

RESEARCH

Open Access



Enhancing the catalytic activity of a novel GH5 cellulase *GtCel5* from *Gloeophyllum trabeum* CBS 900.73 by site-directed mutagenesis on loop 6

Fei Zheng^{1,2}, Tao Tu¹, Xiaoyu Wang², Yuan Wang¹, Rui Ma¹, Xiaoyun Su¹, Xiangming Xie², Bin Yao^{1*} and Huiying Luo^{1*}

Abstract

Background: Cellulases of glycosyl hydrolase (GH) family 5 share a $(\beta/\alpha)_8$ TIM-barrel fold structure with eight $\beta\alpha$ loops surrounding the catalytic pocket. These loops exposed on the surface play a vital role in protein functions, primarily due to the interactions of some key amino acids with solvent and ligand molecules. It has been reported that motions of these loops facilitate substrate access and product release, and loops 6 and 7 located at the substrate entrance of the binding pocket promote proton transfer reaction at the catalytic site motions. However, the role of these flexible loops in catalysis of GH5 cellulase remains to be explored.

Results: In the present study, an acidic, mesophilic GH5 cellulase (with optimal activity at pH 4.0 and 70 °C), *GtCel5*, was identified in *Gloeophyllum trabeum* CBS 900.73. The specific activities of *GtCel5* toward CMC-Na, barley β -glucan, and lichenan were 1117 ± 43 , 6257 ± 26 and 5318 ± 54 U/mg, respectively. Multiple sequence alignment indicates that one amino acid residue at position 233 on the loop 6 shows semi-conservativeness and might contribute to the great catalytic performance. Saturation mutagenesis at position 233 was then conducted to reveal the vital roles of this position in enzyme properties. In comparison to the wild type, variants N233A and N233G showed decreased optimal temperature (-10 °C) but increased activities (27 and 70%) and catalytic efficiencies (k_{cat}/K_m : 45 and 52%), respectively. The similar roles of position 233 in catalytic performance were also verified in the other two GH5 homologs, *TeEgl5A* and *PoCel5*, by reverse mutation. Further molecular dynamics simulations suggested that the substitution of asparagine with alanine or glycine may introduce more hydrogen bonds, increase the flexibility of loop 6, enhance the interactions between enzyme and substrate, and thus improve the substrate affinity and catalytic efficiency.

Conclusion: This study proposed a novel cellulase with potentials for industrial application. A specific position was identified to play key roles in cellulase–substrate interactions and enzyme catalysis. It is of great importance for understanding the binding mechanism of GH5 cellulases, and provides an effective strategy to improve the catalytic performance of cellulases.

Keywords: GH5 cellulase, Saturation mutation, Catalytic activity, Loop region

*Correspondence: binyao@caas.cn; luohuiying@caas.cn

¹ Key Laboratory for Feed Biotechnology of the Ministry of Agriculture, Feed Research Institute, Chinese Academy of Agricultural Sciences, No. 12 South Zhongguancun Street, Haidian, Beijing 100081, People's Republic of China

Full list of author information is available at the end of the article

Background

Lignocellulose is composed of cellulose, hemicellulose, and lignin, and represents the most abundant renewable carbon source on earth [1]. The enzymatic hydrolysis of polysaccharides to monosaccharides is crucial from both viewpoints of cost and efficiency in the current practice of converting lignocellulosic biomass into biofuel. Complete hydrolysis of cellulose requires the cooperative actions of three types of cellulases: endoglucanase (EC 3.2.1.4) that randomly cleaves the internal β -1,4-glycosidic bonds; cellobiohydrolase (exoglucanase; EC 3.2.1.91) that processively acts on the chain termini to release cellobiose; and β -glucosidase (EC 3.2.1.21) that hydrolyzes cellobiose to glucose [2].

Based on the sequence and structure similarity of CAZymes (<http://www.cazy.org>), endoglucanases are grouped into 13 glycoside hydrolase (GH) families, including GH5-9, 12, 44, 45, 48, 51, 74, 124, and 131 [3]. Of these, GH5 is the largest and the most functionally diverse group, and those from fungi are mainly confined into subfamily GH5_5 with *endo*- β -1,4-glucanase activity [4]. So far, six eukaryotic GH5 endoglucanases from *Pirromyces rhizinflata* (PrEglA) [5], *Thermoascus aurantiacus* (TaCel5A) [6], *Hypocrea jecorina* (*Trichoderma reesei*) (TrCel5A) [7], *Ganoderma lucidum* (GlCel5A) [8], *Aspergillus niger* (AnCel5A) [9], and *Penicillium verruculosum* (PDB No. 5I6S) have been resolved. The typical catalytic domain of a GH5 cellulase has a canonical $(\beta/\alpha)_8$ TIM-barrel fold, in which the eight parallel β -strands and eight α -helices are connected by seven $\beta\alpha$ or $\alpha\beta$ loops [10].

The loops that connect secondary structures are frequently located on the protein surface and are critical for substrate specificity and catalytic activity. For example, the mutation T113R of polygalacturonase PG8fn increased the plasticity of T3 loop and caused an improvement of the catalytic efficiency by ~2.4-fold [11]; modifying the loop conformations of two GH6 cellobiohydrolases facilitated the cellulose chain gliding and allowed more occasional endo-cleavages [12, 13]; and deletion of an exo-loop of a bacterial cellobiohydrolase altered its endolytic activity [14]. A few studies also reported the effects of loops on cellulases. For the cellulase Cel12A from *Thermotoga maritima*, a Tyr-to-Gly mutation on a unique loop related to substrate binding led to an increased specific activity by 1.7-fold [15]. The protonation state of the catalytic glutamates of Cel5B from *Clostridium thermocellum*, with or without substrate, is largely governed by the conformational changes of $\beta_3\alpha_3$ loop [16]. When replacing the Phe267 with Ala of cellulase GtCel5E from *Clostridium thermocellum*, its hydrophobic interactions with two other residues were broken, the flexible loop was relocated, and the variant displayed an increased k_{cat} value by fourfold [17]. These

previous studies altogether reveal the importance of loop structures in enzyme catalysis.

Protein engineering is a prevalent method with numerous successes for enzyme improvements [18, 19]. Site-directed mutagenesis based on rational design has been widely used to identify the roles of a specific amino acid residue. In the present study, a novel cellulase of GH5 from *Gloeophyllum trabeum* CBS 900.73, designated GtCel5, was produced in *Pichia pastoris* GS115. GtCel5 with great catalytic performance had asparagine at position 233 of loop 6 ($\beta\alpha$ loop), the same as the structure-resolved homologs GlCel5A (72%, 5D8W) and TrCel5A (69%, 3QR3) of GH5_5 [3, 4]. In contrast, some other GH5 cellulases have glycine at this position. In order to gain insights into the functional role of loop 6 in GH5 cellulases, we created saturation mutants of GtCel5 at position 233 by site-directed mutagenesis. The results were then verified by reverse mutation on two GH5 homologs. Biochemical and bioinformatics analyses indicated that residue 233 on loop 6 is critical for the substrate binding and catalytic efficiency.

Methods

Strains and plasmids

The donor strain *G. trabeum* CBS 900.73 from the CBS-KNAW Fungal Biodiversity Center (Utrecht, the Netherlands) was grown at 30 °C for 3 days in a lignocellulose medium containing (w/v) 5 g/L NaCl, 5 g/L $(\text{NH}_4)_2\text{SO}_4$, 1 g/L KH_2PO_4 , 0.5 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.2 g/L CaCl_2 , 0.01 g/L $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 15 g/L corncob, 15 g/L soybean meal, and 15 g/L wheat bran. Plasmid pPIC9 harboring an ampicillin resistance gene was used for selection in *Escherichia coli* Trans I-T1 (TransGen, Beijing, China). Transformed *E. coli* was maintained on LB medium supplemented with 100 $\mu\text{g}/\text{mL}$ ampicillin and grown at 37 °C. Cellulases TeEgl5A from *Talaromyces emersonii* [20] and PoCel5 from *Prostheccium opalus* (unpublished data) were selected for reverse mutation. Plasmids pPIC9-Teegl5A and pPIC9-Pocel5 containing the cDNA fragments of mature TeEgl5A and PoCel5-encoding sequences were used as the PCR templates. Plasmid DNA was isolated from *E. coli* using a Qiagen Miniprep Kit. *P. pastoris* GS115 (Invitrogen) was maintained on MD plates (2% glucose and 2% agarose) at 30 °C.

