combinatorial variants

In order to further improve the thermostability of afCR-M1, we then gradually introduced additional positive mutations (i.e., L6P, G58D, Q165I, F108Y, Y109F, T117P, T199S, T251C, E349V, and K351E) (see Table 2). We found that thermostability improved with subsequent mutations and that these variants retained 80% or higher activity compared to afCR-M1. Obviously, during the experiment, the half-life of the mutant at 55 ºC increased continuously with the improvement of the thermostability, the half-lives of some mutants became very long. For example, the mutant M2-4 (D17V/T199S) has an half-life ~700 min at 55 ºC, which made it was difficult to be measured accurately. Therefore, we adjusted the temperature to 57 ºC in the subsequent experiment to decrease the difficulty of our experiment.
Table 2

| Enzyme | Mutation       | 57 °C_t/2 (min) | Fold Improvement | Relative Activity (%) |
|--------|----------------|-----------------|-------------------|-----------------------|
| M0     | I304L/F395V    | 2               | 1                 | 100                   |
| M1     | M0 + D17V      | 40              | 20                | 105 ± 7.71            |
| M2-1   | M1 + L6P       | 142             | 71                | 104.76 ± 18.10        |
| M2-2   | M1 + T251C     | 71              | 35.5              | 118.09 ± 3.25         |
| M2-3   | M1 + K351E     | 101             | 50.5              | 161.90 ± 1.90         |
| M2-4   | M1 + T199S     | 210             | 105               | 111.43 ± 2.31         |
| M3-1   | M2-4 + T251C   | 599             | 299.5             | 112.38 ± 2.86         |
| M3-2   | M2-4 + F108Y   | 482             | 241               | 106.67 ± 4.76         |
| M3-3   | M2-4 + K351E   | 859             | 429.5             | 157.61 ± 5.30         |
| M3-4   | M2-4 + L6P     | 1258            | 629               | 123.81 ± 1.90         |
| M4-1   | M3-4 + F108Y   | 2498            | 1249              | 134.28 ± 2.86         |
| M4-2   | M3-4 + T251C   | 3371            | 1685.5            | 110.47 ± 1.90         |

Consequently, among the double mutants, D17V/T199S (afCR-M2-4) showed the highest thermostability, with a half-life of 210 min at 57 °C (~105-fold higher than that of afCR-M0) (Table 2). We then used it as a template to individually introduce other mutations (L6P, T251C, F108Y, Y109F, K351E). All triple mutant variants exhibited significant improvements in thermostability, with the highest increase found in mutant M3-4 (D17V/T199S/L6P), which had a half-life of 1258 min at 57 °C (~629-fold higher than afCR-M0) (Table 2). We then generated variants of M3-4 by introduction of T251C, F108Y, Y109F, and K351E, respectively. This third round of mutagenesis produced variant M4-2 (D17V/T199S/L6P/T251C), which exhibited the highest stability of all variants up to this point, i.e., a half-life of 3371 min at 57 °C (~1685-fold higher than afCR-M0). All the above triple and quadruple mutants retained similar catalytic activity to that of afCR-M0 in addition to showing improved thermostability. During the combination process, we also discarded some mutation sites that led to decreased activity or failed to significantly improve thermostability, such as D17V/G58D, D17V/Q165I, D17V/T117P, and D17V/E349V.

Notably, we found that some single-site mutants together produced a synergistic effect on thermostability. For example, compared with the single-site mutant D17V, the addition of T199S (D17V/T199S) resulted in another 5-fold increase in thermostability, while T199S alone led to a 1.2-fold increase compared to afCR-M0 (Table 1). Further combination of D17V/T199S with L6P produced another 6-fold increase in thermostability, whereas alone it only provided a 1.64-fold increase compared to afCR-M0 (Table 1). Thus, these results demonstrated that these amino acid residues function together resulting in a synergistic improvement to thermostability. However, we also observed that some combinations of single-site mutants resulted in antagonistic effects on the thermostability. For example, compared with the L6P (half-life of 19 min at 55 °C) and G58D (half-life of 17 min at 55 °C) single mutation variants, the L6P/G58D double mutant had reduced thermostability (half-life of 6 min at 55 °C),
which was even lower than that of the afCR-M0 template. Further exploration of the contributions by each of these residues to overall thermostability will help to clarify the molecular mechanisms underlying the synergistic effects of these mutations, which will provide a useful reference for CR engineering through combinatorial mutagenesis.

