Applied Microbiology and Biotechnology
https://doi.org/10.1007/s00253-020-10831-5

BIOTECHNOLOGICALLY RELEVANT ENZYMES AND PROTEINS

Characterization of two GH5 endoglucanases from termite microbiome using synthetic metagenomics

Emiliano Ben Guerrero[1] · Rubén Marrero Díaz de Villegas[1] · Marcelo Abel Soria[2] · M. Paz Santangelo[1] · Eleonora Campos[1] · Paola M. Talia[1]

Received: 26 May 2020 / Revised: 5 August 2020 / Accepted: 11 August 2020

Abstract

Here, we characterize two novel GH5 endoglucanases (GH5CelA and GH5CelB) from an uncultured bacterium identified in termite gut microbiomes. Both genes were codon-optimized, synthetized, cloned, and expressed as recombinant proteins in *Escherichia coli* for subsequent purification. Both enzymes showed activity on the *p*NPC and barley β-glucan substrates, whereas GH5CelB also showed low activity on carboxymethyl cellulose. The optimum conditions for both enzymes were an acid pH (5) and moderate temperature (35 to 50 °C). The enzymes differed in the kinetic profiles and patterns of the generated hydrolysis products. A structural-based modeling analysis indicated that both enzymes possess a typical (β/α)8-barrel fold characteristic of GH5 family, with some differential features in the active site cleft. Also, GH5CelB presents a putative secondary binding site. Furthermore, adjacent to the active site of GH5CelA and GH5CelB, a whole subdomain rarely found in GH5 family may participate in substrate binding and thermal stability.

Therefore, GH5CelA may be a good candidate for the production of cello-oligosaccharides of different degrees of polymerization applicable for feed and food industries, including prebiotics. On the other hand, GH5CelB could be useful in an enzymatic cocktail for the production of lignocellulosic bioethanol, because of the production of glucose as a hydrolysis product.

Key Points
- *Synthetic metagenomics is a powerful approach for discovering novel enzymes.*
- *Two novel GH5 endoglucanases from nonculturable microorganisms were characterized.*
- *Structural differences between them and other GH5 endoglucanases were observed.*
- *The enzymes may be good candidates for feed, food, and/or bioethanol industries.*

Keywords GH5 · Endoglucanase · Biochemical characterization · Termites · Synthetic metagenomics

Introduction

In recent years, worldwide researchers have shown interest in the development of robust biocatalysts for both the hydrolysis of lignocellulosic biomass and the synthesis of value-added bioproducts (Bastien et al. 2013; Kuhad et al. 2011). Lignocellulosic biomass consists mainly of three major polymers (cellulose, hemicellulose, and lignin) and presents an abundant renewable feedstock for many biorefinery processes.

Cellulose, a major component of plant cell wall, is the most abundant biopolymer on earth. Its degradation is carried out by the enzymes called "cellulases," which are responsible for the hydrolysis of β-1,4-linkages present in cellulose (Wilson 2011). The structure of

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00253-020-10831-5) contains supplementary material, which is available to authorized users.

✉ Paola M. Talia
talia.paola@inta.gob.ar; taliapaolam@gmail.com

[1] Instituto de Agrobiotecnología y Biología Molecular (IABIMO), Instituto Nacional de Tecnología Agropecuaria (INTA), Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Dr. N. Repetto y Los Reseros s/n, 1686 Hurlingham, Provincia de Buenos Aires, Argentina

[2] Cátedra de Microbiología Agrícola, Facultad de Agronomía, Universidad de Buenos Aires, INBA-CONICET, Ciudad Autónoma de Buenos Aires, Argentina

Published online: 20 August 2020

cellulose consists of crystalline and amorphous regions. At least three different types of synergically acting enzymes are required for the complete hydrolysis of cellulose: endo-β-1,4-glucanases (EG; EC 3.2.1.4), cellobiohydrolases or exoglucanases (CBHs; EC3.2.1.91), and β-glucosidases (EC; 3.2.1.21). Endoglucanases produce random internal cuts within the amorphous region in the cellulose molecule, yielding cello-oligosaccharides of various lengths and thereby creating new reducing ends. CBHs display an exo-type attack of polymeric substrates, and the major product of their action on cellulose is cellobiose. These soluble cellobioses, along with cello-oligosaccharides, are then hydrolyzed by β-glucosidases to glucose (Lynd et al. 2002; Soni et al. 2010).

Endo-β-1,4-glucanases have a biotechnological potential in various industrial applications (Kuhad et al. 2011) and belong to any of several glycoside hydrolase families (GH5, GH6, GH7, GH8, GH9, GH12, GH44, GH45 GH48, GH51, GH74, and GH124) according to amino acid sequence similarity (Wu and Wu 2020; Tang et al. 2014; Kim et al. 2016).

To date, more than 20 different enzymatic activities have been reported in GH5 family such as endoglucanases, cellobiohydrolases, mannanases, xylanases, and xyloglucanases, among others, which suggests that this family could have a wide range of applications in different industries (Gautam et al. 2010; Zhang et al. 2006; Aspeborg et al. 2012). All GH5 enzymes share a common $(\beta/\alpha)_8$-barrel fold. Furthermore, more than 30 structures of GH5 are currently available in the protein databank (PDB), with different substrate specificities that may be produced by significant variations in the surface loops (Zhang et al. 2006).

Although in recent years researchers have extended the identification and knowledge of new enzymes or their activities, currently most of these enzymes originated from culturable microorganisms, and only a few derived from nonculturable microorganisms. In the last few years, several insect-inhabiting microorganism enzymes have been identified by the process known as "synthetic metagenomics," which involves identifying, synthesizing, and expressing enzymes from metagenomes (Joynson et al. 2017; Liu et al. 2019; Shi et al. 2013; Xia et al. 2017).

In a previous study, we explored microbiomes from guts of two termites and therefore identified many (hemi)cellulolytic genes (Romero Victorica et al. 2020). To explore these results, here we performed a targeted synthetic metagenomic approach. Two codifying genes for GH5 cellulases were selected, codon-optimized, and expressed in *Escherichia coli* to characterize their endoglucanase and cellobiohydrolase activities.

Materials and methods

Identification, heterologous expression, and protein purification of two GH5-enconding genes

In a previous analysis of our work group, several candidate genes encoding for lignocellulose degradation were identified by shotgun metagenomic sequencing of microbiomes from *Cortaritermes fulviceps* and *Nasutitermes aquilinus* termite guts (Romero Victorica et al. 2020). From the assembled contigs, we selected three predicted GH5-encoding genes KBCPBGKF_00151, KBCPBGKF_26578, and KBCPBGKF_04580 henceforth referred to as GH5CelA, GH5CelB, and GH5CelC, respectively. These GH-encoding genes were selected on the basis of the following criteria: (a) identified as part of the GH5 family, which is the most abundant in these metagenomes, (b) have a complete coding sequence with identifiable start and stop codons and a complete open reading frame, (c) belong to a non-cultivable organism, (d) identified in the gut metagenomes of *N. aquilinus* and *C. fulviceps* analyzed by Romero Victorica et al. (2020), and (e) group together in the phylogenetic analysis of GH5 family.

A BLASTP analysis of GH5CelA, GH5CelB, and GH5CelC was performed against reference protein sequences (refseq) available in the NCBI database (http://blast.ncbi.nlm.nih.gov/) using ClustalW program from MEGA software v6.0 (Hruska and Kaevska 2012).

Open reading frames of endoglucanases GH5CelA and GH5CelB were codon-optimized (GH5CelA-CO and GH5CelB-CO) for expression in *Escherichia coli* and subsequently synthesized and cloned into pET28a+ vector (Novagen, Birmingham, UK). *E. coli* Shuffle (NEB, MA, USA) and Rossetta DE3 (Novagen, Birmingham, UK) competent cells were transformed with the recombinant plasmids pET28a-GH5CelA-CO and pET28a-GH5CelB-CO, respectively. In addition, both competent cells were transformed with pET28a-GH5CelC. Recombinant clones were selected for protein expression and purification. Briefly, protein expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16 h at 20 °C. After cell lysis and sonication (six pulses of 10 s, 28% amplitude), recombinant proteins were purified in the soluble fraction by immobilized metal affinity chromatography (IMAC) with Ni-NTA agarose resin (Qiagen, Venlo, Netherlands), using 20 mM Tris-Hcl pH 7, 20 mM KCl, 1 mM EDTA, 250 mM imidazole pH 7, and 0.1% (v/v) Igepal (Sigma-Aldrich, St. Louis, USA) as elution buffer. The purified proteins were conserved at 4 °C.

SDS-PAGE, western blot, and zymogram analysis of the recombinant proteins

SDS-PAGE was performed in a 12% (w/v) polyacrylamide gel according to Laemmli (1970). The purified enzyme samples mixed with the same volume of loading buffer were boiled at 100 °C for 5 min and subsequently subjected to SDS-PAGE. The gel was stained with Coomassie Brilliant Blue R-250 and destained with a destain solution (50% methanol, 10% acetic acid, and 40% H_2O) for 3–5 h. For western blotting, SDS-PAGE-separated proteins were blotted onto a nitrocellulose membrane Hybond C-Extra (GE Healthcare Life Science, Rahway, USA) by using Mini Trans-Blot™ (Bio-Rad, Irvine, USA) according to the manufacturer's specification. The separated recombinant proteins were detected with the anti-His mouse antibody (GE Healthcare Life Science, Rahway, USA) and the anti-mouse AP conjugate goat antibody (Sigma-Aldrich, St. Luis, USA) using the BCIP/NBT substrate. A broad range of prestained protein ladder (10–180 kDa) (Thermo Scientific, Waltham, USA) was used as a molecular weight marker.

For the zymogram analysis, the purified enzyme samples were loaded into 12% SDS-PAGE containing 0.5% CMC. After electrophoresis, the gel was treated according to Ben Guerrero et al. (2015). Briefly, the gels were washed twice with 0.04 M Tris-HCl, pH 7.6 for 1 h each, and incubated at 4 °C overnight. Then, the gels were washed again and incubated at 37 °C for 2 h. Finally, they were stained with Congo red for 15 min and subsequently destained in 1 M NaCl.

Analysis of hydrolysis reaction products by thin layer chromatography

The hydrolysis patterns of cello-oligosaccharides, CMC, and barley β-glucan were qualitatively analyzed by thin layer chromatography (TLC) in silica gel plates (GE Healthcare Life Science, Rahway, USA), using butanol/acetic acid/water (2:1:1) as solvents and revealed by water/ethanol/sulfuric acid (20:70:3) with 1% (v/v) orcinol solution, over flame. Glucose (G1), cellobiose (G2), cellotriose (G3), cellotetraose (G4), cellopentaose (G5), and cellohexaose (G6) (Megazyme) (Sigma-Aldrich, St. Louis, USA) were used as standards.

Enzyme activity assays

The cellobiohydrolase activity was assayed using 4-nitrophenyl-β-D-cellobioside (*p*NPC) (Sigma-Aldrich, St. Louis, USA). Reactions of 0.1 mL containing 2.5 mM of substrate were prepared in 50 mM citrate/phosphate buffer pH 6.5 and the properly diluted enzyme solution. Mixtures were incubated at 45 °C for 20 min, and the reaction was stopped by adding 0.5 mL of 2% Na_2CO_3. The released *p*-nitrophenol (*p*NP) concentration was calculated according to a standard curve measuring the absorbance at 410 nm. Enzyme activities were expressed as IU/mg of protein. One international unit (IU) of cellobiohydrolase activity was defined as the amount of enzyme that liberates 1 μmol *p*NP per minute at the assayed conditions.

