Int. J. Mol. Sci. **2006**, *7*, 571-589

International Journal of
Molecular Sciences
ISSN 1422-0067

www.mdpi.org/ijms/

Identification and molecular modeling of a family 5 endocellulase from *Thermus caldophilus* GK24, a cellulolytic strain of *Thermus thermophilus*

Dooil Kim [1,2,*], Bo Hyun Park [1], Bo-Won Jung [1], Mi-kyung Kim [1], Suk-In Hong [2,3] and Dae-Sil Lee [1,*]

1 Systems Microbiology Research Center, KRIBB, Daejeon 305-806, Korea

2 Department of Biomicrosystem Technology, Korea University, Seoul 136-701, Korea

3 Department of Chemical and Biological Engineering, Korea University, Seoul 136-701, Korea

* Author to whom correspondence should be addressed. Tel: +82-42-860-4135; Fax: +82-42-860-4597; E-mail address: dikim@kribb.re.kr

Received: 20 November 2006 / Accepted: 14 December 2006 / Published: 18 December 2006

Abstract: The genome of *T. caldophilus* GK24 was recently sequenced and annotated as 14 contigs, equivalent to 2.3 mega basepairs (Mbp) of DNA. In the current study, we identified a unique 13.7 kbp DNA sequence, which included the endocellulase gene of *T. caldophilus* GK24, which did not appear to be present in the complete genomic sequence of the closely related species *T. thermophilus* HB27 and HB8. Congo-red staining revealed a unique phenotype of cellulose degradation by strain GK24 that was distinct from other closely related *Thermus* strains. The results showed that strain GK24 is an aerobic, thermophilic, cellulolytic eubacterium which belongs to the group *T. thermophilus*. In order to understand the mechanism of production of cellobiose in *T. caldophilus* GK24, a three-dimensional model of the endocellulase, *Tc*Cel5A, was generated based on known crystal structures. Using this model, we carried out a flexible cellotetraose docking study.

Keywords: Endocellulase, *Thermus caldophilus* GK24, Glycosyl hydrolase family 5, Homology modeling, Molecular dynamics, Congo-red staining

1. Introduction

Members of the genus *Thermus* are hyperthermophilic, non-sporulating, Gram-negative bacteria that grow aerobically at an optimum growth temperature of 70°C-75°C [1]. Phylogenetic analysis revealed that *Thermus* species belong to the Deinococcus-Thermus phylum [2]. There are several validated *Thermus* species, including *T. aquaticus* (the prototype species) [3], *T. thermophilus*, *T. filiformis*, *T. scotoductus*, *T. oshimai*, *T. antranikianus*, *T. igniterrae*, and *T. brockianus*. However, there are many *Thermus* species that have been isolated and not taxonomically characterized. Among these is the species *T. caldophilus* GK24, which was originally isolated from the Kawamata hot spring, Tochigi-ken, Japan [4]. Strain GK24, along with another Japanese isolate, *T. flavus* AT62, was previously grouped as a subspecies of *T. thermophilus* by DNA-DNA homology [5-8].

Cellulose is an unbranched glucose polymer composed of D-glucose units linked by a 1,4-β-D-glucosidic bond. It is degraded by enzymes produced by both fungi and bacteria. As cellulose is a very stable polymer, effective hydrolysis of it requires the synergistic action of several enzymes, including endo-β-1,4-glucanases, exo-β-1,4-glucanases (or cellobiohydrolase) and β-glucosidases [9-22]. The mechanism of cellulose hydrolysis by cellulases has been studied extensively [23-26].

This general classification is still widely accepted, although several observations suggest that it is an oversimplified description of a complex set of interactions by cellulases. For example, not all endoglucanases appear to be able to act synergistically with exoglucanases. Recently, it was suggested that the concepts of processive and non-processive hydrolysis of cellulose are necessary when discussing the action of cellulases [27]. Processive cellulases hydrolyze cellulose chains by continuous removal of cellulobiose units while remaining bound to the substrate. The non-processive cellulases randomly hydrolyze cellulose chains once, then dissociate from the substrate, and recombine for another cycle of hydrolysis, resulting in a rapid decrease in polymer length. The classification of processive and non-processive cellulases is based on whether the enzyme engages in a continuous cycle of hydrolysis without disassociating from the substrate after each cycle, or dissociates from the substrate and recombines for each cycle of hydrolysis. On the other hand, the classification of endo- and exo-cellulases is based on the position of the cellulase binding site, either internally, or at one of the substrate ends, respectively.

All glycosyl hydrolases catalyze stereoslective hydrolysis, meaning that the configuration about the anomeric centre is either inverted or retained upon hydrolysis of the glycosidic linkage. Details of both reaction mechanisms are described by Tomme *et al*. [12]. Retaining enzymes, but not inverting enzymes, also have transglycosylation activity [28]. The catalytic amino acid residues in both types of enzymes can be tentatively identified by chemical modification or site-directed mutagenesis, in conjunction with three-dimensional (3D) structural analysis. X-ray crystallographic analysis of more than 20 celluases and xylanases, representing several families of glycosyl hydrolases (families 5, 6, 7, 8, 9, 10, 11, 12, 45, and 48), have been reported or actively studied [29].

In fact, most thermostable cellulases reported to date have been isolated from thermophilic microorganisms, such as *Acidothermus cellulolyticus* [30-32], *Clostridium thermocellum* [33-34], *Thermomonospora fusca* [35], *Streptomyces albaduncus* [36], *Thermotoga neapolitana* [37].

In this report, we identified a new strain of thermophilic cellulolytic microorganism, *Thermus caldophilus* GK24, as a source of a novel thermostable endocellulase. We describe modeling of the

Thermus caldophilus GK24 endocellulase *Tc*Cel5A, which belongs to Family 5, in complex with cellotetraose (G4). The results provide insight into the structural basis for complex formation between *Tc*CelA and G4, and provide a better understanding of the molecular basis of the interaction between *Tc*CelA and G4.

2. Materials and Methods

2.1. Sequencing of a whole genome shot gun library of Thermus calodphilus GK24

A random shotgun library was constructed using pGEMT-EASY (Promega) vectors. Genomic DNA was sonicated or fragmented by a Nebulizer (cat. 7025-05 Invitrogen) to generate genomic fragments of 4~5kb. *E. coli* were transformed with the *T. caldophilus* GK24 genomic library, then individual colonies were isolated using Q-Pix (Genetix) and cultured. Solid phase extraction kits optimized for DNA were used to prepare plasmid DNA. Sequencing reactions were carried out using a Big Dye Sequencing kit protocol that came with the ABI 3700 Genetic Analyzer, according to the manufacturer's instructions. Gap closing combinatorial PCR using primer walking was performed to read the complete genomic DNA sequence [38].

2.2. Data Processing, Sequence Assembly, and Annotation

Initially, the raw sequence data of the shotgun library was assembled using the Phred/Phrap/Consed packages [39-40]. Base-calling was done using the Phred program. The sequence data, together with the base-calling quality information for each base, were then used to assemble the whole genome using the Phrap program. The resulting contigs were further visualized and examined using the Consed program. Gap-closing was then performed using manual editing and the Autofinish function incorporated into the Consed package [41-42]. Identification of ORF's, BLAST searches, and annotation were performed using an in-house bioinformatics server installed with the appropriate software. We compared *T. caldophilus* GK24 [43] with the complete *T. thermophilus* HB27 [44] and HB8 [45] genomes using the MUMer 3.0 package [46] and perl-based in-house analysis tools.

2.3. Detection of endocellulase activity

Six *Thermus* strains were screened for possible endocellulase activity on medium agar plates containing 1% CM-cellulose using the Congo red staining method [47]. Endocellulase activity was observed as the formation of a clear halo around a bacterial colony. The *Thermus* strains used in this study were *T. aquaticus* YT1 (ATCC 25104), *T. filiformis* Wai33 A1 (ATCC 43280), *T. thermophilus* HB8 (ATCC 27634), *T. thermophilus* HB27 (ATCC BAA-163), *T. flavus* AT62 (ATCC 33923), and *T. caldophilus* GK24. Cells were grown aerobically at 70°C with shaking in Castenholz medium (ATCC medium 461) and plated on the same medium containing 2% agar (Difco). Where indicated, media was supplemented with 0.1% carboxymethyl-cellulose (CMC). The agar medium was immersed in 0.5% (w/v) Congo red for 15 min and destained with 1 M NaCl for 15 min. Active bands appeared as yellow halos on a red background. Samples were treated with 0.1 N HCl to turn the background dark blue, and photographed.