2.6 Kinetic and thermodynamic stability of the improved afCR combinatorial variants

To assess the kinetic stability of the variants, activity was measured across a range of temperatures (35-70 °C) to determine the temperatures at which enzyme activity was reduced by 50% after 15 min of incubation ($T_{50}^{15}$) (Fig. 4A). We found that the residual activity of M0 was only 7% after incubation at 60 °C for 15 min (Fig.4A), whereas, M1 retained 40% activity, M2-4 and M3-4 retained 80% activity, M4-1 retained 68% activity, and M4-2 retained 50% activity. These results thus indicated that the $T_{50}^{15}$ of these variants was 3.6 to 6.7 °C higher than that of the afCR-M0 (Table 3). Notably, although M4-1 exhibited the highest half-life ($T_{1/2}$), it had a lower $T_{50}^{15}$ value than that of M3-4. Since $T_{1/2}$ and $T_{50}^{15}$ are both primary indicators of protein kinetic stability, this result also demonstrated that the improvements in stability among the combinatorial variants were not due to the simple additive effects of single mutations. However, to determine the contribution of each mutation to the overall effect on thermostability requires detailed exploration of the relationship between protein structure and external reaction conditions in determining protein kinetic stability [42].

Table 3. Kinetic and thermodynamic properties of afCR mutants.

| Enzyme | Mutation       | $T_{50}^{15}$ (°C) | $T_m$ (°C) |
|--------|----------------|-------------------|------------|
| M0     | I304L/F395V    | 55.5 ± 0.1        | 59.6 ± 0.40 |
| M1     | M0 + D17V      | 59.06 ± 0.3       | 62.12 ± 0.02 |
| M2-4   | M1 + T199S     | 61.35 ± 0.2       | 63.93 ± 0.30 |
| M3-4   | M2-4 + L6P     | 62.21 ± 0.12      | 64.94 ± 0.20 |
| M4-1   | M3-4 + F108Y   | 61.0 ± 0.15       | 64.35 ± 0.31 |
| M4-2   | M3-4 + T251C   | 59.7 ± 0.1        | 64.24 ± 0.23 |

To further examine how the combined mutations affected thermodynamic stability, the melting temperatures ($T_m$) was determined by nanoDSF. The underlying principle of this assay is that the 350/330 nm emission ratio for tryptophan fluorescence of a given protein can indicate the temperature at which the protein unfolds. The $T_m$ values for M1, M2-4, M3-4, M4-1, and M4-2 ranged from 2 to 5.3 °C higher than that of afCR-M0 (Fig.4B). Interestingly, unlike the changes in kinetic stability, the melting temperatures of the combined mutation variants increased only slightly over that of the template protein. In previous studies, some researchers has proposed that the unfolding free-energy barrier and the unfolding rate in protein are the key factors of the thermodynamic stability [43, 44]. Therefore, although the denaturation and unfolding process of the enzyme may be related, the results clearly reflected the
different process for the enzyme in catalyzed reaction [45], and when measuring enzyme stability, kinetic stability, it should be differentiated from the thermodynamic stability.

