CrossMark

Rational engineering of Cel5E from *Clostridium thermocellum* to improve its thermal stability and catalytic activity

Ibrahim Torktaz[1] · Ali Asghar Karkhane[1] · Jafar Hemmat[2]

Abstract

The *celH* gene from *Clostridium thermocellum* encodes a protein containing 900 residues and three components, including Cel5E, Lic26a, and carbohydrate-binding domains. Cel5E is a member of the glycoside hydrolase-5 family and is a bifunctional xylanase/cellulase enzyme. We targeted a semi-hydrophobic pocket near the Cel5E active site and theoretically screened mutated variants for enhanced levels of thermal stability. Cel5E mutations were inserted into *celH* by overlapping polymerase chain reaction, followed by expression of wild-type and mutant enzymes in *Escherichia coli* BL21 (DE3) and purification by affinity chromatography. Thermal-stabilizing mutations were subjected to molecular dynamics simulation, and measurement of the in vacuo potential energy, van der Waals forces, electrostatic interactions, and net nonbonded potential energies obtained an overall binding affinity of -64.964 KJ/mol for wild-type Cel5E and -176.148, -200.921, and -120.038 KJ/mol for the N94F, N94W, and E133F mutants, respectively. Additionally, the N94W, N94F, E133F, and N94A variants exhibited 1.92-, 1.29-, 1.1-, and 1.15-fold better carboxymethyl cellulase (CMCase) and 1.46-, 1.29-, 1.11-, and 1.12-fold better β-glucanase activity on barley β-glucan relative to the wild-type enzyme. Interestingly, the optimal temperature for CMCase activity by the N94W variant was shifted 2 °C higher than that for the wild-type enzyme. Mutated variants showed improved CMCase and β-glucanase activity and shifted toward higher temperature of maximum activity.

Keywords *Clostridium thermocellum* · *celH* · Cel5E · Rational mutagenesis · Activity

Introduction

In the cellulose hydrolysis industry, pretreatment of lignocellulose materials and enzymatic catalysis represent two costly stages. Because enzymatic catalysis is usually performed at high temperatures, commercialized cellulase enzymes are mostly derived from thermophilic or thermostable sources or are naturally thermotolerant. To reduce costs and increase efficiency, development of new enzymes with higher levels of thermal stability and enhanced catalytic activity is needed. The extracellular thermophilic cellulase from *Trichoderma reesei* is used in the pulp and paper industry (Bergquist et al. 2002), whereas thermophilic glycoside hydrolases from *Aspergillus* spp. are widely used for commercial purposes (Mehboob et al. 2014; Moretti et al. 2012). Additionally, thermostable *Dictyoglomus* cellulases are also commercialized (Peacock et al. 2013).

Currently, most commercial cellulose-degrading enzymes are engineered proteins exhibiting enhanced thermal stability and activity. In a study by Escovar-Kousen et al. (2004), Cel9A from *Thermobifida fusca* showed increased activity by 40% following rational design of the F476Y mutation. Additionally, Vu and Kim (2012) described a 7.93-fold enhanced *Bacillus amyloliquefaciens* endoglucanase activity after a single amino acid substitution (E289V). Moreover, a glycoside hydrolase (GH)5 family member and thermostable endoglucanase encoded by *celH* from *Clostridium*

✉ Ali Asghar Karkhane
  Karkhane@nigeb.ac.ir

✉ Jafar Hemmat
  J.hemmat@gmail.com

1 National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran

2 Biotechnology Department, Iranian Research Organization for Science and Technology (IROST), Tehran, Iran

thermocellum is a component of the cellulosome complex and used for biomass conversion and biofuel applications (Hirano et al. 2016). This enzyme was first sequenced by Yague et al. (1990), revealing its ability to hydrolyze carboxymethylcellulose, p-nitrophenyl-β-D-cellobioside, methylumbelliferyl- β-D-cellobioside, barley β-glucan, and larchwood xylan. This enzyme contains 900 residues, including a 44-residue N-terminal signal-peptide sequence, cohesion-binding site, dockerin-binding site, carbohydrate-binding domain, ion-binding site, GH-conserved domains, and Pro/Thr-rich linker regions (Yague et al. 1990). The GH domains of endoglucanase H are categorized as belonging to family 5 (Cel5E) in the N- terminal region and family 26 (Lic26a) in the C-terminal region of the protein. Moreover, its carbohydrate-binding domain shows high levels of similarity to carbohydrate-binding module family 11, with residues in the N- and C-termini unnecessary for Cel5E catalytic activity (Yague et al. 1990). Therefore, for pretreated substrates, Cel5E can perform endohydrolysis of $(1 \geq 4)$-β-D-glucosidic linkages in cellulose, as well as xylanase activity, with the carbohydrate-binding module influencing its substrate specificity (Ichikawa et al. 2015). The crystal structure of the Cel5E active site contains 414 amino acids (Yuan et al. 2015) is highly similar to that of other GH5 family members, such as *Clostridium cellulolyticum* endoglucanases (PDB: 1EDG and 3AMC) (Wu et al. 2011). Although the Cel5E active site is responsible for endoglucanase activity, this activity was not observed in recombinant variants expressed in *Escherichia coli* (Yuan et al. 2015). The structure of Cel5E contains a TIM barrel $(\alpha/\beta)8$ architecture common in other GH5 family members, with two Glu residues predicted as active site residues (UniProt: P16218). Within the TIM barrel structure, there is a semi-hydrophobic pocket comprising β2, β3, and β4 strands located near the active site and catalytic residues. Cel5E as a single catalytic subunit exhibits specific activity of ~ 300 U/mg on carboxymethyl cellulose (CMC) and 1200 U/mg on barley β-glucan (Yuan et al. 2015); however, when the enzyme is linked to a carbohydrate-binding domain, the activity of the complex can increase by up to tenfold (Ichikawa et al. 2015). Therefore, the catalytic subunit and the carbohydrate-binding domain represent two targets for activity improvement.

In this study, we targeted the Cel5E active site and performed rational site-directed mutagenesis within the semi-hydrophobic pocket in order to increase the hydrophobicity level of the pocket to (i) increase protein internal forces to enhance enzyme stability against thermal denaturation and (ii) influence active-site conformation to enhance specific activity.

Materials and methods

Software and servers

Prediction of ligand-binding sites was performed on the DEPTH server (http://mspc.bii.a-star.edu.sg/tankp/run_depth.html) (Tan et al. 2013). Prediction of protein stability following point mutation was performed using PoPMuSiC (http://dezyme.com/) (Dehouck et al. 2011), I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) (Capriotti et al. 2005), iStable (http://predictor.nchu.edu.tw/iStable/indexSeq.php) (Chen et al. 2013), and MUpro (http://mupro.proteomics.ics.uci.edu) web servers (Cheng et al. 2006). Molecular-docking studies were performed using AutoDock and Molegro Virtual Docker (MVD) (Thomsen and Christensen 2006). Energy minimization and molecular dynamics simulations were performed using GROMACS 4.6.5 software and the GROMOS 9653a6 force field (Pronk et al. 2013). Protein root mean square deviation (RMSD) was calculated using UCSF Chimera software (Pettersen et al. 2004). Mutations were simulated on the protein structure using modeler version 9.12 software (Fiser and Sali 2003). Enzyme kinetic parameters were calculated using SigmaPlot version 12.5 (Systat Software, Inc., San Jose, CA, USA; www.systatsoftware.com).

Mutation selection

The GH from *C. thermocellum* is a thermostable enzyme and harboring several stabilizing regions. To predict areas responsible for thermostability, we used the SCide algorithm (Dosztanyi et al. 2003). Following a strategy to enhance thermal stability, we performed systematic mutations in the protein structure, with every amino acid position mutated to other amino acids, followed by assessment of thermal stability for each mutation. This process was performed on the PoPMuSiC web server, which is a trained neural network (Dehouck et al. 2011). The top ranked mutations predicted to enhance thermal stability by > 0.9-fold were selected for further complementary in silico validation using neural-network-based approaches. The top predicted stabilizing mutations were evaluated by I-Mutant 2.0, MUpro, and Istable, with overlapping data considered theoretically thermal-stabilizing mutations. To assess the effect of these mutations on protein structure, we generated three-dimensional (3D) models of the mutants using Modeller 9.12 software and a template structure (PDB: 4U3A). Mutated structures were optimized by the variable target-function method and the conjugate-gradient technique. Optimized mutant structures were evaluated for predictive accuracy using the Q-Mean (Benkert et al. 2008) and ProSA (Wiederstein and Sippl 2007) algorithms. Verified structures were used for protein-ligand-interaction studies.