2.4. Purification of the enzyme

The culture supernatant from 3,000 ml culture of *T. caldophilus* GK24 was added with ammonium sulfate at 80% saturation. The precipitate was collected by centrifugation (13,000×g at 4ºC for 30 min), dissolved in a small volume of 20 mM acetate buffer, and then added with ammonium sulfate at 20% saturation. This enzyme solution was applied onto a column of Butyl-Toyopearl 650M (4.5 × 6.4 cm) (Tosoh Co.) equilibrated with 20 mM acetate buffer containing ammonium sulfate (20% saturation). The column was washed with 1,000 ml of the same buffer, and the enzyme was eluted with a linear gradient of ammonium sulfate from 20% to 0% (total volume of 1,000 ml). The active fractions were collected, and the protein was precipitated with ammonium sulfate at 80% saturation. The precipitate was collected by centrifugation, dissolved in a small volume of 20 mM acetate buffer, and then desalted by Bio-gel P-4 (2.5×20 cm) (Bio-Rad Laboratories Japan) equilibrated with 20 mM acetate buffer. The desalted enzyme solution was applied onto a column of DEAE-Toyopearl 650M (4.5×3.2 cm) equilibrated with 20 mM acetate buffer. The column was washed with 500 ml of the same buffer, and the enzyme was eluted with a linear gradient of NaCl from 0 to 0.5 M. The active fractions were collected, and the protein was precipitated with ammonium sulfate at 80% saturation. The precipitate was collected by centrifugation, dissolved in a small volume of 20 mM acetate buffer, and then applied onto a Superdex 75pg 16/60 (Pharmacia LKB Biotechnology) with equilibrated with 20 mM acetate buffer (pH 5.5) containing 150 mM NaCl. Pharmacia FPLC system (PUMP: P-500; LIQUID CHROMATOGRAPHY CONTROL: LCC-500; SINGLE PATH UV-MONITOR: UV-1; FRACTION COLLECTOR: FRAC-100, Pharmacia LKB Biotechnology) was used. The enzyme was eluted with the same buffer. The active fractions were collected, and the protein was precipitated with ammonium sulfate at 80% saturation. The precipitate was collected by centrifugation, dissolved in a small volume of 5 mM acetate buffer, and then desalted by Bio-gel P-4 (1.5×20 cm) equilibrated with 5 mM acetate buffer.

2.5. Activities of enzyme against various substrates

The following substrates were used: Alkali swollen cellulose (ASC), carboxymethyl-cellulose (CMC, degree of average polymerization, 500; Wako Pure Chemical), insoluble cellooligosaccharide (ICOS), cellotriose, cellotetraose, cellopentaose and cellohexaose. Concentration of substrates was adjusted to 0.5% in 100 mM acetate buffer (pH 7.0). The reaction mixture consisting of 0.1 ml of substrate solution and 0.1 ml of enzyme solution was incubated at 70ºC for 30 min and the enzyme activity was measured by Somogyi-Nelson method. One unit of enzyme activity was defined as the amount of enzyme that released 1 μmol reducing sugar as glucose per min. Cellotriose, cellotetraose, cellopentaose and cellohexaose were prepared by the method described by Miller [56].

2.6. Measurement of the optimum temperature

The optimum temperature for the enzyme activity was determined as follows; the enzyme was added to 0.5% cellohexaose in 100 mM acetate buffer (pH 7.0). After a 30 min reaction at various temperatures, each activity was measured.

2.7. Homology modeling of T. caldophilus GK24 Endocellulase

Homology modeling is the method of choice when a clear relationship based on sequence homology between the target protein and at least one other known structure exists. This approach is based on the assumption that the tertiary structures of two proteins will be similar if their amino acid sequences are related [48]. Homologous sequence searching and alignment were carried out using FASTA [49] and the ClustalW program [50]. Sequence identity between *Tc*Cel5A and the reference protein *Ac*Cel5A, from *Acidothermus cellulolyticus* (PDB code **1ECE**) was 35%. When modeling *Tc*Cel5A, some residues at the N- and C-termini of the original sequence were removed, since there were no good templates for these fragments, and these residues were far away from the catalytic domain. For modeling the 3D structure of *Tc*Cel5A, the program MODELER [51-53] was used. MODELER is an implementation of an automated approach to comparative modeling based on satisfaction of spatial restraints [52-54]. The initial structure was revised by refining loops and rotamers, checking bonds, and adding hydrogen atoms, then MM and MD simulations with the aid of the CVFF force field were used to optimize the initial structure. For energy minimization, 20,000 iterations of the steepest descent calculation were performed, then the conjugated gradient calculation was carried out until convergence at 0.005 $kcalmol^{-1}Å^{-1}$. After the above simulations, the final model was obtained by carrying out MD simulations using the Discover 3 software package [55] in Insight II software package developed by Biosym Technologies. The explicit solvent model for water, TIP3P, was also used. The final homology model of *Tc*Cel5A was solvent with a 10 Å water cap from the center of mass. Finally, a conjugate gradient energy minimization of the full protein was performed until the root mean-square (RMS) energy gradient was lower than 0.001 $kcalmol^{-1}Å^{-1}$. After this step, the quality of the initial model improved.

2.8. Identification of the binding site of TcCel5A

The Binding-Site module is a suite of programs in Insight II for identifying and characterizing protein binding sites and functional residues in proteins. We used the ACTIVESITE-SEARCH program [55] to search for protein binding sites by locating cavities in the *Tc*Cel5A structure, which were then used to guide protein-ligand docking experiments. The predicted binding site of *Tc*Cel5A was identified by comparing the conserved residues in endocellulases from *A. cellulolyticus* [32] and *C. thermocellum* [34] and combining the results.

2.9. Molecular modeling of cellotetraose docking into theTcCel5A binding site

During molecular docking, molecules fit together in a favorable configuration to form a complex system. The 3D structure of cellotetraose (G4) was generated using the BUILDER program [55], and the geometry was optimized. For the purpose of capturing the interaction of *Tc*Cel5A with G4, the advanced docking program Affinity [55] was used to perform automated molecular docking. A combination of Monte Carlo type and stimulated annealing procedures for docking a guest molecule with a host were employed to find the optimal structures of the *Tc*Cel5A-G4 interaction, based on the energy of the *Tc*Cel5A-G4 complex. A key feature was that the bulk of *Tc*Cel5A, defined as atoms not in the specified binding site, were held rigid during the docking process, while the binding site atoms

and G4 atoms were allowed to move. The docked complexes of *Tc*Cel5A with G4 were further refined according to the criteria of interacting energy combined with geometrical matching quality. These complexes were then used as the starting conformation for energetic minimization and geometrical optimization to finalize the models of the *Tc*Cel5A-G4 docking interaction.

