Modulation of the catalytic activity and thermostability of a highly thermostable GH7 endoglucanase by engineering the key loop B3

Fei Zheng  
Beijing Forestry University

Junzhao Yang  
Beijing Forestry University

Qingyang Sun  
Beijing Forestry University

Xinrui Zhang  
Beijing Forestry University

Ruilin Li  
Beijing Forestry University

Xiaoyu Wang  
Chinese Academy of Medical Sciences and Peking Union Medical College

Xiangwei He (hexiangwei@bjfu.edu.cn)  
Beijing Forestry University

Guozhu Zhao  
Beijing Forestry University

Research Article

Keywords: Endoglucanase, Glycoside hydrolase (GH) family 7, Loop, Molecular dynamics (MD) simulation, Thermostability

Posted Date: March 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2667966/v1

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

Background

The cellulases of glycoside hydrolase family 7 (GH7) are confined into two main types, endoglucanase and cellobiohydrolase, based on their subtle differences in loop structures. In the viewpoint of evolution, the loop regions of GH7 cellulases exhibit a more pronounced effect on enzyme properties.

Results

A thermophilic endoglucanase of GH7, \(Tt\)Cel7, having a long 18 amino acid loop B3 was identified in Thermothelomyces thermophilus ATCC 42464. It was successfully obtained with heterologous expression and then purified for activity assays. The recombinant \(Tt\)Cel7 was distinguished for the excellent thermostability at 90°C (>30% residual activity after 1-h incubation). When truncated the loop B3 or mutated C220A to remove the disulfide bond on loop B3, both the \(Tt\)Cel7 variants showed decreased catalytic efficiency, but the \(\Delta B3\) showed improved thermostability, retaining higher residual activities (9–44%) at 70–90°C than the wild type. Based on the molecular dynamics (MD) simulation analysis, both the loops B1 and A3 of \(\Delta B3\) swing toward the catalytic center, which contributes to the reduced cleft space and more rigid structure; instead, the structural rigidity of C220A was decreased as an \(\alpha\)-helix was introduced into the loop B3 due to the deletion of disulfide bond.

Conclusions

Two structural elements related to catalysis and thermostability of GH7 cellulases were identified in this study through structure-directed enzyme modulation. Of them, the loop B3 of \(Tt\)Cel7 possibly stretches the catalytic pocket, making the catalytic tunnel more open and the protein structure more flexible for efficient catalysis. Additionally, the disulfide bond in loop B3 stabilizes the loop structure and keeps it in a highly active and stable state. This strategy casts an insight into the engineering of GH7 endoglucanases for potential commercialization.

Background

Sustainable production of commodity chemicals from cellulotic biomass has been attracting significant research attention, due to the population explosion, resource exploitation, and environmental pollution. During the bioconversion of cellulotic components into fermentable sugars [1–3], cellulases are the crucial enzymes that catalyze the cleavage of \(\beta\)-1,4-glucoside bonds of cellulose. According to the catalytic mechanism, cellulases can be classified into three major types: (i) endoglucanase, which cleaves glycosidic bonds at disordered sites on cellulose substrates in a non-processive fashion; (ii) cellobiohydrolase, which acts on the reducing and non-reducing ends of cellulose chains (the reducing end is able to form an open-chain aldehyde, whereas the non-reducing end is not able to isomerize into an
open-ring form), and (iii) β-glucosidase, which cleaves soluble cellobiose into glucose. These cellulases can work synergistically to effectively degrade cellulose [4–6]. Of the carbohydrate active enzymes, glycoside hydrolases of family 7 (GH7s) consisting of cellobiohydrolases and endoglucanases are the powerhouse of cellulose degradation in nature [7, 8].

The catalytic domains of GH7 endoglucanases and cellobiohydrolases differ greatly in the loop length. Cellobiohydrolases have eight flexible loops that form a tunnel-shaped substrate-binding site along with up to eleven carbohydrate-binding subsites [9], while endoglucanases have fewer subsites and more exposed catalytic cleft due to the shorter loops in this region [10]. Studies have shown that the length of loops covering the catalytic cleft play an important role in tuning processivity, binding, and endolytic activity of enzymes [11]. For example, Yang et al. clarified the function of loop 3 in GH12s by introducing and deleting a loop, and revealed that the longer loop 3 may strengthen the hydrogen network interactions between the substrate and protein and consequently increase the turnover rate ($k_{\text{cat}}$) [12]. Moreover, specific residue sites within the cellulase loop structure have been identified to play critical roles in cellulase-substrate interactions and enzymatic catalysis. For example, using multiple sequence alignment and saturation mutagenesis, Zheng et al. revealed the important role of amino acid residue at position 233 derived from loop 6 in semi-conservation of GH5 cellulases [13]. The loop structure is also crucial in the catalytic efficiency of GH7 cellobiohydrolases. Divne et al. co-crystallized $Tr$Cel7A in complex of cellotetraose, cellopentaose, and cellohexaose, and confirmed that loop B3 is the most flexible region that interacts with glucosyl residues in cellooligosaccharide with the backbone displacement of up to 2 Å [14]. von Ossowski et al. indicated loop B3 as a key contributor to the high synthesis capacity of $Tr$Cel7A as the eight residues (245GTYSNDNY252 of loop B3)-truncated variant showed significantly decreased activity against crystalline cellulose [15]. However, few studies have elucidated the function of the loop region in GH7 endoglucanases.

