

The first crystal structure of a family 45 glycoside hydrolase from a brown-rot fungus, *Gloeophyllum trabeum* GtCel45A

Laura Okmane , Louise Fitkin, Mats Sandgren and Jerry Ståhlberg 

Department of Molecular Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

Keywords

Cel45A; cellulase; cellulose; endoglucanase; glycoside hydrolase GH45

Correspondence

L. Okmane, Department of Molecular Sciences, Swedish University of Agricultural Sciences, PO Box 7015, SE-750 07 Uppsala, Sweden

E-mail: laura.okmane@slu.se

(Received 31 August 2023, revised 12 January 2024, accepted 22 January 2024)

doi:10.1002/2211-5463.13774

Here we describe the first crystal structure of a beta-1,4-endoglucanase from a brown-rot fungus, *Gloeophyllum trabeum* GtCel45A, which belongs to subfamily C of glycoside hydrolase family 45 (GH45). GtCel45A is ~ 18 kDa in size and the crystal structure contains 179 amino acids. The structure is refined at 1.30 Å resolution and R_{free} 0.18. The enzyme consists of a single catalytic module folded into a six-stranded double-psi beta-barrel domain surrounded by long loops. GtCel45A is very similar in sequence (82% identity) and structure to PcCel45A from the white-rot fungus *Phanerochaete chrysosporium*. Surprisingly though, initial hydrolysis of barley beta-glucan was almost twice as fast in GtCel45A as compared to PcCel45A.

Brown rot fungi are major wood decomposers in coniferous forests. They are considered generalists or gymnosperm specialists in regard to host preference [1]. As the name suggests, a visual consequence of the degradation is browning of residual wood. That is because brown rot fungi primarily depolymerize cellulose and hemicellulose, and leave most of the lignin in the residues [2,3]. Substrate degradation is achieved by secretion of various enzymes, such as cellulases, LPMOs and xylanases [4–6]. In comparison to white rot fungi, brown rot fungi contain less glycoside hydrolase and LPMO encoding genes, yet glycoside hydrolase family 45 (GH45) subfamily C genes appear to be more common in brown rot fungi [7]. The brown rot fungus *Gloeophyllum trabeum* is known to secrete all types of the aforementioned enzymes [8,9].

Fungal cellulases and their use as tools in utilizing industrial lignocellulosic waste have been a research topic for more than 60 years. Perhaps the most studied is the application of fungal cellulases to generate

biofuels from cellulosic biomass, often carried out by introducing potent fungal cellulases in the enzyme mixtures. However, a rapid and complete cellulose degradation is not always the desired characteristic in an industrial cellulase. For example, for cellulases used in washing powders, mild cellulolytic activity is desirable to avoid bulk degradation of intact cotton fibers. A common cellulase used for this purpose is the GH45 cellulase from *Humicola insolens*, HiCel45A [10–13].

GH45 enzymes are generally small and have the smallest catalytic modules among GH families. Although often called cellulases, they commonly have broad specificity and hydrolyze other cell wall polysaccharides more readily (e.g. beta-glucan, glucomannan), with low activity on insoluble cellulose [10,14]. Currently, ~ 570 entries are listed in GH45 in the CAZy database, ~ 40 from bacteria, the great majority from fungi, and fewer from other eukaryotes (e.g. mollusks, insects, crustaceans, nematodes).

Abbreviations

CBM, carbohydrate binding module; GH45, glycoside hydrolase family 45; LPMO, lytic polysaccharide monooxygenase; RMSD, root mean square deviation.

Phylogenetic analyses have divided GH45 family into three subfamilies. The subfamilies are referred to as either A, B, and C or according to a CAZy database classification as 1, 2, and 3, respectively [10,15,16]. The division of GH45 in subfamilies A, B, C predates the CAZy division. To this date, subfamily A, which includes HiCel45A, has been the most studied and subfamily C the least. Subfamilies B and C are more similar than subfamily A, to each other and to non-hydrolytic cell wall-active proteins called expansins, loosenins, and swollenins [17]. All these proteins contain a conserved double-psi beta-barrel domain (DPBB), also known as the GH45-like domain. Loosenins and most GH45s consist of this domain alone, while expansins and swollenins also contain a beta-sandwich domain (CBM63) in combination with the DPBB. Swollenins and some GH45s (e.g. HiCel45A) have an additional carbohydrate-binding module attached with a flexible linker. There are also examples of GH45s with more complex domain architectures.

GH45 enzymes from all three subfamilies employ an inverting hydrolysis mechanism, with an aspartate residue acting as catalytic acid, which is also conserved in expansins, loosenins and swollenins [14]. However, the Asp proposed to act as catalytic base in subfamilies A and B, is missing in subfamily C, as well as in the non-hydrolytic proteins [15]. Instead, an asparagine at another position has been proposed to act as catalytic base in GH45 subfamily C, based on neutron diffraction structure studies of *PcCel45A* from the white-rot fungus *Phanerochaete chrysosporium* [18]. Although the mechanism of subfamily C may differ from A and B, subfamilies B and C are more alike in regard to reaction product profile [14].

Crystal structures of nine GH45 enzymes are available in the PDB ([rcsb.org](https://www.rcsb.org)), six from kingdom Fungi (ascomycetes and basidiomycetes) and three from Metazoa (gastropod, bivalve, springtail), but no structure from brown rot fungi. There are structures for six members of subfamily A, two members of B (from mollusks) and only one of subfamily C, *PcCel45A*. Here we describe a second structure of a GH45 subfamily C enzyme—*GtCel45A* from the brown-rot fungus *Gloeophyllum trabeum*.