2.7 Molecular mechanisms underlying higher thermostability conferred by individual mutations

Given our results of the higher thermostable mutants, we next investigated the mechanistic interactions associated with individual mutations that improved thermostability among the positive mutants using afCR structural homology modeling. To this end, we first checked if mutations introduced new interactions (Fig.5 and Table S3) and found that the F108Y substitution led to the formation of a new hydrogen bond with the Y68 side chain hydroxyl group (Fig.5A). Similarly, the Y109F mutation was observed to form new π-π interactions with residues F133 and F108, which subsequently increased the stability of the two helical structures in which the residues are located (Fig.5B). By contrast, the K351E mutation resulted in the loss of a hydrogen bond between that residue and E349. However, this mutation resulted in new salt bridge interactions and an H-bond formation between P352 and L350 (Fig.5C). In addition, the T251C substitution introduced a new sulfhydryl group and may create a new disulfide bond with the neighboring C175 (Fig.S5A).

We then examined changes in hydrophobic and hydrophilic substitutions among the mutants and found that D17V mainly improved hydrophobic packing in the protein interior and facilitated the interaction network between the neighboring residues His12, Asn13, Lys16, Trp90, and Arg91 (Fig.S5B). Furthermore, we found that Q165I and E349V conversions, mutations similar to D17V, also stabilized the enzyme through increased hydrophobicity, using similar mechanisms as that of D17V. The advantages provided by these mutations apparently contradict conventional mutation theory, in that hydrophobic to hydrophilic amino acid conversion is more likely to improve protein thermostability. For example, the G58D and T199S mutations mainly contributed to higher stability by increasing the hydrophilicity of the protein surface, as well as the inter-helix hydrophobic interactions. In addition, introduction of proline substitutions has been previously shown to be a trend in the stabilization of proteins [46]. The leucine to proline substitution in the L6P mutation (Fig.S5C) decreased the conformational entropy of local unfolded protein, which resulted in the protein space structure more rigid.

In summary, this analysis showed that individual mutations incurred a wide range of alterations to afCR-M0 structure and internal and external interactions, including changes in hydrogen bonds, salt bridge interactions, π-π interactions and disulfide bonds (F108Y, Y109F, K351E, T251C), increased hydrophilic interactions at the protein surface (T199S, G58D), reduction in the conformational entropy of local unfolded proteins (L6P), and improved hydrophobic packing in the protein interior or increased the interaction network (D17V, Q165I, E349V). Moreover, the introduction of the above interactions can improve the thermostability of proteins has been verified in other studies [47-49]. Through analysis of these mutations, we have modified our current understanding of the interaction mechanisms by which different types of amino acid conversions may improve thermostability and thereby provide guidance for engineering higher thermostability in other proteins.
3. Conclusions

In this study, we reported a new enzyme with excellent properties and characterized its enzymatic function. Moreover, we demonstrated the enhancement of afCR thermostability, guided by an improved non-biased phylogenetic consensus method, through combinatorial mutagenesis of several targeted residues in afCR led to significant improvements in thermostability. Specifically, afCR mutant M4-2 exhibited a 4.2 °C increase in the T$_{50}$ and ~1700-fold enhanced half-life at 57 °C than that of afCR-M0, while retaining 80% or higher activity. Consequently, these results showed that this design strategy for improvement of thermostability was effective, to providing a highly thermostable enzyme resource for creatinine detection, in addition can provide guidance for similar improvements in other commercially valuable enzymes.

4. Materials And Methods

4.1 Experimental operations

4.1.1 Media and Reagents

The afCR gene (BAA88830.1) was synthesized at GenScript Crop (Nanjing, China) and cloned into a pANY1 expression vector (gift from College of Biosciences and Biotechnology, Shenyang Agricultural University) [50]. Restriction enzymes and T4 ligase were purchased from New England Biolabs (Ipswich, MA). QIAquick™ PCR purification kits were purchased from Qiagen (Hilden, Germany). E.coli BL21 (DE3) competent cells were purchased from WEIDI (Shanghai, China). E. coli was routinely cultured overnight at 37 °C in 2×YT broth containing Bacto tryptone (1.6%, w/v), Bacto yeast extract (1%, w/v), and sodium chloride (0.5%, w/v) or on 2×YT agar plates with (in both cases) 50 μg/ml kanamycin.