3. Results and discussion

3.1. Endocellulase activity of T. caldophilus GK24

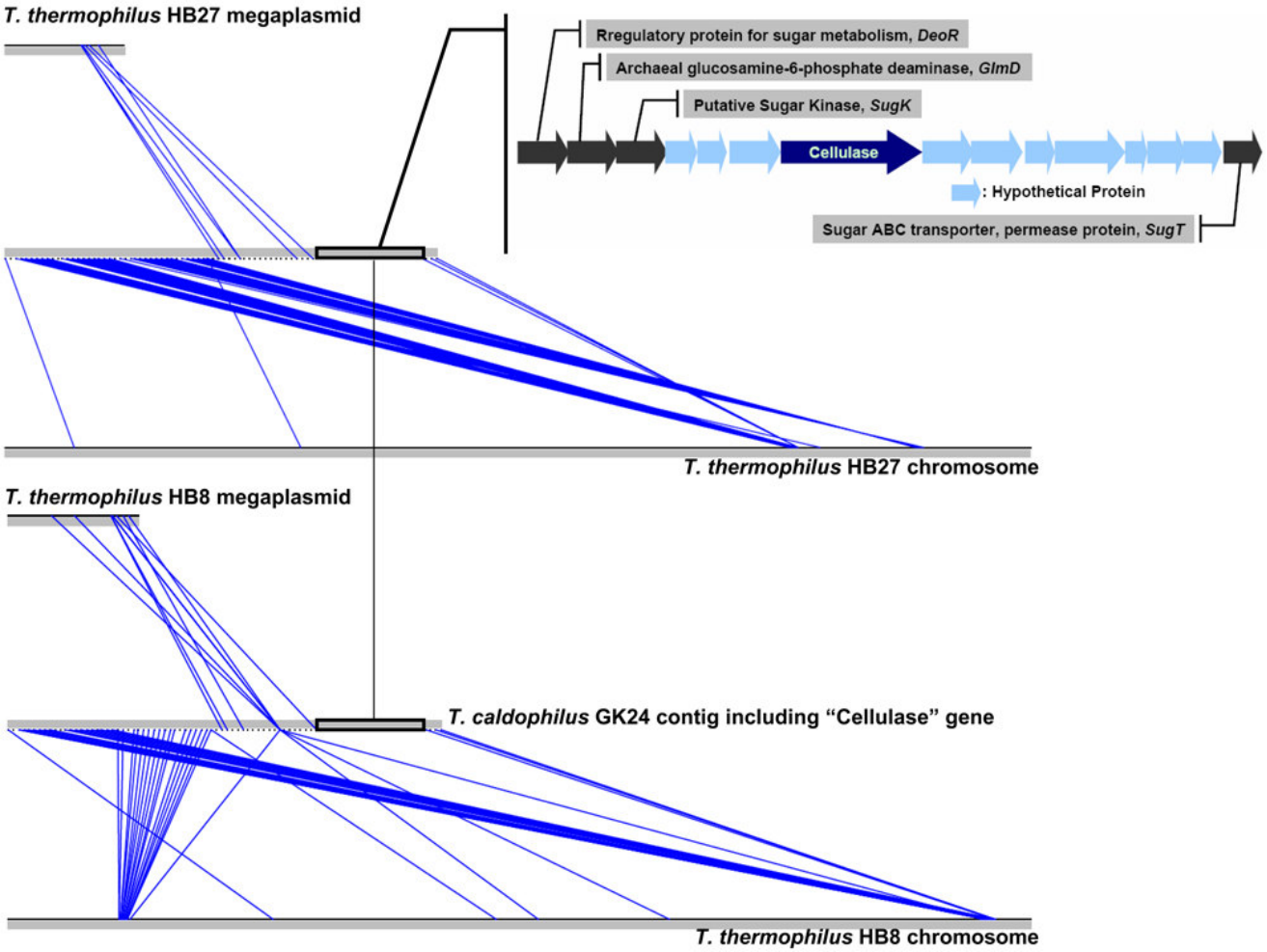

Figure 1. Linear comparison of the genomes of (A) *T. thermophilus* HB27 and *T. caldophilus* GK24 and (B) *T. thermophilus* HB8 and *T. caldophilus* GK24. The blue lines represent similar protein sequences (BLASTP search, > 60% similarity) between organisms. Gene organization in the 13.7 kbp region, including *Tc*Cel5A (blue arrow), of the *T. caldophilus* GK24 genome. Arrows indicate open reading frames, and their predicted translation products are indicated above the arrows. Black arrows indicate open reading frames of previously characterized genes (*deoR*, *glmD*, *sugK*, and *sugT*). Blue arrows indicate uncharacterized open reading frames found in the genome of *T. caldophilus* GK24.

We identifed a unique cellulolytic activity, and the putative endocellulase gene, in *T. caldophilus* GK24. Sequence analysis of 14 contigs, comprising the ~ 2.3 Mbps of the *T. caldophilus* GK24 genome, revealed a 13.7 kbp DNA region that was absent in the complete genome sequences of the closely related species *T. thermophilus* HB27 and HB8 [44-45]. Within this 13.7 kbp DNA region

(Figure 1), we identified 15 potential open reading frames (ORFs) which appeared to encode enzymes involved in carbohydrate metabolism. Four of the ORFs were predicted to encode a regulatory protein for sugar metabolism (*DeoR*), glucosamine-6-phosphate deaminase (*GlmD*), a putative sugar kinase (*SugK*), and a sugar ABC transporter, permease protein (*SugT*). One ORF appeared to encode a β-1,4-glucanase with homology to the Family 5 glycosyl hydrolases. Of particular interest, one of the putative ORFs was predicted to encode a β-glucanase homologue. Based on sequence data, we examined the degradation of cellulosic material by *T. caldophilus* GK24 using Congo-red staining. As seen in Figure 2, strain GK24 had a unique phenotype of cellulose degradation that was distinct from other closely related *Thermus* strains (*T. thermophilus* HB27/HB8, *T. flavus* AT62, *T. aquaticus* YT1, and *T. filiformis*) (Figure 2). Thus, we were able to identify a unique phenotype of cellulolytic degradation by *T. caldophilus* GK24 that correlated with its unique genotype.

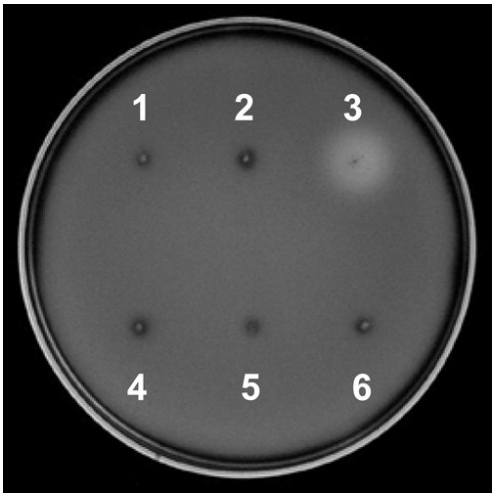

Figure 2. Detection of endocellulase activity in several *Thermus* strains. Cells were incubated on 0.1% CMC agar plates at 70°C for 40 h. 1, *T. thermophilus* HB27; 2, *T. thermophilus* HB8; 3, *T. caldophilus* GK24; 4, *T. filiformis*; 5, *T. aquaticus* YT1; 6, *T. flavus* AT62.

3.2. Purification and Activity of TcCel5A against various substrates

The culture supernatant of *T. caldophilus* GK24 was put onto a Butyl-Toyopearl 650M (4.5 × 6.4 cm). Enzyme fraction was pooled and further purified by DEAE-Toyopearl 650M and Superdex 75 pg 16/60. The molecular weight of *Tc*Cel5A was estimated to be about 44 kDa by SDS-PAGE (Figure 3A).

Table 1. Activity of *Tc*Cel5A against various substrates.

Substrates	*Tc*Cel5A (units/mg)
Carboxymethyl-cellulose (CMC)	23.2
Alkali swollen cellulose (ASC)	9.16
Insoluble cellooligosaccharide (ICOS)	8.21
Cellotriose	N.D.[a]
Cellotetraose	21.9
Cellopentaose	16.5
Cellohexaose	7.84

[a]N.D.: not detected

*Tc*Cel5A showed the highest activity against cellotetraose. The activities of *Tc*Cel5A against various substrates are summarized in Table 1. The degradation of ASC by *Tc*Cel5A was analyzed by thin-layer chromatography. *Tc*Cel5A produced cellobiose and cellotriose (Figure 3B).

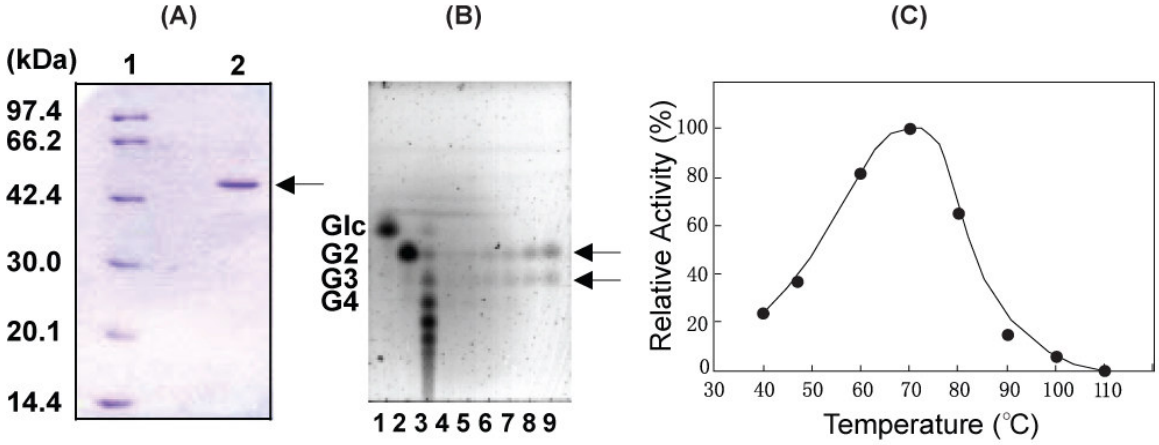

Figure 3. (A) SDS-PAGE analysis of purified *Tc*Cel5A. Purified enzymes were electrophoresed on a 12.5% SDS-PAGE. Lane 1, standard marker proteins; lane 2, purified enzymes. (B) Analysis of degradation of ASC by *Tc*Cel5A. The reaction mixture consisting of 0.1 ml of 0.5% ASC in 100 mM acetate buffer (pH 7.0) and 0.1 ml of enzyme was incubated at 70ºC. After incubation for 10 min, 30 min and 18 hours, the reaction mixture was analyzed by thin-layer chromatography. Lane 1: Glucose (standard), lane 2: Cellobiose (standard), lane 3: Cellooligosaccharide (standard), lane 4 and 5: 10 min reaction mixture, lane 6 and 7: 30 min reaction mixture, lane 8 and 9: 18 hours reaction mixture. (C) Effect of temperature on the enzyme activity of *Tc*Cel5A.