By far five crystal structures of GH7 endoglucanases have been resolved, including $Fo$Cel7B (Fusarium oxysporum; 1OVW), $Tr$Cel7B (Trichoderma reesei; 1EG1), $Hi$Cel7B (Humicola insolens; 6YOZ), $Th$Cel7B (Trichoderma harzianum CBS 226.95; PDB ID: 5W0A), and $Re$Cel7B (Rasamsonia emersonii CBS 394.64; 6SU8). These structures share a common β jelly roll fold with two largely antiparallel β-sheets packing face to face to form a curved β-sandwich [16–20]. The catalytic cleft is covered by loops, in which the loop B3 (referred as the “exo” loop in the original publication) is located at the tunnel where the substrate binds, forming the roof of active site tunnel at the catalytic center [15]. Loop B3 can be categorized into two types based on the amino acid sequence length: a long chain of 18 amino acid residues as in $Fo$Cel7B, and a short chain of less than 5 residues in $Tr$Cel7B. Most GH7 endoglucanases with a long chain loop B3 have poor thermostability [21–23], suggesting the possible role of loop B3 in structural rigidity. Moreover, disulfide bonds can improve the protein rigidity and thermostability in a fluctuating cellular environment [24–26] or contribute to protein activity by stabilizing active protein conformations [27–29]. For example, the introduction of disulfide bonds into the unstable flexible region or the N-terminus enhanced the activity half-life and melting temperature of xylanases [30]. Therefore, the loop B3 and disulfide bond might represent key factors in enzyme catalysis and stability of GH7 endoglucanases.
The highly thermostable fungus *Thermothelomyces thermophilus* ATCC 42464 is an exceptionally powerful cellulolytic origin of organism that synthesizes a complete set of enzymes necessary for the breakdown of cellulose [31]. In this study, an endoglucanase, designated *Tt*Cel7, was identified, and two variants were constructed through mutating cysteine 220 to alanine to destroy the original disulfide bond (C220–C225) and truncating 14 amino acid residues \((2^{16}\text{GPYLCEGAECEF}DG^{229})\) of loop B3, which is the extra part than *Tr*Cel7B. These mutations caused structural variability of the loop B3 as well as alteration in the catalytic efficiency, substrate binding capacity, and stability of *Tt*Cel7. The results provide a new strategy applicable for enzyme molecule modification.

**Results And Discussion**

### Sequence and structural analysis

A GH7 endoglucanase-encoding gene, *Ttcel7* (gene ID: 11506134, GenBank accession no.: XP_003664606.1) was identified in the genome of *T. thermophilus* ATCC 42464. The deduced amino acid sequence consists of 456 amino acid residues, including a putative signal peptide of 20 residues and a catalytic domain of GH7. Based on the analysis of ExPASy-Prot Param tool, the theoretical molecular weight and the theoretical isoelectric point \((pI)\) of mature *Tt*Cel7 are 47.2 kDa and 4.75, respectively. Multiple sequence alignments (Fig. 1a) revealed that the loop B3 of *Tt*Cel7 and other GH7 endoglucanases differed distinctively in the length. The catalytic triad is highly conserved, including the E194 (nucleophile), D196, and E199 (acid/base, numbering without the signal peptide sequence).

Using AlphaFold 2 to predict the structure of *Tt*Cel7, it is composed of a β jelly roll with six loops forming a tunnel for substrate recognition and binding as other GH7 cellulases (Fig. 1b). Further structural superimposition of homology-modeled *Tt*Cel7 and *Tr*Cel7B demonstrates that the loop B3 of *Tt*Cel7 has 14 more amino acid residues \((2^{16}\text{GPYLCEGAECEF}DG^{229})\) than that of *Tr*Cel7B and contains a β-fold, an α-helix and a disulfide bond.

### Expression and purification of *Tt* Cel7 and its variants

The gene fragments coding for mature *Tt*Cel7 and its loop B3-truncated and C220A variants were cloned into the pPIC9 vector. Recombinant enzymes were successfully produced in the competent cells of *Komagataella phaffii* GS115 after 48 h methanol induction. The apparent molecular weights of the enzymes as determined by SDS-PAGE (Fig. S1) were larger than the theoretical masses (*Tt*Cel7: 47.2 kDa; *C220A*: 47.2 kDa, ΔB3: 45.7 kDa). The extra molecular masses might be ascribed to possible glycosylation occurred in the N-glycosylation sites (N57, N244, and N371).