Recombinant expression of *GtCel45A*, biochemical characterization and activity profiling have been published previously, in comparison with a subfamily A enzyme, MtGH45 from the ascomycete *Myceliophthora thermophila* [19]. Activity optima for *GtCel45A* were at pH 5 and 65 °C, and activities were in a similar range to MtGH45 on carboxymethyl cellulose, beta-glucan, lichenan, and Avicel substrates.

Materials and methods

Cultivation, expression, and purification

Spores of *Aspergillus nidulans* A773 strain expressing *GtCel45A* (GenBank: [EPQ56593](https://www.ncbi.nlm.nih.gov/nuccore/EPQ56593)) were kindly provided by Dr. Fernando Segato, University of Sao Paulo, Brazil. The generation of constructions is described by Berto *et al.* [19]. Approximate protein size is 18.4 kDa (183 aa), the calculated pI is 4.5.

A spore suspension in water was made from a 13-day old sporulating culture growing on potato dextrose agar (PDA) plates. To prepare a pre-culture, 50 mL of minimal medium (70 mM NaNO₃, 7 mM KCl, 6 mM KH₂PO₄, 6 mM K₂HPO₄, 4.3 mM MgSO₄, 20 μM FeSO₄·7 H₂O, 80 μM ZnSO₄·7 H₂O) supplemented with pyridoxine (1 mg·L⁻¹) and maltose (3%) were inoculated with 200 μL of the spore suspension. The pre-culture was incubated in shake flasks at 25 °C, 120 rpm for 10 days.

Protein expression was done in the same medium at 30 °C in 1 L shake flasks, 4 × 350 mL cultures, shaking at 75 rpm. Cultures were harvested after 7 days, twice filtered (through 1 μm GF/B Whatman glass-fiber filter, followed by vacuum filtration through a 0.45 μm PES membrane).

The culture filtrate was concentrated using a VivaFlow 200 crossflow cassette (5000 MWCO) into 10 mM sodium phosphate buffer, pH 6.5. The concentrated protein solution was then loaded on an anion exchange column (DEAE Sepharose CL-6B, GE Healthcare, Uppsala, Sweden), with 50 mM sodium phosphate, pH 6.5, as the loading buffer and 50 mM sodium phosphate, 1 M NaCl, pH 6.5, as the elution buffer (gradient elution: 400 mL 0–50%; 200 mL 50–100%; 200 mL 100%). SDS/PAGE was run between purification steps to track the protein. Fractions containing ~ 18 kDa large protein were pooled and selected for next purification step. The IEC was followed by size exclusion chromatography (Superdex 75 HiLoad 16/600, GE Healthcare) in 10 mM sodium acetate, pH 5.0.

Protein concentration was determined at 280 nm with a NanoDrop UV-Vis Spectrophotometer using an extinction coefficient of 0.1% = 1.71, which was calculated by Protein Identification and Analysis Tools on the ExPASy Server [20]. The purified protein was concentrated to 14.12 mg·mL⁻¹ in 700 μL.

Protein crystallization

Initial screening was carried out using commercially available sparse matrix screen JCSG-plus HT-96 (Molecular Dimensions, Newmarket, Suffolk, UK). The following was identified as the hit condition: 12% (w/v) PEG 3350, 0.1 M HEPES, pH 7.5, 5 mM NiCl₂, 5 mM CoCl₂, 5 mM MgCl₂, 5 mM CdCl₂.

The condition was further optimized in hanging drops to 0.1 M HEPES, pH 7.5, 20% PEG 3350 and 3 mM NiCl₂

with protein solution concentration of 14 mg·mL⁻¹. Protein and reservoir proportion was 1 : 1 in hanging drops.

To reassure the identity of the crystallized protein, a protein crystal was dissolved in distilled water and the sample analyzed by SDS/PAGE (Fig. S2).

Data collection and structure determination

X-ray diffraction data were acquired at ID30B beamline at European Synchrotron Radiation Facility (ESRF), Grenoble, France. 7000 images were collected with an oscillation angle of 0.10° and X-ray exposure time of 0.02 s with 4.9% transmission.

4000 images were used and processed using xds software package [21]. The structure was solved by molecular replacement with phenix.phaser from PHENIX suite [22] using the mature protein sequence of Cel45A from *G. trabeum* (UniProtKB: S7QB86) and coordinates from PDB entry 5KJO (*P. chrysosporium* Cel45A) as search model. The structure was refined using phenix.refine from PHENIX suite. Visual inspection and real space refinement were carried out using COOT [23]. Statistics are summarized in Table 1. The structure was deposited in the Protein Data Bank under ID: 8BZQ.

Protein structure figures were prepared with PYMOL Molecular Graphics System, Version 2.0 Schrödinger (LLC, New York, NY, USA). Polar contacts were identified in PYMOL. Secondary structure plot was created with EMBL-EBI online service PDBSum, where the secondary structure

motifs are computed by v.3.0 of Gail Hutchinson's PROMOTIF program [24].

Bioinformatic analysis

The protein-protein BLAST algorithm of the blastp suite was used to search for similar sequences [25–27]. The following sequences were used for query: *Mytilus edulis* MeCel45A, *Ampullaria crosseana* AcCel45A, *Humicola insolens* HiCel45A, *P. chrysosporium* PcCel45A, *Trichoderma reesei* TrCel45A. The top 100 sequences with minimum 50% identity were collected and later revised to remove incomplete or identical sequences.