3.3. Homology modeling of TcCel5A

Identification and alignment of eight endocellulases with homology to *Tc*Cel5A were done using FASTA [49], and ClustalW [50] (Figure 4). When the sequence of *Tc*Cel5A was compared to all known proteins in the Protein Data Bank (PDB), we found that *Ac*Cel5A had the highest sequence identity (35%) with *Tc*Cel5A, so *Ac*Cel5A was used to model the 3D structure of *Tc*Cel5A. All of the amino acid side chains of *Tc*Cel5A were set using the AUTO_ROTAMER program [55], which uses the library proposed by Ponder and Richards [57]. In order to eliminate steric contacts and achieve a stable conformation, energy minimizations of 20,000 iterations and dynamics simulations of 150 ps were performed. The variation of potential energy with time during the 150 ps of molecular dynamics simulation is plotted in Figure 5(A). The potential energy fell rapidly in the first 20 ps, then decreased after that with very low deviation between successive steps. The dynamics simulations tended to equilibrium at 150 ps, thus, we chose the conformation at 150 ps as the final 3D structure for further analysis.

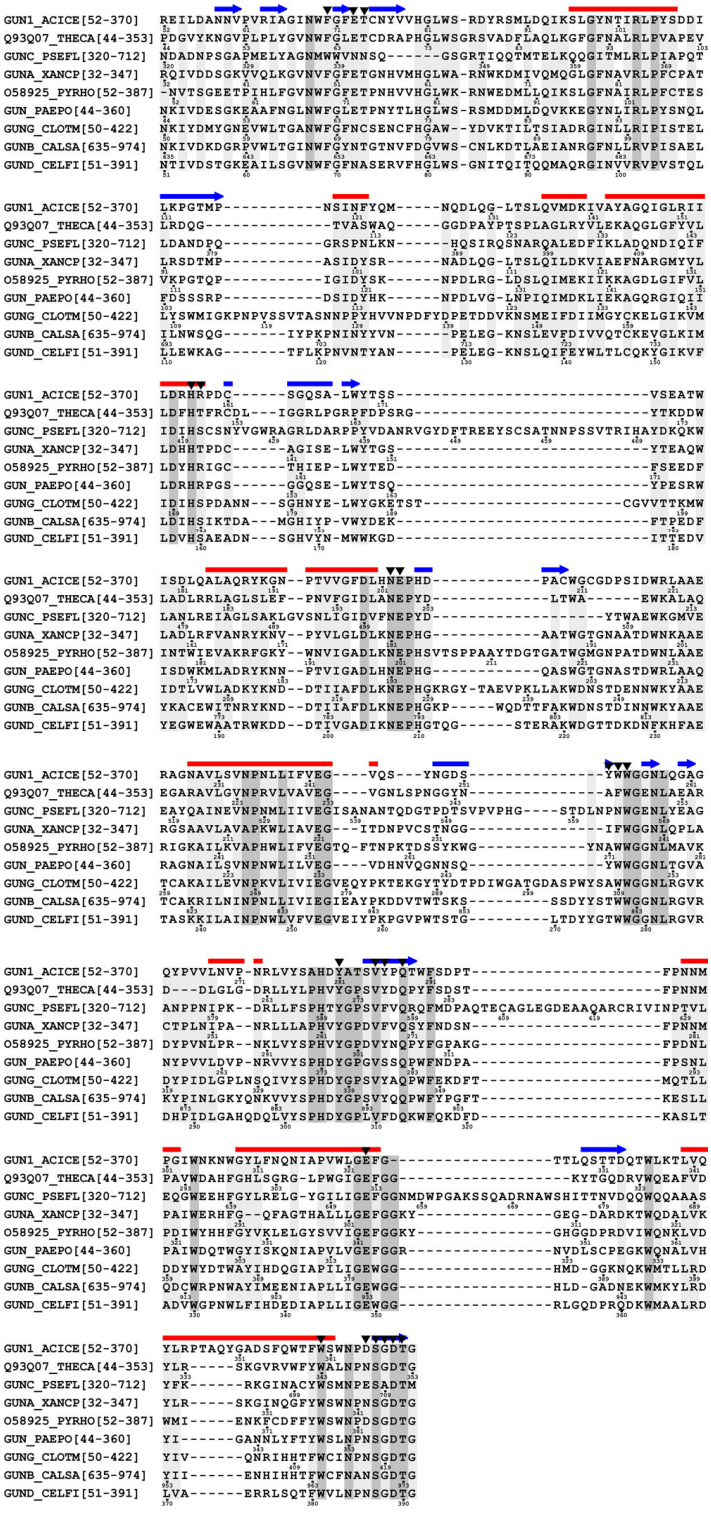

Figure 4. Structure-based sequence alignment of nine family 5 glycosyl hydrolases, two of which have known structures. The secondary structural elements (β-strand, blue; α-helix, red) are indicated above the sequences. The 21 active site residues are indicated by inverted triangles. The protein sequences, with their SwissProt or PDB access codes and selected sequence numbers indicated in parentheses, are: *Adicothermus cellulolyticus* (Entry name:GUN1_ACICE, PDB:**1ECE**, Sequence No:52-370); *T. caldophilus GK24* (Q93Q07_THECA, 44-353); *Pseudomonas fluorescens* (GUNC_PSEFL, 320-712); *Xanthomonas camprestrs pv. campestris* (GUNA_XANCP, 32-347); *Pyrococcus horikoshii* (O58925_PYRHO, 52-387); *Paenibacillus polymyxa* (GUN_PAEPO, 44-360); *Clostrdium thermocellum* (GUNG_GLOTM, **1CEM**, 50-422); *Caldicellulosiruptor saccharolyticus* (GUNB_CALSA, 635-975); *Celllomonas fimi* (GUND_CELFI, 51-391).

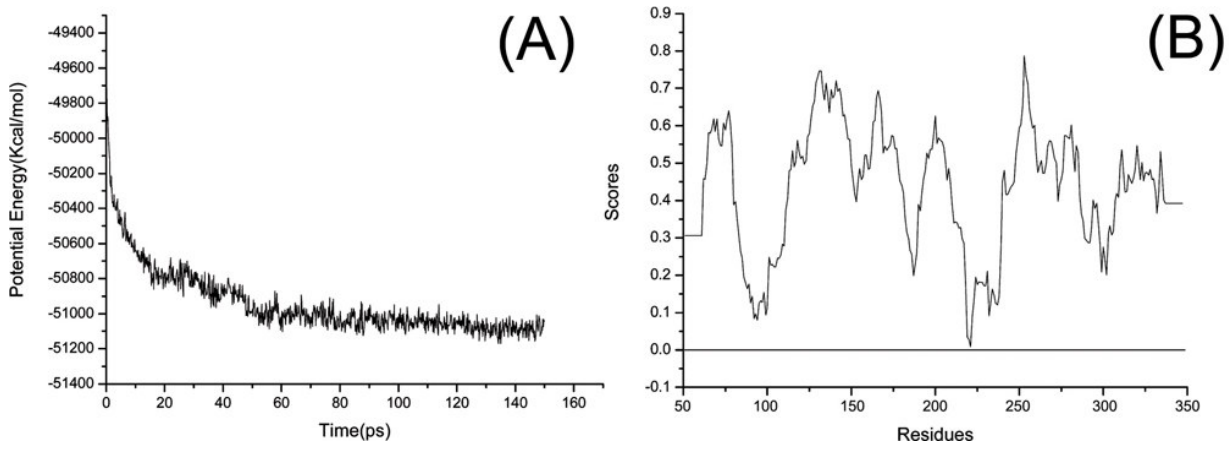

Figure 5. (A) The potential energy of *Tc*Cel5A as a function of simulation time (ps). (B) Evaluation of the final structure of *Tc*Cel5A by PROFILE-3D.