The endoglucanase activity was assayed in 1% (w/v) of carboxymethyl cellulose (CMC) (Sigma-Aldrich, St. Louis, USA), in a final reaction volume of 0.2 mL at 400 rpm for 15 min in a Thermomixer (Eppendorf, Hamburg, Germany). Reducing sugars released from CMC hydrolysis were measured using the 3,5-dinitrosalicylic acid (DNS) assay (Miller 1959) with glucose as standard.

The optimal pH was studied using sodium citrate (pH 3–4), sodium phosphate (pH 5–8), and glycine-NaOH (pH 8.5–10) buffers at 45 °C, and the effect of temperature was evaluated by incubation at pH 6.5 at temperatures between 30 and 70 °C.

The thermal stability was evaluated by pre-incubating the enzymes at 30, 35, 40, 45, and 50 °C from 0 to 24 h. In addition, kinetic parameters were determined under optimal assay conditions using 0–20 mg/mL of *p*NPC or CMC as substrate, by fitting the models to data with the software GraphPad Prism v 6.0 (http://www.graphpad.com/scientific-software/prism/). Other substrates were also used for testing enzyme specificity: 1% (w/v) Avicel (Fluka), 1% (w/v) bacterial microcrystalline cellulose (BMCC), 1% (w/v) phosphoric acid-swollen cellulose (PASC), 0.5% (w/v) beechwood xylan, 1% (w/v) galactomannan, 1% (w/v) barley β-glucan, 1% (w/v) laminarin, 2.5 mM *p*NP-glucopyranoside, and 2.5 mM *p*NP-xylopyranoside.

Phylogenetic analysis and structural modeling of GH5CelA and GH5CelB

A fasta file for GH5 family was prepared with the amino acid sequences from the annotation obtained in our previous study (Romero Victorica et al. 2020) along with reference sequences with structural information available in CAZy (http://www.cazy.org). Then, the sequences were aligned with Muscle (Edgar 2004) (http://www.ebi.ac.uk/Tools/msa/muscle/), and neighbor-joining phylogenetic trees were built using the Phangorn package in R (Schliep 2011). In all cases, the branch support was calculated by bootstrapping with 1000 resampling iterations.

GH5CelA and GH5CelB proteins were aligned with complete amino acid sequences of GH5 family and subfamilies 5, 22, 25, 26, 36, 37, and 39 available in CAZy database (http://www.cazy.org), using the ClustalW from MEGA program version 6.0 (Hruska and Kaevska 2012). Maximum likelihood (ML) trees were built assuming a Poisson distribution. Poisson substitution model was used because it was the best-fit model according to the analysis implemented in MEGA. The branch support was calculated by bootstrapping, by performing 1000 resampling iteration. The subfamilies were

selected based on the following criteria: (1) shows similar characteristics as observed experimentally; (2) has bacterial origin endo-β-1,4-glucanase, cellodextrinases, and/or active on β-glucan polymers; and (3) lacks carbohydrate-binding module (CBM). Homology modeling by Iterative Threading Assembly Refinement (I-TASSER) (Skerman et al. 1980) (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) was used to generate three-dimensional models of GH5CelA and GH5CelB. The models were constructed by retrieving structure of proteins with similar folds from PDB (Protein Data Bank) library (http://www.rcsb.org) by LOMETS (Wu and Zhang 2007). The fragments from templates with the highest significance level (by Z-score) in the alignments were reassembled into a full-length model.

The final model resulted from a second simulation round, in which the global topology was refined. The confidence of each model was quantitatively measured by C-score (confidence score), TM-score (template modeling score), and RMSD (root-mean-square deviation). The C-score is typically in the range of − 5 to 2, where the highest scores show the highest quality values. A C-score > − 1.5 indicates a model of correct topology. TM-score has values between 0 and 1, where 1 indicates a perfect match between two structures (Roy et al. 2010). A TM-score > 0.5 indicates a model of correct topology, and a TM-score < 0.17 means a random similarity. RMSD is the root-mean-square deviation of atomic positions, which is the measure of the average distance between the atoms of superimposed proteins.

A functional analysis was conducted by COACH (combined results from COFACTOR, TMSITE, and SSITE programs), within I-TASSER, which included prediction of the protein-ligand-binding site. Pocket size prediction and measurement were made by Depth (r1) and CASTp (r2) web servers, respectively. Chimera software v1.11 (Tsukamura et al. 1983) was used to perform model manipulation, imaging, as well as the mapping of pocket predictions and atomic partial charge coloring onto the molecular surfaces of enzymes (Tan et al. 2013; Tian et al. 2018).

Statistical analysis

All the assays were carried out at least in triplicate. Data were expressed as the mean ± one standard deviation of the triplicate measurement. Enzymatic activity data were analyzed for statistical significance by a one-way analysis of variance (ANOVA) and post-test Tukey's multiple comparison using GraphPad Prism, v5.01 (GraphPad Software Inc., San Diego, CA).

Nucleotide sequence accession numbers

The amino acid sequence of GH5CelA and GH5CelB were deposited in GenBank database with the accession numbers MK636676 and MK636677, respectively. In addition, the nucleotide sequences of the codon-optimized GH5CelA-CO and GH5CelB-CO genes were deposited in GenBank database with the accession numbers MT610910 and MT610911.

Results

Selection, identification, cloning, and sequence analysis of three endoglucanase genes

According to the BLASTP analysis, the amino acid sequence of GH5CelA had 79% identity and 98% coverage with a GH5 cellulase from an uncultured bacterium (ABW39346), whereas GH5CelB showed 84% identity and 99% coverage with a GH5 from an uncultured bacterium (ABW39352). Finally, GH5CelC showed 72% identity and 96% coverage with a GH5 from uncultured bacterium (ABW39345). A multiple amino acid sequence alignment was performed with GH5 endoglucanase sequences from nonculturable bacteria, previously identified in the metagenomic analysis of the termite gut microbiota *Nasutitermes* (Warnecke et al. 2007). The conserved signature sequence of the GH5 family was between amino acids 139–148 (VVYEVLNEPH) and amino acids 123–132 (VIYEILNEPH) for GH5CelA and GH5CelB, respectively. Both enzymes showed Glu catalytic residues corresponding to family 5 (Supplemental Fig. S1). Regarding GH5CelC, the conserved signature sequence of the GH5 family was between amino acids 152–161 (VLFESLNEPV) (data not shown).

To characterize the enzymatic activity of the three putative endoglucanases, we cloned and recombinantly expressed GH5CelA-CO, GH5CelB-CO, and GH5CelC as His6X N-terminal fusion proteins. Therefore, rGH5CelA and rGH5CelB were detectable in the soluble fraction, whereas GH5CelC was mainly insoluble and occurred in the form of inclusion bodies (data not shown). We thus continued with the characterization of rGH5CelA and rGH5CelB, which were detected as monomeric proteins in non-denaturing gels (data not shown), with a molecular weight of 40 kDa, according to a SDS-PAGE analysis (Fig. 1a).

Enzymatic properties of GH5CelA and GH5CelB

The substrate specificity was determined by incubating the purified enzymes with different cellulosic and hemicellulosic substrates. The purified rGH5CelA was active on *p*NP-cellobioside (*p*NPC) (42.3 ± 2.0 IU/mg) and barley β-glucan (0.95 ± 0.02 IU/mg). On the other hand, rGH5CelB was active on *p*NPC (29.2 ± 0.4 IU/mg) and barley β-glucan (1.3 ± 0.4 IU/mg) and lightly functional on CMC (0.1 ± 0.01 IU/mg). However, this enzyme was inactive on other hemicellulosic substrates or crystalline cellulose (Table 1). As expected, only

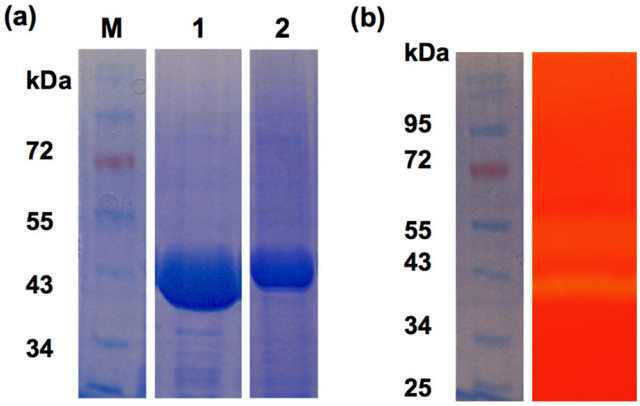

Fig. 1 Analysis of purified rGH5CelA and rGH5CelB. Soluble IMAC purification (12% SDS-PAGE) stained with Coomassie blue (**a**) rGH5CelA (line 1) and rGH5CelB (line 2). (**b**) CMC zymography (12% SDS-PAGE, 0.5% carboxymethyl cellulose) of rGH5CelB. Prestained protein marker (M)

rGH5CelB displayed endoglucanase activity in zymograms using CMC as substrate (Fig. 1b).

The mode of action of rGH5CelA and rGH5CelB was determined by using CMC, barley β-glucan, and different cello-oligosaccharides as substrates and performing TLC to assess the hydrolysates (Fig. 2). The TLC results showed that rGH5CelA hydrolyzed G4 by releasing G2 as a product (Fig. 2a). In contrast, rGH5CelB hydrolyzed G3 by releasing G1 and G2, as well as G4, which was hydrolyzed to G2 (Fig. 2b). In addition, a faint spot of G3 was observed after G4 hydrolysis.

The hydrolysate profile of barley β-glucan by rGH5CelA exhibited oligomers > 6, whereas cellobiose and oligomers of 3–6 glucose residues appeared at early reaction stages. During hydrolysis, degree of polymerization of oligomers > 6 decreases, whereas oligomers of 2–4 glucose residues increased (Fig. 2c). On the other hand, rGH5CelB presented a different hydrolysis profile on barley β-glucan. In the early phase of hydrolysis, the main products were oligomers of 3–4 glucose residues and to a lesser extent glucose, cellobiose, and oligomers of 5 glucose residues. During the course of the reaction, the oligomers with higher degree of polymerization decreased, whereas glucose, cellobiose, and triose were gradually accumulated (Fig. 2d).

The main reaction products of CMC hydrolysis were cello-oligosaccharides with a high degree of polymerization, whereas throughout the reaction, cello-oligosaccharides < 4 were accumulated (Fig. 2e).

Enzyme characterization

The optimal temperature and pH of each enzyme were tested to further characterize their activities under different reaction conditions. We used *p*NPC as substrate, because of its highest levels of activity (Fig. 3). The enzyme rGH5CelA showed the highest cellobiohydrolase activity at pH 4.5–5 ($P < 0.0001$) and 35 °C ($P < 0.0001$), with more than 60% of activity at temperatures ranging from 30 to 35 °C (Fig. 3a–c). On the other hand, rGH5CelB showed the highest cellobiohydrolase activity at pH 5 ($P < 0.0001$) and 45 °C ($P < 0.0001$), with more than 60% of activity at pH range from 5 to 6 and temperatures from 40 to 50 °C. At 40 °C, the enzyme still retained more than 20% of cellobiohydrolase activity after 16 h (Fig. 3a, b, and d).