For each subfamily, a multiple sequence alignment was generated by MUSCLE (version 5) tool for multiple alignment [28]. Consensus sequences were created using UGENE (version 41.0) [29–31], with consensus type “strict” and 90% threshold. Minimum of 50 sequences were used for the generation of each consensus sequence. Visualization of consensus sequence alignment was carried out using ESPRIT 3.0 [32]. To determine the percentage of sequence identity between *Gt*Cel45A and *Pc*Cel45A, and Cel45A from *Gymnopilus dilepis*, pairwise sequence alignments were made using EMBOSS Needle available online at the EMBL-EBI server (ebi.ac.uk) [33]. T-COFFEE (Version 11.00) was used to create a structure-based multiple sequence alignment [34]. Domain identification in Cel45A from *G. dilepis* was carried out with ScanProsite tool available online at Swiss Bioinformatics Resource Portal (expasy.org) [35].

To mark the secondary sequence elements, the following crystal structures were visualized in PYMOL and used as a template for the corresponding subfamily: HiCel45A (PDB ID: 2ENG), MeCel45A (PDB ID: 1WC2), *Gt*Cel45A (PDB ID: 8BZQ), and a homology model for TrCel45A. A homology model for GdCel45A (GenBank: PPQ98991.1) was created using RoseTTAFold [36]. PYMOL was used to generate the 3D structure images of all GH45 enzymes in this paper. All figures were made ready for publication using a vector design application AFFINITY DESIGNER (version 1.10, 8ab56).

For the modeling of cellulose chain binding in *Gt*Cel45A, *Gt*Cel45A structure was superposed with the structure of *Pc*Cel45A in complex with two cellopentaose molecules, binding in -5 to -1 and +1 to +5, respectively (PDB: 3X2M). To improve cellulose (G10) fit in the substrate binding area, adjustments were made using Coot. The glucose residue in the -1 subsite was replaced with a distorted -1 Glc from an MD simulation model of cello-oligo binding in HiCel45A [37]. The Asn95 sidechain clashed with the +1 Glc unit and was therefore rotated to the orientation that the corresponding residue has in the *Pc*Cel45A complex structure (PDB: 3X2M). The G10 chain was merged with the *Gt*Cel45A structure regularized in Coot. The resulting structure was refined in PHENIX. The G10 chain was reconstructed by merging the -5 to +1 Glc units from

Table 1. Data collection and refinement statistics. Values taken from the validation report for the deposited structure.

Property	Value
Space group	P 1, 21, 1
Cell constants	28.97 Å, 48.41 Å, 45.79 Å
<i>a</i> , <i>b</i> , <i>c</i> , α , β , γ	90.00°, 92.21°, 90.00°
Resolution (Å)	45.76–1.30
% Data completeness (in resolution range)	99.7 (45.76–1.30)
$\langle I/\sigma(I) \rangle^1$	9.33 (at 1.30 Å)
Refinement program	PHENIX 1.19.2_4158
<i>R</i> , <i>R</i> _{free}	0.166, 0.183
<i>R</i> _{free} test set	974 reflections (3.13%)
Wilson B-factor (Å ²)	9.6
Anisotropy	0.149
Bulk solvent <i>k</i> _{sol} (e/Å ³), <i>B</i> _{sol} (Å ²)	0.34, 38.7
<i>L</i> -test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$
Estimated twinning fraction	0.035 for h, -k, -l
<i>F</i> _o , <i>F</i> _c correlation	0.96
Total number of atoms	1481
Average B, all atoms (Å ²)	13.0

the refined structure (BGC210-205; including glycosidic oxygen O4 from BGC204) with the +2 to +5 units from the *PcCel45A* complex (BGC204-GLC201).

Enzyme activity assays

GtCel45A and *PcCel45A* at 0.1 μM concentration were each incubated at 30 $^{\circ}\text{C}$, 400 rpm, in 0.1 M sodium acetate buffer pH 5.0 with 0.1% barley beta-glucan (purity \sim 95%, Megazyme, Wicklow, Ireland) dissolved in water. 1 M sodium hydroxide solution was added in a 1 : 1 ratio to the samples to stop the reaction, which was followed by cooling the samples on ice.

Reducing sugar formation was determined colorimetrically by adding p-hydroxybenzoic acid hydrazide (PHBAH, Sigma) solution: 0.1 M PHBAH; 0.2 M sodium potassium tartrate; 0.5 M sodium hydroxide solution [38]. The PHBAH solution was prepared directly before use and added in 1 : 1 ratio to the samples. Samples containing the PHBAH reagent were immediately incubated at 95 $^{\circ}\text{C}$ for 15 min and then cooled on ice for 10 min. Samples were kept at room temperature for 5 min before absorption was read at 410 nm.

Results and Discussion

Overall structure of *GtCel45A*

The final crystal structure model of *GtCel45A* was refined at 1.3 \AA resolution and R/R_{free} values

0.166/0.183. The space group was P21 with 1 protein chain per asymmetric unit. Data collection and refinement statistics are summarized in Table 1. The structure has been deposited in the Protein Data Bank under PDB ID: 8BZQ.

GtCel45A has a globular structure, with dimensions approximating to 26 \AA \times 37 \AA \times 48 \AA . It contains a six-stranded beta barrel with a double-psi beta barrel (DPBB) fold, also known as the N-terminal domain of expansins, surrounded by six short helices and seven long loops. An open substrate binding groove spans across the surface and is 48 \AA in length, with the widest part of around 14 \AA near the center and depth around 12 \AA (Fig. S1). The first four amino acids that are present in the mature protein sequence (LEER) are not visible in the crystal structure.

The secondary structure along the protein sequence is depicted in Fig. 1 and the 3D structure with labeled secondary structure elements is shown in Fig. 2A.

In the following, beta strands are assigned letter β and numbered based on the order in the amino acid sequence. Loops in the structure are assigned a letter depending on their position relative to the substrate binding area (Fig. 2A). Loops positioned above the substrate binding area (where reducing end of the cellulose chain is to the right side relative to the active site) are assigned capital letter A, loops below – capital letter B. Loop numbering is based on their order in the amino acid sequence.