The final structure was further checked using the PROFILE-3D and PROSTAT programs [54]. The results of PROFILE-3D analysis are presented in Figure 5(B). All residues were scored positive, meaning that the positions of these residues in *Tc*GalK model structure were reasonable. The PROSTAT program was used to check dihedral angles, bond lengths, and bond angles in the structures of *Tc*Cel5A and *Ac*Cel5A, and the main results are listed in Table 2. The U-W values for *Ac*Cel5A and *Tc*Cel5A showed that 96.7% and 95.6% of the amino acid residues, respectively, fell within the favored regions of the Ramachandran plot. Of the 308 residues examined in *Tc*Cel5A, eight bond angles deviated from the formal values, while two bond angles out of 317 amino acids examined had greater values than the reference values for *Ac*Cel5A. The RMS deviation (RMSD) of C^{α} atoms between *Tc*Cel5A and *Ac*Cel5A was 0.61 Å, which is in a reasonable range. Overall, the final 3D structure of *Tc*Cel5A appeared to be reliable, based on comparison of its structure with that of *Ac*Cel5A and the assessments mentioned above.

Table 2. The results of dihedral angles, bonds, and angles of *Ac*Cel5A and *Tc*Cel5A checked by PROSTAT program.

Target protein	*Ac*Cel5A	*Tc*Cel5A
%U–W Angles in core Ramachandran region	96.7%	95.6%
Number bond distances with significant deviations	0	0
Number bond angles with significant deviations	2	8
Number residues examined	317	308

The final structural model of *Tc*Cel5A is presented in Figure 6(A). It includes 8 α-helices and eight β-sheets. As seen in Figure 6, the predicted structure of *Tc*Cel5A is typical for members of family 5 endocellulases. Active site residues included a nucleophilic glutamate (Glu312) at the C-terminus of β-strand 7, and the general acid-base pair, asparagine (Asp199)-glutamate (Glu200), at the C-terminus of β-strand 4. During MD simulations, we found that the most variable domain of *Tc*Cel5A was the activation loop between β-strands 4 and 5, indicating that this domain of family 5 endocellulases is the

most variable part of these proteins, and represents a suitable candidate region for determining substrate selectivity.

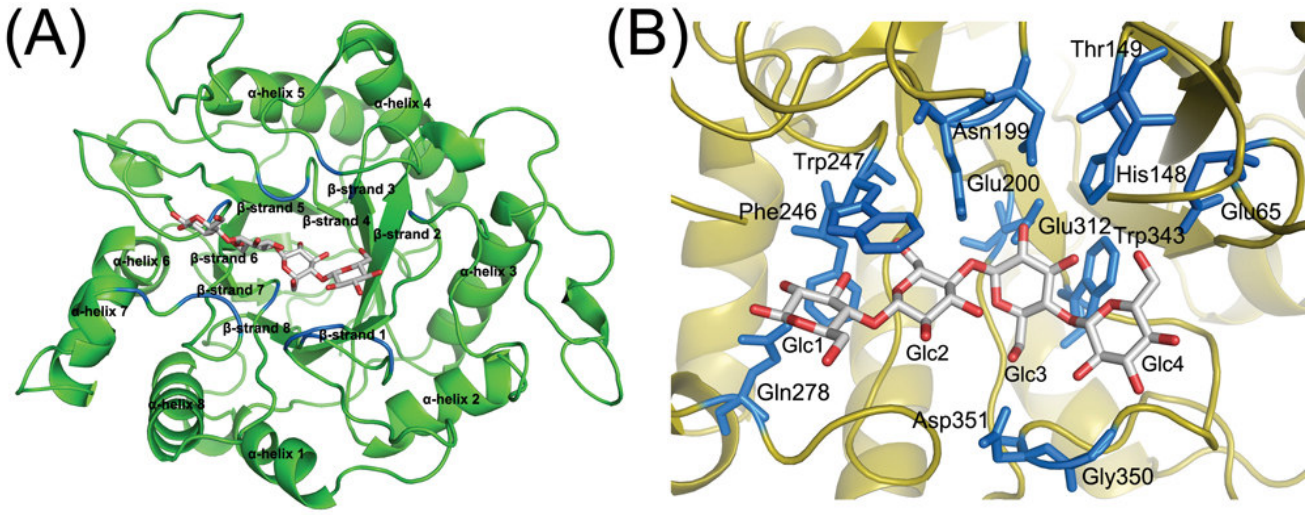

Figure 6. (A) The final structure of the *Tc*Cel5A-G4 complex. (B) Active site residues of *Tc*Cel5A that bind G4.

3.4. Identification of the binding site of TcCel5A

*Tc*Cel5A folds into a typical $(\alpha/\beta)_8$ structure, with the carboxy-terminal ends of β-strands 4 and 7 contributing to the substrate binding cleft. There were no significant differences in domain orientations between the *Ac*Cel5A-G4 complex and the *Tc*Cel5A-G4 complex. The conserved sequence and structure of endocellulases from two different species is not surprising, given that their main biological functions are similar. Similarly, we would predict that G4 binds in a similar manner to both structures. Endocellulases that bind G4 have a classic 4/7 superfamily fold [58-60], which includes the activation loop, and contains the key catalytic residues for hydrolysis of cellulose *i.e.*, the arginine/glutamate acid-base pair and glutamate at the termini of β-strands 4 and 7. This is also a characteristic feature of the sugar linkage-binding site [29, 61-62]. To identify residues involved in the interaction between the active site of *Tc*Cel5A and its substrates, we defined a subset of binding pocket residues as residues in which any atoms were within 4.5 Å of G4. The binding-site was searched using Binding-Site software [55], which was also used to guide the *Tc*Cel5A-G4 docking experiment, and the residues comprising the binding pocket of *Tc*Cel5A are displayed in Figure 6(B), and listed in Table 3. The binding pocket makes up the activation loop consisting of Asn199, Glu200, Ala245, Phe246, Trp247, Tyr271, Val275, Tyr276 from the C-terminal loop between *β*-strands 4 and 5, and Glu312, Trp343, Asn348, Ser349, Gly350, Asp351, Thr352 from the C-terminal loop between *β*-strands 7 and 8. Other endocellulases have similar overall structures and almost superimposable catalytic clefts that include the 4/7 superfamily cellulose binding domain [29, 61-62]. Based on sequence alignment, Asn161/Glu162/Glu282 in *Ac*Cel5A corresponded to Asn199/Glu200/Glu312 in *Tc*Cel5A. With the exception of Thr149, Ala245, Phe246, and Asn348, other residues comprising the binding site in *Tc*Cel5A were conserved in *Ac*Cel5A.

3.5. Explicit characterization of the TcCel5A-G4 complex

Table 3. Hydrogen bonds between *Tc*Cel5A and cellotetraose.

Cellotetraose	Atom	Class[a]	Residue	Atom	Class[a]	Dist[b]	Surf[c]
Glc1	O1	I	Tyr276	OH	I	4.8	0.2
	O1	I	Gln278	OE1	II	4.9	1.4
	O2	I	Gln278	NE2	III	3.0	16.5
	O2	I	Ala245	O	II	3.7	15.4
	O2	I	Gln278	OE1	II	4.1	0.3
	O3	I	Trp247	N	III	3.6	1.2
	O5	II	Trp247	NE1	III	4.4	1.4
Glc2	O2	I	Val275	O	II	5.5	0.2
	O6	I	Trp247	N	III	3.0	26.7
	O6	I	Trp247	O	II	3.2	10.4
	O6	I	Glu200	OE2	II	3.4	0.2
Glc3	O2	I	Glu312	OE1	II	3.0	10.4
	O2	I	Glu312	OE2	II	3.1	9.0
	O2	I	Glu200	OE2	II	3.2	5.5
	O2	I	Glu200	OE1	II	3.3	1.9
	O2	I	Asn199	ND2	III	3.3	5.2
	O3	I	His148	NE2	I	3.0	18.4
	O4	II	Trp343	NE1	III	3.7	0.7
	O6	I	Asp351	OD1	II	2.8	13.7
	O6	I	Asp351	OD2	II	4.2	0.3
Glc4	O2	I	Gly350	N	III	3.6	9.4
	O2	I	Asp351	OD1	II	4.2	0.5
	O3	I	Gly350	N	III	3.3	14.7
	O3	I	Asn348	O	II	3.7	7.4
	O4	I	Thr149	OB	III	4.5	2.6
	O4	I	Asn348	O	II	4.7	0.7
	O6	I	Glu65	OE2	II	2.7	13.7
	O6	I	Thr149	OB	III	3.4	8.1
	O6	I	His148	NE2	I	3.6	7.1

[a]I is Hydrophilic atom type : N and O that can donate and accept hydrogen bonds, II is Acceptor : N or O that can only accept a hydrogen bond, III is Donor : N that can only donate a hydrogen bond. [b]Dist is distance (Å) between the cellotetraose and *Tc*Cel5A atoms. [c]Surf is contact surface area (Å^2) between the cellotetraose and *Tc*Cel5A atoms.