Molecular simulation

The cellulose crystal structure was obtained from a complex structure (PDB: 3AOF). Cellulose was docked to the active site of the wild-type and mutant Cel5E structures in the same conditions and coordinates, which were obtained by superimposing our modeled structures onto the crystallographic structures of *Thermotoga maritima* Cel5A in complex with a cellobiose substrate (PDB: 3AMG) and *T. maritima* Cel5A in complex with mannotriose (PDB: 3AOF). Additionally, a computational method using the DEPTH server was used to determine the exact coordinates of the binding site (Tan et al. 2013). In the predicted binding site, E137 and E242 as catalytic residues were present. Molecular docking simulations were performed using MVD and PYRX software (Dallakyan and Olson 2015). MVD includes the MolDock (Thomsen and Christensen 2006) and PLANTS (Korb et al. 2009) scoring functions, and PYRX is a graphical interface for Autodock and Autodock Vina, which use a Lamarckian genetics algorithm as the scoring function (Morris et al. 2009). Charges were calculated and added to the protein and ligand by MVD, and the best-designed structures underwent molecular dynamics simulations using GROMACS version 4.6.5 and the gromos9653a6 force field (Hess et al. 2008). The ligand structures were parameterized by the PRODRG web server (Schuttelkopf and van Aalten 2004), and the force field simulated a dodecahedral water box with a 1.5-nm distance used for the solvation step. The system was neutralized using Na or Cl ions, and after a 50-ns simulation, binding affinities were calculated using the g_mmpbsa package (Kumari et al. 2014).

Bacterial strains, plasmids, and culture

E. coli DH5α cells were used for cloning, and *E. coli* BL21 (DE3) sigma CMC0014 was used for protein expression (Sigma-Aldrich, St. Louis, MO, USA). Cells were cultured in Luria-Bertani (LB) broth supplemented with 100 μg/mL kanamycin at 37 °C until the optical density at 600 nm (OD_{600}) reached 0.6. The culture was centrifuged at 2000g for 5 min, and the supernatant was replaced with an equivalent volume of fresh LB medium. To induce recombinant protein expression, 0.5 mM isopropyl-β-thiogalactopyranoside (IPTG) was added, followed by incubation at 30 °C for overnight. The pGH plasmid (GV0108-A; Generay Biotech, Shanghai, China) was used as the cloning vector, and the pET-26b(+) vector (Novagen; Merck, Darmstadt, Germany) was used as the expression vector.

Gene cloning and mutagenesis

The amino acid sequence of the GH enzyme from *C. thermocellum* obtained from the NCBI database (accession No. 500163419) included the catalytic domain, with both N- and C-terminal linkers added to the catalytic core of the protein. Residues 325 through 630 were selected for gene cloning and expression studies. *E. coli*-optimized codons were generated using the OPTIMIZER web server (Puigbo et al. 2007). The optimized gene was inserted into the pGH vector using the cloning sites *Nco*I-*Xho*I. We performed site-directed mutagenesis by overlap extension using the polymerase chain reaction (PCR), as described previously (Ho et al. 1989). Mutant primers were designed using SnapGene software (GSL Biotech, Chicago, IL, USA). The mutant segments were confirmed by sanger sequencing and cloned into the pET-26b(+) vector using the same *Nco*I-*Xho*I sites, followed by transformation into *E. coli* BL21 (DE3) competent cells.

Expression and purification of wild-type and mutant proteins

Wild-type and mutant genes were cloned at a point immediately following the PelB leader sequence. Transformed *E. coli* BL21 (DE3) cells harboring the pKHT-26-Native (GenBank: MG717496), pKHT-26-N94A (GenBank: MH404258), pKHT-26-N94W (GenBank: MG717499), pKHT-26-E133A (GenBank: MH404257), pKHT-26-N94F (GenBank: MG717498), and pKHT-26-E133F (GenBank: MG717497) recombinant plasmids were cultured in 10 mL LB medium containing kanamycin (100 mM). The starter cultures were incubated at 37 °C overnight and used to inoculate 100-mL primary cultures, which were then incubated at 37 °C to an OD_{600} of 0.6, followed by addition of 0.4 mM IPTG to induce protein expression at 30 °C overnight. Cells were separated by centrifugation at 4 °C, and the supernatant was collected. For purification of wild-type and mutant enzymes, nickel-affinity chromatography was performed by adding binding buffer (10 mM imidazole, 300 mM NaCl, and 50 mM NaH_2PO_4) to the collected supernatant and loading the supernatant onto a 3-mL nickel-nitrilotriacetic acid column equilibrated with the same binding buffer. The column was washed with wash buffer (20 mM imidazole, 300 mM NaCl, and 50 mM NaH_2PO_4), and products were eluted with elution buffer (250 mM imidazole, 300 mM NaCl, and 50 mM NaH_2PO_4). Eluted products were evaluated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Enzyme activity assay

Enzyme activity was assessed using the 3,5-dinitrosalicylic acid (DNS) method using CMC-Na (Sigma Aldrich) and barley β-glucan (G6513; Sigma-Aldrich) as substrates. Purified

enzyme (100 μL) was added to 900 μL of substrate solution containing 1.0% (*w/v*) CMC-Na diluted in 50 mM phosphate buffer (Gusakov et al. 2011). The reaction mixture was incubated at 60 °C for 10 min, and the enzymatic reaction was terminated by adding 1 mL DNS, followed by boiling for 5 min. Released reducing sugars were determined by measuring the absorption of the solution at 540 nm. One unit was defined as the amount of enzyme required to release 1 μmol of glucose-reducing-sugar equivalents per min. The activities of the wild-type and mutant enzymes were measured in a temperature range of 40 to 80 °C and a pH range of 3 to 11.

Results

Sequence and structure analyses of the GH5 endoglucanase from *C. thermocellum*

Because the catalytic core of the GH5 endoglucanase from *C. thermocellum* (residues 344–604) is not expressed in soluble and active forms of the enzyme

(Yuan et al. 2015), we added N- and C-terminal linker regions to enhance enzyme solubility and activity. Therefore, a sequence containing 306 amino acids (325–630) was used for this study. Although the N- and C-terminal linkers were added to the catalytic core, the selected sequence contained neither the N-terminal glycosyl hydrolase family 26 conserved domain (conserved domain database: 304502) nor the C-terminal carbohydrate-binding domain family 11 region (conserved domain database: 281426). After re-numbering the residues, the catalytic residues in the active site of our truncated protein included residues Glu137 and Glu242. Superimposition and other analytical methods indicated that the binding site coordinates were as follows: X, − 0.92; Y, 35.29, and Z, − 5.66. Additionally, residues present in and around the substrate-binding site were predicted as Asn277, Trp275, Glu137, Tyr173, His196, Trp174, Asn175, Tyr197, Asp199, Phe195, Tyr201, Met245, Tyr198, His205, Met287, Glu242, Asn285, His97, His96, and Asn136. The structure of the truncated catalytic core comprised six α helices and two partial α-helix-like structures in the N- and

