Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Biochemical and structural characterisation of a family GH5 cellulase from endosymbiont of shipworm P. megotara

Madan Junghare (madan.junghare@nmbu.no)
Norwegian University of Life Sciences
Tamilvendan Manavalan
Norwegian University of Life Sciences
Lasse Fredriksen
Norwegian University of Life Sciences
Ingar Leiros
The Arctic University of Norway
Bjørn Altermark
The Arctic University of Norway
Vincent G.H. Eijsink
Norwegian University of Life Sciences
Gustav Vaaje-Kolstad
Norwegian University of Life Sciences

Research Article

Keywords: GH family 5, endo/exo-1-4-β-Glucanase, shipworm, endosymbiont, β-glucan barely, enzymatic hydrolysis

Posted Date: January 19th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2475918/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Biotechnology for Biofuels and Bioproducts on April 4th, 2023. See the published version at https://doi.org/10.1186/s13068-023-02307-1.

Abstract

Background

Cellulases play a key role in enzymatic conversion of plant cell-wall polysaccharides into simple and economically relevant sugars. The discovery of novel cellulases from exotic biological niches is of interest as they may present properties that are valuable in biorefining of lignocellulose.

Results

We have characterized a glycoside hydrolase 5 (GH5) domain of a bi-catalytic GH5-GH6 multidomain enzyme from the unusual bacterial endosymbiont *Teredinibacter waterbury* of the wood-digesting shipworm *Psiloteredo megotara*. The cellulase enzyme, *Tw*Cel5, was produced with and without a native C-terminal family 10 carbohydrate-binding module belongs to GH5, subfamily 2. Both variants showed hydrolytic endo-activity on soluble substrates such as, β-glucan, carboxymethylcellulose and konjac glucomannan. However, low activity was observed towards crystalline form of cellulose. Interestingly, when co-incubated with a cellulose active LPMO, a clear synergy was observed that boosted hydrolysis of crystalline cellulose. The crystal structure of the GH5 catalytic domain was solved to 1.0 Å resolution and revealed a substrate binding cleft containing a putative +3 subsite, which is uncommon in this enzyme family. The enzyme *Tw*Cel5 was active in a wide range of pH and temperatures and showed high tolerance for NaCl.

Conclusions

This study provides an important advance on discovery new enzymes from shipworm and shed new light on biochemical and structural characterization of cellulolytic cellulase and showed boost in hydrolytic activity of cellulase on crystalline cellulose when co-incubated with cellulose active LPMO. These findings will be relevant for the development of future enzyme cocktail that may be useful for the biotechnological conversion of lignocellulose.

Background

The desire to reduce the consumption of fossil fuels has sparked massive interest in searching alternative renewable energy sources, including lignocellulosic biomass. Plant-based lignocellulosic biomass is an abundant polymeric material that may be exploited for the production of renewable energy thus helping to reduce the consumption of fossil fuels. Lignocellulose consists of polysaccharides, including cellulose, hemicelluloses, and pectin, as well as lignin, a complex aromatic polymer. Cellulose is the most important structural component in plant biomass with an estimated global production of more than 1.5×10^{12} tons per year [1]. Cellulose is a homopolymer of glucose, linked by β-1,4-glycosidic bonds [2, 3]. Cellulose chains associate into insoluble, often crystalline fiber structures, which make the material strong and difficult for biodegradation. Intermolecular interactions with the other polymeric cell wall compounds add to this strength and recalcitrance. Consequently, although cellulose is an attractive

source of green energy, its exploitation is intricate due to its resistance to enzymatic depolymerization [4]. Indeed, an efficient conversion of lignocellulosic biomass, or even relatively pure cellulose fibers, requires multiple enzymes acting synergistically to deconstruct the feedstock and to generate monomeric sugars that can subsequently be converted to fuels and chemicals [5].

Enzymatic cellulose depolymerization is catalysed by cellulases and lytic polysaccharide monooxygenases (LPMOs) acting in concert [5]. The cellulases include endo- and exo-acting glycosyl hydrolases (GHs) that are though to act synergistically because, for instance, endo-β-1,4-glucanases hydrolyze internal glycosidic bonds to generate new chain ends on which exo-β-1,4-glucanases, also known cellobiohydrolases, can act. Despite decades of research, there is still a need for novel and efficient cellulases that may help improving sustainability and economy of biorefining processes. Until now, most cellulose-active enzymes have been isolated and characterised from wood-decaying fungi and soil bacteria [6–9]. Cellulose-degrading higher organisms may use symbiotic microbes as a source of enzymes for biomass degradation, such as the marine shipworm which is an eminent lignocellulose degrader [10–12], remain largely unexplored. In the present study, we analysed marine wood-digesting bivalve molluscs called shipworms, which feeds on submerged wood in the ocean.