4.1.2 Protein expression and purification

Cells expressing recombinant afCR were cultivated in 2×YT medium with kanamycin (50 μg/ml) at 37 °C and rotary shaking at 220 rpm. For the expression of afCR, the E. coli cells were induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration of 0.5 mM when the OD$_{600}$ value reached 0.6-0.8, then they were further cultured at 20 °C for 16 h. The collected cells were washed and resuspended in binding buffer (25 mM Tris-HCL, pH 8.0 containing 200 mM NaCl and 20 mM imidazole). Resuspended cells were lysed by an High Pressure Cell Disruptor (Union-Biotech, ShangHai, China) followed by centrifugation at 12000 g for 20 min to remove cell debris. The supernatant was loaded onto a pre-equilibrated Ni-NTA column (GE, USA), and proteins were eluted with a gradient of imidazole (from 20 to 200 mM). The purity of the collected fractions was analyzed by SDS-PAGE. Fractions containing the pure target protein were combined and desalted by ultrafiltration. The purified proteins were concentrated and stored in phosphate buffer saline (10 mM, pH 7.5) at -80 °C.

4.1.3 Activity measurement
The activity of afCR was measured by continuous coupled enzymatic assays, based on the action of sarcosine oxidase and horseradish peroxidase. The enzyme was appropriately diluted to 1 mg/ml with phosphate buffer (10 mM, pH 7.5). Enzyme activity was determined by adding 50 μl enzyme to 950 μl substrate solution. The substrate solution was composed of 500 μM creatine, 0.45 mM 4-AA (4-Aminoantipyrine), 0.5 mM TOOS (N-Ethyl-N-(2-hydroxy-3-sulfopropyl)-3-methylaniline) and phosphate buffered saline (10 mM, pH 7.5), which was incubated at 37 °C. The change in ultraviolet absorption at 555 nm in the enzyme reaction system was monitored using a UV2550 spectrophotometer (Shimadzu). One unit of activity was defined as the amount of enzyme producing 1 μM hydrogen peroxide per minute.

4.1.4 Thermostability assay

For determination of enzyme thermostability, the concentration of the purified enzyme was diluted to 1.0 mg/ml in phosphate buffered saline (10 mM, pH 7.5). \( T_{1/2} \) values were determined by incubating the purified enzyme at 55 ºC or 57 ºC for different time periods. Their residual enzymatic activities were assayed at 37 ºC as described above. The half-life was calculated by fitting the linear part of the curves: \( T_{1/2} = -\ln (2)/k \), where \( k \) is the slope of the straight line, and plotting the natural logarithm of the residual relative activity of the enzyme versus heat treatment time.

\( T_{50} \) values were determined by incubating the purified enzyme over a range of temperatures from 35 ºC to 70 ºC for 15 min. After incubation, the sample solution was then cooled immediately in an ice bath and the remaining afCR activity was measured as described above. The \( T_{50} \) value is the temperature at which enzymatic activity is reduced to 50% after 15 min of heat treatment. The activity measured at 35 ºC was considered to be 100%. The data was analyzed by calculating the inflection point of a fit of the residual activity at certain temperatures to a sigmoidal plot (sigmoidal Boltzmann fit using Origin 9.0).

4.1.5 Differential scanning fluorimetry

DSF experiments were performed on a nanoDSF device (Prometheus NT.48, NanoTemper Technologies GmbH). All samples were diluted to with phosphate buffered saline (10 mM, pH 7.5) to a final concentration of 1 mg/ml and loaded into high sensitivity capillaries. The protein unfolding process was subjected to a thermal ramp (20-95 ºC, 1 ºC/min). Data analysis was performed using the Prometheus PR ThermControl software. The Tm-value was determined by fitting the tryptophan 350/330 nm fluorescence emission ratio using a polynomial function in which the maximum slope is indicated by the peak of its first derivative.

4.1.6 Measurement of optimal pH and temperature for activity and stability

The optimal pH was determined by measuring afCR activity at 37 ºC and at pH 4.5 to 10.0 for 60 min. The optimal temperatures of the WT afCR were determined at pH 7.0 for 60 min in the temperature range of 25-55 ºC.