Gene cloning

Fungal RNA was isolated and purified from 100 mg of 3-day-old mycelia grown in the lignocellulose medium using the SV Total RNA Isolation System (Promega). cDNAs were synthesized in vitro using the ReverTra Ace-a-TM kit (TOYOBO, Osaka, Japan) with total RNA as the template. Amplification of the GtCel5 gene was carried out using the oligonucleotide primer

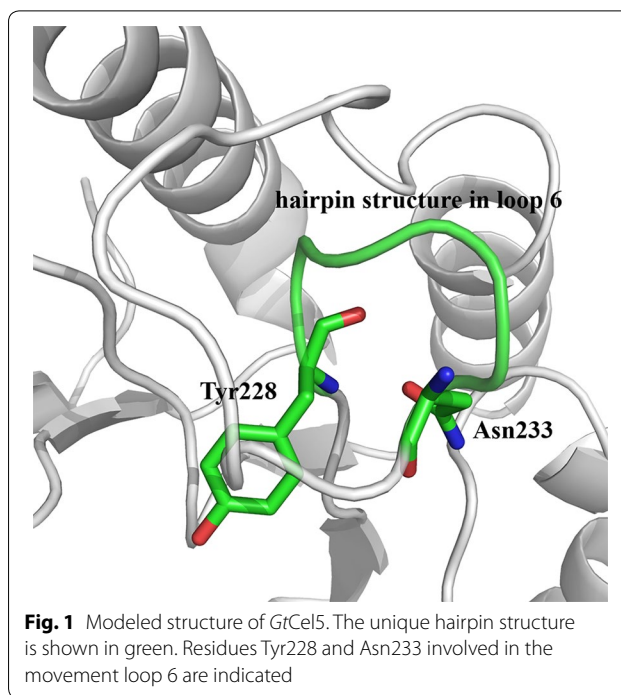
set *GtCel5*-F and *GtCel5*-R (5'-CCGAATTCGCC GCGCTCTCTCCGAGAGTGACA-3', and 5'-ACT GCGGCCGCTCATGCGTTGCAATCGGAGCCAA GCA-3', with the *EcoRI* and *NotI* restriction sites underlined). The 50- μ L PCR contained 10 μ g of cDNA template, 5 μ M of each primer, 1 mM of dNTPs, 5 μ L of 10 \times PCR buffer, and 1 μ L of Taq DNA polymerase (Fermentas; 2.5 U/ μ L). The specific PCR products were digested with *EcoRI* and *NotI* to create sticky ends and ligated to the *EcoRI*–*NotI*-digested vector pPIC9 using T4 DNA ligase (New England Laboratory). The constructed recombinant plasmid pPIC9-*Gtcel5* was then transformed into *E. coli* Trans I-T1 competent cells. Positive transformants were sequenced for verification.

Sequence analysis

The DNA and amino acid sequences were analyzed using the BLASTx and BLASTp programs (<http://www.ncbi.nlm.nih.gov/BLAST/>), respectively. The introns, exons, and transcription initiation sites were predicted using the GENSCAN Web Server (<http://genes.mit.edu/GENSCAN.html>). SignalP 3.0 was used to predict the signal peptide sequence (<http://www.cbs.dtu.dk/services/SignalP/>). The potential *N*-glycosylation sites were predicted online (<http://www.cbs.dtu.dk/services/NetNGlyc/>). Sequence assembly and estimation of the molecular mass and *pI* of the mature peptide were performed using the Vector NTI Suite 10.0 software (Invitrogen). MEGA 4.0 was used for inferring the phylogenetic relationship of GH5 cellulases [21].

Selection of the mutation site and site-directed mutagenesis

Loop 6 of *GtCel5* is in close proximity to the catalytic pocket and the component residues YLSDSN are capable of forming a unique hairpin structure (Fig. 1). Identification and multiple sequence alignment of 51 fungal cellulases of GH5 were conducted using the FASTA [22] and ClustalW [23] algorithms. Based on the structure and sequence analysis of loop 6, a key residue probably related to *GtCel5* functionality was identified, and selected for saturation mutagenesis. With recombinant plasmids pPIC9-*Gtcel5*, pPIC9-*Teegl5A*, and pPIC9-*Pocel5* as the templates, the mutants were first constructed by overlap PCR for preliminary screening. Reverse mutations of G216A and G216N of *Teegl5A* and G210A and G210N of *Pocel5* were performed using the Fast Mutagenesis System Kit (TransGen) with 30 amplification cycles. The primer pairs used in this study are listed in Additional file 1: Table S1.



Enzyme expression and purification

Recombinant plasmids containing the gene fragments coding for the wild-type and mutant enzymes of *GtCel5*, *Teegl5A*, and *Pocel5* were then linearized with *BglII* for transformation into *P. pastoris* GS115. The positive transformants were screened on MD plates. Ninety-six positive transformants of each enzyme were selected to grow in 3 mL BMGY at 30 °C for 48 h, collected, and resuspended in 1 mL BMMY containing 0.5% methanol for 72-h enzyme induction at 30 °C. The culture supernatants of each transformant were collected by centrifugation at 12,000 $\times g$ for 10 min at 4 °C and examined by activity assay. The transformants showing the highest cellulase activities were inoculated into 30 mL YPD and incubated at 30 °C, 200 rpm for 48 h, and transferred into 400 mL BMGY in 1-L Erlenmeyer flasks for 48-h growth. Cells were then harvested by centrifugation at 4500 $\times g$ for 5 min at 4 °C and resuspended in 200 mL of BMMY containing 0.5% (v/v) methanol for 48 h at 30 °C for induction.

Cell-free cultures were collected by centrifugation at 12,000 $\times g$ for 10 min at 4 °C. Further purification was performed using the HiTrap Q HP anion exchange column (Amersham Biosciences, Uppsala, Sweden). Binding buffer was composed of 10 mM sodium phosphate (pH 7.5). Elution was performed using a linear gradient of 0–1 M sodium chloride in the same buffer. The purities of the enzymes were checked with 12% SDS–polyacrylamide gel electrophoresis (PAGE) and Coomassie

blue staining. *Endo*- β -*N*-acetylglucosaminidase H (*Endo* H) from New England Biolabs was used to remove *N*-glycosylation according to the manufacturer's instructions. Purified proteins were quantified using the Bradford protein assay kit (Bio-Rad) and then used for enzyme characterization.

Cellulase activity assay

CMC-Na (medium viscosity) from Sigma-Aldrich at a concentration of 10 mg/mL was used as the substrate. The assay mixtures contained 900 μ L of substrate solution in 100 mM McIlvaine buffer (optimal pH) and 100 μ L of appropriately diluted enzyme. The reaction mixtures were incubated at optimal temperature for 10 min, followed by the addition of 1.5 mL 3,5-dinitrosalicylic acid (DNS) and incubation in a 100 °C water-bath for 5 min [24]. The amounts of reducing sugar released were measured at 540 nm, and one unit of the cellulase activity was defined as the amount of enzyme that released 1 μ mol of reducing sugar per minute.

Biochemical characterization

CMC-Na was used as the substrate for enzyme characterization. The buffers used were 100 mM KCl–HCl (pH 1.0–3.0), 100 mM citric acid– Na_2HPO_4 (pH 2.2–7.0), 100 mM Tris–HCl (pH 8.0–9.0), and 100 mM glycine–NaOH (pH 9.0–12.0). The pH–activity profile of each enzyme was determined at optimal temperature in buffers of pH 2.2–8.0. The temperature–activity profile of each enzyme was determined at optimal pH over the temperature range from 40 to 90 °C. For pH stability, each enzyme was preincubated at 37 °C for 1 h in buffers of different pH (1.0–12.0) and subjected to the residual activity assay under standard conditions as described above. For thermostability assay, each enzyme (approximately 100 μ g/mL) was preincubated at 60 or 70 °C for 0–60 min, and aliquots of 100 μ L were withdrawn at specific time points for residual activity assay.