### Biochemical characterization of *Tt* Cel7 and its variants

The enzymatic properties of *Tt*Cel7 and its variants were determined using CMC-Na as the substrate. Over the pH range of 3.0–9.0, *Tt*Cel7 showed the optimal activities at pH 5.0 (Fig. 2a). The pH stability was assessed at pH 3.0–12.0. *Tt*Cel7 remained highly active (> 90% of the initial enzyme activities) over pH
4.0–12.0 (Fig. 2b). The optimal temperature for \( Tt\)Cel7 activity was observed at 60°C (Fig. 2c). Thermostability at 70°C, 80°C, and 90°C were also assayed (Fig. 3a-c). It was shown that \( Tt\)Cel7 has excellent thermostability, retaining over 30% of the initial activities after 1 h incubation at 80°C and 90°C.

Both the mutants of \( Tt\)Cel7 showed the optimal activities at pH 5.0, but the pH tolerance ranges of the variant enzymes were narrowed (Fig. 2a). \( \Delta B3 \) was consistent with the wild type and both remained more than 90% of the initial enzyme activities over pH 4.0–12.0 (Fig. 2b), while C220A only maintained most of its initial activity at pH 4.0–8.0, and only retained 54% enzymatic activity at pH 12.0. The optimal temperature of C220A and \( \Delta B3 \) was observed at 50°C, which is 10°C lower than the wild type. C220A had similar thermostability to the wild type, while \( \Delta B3 \) was more prominent in thermostability, retaining 9–44% higher residual activities than the wild type and C220A under the same conditions (Fig. 3a-c).

GH7 endoglucanases are mostly mesophilic and acidic, displaying maximum enzyme activities at the temperature range of 50–60°C and pH range of pH 4.0–5.0 [32–36]. However, \( Tt\)Cel7 demonstrated higher thermostability than other GH7 endoglucanases. For example, the \( Af\)-EGL7 from \( Aspergillus fumigatus \) retained no residual activity at temperatures above 60°C after 30 min incubation [37]; the CelA and CelB from \( Aspergillus oryzae \) KBN616 were quickly inactivated after incubation for 10 min at temperatures above 60°C and 55°C, respectively [21]; and the EG from \( Humicola grisea \) var. \( thermoidea \) lost its initial activity when heated at 60°C for 10 min [22]. In comparison to the GH7 endoglucanases mentioned above that dramatically lost activities at 60°C and above, \( Tt\)Cel7 with excellent thermostability at 90°C is more valuable in commercial prospects. Interestingly, among the endoglucanases with poor thermostability, the loop B3 of CelA, CelB, and EG belongs to the long chain type. Although both the GH7 endoglucanase \( Mt\)EG7a and \( Tt\)Cel7 are derived from \( T. thermophilus \) ATCC 42464 [32], they share 45% sequence similarity and vary in the length of loop B3 and thermostability. The \( Mt\)EG7a having a loop B3 of the short chain type showed even better thermostability than \( Tt\)Cel7, retaining more than 40% residual activity after being treated at 80°C for 8 h. When truncated the long loop of \( Tt\)Cel7, the variant \( \Delta B3 \) retained 75% of its initial activity after incubation at 80°C for 1 h. Therefore, the length of loop B3 probably defines the thermostability of GH7 endoglucanases. In general, loops represent a flexible element in protein; for example, thermophilic enzymes have shorter loops than the mesophilic counterparts. Short loops are also related with enhanced protein stability due to the less of entropic loss [38]. The excellent thermostability of \( Mt\)EG7a as a short chain type endoglucanase and \( Tt\)Cel7’s truncated variant \( \Delta B3 \) confirms the role of loop B3 in structural rigidity.

The specific activities of \( Tt\)Cel7 and its variants were calculated using 10 mg/mL CMC-Na, barley β-glucan, or lichenan as the substrate (Table 1 in the Additional file). \( Tt\)Cel7 showed the highest activity towards barley β-glucan (790 ± 1 U/mg), followed by lichenan (493 ± 10 U/mg) and CMC-Na (38 ± 2 U/mg), which is consistent with previous reports [32, 33, 39]. The lower activities towards CMC-Na might be contributed to the highly substituted side chains, which interferes the substrate binding and consequently enzymatic activity [40]. The variants C220A and \( \Delta B3 \) showed similar substrate preference to the wild type but decreased specific activities (by 34–74%).
The kinetic parameters of *Tt*Cel7 and its variants were determined under optimal assay conditions with different concentrations of CMC-Na (Table 1 in the Additional file). The $K_m$, $V_{max}$, and $k_{cat}/K_m$ values of *Tt*Cel7 were $8.8 \pm 0.8$ mg/mL, $143 \pm 7$ µmol/min/mg, and $12.8 \pm 0.1$ mL/mg/s respectively. In comparison to the wild type, variants showed much lower substrate affinity and catalytic efficiency, as reflected by the higher $K_m$ values, and lower $k_{cat}/K_m$ values. The $k_{cat}/K_m$ was reduced to $6.2 \pm 0.6$ and $4.9 \pm 0.6$ mL/mg/s, respectively.