As mentioned above, G4 was selected for this purpose. G4 is a significant endocellulase substrate, hydrolysis of which plays an important role in producing cellobiose. The molecular structure of G4 consists of four glucose molecules, Glc1-Glc2-Glc3-Glc4. The 3D structure of cellotetraose was generated using the BUILDER program, and the geometry was further optimized by using the DISCOVER 3 program. Hydrogen bonds play an important role in the structure and function of biological molecules, and particularly in enzymatic catalysis. The hydrogen bonds of the *Tc*Cel5A-G4 complex are listed in Table 3. There were 29 hydrogen bonds between G4 and *Tc*Cel5A. Glc3 of G4 formed hydrogen bonds with the O2 atom at the C2 position and with the side-chain amide group of Asn199. The Glc3 O2-hydroxyl group of G4 formed two hydrogen bonds with the side-chain carboxyl groups of Glu200 and Glu312. The Glc3 O3-, O4- and O6-hydroxyl groups formed four hydrogen

bonds with His148, Trp343, Asp351 of *Tc*Cel5A, and the Glc2 O2- and O6-hydroxyl groups formed three hydrogen bonds with Trp247 and Val275. To determine the key residues of the binding pocket in this structural model, the total interaction energies for G4 and each individual amino acid in *Tc*Cel5A were calculated. This method, along with the distance from the substrate, can clearly show the relative significance for every residue. The total interaction energies (lower than -1.0 kcal/mol) for G4 and every amino acid of *Tc*Cel5A are listed in Table 4. Glu312, Glu200, Asn199, Asp351, His148, and Trp343 had strong interaction energies with G4, with Glu312 having the strongest interaction energy (-35.434 kcal/mol). Another key residue was Glu200, which strongly interacted with G4 via a hydrogen bond, with an interaction energy of -13.466 kcal/mol. Both Glu200 and Glu312 also participated in strong electrostatic interactions with G4, indicating that they play important roles in the interaction between the G4 and *Tc*Cel5A. Asn199, Asp351, His148 and Trp343 in *Tc*Cel5A also appeared to be important determinants in binding, as they exhibited strong interactions with G4. As seen in Tables 3 and 4, this type of structure-based analysis can serve as a guide to the selection of candidate sites for further experimental studies and site directed mutagenesis.

Table 4. The total interaction (E_{total}), van der Waals (E_{vdw}) and electrostatic energy (E_{ele}) (unit: kcal/mol) between the cellotetraose and the residue of *T. caldophilus* GK24 Cel5A (*Tc*Cel5A).

Residue	Dist[a]	Surf[b]	HB[c]	AR[d]	PH[e]	DC[f]	E_{vdw}	E_{ele}	E_{total}
Glu312*	3.0	27.9	+	-	-	-	2.199	-37.633	-35.434
Glu200*	2.5	39.5	+	-	-	+	1.007	-14.473	-13.466
Asn199*	3.3	5.2	+	-	-	-	-5.176	0.385	-4.791
Asp351*	2.7	31.6	+	-	-	-	-3.398	-1.268	-4.666
His148*	2.8	31.1	+	-	-	-	-3.496	-0.757	-4.253
Trp343*	3.0	26.4	+	-	-	-	-2.619	-0.845	-3.464
Trp247*	3.0	18.7	+	-	-	-	-2.667	-0.432	-3.099
Glu65*	3.0	108.1	+	-	-	+	-2.702	-0.166	-2.868
Gly350*	3.3	28.9	+	-	-	-	-1.324	-1.428	-2.752
Gln278*	3.5	26.0	+	-	-	-	-1.297	-1.434	-2.731
Asn348	3.7	8.1	+	-	-	-	-2.931	0.268	-2.663
Thr149*	3.4	29.4	+	-	-	-	-1.108	-1.435	-2.543
Phe246*	3.2	37.3	-	-	-	+	-2.332	-0.166	-2.498
Ala245	3.7	15.9	+	-	-	-	-2.447	0.106	-2.341
Ser349*	3.7	15.9	-	-	-	-	-2.080	-0.257	-2.337
Phe62*	3.7	30.4	-	-	-	-	-2.184	-0.127	-2.311
Tyr271*	3.7	36.5	-	-	-	-	-1.796	-0.513	-2.309
Tyr276*	3.9	49.5	+	-	-	+	-4.953	2.942	-2.011
Val275*	4.1	22.4	+	-	-	+	-1.790	0.135	-1.655
Thr352*	4.3	17.3	-	-	-	-	-1.690	0.059	-1.631
Thr66*	4.3	8.8	-	-	-	+	-1.414	-0.052	-1.466
His269*	4.7	0.9	-	-	-	+	0.516	-1.858	-1.342
Total							-43.682	-58.949	-102.631

[a]Nearest distance (Å) between the cellotetraose and the residue of *T. caldophilus* GK24 Cel5A (*Tc*Cel5A). [b]Contact surface area (Å^2) between the substrate and the residue of *Tc*Cel5A. [c]Hydrophilic - hydrophilic contact (hydrogen bond). [d]Aromatic-aromatic contact. [e]Hydrophobic - hydrophobic contact, [f]Hydrophobic - hydrophilic contact (destabilizing contact). +/- indicates presence/absence of a specific contacts. * indicates residues contacting ligand by their side chain (including C^{α} atoms).

3.6. Proposed mechanism of hydrolysis by Thermus caldophilus GK24 endocellulase

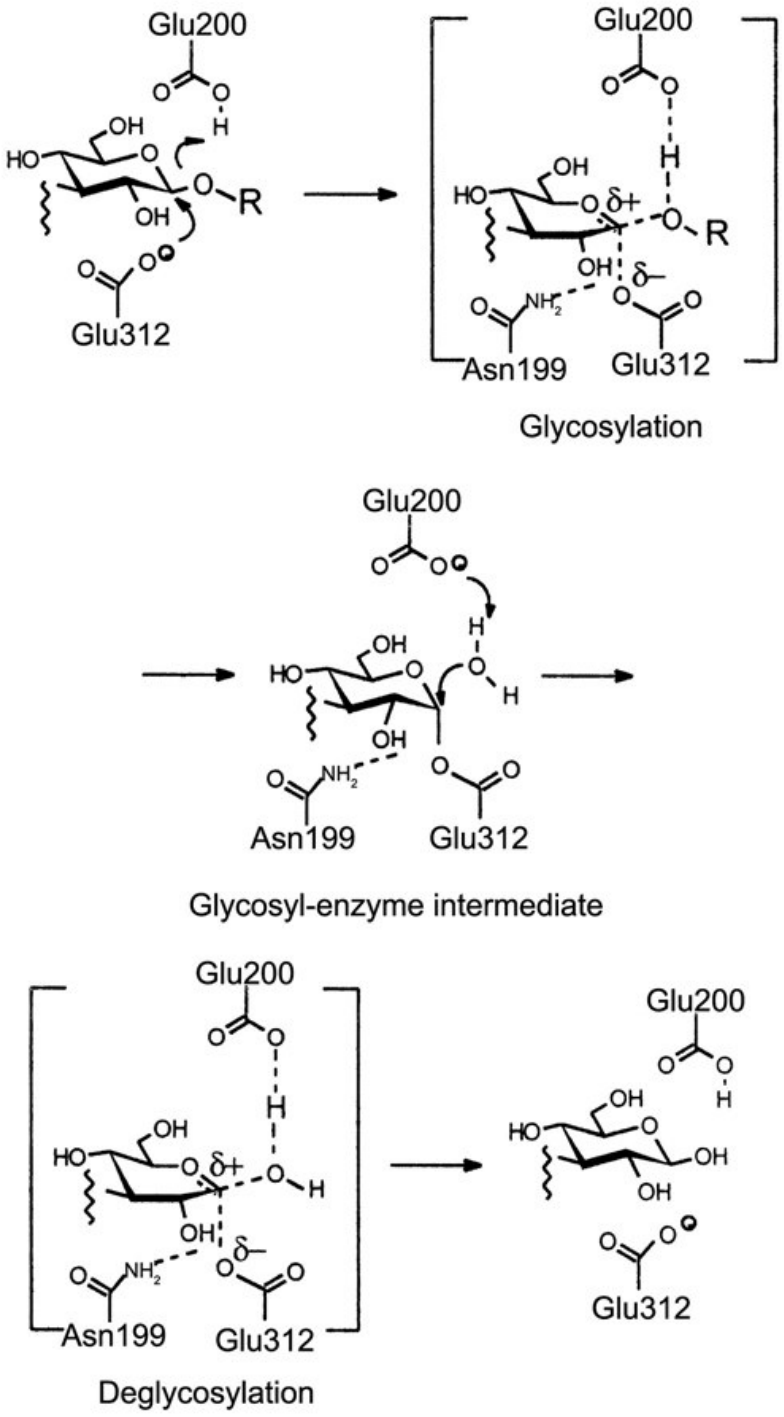

Figure 7. Proposed mechanism of hydrolysis by *Thermus caldophilus* GK24 endocellulase, a retaining glycosidase, involving a double-displacement reaction.