Substrate specificity

Polysaccharides from Sigma-Aldrich and Megazymes (Wicklow, Ireland) containing different glycosidic linkages, including CMC-Na, barley β -glucan, lichenan, laminarin, konjac glucomannan, Avicel, locust bean gum, xylan, and filter paper, were used to test the substrate specificity of *GtCel5* under standard conditions. The specific activities of *GtCel5* variants toward barley β -glucan and CMC-Na were also determined and compared to that of the wild type.

Kinetic assays

Kinetic parameters of the enzymes were derived from the reactions under optimal conditions with 0.125–10 mg/

mL CMC-Na as the substrate. Initial velocities were determined by measuring the production rates of reducing sugar with the DNS method. The kinetic parameters (apparent K_m and k_{cat}) were calculated using the GraphPad Prism 6.0 (<http://www.graphpad.com/scientific-software/prism/>) and the nonlinear regression algorithm embedded in the enzyme kinetics module. The catalytic efficiency (k_{cat}/K_m) of each enzyme was then calculated.

Bioinformatic analyses

Discovery Studio 2017 software was used for automated comparative modeling of *GtCel5* and its variants with *TrCel5A* (3QR3, 69% identity) as the template. To explore the possible roles of site-directed mutagenesis at position 233, molecular dynamic (MD) simulation was conducted to compare the dynamic properties of monomeric *GtCel5* and its variants N233A, N233D, and N233G. All of the MD simulations were carried out using the Amber 14 package at a temperature of 323K for 20 ns. Force field ff99SB with the TIP3P water model was used to describe the systems [25–27]. All protein atoms were at least 12 Å from the edge of the water box. The systems had net negative charges and were neutralized by addition of sodium ions with the Amber tool program [28]. Prior to the MD simulations, each system was carried out with 10,000 steps of steepest descent for energy minimizations. The trajectories of the first 5 ns were treated as equilibration periods, and the trajectories of the last 15 ns were used for data analyses. The root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) values of the C_α atoms calculated from the equilibrium state were plotted as a function of residue number.

To analyze the interactions between enzyme and substrate, cellotetraose (G4) was docked to *GtCel5* and its variants N233A and N233G, respectively, using YASARA software (<http://www.yasara.org>). MD simulations of the enzyme–cellotetraose complex were then carried out at a temperature of 323K for 20 ns. The Amber force fields ff99SB and GLYCAM_06 [26, 29] were employed to model the cellulase and cellotetraose, respectively. Five thousand snapshots taken from the last 5-ns MD trajectories were used for molecular mechanics/Poisson Boltzmann and surface area continuum solvation (MM/PBSA) calculations. The binding free energy between ligand and protein was calculated by the Amber14. The ΔG value was determined according to the equation: $\Delta G = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}}$. The contributions of internal, electrostatic, and van der Waals' energy to ΔG were analyzed using the force field (http://ambermd.org/tutorials/advanced/tutorial3/py_script/section2.htm) [30].

Hydrogen bond is one of the most important directional intermolecular interactions [31]. Putative hydrogen bonds were assigned based on two geometric criteria: the

distance of less than 3.5 Å and the angle larger than 120° between the acceptor and the hydrogen donor. Visualization and figure preparation of the three dimensional molecules were performed using the PyMOL version 1.7.2.1 (Delano Scientific).

Results

Cloning and sequence analysis of *Gtcel5*

A GH5 cellulase-encoding gene, *Gtcel5* (GenBank Accession No. XP_007867902), was identified in the genome of *G. trabeum* CBS 900.73. The *Gtcel5* contains 1415 base pair that is composed of 7 exons and 6 introns. Deduced *GtCel5* contained 359 amino acid residues including a putative N-terminal signal peptide of 20 amino acids. The catalytic domain showed the highest amino acid sequence identity of 73% with the Cel4 from *Polyporus arcularius* and the endo- β -1,4-glucanase from *Sporotrichum thermophile*, and 72% with the structure-resolved *GtCel5A* (5D8W) [8] and 69% with the structure-resolved *TrCel5A* (3QR3) [7]. As most fungal cellulases of GH5 (EC3.2.1.4) are classified into subfamily GH5_5 [4], *GtCel5* is closely related to cellulases 5D8W and BAF75943 belonging to the same subfamily (see Additional file 1: Fig. S1). Homology modeling indicated that *GtCel5* folds into a typical $(\beta/\alpha)_8$ structure and contains eight highly conserved residues of GH5 enzymes, including Arg67, His111, Asn154, Glu155, His226, Tyr228, Glu267, and Trp300 (numbering without the signal peptide sequence).

Selection of the mutagenesis site in *GtCel5*

Loop regions are proposed to play vital roles in the interactions between TIM-barrel enzymes and substrate [11, 13, 32, 33]. Structure and sequence analysis (Figs. 1, 2) indicated that the Tyr228 and Asn233 of *GtCel5* might be the key switch residues to control the movement of loop 6. The conformational plasticity of this hairpin structure might affect the catalytic performance of *GtCel5*. Tyr228 was highly conserved, while Asn233 showed variation. Therefore, we selected N233 for saturation mutagenesis to investigate the effects of the residue at position 233 on catalytic efficiency of GH5 cellulases.

Production of *GtCel5* and its mutants in *P. pastoris*

GtCel5 and its 19 mutant enzymes were successfully expressed in *P. pastoris* GS115. One protein band of *GtCel5* with the apparent molecular weight of approximately 40 kDa was detected on the SDS-PAGE (Additional file 2: Fig. S2), which was higher than the calculated value (35.7 kDa). After *Endo* H digestion, the *N*-deglycosylated *GtCel5* decreased to approximately 36 kDa. Mutant enzymes had similar apparent molecular weights,

and showed a single band with expected molecular mass after *Endo* H treatment (data not shown).

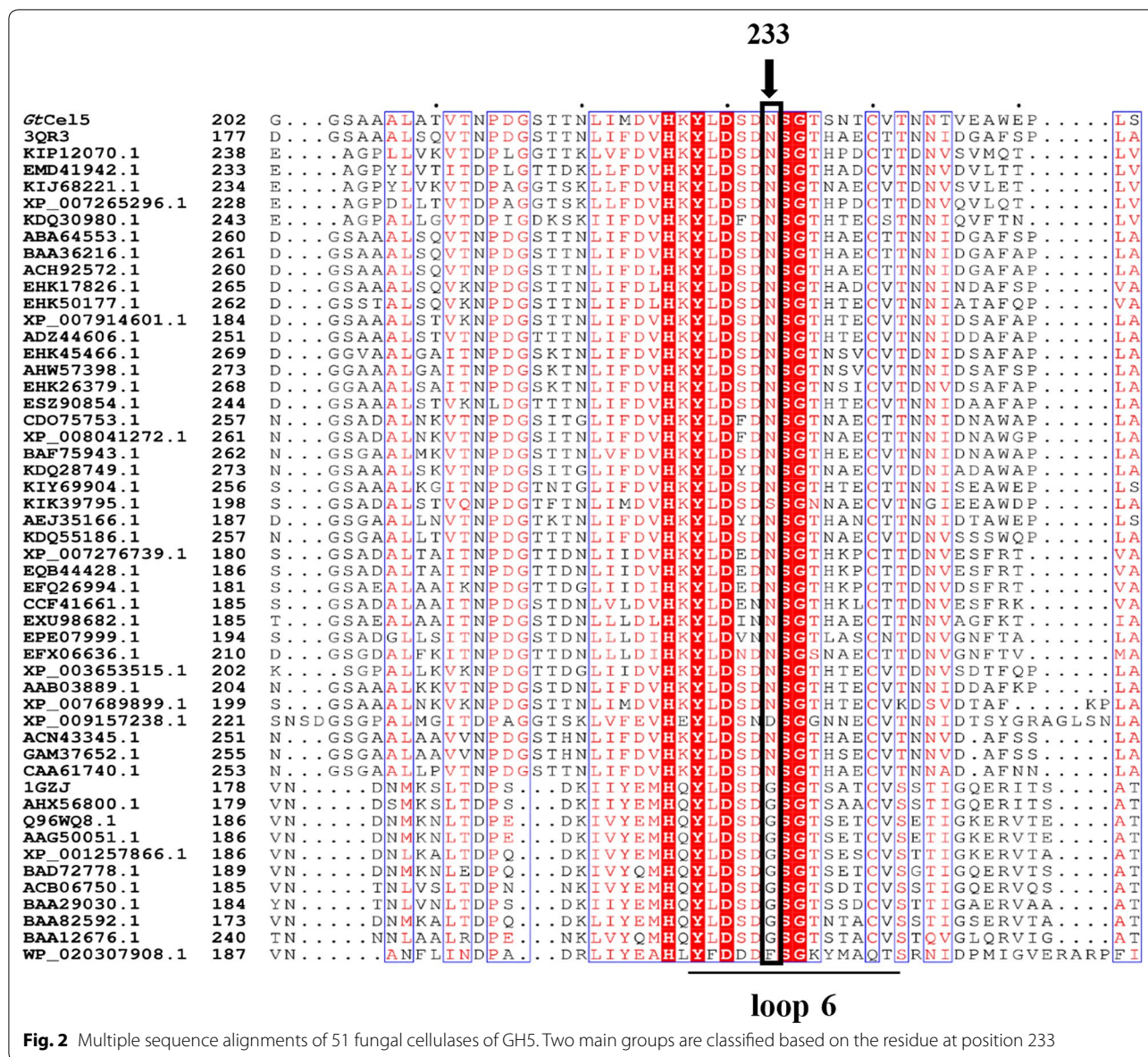
Comparison of the enzymatic properties between *GtCel5* and its variants