The enzymatic activities of *Tt*Cel7 and its variants under high-salt conditions were also assayed. As shown in Fig. 4, the relative specific activities of *Tt*Cel7 gradually decreased along with the increase of salt concentration, reducing to 37% in the presence of 4 M NaCl; instead, it was highly halo-stable, remaining approximately the initial activity after incubation in 5 M NaCl. In comparison to the wild type, C220A and ΔB3 showed improved and similar halotolerance respectively, but improved halo-stability (22% and 20% higher). Studies have revealed that a high negative electrostatic potential on the surface is a specific characteristic of halophilic proteins [41]. These excess acidic amino acid residues on protein surface are capable to bind more water-binding molecules and metal ions, which contribute to the maintenance of protein activity [42, 43]. In the case of *Tt*Cel7, it has 64 negatively charged residues, in which aspartic acid and glutamic acid account for 14.7% of the total residues. The large number of acidic surface residues contributes to the lower $pI$ (4.75) of *Tt*Cel7, which is far lower than that of most enzymes, and the acidic amino acid residues on the surface may form a hydration shell to protect and maintain the protein's structural integrity in the presence of salt. Surprisingly, although the *Tt*Cel7 and its variants have no significant difference in the surface electrostatic potentials, local mutations in loop B3 made the variants more salt tolerant. Thus, an insight into the structure-based mechanism should be cast to undermine the improved enzyme properties.

**Molecular dynamics (MD) simulation analysis**

Ramachandran diagrams of the AlphaFold 2 simulated structures showed that most amino acid residues of the *Tt*Cel7 and its variants were in the allowable regions with great model quality and reliability (Fig. S2). MD of the three enzymes were performed at 343 K for 100 ns. Values of the root mean square deviation (RMSD) and root mean square fluctuation (RMSF) of the backbone atom, $C_\alpha$, $N$, and carbonyl C were monitored during the simulations for estimating thermal fluctuation of the protein conformation. Based on the results of RMSD (Fig. 5a–c), *Tt*Cel7 reached equilibrium after 50 ns with the average RMSD value of 2.0 Å, which was roughly similar to C220A (1.8 Å) and ΔB3 (2.0 Å). The enzymes exhibited similar RMSF trends in the appropriate cleft (Fig. 5d–f), with higher fluctuations in the loop regions surrounding the cleft. The RMSF values of residues 214–218 of C220A increased to 4 Å in the loop B3, while the corresponding residues of *Tt*Cel7 reached about 1.5 Å (Fig. 5d–f, red dashed box). The flexibility of residues at position 257–266 adjacent to loop B3 increased to the maxima, 5 Å, due to the lack of partial residues in variant ΔB3, while the RMSF values are 2 Å and 3 Å in the corresponding positions in wild type and C220A, respectively (Fig. 5d–f, blue dashed box). It indicated that either the removal of disulfide bond C220-225 or the truncation of loop B3 caused significant structural fluctuations in this region. As presented in Fig. 6a–c, the average structure of mutations C220A and ΔB3 had little effect on
the integral structure of the enzyme, indicating that the catalytic function of the variants remained intact. Detection the loop B3 of the average structure reveals that removal of the disulfide bond C220-C225 caused the formation of an α-helix in the loop B3, and truncation of the 14 residues of loop B3 made the cavity in this zone more open, which caused conformational changes of the loop B3. Principal component analysis (PCA) demonstrated the distributions of different dynamic states of the *Tt*Cel7 and its variants (Fig. S3a-c), the proteins share five states of different structure stability. According to the dynamical cross correlation matrix (DCCM) coefficients (Fig. S3d-f), the motions of the loop B1 enhanced the interactions with its adjacent residues. And the motions of loop B3 were closely related with the residues of loop A1, suggesting the interactions between distant regions.

The altered enzymatic properties of the two variant enzymes are well explained by the MD results. Based on the energy calculation of MD simulation, it was concluded that a complex network of hydrogen bonding interactions forms around the loop B3 of wild type (Fig. 7). Of them, C225 forms hydrogen bonds with G222 and K233, respectively, and forms a disulfide bridge with the sulphydryl group of C220; and K233 also forms hydrogen bonds with E226 and C231, respectively. However, due to the lack of the pulling effect of the loop B3 and hydrogen bonds, conformational changes of the residues in ∆B3 are induced, and new hydrogen bonds are formed to maintain structural stability. For instance, in Fig. 7b, the side chain of residue K233 was rotated by 120° to form new hydrogen bonds with the catalytic residue E194 and the neighboring acidic amino acid residue D232, which reduced the Cα distance of the two catalytic residues (E194 and E199) from 12.7 Å to 11.7 Å, making the space for the catalytic cleavage smaller.