4.1.7 Effects of various metal ions on enzyme activity
The effects of metal ions on enzyme activity were investigated using EDTA, CoCl$_2$, FeSO$_4$, FeCl$_3$, MnSO$_2$, ZnCl$_2$, CaCl$_2$, MgCl$_2$, CuSO$_4$, and NaN$_3$. WT afCR was preincubated in phosphate buffer (10 mM, pH 7.5) containing: Co$^{2+}$, Fe$^{2+}$, Fe$^{3+}$, Mn$^{2+}$, Zn$^{2+}$, Ca$^{2+}$, Mg$^{2+}$, Cu$^{2+}$, and Na$^+$ at a final concentration of 1mM for 3 min at 37 °C. After incubation, a residual enzymatic activity was measured as described above.

4.1.8 Site-directed mutagenesis

In this work, a two-site variant of afCR-M0 (I304L/F395V), with an improved $K_M$ value compared to wild type afCR, was used as the template protein as a basis for subsequent mutations. Site-directed mutagenesis was performed according to the standard QuikChange™ Site-Directed Mutagenesis kit. All mutations were constructed using whole plasmid PCR-based site-directed mutagenesis and the primers containing the mutation site are shown in Table S2. We utilized the pANY1 plasmid harboring afCR gene as a template and the correct mutation sites were confirmed by DNA sequencing. For the construction of combinatorial mutants, the most thermostable mutants in each round of construction were used as templates for the next round of site-directed mutagenesis.

4.2 Consensus approach

4.2.1 Multiple sequence alignment and Phylogenetic analysis

Using wild-type afCR sequence as a query for blastP searches of the NCBI database, we acquired 45 homologous sequences (identity >50 %). These sequences were aligned using the ClustalX 2.1 software package. Duplicate sequences, excessively long and excessively short were excluded from further analysis [51]. Finally, sequences of twenty-four CR homologs were selected for phylogenetic reconstruction and the sequences was displayed using ESPript3.0 (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi).

Unrooted phylogenies were generated with the phylogenetic module in MEGA7.0 [52] using the Neighbor-Joining method [53]. To construct rooted trees for calculation of sequence weight, we added the outgroup sequence KJS40474.1 to the basal nodes of the unrooted tree (Fig.3). Here, bias in the natural sequences was balanced by applying a weight to the sequence branch in the tree. The weight of each sequence was calculated according to the branch length in the phylogenetic tree (branch lengths shown in Fig.S4). The sequence weight was equal to the weighted sum of the average distance of root node branches in the sequence module and the calculation followed Equation ($W_a$):

$$W_a = \frac{X_1}{3} + \frac{X_2}{2} + X_4$$
which the $W_a$ was the weight of the target sequence $a$, $X_i$ represents the total branch length of the tree, and the outgroup sequence $d$ was used to add root nodes to the unrooted tree to construct the weighted phylogenetic tree.

4.2.2 Consensus sequence

Sequence alignment was performed using ClustalX 2.1 and the Phylogenetic tree was constructed using the same method as above with MEGA 7.0 Software Package [51]. The consensus sequence was calculated by the python script (10.5281/zenodo.3949790).

4.2.3 Structural homology modeling for improved variants

The structural homology models of afCR were constructed using the SWISS-MODEL [54] package based on the crystal structure of creatinase from Actinobacillus (PDB:1KP0, 75% identical to afCR). Amino acid residues were mutated using the Mutagenesis module of PyMOL software for subsequent research. All drawings were performed using PyMOL software.

**Declarations**

**Acknowledgments**

We thank College of Biosciences and Biotechnology, Shenyang Agricultural University for providing the pANY1 plasmids.

**Funding**

This study was financially supported by the National Natural Science Foundation of China (Grant number 21627812, 31670791, 31900911 and 21627812), Jiangsu Basic Research Plan of Natural Science (Grant Number SBK2019041665) and China Postdoctoral Science Foundation (Grant Number 2019M660129).