When using CMC-Na as the substrate, *GtCel5A* showed the highest activity at pH 4.0 and remained more than 30% active at pH values between 2.2 and 6.0 (Fig. 3a). This pH-activity profile is similar to those of most fungal cellulases. The variants except for N233V showed similar pH optima to the wild type, and the optimal pH of N233V shifted to pH 5.0 (Fig. 3a). As shown in Fig. 3b, *GtCel5A* had an optimal temperature of 70 °C. All of the variants except for N233D were optimally active at 50 or 60 °C, which was 10–20 °C lower than the wild type, while N233D had similar optimal temperature to the wild type. In comparison with the wild type, all variants showed decreased activities except for N233A, N233G, N233S, and N233D (Fig. 3a, b). The stabilities of *GtCel5A* and mutants N233A, N233G, and N233D were also compared. For pH stability, *GtCel5* retained more than 65% of its initial activity after 60-min incubation at 37 °C over a wide pH range (3.0–12.0), while the variants N233D and N233G retained stability over a wider pH range (over 70% activity after 1-h incubation at pH 2.0–12.0) (Fig. 3c). The good stability under both acidic and alkaline conditions makes variants N233D and N233G more favorable for applications in the industries of bioethanol, detergents, and feed additives. *GtCel5* and variants N233A, N233D, and N233G showed similar thermostability (Fig. 3d). The results suggested that the single mutation at position 233 had significant effects on some enzyme properties of *GtCel5*.

Substrate specificities and kinetics of *GtCel5* and its mutants

Of the nine polysaccharide substrates tested, *GtCel5* showed the highest activity on barley β -glucan (6257 ± 26 U/mg) and lichenan (5318 ± 54 U/mg), moderate on CMC-Na (1117 ± 43 U/mg), low toward locust bean gum, and no detectable activity on Avicel, filter paper, xylan, and laminarin. These results indicated that *GtCel5* had no activity on crystalline cellulose and β -1,3 glycosidic linkages. Using CMC-Na as the substrate, *GtCel5* had the K_m , V_{max} , and k_{cat} values of 4.5 ± 0.3 mg/mL, 1475 ± 71 μ mol/min/mg, and 878 ± 44 /s, respectively, according to the Lineweaver–Burk plot.

CMC-Na was selected as the substrate to compare the specific activities and kinetic values of *GtCel5* and its variants (Table 1). For the substitutions at position 233, the main outcome was a significant decrease in specific activity (1.4–138.6-fold) and k_{cat} (0.7–66.5-fold). A few of the variants also showed an increase of K_m (variants N233M,



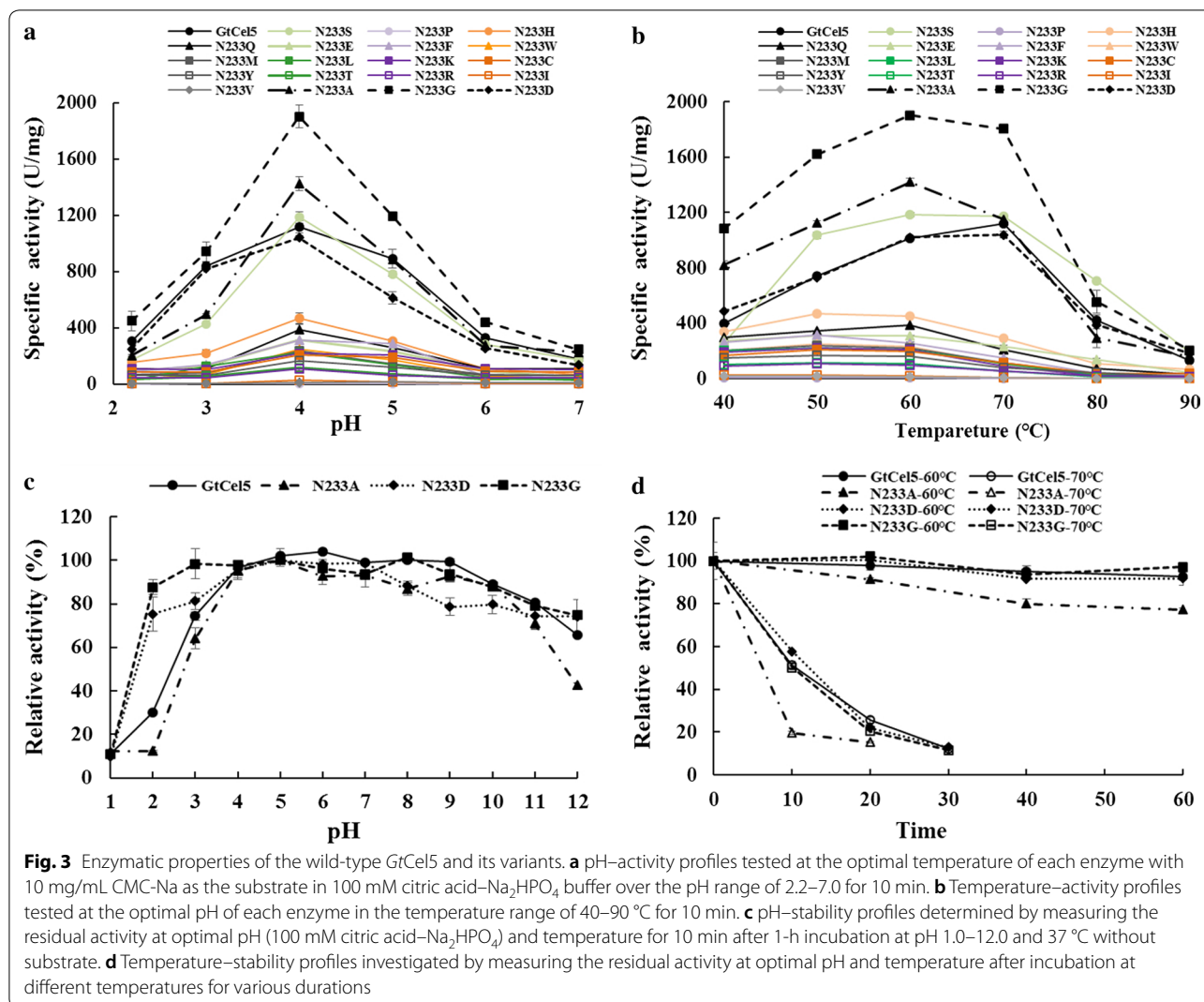


Table 1 The kinetic values of *GtCel5*, *TeEgl5A*, *PoCel5*, and their mutants with CMC-Na as the substrate