The shipworms are marine molluscs of the order *Myida* and the family *Teredinidae* (also called “termites of the sea”). They are wood-boring bivalves found throughout the world’s oceans [12]. They are notorious for boring into wooden structures immersed in sea water, where they settle on and begin to excavate into wood as larvae that grow eventually to become elongated worms [13]. Thus, shipworms are unique because only few organisms have evolved the ability to feed on woody biomass as the sole nutrient source [7]. Majority of shipworm species possess a simple digestive system with a large caecum and a short intestine. Previously it was suggested that shipworms may present few microbes in their digestive systems (cecum) that help in wood digestion [14]. However, later it was reported that endosymbiotic bacteria residing in a specialized region of the gill tissue have been shown to fix atmospheric nitrogen [15] but also produce variety of carbohydrate active enzymes (CAZymes) that function in lignocellulose deconstruction in shipworms [16]. In addition to this, shipworms themselves also produce several endogenous CAZymes that are secreted by a specialized digestive gland that finally accumulate in cecum for lignocellulose digestion [10, 11]. Because of these unique features, shipworms are an attractive target for the discovery of new CAZymes for depolymerization of lignocellulose.

The genome analysis of endosymbiotic bacteria in combination with proteome analysis from cecum of shipworms revealed that endosymbionts harbour large number of cellulases that are classified into several glycosyl hydrolase (GH) families with different substrate specificities [16–18]. For instance, GH5 which is a large protein family contains not only endo-glucanases (EC 3.2.1.4), but also β-mannanases (EC 3.2.1.78), exo-1,3-glucanases (EC 3.2.1.58), endo-1,6-glucanases (EC 3.2.1.75), xylanases (EC 3.2.1.8), endo-glycoceramidases (EC 3.2.1.123) and xanthanase [19]. In the quest for novel cellulases, we have performed the identification, cloning, expression, functional analysis and structural characterization of a cellulase belonging to glycoside hydrolase family 5 (GH5). This protein, share high homology to putative GH5 cellulase from the shipworm endosymbiont *Teredinibacter waterburyi*, named *Tw*Cel5, is part of a

large multi-domain cellulase comprised of a N-terminal GH5 domain, followed by three CBM10 domains separated by serine-rich linker region and a C-terminal GH6 domain. In this study, we have produced, and functionally compared two variants of the GH5 enzyme: the catalytic domain only (*Tw*Cel5$_{CAT}$) and the catalytic domain connected to a C-terminal CBM10 (*Tw*Cel5$_{CBM}$). Next to functional characterization of the two variants, we also explored the potential synergistic effect between GH5 cellulase and Cels2, a cellulose-active bacterial LPMO [20], boosted activity for depolymerization of crystalline cellulose. The results showed that *Tw*Cel5 is an endo-β-1,4-glucanase with catalytic properties profile that renders it as a potentially attractive industrial biocatalyst for cellulose bioconversion.

Results And Discussion

Sequence analysis and structural modelling

The 3312 nucleotides gene sequence (GenBank; OP793796) encode a multidomain protein (WAK85940.1), *Tw*Cel5-6, consisting of 1103 amino acids residues that possess a putative N-terminal signal peptide for protein secretion. The deduced amino acid sequence analysis using Interpro classified *Tw*Cel5-6 protein into putative N-terminal glycosyl hydrolase family 5 (GH5) domain (amino acids 15–322) and a C-terminal glycosyl hydrolase family 6 (GH6) domain (amino acids 690–1103), which are interspaced by 368 amino acids long region encoding three cellulose binding CBM10 modules (Fig. 1A). The five modules are connected by poly-serine linkers that are thought to be flexible, disorganized spacers [21]. A BLAST search using the deduced amino acid sequence of *Tw*Cel5-6 revealed the closest 99.49% sequence identity with a hitherto uncharacterized glycosyl hydrolase (WP_223144885.1) from the shipworm gill endosymbiont *Teredinibacter waterburyi* [22]. Notably, the latter partial protein only comprised of 773 amino acids residues, contain N-terminal putative GH5 domain (amino acids 15–322) and incomplete C-terminal GH6 domain (amino acids 688–773) which are interspaced by three cellulose binding CBM10 modules. Alphafold2 structure prediction [23] of *Tw*Cel5-6 through Colabfold [24] showed, as predicted, an unstructured N-terminus (most likely the signal peptide) followed by five modules all connected by flexible linkers (Fig. 1B; see additional file 1: Fig. S1 for prediction quality). Interestingly, Alphafold modelling predicts that the two first CBM10 modules contain a disulfide bridge that connects the end of the N-terminal linker with the start of the C-terminal linker (Fig. 1C).