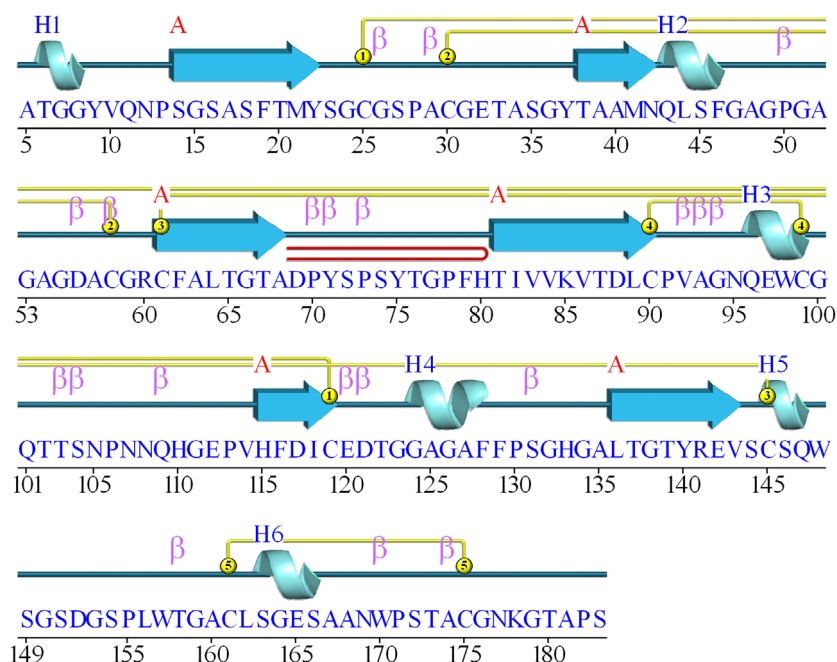


Fig. 1. Secondary structure plot for *GtCel45A*. Arrows represent beta strands (A), spirals—alpha helices (H), number connections show cysteine disulfide pairings, letter β indicates a beta turn motif, a beta hairpin motif is depicted in red.