Based on previous work [16,23-24,28-29,30], we propose the following nucleophilic catalytic mechanism for *Tc*Cel5A. As shown in Figure 7, G4 is stably bound in the center of the active site in *Tc*Cel5A through the formation of 29 hydrogen bonds (Table 3). During the glycosylation step, Glu312 acts as a nucleophile, attacking the C1 position of Glc3, displacing aglycon in an inverting reaction assisted by proton transfer to the glycosidic oxygen, resulting in the formation of an intermediate with

two hydrogen bonds: one between the carboxyl group of Glu312 and the O atom (glycosidic oxygen) of the Glc3-Glc2 linkage; the other between the carboxyl group of Glu200 and the hydroxyl group in the C2 position of Glc3. The system then goes into an unstable transitional state, followed by the formation of a covalent glycosyl-enzyme (O2(Glc3)-OE1(Glu312)) intermediate. The second step of the reaction mechanism is deglycosylation. The carboxyl group of Asn200 snatches a proton from water. At the same time, the nucleophilic attack at the C1 position of Glc3 by the oxygen atom of water occurs with the help of Asn199 and Glu200. With the break between Glc3 and Glu312, the proton of water is transferred to the oxygen atom of Glu200. Finally, the charge of the system reaches a balance, and hydrolyzation terminates. The final step involves the attack of a water molecule assisted by Asn199 and Glu200, to release free cellobiose, with overall retention of an anomeric configuration. The above four steps proceed via transition states with considerable oxocarbonium-like character.

4. Conclusions

In this work, we have identified the putative cellulase gene of *T. caldophilus* GK24, through full genome sequencing, and demonstrated cellulase activity using Congo-red staining. The purification of cellulase produced by *T. caldophilus* GK24 was done. As for substrate specificity, *Tc*Cel5A showed the highest activity against cellotetraose (G4). *Tc*Cel5A exhibited the high enzyme activity on CMC and the product of degradation of ASC was cellobiose and cellotriose. From these results, we concluded that *Tc*Cel5A is an endo-type of cellulase. We constructed a 3D structural model of *Tc*Cel5A using the Insight II/Homology module. After energy minimization and molecular dynamics simulations, a refined structure was obtained. This structure was then used to perform a docking experiment using the substrate G4. As a result of simulated docking analysis, a model for the structure of the *Tc*Cel5A-G4 complex was obtained. The results indicated that several conserved amino acid residues in *Tc*Cel5A play important roles in maintaining a functional protein conformation, and are directly involved in binding to cellotetraose. Identification of specific interactions between *Tc*Cel5A with G4 is important for our understanding of the mechanism of binding of *Tc*Cel5A and G4. As is well known, hydrogen bonds play an important role in the structure and function of biological molecules, in particular, in enzymatic catalysis.

Acknowledgements

This work was supported by a grant from the KRIBB Research Initiative Program, Korea.

References and Notes

1. da Costa, M. S.; Nobre, M. F.; Rainey, R. A. The genus *Thermus*, (Boone D. R. and Castenholtz R. W., eds.), Bergey's manual of systematic bacteriology, 2nd ed., vol. 1. Springer, New York, N.Y. 2001.
2. Weisburg, W. G.; Giovannoni, S. J.; Woese, C. R. The Deinococcus-Thermus phylum and the effect of rRNA composition on phylogenetic tree construction. *Syst. Appl. Micorbiol.* **1989**, *11*, 128-134.

3. Brock, T. D.; Freeze, H. *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *J. Bacterial.* **1969**, *98*, 289-297.
4. Taguchi, H.; Yamashita, M.; Matsuzawa, H.; Ohta, T. Heat-stable and fructose 1,6-bisphosphate-activated L-lactate dehydrogenase from an extremely thermophilic bacterium. *J. Biochem.* (*Tokyo*) **1982**, *91*, 1343-1348.
5. Oshima, T.; Imahori, K. Isolation of an extreme thermophile and thermostabilities of its transfer ribonucleic acid and ribosomes. *J. Gen. Appl. Microbiol.* **1971**, *17*, 513-317.
6. Manaia, C. M.; Hoste, B.; Gutierrez, M. C.; Gillis, M.; Ventosa, A.; Kersters, K.; da Costa, M. S. Halotolerant *Thermus* strains from marine and terrestrial hot springs belong to *Thermus thermophilus* nom. rev. emend. *Syst. Appl. Microbiol.* **1994**, *17*, 526-532.
7. Williams, R. A.; Smith, K. E.; Welch, S. G.; Micallef, J.; Sharp, R. J. DNA relatedness of *Thermus* strains, description of *Thermus brockianus* sp. nov., and proposal to reestablish *Thermus thermophilus*. *Int. J. Syst. Bacteriol.* **1995**, *45*, 495-499.
8. Nobre, M. F.; Truper, H. G.; da Costa, M. S. Transfer of *Thermus ruber*, *Thermus silvanus* and *Thermus chliarophilus* to *Meiothermus* gen. nov. as *Meiothermus ruber* comb. nov., *Meiothermus silvanus* comb. nov. and *Meiothermus chliarophlus* comb. nov., respectively, and emendation of the genus *Thermus*. *Int. J. Syst. Bacterial.* **1996**, *46*, 604-606.
9. Coughlan, M. P. Cellulose degradation by fungi, in Microbial Enzymes and Biotechnology (W. M. Fogarty and C. T. Kelly, Eds.) Elsevier Applied Science, London. 1990.
10. Ito, S. Alkaline cellulases from alkaliphilic *Bacillus*: enzymatic properties, genetics, and application to detergents. *Extremophiles* **1997**, *1*, 61-66.
11. Beguin, P.; Aubert J. A. The biological degradation of cellulose. *FEMS Microbiol. Rev.* **1994**, *13*, 25-58.
12. Tomme, P.; Warren, R. A. J.; Gilkes, N. R. Cellulose Hydrolysis by Bacteria and Fungi. *Adv. Microbial Physiol.* **1985**, *37*, 1-81.
13. Gilkes, N. R.; Kilburn, D. G.; Miller, Jr. R. C.; Warren, R. A. J. Structural and functional analysis of a bacterial cellulase by proteolysis. *J. Biol. Chem.* **1989**, *264*, 17802-17808.
14. Gilkes, N. R.; Henrissat, B.; Kilburn, D. G.; Miller, Jr. R. C.; Warren, R. A. J. Domains in microbial β-1,4-glycanase: Sequence conservation, function, and enzyme families. *Microbiol. Rev.* **1991**, *55*, 303-315.
15. Stone, J. E.; Scallan, A. M.; Donefer, E.; Ahlgren, E. Cellulases and their application (R. F. Gould Ed.), American Chemical Society, Washington DC. 1969.
16. Fan, L. T.; Lee, Y. H.; Beardmore, D. H. Mechanism of the enzymatic hydrolysis of cellulose: Effects of major structural features of cellulose on enzymatic hydrolysis. *Biotechnol. Bioeng.* **1980**, *22*, 177-199.
17. Cowling, E. B. Physical and chemical constraints in the hydrolysis of cellulose and lignocellulosic materials. *Biotechno. Bioeng. Symp.* **1975**, *5*, 163-181.
18. Grethlein, H. E. The effect of pore size distribution on the rate of enzymatic hydrolysis of cellulosic substrates. *Bio/Technology* **1985**, *3*, 155-160.
19. Gama, F. M.; Teixeira, J. A.; Mota, M. Cellulose morphology and enzymatic reactivity: A modified solute exclusion technique. *Biotechnol. Bioeng.* **1994**, *43*, 381-387.