Protein expression, purification and activity screening of TwCel5

Attempts to produce the full-length protein and several truncated variants failed (data not shown), except for the GH5 catalytic domain alone, *Tw*Cel5$_{CAT}$, and the catalytic domain connected to one CBM10, *Tw*Cel5$_{CBM}$, which were both expressed in soluble and active form. cDNA encoding the GH5 catalytic module of the full-length protein was cloned without the predicted signal peptide (Fig S1). Both proteins carried an affinity tag, such as *Tw*Cel5$_{CAT}$ contained C-terminal 6xHis-tag and *Tw*Cel5$_{CAT}$ contained N-

terminal V5-6xHis-tag that were purified to homogeneity by using affinity chromatography followed by size exclusion chromatography. SDS-PAGE analysis of the purified proteins showed single homogenous protein bands migrating at around ~ 35 kDa and ~ 45 kDa for *Tw*Cel5$_{CAT}$ and *Tw*Cel5$_{CBM}$ (Fig. 2A), which is in accordance with the calculated theoretical protein masses of 33.7 kDa and 44.9 kDa, respectively.

Based on CAZy database (available at http://www.cazy.org/GH5.html), the GH5 family contains a wide range of enzymes acting on diverse β-linked substrates, including cellulose. Initial activity screening experiments indicated that the two enzyme variants of *Tw*Cel5 was able to hydrolyse amorphous cellulose (PASC), but only negligible hydrolytic activity observed towards crystalline Avicel and Whatman® paper (Fig. 2B). The absence of enzymatic activity toward crystalline cellulosic substrates, indicates that the *Tw*Cel5 variants have a preference for the amorphous regions of cellulose PASC similar to, e.g., the KG35 (GH5) endo-β-1,4-glucanase that was obtained from the black-goat rumen [25]. The two variants showed comparable activities for all three substrates, which is somewhat surprising since one would expect that the presence of a CBM10 in *Tw*Cel5$_{CBM}$ is beneficial for activity on insoluble substrates, particularly at the low substrate concentrations used in this study [26–28]. The similarity in activity was confirmed by recording comparative progress curves for PASC (Fig. 2C). This result may be taken to indication that the CBM10 has little affinity for PASC substrate than cellulose or that function of this domain cannot be observed with this truncated variant of *Tw*Cel5-6.

Enzymatic activity on other polysaccharides

The substrate specificity of *Tw*Cel5 was assessed by measuring the hydrolytic activity on nine different soluble cellulosic and hemicellulosic substrates. Both variants showed highest activity for reactions with mixed linkage β-1,3, β-1,4 β-glucan (Fig. 3). Furthermore, the two enzyme variants also exhibited (lower) activity on CMC and konjac glucomannan that contain β-1,4 glycosidic linkages, although the enzyme without CBM produced slightly higher amounts of reducing sugars. Both enzymes were inactive on hemicellulosic substrates includes xylan, arabinogalactan, arabinan, gum arabic, xyloglucan, and lichenan (Fig. 3). It thus seems clear that the two enzyme variants prefer substrates with β-1,4-linked glucose monomers. This is in line with previous findings for several GH5 enzymes that showed similar substrate specificities [29–31]. Of note, the enzyme showed no detectable activity toward ivory nut mannan (data not shown), which is a pure mannose polymer linked via β-1,4 glycosidic linkage. It is thus likely that β-1,4 glycosidic bonds between glucose containing substrates are the target for the *Tw*Cel5 enzyme when hydrolysing konjac glucomannan, which consists of β-1,4-linked mannose, and glucose residues in a 60:40 ratio. These biochemical properties highlight the high specificity of *Tw*Cel5 towards glucans for *P. megotara* in the context of metabolism and nutrition.