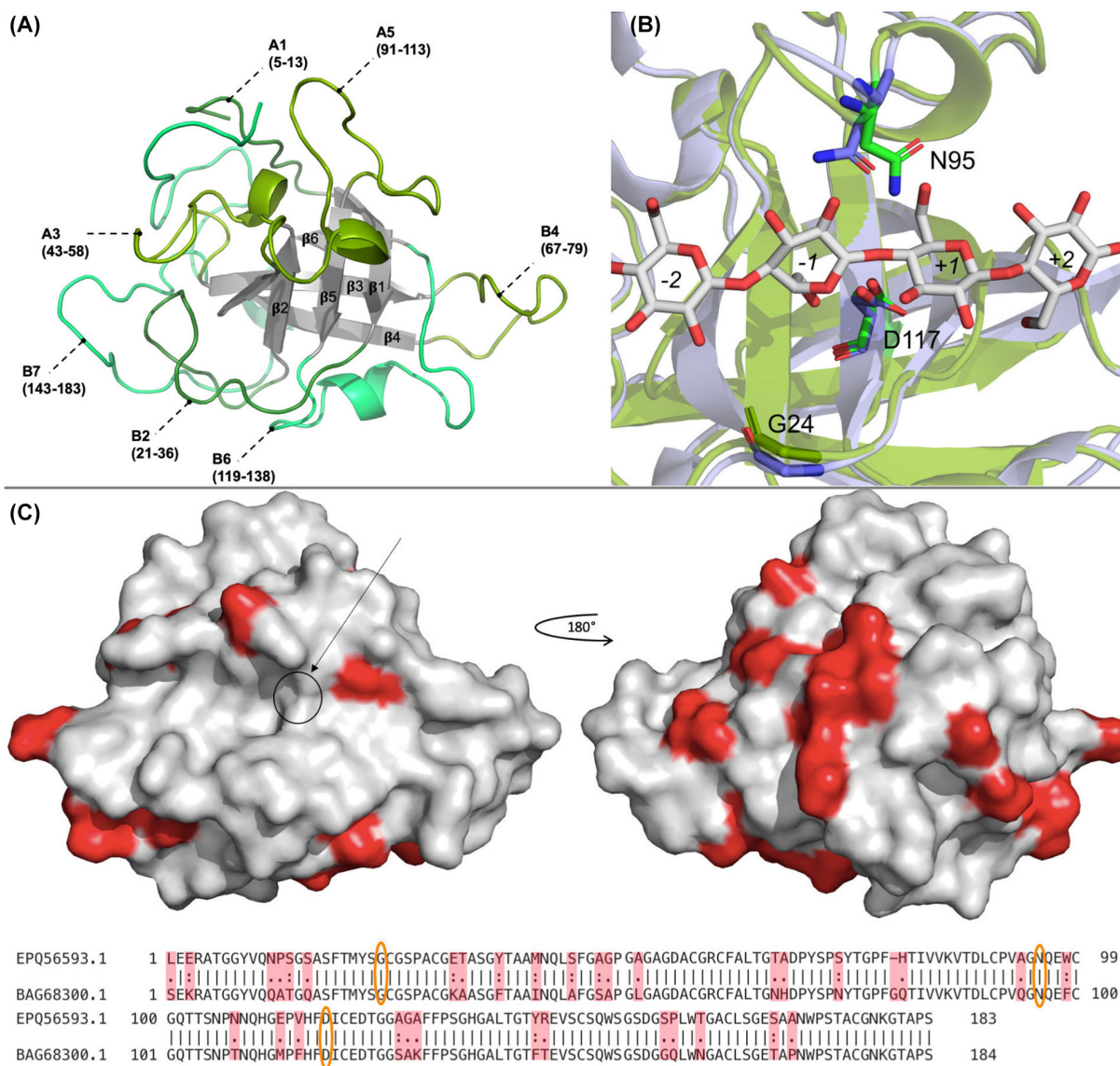


Fig. 2. Overall structure of GtCel45A. (A) Ribbon drawing with numbering of loops, corresponding amino acid range indicated in brackets; (B) Superposed active site of GtCel45A (green) and PcCel45A (blue) with substrate bound; (C) GtCel45A surface structure. Residues in GtCel45A that are identical to residues in PcCel45A are indicated in red on the surface structure of GtCel45A. Catalytic center indicated with an arrow. Residues positioned at the location of catalytic residues in subfamily A are circled in orange in the sequence alignment. [Correction added on 08 February, 2024, after first online publication: The residue numbering of loop A4 in Figure 2A has been corrected in this version.]

The structure model begins with loop A1, which contains a three amino acid long alpha helix. Five amino acids connect the helix to the first beta strand ($\beta 1$), which is connected to the next beta strand ($\beta 2$) via 15 amino acid loop (B2) (Fig. 2A). The secondary structure pattern of *helix* – β -strand – β -strand is repeated once more (Fig. 1), where at least 12 amino acid long loops connect the secondary structure elements. This pattern is followed by two repetitions of *helix* – β -strands. Two helices and long loops make up the C-terminal part of the structure.

On loop B2, shortly after strand $\beta 1$, is located Tyr22 – a residue conserved in GH45 enzymes, as well as expansins. Gly24 on loop B2 is at the same position as the catalytic base in GH45 subfamily A and B members, for example Asp10 in HiCel45A and Asp24 in MeCel45A. On loop 5 (A5), between strands $\beta 4$ and $\beta 5$, is located an Asparagine (Asn95) residue (Fig. 2A). An asparagine in this position is regarded as the assisting residue in subfamily A enzymes. The putative catalytic acid Asp117 is found at the end of $\beta 5$.

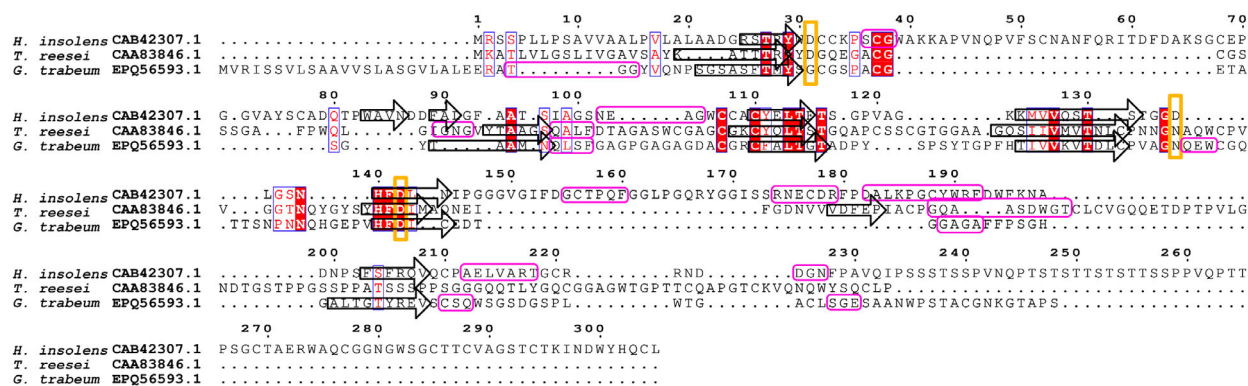


Fig. 3. Structure-based sequence alignment of *H. insolens* Cel45A, *T. reesei* Cel45A and *G. trabeum* Cel45A. Alignment visualized in ESPRIT 3.0. Secondary structure elements are represented as rounded rectangles (helices) and arrows (β -strands). Character coloration according to ESPRIT 3.0: filled red box and a white character indicate strict identity; red character—similarity within a group. Active site residues of subfamily A and residues at the corresponding location in other subfamilies are marked with a yellow frame.

The *Gt*Cel45A structure contains five disulfide bridges formed by cysteine pairing. All disulfide bonds appear to become reduced with the increase of X-ray exposure time during data collection. Two of the cysteine pairs (Cys61/Cys145, Cys161/Cys175) are found at either end of loop B7. Loop B7 carries Trp157 in the same position as Trp154 in *Pc*Cel45A (structure with two cellopentaose molecules, PDB ID: 3x2m) and similar to Trp64 in *Me*Cel45A, which function as sugar-binding platform in subsite -4 .

The position of loop A5, which carries the putative alternate base Asn95, is fastened by Cys90/Cys99, which positions the Asn95 toward the catalytic acid Asp117. Molecular dynamics simulations have shown that in the absence of a substrate, a corresponding loop in *Pc*Cel45A is able to enclose toward Asp117 (Asp114 in *Pc*Cel45A), forming a polar contact between Asn95 and Asp117 (Asn92 and Asp114 in *Pc*Cel45A) [39].

The following residues have poor electron density indicated by relatively low real-space correlation coefficient ($CC < 0.9$) and relatively high isotropic temperature factors (B-factors; $B_{iso} > 18.0$): Pro50, Ala93, Gly94, Asn105, Pro130, Ser131, Gly132, His133, Thr148, Pro171, Ser172, Thr173, Thr180, Ser183. The low CC and high B factors for residues on the C-terminal loops B6 and B7, specifically residues 130–133 and 171–183, possibly indicate dynamic disorder thus flexibility of loops B6 and B7 [40]. The electron density around Ser16 suggests static disorder with alternate conformation, assigned occupancies 0.7 and 0.3.