Fig. 1 The predicted structure of the truncated form of glycoside hydrolase GH5 from *Clostridium thermocellum* using meta-prediction approach. **a** The structure of truncated enzyme consisting of six alpha helixes plus two partial alpha like structures in the N and C terminal regions with eight beta sheets. **b** The energy profile per residue of predicted model that is calculated by QMEAN. **c** The assessment of predicted structure quality by ProSA Z score of − 8.77 indicated that our predicted 3D model has X-ray diffraction quality. **d** QMEAN, DisCo, and QMEANDisCo scores indicated that the predicted model has x-ray diffraction quality
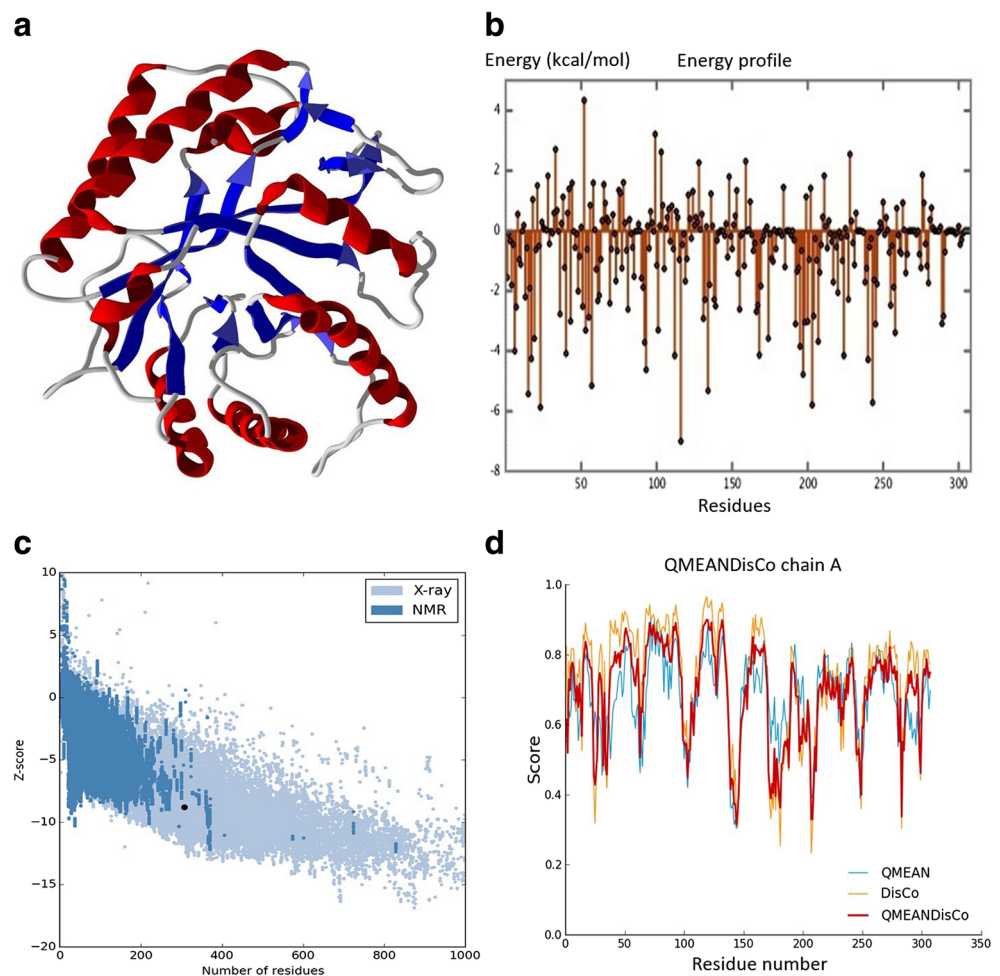
C-terminal regions along with eight β sheets (Fig. 1a). Assessment of predicted structure quality returned a Z-score of -8.77, indicating that our predicted 3D model had X-ray diffraction quality and was similar to that found in crystallographic structures (Fig. 1c, d).

Mutagenesis

For rational design, we theoretically predicted candidate substitutions using pre-trained artificial neural networks, which were chosen because the data used for their training were experimentally verified. Among the predicted mutations, we selected substitutions exhibiting reduced fluctuations within the Cel5E hydrophobic pocket, as previously described (Badieyan et al. 2012; Pikkemaat et al. 2002). The resulting predicted thermal-stabilizing mutations are described in Table 1. The 3D conformers of the top 50 mutants were predicted and optimized, followed by quality verification by Z-score analysis (data not shown). Mutant structures were then used for molecular-docking analyses performed in a water box equilibrated with neutralizing ions. Among the top mutations predicted to increase both thermal stability and ligand-binding affinity (Table 1), we selected residues at positions 94 and 133, which were unanimously predicted as hot spots for thermal-stability modification. Mutations resulting in considerable enhancement in both ligand-binding affinity and thermal stability included E133F, E133A, A94A, N94W, and N94F. These variants underwent a 50-ns molecular dynamics simulation, with Fig. 2 showing the backbone RMSD plot of the wild-type and each of the mutant variants. Results showed that the simulated constructs reached an equilibrated state within 50 ns, and that the ligands were stable within the mutant proteins. We then calculated ligand-binding affinity, including in vacuo potential energy, Van der Waals forces, electrostatic interactions, and net non-bonded potential energy between protein and ligand, as well as the polar solvation energy of the protein-inhibitor complex (Fig. 3 and Table 2). The binding affinity was -64.964 KJ/mol for the wild-type variant as compared with -176.148, -200.921, -120.038, -101.26, and -123.76 KJ/mol for N94F, N94W, E133F, N94A, and E133A, respectively, indicating enhanced ligand-binding affinities in all of the mutants. Similar results were obtained using an alternative method (MVD; Table 2).

Expression, purification, and enzyme assays

In vitro experiments were performed in order to verify the predicted increases in specific activity of the mutant enzymes. Mutant primers (Table 3) were used for overlapping PCR, followed by transformation and expression

of the wild-type and mutant variants in *E. coli* BL21 (DE3) cells. The recombinant proteins were isolated and purified by nickel-affinity chromatography and assessed by SDS-PAGE (Supplemental Fig. S1). Purified proteins were used for specific activity assays. To determine the optimal assay conditions, wild-type and mutant enzymes were subjected to pH ranges of 3 to 11 (Fig. 4), and temperature ranges from 50 to 70 °C, as the enzyme is naturally thermostable, with optimal activity observed at 60 °C (Yuan et al. 2015). The pH assay indicated maximal activity at 537 U/mg for the wild-type enzyme at pH 7 in the presence of CMC and 1164 U/mg in the presence of barley β-glucan. Notably, this activity was obtained in a single subunit comprising the catalytic core of Cel5E and excluding the carbohydrate-binding domain. Although the activity of our truncated variant was 1.7-fold higher than previously reported truncated forms (Yuan et al. 2015), the carbohydrate-binding domain can increase the overall activity of the enzyme by up to tenfold (Ichikawa et al. 2015). Moreover, at acidic pH levels of 3 and 4 and alkaline pH levels of 10 and 11, enzyme activity was undetectable in wild-type Cel5E. Interestingly, low levels of CMCase activity (39 U/mg) were detected in the E133F mutant at pH 10. It is possible that substituting Glu133 with a phenylalanine resulted in an overall positive charge of the enzyme. A previous study reported altered activity across different pH ranges (Mahadevan et al. 2008), where five single amino acid substitutions resulted in a shift in the optimal pH from 5 to 5.4 in Cel5A from *T. maritima*. Although considerable variation in the overall charge of a protein occurs at pH 10, we speculated that modification of the hydrophobic pocket in the core of the protein might increase the hydrophobicity of the enzyme core, thereby leading to increased enzyme stability at an alkaline pH. In all of the mutant enzymes, we observed 4% activity at pH 10 and pH 4 relative to no activity observed in the wild-type enzyme. This result was likely due to the structural stability promoted by alteration of the hydrophobic pocket located at the center of the enzyme.

Thermal stability and enzyme kinetics

To measure residual activities, the purified enzymes were incubated in 0.1 M phosphate buffer for 4 h at the optimal pH 7 and temperatures (60 and 62 °C). The wild-type enzyme retained its residual CMCase activity (~88%) at 60 °C and 86% at 62 °C, whereas activity in the presence of barley β-glucan was 87 and 82% at these respective temperatures. For the N94W variant, residual CMCase activity was 95 and 92% at 60 and 62 °C, respectively, and 94 and 91% in the presence of barely β-glucan at