The products analysis from enzymatic hydrolysis

To gain more insight into the mode of action of *Tw*Cel5, products generated by the enzyme after 60-minute incubation with barley β-glucan and konjac glucomannan were analysed by MALDI-TOF mass spectrometry. The MALDI-TOF MS spectra showed that *Tw*Cel5$_{CAT}$ hydrolysed barley β-glucan to products ranging from a degree of polymerization from 4 (DP4) to DP14 (rather than only low DP products), suggesting that the enzyme mostly attacks internal glycosidic linkages (Fig. 4A, upper panel). Notably, products with DP6, DP9 and DP12 were not detected. Although this does not exactly shows how the β-glucan is hydrolysed, (barely β-glucan is a linear homopolysaccharide of consecutively linked β-(1,4)-glucosyl residues, i.e., oligomeric cellulose segments, that are separated by single β-(1,3)-linkages [32]. However, it does show that certain linkages, likely β-(1,3) linkages, are not cleaved by the enzyme. The product spectrum for konjac glucomannan showed a continuum of oligosaccharides ranging from DP4 to DP14 (Fig. 4A, lower panel), which one would expect if the ratio of glucose to mannan distribution is random, which is the case [33]. Konjac glucomannan contains about 5–10% of the acetylated sugars [34], and the oligosaccharides observed at masses m/z 2189 ([DP13 + Na + acetyl]+) and 2352 ([DP14 + Na + acetyl]+), indeed suggest the presence of acetylation. Commensurate with these analyses, products generated from PASC was analysed by HPAEC-PAD showed a major cellobiose peak and lesser peaks for glucose, and cellotriose (Fig. 4B), as is commonly observed for cellulases. In summary, the observed activities, the distribution of products and the cleavage patterns strongly indicate that *Tw*Cel5 is a β-(1,4)-endoglucanase with a mode of action resembling that of other known GH5 enzymes [30, 35, 36].

TwCel5 displays broad pH and temperature stability

Using barley β-glucan as substrate, the pH and temperature stability of *Tw*Cel5$_{CAT}$ and *Tw*Cel5$_{CBM}$ were determined to further characterize their optimal activities under different physicochemical conditions. When incubated at 30 ºC, both variants were active over a broad pH range, from 5 to 8, with maximum activity observed at around pH 7.0–7.5 (Fig. 5). More than 70% activity was retained at between pH 4.5 and 9.6, whereas activity was drastically reduced when moving to more extreme pHs (Fig. 5). Relatively broad pH optima are commonly observed for several known GH5 enzymes, e.g., cellulase 5 from sugarcane soil metagenome [30], and endo-cellulases [31, 37], although GH5 enzymes with slightly acidic pH optima have also been reported [38, 39]. It is worth noting that the enzyme remained active for at least 60 minutes when incubated at from pH 4.0 to pH 10.6, suggesting that the protein is quite stable in a wide pH range.

Furthermore, *Tw*Cel5 activity increased in a temperature dependent manner reaching maximum activity at temperatures between 40–50°C range (Fig. 6). At temperature above 50°C signs of enzyme inactivation became noticeable during the 60 minutes incubation period and this effect became stronger for the enzyme connected with the CBM that showed slightly low activity. It is worth noting that *Tw*Cel5$_{CAT}$ retained over 50% hydrolytic activity at temperatures ranging from 10°C to 60°C, which suggests that this protein is both cold-adapted and moderately thermotolerant (Fig. 6). Although, shipworms are adapted to a cold environment, it has previously been reported that enzymes from shipworm symbionts exhibit biochemical properties similar to *Tw*Cel5$_{CAT}$ for instance a cellulolytic endo-1,4-β-glucanase obtained

from shipworm *Lyrodus pedicellatus* [37]. It is also interesting observe that, to the best of our knowledge, the combination of being active at rather alkaline pH and being thermo-tolerant is rather rare and unusual enzyme feature [40–42]. As *Tw*Cel5 is originating from a marine shipworm symbiont that reside in ocean water, we assessed the impact of NaCl salt concentration on enzyme hydrolytic activity. Using standard assay conditions, we found that the no impact of 0–1.5 M NaCl concentration on enzyme activity (additional file 1: Fig. S2). In conclusion, *Tw*Cel5 is a salt-tolerant β-1,4-endo-glucanase capable of function in a wide pH and temperatures ranges, which makes it an interesting candidate relevant for industrial applications.