GH45 subfamily C

The other molecular structure which has been deposited in PDB and belongs to subfamily C is *Pc*Cel45A. *Gt*Cel45A has 82% sequence identity with *Pc*Cel45A

and 88% sequence similarity. The crystal structure of *Gt*Cel45A consists of 179 amino acids while *Pc*Cel45A – of 180. At the end of the loop B4, *Gt*Cel45A has His80, while two residues, Gly76 and Gln77, are located at this position in *Pc*Cel45A.

A total of 31 residues differ in the crystal structure of *Gt*Cel45A relative to the crystal structure of *Pc*Cel45A. Most are away from the substrate binding groove (Fig. 2C), three can be found in its periphery (Ala93/Gln90, Trp98/Phe95, Thr158/Asn155 in *Gt*Cel45A/*Pc*Cel45A respectively), but none at the catalytic site. All available *Pc*Cel45A crystal structures to date are missing the first four N-terminal residues present in the mature protein sequence.

The subfamily C enzymes are unique in GH45 family due to the ability to achieve hydrolysis despite lacking a traditional catalytic base residue in the active site (Fig. 2B). Both subfamily A and subfamily B enzymes have aspartic acid residues acting as a catalytic base in the active site, while subfamily C has a glycine residue at this position.

It has been suggested that subfamily C members utilize an asparagine residue (Asn95 in *Gt*Cel45A) in their catalytic mechanism, instead of using the residue corresponding to the position of catalytic base in the other two subfamilies. The Asn95 is positioned on a loop enclosing toward the active site. Asparagines are conserved at this location in subfamily B and C, and aspartic acid—in subfamily A (Fig. 3). Mutation of the asparagine residue at this position in the subfamily C enzymes from *P. chrysosporium* and *Fomitopsis palustris* leads to a significant decrease in their hydrolytic activity [41].

The catalytic mechanism attributed to the hydrolytic properties of subfamily C is the “Newton’s cradle” proton relay mechanism which originally was proposed in *Pc*Cel45A. This mechanism involves a proton relay

through hydrogen bonding network from the catalytic acid to the proposed catalytic base. In *PcCel45A*, it involves 9 amino acids: Asn92, Gln93, Cys96, Phe95, Asn105, Ser14, His112, Thr16, Asp114 [18]. *GtCel45A* has conserved most of the amino acids involved in the “Newton’s cradle” proton relay mechanism (Asn95, Gln96, Cys99, Trp98, Asn108, Ser18, His115, Thr20, Asp117), except for residue Phe95 (Trp98 instead). According to this mechanism, the imidic acid form of Asn95 could act as a general base in *GtCel45A*.

MD simulations [39] with celloheptaose molecule bound in the substrate-binding groove have shown that residues interacting with substrate the strongest in *PcCel45A* are Asp85 (subsite -2), Asn92 (subsite -1), Trp154 (subsite -4). The corresponding residues in *GtCel45A* (Asp88, Asn95, Trp157) are located in the same positions. Other residues assumed to interact with celloheptaose in *PcCel45A* (Gly47, Leu86, Tyr18, Asp114, Thr16, Met17, His107, Gly131, Tyr67) are also the same in *GtCel45A* (Gly51, Leu89, Tyr22, Asp117, Thr20, Met21, His110, Gly134, Tyr71).

Consensus sequences show that 14 residues are strictly conserved in all of the GH45 subfamilies known to date (Fig. S3). Subfamily B sequences appear less conserved, presumably due to the larger diversity of organisms expressing subfamily B enzymes – mollusks and fungi.

As mentioned previously, the absence of an acidic residue at the catalytic base position distinguishes subfamily C from the other two subfamilies (Fig. S3). However, there is further diversity among these enzymes which requires investigation. Most subfamily C enzymes are one domain proteins, but at least one appears to have a CBM, *Gymnopilus dilepsis* Cel45A (GenBank: PPQ98991.1) [42], which has an N-terminal CBM1 domain connected by a linker to the catalytic domain (Fig. S4). After removal of the CBM1 and linker from the sequence, *G. dilepis* Cel45A (*GdCel45A*) becomes 75% sequence identical to *GtCel45A*.

Thus far studies regarding enzymatic activity have not been carried out at identical conditions, therefore a uniform opinion on the catalytic activities of subfamily C enzymes cannot be made. It is clear that GH45 subfamily enzymes are able to hydrolyze CMC, barley beta-glucan and glucomannan [14,15,19,41]. Studies show that the pH optimum is around pH 5, when determined on CMC, and some of GH45 subfamily C enzymes appear stable at 50 °C [19,41].

We compared the hydrolytic activity of *GtCel45A* and *PcCel45A* on barley beta-glucan at 30 °C, pH 5. Despite the high sequence and structural similarity, *GtCel45A* exhibited ~ 3.8 times higher initial rate – it took 2.15 min for 0.1 μM *GtCel45A* to produce 112 ± 19 μM of