Enzymes	Specific activity (U/mg)	K_m (mg/mL)	V_{max} (μ mol/min/mg)	k_{cat} (/s)	k_{cat}/K_m (mL/s/mg)
<i>GtCel5</i>	1117 ± 43	4.5 ± 0.3	1475 ± 71	878 ± 44	195 ± 6
N233G	1901 ± 21	3.6 ± 0.4	1802 ± 96	1072 ± 63	297 ± 8
N233A	1419 ± 22	3.7 ± 0.3	1742 ± 96	1036 ± 57	283 ± 4
N233D	1185 ± 35	4.4 ± 0.7	1511 ± 98	899 ± 87	202 ± 5
N233S	1039 ± 26	7.5 ± 0.2	2302 ± 112	1370 ± 74	184 ± 3
<i>TeEgl5A</i>	732 ± 10	7.4 ± 0.3	1014 ± 54	600 ± 24	81.1 ± 7.5
<i>TeEgl5A</i> _G216A	523 ± 30	10.9 ± 0.4	1255 ± 94	732 ± 43	61.2 ± 6.7
<i>TeEgl5A</i> _G216N	368 ± 15	11.3 ± 0.6	958 ± 76	559 ± 33	49.4 ± 4.3
<i>PoCel5</i>	401 ± 3	5.6 ± 0.3	534 ± 28	311 ± 37	55.1 ± 5.2
<i>PoCel5</i> _G210A	374 ± 2	5.5 ± 0.4	599 ± 36	337 ± 43	62.5 ± 12.1
<i>PoCel5</i> _G210N	166 ± 4	6.9 ± 0.6	280 ± 32	163 ± 24	23.7 ± 3.2

Table 2 The substrate specificities of *GtCel5* and its variants

Enzymes	CMC-Na (U/mg)	Barley β -glucan (U/mg)
<i>GtCel5</i>	1117 \pm 43	6257 \pm 26
N233G	1901 \pm 21	8383 \pm 57
N233A	1419 \pm 22	7866 \pm 56
N233D	1185 \pm 35	5869 \pm 33
N233S	1039 \pm 26	5474 \pm 43

effect on the pH–activity and temperature–activity profiles of *TeEgl5A* and *PoCel5*. However, great changes were detected on the catalytic performance of variants, as the specific activities of *TeEgl5A_G216N* and *PoCel5_G210N* decreased to 50 and 41% of the wild types (Table 1). The kinetic values of the variants showed similar trends, i.e., decreased or similar substrate affinity and catalytic efficiencies. The results suggested that glycine at position 233 on loop 6 does make a contribution to the catalytic performance of GH5 cellulases.

Homology modeling and MD simulation

To determine the structural changes caused by mutation at position 233, the modeled structures of *GtCel5* and its variants N233A, N233D, and N233G with and without substrate (G4) were constructed. MD simulations of 20 ns at 323K were then performed. The RMSD of C_{α} atoms tended to be at equilibrium after 5 ns, and thus, the simulation trajectory of the last 5 ns was selected for further analysis. As shown in Fig. 4a, the RMSD values of variants N233A and N233G were lower than that of *GtCel5*, suggesting that variants have more stable conformations than the wild type. Moreover, the average RMSF values

of N233A (1.84 Å) and N233G (1.87 Å) at loop 6 were higher than that of *GtCel5* (1.74 Å) and N233D (1.46 Å) (Fig. 4b). These results suggested that loop 6, containing the position 233 is more flexible in variants N233A and N233G than in the *GtCel5* and mutant N233D, which may affect the interaction between enzyme and substrate.

The conformations of *GtCel5* and its variants in the force field AMBER99SB were chosen for the analysis of putative hydrogen bonds. As shown in Fig. 5 and Table 3, the Asn233 of *GtCel5* and the Asp233 of N233D formed two hydrogen bonds with Tyr228 and Asp230, and the Tyr228 formed one more hydrogen bond with the catalytic residue Glu267. These three hydrogen bonds had the occupancy rates of 20–40%. However, in variants N233A and N233G, Ala233, and Gly233 form only one hydrogen bond with Asp230; the occupancy rates were 35 and 53%, respectively; and the hydrogen bond between Tyr228 and Glu267 was absent. The results confirmed that mutation at position 233 has significant effects on the local hydrogen-bonding network.

Interactions between residue 233 and the substrate

The interactions between residue 233 and the substrate were analyzed using the YASARA software. As shown in Fig. 6, one hydrogen bond was formed between the A233@O and G4@H6O in variant N233A–cellotetraose complex or G233@O and G4@H6O in N233G–cellotetraose complex. The occupancy rates of these hydrogen bonds were up to 37 and 45%, respectively. However, this hydrogen bond was absent in the *GtCel5*. These results are in accordance with the increased catalytic efficiencies of the two variants. Based on the MM/PBSA calculations, the wild-type *GtCel5* has a binding free energy (ΔG) of -2.7 ± 0.2 kcal/mol, while variants N233A and N233G exhibit much lower ΔG values (-22.2 ± 0.2 and

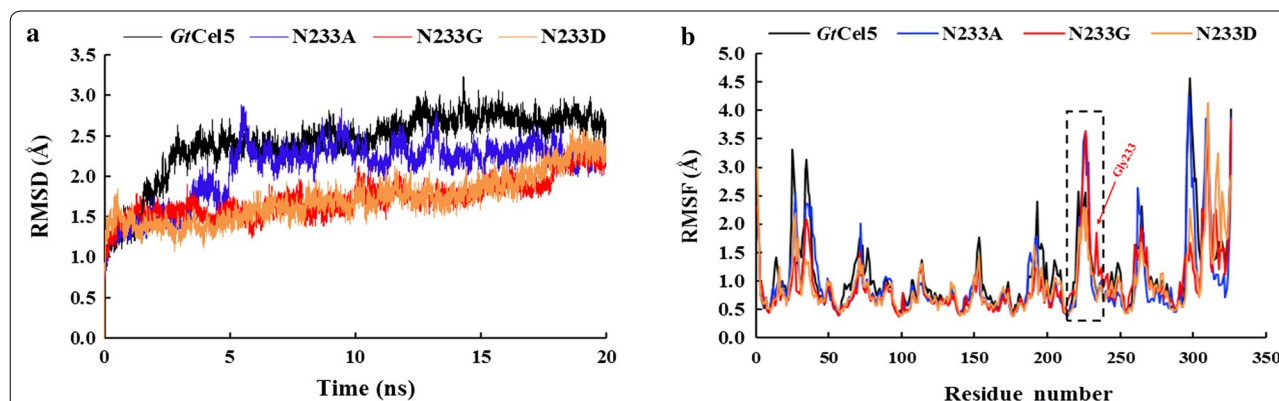
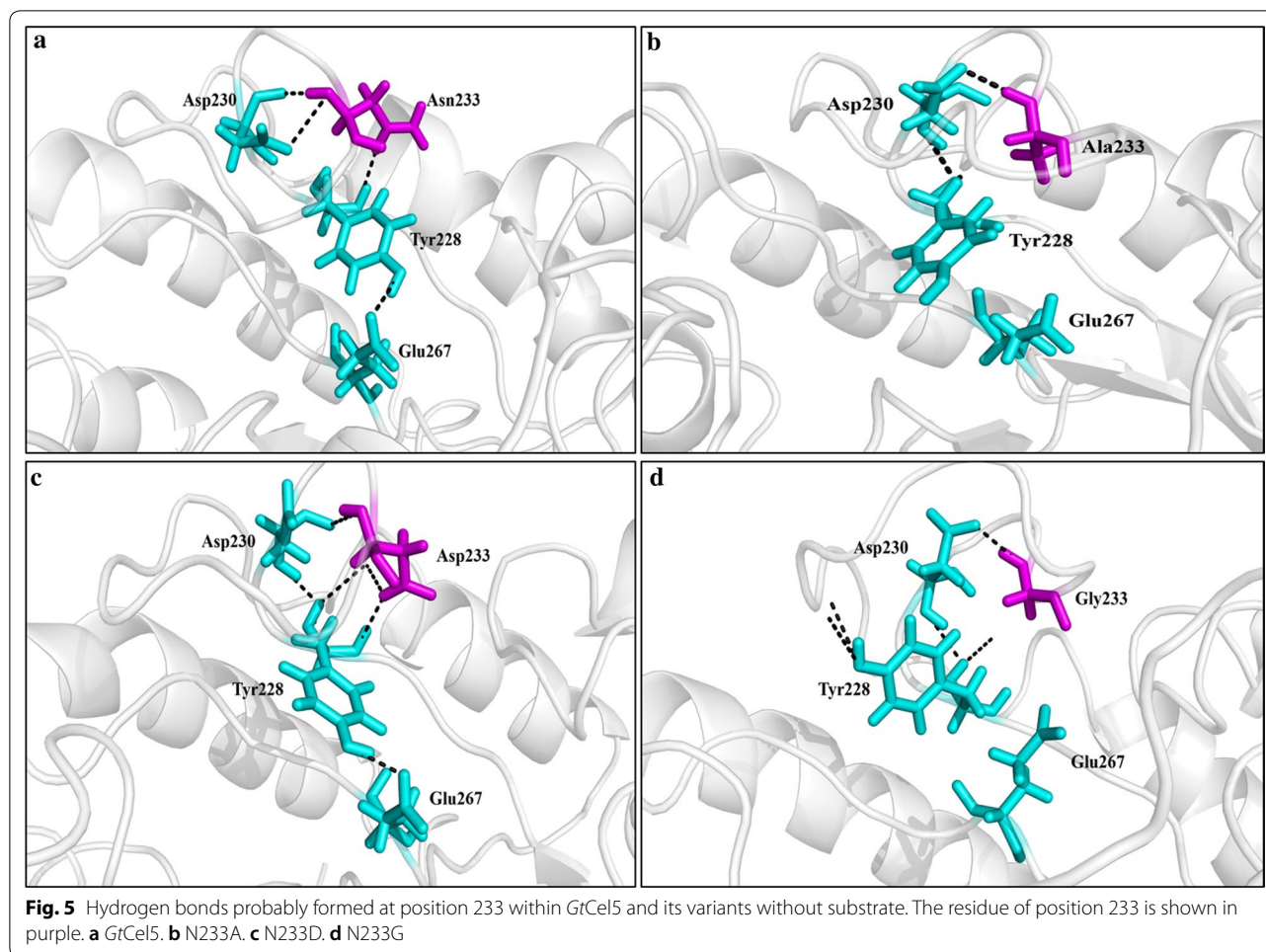