Table 1 Predicted protein stability changes upon point mutation

Mutation	DDG (PupMusic)[*]	DDG (Imutant)[**]	MUpro	Istable
E133C	−1.46	0.76	Decrease	Decrease
E133F	−2.53	1.24	Increase	Increase
E133H	−1.44	−0.63	Decrease	Decrease
E133I	−2.19	0.83	Increase	Increase
E133L	−1.77	0.90	Increase	Increase
E133M	−1.65	1.16	Increase	Increase
E133V	−1.62	−0.15	Increase	Increase
E133W	−1.77	0.22	Increase	Increase
E133Y	−2.27	0.66	Increase	Increase
E137F	−0.77	1.18	Increase	Increase
E137I	−0.90	0.74	Decrease	Increase
E137Y	−0.74	0.76	Decrease	Decrease
N18I	−0.78	1.74	Decrease	Decrease
V181N	−0.75	−3.65	Decrease	Decrease
P186E	−0.75	−1.66	Decrease	Decrease
P186I	−0.81	−1.29	Decrease	Decrease
P186Q	−0.86	−2.25	Decrease	Decrease
P186T	−0.78	−2.05	Decrease	Decrease
P186V	−1.02	−1.54	Decrease	Decrease
P189F	−0.96	−1.15	Decrease	Decrease
P189I	−0.77	−1.17	Decrease	Decrease
P189W	−0.94	−2.33	Decrease	Decrease
P189Y	−0.95	−2.11	Decrease	Decrease
W211G	−0.76	−2.11	Increase	Increase
D216F	−0.78	−0.17	Decrease	Increase
E242F	−0.88	0.82	Decrease	Increase
E242W	−0.78	−0.59	Decrease	Increase
E242Y	−0.81	0.06	Decrease	Increase
S282G	−0.76	−1.61	Decrease	Decrease
S32G	−1.20	−1.62	Null	Decrease
N50F	−0.82	0.15	Increase	Increase
N50 W	−0.93	−0.45	Decrease	Decrease
N50Y	−1.07	−0.36	Decrease	Decrease
T91I	−0.83	−1.94	Increase	Increase
T91 V	−0.76	−2.14	Increase	Increase
N94F	−1.37	0.52	Decrease	Increase
N94I	−1.19	1.67	Decrease	Increase
N94 L	−0.96	0.93	Null	Increase
N94 M	−0.89	1.42	Increase	Increase
N94 V	−1.00	0.62	Increase	Increase
N94 W	−1.26	0.22	Null	Increase
N94Y	−1.21	0.10	Increase	Increase
S95C	−0.76	−2.06	Decrease	Decrease
S95F	−1.15	−1.02	Decrease	Decrease
S95I	−0.95	−1.27	Decrease	Increase
S95 L	−0.87	−1.75	Increase	Increase

[*]The negative scores in PupMusic indicates increase in the thermal stability

[**]The positive scores in Imutant indicates the stability incensement. The Imutant data retrieved from Imutant V2

Fig. 2 The root mean square deviation of native, E133F, E133A, A94A, N94W, and N94F after 50 ns molecular dynamics simulation by GROMACS package using gromos 9653a6 force filed. **a** The backbone RMSD of native model within 50 ns MD simulation reached plateau state. **b** The backbone RMSD of E133F mutant reached plateau state during 50 ns MD simulation and indicated similar fluctuations in protein backbone like native model. **c** The backbone RMSD of N94F mutant indicated similarity to native model and reached plateau state during 50 ns MD simulation. **d** N94W model reached plateau state after 35 ns MD simulation. **e** N94A model has reached plateau state after 20 ns. **f** E133A model reached plateau state after 35 ns
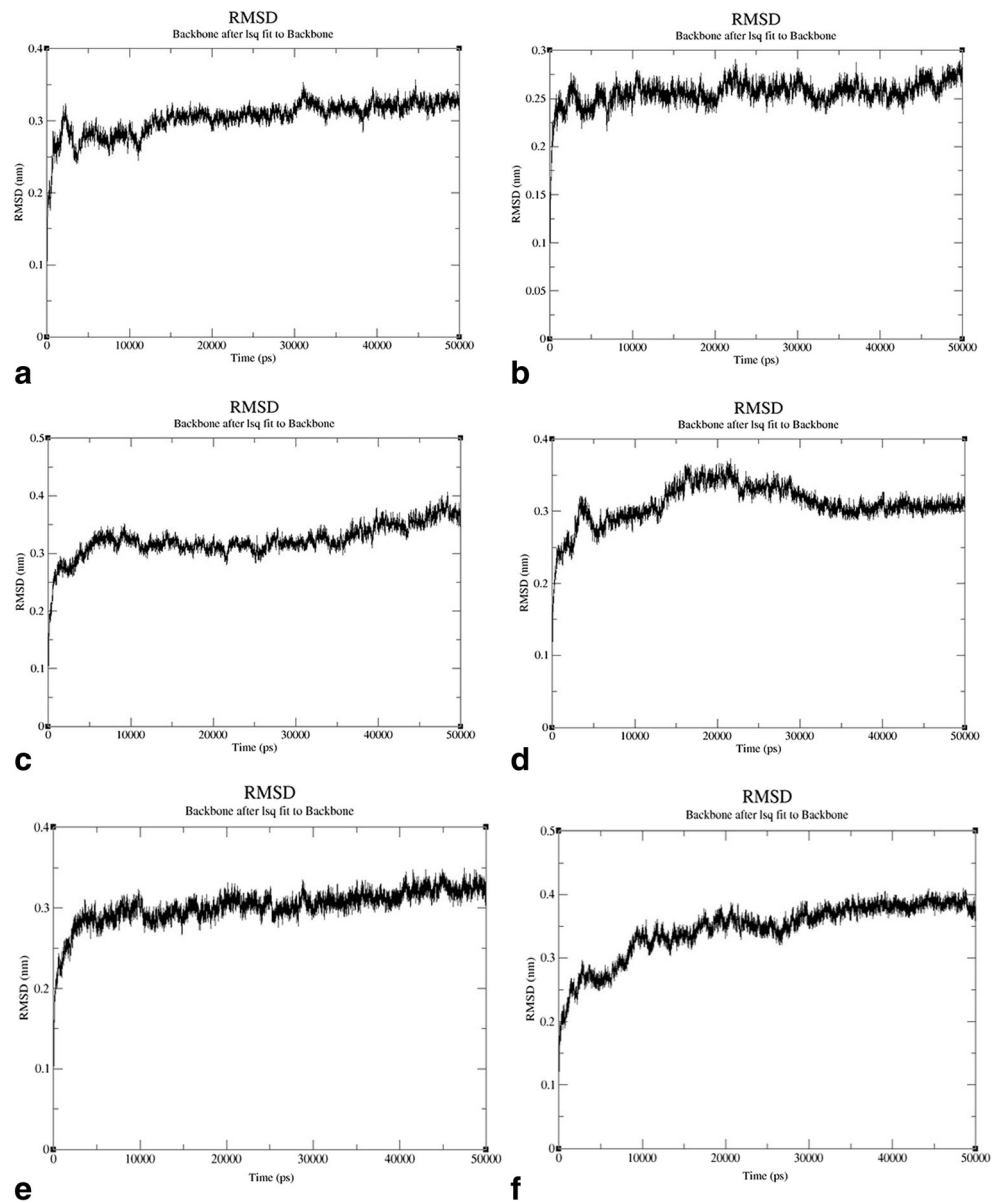
these respective temperatures. This result indicated that the N94W mutant showed higher thermostability at elevated temperatures relative to most GH enzymes. Additionally, N94F N94A retained residual CMCase activities of 94 and 91% and 91 and 90% at 60 and 62 °C, respectively, and 91 and 89% and 90 and 88%, respectively, in the presence of barley β-glucan at the respective temperatures. For E133F and E133A, the residual CMCase activities at 60 and 62 °C were 94 and 93% and 87 and 84%, respectively, and 92 and 91% and 86 and 83%, respectively, in the presence of barley β-glucan and at the respective temperatures. Results for the kinetics parameters K_m and V_m are shown in Table 2.

To calculate the temperature allowing maximal enzymatic activity (T_{max}) and the temperature at which 50% of