A kinetic analysis determined on *p*NPC showed a hyperbolic kinetic profile for both enzymes (fitting Michaelis-Menten function). The K_M and V_{max} of rGH5CelA were 1.8 mM and 120.1 μmol min^{-1} mg^{-1}, respectively. In addition, the turnover number for this enzyme was 81.9 S^{-1}, and thus, the catalytic efficiency (K_{cat}/K_M) was 45.7 $mM^{-1}S^{-1}$ (Fig. 3e). For rGH5CelB, the K_M was 3.2 mM and the V_{max} was 50.47 μmol min^{-1} mg^{-1}, whereas the turnover number for

Table 1 Substrate specificity of rGH5CelA and rGH5CelB

Substrate	Activity	Glycosidic bond	rGH5CelA (IU/mg)	rGH5CelB (IU/mg)
Carboxymethyl cellulose	Endoglucanase	(β-1,4) Glc	ND	0.1 ± 0.01
Avicel	Exoglucanase	(β-1,4) Glc	ND	ND
Bacterial cellulose	Exoglucanase	(β-1,4) Glc	ND	ND
PASC	Exoglucanase	(β-1,4) Glc	ND	ND
Beechwood xylan	Xylanase	(β-1,4) Xyl	ND	ND
Galactomannan	Endomannanase	(β-1,4) Man	ND	ND
Barley β-glucan	β-Glucanase/lichenase	(β-1,4/β-1,3) Glc	0.95 ± 0.02	1.3 ± 0.4
Laminarin	Laminarase	(β-1,3) Glc	ND	ND
*p*NP-cellobioside	Cellobiohydrolase	β-Glc	42.3 ± 2.0	29.2 ± 0.4
*p*NP-glucopyranoside	β-Glucosidase	β-Glc	ND	ND
*p*NP-xylopyranoside	Xylosidase	β-Xyl	ND	ND

ND not detected, *IU* international units (μmol of product/min of reaction), *Glc* glucose, *Xyl* xylose, *Man* mannose

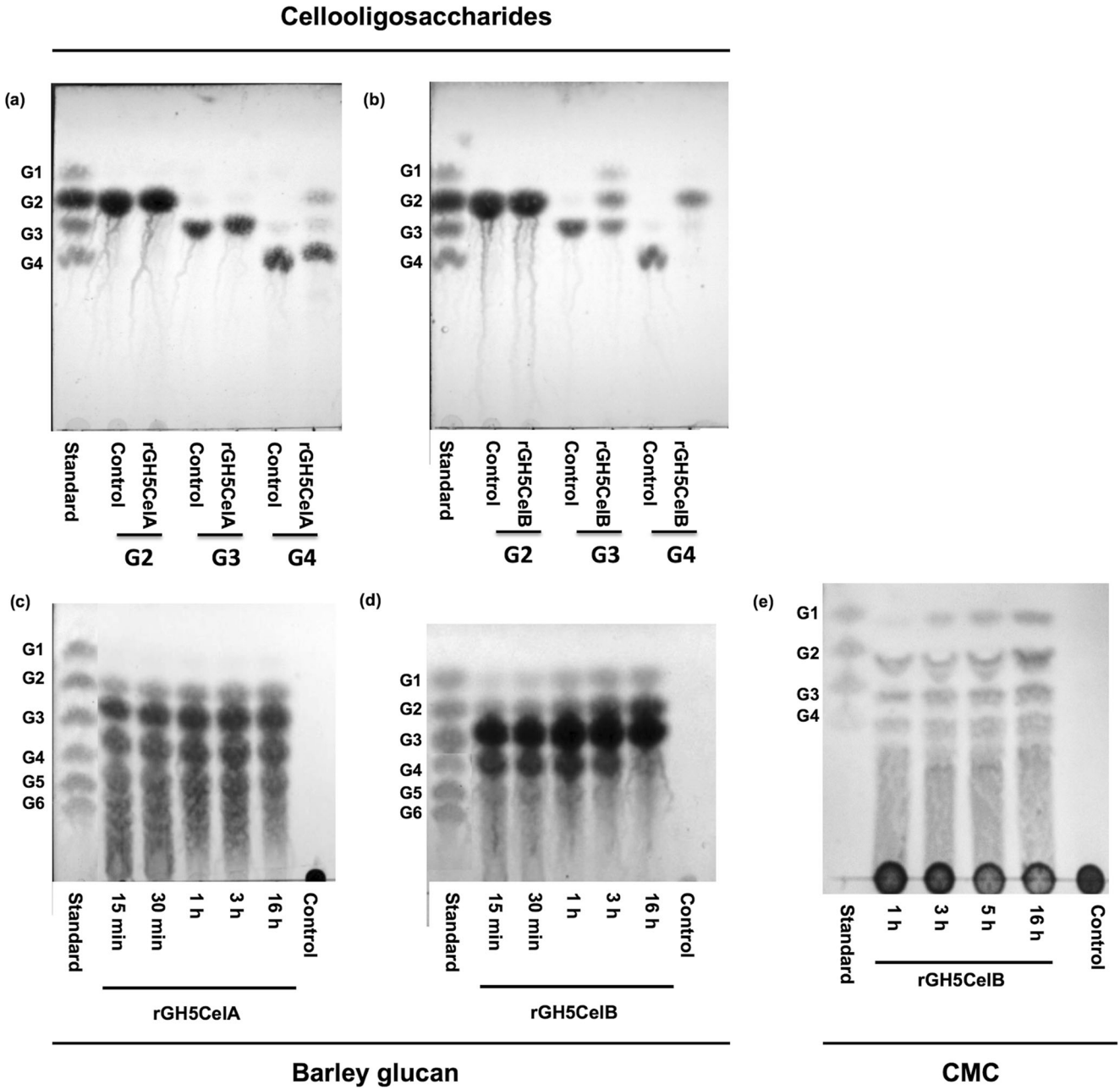

Fig. 2 Thin layer chromatography analysis of hydrolysis products of rGH5CelA and rGH5CelB. Hydrolysis products from cello-oligosaccharides by (**a**) rGH5CelA and (**b**) rGH5CelB. Time course degradation of barley β-glucan by (**c**) rGH5CelA and (**d**) rGH5CelB. Time course degradation of CMC by rGH5CelB (**e**). Standard marker for glucose (G1), cellobiose (G2), cellotriose (G3), cellotetraose (G4), cellopentaose (G5), and cellohexaose (G6). Control: substrate without enzyme

this enzyme was 33.4 S^{-1}. Thus, the catalytic efficiency (K_{cat}/K_M) was 10.50 $mM^{-1}S^{-1}$ (Fig. 3f).

In addition, we determined the activity profile of rGH5CelB using CMC (Fig. 4a, b). Although the relative activity was very low with this substrate (Table 1), we performed the characterization because the enzyme was active in zymograms and because TLC analysis evidenced accumulation of low molecular weight cellodextrins (GP < 4). In this case, rGH5CelB showed the highest activity at pH 5 and 5.5 and an optimal temperature of 45 °C, with more than 60% activity in a temperature range between 30 and 50 °C. The enzyme retained more than 90% and 80% activity for 2 h at 40 °C and 45 °C, respectively. After 16 h the activity declined to 40% at 40 °C and lost its activity at 45 °C. The enzyme was completely inactive after 2 h at 50 °C (Fig. 4c). The kinetic analysis on CMC showed a sigmoidal profile (non-Michaelis-Menten behavior), thus indicating cooperative binding (Fig. 4d). For rGH5CelB, the K_{half} was 704.3 mM and the V_{max} was

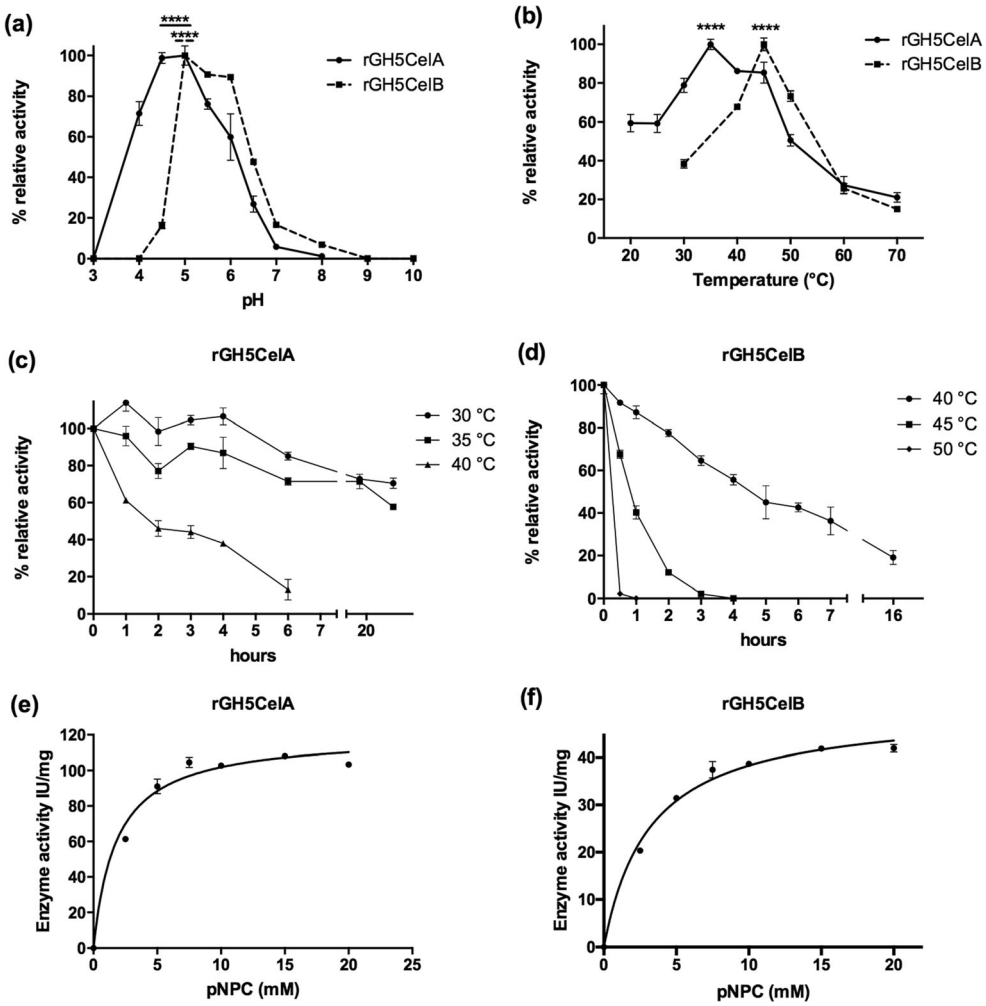

Fig. 3 Cellobiohydrolase profile activity of rGH5CelA and rGH5CelB. Optimal pH condition (**a**), temperature (**b**), thermal stability (**c**, **d**), and kinetic analysis (**e**, **f**) were evaluated using *p*NPC as substrate. The results correspond to mean and standard deviations of technical triplicates. Two independent biological replicate assays were performed with equivalent results. *IU* international units: μmol/min. Enzymatic activity data were analyzed for statistical significance with a one-way ANOVA and post-test Tukey's multiple comparison. ****$P <$ 0.0001