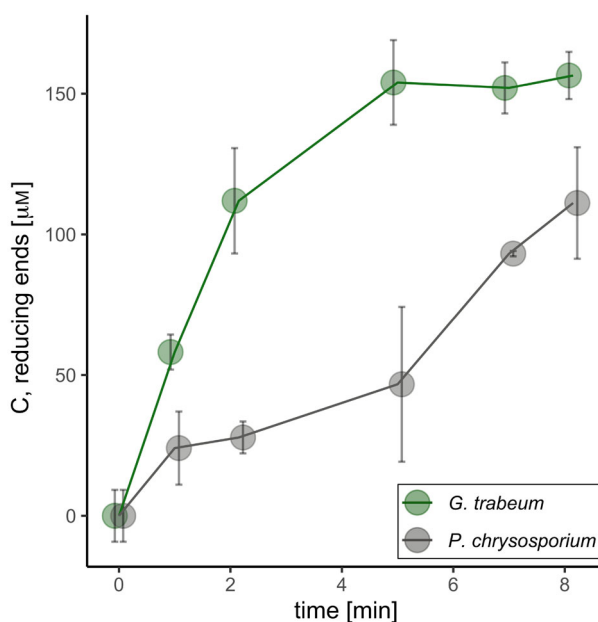


Fig. 4. Hydrolytic activity comparison of 0.1 μM *GtCel45A* and 0.1 μM *PcCel45A* on 0.1% barley beta-glucan, at +30 °C, pH 5.0. Error bars represent SD.

reducing ends and 8.15 min for 0.1 μM *PcCel45A* to produce 111 ± 20 μM. Moreover, *GtCel45A* reached a plateau faster than *PcCel45A* (Fig. 4). At present we do not have any explanation for that difference.

Conclusions

Cel45A from the brown-rot fungus *G. trabeum* is a GH45 subfamily C member with a nearly identical structure to the Cel45A from white-rot fungus *P. chrysosporium*, and yet shows higher catalytic activity. The diversity within GH45 subfamily C would benefit from elucidation, as there is minimum one subfamily C member which has a CBM1 domain, the Cel45A from *G. dilepis*. The GH45 family is strictly conserved regarding 14 residues.

Acknowledgements

We thank Dr. Fernando Segato, University of Sao Paulo, Brazil, for providing spores of the *GtCel45A* expressing *A. nidulans* strain. This work was supported by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas), Grant Number 2017-01130.

Conflict of interest

The authors declare no conflict of interest.

Peer review

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/2211-5463.13774>.

Data accessibility

The structural data that support these findings are openly available in the wwPDB at <https://doi.org/10.2210/pdb8BZQ/pdb>. The data that support the findings of this study are openly available in enzyme nomenclature database – ENZYME at <https://enzyme.expasy.org>, Enzyme Commission number: EC3.2.1.4.

Author contributions

LO: Conceptualization, Validation, Methodology, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision. LF: Validation, Methodology, Formal analysis, Investigation, Data Curation. MS: Writing – Review & Editing, Resources, Supervision, Project administration, Funding acquisition. JS: Conceptualization, Validation, Methodology, Formal analysis, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

References

- Krah F-S, Bäessler C, Heibl C, Soghigian J, Schaefer H and Hibbett DS (2018) Evolutionary dynamics of host specialization in wood-decay fungi. *BMC Evol Biol* **18**, 119.
- Chirkova J, Irbe I, Andersons B and Andersone I (2006) Study of the structure of biodegraded wood using the water vapour sorption method. *Int Biodeterior Biodegradation* **58**, 162–167.
- Jin L, Sellers T, Schultz TP and Nicholas DD (1990) Utilization of lignin modified by Brown-rot fungi. I. Properties of flakeboard produced with a brown-rotted lignin modified phenolic adhesive. *Holzforschung* **44**, 207–210.
- Kojima Y, Várnai A, Eijsink VGH and Yoshida M (2020) The role of lytic polysaccharide monoxygenases in wood rotting basidiomycetes. *Trends Glycosci Glycotechnol* **32**, E135–E143.
- Ritschkoff A-C, Buchert J and Viikari L (1992) Identification of carbohydrate degrading enzymes from the brown-rot fungus, *Gloeophyllum trabeum*. *Material Und Organismen* **27**, 19–29.
- Yoon J-J, Cha C-J, Kim Y-S, Son D-W and Kim Y-K (2007) The brown-rot basidiomycete *Fomitopsis palustris* has the endo-glucanases capable of degrading microcrystalline cellulose. *J Microbiol Biotechnol* **17**, 800–805.
- Hori C, Gaskell J, Igarashi K, Samejima M, Hibbett D, Henrissat B and Cullen D (2013) Genomewide analysis of polysaccharides degrading enzymes in 11 white- and brown-rot Polyporales provides insight into mechanisms of wood decay. *Mycologia* **105**, 1412–1427.
- Hegnar OA, Petrovic DM, Bissaro B, Alfretdsen G, Várnai A and Eijsink VGH (2019) pH-dependent relationship between catalytic activity and hydrogen peroxide production shown via characterization of a lytic polysaccharide monoxygenase from *Gloeophyllum trabeum*. *Appl Environ Microbiol* **85**, e02612-18.
- Mansfield SD, Saddler JN and Gübitz GM (1998) Characterization of endoglucanases from the brown rot fungi *Gloeophyllum sepiarium* and *Gloeophyllum trabeum*. *Enzyme Microb Technol* **23**, 133–140.
- Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, Sandgren M, Ståhlberg J and Beckham GT (2015) Fungal cellulases. *Chem Rev* **115**, 1308–1448.
- Olsen HS and Fallholt P (1998) The role of enzymes in modern detergency. *J Surfactant Deterg* **1**, 555–567.
- Kan CW, Yuen CWM and Wong WY (2011) Optimizing color fading effect of cotton denim fabric by enzyme treatment. *J Appl Polym Sci* **120**, 3596–3603.
- Niyonzima FN (2021) Detergent-compatible fungal cellulases. *Folia Microbiol* **66**, 25–40.
- Okmane L, Nestor G, Jakobsson E, Xu B, Igarashi K, Sandgren M, Kleywegt GJ and Ståhlberg J (2022) Glucomannan and beta-glucan degradation by *Mytilus edulis* Cel45A: crystal structure and activity comparison with GH45 subfamily a, B and C. *Carbohydr Polym* **277**, 118771.
- Igarashi K, Ishida T, Hori C and Samejima M (2008) Characterization of an endoglucanase belonging to a new subfamily of glycoside hydrolase family 45 of the basidiomycete *Phanerochaete chrysosporium*. *Appl Environ Microbiol* **74**, 5628–5634.
- Couturier M, Feliu J, Haon M, Navarro D, Lesage-Meessen L, Coutinho PM and Berrin J-G (2011) A thermostable GH45 endoglucanase from yeast: impact of its atypical multimodularity on activity. *Microb Cell Fact* **10**, 103.
- Cosgrove DJ (2015) Plant expansins: diversity and interactions with plant cell walls. *Curr Opin Plant Biol* **25**, 162–172.
- Nakamura A, Ishida T, Kusaka K, Yamada T, Fushinobu S, Tanaka I, Kaneko S, Ohta K, Tanaka H, Inaka K *et al.* (2015) “Newton’s cradle” proton relay with amide–imidic acid tautomerization in inverting cellulase visualized by neutron crystallography. *Sci Adv* **1**, e1500263.
- Berto GL, Velasco J, Tasso Cabos Ribeiro C, Zanphorlin LM, Noronha Domingues M, Tyago