In addition to the position change of the residues in the vicinity of loop B3, the positions of the residues at the entrance and exit of the catalytic cleft also shifted significantly due to the lack of loop B3. Figure 8a and 8b demonstrates that the catalytic cavities and pockets of ∆B3 are significantly narrower than those of the wild type. As illustrated in Fig. 8c and 8d, residues W52, Q172, K233, N234, R242, G351, D352, N353, W356, and E361 all move toward the catalytic center. At the entrance of the catalytic tunnel, the orientation of the W52 side chain in the loop B1 changes and becomes closer to the catalytic center. Moreover, a hydrogen bond is formed between D352 in loop A3 and Q172 in proximity to loop B3, causing the catalytic pocket to be covered by the surrounding amino acids and the entrance of the catalytic tunnel becomes narrower. Meanwhile, at the exit of the catalytic tunnel, amino acid orientation change is also detected, which induces a number alterations of hydrogen bonds. In the wild type, R242 forms hydrogen bonds with W356 and N239 (Fig. 8d), while in the ∆B3, a novel complex dense hydrogen bonding network is formed between R242 at the exit of the catalytic tunnel. As shown in Fig. 8d, R242 forms hydrogen bonds with A237 and N239 above the cleft and E361 below the cleft, which enables the catalytic tunnel exit to be closed. Although R242 does not form a hydrogen bond with W356, the pi bond of W356 is rotated by 50°, making the product production site more confined and the catalytic cleft more hydrophobic.

In summary, the hydrogen bonding network formed between amino acids in the loop B3 makes *Tt*Cel7 have a more open catalytic cleft and able to recruit more substrates and small molecules, which may
explain its high catalytic efficiency. The truncation of the loop B3 generates variations in the position of amino acid residues and the number of hydrogen bonds makes the catalytic cleft narrower at both ends and amino acid residues closer to the catalytic cleft, which makes the structure of variant ΔB3 more compact. It is widely acknowledged that the compactness of a protein is one of the crucial factors of high stability [44]. This may explain the high thermostability of the variant ΔB3. As the catalytic cleft becomes smaller or narrower, it may interfere with the grasping and releasing of ΔB3 to the substrate, which leads to a smaller $k_{cat}$ value. It might further decrease the protein-ligand interactions in the substrate-binding sites, hence reducing the affinity of the enzyme to the substrate.

A highly conserved disulfide bond was identified in the GH7 endoglucanases of long loop B3 type, which is absent in the ones of short chain type. Disulfide bonds of proteins are crucial for the structural and functional stabilities [45]. For instance, both the optimal temperature and $T_m$ values of the disulfide bond-removed variant of cellulase TaCel45 were decreased, and the thermal refolding ability retained less than 50% of the wild type [46]; instead, the thermostability and hydrolytic efficiency of xylanase PjxA were enhanced upon the introduction of a disulfide bridge [47]. The effect of disulfide bond C220-C225 in loop B3 on endoglucanase TtCel7 was also determined by construction of variant C220A. In comparison to the wild type, the optimum temperature, substrate affinity, and specific activity of C220A were reduced. Due to the lack of disulfide bond C220-C225, the loop B3 became more flexible, and the number and area of cavities and pockets increased (Fig. 9a and 9b). Moreover, C225 together with its neighboring amino acid residues (222GAECEF227) forms an α-helix when C225 is free from the disulfide bond in variant C220A (Fig. 9c). In addition, the average structure exhibited a significant variation in the loop B1 of C220A (Fig. 9d). The amino acid residues D51, W52, and G53 in loop B1 of C220A all migrated toward the active center. Ca atom of the wild-type W52 moved 12.2 Å towards the catalytic center, which narrowed the catalytic cleft of C220A. The decreases in C220A catalytic activity and substrate affinity may be explained by a narrower catalytic cleft. The more flexible loop B3 may facilitate the recruitment of salt molecules to the active site, thus improving the catalytic performance in the presence of high concentrations of salt.