the maximal activity of an enzyme is retained (T_{50}), activity at a temperature range of 40 to 80 °C was assayed (Fig. 5). The N94F mutant showed similar activity as that of the wild-type enzyme, with T_{max} values in the presence of CMC (541 U/mg) and barley β-glucan (1168 U/mg) of 60.4 and 60.7 °C, respectively, and T_{50} values for CMC (270 U/mg) and barley β-glucan (584 U/Mg) of 51.6 and 50.4 °C, respectively. The N94 W variant showed T_{max} values in the presence of CMC (693 U/mg) and barley β-glucan (1517 U/mg) of 60.5 and 60.6 °C, respectively, and T_{50} values for CMC (346 U/mg) and barley β-glucan (758 U/mg) of 51 and 50.9 °C, respectively. Both enzymes lost between 2 and 6% activity following increases in reaction temperature of 2 °C, and at 66 °C, 8–43% loss of activity was observed for all enzymes.
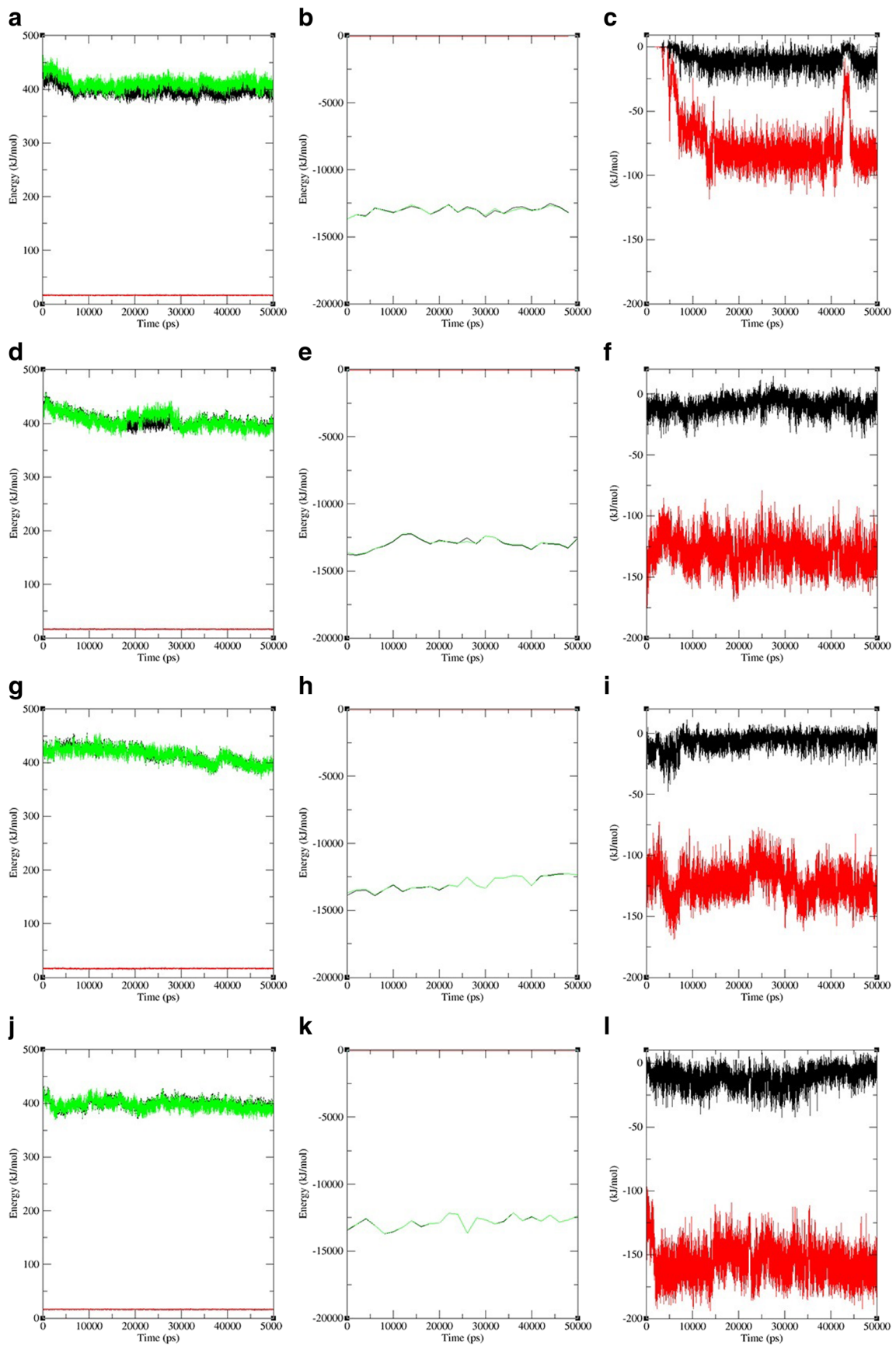
◀ **Fig. 3** Energy analysis of ligand binding in the active site of truncated structure of glycoside hydrolase GH5 from *Clostridium thermocellum*. **a** Apolar solvation energy of native protein, black: protein energy, red: ligand energy, green: protein-ligand energy. **b** Lennard-Jones and coulomb binding affinity of ligand-native model. black: coulomb energy, red: Lennard-Jones energy. **c** Apolar solvation energy of N94F mutant. **d** Lennard-Jones and coulomb binding affinity of ligand-N94F model. **e** Apolar solvation energy of N94W mutant. **f** Lennard-Jones and coulomb binding affinity of ligand-N94W model. **g** Apolar solvation energy of E133F mutant. **h** Lennard-Jones and coulomb binding affinity of ligand-E133F model. **i** Apolar solvation energy of N94A mutant. **j** Lennard-Jones and coulomb binding affinity of ligand-N94A model. **k** Apolar solvation energy of E133A mutant **l** Lennard-Jones and coulomb binding affinity of ligand-E133A model

E133F exhibited 1.16-fold higher activity relative to the wild-type enzyme and T_{max} values of 60.7 °C for both CMC (634 U/mg) and barley β-glucan (1306 U/mg) and T_{50} values of 50.1 and 50.2 °C for CMC and barley β-glucan, respectively. Additionally, the activity of the E133F variant suggested better thermal stability at 66, 68, and 70 °C relative to both the wild-type and N94F variants. The E133A variant showed T_{max} values in the presence of CMC (331 U/mg) and barley β-glucan (641 U/mg) of 60.4 and 60.1 °C, respectively, and T_{50} values for CMC and barley β-glucan of 49.7 and 50.3 °C, respectively. The N94A variant showed T_{max} values in the presence of CMC (623 U/mg) and barley β-glucan (1314 U/mg) of 60.6 and 60.8 °C, respectively, and T_{50} values for CMC and barley β-glucan of 50.4 and 50.7 °C, respectively. N94W showed a unique pattern associated with both thermal stability and catalytic activity, with CMCase activity reaching 1014 U/mg at a T_{max} value of 62.3 °C, which was 2.3 °C higher than that of the wild-type enzyme. Interestingly, the T_{max} value in the presence of barley β-glucan reached 1714 U/mg at 60.9 °C, with T_{50} values for CMC and barely β-glucan of 52.1 and 52.4 °C.

Discussion

The results of point mutations were comparable with those previously reported for protein engineering Cel12B from *T. maritima*, with results reaching 1.44- and 1.43-fold higher activity levels in the E225H and K207G variants, respectively. Moreover, the E225H/K207G double mutant and the E225H/K207G/D37V triple mutant exhibited 1.77- and 1.87-fold higher catalytic activities, respectively (Zhang et al. 2015). Interestingly, Vu and Kim (2012) engineered the endoglucanase from *B. amyloliquefaciens* by random and side-directed mutagenesis, finding a E289V variant enhanced enzyme activity by 7.93-fold (Vu and Kim 2012), which was comparable with our findings in Cel5E, where enhancing the hydrophobic core of the protein was augmented by mutating N94 and E133 to hydrophobic residues. A previous study by Miyazaki et al. (2006) on *Bacillus subtilis* family-11 xylanase reported that an N8F mutation increased hydrophobicity, thereby enhancing thermostability. However, potential steric hindrance created by the substituted residue is an important factor. Because our selected pocket for site-directed mutagenesis was located near the active site, such alterations would likely influence the active site, causing variation in specific activity. Although substitution with a hydrophobic residue in this region-enhanced enzyme thermostability, side-chain size was also a factor. Kellis et al. (Kellis et al. 1988) reported that in the ribonuclease from *B. amyloliquefaciens*, reducing the size of hydrophobic side chains significantly decreased enzyme thermostability. Therefore, we used phenylalanine and tryptophan for our substitutions in order to maximize potential hydrophobic interactions. Although interactions within the protein core in an aqueous

Table 2 Kinetic parameters and theoretical binding energy of native and mutant models calculated by MolDock and g-mmpbsa