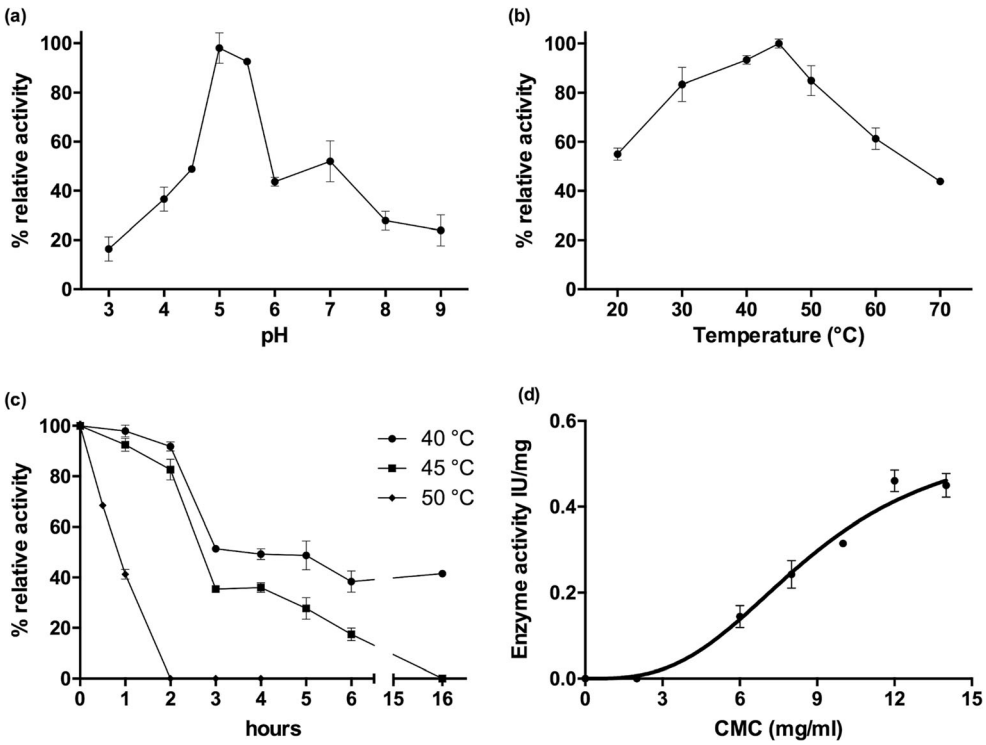

Fig. 4 Endoglucanase profile activity of rGH5CelB. Optimal pH condition (**a**), temperature (**b**) thermal stability, (c) and kinetic analysis (**d**) were evaluated using CMC as substrate. The results correspond to mean and standard deviations of technical triplicates. Two independent biological replicate assays were performed with equivalent results. *IU* international units: μmol/min

0.57 μmol min^{-1} mg^{-1}, with a Hill coefficient (η_H) of 3.015. These results suggest the possible presence of secondary binding sites (SBSs) in rGH5CelB for this substrate.

Phylogenetic analysis and structure prediction of GH5CelA and GH5CelB

Supplemental Fig. S2 illustrates a phylogenetic tree constructed based on protein sequences for the most abundant cellulolytic family (GH5) identified in the gut microbiome analysis of two Argentinean termite species (*C. fulviceps* and *N. aquilinus*). In general, the protein sequences tended to group together according to the termite species from which they were obtained. Likewise, in every GH family, several protein sequences retrieved from *C. fulviceps* and *N. aquilinus* were closely related between them. Some of these clusters included reference sequences of bacterial or fungal origin, although these reference sequences were more distantly related.

The phylogenetic analysis, with the selected subfamilies, revealed that GH5CelA and GH5CelB grouped together and with sequences of subfamily 25, bootstrap value (BS) of 50 (Fig. 5a). Because the BS is not high, this value neither allows us to confirm that both enzymes are part of subfamily 25 nor to ensure that they are part of a different subfamily. Subsequently, 80 amino acid sequences of the GH5 subfamily 25 together with GH5CelA and GH5CelB were used to construct a matrix. This phylogenetic analysis revealed that GH5CelA and GH5CelB grouped together to each other but separately from the other GH5 subfamily 25 amino acid sequences (Fig. 5b).

Regarding the structural modeling, the model accuracy yields a C-score of − 0.13 and a TM-score of 0.70. In addition, according to this analysis, the GH5CelA model displays a structural similarity of 1 Å (RMSD) with the endoglucanase 5 1CEN from *Clostridium thermocellum*. GH5CelA showed a common (β/α)(8) TIM barrel fold of GH5 family. Furthermore, two glutamic acid residues E146 and E288 and the histidine H204 probably involved in the catalytic site were identified in the catalytic triad (Fig. 6a–c). Concerning the GH5CelB molecular model, this enzyme also exhibits an 8-fold TIM barrel [(β/α)8] molecular structure, with a C-score of 0.8 and a TM-score of 0.8. In addition, GH5CelB showed a high structural similarity (an RMSD of 0.99 Å) with the endoglucanase 5 1CEN from *C. thermocellum* (Fig. 6a–c). The glutamic acid residues E130 and E272 and the histidine H188 were identified as those involved in the catalytic mechanism (Fig. 6a).

Compared with the canonical TIM barrel fold, the structures of GH5CelA and GH5CelB have an additional loop region consisting of residues 208–258 and residues 192–242, respectively. The 51 residue subdomain is well formed and contains 4α-helices, inserted between the core β strand 6 and helix 6 of the barrel, near to the substrate-binding cleft. This subdomain extends the top of the barrel on one side creating a deep substrate-binding cleft. The structural similitude was found with the crystal structure of the endoglucanase CelC from *C. thermocellum* (PDB 1CEC) (Dominguez et al. 1995) (Supplemental Fig. S3).

A comparison of both models revealed that GH5CelA and GH5CelB outline some differential features that would support the distinctive enzymatic behavior of both enzymes. Indeed, their active sites share six canonical residues (Fig. 6b), but the site of GH5CelB is a wide and deep cleft that is also lined for aromatic residues locates: H188, Y190, H89, W306, and F276. In the case of GH5CelA, the aromatic residues H204, Y206, H105, W322, and Y292 are located along the active site cleft and provide a favorable platform for cellulose binding (Fig. 6b–c).

The subsite-binding cleft of GH5CelA clearly involves aromatic residues W322 (− 1), W218 (+ 1) and potentially F328 (− 2) and Y182 (+ 2). The binding cleft of GH5CelB is sufficiently long to bind four or possibly five glycosyl units. Besides, the cellotriose-bound structure clearly identifies three of the glucose-binding subsites at W322 (− 1), H89 (− 1), W202 (+ 1), and H193 (+ 2), as well as the potential fourth W306 (− 2) (− 1) and subsite (− 3) sustained by ionic interactions in GH5CelB (Fig. 6a).

According to measurements of the active site, the cleft of GH5CelA presented a smaller area and volume (759.87 $Å^2$ and 1107.29 $Å^3$) than the clef of GH5CelB (1105.05 $Å^2$ and 1729.35 $Å^3$). A comparison of the surface electrostatic potential of substrate-binding cleft site between the two enzymes revealed a significant increase in negatively charged Y292 and D327 residues in GH5CelA in relation to GH5CelB. Other residues that contribute for the net negative charged surface of GH5CelB (in relation to GH5CelA) were the polar K13 and R18 and neutral N165, F276, and G311. Thus, the negative electrostatic potential of GH5CelB is consistent with the acidic pI (6.0) of the protein (Fig. 6d–e and g–h).

The best scored ligand-binding site identified in GH5CelA was delimited by residues 29, 105, 106, 145, 146, 206, 213, 218, 288, 322, and 328, similar to PDB 3AZR, which is a crystallized GH5 in a complex with cellobiose (Fig. 6f). On the other hand, the high-rated ligand-binding site identified in GH5CelB was defined by residues 13, 89, 90, 129, 130, 190, 272, and 306, similar to the catalytic site at PDB, a crystallized GH5 in a complex with cellotriose (Fig. 6i). Furthermore, for GH5CelB, a potential SBS was outlined in structural similitude with a *Candida albicans* cell wall-associated exo-β-1,3-glucanase (Exg; PDB 2PC8) (Patrick et al. 2010). Delineated by residues 180, 181, 183, 265, 266, and 267, the GH5CelB SBS does not include a canonical tryptophan located at the

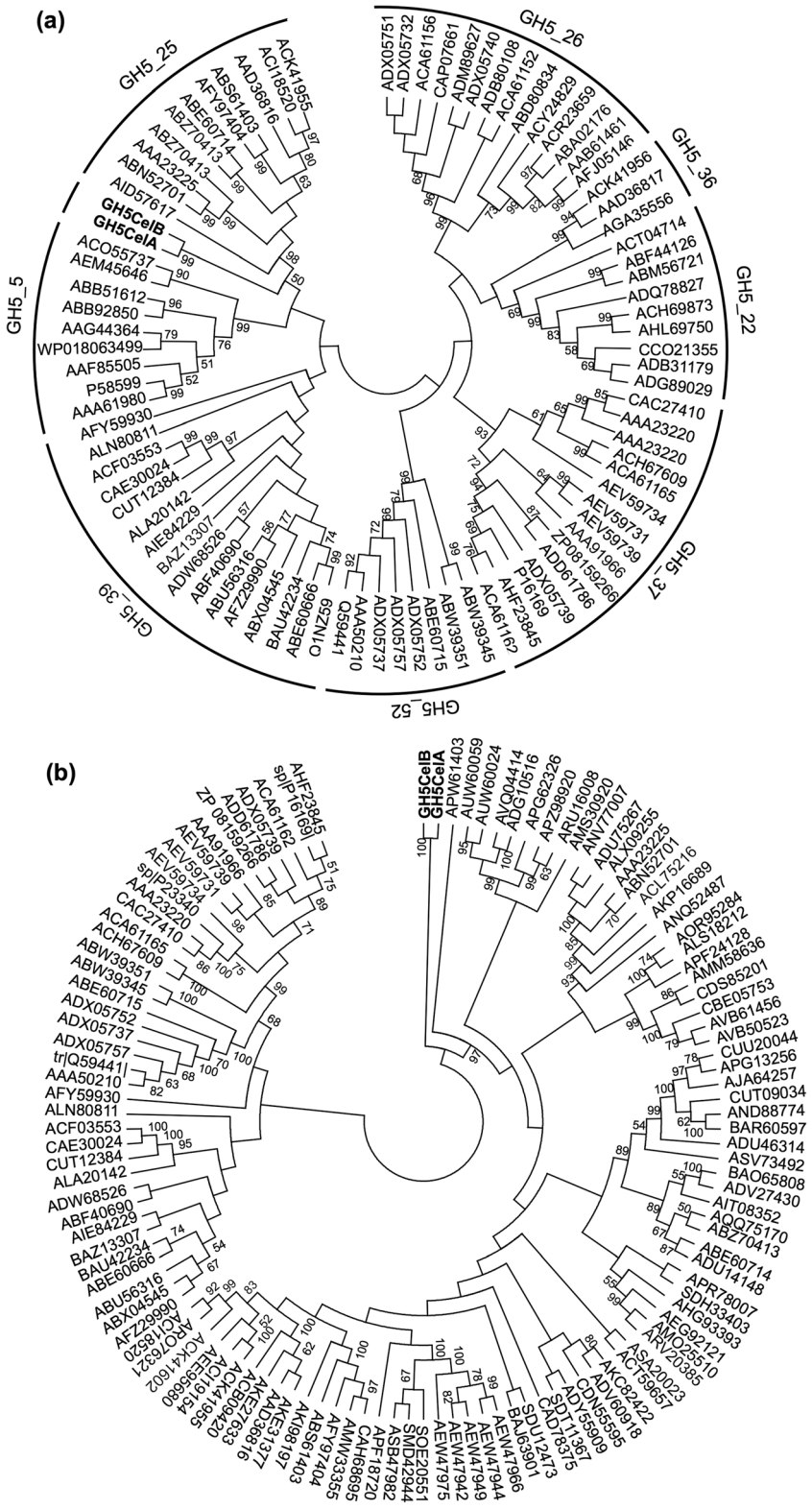

Fig. 5 Maximum likelihood tree obtained with MEGA v6 using GH5CelA and GH5CelB amino acid sequences and protein sequences of different subfamilies of the GH5 family (**a**) and amino acid sequences of subfamily 25 deposited in the CAZY database (**b**)