- Murakami M, Polikarpov I, de Oliveira LC, Ferraz A and Segato F (2019) Functional characterization and comparative analysis of two heterologous endoglucanases from diverging subfamilies of glycosyl hydrolase family 45. *Enzyme Microb Technol* **120**, 23–35.
- 20 Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD and Bairoch A (2005) Protein identification and analysis tools on the ExpASY server. In *The Proteomics Protocols Handbook* (Walker JM, ed.), pp. 571–607. Humana Press, Totowa, NJ.
- 21 Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132.
- 22 Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, Hintze B, Hung L-W, Jain S, McCoy AJ *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877.
- 23 Emsley P and Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.
- 24 Hutchinson EG and Thornton JM (1996) PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci* **5**, 212–220.
- 25 Camacho C, Boratyn GM, Joukov V, Vera Alvarez R and Madden TL (2023) ElasticBLAST: accelerating sequence search via cloud computing. *BMC Bioinformatics* **24**, 117.
- 26 Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
- 27 Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ and Madden TL (2012) Domain enhanced lookup time accelerated BLAST. *Biol Direct* **7**, 12.
- 28 Edgar RC (2022) High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *bioRxiv* [10.1101/2021.06.20.449169](https://doi.org/10.1101/2021.06.20.449169) [PREPRINT]
- 29 Rose R, Golosova O, Sukhomlinov D, Tiunov A and Proserpi M (2019) Flexible design of multiple metagenomics classification pipelines with UGENE. *Bioinformatics* **35**, 1963–1965.
- 30 Golosova O, Henderson R, Vaskin Y, Gabrielian A, Grekhov G, Nagarajan V, Oler AJ, Quiñones M, Hurt D, Fursov M *et al.* (2014) Unipro UGENE NGS pipelines and components for variant calling, RNA-Seq and CHIP-Seq data analyses. *PeerJ* **2**, e644.
- 31 Okonechnikov K, Golosova O, Fursov M and The UGENE Team (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167.
- 32 Robert X and Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320–W324.
- 33 Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A and Lopez R (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* **50**, W276–W279.
- 34 Notredame C, Higgins DG and Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205–217.
- 35 Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L and Xenarios I (2013) New and continuing developments at PROSITE. *Nucleic Acids Res* **41**, D344–D347.
- 36 Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD *et al.* (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876.
- 37 Bharadwaj VS, Knott BC, Ståhlberg J, Beckham GT and Crowley MF (2020) The hydrolysis mechanism of a GH45 cellulase and its potential relation to lytic transglycosylase and expansin function. *J Biol Chem* **295**, 4477–4487.
- 38 Lever M (1972) A new reaction for colorimetric determination of carbohydrates. *Anal Biochem* **47**, 273–279.
- 39 Godoy AS, Pereira CS, Ramia MP, Silveira RL, Camilo CM, Kadowaki MA, Lange L, Busk PK, Nascimento AS, Skaf MS *et al.* (2018) Structure, computational and biochemical analysis of pc Cel45A endoglucanase from *Phanerochaete chrysosporium* and catalytic mechanisms of GH45 subfamily C members. *Sci Rep* **8**, 3678.
- 40 Sun Z, Liu Q, Qu G, Feng Y and Reetz MT (2019) Utility of B-factors in protein science: interpreting rigidity, flexibility, and internal motion and engineering thermostability. *Chem Rev* **119**, 1626–1665.
- 41 Cha J-H, Yoon J-J and Cha C-J (2018) Functional characterization of a thermostable endoglucanase belonging to glycoside hydrolase family 45 from *Fomitopsis palustris*. *Appl Microbiol Biotechnol* **102**, 6515–6523.
- 42 Reynolds HT, Vijayakumar V, Gluck-Thaler E, Korotkin HB, Matheny PB and Slot JC (2018) Horizontal gene cluster transfer increased hallucinogenic mushroom diversity. *Evol Lett* **2**, 88–101.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Measurements of GtCel45A dimensions expressed in Å and depicted as dashed lines.

Fig. S2. SDS/PAGE analysis of a dissolved protein crystal shows a single band at ~18 kDa, the expected

molecular weight of GtCel45A, thus confirming the identity of the purified and crystallized protein.

Fig. S3. (A) An alignment of GH45 subfamily A, B and C consensus sequences; (B) An alignment of subfamily B consensus sequence, subfamily B consensus

sequence in phylum Ascomycota, and phylum Mollusca.

Fig. S4. Homology structure model of the bimodular GH45 subfamily C enzyme GdCel45A from *Gymnopilus dilepis* with linker and CBM1.