Fig. 4 MD simulation analysis of the wild-type *GtCel5* and its variants, N233A, N233D, and N233G, using the force field of AMBER99SB. The data are collected at 323K for a minimum of 20 ns. **a** The RMSD values. The trajectories of first 5 ns were treated as equilibration periods, and the trajectories of the last 15 ns were used for data analyses. **b** The RMSF values. The loop 6 region is indicated by the dashed box



–32.4 $k_{\text{cal}}/\text{mol}$, respectively). Moreover, the binding energies of *GtCel5* and its variants at position 233 were also calculated. As shown in Fig. 7, N233G showed lower binding energy than that of N233A and *GtCel5*. These findings revealed a stronger interaction between the substrate and N233G.

Discussion

GH5 is a large GH family containing enzymes with broad substrate specificity and various activities, and those from fungi are generally acidic and mesophilic [4]. In this study, an acidic, mesophilic GH5 cellulase was identified in *G. trabeum* CBS 900.73. Based on the key amino acid residue at position 233 of loop 6, the 51 fungal cellulases of GH5 were classified into two main groups: one with asparagine as shown in *GtCel5* and 3QR3 [7], and the other with glycine, as in cellulases *TeEgl5A*, *PoCel5* and 1GZJ [6]. The roles of the residue at position 233 were then revealed in *GtCel5* by saturation mutagenesis, which were further verified by reverse mutation in *GtCel5* homologs *TeEgl5A* and *PoCel5*.

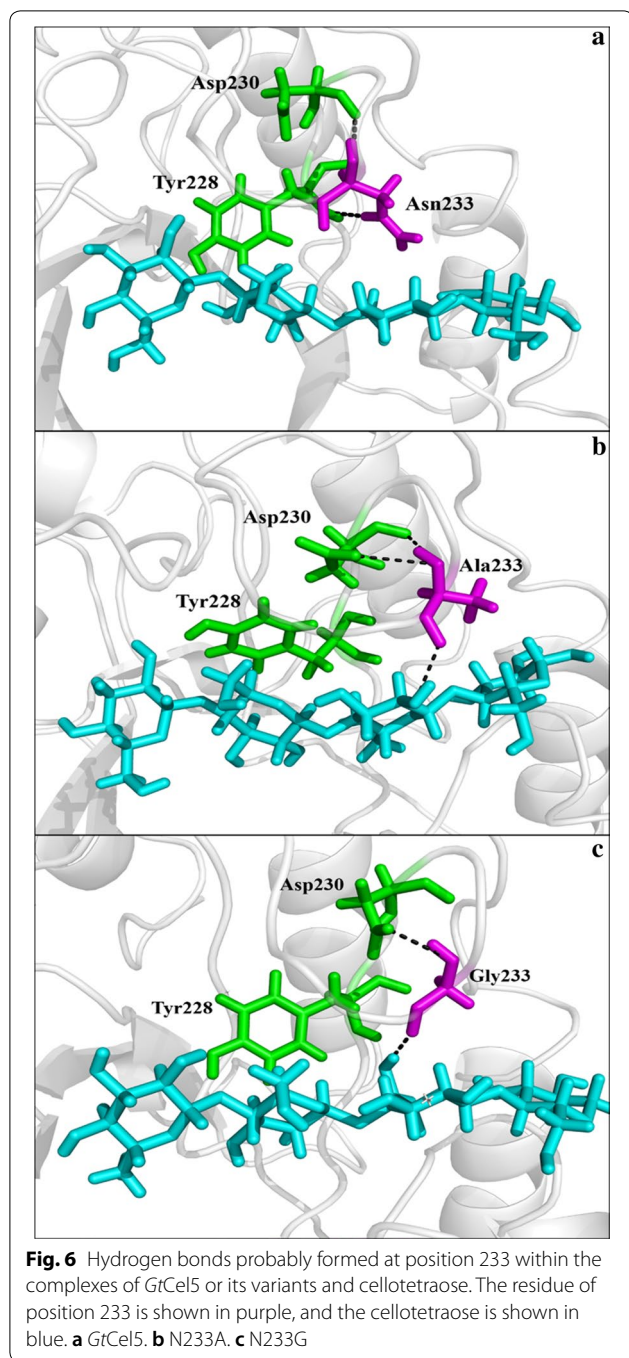
Mobile surface loops have been found to play key roles in protein functions. For example, the thermostability and activity of cellobiohydrolase *TeCel7A* were improved by introducing more disulfide bridges to the loop structures [34]. As for the *N*- α -acetyl transferase from *Sulfolobus solfataricus*, changing the residues of the loop region between sheets β_3 and β_4 destroyed the hydrogen bond network and caused a decrease of 3–7 °C in the protein melting temperature [35]. In the present study, the residue at position 233 was found to have effects on both thermal adaptation and catalytic efficiency of *GtCel5*. The temperature–activity profiles of *GtCel5* and its variants showed great variations. Bioinformatic analysis indicated that the local hydrogen bond network of loop 6 (Fig. 5) varied in the enzymes, which probably contribute to the thermal adaptability.