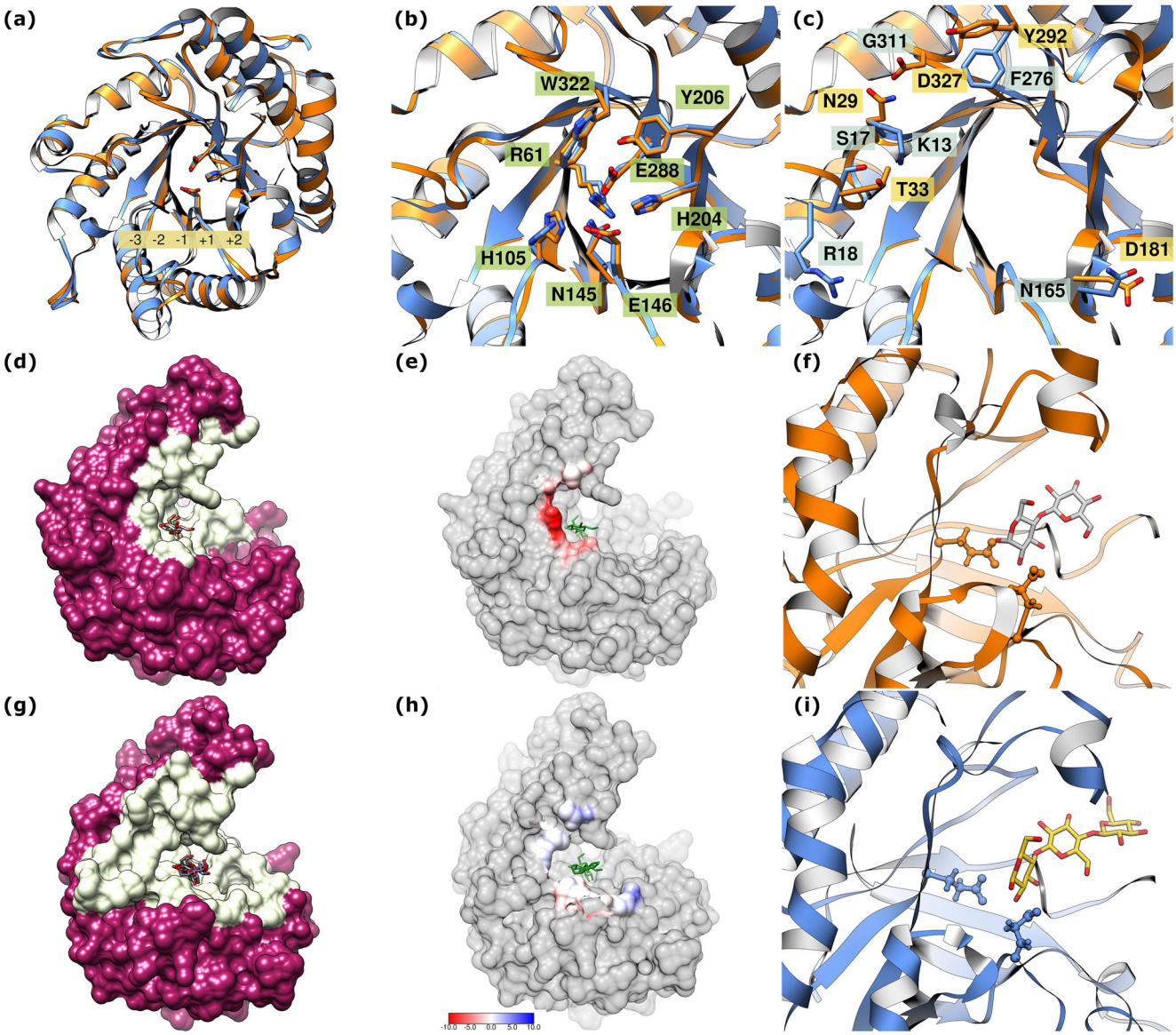

Fig. 6 Overall structure and active site of GH5CelA and GH5CelB. The superimposed models of GH5CelA and GH5CelB are colored in orange and blue, respectively (**a**, **b** and **c**). Close-up stereoview of the substrate-binding cleft displaying the locations of glucose-binding subsites (**a**). The common amino acid residues found in the active site (which are predicted to interact with the ligands) (**b**), orange and blue boxes display differential amino acids lining the active site in in accordance with the colors assigned to the protein (**c**). Stereoview of the molecular surface shows the active site binding cleft (white) of GH5CelA complex with a cellobiose molecule (**d**). The electrostatic potentials on the surface of active site binding cleft of GH5CelA (red and blue colors) indicate from − 10 to 10 kcal/mol, respectively (**e**). Close-up view of GH5CelA shows the substrate-binding cleft in complex with a cellobiose molecule (**f**). Stereoview of the molecular surface shows the active site binding cleft (white) of GH5CelB complex with a cellotriose molecule (**g**). Active site and electrostatic potential surface charge of GH5CelB (**h**). Substrate-binding cleft of GH5CelB in complex with a cellotriose molecule (**i**)

β7 N-terminal end of some GH5 subfamily 9 enzymes (like Exg 2PC8).

Discussion

In this work, we characterized the enzymatic activities of GH5CelA-CO and GH5CelB-CO retrieved from termite gut microbiome.

To date, most GH obtained from shotgun metagenomic approach have been acquired through synthetic metagenomics (Cheng et al. 2016; Maruthamuthu et al. 2017; Patel et al. 2019; Wierzbicka-Woś et al. 2019). However, most of these enzymes derived from culturable microorganisms. We thus focus on two endoglucanases derived from nonculturable microorganisms, since this approach represents a powerful source for discovering novel enzymes.

The profile of sugars released from the hydrolysis of different cello-oligosaccharides used as substrates was different for the two enzymes. rGH5CelA produced oligomers larger than six glucose residues, while rGH5CelB produced smaller oligomers of 3–4 glucose residues. Both enzymes mostly attack internal glycosidic bonds. Similar results were observed for other GH5 endo-β-1,4-glucanases (Hendricks et al. 1995; Nacke et al. 2012; Schinsky et al. 2000; Vogel 2008).

The biochemical characterization using *p*NPc as substrate revealed that rGH5CelA is active at a moderate range of pH and moderate to low temperature (with pH 4–5, 25–40 °C as optimal conditions). At these conditions, the enzyme remained active (60%) after 24 h of incubation at 30 and 35 °C. On the other hand, the optimal values of GH5CelB activity were in a range of pH 5–6 and temperatures from 40 to 50 °C. The endoglucanase optimum activity of rGH5CelB was at pH 5 and in a range of temperature, maintaining more than 50% of relative activity at 40 °C. The optimum pH and temperature values obtained for rGH5CelA and rGH5CelB were within the range described by other authors for GH5 cellobiohydrolases and endoglucanases (Hartmans et al. 2006; He et al. 2019; Kim et al. 2018; Liu et al. 2011; Park et al. 2011; Song et al. 2017).

Regarding the enzymatic activity, rGH5CelA (~ 42 IU/mg) and rGH5CelB (~ 29 IU/mg) presented similar range activity values than those reported for some purified microbial endoglucanases and cellobiohydrolases such as Rucel5B and CBH (Lee et al. 2011; Yamane et al. 1970). Both enzymes presented higher activities than fungal cellobiohydrolases such as Exo2b (Hendricks et al. 1995), cbh1 from *Trichoderma virens* (Wahab et al. 2018), CBH from *Fomitopsis pinicola* KMJ812 (Shin et al. 2010), and CBH2 from *Schizophyllum commune* (Kondaveeti et al. 2019).

GH5CelA and GH5CelB had efficient catalytic activities towards substrates with a β-1,4-glycosidic bonds (barley β-glucan and *p*NP-cellobioside), and GH5CelB also showed activity against CMC. This finding suggests that the catalytic activity of GH5CelA and GH5CelB is specific for the β-1,4-glucan linkage in the amorphous region of the cellulose and hemicellulose.

Both enzymes hydrolyzed *p*NP-cellobioside, which is often used as a model substrate for exo-β-glucanases, although with less catalytic efficiency. However, differentiation between exo- and endo-glucanases based on small soluble substrates does not appear to be a very useful method (Claeyssens and Henrissat 1992; Rahman et al. 2002). In addition, the enzymes produced oligomers of 2 and 3 glucose residues from tetraose (GH5CelA) and glucose and cellobiose from triose and tetraose (GH5CelB), which suggests that it is not a CBH (cellobiohydrolase), as assumed by the enzyme nomenclature. Indeed, the CBH produces only glucose or cellobiose in the hydrolysates of cellulose (Kim 1995; Lynd et al. 2002). Thus, the different products obtained from the hydrolysis of CMC by GH5CelB such as glucose, cellobiose, cellotriose, and cellotetraose support the idea that the enzyme may be an endoglucanase. Therefore, classification of GH5CelA and GH5CelB as endoglucanase is valid.

In general, cellulose-binding domains (CBM) are involved in the degradation of crystalline cellulose, and a cellulase without CBM can hydrolyze only the amorphous form of cellulose (Bolam et al. 1998; Lee and Lee 2014; Lynd et al. 2002). In the sequence analysis of both enzymes, no CBM was identified, which could explain the lack of enzymatic activity in crystalline cellulose forms such as Avicel, PASC, and BMCC.

The rGH5CelA and rGH5CelB kinetic profiles on *p*NPC under optimal pH and temperature were fitted to Michaelis-Menten function. Instead, rGH5CelB showed a sigmoidal kinetic profile on CMC as substrate, thus indicating positive cooperative binding. This result suggests the presence of secondary binding sites (SBSs). For small substrates, binding to the SBS and the active site appears to be independent of each other. By contrast, for larger substrates, cooperative binding occurs and leads to improved hydrolysis of these substrates (Ludwiczek et al. 2007). In several GHs, the occurrence of SBSs is a general strategy to compensate the absence of additional CBMs (Cuyvers et al. 2012; Cockburn and Svensson 2013).

The results of the amino acid sequence analysis showed high identity with sequences of enzymes of the GH5 family previously identified in the metagenomic analysis of the gut microbiota of *Nasutitermes* genus (Warnecke et al. 2007). Despite the high level of similarity in the catalytic domain with other GHs of the GH5 family, the phylogenetic analysis revealed that GH5CelA and GH5CelB are phylogenetically separated from the subfamilies described so far. This finding indicates that they could be part of a new subfamily not described yet.