Energy (kJ/mol)	Native	N94F	N94W	E133F	N94A	E133A
van der Waal energy	− 45.285	− 132.453	− 147.397	− 149.614	− 111.236	− 114.490
Electrostattic energy	− 5.139	− 9.108	− 15.901	− 4.525	− 6.193	− 9.543
Polar solvation energy	46.058	63.258	66.167	51.346	54.812	57.612
Apolar energy	− 60.598	− 97.846	− 103.789	− 17.244	− 38.643	− 56.889
Total binding energy by g-mmpbsa	− 64.964	− 176.148	− 200.921	− 120.038	− 101.26	− 123.76
MolDock score:	− 70.06	− 163.38	− 198.72	− 114.96	− 98.43	− 119.32
$K_{m\ (CMC)}$ mg/ml	2.3 ± 0.2	1.6 ± 0.2	1.2 ± 0.2	1.7 ± 0.2	1.9 ± 0.2	4.1 ± 0.3
$K_{cat\ (CMC)}$/min	1716 ± 72	1990 ± 77	2176 ± 84	1912 ± 76	1814 ± 70	936 ± 39
$V_{Max\ (CMC)}$ pmol min^{-1} mg^{-1}	1865 ± 78	2187 ± 84	2343 ± 90	2141 ± 85	1977 ± 76	1020 ± 42
$K_{m\ (barely\ β\ glucan)}$ mg/ml	1.6 ± 0.2	1.3 ± 0.2	0.9 ± 0.2	1.4 ± 0.2	1.5 ± 0.2	4.6 ± 0.3
$K_{cat\ (barely\ β\ glucan)}$/min	1964 ± 52	2156 ± 80	2361 ± 86	2097 ± 78	2149 ± 78	1004 ± 43
$V_{Max\ (barely\ β\ glucan)}$ pmol min^{-1} mg^{-1}	2134 ± 56	2369 ± 88	2542 ± 93	2348 ± 87	2342 ± 85	1094 ± 47

Table 3 The descriptions of mutant primers, which used for mutagenesis purpose. For mutagenesis purpose, two PCR steps has been applied

Primer name	Sequence	Nucleotides	GC content	Molecular weight	TM
E133f-Forward	CTGCTGTTCTTCAT CATGAACGAACCG	27-mer	48% GC	8210.4 Da	52
E133F-Reverse	CGTTCATGATGAAG AACAGCAGGTTT TCAG	30-mer	43% GC	9270.1 Da	52
N94F-Forward	ACCATCATCTTCTC TCACCACGAC	24-mer	50% GC	7152.7 Da	55
N94F-Reverse	GAGAAGATGATGGT AACGAAACCAC	25-mer	44% GC	7757.1 Da	55
N94 W-Forward	TACCATCATCTGGT CTCACCACGACG ACTGGA	32-mer	53% GC	9729.4 Da	56
N94 W-Reverse	GTGGTGAGACCAGA TGATGGTAACGA AACCACGAG	35-mer	51% GC	10,903.2 Da	56
N94A-Forward	TTACCATCATCGCA TCTCACCACGAC GAC	29-mer	52% GC	8726.7 Da	53
N94A-Reverse	AGATGCGATGATGG TAACGAAACCAC	26-mer	46% GC	8037.3 Da	53
E133A-Forward	CTGCTGTTCGCAAT CATGAACGAACCG	27-mer	52% GC	8244.4 Da	59
E133A-Reverse	CGTTCATGATTGCG AACAGCAGGTTT TCA	29-mer	45% GC	8907.9 Da	59

The TM that is provided in the table is the TM which is calculated in the attached with template. Because primers are mutant and do not completely bind to template, the single primer TMs are different with the TM of template-primer complex

environment is complex, it is an accepted principle that thermal stability will increase because of increasing hydrophobic interactions within the protein core (Arakawa and Tokunaga 2004; Kellis et al. Kellis et al. 1988; Privalov and Gill 1988; Radhakrishna et al. 2013).

Structural analysis of mutant enzymes

In this study, we substituted polar residues with hydrophobic residues within a hydrophobic pocket. Additionally, alanine was used as a small side chain in order to expand the volume of the active site. Replacements with hydrophobic residues increased the volume of this pocket, which might have increased internal interactions at the substitution sites. At position 94 located in the middle of the β3 strand, surrounding residues within a 4-Å radius included Asn136, Met135, Ser95, Glu133, Pro54, Val55, Ile53, Ile93, and Ile92, which are mostly hydrophobic. At position 133, the Glu residue is located within a semi-hydrophobic pocked surrounded by Asn94, Thr194, Arg52, Ser95, Ile134, Ile168, Ile92, Ile93, Phe132, Leu131, Met135, and Asn136. Interestingly, Glu137, which is a catalytic residue, is not located within 4 Å of Glu133. As shown in Fig. 6, E133 and N94 are located in close proximity. Theoretically, substitution of a polar residue with nonpolar residues would

enhance internal stability of this area. As shown in Supplemental Fig. S2, the root mean square fluctuation (RMSF) plot of the wild-type protein was 0.1778 nm around N94 and 0.1683 nm E133, which is not in equivalence with their nearby residues and suggests that they are sites of instability in the protein. The N94F and N94W mutations resulted in decreased RMSF values of 0.0619 and 0.0615 nm, revealing decreased fluctuation and increased local stability. This finding was comparable with the results of a study by Xie et al. (2014) showing that the half-life of lipase B from *Candida antarctica* increased 13-fold following D223G/L278 M substitutions, which enhanced overall enzyme rigidity by decreasing the fluctuation of active site residues (Xie et al. 2014).

Additionally, the RMSF associated with E133 was 0.1683 nm and changed to 0.058 nm in the E133F mutant. This supported our hypothesis that local stability can be enhanced by expanding the hydrophobic pocket. Moreover, the RMSF pattern observed between wild-type and mutant enzymes was similar; however, the mutant enzymes showed unique levels of variation, suggesting that the mutations affected both the local pocket and their surrounding area, resulting in a global effect on protein structure through altered rigidity and flexibility. As shown in Supplemental Fig. S2, decreased variations in flexibility

Fig. 4 The activity vs. pH diagram of native enzyme in comparison with mutant types tested in CMC (**a**) and barley β glucan (**b**). The optimum pH was detected as 7 for native enzyme as well as mutant types. The PI of the native enzyme was estimated as 5.14 while E133F mutant was 5.21. In addition, the net charge of native, N94W and N94F in the pH of 7 was estimated at − 10.8 while the E133F mutant reached − 9.8. In all of the mutant enzymes, an un-ignorable activity around 4% was observed in both strict alkaline (pH = 10) and strict acidic (pH = 4) conditions
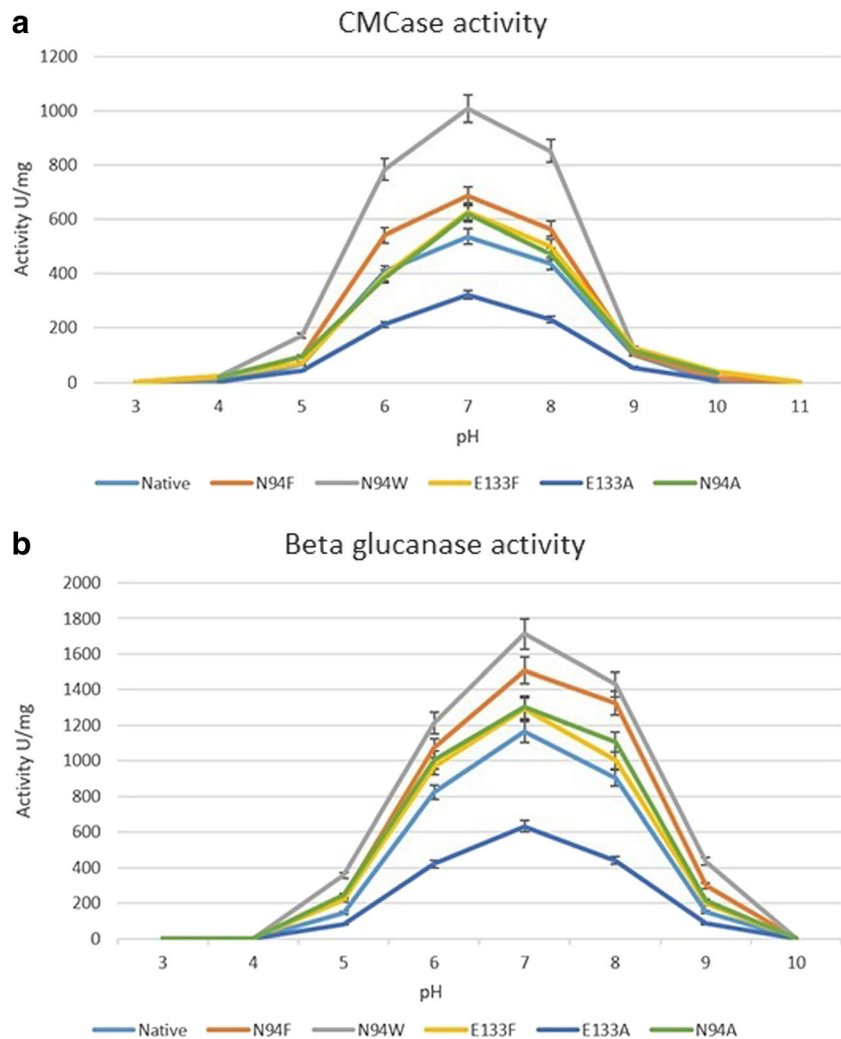
and rigidity were observed in the mutant enzymes at positions 144, 214, and 282 and was accompanied by increases in RMSF values associated with several positions. Although there are correlations between thermostability and protein rigidity, thermostability is a multifactorial phenomenon, with compactness, overall hydrophobicity, helical content, total charge, and exposed surface area key factors (Nick Pace et al. 2014).