**Conclusion**

The thermostable GH7 endoglucanase, TtCel7, derived from *T. thermophilus* ATCC 42464 has distinguished enzymatic and structural characteristics. It’s acidic, mesophilic, and highly thermostable, with a flexible long chain of loop B3 containing a conserved disulfide bond C220-C225. Loop truncation and site mutation experiments revealed the crucial roles of loop B3 and disulfide bond C220-C225 in the size of the enzyme catalytic cleft opening, catalytic performance, and stability of TtCel7. Truncation of the loop B3 makes the catalytic cleft narrower, resulting in a more compact structure and improved thermostability. Although elimination of the disulfide bond C220-C225 also results in a narrower catalytic cleft and reduced catalytic efficiency, the flexibility of the loop B3 is enhanced, and the halotolerance is improved. This study expands the enzymatic library of highly thermostable endoglucanases, and provides a new perspective for the modulation of GH7 counterparts with long loop B3.
Methods

Media

Media were prepared according to the standard recipes. Liquid Luria-Bertani (LB) medium: 0.5% (w/v) yeast extract, 1% tryptone, 1% NaCl, autoclaved at 121°C for 20 min. Solid LB medium: 0.5% yeast extract, 1% tryptone, 1% NaCl, 1.5% agar, autoclaved at 121°C for 20 min. Solid minimal dextrose medium: 1.5% agarose, 2% glucose, autoclaved at 115°C for 25 min, 13.4% yeast nitrogen base without amino acids (YNB), 1‰ biotin added after sterilization. Liquid yeast extract peptone dextrose medium (YPD): 1% yeast extract, 2% tryptone, 2% glucose, autoclaved at 115°C for 25 min. Liquid buffered glycerol-complex medium (BMGY): 1% yeast extract, 2% tryptone, 1% glycerol, autoclaved at 121°C for 20 min, 13.4% YNB, 1‰ biotin added after sterilization. Liquid buffered methanol-complex (BMMY) medium: 1% yeast extract, 2% tryptone, autoclaved at 121°C for 20 min, 13.4% YNB, 1‰ biotin, 0.5% methanol added after sterilization. The final concentration of biotin is 40 µg/mL.

Strains And Plasmid

The cloned host *Escherichia coli* Trans1-T1 was purchased from TransGen (Beijing, China) and was used for cloning the target genes. Expression plasmid pPIC9 and expression host *K. phaffii* GS115 from Invitrogen (Carlsbad, CA, USA) were used for vector construction and expression of exogenous genes. The gene fragments coding for *Tt*Cel7 was synthesized by Ruibiotech (Beijing, China).

Chemicals

The kits for gel purification, gene cloning, plasmid extraction, and fast mutagenesis were purchased from TransGen. Primers were synthesized by Ruibiotech. Sodium carboxymethyl cellulose (CMC-Na) was purchased from Sigma-Aldrich (St. Louis, MO, USA). Barley β-glucan and lichenan were purchased from Megazyme (Wicklow, Ireland). Yeast extract and tryptone were purchased from Thermo (Waltham, MA, USA). Agarose was purchased from Biowest (Loire valley, France). D(+)-Glucose and agar were purchased from Biotopped (Beijing, China). And YNB was purchased from Lablead (Beijing, China).

Sequence And Structure Analysis

The theoretical molecular weight and *pI* of the *Tt*Cel7 were predicted using the ExPASy-ProtParam tool (https://web.expasy.org/protparam/). Signal peptide prediction was achieved by the SignalP 5.0 (https://services.healthtech.dtu.dk/service.php?SignalP-5.0). Homology analysis of protein sequences was performed using the BLASTp. Smart (http://smart.embl-heidelberg.de/) was used to predict the domains of proteins. Multiple sequence alignments were performed with MEGA 11 using the ClustalW and were prettified using the ESPript3 (http://escript.ibcp.fr/ESPript/ESPript/index.php). Homology-based
three-dimensional modeling of the proteins was conducted using the AlphaFold 2. Pymol was used to visualize the modeled structures and to prepare the figures.

**Primer Design, PCR, And Cloning**

The primer-based mutagenesis strategy was used to produce variants [48]. Site/truncation mutation using the methylated plasmid as a template was performed using the partial overlapping primers (including mutation sites or regions). Briefly, four template-specific primers (Table S1 in the Supplementary materials) were designed using the Vector NTI Suite 10.0 (Invitrogen, CA, USA). PCR products were separated using 1% \((w/v)\) agarose gel electrophoresis, and digested by DMT enzyme (Beijing, China) for 1 h. The gel-purified specific DNA fragments were ligated to an EcoR -Not digested pPIC9, and the recombinant products were transformed into *E. coli* Trans1-T1 competent cells. After overnight culture in LB medium containing 50 µg/mL ampicillin, correct recombination was verified by colony PCR and DNA sequencing using the universal primers 5’AOX and 3’AOX.