**Author information**

**Affiliations**

1. Institute of Biothermal Science and Technology, University of Shanghai for Science and Technology, Shanghai, 200093, People's Republic of China.

Xue Bai (baixue62986@163.com), Daixi Li (dxli75@126.com)

2. CAS Key Lab of Bio-Medical Diagnostics, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, Jiangsu, People's Republic of China.

Fuqiang Ma (mafuqiang318@sibet.ac.cn)
Author contributions

X.B performed most assays. F.Q.M provided assistance in mutant designing. M.J.L and X.D performed the enzyme activity tests. G.Y.Y, D.X.L and Y.F designed the experiments and prepared the manuscript. All authors read and approved the final manuscript.

+Xue Bai, Daixi Li and Fuqiang Ma joint first authors.

Corresponding author

*Correspondence to Daixi Li (dxli75@126.com) or Guangyu Yang (yanggy@sjtu.edu.cn).

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Ethics approval and consent to participate

Not applicable.

Publisher's Note:

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1]. Berberich J, Yang L, Bahar I, Russell A. A stable three enzyme creatinine biosensor. 2. Analysis of the impact of silver ions on creatine amidinohydrolase. Acta Biomaterialia. 2005;1(2):183-91.
[2]. Zhi Q, Kong PY, Zang JT, Cui YH, et al. Biochemical and molecular characterization of a novel high activity creatine amidinohydrolase from Arthrobacter nicotianae strain 02181. Process Biochemistry. 2009;44(4):460-5.

[3]. Weber JA, van Zanten AP. Interferences in current methods for measurements of creatinine. Clinical Chemistry. 1991;37(5):695-700.

[4]. Appleyard G, Woods DD. The pathway of creatine catabolism by Pseudomonas ovalis. J Gen Microbiol. 1956;14(2):351-65.

[5]. Judith S, Gerald B, Gunter S. Stabilization of creatinase from Pseudomonas putida by random mutagenesis. Protein Sci. 1993;2(10):1612-20.

[6]. Afshari E, Amini-Bayat Z, Hosseinkhani S, Bakhtiari N. Cloning, Expression and Purification of Pseudomonas putida ATCC12633 Creatinase. Avicenna J Med Biotech. 2017;9(4):169-75.

[7]. Gihaz S, Bash Y, Rush I, Shahar A, Pazy Y, Fishman A. Bridges to Stability: Engineering Disulfide Bonds Towards Enhanced Lipase Biodiesel Synthesis. ChemCatChem. 2019;12(1):181-92.

[8]. Gribenko A, Patel M, Liu J, McCallum SA, Wang C, Makhatadze GI. Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. Proc Natl Acad Sci U S A. 2009;106(8):2601-6.

[9]. Silva IR, Jers C, Otten H, Nyffenegger C, Larsen DM, Derkx PM, et al. Design of thermostable rhamnogalacturonan lyase mutants from Bacillus licheniformis by combination of targeted single point mutations. Appl Microbiol Biotechnol. 2014;98(10):4521-31.

[10]. Edwardraja S, Sriram S, Govindan R, Budisa N, Lee S-G. Enhancing the thermal stability of a single-chain Fv fragment by in vivo global fluorination of the proline residues. Molecular BioSystems. 2011;7(1):258-65.

[11]. Benedix A, Becker CM, de Groot BL, Caflisch A, Böckmann RA. Predicting free energy changes using structural ensembles. Nat Methods. 2009;6(1):3-4.

[12]. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 2005;33(Web Server issue):W306-10.

[13]. Heselpoth RD, Yin Y, Moul J, Nelson DC. Increasing the stability of the bacteriophage endolysin PlyC using rationale-based FoldX computational modeling. Protein Eng Des Sel. 2015;28(4):85-92.

[14]. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins. 2006;62(4):1125-32.
[15]. Liu B, Zhang J, Fang Z, Gu L. Enhanced thermostability of keratinase by computational design and empirical mutation. Journal of Industrial Microbiology & Biotechnology. 2013;40(7):697-704.