As the best variants, the catalytic performances of N233A and N233G were compared to that of commercial cellulases. When using CMC-Na as the substrate, the specific activities of the widely used cellulase Cel5A from *Hypocrea jecorina* (*Trichoderma reesei*) [36] and

Table 3 Comparison of the hydrogen bond occupancy rates of GfCel5 and its mutants, N233A and N233G, during the last 15 ns trajectories

Residue ^a	GfCel5			N233D			N233A			N233G		
	Donor	Acceptor	Occupancy rate (%)	Donor	Acceptor	Occupancy rate (%)	Donor	Acceptor	Occupancy rate (%)	Donor	Acceptor	Occupancy rate (%)
228/233	Y228@H	N233@OD1	32	Y228@H	D233@OD2	77	-	-	-	-	-	-
230/233	N233@H	D230@O	20	D233@H	D230@O	13	A233@H	D230@OD2	53	G233@H	D230@OD1	35
228/267	Y228@HH	E267@OE2	40	Y228@HH	E267@OE1	36	-	-	-	-	-	-

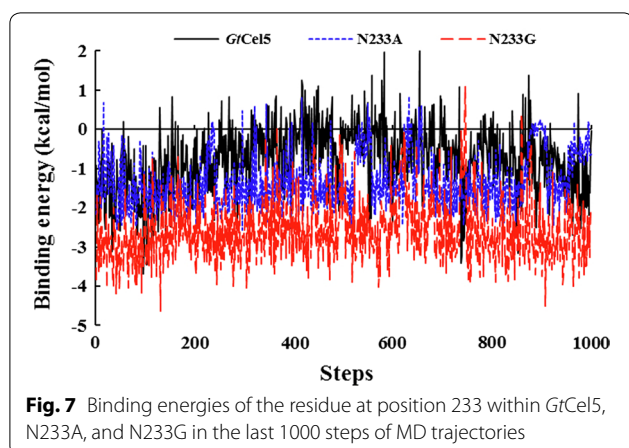
^a The hydrogen bonds around position 233 that exist only in wild-type GfCel5 and N233D are indicated in italics



the commercial cellulase from *Thermotoga maritima* (Magazyme) are 215.6 and 245 U/mg, respectively, which were much lower than those of N233A (1419 U/mg) and N233G (1901 U/mg). However, other variants had similar or decreased activities (Fig. 3). MD simulation analyses indicated that variants N233G and N233A have higher RMSF values in the region of loop 6, which are correspondent to the improved loop flexibility, especially

in N233G. Glycine without side chain has been found to contribute to conformational flexibility of some loop regions, and consequently has effects on enzymatic catalysis and substrate binding. For example, a glycine-rich loop is postulated to undergo conformational change for substrate binding in the mitochondrial-processing peptidase [36], while residue G76 contributes to the active-site loop flexibility of a pepsin [37]. Variants N233A and N233G with more flexible loop 6 showed improvements in substrate affinity (decreased K_m values), turnover rate (increased k_{cat} values), and catalytic efficiency (increased k_{cat}/K_m values) (Table 1), which confirmed the effects of alanine and glycine on the loop conformation. Moreover, MD calculation indicated that the improved flexibility of the loop 6 probably affects the hydrogen-bonding network near the active site indirectly (Fig. 5, Table 3). As a result, the conformational freedom of catalytic Glu267 is reduced. However, without the steric hindrance caused by the hydrogen bond between Tyr228 and Glu267, variants N233A and N233G probably experienced a conformational change of the catalytic pocket. Consequently, these variants having higher mobility at loop 6 and a different hydrogen bond pattern at the active site may bind substrates more easily and thus catalyze the hydrolysis of substrate more efficiently.

To the best of our knowledge, hydrogen bonds are also crucial in substrate recognition and binding [38, 39]. Therefore, we also investigated the hydrogen bonds between the enzyme and substrate. MD analysis of the enzyme–substrate complex dynamics indicated that the Asn233 of *GtCel5* has no direct ligand contact with G4, while Ala233 or Gly233 of variants N233A and N233G was more likely to form a hydrogen bond with G4 with higher occupancy rates. Although this hydrogen bond was also identified in N233D, the occupancy rate was much lower (27%). This result is in agreement with the increased catalytic efficiencies of variants N233A and N233G. Similar results have been reported in the *TrCel7A* from *T. reesei*, in which hydrogen bond interaction exists in the whole catalytic process and plays a role of special importance in stabilizing the intermediate state and improvement of the catalytic performance [40]. Besides, in *TLXyn10A_P* from *Talaromyces leycetanus*, G149D on the loop 4 is able to form a hydrogen bond with substrate and probably plays a major role in the improvement of catalytic performance [41]. To analyze the binding affinity of substrate and enzyme, we performed MM/PBSA calculations, and found that the binding energies of *GtCel5* and its variants are in the order of N233G < N233A < *GtCel5*. These data are correspondent to the experimental work that showed N233G and N233A having higher affinity with cellotetraose than with *GtCel5* (Tables 1, 2). Therefore, the substitution of



Asn233 with alanine or glycine might cause the enzyme to form more stable hydrogen bonds with the substrate and improve the interactions between enzyme and substrate, and consequently enhance the substrate's binding and catalytic efficiencies.

Conclusions

In the present study, an acidic, mesophilic cellulase of GH5 was identified in *G. trabeum* CBS 900.73 and produced in *P. pastoris* GS115. Structure and sequence analyses indicated that the residue at position 233 on loop 6 plays a crucial role in the catalytic performance of GH5 cellulases. By increasing the local hydrogen bond interactions around the residue at position 233 and between the enzyme and substrate, the substrate affinity was enhanced, as was the catalytic efficiency. Considering the significance of GH5 cellulases in biomass conversion, the findings are valuable for the protein engineering of GH5 cellulases in the viewpoints of research, development, and industrial applications.

Additional files

Additional file 1: Table S1. Primers used in this study.

Additional file 2: Fig. S1. The phylogenetic analysis of GrCel5 and other GH5 enzymes of bacterial and fungal sources. GenBank accession numbers or PDB numbers are shown. **Fig. S2.** SDS-PAGE analysis of GrCel5 and its variants. M, the molecular weight markers; 1, 4, 7 and 10, the crude enzymes; 2, 5, 8 and 11, the purified enzymes; 3, 6, 9 and 12, the deglycosylated enzymes with *Endo* H treatment. **Fig. S3.** SDS-PAGE analysis of *TeEgl5A*, *PoCel5* and their variants. M, the molecular weight markers; 1, 4 and 7, the crude enzymes; 2, 5 and 8, the purified enzymes; 3, 6 and 9, the deglycosylated enzymes with *Endo* H treatment. **Fig. S4.** Enzymatic properties of the wild type *TeEgl5A*, *PoCel5* and their variants. (A) pH-activity profiles of *TeEgl5A* and its variants tested at the optimal temperature of each enzyme (90 °C) over the pH range of 3.0–7.0 for 10 min. (B) Temperature-activity profiles of *TeEgl5A* and its variants tested at the optimal pH of each enzyme in the temperature range of 50–95 °C for 10 min. (C) pH-activity profiles of *PoCel5* and its variants tested at the optimal temperature (60 °C) over the pH range of 3.0–8.0 for 10 min. (D) Temperature-activity profiles of *PoCel5* and its variants tested at the optimal pH of each enzyme in the temperature range of 40–90 °C for 10 min.

Abbreviations

CMC-Na: carboxymethylcellulose sodium; G4: cellotetraose; MD: molecular dynamics; RMSD: root-mean-square deviation; RMSF: root-mean-square fluctuation; MM/PBSA: molecular mechanics/Poisson Boltzmann and surface area continuum solvation.

Authors' contributions

FZ performed the experiments and wrote the manuscript. TT and XW participated in the bioinformatic analysis. YW, XS, XX, and RM analyzed the data and revised the manuscript. BY and HL conceived and designed the experiments and revised the manuscript. All the authors read and approved the final manuscript.

Author details

¹ Key Laboratory for Feed Biotechnology of the Ministry of Agriculture, Feed Research Institute, Chinese Academy of Agricultural Sciences, No. 12 South Zhongguancun Street, Haidian, Beijing 100081, People's Republic of China.

² College of Biological Sciences and Biotechnology, Beijing Forestry University, Beijing 100083, People's Republic of China.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Not applicable.

Availability of data and materials

The dataset supporting the conclusions of this article are included within the article and its Additional files 1 and 2.

Funding

This work was supported by the National Natural Science Foundation of China (No. 31572446), the National High-Tech Research and Development Program of China (863 Program, No. 2013AA102803), the National Chicken Industry Technology System of China (No. CARS-42), and the National Key Research and Development Program of China (No. 2016YFD0501409).