Through a modeling analysis, we detected structural differences between the active site and cellulose-binding cleft of GH5CelA and GH5CelB that could explain the different behaviors observed between both enzymes. The structural modeling revealed that both enzymes have a typical barrel structure (β/α) 8 (TIM barrel) consisting of a core composed of eight parallel β-sheets surrounded by eight α-helices, where the active site has a cleft topology (Delsaute et al. 2013; Dominguez et al. 1995). This analysis allowed us to identify the key residues that are part of the active site.

A unique feature of GH5CelA and GH5CelB is an additional helical subdomain of 51 amino acid residues inserted between β-strand 6 and α-helix 6, on the cover upper side to the substrate-binding cleft. Although this structural element is commonly encountered in several GH5 enzymes, we have found a single structural homolog at the crystallized endoglucanase CelC from *C. thermocellum* (PDB 1CEC) (Dominguez et al. 1995, 1996). The endoglucanase EglA

(PDB 3AYR, subfamily 5) from *Piromyces rhizinflata* presents a loop containing a disulfide bond adjacent to the active site (Tseng et al. 2011). This loop with its high flexibility seems to have an important role in substrate binding and hydrolysis (Dominguez et al. 1996) as well as to be related to thermal stability (Badieyan et al. 2012). Studies of thermostable GH5 cellulases, *Tm*_Cel5A (subfamily 25) from *Thermotoga maritima* and *Bacillus subtilis* 168, indicate that the presence of additional structural elements at the N terminus contributes to conformational stability of the proteins (Pereira et al. 2010; Santos et al. 2012). In CelE1 (PDB 4M1R), the extended loop promotes additional intramolecular contacts that might be related to the significant thermal stability (Alvarez et al. 2013). Although to understand the exact role of this subdomain crystallized studies would be needed.

Generally, the hydrolytic mechanism of GH5 family members relies on two strictly conserved glutamate residues, the catalytic acid/base and the nucleophile (Badieyan et al. 2012; Dominguez et al. 1995; Henrissat et al. 1995; Zhang et al. 2006). However, the catalytic triad (Glu-His-Glu) located in the substrate-binding site has shown to be relevant for catalysis and probably the histidine function as an intermediate for the electron transfer network between the typical Glu-Glu catalytic modules (Wierzbicka-Woś et al. 2019; Zheng et al. 2012). In addition, four other amino acids (Arg, Asn, His, Trp) located in the substrate-binding pocket are highly conserved in the GH5 family (Hilge et al. 1998).

All residues of GH5CelA (E146 (catalytic acid/base), E288 (nucleophile), R61, N145, H204, and W322) and GH5CelB (E130 (catalytic acid/base), E272 (nucleophile), R45, N129, H188, and W306), identified by I-TASSER, were located in the canonical position previously reported in the closest structural neighbors PDBs 4HU0. The identification of these six strictly conserved residues in GH family 5 members provides additional evidence indicating that GH5CelA and GH5CelB are glycosyl hydrolases of family 5 (Hilge et al. 1998; Pereira et al. 2010; Wang et al. 1993).

One of the major features of GH5 cellulases is the large open active site cleft located along the catalytic face of the (β/α)8 barrel. The cleft contains several distinct glucose-binding sites and is large enough to accommodate a single strand of cellulose (Caines et al. 2007; Dominguez et al. 1996; Ducros et al. 1995; Gloster et al. 2007; Lo Leggio and Larsen 2002).

In GH5CelA and GH5CelB, eight strictly conserved residues that coincide with the reports for the GH5 family members were identified (Bianchetti et al. 2013; Lo Leggio and Larsen 2002; Wang et al. 1993). In the substrate-binding site, some aromatic residues of GH5CelA (H204, Y206, H105, W322, and Y292) and GH5CelB (H188, Y190, H89, W306, and F276), as well as polar residues, are located along the active site cleft and provide a favorable platform for cellulose binding through stacking (Brandl et al. 2001; Ramirez-Gualito et al. 2009; Vyas et al. 1991) (Fig. 6a–b). These residues may play the same roles in substrate recognition in GH5 family (Wang et al. 1993).

By TLC assay, we could describe the possible interactions between the subsites of both enzymes and cellobiose, cellotriose, and cellotetraose. GH5CelA was unable to hydrolyze cellobiose or cellotriose, which suggests that cellobiose joins subsites − 2 to − 1, whereas cellotriose joins subsites − 2 to + 1 of the enzyme, by-passing the catalytic machinery in a non-productive manner bind in a similar way (− 2 to − 1 subsites). Cellotetraose was partially hydrolyzed by GH5CelA producing only cellobiose. According to these results, GH5CelA hydrolyzed products indicate that the carbohydrate-binding cleft may be composed of four subsites for glucopyranose units, which should correspond to the − 2, − 1, + 1 and + 2 subsites as inferred from a comparison with other family GH5 enzyme substrate complexes (PDBs 3AZR, 1ECE, and 5D8Z) (Rahman et al. 2002; Sakon et al. 1996; Wu et al. 2011) (Fig. 7a).

Furthermore, GH5CelB, like GH5CelA, was unable to hydrolyze cellobiose, but hydrolyzed cellotriose with formation of cellobiose and glucose. This suggests that cellotriose binds to the − 2 to + 1 and − 1 to + 2 subsites. Cellotetraose was totally degraded by GH5CelB. Since cellotetraose bound to the high affinity site (− 3 to − 1 subsites), then glucose and cellotriose would be the only products formed. However, hydrolysis of cellotetraose produced cellobiose, thus demonstrating substantial binding to the − 2 to + 2 subsites (Fig. 7b).

The interactions between cellotriose and the − 3 to − 1 subsites of GH5CelB are similar to the interactions observed in enzyme substrate complexes of EngD15 (PDB 3NDZ), CelAcd16 (PDB 3AYS), and Tm_Cel5A (PDB 3MMU) (Bianchetti et al. 2013; Pereira et al. 2010; Tseng et al. 2011). However, a structural comparison with metagenome-derived Cel5A (PDB 4HU0) and SdGluc5_26A (PDB 5A8O) containing a cellotetraose molecule bound to subsites − 2 to + 2 (Lafond et al. 2016; Telke et al. 2013).

GH5 family cellulases hydrolytically cleave the glycosidic linkages of their substrates using the double-displacement retaining mechanism. Many GH5 are described to be purely hydrolytic enzymes, while some are reported to be transglycosylases. Several examples of transglycosylation activity have been reported in GH5 endo-1,4-β-glucanases (Dingee and Anton 2010; Delsaute et al. 2013; Dutoit et al. 2019). In this study, the presence of a faint cellotriose spot in the hydrolysis of cellotetraose could result in the ability of GH5CelA and GH5CelB to act as a transglycosylase.

Thus, the analysis of the structure of the substrate and the amino acid residues in the active site of the GH5 enzymes demonstrates that different properties of the active sites, even by one distinct amino acid, can lead to different substrate-binding specificities (Schagerlof et al. 2007; Tailford et al. 2009; Vlasenko et al. 2010).

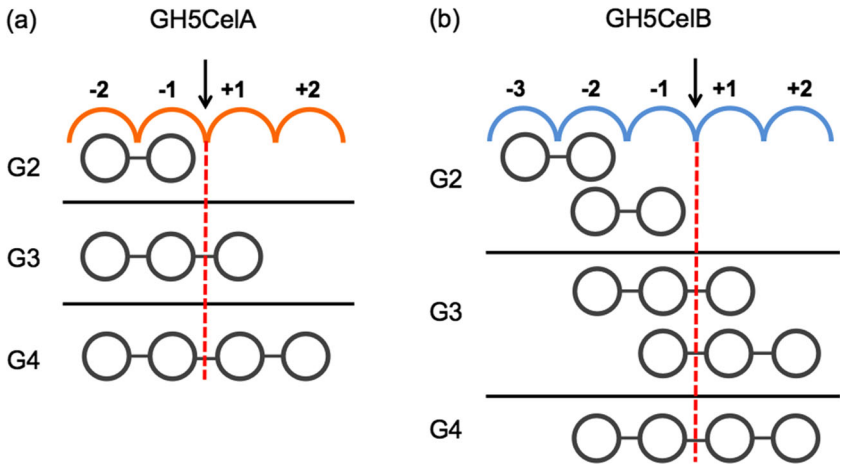

Fig. 7 The possible interactions between the subsites of GH5CelA and GH5CelB. Schematic representation of gluco-oligosaccharide accommodation into the active site of GH5CelA (**a**) and GH5CelB (**b**). Bond cleavage occurs between subsite − 1 and + 1. G2, cellobiose; G3, cellotriose; G4, cellotetraose

Nowadays, endoglucanases have become of special interest due to their possible industrial applications. Same examples are the use of these enzymes in bioethanol production by conversion of lignocellulosic biomasses into fermentable sugars, in food industry for removal of β-glucan to reduce viscosity and increase the filtration rate in beer production, and in animal feed industry by improvement in digestibility of diets based on soybean meal or barley used in chickens and pigs (Anwar et al. 2014; Kim et al. 2013; Pan et al. 2018; Sun et al. 2012). Their ability to produce cello-oligosaccharides from biomass is also a feature of biotechnological interest. Researchers have reported these oligosaccharides as potential prebiotics that could regulate intestinal bacterial composition, modify fermentative processes, and possibly improve host health (Cheng et al. 2017; Jiao et al. 2014; Zhao et al. 2012). Due to the β-1,4-glucosidic linkage between glucose monomers, cello-oligosaccharides are resistant to host digestive enzymes and in turn serve as a substrate for intestinal microbiota (Hasunuma et al. 2011; Kido et al. 2016; Otsuka et al. 2004). The endo-β-1,4-glucanases characterized in this work could be used in the development of value-added products. Particularly, rGH5CelA could be used for their production of cello-oligosaccharides of different degrees of polymerization. Conversely, rGH5CelB could be used in an enzymatic complex for the production of lignocellulosic bioethanol, since this enzyme produces glucose as a hydrolysis product.

Acknowledgments MPS, EC, and PMT acknowledge CONICET as career research members. EBG and RMDV belong to the National Institute Agriculture Technology (INTA). The authors are grateful to Dr. Julia Sabio y García for linguistic improvement of the manuscript.

Authors' contributions EBG conceived the experiments, analyzed the results, and contributed to manuscript writing. RMDV conceived and analyzed the protein modeling results and contributed to manuscript writing. MAS analyzed the results. MPS conceived the experiments and contributed to manuscript writing. EC contributed to manuscript writing. PMT conceived and conducted the experiments, analyzed the results, and was responsible for writing the manuscript. All authors reviewed the manuscript.

Funding information This study was funded by grants from the Instituto Nacional de Tecnología Agropecuaria (INTA) (PNAIyAV-1130034) and Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) Proyectos de Investigación Científica y Tecnológica (PICT) 2013 No.1454 (Argentina) and 2018 No. 4149.