[16]. Rohl C, Strauss C, Misura K, Baker D. Protein structure prediction using ROSETTA. Methods in enzymology. 2004;383:66-93.

[17]. Magliery TJ, Regan L. Beyond consensus: statistical free energies reveal hidden interactions in the design of a TPR motif. J Mol Biol. 2004;343(3):731-45.

[18]. Forrer P, Binz HK, Stumpp MT, Pluckthun A. Consensus design of repeat proteins. Chembiochem. 2004;5(2):183-9.

[19]. Sullivan BJ, Durani V, Magliery TJ. Triosephosphate isomerase by consensus design: dramatic differences in physical properties and activity of related variants. J Mol Biol. 2011;413(1):195-208.

[20]. Steipe B, Schiller B, Pluckthun A, Steinbacher S. Sequence statistics reliably predict stabilizing mutations in a protein domain. J Mol Biol. 1994;240(3):188-92.

[21]. Lehmann M, Pasamontes L, Lassen SF, Wyss M. The consensus concept for thermostability engineering of proteins. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology. 2000;1543(2):408-15.

[22]. Lehmann M, Loch C, Middendorf A, Studer D, Lassen SF, Pasamontes L, et al. The consensus concept for thermostability engineering of proteins: further proof of concept. Protein engineering. 2002;15(5):403-11.

[23]. Porebski BT, Nickson AA, Hoke DE, Hunter MR, Zhu L, McGowan S, et al. Structural and dynamic properties that govern the stability of an engineered fibronectin type III domain. Protein Eng Des Sel. 2015;28(3):67-78.

[24]. Aerts D, Verhaeghe T, Joosten H-J, Vriend G, Soetaert W, Desmet T. Consensus engineering of sucrose phosphorylase: The outcome reflects the sequence input. Biotechnology and Bioengineering. 2013;110(10):2563-72.

[25]. Amin N, Liu AD, Ramer S, Aehle W, Meije D, Metin M, et al. Construction of stabilized proteins by combinatorial consensus mutagenesis. Protein Engineering, Design & Selection. 2004;17(11):787-93.

[26]. Jackel C, Bloom JD, Kast P, Arnold FH, Hilvert D. Consensus protein design without phylogenetic bias. J Mol Biol. 2010;399(4):541-6.

[27]. Bloom JD, Glassman MJ. Inferring stabilizing mutations from protein phylogenies: application to influenza hemagglutinin. PLoS Comput Biol. 2009;5(4):e1000349.
[28]. Koyama Y, Kitao S, Yamamoto-Otake H, Suzuki M, Nakano E. Cloning and Expression of the Creatinase Gene from Flavobacterium sp. U-188 in Escherichia coli. Agricultural and Biological Chemistry. 2014;54(6):1453-7.

[29]. Hoeffken HW, Knof SH, Bartlett PA, Huber R, Moellering H, Schumacher G. Crystal structure determination, refinement and molecular model of creatine amidinohydrolase from Pseudomonas putida. Journal of Molecular Biology. 1988;204(2):417-33.

[30]. Hong MC, Chang JC, Wu ML, et al. Expression and export of Pseudomonas putida NTU-8 creatinase by Escherichia coli using the chitinase signal sequence of Aeromonas hydrophila. Biochem Genet. 1998;36(11-12):407-15.

[31]. Nishiya Y, Toda A, Imanaka T. Gene cluster for creatinine degradation in Arthrobacter sp. TE1826. Mole Gen Genet MGG. 1998;257(5):581-6.

[32]. Suzuki K, Sagai H, Sugiyama M, et al. Molecular cloning and high expression of the Bacillus creatinase gene in Escherichia coli. J Ferment Bioeng. 1993;76(2):77-81.

[33]. Kensa R, Oka I, Ando M, Yoshimoto T, et al. Creatinine Amidohydrolase (Creatininase) from Pseudomonas putida. J Biochem. 1979;86(4):1109-17.