Structure comparison of wild-type Cel5E with Cel5E mutants

Differences in the stability of enzyme homologs can occur as a result of relatively few sequence variations (Eijsink et al. 2004). Because we targeted both enzyme thermostability and activity, we incorporated point mutations in order to influence enzyme activity. Superimposition of wild-type and mutant structures was performed using the average structure extracted from the final 10 ns of the equilibrated state of the molecular dynamic simulation as the reference structure for the final functional form of the enzyme. As

shown in Supplemental Fig. S2, the overall structure did not undergo alterations; however, local changes in the hydrophobic pocket represented the primary source of altered thermostability. Moreover, improved catalytic activity would also be expected because of increased ligand-binding affinity. Interestingly, the distance between the catalytic residues E137 and E242 was 5.19 Å in the wild-type enzyme, whereas this distance in the N94F, N94W, and E133F variants was 5.17, 6.28, and 5.26 Å, respectively, which is comparable with the average distances observed in GH5 enzyme. This suggested that the mutations changed the active site cavity, but maintained the distance between key residues within a functional range. Moreover, based on our enzyme activity assays, these structural alterations to the hydrophobic pocket influenced the distance between E137 and E242 and catalytic activity. The size of the active site cavity in the wild-type enzyme was 65.7 Å, whereas that of the N94F, N94W, and E133F variants was 69.4, 72.3, and 64.1 Å, respectively. For E133A, the distance between these catalytic residues decreased to 4.75 Å in E133A mutant relative to the wild-type enzyme, and the local RMSD value

Fig. 5 Activity vs. temperature diagram of native enzyme in comparison with mutant types indicated that all of the mutant enzymes could reach higher activity levels in the optimum temperature of 60 °C in both CMC (**a**) and barley β glucan (**b**). N94W has indicated a unique thermal stability pattern and could reach its optimum CMCase activity at 62 °C. Although E133F could not reach higher activity levels than N94F in the optimum temperature, its stability pattern indicated higher levels of activity in 64–70 °C
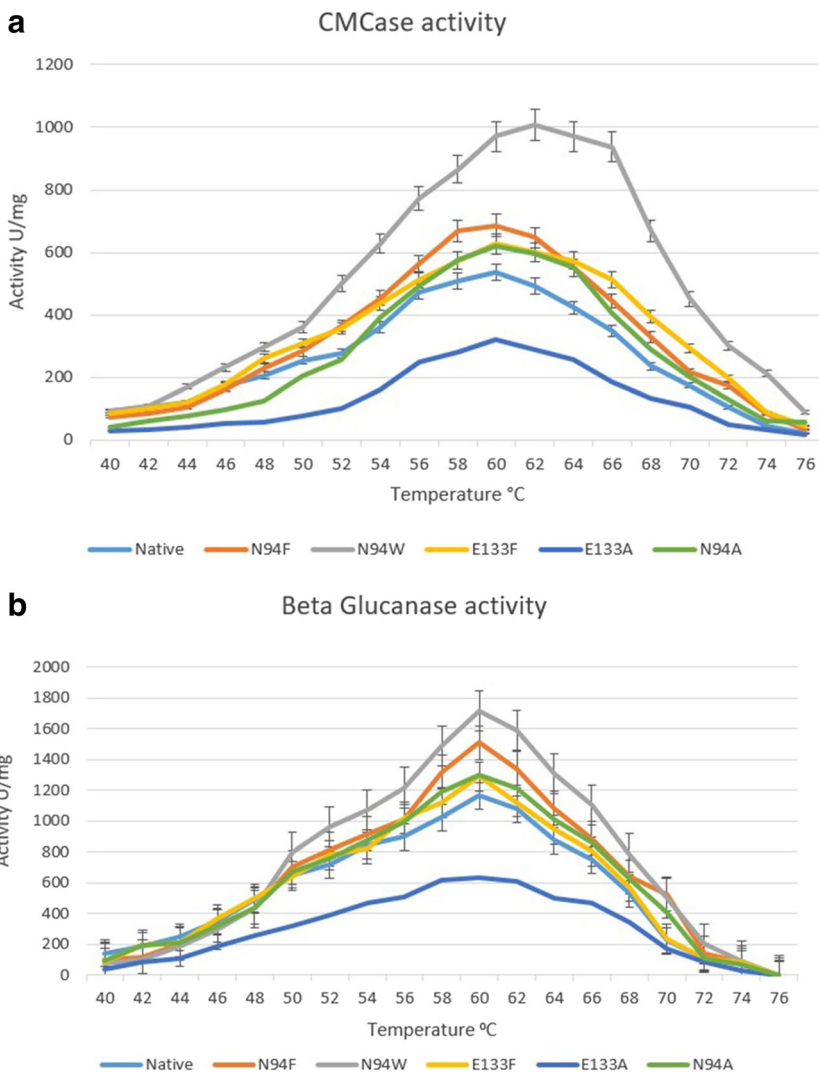
(residue 125–140) differed from that of the global value (0.3994 vs. 0.16238 Å). Therefore, it is probable that structural alteration of the active site explained the decreased catalytic efficiency of the E133A.

In summary, we generated three mutant and truncated forms of Cel5E that showed enhanced CMCase levels in acidic and alkaline pH conditions. Molecular dynamics simulations revealed that the mutant enzymes harbored

Fig. 6 Position of N94 and E133 in a hydrophobic pocket in the middle of glycoside hydrolase GH5 from *Clostridium thermocellum*. By substituting polar amino acids with nonpolar, the hydrophobic pocket was expanded and lead to higher levels of thermal stability
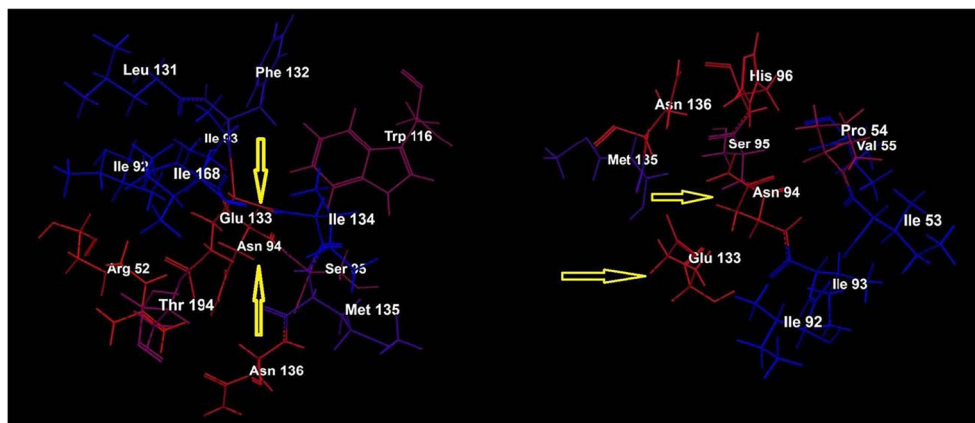
an expanded hydrophobic pocket, which enhanced overall enzyme thermostability. Furthermore, this structural alteration improved the catalytic activity of most of the mutants. These results described three novel engineered enzymes available for commercial applications.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval This study does not contain any experiments involving human participants or animals performed by any of the authors.