Compliance with ethical standards

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

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Alvarez TM, Goldbeck R, dos Santos CR, Paixao DAA, Goncalves TA, Cairo JPLF, Ferreira Almeida R, Oliveira Pereira I, Jackson G, Cota J, Buchli F, Citadini AP, Ruller R, Polo CC, Oliveira Neto M, Murakami MT, Squina FM (2013) Development and biotechnological application of a novel endoxylanase family GH10 identified from sugarcane soil metagenome. PLoS One 8(7):e70014. https://doi.org/10.1371/journal.pone.0070014

Anwar Z, Gulfraz M, Irshad M (2014) Agro-industrial lignocellulosic biomass a key to unlock the future bio-energy: a brief review. J Radiat Res Appl Sci 7(2):163–173

Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol Biol 12(1):186

Badieyan S, Bevan DR, Zhang C (2012) Study and design of stability in GH5 cellulases. Biotechnol Bioeng 109(1):31–44

Bastien G, Arnal G, Bozonnet S, Laguerre S, Ferreira F, Fauré R, Henrissat B, Lefèvre F, Robe P, Bouchez O, Noirot C, Dumon C, O'Donohue M (2013) Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. Biotechnol Biofuels 6:78. https://doi.org/10.1186/1754-6834-6-78

Ben Guerrero E, Arneodo J, Bombarda Campanha R, Abrão de Oliveira P, Veneziano Labate MT, Regiani Cataldi T, Cataldi A, Labate CA, Martins Rodrigues C, Talia P (2015) Prospection and evaluation of cellulolytic and hemicellulolytic enzymes using untreated and pretreated biomass in two argentinean native termites. PLoS One 10:e0136573. https://doi.org/10.1371/journal.pone.0136573

Bianchetti CM, Brumm P, Smith RW, Dyer K, Hura GL, Rutkoski TJ, Phillips GNJ (2013) Structure, dynamics, and specificity of endoglucanase D from *Clostridium cellulovorans*. J Mol Biol 425(22):4267–4285. https://doi.org/10.1016/j.jmb.2013.05.030

Bolam DN, Ciruela A, McQueen-Mason S, Simpson P, Williamson MP, Rixon JE, Boraston A, Hazlewood GP, Gilbert HJ (1998) *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. Biochem J 331(3):775–781

Brandl M, Weiss MS, Jabs A, Suhnel J, Hilgenfeld R (2001) C-H…pi-interactions in proteins. J Mol Biol 307:357–377

Caines ME, Vaughan MD, Tarling CA, Hancock SM, Warren RA, Withers SG, Strynadka NC (2007) Structural and mechanistic analyses of endo-glycoceramidase II, a membrane-associated family 5 glycosidase in the Apo and GM3 ganglioside-bound forms. J Biol Chem 282:14300–14308

Cheng J, Huang S, Jiang H, Zhang Y, Li L, Wang J, Fan C (2016) Isolation and characterization of a non-specific endoglucanasefrom a metagenomic library of goat rumen. World J Microbiol Biotechnol 32(12):12. https://doi.org/10.1007/s11274-015-1957-4

Cheng W, Lu J, Li B, Lin W, Zhang Z, Wei X, Sun C, Chi M, Bi W, Yang B, Jiang A, Yuan J (2017) Effect of functional oligosaccharides and ordinary dietary fiber on intestinal microbiota diversity. Front Microbiol 8:1–11

Claeyssens M, Henrissat B (1992) Specificity mapping of cellulolytic enzymes: classification into families of structurally related proteins confirmed by biochemical analysis. Protein Sci 1:1293–1297

Cockburn D, Svensson B (2013) Surface binding sites in carbohydrate active enzymes: an emerging picture of structural and functional diversity. Carbohydr Chem 39:2014–2221

Cuyvers S, Dornez E, Delcour JA, Courtin CM (2012) Occurrence and functional significance of secondary carbohydrate binding sites in glycoside hydrolases. Crit Rev Biotechnol 32(2):93–107

Delsaute M, Berlemont R, Dehareng D, Van Elder D, Galleni M, Bauvois C (2013) Three-dimensional structure of RBcel1, a metagenome-derived psychrotolerant family GH5 endoglucanase. Acta Crystallogr Sect F 69(8):828–833

Dingee J, Anton AB (2010) The kinetics of *p*-nitrophenyl-β-D-cellobioside hydrolysis and transglycosylation by *Thermobifida fusca* Cel5Acd. Carbohydr Res 345:2507–2515

Dominguez R, Souchon H, Spinelli S, Dauter Z, Wilson KS, Chauvaux S, Béguin P, Alzari PM (1995) A common protein fold and similar active-site in 2 distinct families of beta-glycanases. Nat Struct Biol 2(11):983–989

Dominguez R, Souchon H, Lascombe M, Alzari PM (1996) The crystal structure of a family 5 endoglucanase mutant in complexed and uncomplexed forms reveals an induced fit activation mechanism. J Mol Biol 257:1042–1051

Ducros V, Czjzek M, Belaich A, Gaudin C, Fierobe HP, Belaich JP, Davies GJ, Haser R (1995) Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. Structure 3:939–949

Dutoit R, Delsaute M, Collet L, Vander Wauven C, Van Elder D, Berlemont R, Richel A, Gallenib M, Bauvoisa C (2019) Crystal structure determination of *Pseudomonas stutzeri* A1501 endoglucanase Cel5A: the search for a molecular basis for glycosynthesis in GH5_5 enzymes. Acta Crystallogr Sect D 75:605–615

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797. https://doi.org/10.1093/nar/gkh340

Gautam S, Bundela P, Pandey A, Jamaluddin H, Awasthi M, Sarsaiya S (2010) Optimization of the medium for the production of cellulase by the *Trichoderma viride* using submerged fermentations. Int J Environ Sci 1:656–665

Gloster TM, Ibatullin FM, Macauley K, Eklof JM, Roberts S, Turkenburg JP, Bjornvad ME, Jorgensen PL, Danielsen S, Johansen KS, Borchert TV, Wilson KS, Brumer H, Davies GJ (2007) Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12. J Biol Chem 282:19177–19189

Hartmans S, De Bont J, Satackbrandt E (2006) The genus *Mycobacterium*-nonmedical. Prokaryotes 3:889–918

Hasunuma T, Kawashima K, Nakayama H, Murakami T, Kanagawa H, Ishii T, Akiyama K, Yasuda K, Terada F, Kushibiki S (2011) Effect of cellooligosaccharide or synbiotic feeding on growth performance, fecal condition and hormone concentrations in Holstein calves. Anim Sci 82(4):543–548

He B, Jin S, Cao J, Mi L, Wang J (2019) Metatranscriptomics of the Hu sheep rumen microbiome reveals novel cellulases. Biotechnol Biofuels 12(153):153

Hendricks C, Doyle J, Hugley B (1995) A new solid medium for enumerating cellulose-utilizing bacteria in soil. Appl Environ Microbiol 61:2016–2019

Henrissat B, Callebaud I, Fabrega S, Lehn P, Mornon J-P, Davies G (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc Natl Acad Sci USA 92:7090–7094. https://doi.org/10.1073/pnas.92.15.7090

Hilge M, Gloor SM, Rypniewski W, Sauer O, Heightman TD, Zimmermann W, Winterhalter K, Piontek K (1998) High-resolution native and complex structures of thermostable b-mannanase from *Thermomonospora fusca* - substrate specificity in glycosyl hydrolase family 5. Structure 6:1433–1444. https://doi.org/10.1016/s0969-2126(98)00142-7

Hruska K, Kaevska M (2012) Mycobacteria in water, soil, plants and air: a review. Vet Med 57(12):623–679

Jiao LF, Song ZH, Ke YL, Xiao K, Hu CH, Shi B (2014) Cello-oligosaccharide influences intestinal microflora, mucosal architecture and nutrient transport in weaned pigs. Anim Feed Sci Technol 195:85–91

Joynson R, Pritchard L, Osemwekha E, Ferry N (2017) Metagenomic analysis of the gut microbiome of the common black slug *Arion ater* in search of novel lignocellulose degrading enzymes. Front Microbiol 8:2181. https://doi.org/10.3389/fmicb.2017.02181

Kido K, Tejima S, Nagayama H, Uyeno Y, Ide Y, Kushibiki S (2016) Effects of supplementation with cellooligosaccharides on growth performance of weaned calves on pasture. Anim Sci J 87(5):661–665

Kim CH (1995) Characterization and substrate specificity of an endo-β-1, 4-D-glucanase I (Avicelase I) from an extracellular multienzyme complex of *Bacillus circulans*. Appl Environ Microbiol 3:959–965

Kim YR, Kim EY, Lee JM, Kim JK, Kong IS (2013) Characterisation of a novel *Bacillus* sp. SJ-10 β-1,3-1,4-glucanase isolated from jeotgal, a traditional Korean fermented fish. Bioprocess Biosyst Eng 36(6):721–727

Kim DY, Lee MJ, Cho HY, Lee JS, Lee MH, Chung CW, Shin DH, Rhee YH, Son KH, Park HY (2016) Genetic and functional characterization of an extracellular modular GH6 endo-β-1,4-glucanase from an earthworm symbiont, *Cellulosimicrobium funkei* HY-13. Anton Leeuw Int J G 109:1–12

Kim DR, Lim HK, Hwang IT (2018) Identification and functional characterization of an endoglucanase KRICT PC-001 from *Paenibacillus terrae* HPL-003. Appl Biotechm 54:616–623

Kondaveeti S, Patel SKS, Woo J, Wee JH, Kim S-Y, Al-Raoush RI, Kim I-W, Kalia VC, Lee J-K (2019) Characterization of cellobiohydrolases from *Schizophyllum commune* KMJ820. Indian J Microbiol 68:160–166. https://doi.org/10.1007/s12088-019-00843-9

Kuhad RC, Gupta R, Singh A (2011) Microbial cellulases and their industrial applications. Enzyme Res 2011:1–10. https://doi.org/10.4061/2011/280696

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Lafond M, Sulzenbacher G, Freyd T, Henrissat B, Berrin JG, Garron ML (2016) The quaternary structure of a glycoside hydrolase dictates specificity towards beta-glucans. J Biol Chem 291:7183–7194

Lee SH, Lee HE (2014) Cloning and characterization of a multidomain GH10 xylanase from *Paenibacillus* sp. DG-22. J Microbiol Biotechnol 24(11):1525–1535. https://doi.org/10.4014/jmb.1407.07077

Lee KM, Moon HJ, Kalyani D, Kim H, Kim IW, Jeya M, Lee JK (2011) Characterization of cellobiohydrolase from a newly isolated strain of *Agaricus arvencis*. J Microbiol Biotechnol 21(7):711–718

Liu D, Zhang R, Yang X, Xu Y, Thang Z, Tian W, Shen Q (2011) Expression, purification and characterization of two thermostable endoglucanases cloned from a lignocellulosic decomposing fungi *Aspergillus fumigatus* Z5 isolated from compost. Protein Expr Purif 79:176–186

Liu N, Li H, Chevrett MG, Zhang L, Cao L, Zhou H, Zhou X, Zhou Z, Pope PB, Currie CR, Huang Y, Wang Q (2019) Functional metagenomics reveals abundant polysaccharide degrading gene clusters and cellobiose utilization pathways within gut microbiota of a wood-feeding higher termite. ISME J 13:104–117. https://doi.org/10.1038/s41396-018-0255-1

Lo Leggio L, Larsen S (2002) The 1.62 A structure of *Thermoascus aurantiacus* endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5. FEBS Lett 523:103–108

Ludwiczek ML, Heller M, Kantner T, McIntosh LP (2007) A secondary xylan-binding site enhances the catalytic activity of a single-domain family 11 glycoside hydrolase. J Mol Biol 373(2):337–354

Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506–577

Maruthamuthu M, Jiménez DJ, van Elsas JD (2017) Characterization of a furan aldehyde-tolerant β-xylosidase/α-arabinosidase obtained through a synthetic metagenomics approach. J Appl Microbiol 123(1):145–158. https://doi.org/10.1111/jam.13484

Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31(3):426–428

Nacke H, Engelhaupt M, Brady S, Fischer C, Tautzt J, Daniel R (2012) Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. Biotechnol Lett 34:663–675

Otsuka M, Ishida A, Nakayama Y, Saito M, Yamazaki M, Murakami H, Nakamura Y, Matsumoto M, Mamoto K, Takada R (2004) Dietary supplementation with cellooligosaccharide improves growth performance in weanling pigs. Anim Sci J 75(3):225–229

Pan L, Farouk MH, Qin G, Zhao Y, Bao N (2018) The influences of soybean agglutinin and functional oligosaccharides on the intestinal tract of monogastric animals. Int J Mol Sci 19(2):544. https://doi.org/10.3390/ijms19020554

Park JI, Kent MS, Datta S, Holmes BM, Huang Z, Simmons BA, Sale KL, Sapra R (2011) Enzymatic hydrolysis of cellulose by the cellobiohydrolase domain of CelB from the hyperthermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Bioresour Technol 102:5988–5994

Patel M, Patel HM, Dave S (2019) Determination of bioethanol production potential from lignocellulosic biomass using novel Cel-5m isolated from cow rumen metagenome. Int J Biol Macromol 153:1099–1106. https://doi.org/10.1016/j.ijbiomac.2019.10.240

Patrick WM, Nakatani Y, Cutfield SM, Sharpe ML, Ramsay RJ, Cutfield JF (2010) Carbohydrate binding sites in *Candida albicans* exo-β-1,3-glucanase and the role of the Phe-Phe "clamp" at the active site entrance. FEBS J 277(21):4549–4561

Pereira JH, Chen Z, McAndrew RP, Sapra R, Chhabra SR, Sale KL, Simmons BA, Adams PD (2010) Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic *Thermotoga maritima*. J Struct Biol 172:372–379

Rahman M, Bhuiyan SH, Nirasawa S, Kitaoka M, Hayashi K (2002) Characterization of an endo-β-1,4-glucanase of *Thermotoga maritima* expressed in *Escherichia coli*. J Appl Glycosci 49(4):487–495

Ramirez-Gualito K, Alonso-Rios R, Quiroz-Garcia B, Rojas-Aguilar A, Diaz D, Jimenez-Barbero J, Cuevas G (2009) Enthalpic nature of the CH/pi interaction involved in the recognition of carbohydrates by aromatic compounds, confirmed by a novel interplay of NMR, calorimetry, and theoretical calculations. J Am Chem Soc 131:18129–18138

Romero Victorica M, Soria MA, Batista-García RA, Ceja-Navarro JA, Vikram S, Ortiz M, Ontañon O, Ghio S, Martínez-Ávila L, Quintero García OJ, Etcheverry C, Campos E, Cowan D, Arneodo J, Talia PM (2020) Neotropical termite microbiomes as sources of novel plant cell wall degrading enzymes. Sci Rep 10(3864) doi: https://doi.org/10.1038/s41598-020-60850-5

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5(4):725–738. https://doi.org/10.1038/nprot.2010.5

Sakon J, Adney WS, Himmel ME, Thomas SR, Karplus PA (1996) Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. Biochemistry 35:10648–10660

Santos CR, Paiva JH, Sforca ML, Neves JL, Navarro RZ, Cota J, Akao PK, Hoffmam ZB, Meza AN, Smetana JH, Nogueira ML, Polikarpov I, Xavier-Neto J, Squina FM, Ward RJ, Ruller R, Zeri AC, Murakami MT (2012) Dissecting structure-function-stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168. Biochem J 441:95–104

Schagerlof U, Schagerlof H, Momcilovic D, Brinkmalm G, Tjerneld F (2007) Endoglucanase sensitivity for substituents in methyl cellulose hydrolysis studied using MALDI-TOFMS for oligosaccharide analysis and structural analysis of enzyme active sites. Biomacromolecules 8:2358–2365

Schinsky M, McNeil M, Whitney A, Steiger A, Lasker M, Floyd M, Hogg G, Brenner D, Brown J (2000) *Mycobacterium septicum* sp. nov., a new rapidly growing species associated with catheter-related bacteraemia. Int J Syst Evol Microbiol 50:575–581

Schliep KP (2011) Phangorn: phylogenetic analysis in R. Bioinformatics 27:592–593

Shi W, Xie S, Chen X, Sun S, Zhou X, Liu L, Gao P, Kyrpides NC, No E-G, Yuan JS (2013) Comparative genomic analysis of the microbiome of herbivorous insects reveals eco-environmental adaptations: biotechnology applications. PLoS Genet 9(2):1–14. https://doi.org/10.1371/annotation/91a25db3-8127-42c7-baa0-ce398a2857a6

Shin K, Kim YH, Jeya M, Lee JK, Kim YS (2010) Purification and characterization of a thermostable cellobiohydrolase from *Fomitopsis pinicola*. J Microbiol Biotechnol 20(12):1681–1688

Skerman V, McGowan V, Sneath P (1980) Approved lists of bacterial names. Int J Syst Bacteriol 30:225–420

Song YH, Lee KT, Baek JY, Kim MJ, Kwon MR, Kim YJ, Park MR, Ko H, Lee JS, Kim KS (2017) Isolation and characterization of a novel endo-β-1,4-glucanase from a metagenomic library of the black-goat rumen. Braz J Microbiol 48:801–808

Soni S, Batra N, Bansal N, Soni R (2010) Bioconversion of sugarcane bagasse into second generation bioethanol after enzymatic hydrolisis with in-house produced cellulases *Aspergillus* sp. S4B2F. BioRes 5:741–758

Sun J, Wang H, Lv W, Ma C, Lou Z, Yao H, Dai Y (2012) Cloning and expression of a thermostable beta-1,3-1,4-glucanase from *Bacillus amyloliquefaciens* ATCC 23350. Ann Microbiol 62(3):1235–1242

Tailford LE, Ducros VM, Flint JE, Roberts SM, Morland C, Zechel DL, Smith N, Bjornvad ME, Borchert TV, Wilson KS, Davies GJ, Gilbert HJ (2009) Understanding how diverse beta-mannanases recognize heterogeneous substrates. Biochemistry 48:7009–7018

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(1):314–321. https://doi.org/10.1093/nar/gkt503

Tang B, Zhang Y, Yang Y, Song Z, Li X (2014) Expression and functional analysis of a glycoside hydrolase family 45 endoglucanase from *Rhizopus stolonifer*. World J Microbiol Biotechnol 30:2943–2952

Telke AA, Zhuang N, Ghatge SS, Lee SH, Ali Shah A, Khan H, Um Y, Shin HD, Chung YR, Lee KH, Kim SW (2013) Engineering of family-5 glycoside hydrolase (Cel5A) from an uncultured bacterium for efficient hydrolysis of cellulosic substrates. PLoS One 8(6): e65727

Tian W, Chen C, Lei X, Zhao J, Liang J (2018) CASTp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Res 46(1): 363–367. https://doi.org/10.1093/nar/gky473

Tseng CW, Ko TP, Guo RT, Huang JW, Wang HC, Huang CH, Cheng YS, Wang AHJ, Liu JR (2011) Substrate binding of a GH5 endoglucanase from the ruminal fungus *Piromyces rhizinflata*. Acta Crystallogr Sect F Struct Biol Cryst Commun 67:1189–1194

Tsukamura M, Nemoto H, Hiroyuki Y (1983) *Mycobacterium porcinum* sp. nov., a porcine pathogen. Int J Syst Bacteriol 33:162–165

Vlasenko E, Schulein M, Cherry J, Xu F (2010) Substrate specificity of family 5, 6, 7, 9,12, and 45 endoglucanases. Bioresour Technol 101: 2405–2411

Vogel J (2008) Unique aspects of the grass cell wall. Curr Opin Plant Biol 11:301–307. https://doi.org/10.1016/j.pbi.2008.03.002

Vyas NK, Vyas MN, Quiocho FA (1991) Comparison of the periplasmic receptors for L-arabinose, D-glucose/D-galactose, and D-ribose. Structural and functional similarity. J Biol Chem 266:5226–5237

Wahab AFFA, Abdul Karim NA, Ling JG, Hasan NS, Yong HY, Bharudin I, Kamaruddin S, Abu Bakar FD, Murad AMA (2018) Functional characterisation of cellobiohydrolase I (Cbh1) from T*richoderma virens* UKM1 expressed in *Aspergillus niger*. Protein Expr Purif 154:52–61

Wang Q, Tull D, Meinke A, Gilkes NR, Warren RA, Aebersold R, Withers SG (1993) Glu280 is the nucleophile in the active site of *Clostridium thermocellum* CelC, a family A endo-beta-1,4-glucanase. J Biol Chem 268:14096–14102

Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernandez M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P, Leadbetter JR (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450(7169):560–565

Wierzbicka-Woś A, Henneberger R, Batista-García RA, Martínez-Ávila L, Jackson SA, Kennedy J, Dobson ADW (2019) Biochemical characterization of a novel monospecific endo-β-1,4-glucanase belonging to GH family 5 from a rhizosphere metagenomic library. Front Microbiol 10:1342. https://doi.org/10.3389/fmicb.2019.01342

Wilson DB (2011) Microbial diversity of cellulose hydrolysis. Curr Opin Microbiol 14(3):259–263

Wu S, Wu S (2020) Processivity and the mechanisms of processive endoglucanases. Appl Biochem Biotechnol 190(2):448–463. https://doi.org/10.1007/s12010-019-03096-w

Wu S, Zhang Y (2007) LOMETS: a local meta-threading-server for protein structure prediction. Nucl Acids Res 35:3375–3382. https://doi.org/10.1093/nar/gkm25

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

Xia X, Gurr GM, Vasseur L, Zheng D, Zhong H, Qin B, Lin J, Wang Y, Song F, Li Y, Lin H, You M (2017) Metagenomic sequencing of diamondback moth gut microbiome unveils key holobiont adaptations for herbivory. Front Microbiol 8:663. https://doi.org/10.3389/fmicb.2017.00663

Yamane K, Suzuki H, Hirotani M, Ozawa H, Nisizawa K (1970) Effect of nature and supply of carbon sources on cellulase formation in *Pseudomonas fluorescens* var. cellulosa. Biochem 67:9–18. https://doi.org/10.1093/oxfordjournals.jbchem.a129238

Zhang Y, Himmel M, Mielenz J (2006) Outlook for cellulose improvement: screening and selection strategies. Biotechnol Adv 24(5):452–481. https://doi.org/10.1016/j.biotechadv.2006.03.003

Zhao PY, Jung JH, Kim IH (2012) Effect of mannan oligosaccharides and fructan on growth performance, nutrient digestibility, blood profile, and diarrhea score in weanling pigs. J Anim Sci 90(3):833–839

Zheng B, Yang W, Zhao X, Wang Y, Lou Z, Rao Z, Feng Y (2012) Crystal structure of hyperthermophilic endo-β-1,4-glucanase. J Biol Chem 287:8336–8346. https://doi.org/10.1074/jbc.M111.266346