**Enzyme Expression And Purification**

The recombinant plasmids were linearized by the restriction enzyme *Dral* and introduced into the genome of *K. phaffii* GS115 competent cells by electroporation. Positive transformants were screened on minimal dextrose medium plates, and 24 colonies of each enzyme were selected for growth in 3 mL of BMGY for 48 h at 30°C and 200 rpm. The cells were collected and re-suspended in 1 mL of BMMY containing 0.5% \((v/v)\) methanol for expression induction at 30°C and 200 rpm for 72 h. The culture supernatants were collected by centrifugation at 12,000 rpm for 10 min at 4°C and examined by activity assay and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The transformants showing the highest cellulase activities and predicted molecular masses were grown in 30 mL of YPD at 30°C, 200 rpm for 48 h, and directly transferred into 400 mL of BMGY for 48 h growth. Cells were then harvested by centrifugation at 4,500 rpm for 10 min at 4°C and re-suspended in 200 mL of BMMY containing 0.5% \((v/v)\) methanol for induction at 30°C for 48 h. Culture supernatants were concentrated by ultrafiltration using a membrane with a 10 kDa cutoff, and dialyzed overnight in 10 mM Tris-HCl buffer (pH 8.0). The crude enzymes were further purified using the HiTrap Q HP anion exchange column (ÄKTA pure, Umeå, Sweden) and 10 mM Tris-HCl (pH 8.0) as the binding buffer. Elution was performed using a linear gradient of 0–1 M NaCl in the same buffer. The purified proteins were quantified by SDS-PAGE using different concentrations of BSA.

**Cellulase Activity Assay**

Cellulase activity was determined by using the DNS method [49]. The reaction systems containing 450 µL of substrate solution and 50 µL of appropriately diluted enzyme solution were incubated under the specific conditions for 10 min, followed by the addition of 750 µL of DNS reagent and a boiling water bath for 5 min. After cooling to room temperature, the absorbance was measured at 540 nm. Using
glucose as a standard to produce the calibration curve, one unit of cellulase activity (U) was defined as the amount of enzyme required to release 1 µmol glucose equivalent of reducing sugar per minute.

**Enzyme Characterization**

The enzymes were characterized using 1% (w/v) CMC-Na as the substrate.

**The optimal pH and pH stability**

The optimal pH was determined by measuring the enzymatic activity in 100 mmol/L McIlvaine buffer (pH 3.0–7.0) and 100 mmol/L Tris-HCl buffer (pH 8.0–9.0) and 37°C. The maximum of enzymatic activity was defined as 100% to calculate the relative enzymatic activities. pH stability was determined by measuring the residual activity under optimal conditions after incubating the enzymes in 100 mmol/L McIlvaine buffer (pH 3.0–7.0), 100 mmol/L Tris-HCl buffer (pH 8.0–9.0), and 100 mmol/mL glycine-NaOH (pH 10.0–12.0) at 37°C for 1 h. The initial enzymatic activity at time 0 was defined as 100%.

**The optimal temperature and thermostability**

The effect of temperature on enzymatic activity was determined at 40°C to 70°C and optimum pH for 10 min. The highest enzyme activity was set to 100%. To study the thermostability, each enzyme (approximately 100 µg/mL) was preincubated at 70°C, 80°C, and 90°C for 2 min, 5 min, 10 min, 30 min, and 60 min, respectively, and the residual enzymatic activities were determined under optimal conditions. The initial enzymatic activity at time 0 was defined as 100%.

**Substrate specificity and kinetics**

The substrate specificity was determined by measuring the enzymatic activity under optimal conditions against 10 mg/mL of CMC-Na, barley β-glucan, and lichenan, respectively.

The kinetic parameters, Michaelis-Menten constant ($K_m$), maximum initial velocity ($V_{max}$), and catalytic constant ($k_{cat}$), were determined under the optimal conditions of each enzyme for 5 min by using 4–20 mg/mL CMC-Na as the substrate. Using the Lineweaver-Burk plot, the kinetic values and catalytic efficiency ($k_{cat}/K_m$) of each enzyme were calculated respectively.

**Halotolerance and halo-stability**

The halotolerance $Tt$Cel7 and its variants were determined by measuring the enzymatic activities under optimal conditions in the presence of 1–4 M NaCl. The halo-stability was determined by pre-incubating the enzyme in 1–5 M NaCl for 1 h at 37°C and measuring the residual activities under optimal conditions.

**MD Simulation**

The MD simulations were conducted to investigate the different dynamic features between wild type and mutants. The ensemble system was an octahedron filled with TIP3P water molecules with periodic
boundaries. During simulation, the Amber ff19SB force field was applied for a 20 ns isothermal-isobaric (NVT) and a 100 ns isothermal-isochoric (NpT) equilibrations at a temperature of 343 K, and time step was set to 2 fs [50]. The backbone atoms, C\text{\textalpha}, N, and carbonyl C, were included for RMSD and RMSF calculation. The averaged structures of \textit{TtCel7} and its variants were derived from the last 20 ns of NpT trajectories. The relationships of dynamic state and structural stability of each enzyme were assessed using PCA. Moreover, DCCM over the 100 ns NpT simulation was calculated with correlation plus python package to reveal the mutated region caused difference from correlation of amino acid residue pairs.