References

Arakawa T, Tokunaga M (2004) Electrostatic and hydrophobic interactions play a major role in the stability and refolding of halophilic proteins. Protein Pept Lett 11(2):125–132

Badieyan S, Bevan DR, Zhang C (2012) Study and design of stability in GH5 cellulases. Biotechnol Bioeng 109(1):31–44. https://doi.org/10.1002/bit.23280

Benkert P, Tosatto SC, Schomburg D (2008) QMEAN: a comprehensive scoring function for model quality assessment. Proteins 71(1):261–277. https://doi.org/10.1002/prot.21715

Bergquist P, Te'o V, Gibbs M, Cziferszky A, de Faria FP, Azevedo M, Nevalainen H (2002) Expression of xylanase enzymes from thermophilic microorganisms in fungal hosts. Extremophiles 6(3):177–184. https://doi.org/10.1007/s00792-001-0252-5

Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 33(Web Server issue):W306–W310. https://doi.org/10.1093/nar/gki375

Chen CW, Lin J, Chu YW (2013) iStable: off-the-shelf predictor integration for predicting protein stability changes. BMC Bioinformatics 14(Suppl 2):S5. https://doi.org/10.1186/1471-2105-14-S2-S5

Cheng J, Randall A, Baldi P (2006) Prediction of protein stability changes for single-site mutations using support vector machines. Proteins 62(4):1125–1132. https://doi.org/10.1002/prot.20810

Dallakyan S, Olson AJ (2015) Small-molecule library screening by docking with PyRx. Methods Mol Biol 1263:243–250. https://doi.org/10.1007/978-1-4939-2269-7_19

Dehouck Y, Kwasigroch JM, Gilis D, Rooman M (2011) PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. BMC Bioinformatics 12:151. https://doi.org/10.1186/1471-2105-12-151

Dosztanyi Z, Magyar C, Tusnady G, Simon I (2003) SCide: identification of stabilization centers in proteins. Bioinformatics 19(7):899–900

Eijsink VG, Bjork A, Gaseidnes S, Sirevag R, Synstad B, van den Burg B, Vriend G (2004) Rational engineering of enzyme stability. J Biotechnol 113(1–3):105–120. https://doi.org/10.1016/j.jbiotec.2004.03.026

Escovar-Kousen JM, Wilson D, Irwin D (2004) Integration of computer modeling and initial studies of site-directed mutagenesis to improve cellulase activity on Cel9A from *Thermobifida fusca*. Appl Biochem Biotechnol 113(1-3):287–298

Fiser A, Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol 374:461–491. https://doi.org/10.1016/S0076-6879(03)74020-8

Gusakov AV, Kondratyeva EG, Sinitsyn AP (2011) Comparison of two methods for assaying reducing sugars in the determination of carbohydrase activities. Int J Anal Chem 2011:283658. https://doi.org/10.1155/2011/283658

Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4(3):435–447. https://doi.org/10.1021/ct700301q

Hirano K, Kurosaki M, Nihei S, Hasegawa H, Shinoda S, Haruki M, Hirano N (2016) Enzymatic diversity of the *Clostridium thermocellum* cellulosome is crucial for the degradation of crystalline cellulose and plant biomass. Sci Rep 6:35709. https://doi.org/10.1038/srep35709

Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77(1):51–59

Ichikawa S, Yoshida M, Karita S, Kondo M, Goto M (2015) Carbohydrate-binding modules influence substrate specificity of an endoglucanase from *Clostridium thermocellum*. Biosci Biotechnol Biochem 80(1):188–192. https://doi.org/10.1080/09168451.2015.1069696

Kellis JT Jr, Nyberg K, Sali D, Fersht AR (1988) Contribution of hydrophobic interactions to protein stability. Nature 333(6175):784–786. https://doi.org/10.1038/333784a0

Korb O, Stutzle T, Exner TE (2009) Empirical scoring functions for advanced protein-ligand docking with PLANTS. J Chem Inf Model 49(1):84–96. https://doi.org/10.1021/ci800298z

Kumari R, Kumar R, Open Source Drug Discovery C, Lynn A (2014) g_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model 54(7):1951–1962. https://doi.org/10.1021/ci500020m

Mahadevan SA, Wi SG, Lee DS, Bae HJ (2008) Site-directed mutagenesis and CBM engineering of Cel5A (*Thermotoga maritima*). FEMS Microbiol Lett 287(2):205–211. https://doi.org/10.1111/j.1574-6968.2008.01324.x

Mehboob N, Asad MJ, Asgher M, Gulfraz M, Mukhtar T, Mahmood RT (2014) Exploring thermophilic cellulolytic enzyme production potential of *Aspergillus fumigatus* by the solid-state fermentation of wheat straw. Appl Biochem Biotechnol 172(7):3646–3655. https://doi.org/10.1007/s12010-014-0796-3

Miyazaki K, Takenouchi M, Kondo H, Noro N, Suzuki M, Tsuda S (2006) Thermal stabilization of family-11 xylanase by directed evolution. J Biol Chem 281(15):10236–10242

Moretti MM, Bocchini-Martins DA, Silva RD, Rodrigues A, Sette LD, Gomes E (2012) Selection of thermophilic and thermotolerant fungi for the production of cellulases and xylanases under solid-state fermentation. Braz J Microbiol 43(3):1062–1071. https://doi.org/10.1590/S1517-838220120003000032

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30(16):2785–2791. https://doi.org/10.1002/jcc.21256

Nick Pace C, Scholtz JM, Grimsley GR (2014) Forces stabilizing proteins. FEBS Lett 588(14):2177–2184. https://doi.org/10.1016/j.febslet.2014.05.006

Peacock JP, Cole JK, Murugapiran SK, Dodsworth JA, Fisher JC, Moser DP, Hedlund BP (2013) Pyrosequencing reveals high-temperature cellulolytic microbial consortia in great boiling spring after in situ lignocellulose enrichment. PLoS One 8(3):e59927. https://doi.org/10.1371/journal.pone.0059927

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612. https://doi.org/10.1002/jcc.20084

Pikkemaat MG, Linssen AB, Berendsen HJ, Janssen DB (2002) Molecular dynamics simulations as a tool for improving protein stability. Protein Eng 15(3):185–192

Privalov PL, Gill SJ (1988) Stability of protein structure and hydrophobic interaction. Adv Protein Chem 39:191–234

Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29(7):845–854. https://doi.org/10.1093/bioinformatics/btt055

Puigbo P, Guzman E, Romeu A, Garcia-Vallve S (2007) OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res 35(Web Server issue):W126–W131. https://doi.org/10.1093/nar/gkm219

Radhakrishna M, Grimaldi J, Belfort G, Kumar SK (2013) Stability of proteins inside a hydrophobic cavity. Langmuir 29(28):8922–8928. https://doi.org/10.1021/la4014784

Schuttelkopf AW, van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr 60(Pt 8):1355–1363. https://doi.org/10.1107/S0907444904011679

Tan KP, Nguyen TB, Patel S, Varadarajan R, Madhusudhan MS (2013) Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. Nucleic Acids Res 41(Web Server issue): W314–W321. https://doi.org/10.1093/nar/gkt503

Thomsen R, Christensen MH (2006) MolDock: a new technique for high-accuracy molecular docking. J Med Chem 49(11):3315–3321. https://doi.org/10.1021/jm051197e

Vu VH, Kim K (2012) Improvement of cellulase activity using error-prone rolling circle amplification and site-directed mutagenesis. J Microbiol Biotechnol 22(5):607–613

Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35(Web Server issue):W407–W410. https://doi.org/10.1093/nar/gkm290

Wu TH, Huang CH, Ko TP, Lai HL, Ma Y, Chen CC, Cheng YS, Liu JR, Guo RT (2011) Diverse substrate recognition mechanism revealed by *Thermotoga maritima* Cel5A structures in complex with cellotetraose, cellobiose and mannotriose. Biochim Biophys Acta 1814(12):1832–1840. https://doi.org/10.1016/j.bbapap.2011.07.020

Xie Y, An J, Yang G, Wu G, Zhang Y, Cui L, Feng Y (2014) Enhanced enzyme kinetic stability by increasing rigidity within the active site. J Biol Chem 289(11):7994–8006. https://doi.org/10.1074/jbc.M113.536045

Yague E, Beguin P, Aubert JP (1990) Nucleotide sequence and deletion analysis of the cellulase-encoding gene celH of *Clostridium thermocellum*. Gene 89(1):61–67

Yuan SF, Wu TH, Lee HL, Hsieh HY, Lin WL, Yang B, Chang CK, Li Q, Gao J, Huang CH, Ho MC, Guo RT, Liang PH (2015) Biochemical characterization and structural analysis of a bifunctional cellulase/xylanase from *Clostridium thermocellum*. J Biol Chem 290(9): 5739–5748. https://doi.org/10.1074/jbc.M114.604454

Zhang J, Shi H, Xu L, Zhu X, Li X (2015) Site-directed mutagenesis of a hyperthermophilic endoglucanase Cel12B from *Thermotoga maritima* based on rational design. PLoS One 10(7):e0133824. https://doi.org/10.1371/journal.pone.0133824