[34]. Mukherjee G, Banerjee R. Effects of temperature, pH and additives on the activity of tannase produced by a co-culture of Rhizopus oryzae and Aspergillus foetidus. World Journal of Microbiology and Biotechnology. 2005;22(3):207-12.

[35]. Morley KL, Kazlauskas RJ. Improving enzyme properties: when are closer mutations better? Trends Biotechnol. 2005;23(5):231-7.

[36]. Paramesvaran J, Hibbert EG, Russell AJ, et al. Distributions of enzyme residues yielding mutants with improved substrate specificities from two different directed evolution strategies. Protein Eng Des Sel. 2009;22(7):401-11.

[37]. Polizzi KM, Chaparro-Riggers JF, Vazquez-Figueroa E, Bommarius AS. Structure-guided consensus approach to create a more thermostable penicillin G acylase. Biotechnol J. 2006;1(5):531-6.

[38]. Shao ZX, Rainer S, Peter K, Janet K. Variants of Erwiania-type Creatinase. US 6958231 B2. P. 2005-10-25.

[39]. Nikolova PV, Henckel J, Lane DP, Fersht AR. Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. Proc Natl Acad Sci U S A. 1998;95(25):14675-80.

[40]. Rath A, Davidson AR. The design of a hyperstable mutant of the Abp1p SH3 domain by sequence alignment analysis. Protein Sci. 2000;9(12):2457-69.
[41]. Hua Y, Lyu C, Liu C, Wang H, Hu S, et al. Improving the Thermostability of Glutamate Decarboxylase from Lactobacillus brevis by Consensus Mutagenesis. Appl Biochem Biotechnol. 2020;191(4):1456-59.

[42]. Sullivan BJ, Nguyen T, Durani V, Mathur D, et al. Stabilizing proteins from sequence statistics: the interplay of conservation and correlation in triosephosphate isomerase stability. J Mol Biol. 2012;420(4-5):384-99.

[43]. Plaza del Pino IM, Ibarra-Molero B, Sanchez-Ruiz JM. Lower kinetic limit to protein thermal stability: A proposal regarding protein stability in vivo and its relation with misfolding diseases. Proteins: Structure, Function, and Bioinformatics. 2000;40(1):58-70.

[44]. Pey AL, Rodriguez-Larrea D, Bomke S, Dammers S, et al. Engineering proteins with tunable thermodynamic and kinetic stabilities. Proteins. 2008;71(1):165-74.

[45]. Bommarius AS, Paye MF. Stabilizing biocatalysts. Chem Soc Rev. 2013;42(15):6534-65.

[46]. Arnorsdottir J, Sigtryggsdottir AR, Kristjansson MM. Effect of Proline Substitutions on Stability and Kinetic Properties of a Cold Adapted Subtilase. The Journal of Biochemistry. 2008;145(3):325-9.

[47]. Asgeirsson B, Adalbjornsson BV, Gylfason GA. Engineered disulde bonds increase active-site local stability and reduce catalytic activity of a cold-adapted alkaline phosphatase. Biochim Biophys Acta. 2007;1774(6):679-87.

[48]. Vogt G, Woell S, Argos P. Protein thermal stability, hydrogen bonds, and ion pairs. Journal of molecular biology. 1997;269(4):631-43.

[49]. Kumar S, Tsai CJ, Nussinov R. Factors enhancing protein thermostability. Protein Eng. 2000;13(3):179-91.

[50]. Liu X, Li T, Hart DJ, Gao S, et al. A universal mini-vector and an annealing of PCR products (APP)-based cloning strategy for convenient molecular biological manipulations. Biochem Biophys Res Commun. 2018;497(4):978-82.

[51]. Golubchik T, Wise MJ, Easteal S, Jermiin LS. Mind the gaps: evidence of bias in estimates of multiple sequence alignments. Mol Biol Evol. 2007;24(11):2433-42.

[52]. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution. 2016;33(7):1870-4.

[53]. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406-25.

[54]. Waterhouse A, Bertoni M, Bienert S. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-W303.