20. Lee, S. B.; Shin, H. S.; Ryu, D. D. Y. Adsorption of cellulase on cellulose: Effect of physicochemical properties of cellulose on adsorption and rate of hydrolysis. *Biotechnol. Bioeng.* **1982**, *24*, 2137-2153.
21. Woodward, J.; Hayes, M. K.; Lee, N. E. Hydrolysis of cellulose by saturating and non-saturating concentrations of cellulase: Implications for synergism. *Bio/Technology* **1988**, *6*, 301-304.
22. Tanaka, M.; Ikesaka, M.; Matsuno, R. Effect of Pore Size in Substrate and Diffusion of Enzyme on Hydrolysis of Cellulosic Materials with Cellulases. *Biotechnol. Bioeng.* **1988**, *32*, 698-706.
23. Coughlan, M. P. Mechanisms of cellulose degradation by fungi and bacteria. *Anim. Feed Sci. Tech.* **1991**, *32*, 77-100.
24. Enari, T. M.; Niku-Paavola, M. L. Enzymatic hydrolysis of cellulose: is the current theory of the mechanisms of hydrolysis valid?. *CRC Cri. Rev. Biotech.* **1987**, *5*, 67-87.
25. Wood, T. M.; Garcia-Campayo, V. Enzymology of cellulose degradation. *Biodegradation* **1990**, *1*, 147.
26. Woodward, J. Synergism in cellulase systems, *Bioresource Tech.* **1991**, *36*, 67-75.
27. Henrissat, B. Enzymatic cellulose degradation. *Cellulose Commun.* **1998**, *5*, 84-90.
28. Sinnot, M. L. Catalytic mechanisms of enzymic glycosyl transfer. *Chem. Rev.* **1990**, *90*, 1171-1202.
29. Rabinovich, M. L.; Melnick, M. S.; Bolobova, A. V. The structure and mechanism of action of cellulolytic enzymes. *Biochemistry* (*Moscow*) **2002**, *67*, 850-871.
30. Skopec, C. E.; Himmel, M. E.; Matthews, J. F.; Brady, J. W. Energetics for displacing a single chain from the surface of microcrystalline cellulose into the active site of *Acidothermus cellulolyticus* Cel5A. *Protein Eng.* **2003**, *16*, 1005-1015.
31. McCarter, S. L.; Adney, W. S.; Vinzant, T. B.; Jennings, E.; Eddy, F. P.; Decker, S. R.; Baker, J. O.; Himmel, M. E. Exploration of cellulose surface-binding properties of *Acidothermus cellulolyticus* Cel5A by site-specific mutagenesis. *Appl. Biochem. Biotech. - Part A Enzyme Engineering and Biotechnology* 2002, **98-100**, 273-287.
32. Sakon, J.; Adney, W. S.; Himmel, M. E.; Thomas, S. R.; Karplus P. A. Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. *Biochemistry* **1996**, *35*, 10648-10660.
33. Zverlov, V. V.; Schantz, N.; Schwarz, W. H. A major new component in the cellulosome of *Clostridium thermocellum* is a processive endo-β-1,4-glucanase producing cellotetraose. *FEMS Microbiol. Let.* **2005**, *249*, 353-358.
34. Guimarães, B. G.; Souchon, H.; Lytle, B. L.; Wu, J. H. D.; Alzari, P. M. The crystal structure and catalytic mechanism of cellobiohydrolase celS, the major enzymatic component of the *Clostridium thermocellum* cellulosome. *J. Mol. Biol.* **2002**, *320*, 587-596.
35. Zang, S.; Guifang, L.; David, B. W. Characterization of a *Thermomonospora fusca* Exocellulase. *Biochemistry* **1995**, *34*, 3386-3395.
36. Harchand, R. K.; Singh, S. Characterization of cellulase complex of *Streptomyces albaduncus*. *J. Basic Microbiol.* **1997**, *37*, 93-103.
37. Jin-Duck, B.; Dinesh, A. Y.; Douglas, E. E. Purification, Characterization, and Molecular Analysis of Thermostable cellulases CelA and CelB from *Thermotoga neapolitana*. *Appl. Environ. Microbiol.* **1998**, *64*, 4774-4781.

38. Tettelin, H.; Radune, D.; Kasif, S.; Khouri, H.; Salzberg, S. L. Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics* **1999**, *62*, 500-507.
39. Ewing, B.; Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **1998**, *8*, 186-194.
40. Ewing, B.; Hillier, L.; Wendl, M. C.; Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **1998**, *8*, 175-185.
41. Gordon, D.; Abajian, C.; Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **1998**, *8*, 195-202.
42. Gordon, D.; Desmarais, C.; Green, P. Automated finishing with autofinish. *Genome Res.* **2001**, *11*, 614-625.
43. Park, J. H.; Park, B. C.; Koh, S. H.; Kim, J. S.; Koh, J. H.; Yang, M. H.; Kim, Y. S.; Kim, C. H.; Kim, M. H.; Kwon, S. T.; Lee, D. S. Genome mapping of an extreme thermophile, *Thermus caldophilus* GK24. *Genomics inform.* **2003**, *1*, 50-54.
44. Henne, A.; Bruggemann, H.; Raasch, C.; Wiezer, A.; Hartsch, T.; Liesegang, H.; Johann, A.; Lienard, T.; Gohl, O.; Martinez-Arias, R.; Jacobi, C.; Starkuviene, V.; Schlenczeck, S.; Dencker, S.; Huber, R.; Klenk, H. P.; Kramer, W.; Merkl, R.; Gottschalk, G.; Fritz, H. J. The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nat. Biotechnol.* **2004**, *22*, 547-553.
45. http://gib.genes.nig.ac.jp/single/main.php?spid=Tthe_HB8
46. Kurtz, S.; Phillippy, A.; Delcher, A. L.; Smoot, M.; Shumway, M.; Antonescu, C.; Salzberg, S. L. Versatile and open software for comparing large genomes. *Genome Biol.* **2004**, *5*, R12.
47. Teather, R. M.; Wood, P. J. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl. Environ. Microbiol.* **1982**, *43*, 777-780.
48. Tsigelny, I. F. Protein Structural Prediction Bioinformatic approach, International University Line. 2002.
49. Pearson, W. R. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics* **1991**, *11*, 635-650.
50. Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **1994**, *22*, 4673-4680.
51. Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779-815.
52. Sali, A.; Overington, J. P. Derivation of rules for comparative protein modeling from a database of protein structure alignments. *Protein Sci.* **1994**, *3*, 1582-1596.
53. Sali, A.; Potterton, L.; Yuan, F.; Van Vlijmen, H.; Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins* **1995**, *23*, 318-326.
54. Luthy, R.; Bowie, J. U.; Eisenberg, D. Assesment of protein models with three-dimensional profiles. *Nature* **1992**, *356*, 83-85.
55. Discover 3 User Guide, San Diego: MSI, USA. 1999.
56. Miller GL (1963) Method in carbohydrate chemistry Vol III, by R. L. Whistler, Academic Press, New York., 134-139.

57. Ponder, J. W.; Richards, F. M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **1987**, *193*, 775-791.
58. Tull, D.; Withers, S. G.; Gilkes, N. R.; Kilburn, D. G.; Warren, R. A. J.; Aebersold, R. Glutamic acid 274 is the nucleophile in the active site of a "retaining" exoglucanase from *Cellulomonas fimi*. *J. Biol. Chem.* **1991**, *266*, 15621-15625.
59. Wang, Q.; Tull, D.; Meinke, A.; Gilkes, N. R.; Warren, R. A. J.; Aebersold, R.; Withers, S. G. Glu280 is the nucleophile in the active site of *Clostridium thermocellum* CelC, a family A endo-beta-1,4-glucanase. *J. Biol. Chem.* **1993**, *268*, 14096-14102.
60. Bortoli-German, I.; Haiech, J.; Chippaux, M.; Barras, F. Informational Suppression to Investigate Structural Functional and Evolutionary Aspects of the *Erwinia chrysanthemi* Cellulase. *J. Mol. Biol.* **1995**, *246*, 82-94.
61. Planas, A. Bacterial 1,3-1,4-β-glucanases: Structure, function and protein engineering. *Biochim. Biophys. Acta.* **2000**, *1543*, 361-382.
62. Jenkins, J.; Leggio, L. L.; Harris, G.; Pickersgill, R. Beta-glucosidase, beta-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold beta/alpha architecture and with two conserved glutamates near the carboxy-terminal ends of beta-strands four and seven. *FEBS Lett.* **1995**, *362*, 281-285.