Publisher's Note

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

Received: 11 July 2017 Accepted: 14 March 2018

Published online: 22 March 2018

References

- Jönsson LJ, Alriksson B, Nilvebrant NO. Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnol Biofuels*. 2013;6:16.
- Hoshino E, Shiroishi M, Amano Y, Nomura M, Kanda T. Synergistic actions of exo-type cellulases in the hydrolysis of cellulose with different crystallinities. *J Ferment Bioeng*. 1997;84:300–6.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZY) in 2013. *Nucleic Acids Res*. 2014;42:D490–5.
- Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol*. 2013;12:186.
- Tseng CW, Ko TP, Guo RT, Huang JW, Wang HC, Huang CH, Cheng YS, Wang AH, Liu JR. Substrate binding of a GH5 endoglucanase from the ruminal fungus *Piromyces rhizinflata*. *Acta Crystallogr Sect F*. 2011;67:1189–94.
- Lo Leggio L, Larsen S. The 1.62 Å structure of *Thermoascus aurantiacus* endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5. *FEBS Lett*. 2002;523:103–8.
- Lee TM, Farrow MF, Arnold FH, Mayo SL. A structural study of *Hypocrea jecorina* Cel5A. *Protein Sci*. 2011;20:1935–40.

8. Liu G, Li Q, Shang N, Huang JW, Ko TP, Liu W, Zheng Y, Han X, Chen Y, Chen CC, Jin J, Guo RT. Functional and structural analyses of a 1,4- β -endoglucanase from *Ganoderma lucidum*. *Enzyme Microb Technol*. 2016;86:67–74.
9. Yan J, Liu W, Li Y, Lai HL, Zheng Y, Huang JW, Chen CC, Chen Y, Jin J, Li H, Zhong LH, Guo RT. Functional and structural analysis of *Pichia pastoris*-expressed *Aspergillus niger* 1,4- β -endoglucanase. *Biochem Biophys Res Commun*. 2016;475:8–12.
10. Wierenga R. The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett*. 2001;492:193–8.
11. Tu T, Pan X, Meng K, Luo H, Ma R, Wang Y, Yao B. Substitution of a non-active-site residue located on the T3 loop increased the catalytic efficiency of *endo*-polygalacturonases. *Process Biochem*. 2016;51:1230–8.
12. Zou J, Kleywegt GJ, Ståhlberg J, Driguez H, Nerinckx W, Claeysens M, Koivula A, Teeri TT, Jones TA. Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Cel16A from *Trichoderma reesei*. *Structure*. 1999;7:1035–45.
13. Varrot A, Schülein M, Davies GJ. Structural changes of the active site tunnel of *Humicola insolens* cellobiohydrolase, Cel6A, upon oligosaccharide binding. *Biochemistry*. 1999;38:8884–91.
14. Meinke A, Damude HG, Tomme P, Kwan E, Kilburn DG, Miller RCJ, Warren RA, Gilkes NR. Enhancement of the *endo*- β -1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop. *J Biol Chem*. 1995;270:4383–6.
15. Cheng YS, Ko TP, Huang JW, Wu TH, Lin CY, Luo W, Li Q, Ma Y, Huang CH, Wang AH, Liu JR, Guo RT. Enhanced activity of *Thermotoga maritima* cellulase 12A by mutating a unique surface loop. *Appl Microbiol Biotechnol*. 2012;95:661–9.
16. Badleyan S, Bevan DR, Zhang C. A salt-bridge controlled by ligand binding modulates the hydrolysis reaction in a GH5 endoglucanase. *Protein Eng Des Sel*. 2012;25:223–33.
17. Yuan SF, Wu TH, Lee HL, Hsieh HY, Lin WL, Yang B, Chang CK, Li Q, Gao J, Huang CH. Biochemical characterization and structural analysis of a bifunctional cellulase/xylanase from *Clostridium thermocellum*. *J Biol Chem*. 2015;290:5739–48.
18. Bommarius AS, Broering JM, Chaparro-Riggers JF, Polizzi KM. High-throughput screening for enhanced protein stability. *Curr Opin Biotechnol*. 2006;17:606–10.
19. Liang C, Fioroni M, Rodríguez-Ropero F, Xue Y, Schwaneberg U, Ma Y. Directed evolution of a thermophilic endoglucanase (Cel5A) into highly active Cel5A variants with an expanded temperature profile. *J Biotechnol*. 2011;154:46–53.
20. Wang K, Luo H, Bai Y, Shi P, Huang H, Xue X, Yao B. A thermophilic *endo*-1,4- β -glucanase from *Talaromyces emersonii* CBS394.64 with broad substrate specificity and great application potentials. *Appl Microbiol Biotechnol*. 2014;98:7051–60.
21. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007;24:1596–9.
22. Pearson WR. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith–Waterman and FASTA algorithms. *Genomics*. 1991;11:635–50.
23. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22:4673–80.
24. Miller GL. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem (Wash)*. 1959;31:426–8.
25. Lindorff-larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 2010;78:1950–8.
26. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. 2006;65:712–25.
27. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79:926–35.
28. Case DA, Darden TA, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Wang B, Hayik S, Roitberg A, Seabra G, Kolossvai I, Wong KF, Paesani F, Vanicek J, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA. AMBER 11. San Francisco: University of California; 2010.
29. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ. GLYCAM06: a generalizable biomolecular force field. *Carbohydrates*. *J Comput Chem*. 2008;29:622–55.
30. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TR. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res*. 2001;32:889–97.
31. Steiner T. The hydrogen bond in the solid state. *Angew Chem Int Ed*. 2002;41:48–76.
32. Zhai X, Amyes TL, Richard JP. The role of loop-clamping side chains in catalysis by triosephosphate isomerase. *J Am Chem Soc*. 2015;137:15185–97.
33. Zhai X, Go MK, O'Donoghue AC, Amyes TL, Pegan SD, Wang Y, Loria JP, Mesecar AD, Richard JP. Enzyme architecture: the effect of replacement and deletion mutations of loop 6 on catalysis by triosephosphate isomerase. *Biochemistry*. 2014;53:3486–501.
34. Voutilainen SP, Murray PG, Tuohy MG, Koivula A. Expression of *Talaromyces emersonii* cellobiohydrolase Cel7A in *Saccharomyces cerevisiae* and rational mutagenesis to improve its thermostability and activity. *Protein Eng Sel*. 2010;23:69–79.
35. Chang YY, Hsu CH. Multiple conformations of the loop region confers heat-resistance of SsArd1, a thermophilic NatA. *ChemBioChem*. 2015;17:214–7.
36. Nagao Y, Kitada S, Kojima K, Toh H, Kuhara S, Ogishima T, Ito A. Glycine-rich region of mitochondrial processing peptidase alpha-subunit is essential for binding and cleavage of the precursor proteins. *J Biol Chem*. 2000;275:34552–6.
37. Okoniewska M, Tanaka T, Yada RY. The pepsin residue glycine-76 contributes to active-site loop flexibility and participates in catalysis. *Biochem J*. 2000;349:169.
38. Kim D, Bo HP, Jung BW, Kim M, Hong SI, Lee DS. Identification and molecular modeling of a family 5 endocellulase from *Thermus caldophilus* GK24, a cellulolytic strain of *Thermus thermophilus*. *Int J Mol Sci*. 2006;7:571–89.
39. Verdoucq L, Morinière J, Bevan DR, Esen A, Vasella A, Henrissat B, Czjze M. Structural determinants of substrate specificity in family 1 β -glucosidases: novel insights from the crystal structure of sorghum dhurrinase-1, a plant β -glucosidase with strict specificity, in complex with its natural substrate. *J Biol Chem*. 2004;279:31796–803.
40. Li J, Du L, Wang L. Glycosidic-bond hydrolysis mechanism catalyzed by cellulase Cel7A from *Trichoderma reesei*: a comprehensive theoretical study by performing MD, QM and QM/MM calculations. *J Phys Chem B*. 2010;114:15261–8.
41. Wang X, Huang H, Xie X, Ma R, Bai Y, Zheng F, You S, Zhang B, Xie H, Yao B. Improvement of the catalytic performance of a hyperthermostable GH10 xylanase from *Talaromyces leycettanus* JCM12802. *Bioresour Technol*. 2016;222:277–84.