**List Of Abbreviations**
<table>
<thead>
<tr>
<th>abbreviation</th>
<th>full name</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMGY</td>
<td>buffered glycerol-complex medium</td>
</tr>
<tr>
<td>BMMY</td>
<td>buffered methanol-complex</td>
</tr>
<tr>
<td>CMC-Na</td>
<td>sodium carboxymethyl cellulose</td>
</tr>
<tr>
<td>DCCM</td>
<td>dynamical cross correlation matrix</td>
</tr>
<tr>
<td>GH</td>
<td>glycoside hydrolase</td>
</tr>
<tr>
<td>HMM</td>
<td>hidden Markov model</td>
</tr>
<tr>
<td>$k_{cat}$</td>
<td>turnover rate</td>
</tr>
<tr>
<td>$K_m$</td>
<td>Michaelis-Menten constant</td>
</tr>
<tr>
<td>$k_{cat}/K_m$</td>
<td>catalytic efficiency</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>MD</td>
<td>molecular dynamics</td>
</tr>
<tr>
<td>NVT</td>
<td>isothermal-isobaric</td>
</tr>
<tr>
<td>NpT</td>
<td>isothermal-isochoric</td>
</tr>
<tr>
<td>PCA</td>
<td>principal component analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>pI</td>
<td>isoelectric point</td>
</tr>
<tr>
<td>RMS</td>
<td>root mean square</td>
</tr>
<tr>
<td>RMSD</td>
<td>root mean square deviation</td>
</tr>
<tr>
<td>RMSF</td>
<td>root mean square fluctuation</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>$V_{max}$</td>
<td>maximum initial velocity</td>
</tr>
<tr>
<td>YNB</td>
<td>yeast nitrogen base without amino acids</td>
</tr>
</tbody>
</table>

**Declarations**

**Author contributions**

JZY, QYS, and FZ organized and planned the research, and drafted the manuscript. JZY, QYS, XRZ and RLL carried out PCR detection, data analysis and determined the enzyme property. FZ, GZZ, and XWH
were involved in data analysis and paper revision. All authors have read and approved the final manuscript.

**Availability of data and materials**

All data generated or analyzed during this study are included in this published article and its Supplementary materials.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This work was supported by the National Natural Science Foundation of China (32171869 and 31770110)

**Acknowledgements**

Not applicable.

**References**


Figures
Figure 1

Sequence and structure analysis of \textit{TtCel7}. (a) Multiple sequence alignments of the \textit{TtCel7} and GH7 endoglucanases in loop B3. Alignment created with MEGA 11 and rendered with ESPript 3.0 web server with default parameters (www.espript.ibcp.fr). Black connecting wire is disulphide bond. (b) Structural alignment of \textit{TtCel7} and \textit{TrCel7B} (PDB ID: 1EG1). Homology modeling of the structure of \textit{TtCel7} using the AlphaFold. Loops and catalytic residues are highlighted in magenta and cyan for \textit{TtCel7} and \textit{TrCel7B}, respectively.
Figure 2

Enzymatic properties of the *TtCel7* and its variants using 10 mg/mL CMC-Na as the substrate. (a) pH-activity profiles were determined at the optimal temperature of each enzyme. (b) pH-stability profiles were determined by measuring the residual activity under optimal conditions after 1 h incubation at pH 3.0–12.0 and 37°C without substrate. (c) Temperature-activity profiles were determined at the optimal pH of each enzyme.
Figure 3

Thermostability profiles by measuring the residual activity under optimal conditions after incubation at (a) 70°C, (b) 80°C, and (c) 90°C for various durations.
Figure 4

Halotolerance (a) and halostability (b) of the \( Tt\text{Cel7} \) and its variants.

Figure 5

MD simulation analysis of the \( Tt\text{Cel7} \) and its variants. (a-c) RMSD values of \( Tt\text{Cel7}, \text{C220A}, \) and \( \Delta B3 \). (d-f) RMSF values of \( Tt\text{Cel7}, \text{C220A}, \) and \( \Delta B3 \).
Figure 6

Average structure of the (a) TtCel7, (b) C220A, and (c) ΔB3. The color of catalytic residue and loop B3 are magentas and cyan, respectively.

Figure 7

Hydrogen bonds probably formed within (a) TtCel7 and (b) ΔB3. Ca atomic distances and hydrogen bonds are black dashed lines. Amino acids are numbered in accordance with TtCel7.
Figure 8

Cavities and amino acids residues in catalytic tunnels of (a, c) TrCel7 and (b, d) ΔB3. Hydrogen bonds are black dashed lines.
Figure 9

The catalytic tunnel of *Tt*Cel7 and C220A. Cavities and pockets in catalytic tunnels of (a) *Tt*Cel7 and (b) C220A. (c) Loop B3 is highlighted in cyan and magenta for *Tt*Cel7 and C220A, respectively. (d) Residues in catalytic tunnels of *Tt*Cel7 and C220A. Cα atomic distances and hydrogen bonds are black dashed lines.

Supplementary Files

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

- Supplementarymaterials.docx
- Additionalfile.docx