Co-incubation of TwCel5 with LMPO boosted hydrolytic activity

It is well known that shipworm gill endosymbionts produce a multitude of carbohydrate active enzymes, including cellulose active lytic polysaccharide monooxygenases (LPMOs) and diverse glycosyl hydrolases (GHs) to accomplish the efficient wood digestion for nutrition and growth [16, 43]. Recent study using the combination of meta-transcriptome, proteomic and biochemical analysis from wood-degrading shipworm *Lyrodus pedicellatus* reported the expression of gene encoding auxiliary activity family 10 (AA10) LPMO which likely synergise with endogenous as well as endosymbiont multi-domain glycoside hydrolases that functions in the hydrolysis of β-1,4-glucans [10]. It was thus of great interest to determine whether a typical cellulose-active AA10 LPMO, CelS2 from *Streptomyces coelicolor* A3(2) [20, 44] would enhance the hydrolytic activity of *Tw*Cel5$_{CAT}$ or *Tw*Cel5$_{CBM}$ specially on Avicel, a semi-crystalline form of cellulose. As expected, clear synergistic effect was observed on the generation of reducing sugars when CelS2 and *Tw*Cel5$_{CAT}$ or *Tw*Cel5$_{CBM}$ were combined as enzyme cocktail (Fig. 7A). Interestingly, in the reaction with CelS2, the amount of reducing sugars released was higher for enzyme that connected to CBM10, *Tw*Cel5$_{CBM}$ compared to catalytic *Tw*Cel5$_{CAT}$. This may be taken to suggestion that LPMO activity uncovers the regions of crystalline substrate where the CBM10 is partially beneficial for the efficiency of GH5. Of note, a control reaction (Fig. 7B) showed that, potentially apparent, synergy is not just a result of GH5-catalyzed hydrolysis of longer soluble cello-oligomers generated by the LPMO. Our results add to studies showing that cellulose active LPMO boosts the activity of shipworm glycoside hydrolases on crystalline cellulose such as Avicel [43] and support the notion that LPMO action is important for wood depolymerization in shipworms for digestion and nutrition.

Structural analysis of cellulase TwCel5$_{CAT}$

To gain structural insights for *Tw*Cel5$_{CAT}$, bond cleavage pattern and mode of action, we solved and determined the tertiary structure of catalytic module. *Tw*Cel5$_{CAT}$ was crystallized in an apo form and a dataset diffracting to 1.0 Å resolution was collected (Table 1). The structure was determined by molecular replacement using the protein coordinates from the structure of the GH5 cellulase Cel5 (PDB entry 1EGZ)

from *Erwinia chrysanthemi*, a gram-negative plant pathogen [45] as the search model (Table 2). *Tw*Cel5$_{CAT}$ model was refined to R_{work} and R_{free} of 11.42 and 13.30, respectively, and the final model was deposited in the PDB database (PDB identifier 8C10). Structural comparisons using the DALI server [46] revealed highest closest structural matches to the cellulase CelE1 belonging to GH5 family (PDB identifier 4M1R; 67% sequence identity) obtained from a sugarcane soil metagenome [30], and endoglucanase EGZ (PDB identifier 1EGZ; 67% sequence identity) from *Erwinia chrysanthemi* [45, 47], respectively.

The first three-dimensional structure of a CAZY family GH5 subfamily 2 was a cellulase isolated from an alkaline *Bacillus* sp., found in soda lakes [48]. Consistent with the defining family member, *Tw*Cel5$_{CAT}$ exhibits a classical $(\beta/\alpha)_8$-barrel fold (also known as TIM-barrel; Fig. 8A) with two conserved glutamates; Glu157 & Glu245 in the catalytic center (Fig. 8B) that are positioned to promote the expected displacement mechanism characteristic of family 5 cellulases [49]. Thus, from homology to other GH family 5 cellulases, the proton donor is expected to be Glu157, and the nucleophile is Glu228. Furthermore, substrate binding cleft of *Tw*Cel5$_{CAT}$ is homologous to other clade 2 GH5 enzymes, showing conserved aromatic and polar amino acids involved in substrate binding (Fig. 8C). In contrast to most other GH5s subfamily 2, *Tw*Cel5$_{CAT}$ has a tryptophan (Trp226) in the outer region of the reducing end subsites (Fig. 8D). Of the 15 unique structures from GH5s subfamily 2, only three have a Trp in this position, namely the thermostable GsCelA from *Geobacillus* sp. 70PC53 [50], the halotolerant Cel5R cloned from a soil metagenome [51] and Cel5B from *C. hutchinsonii* [52], respectively. Of note, comparison of the Alphafold2 model of *Tw*Cel5$_{CAT}$ with the X-ray crystallographic structure showed RMSD for Ca-carbons of only 0.35 Å (Fig. S3A). Furthermore, the side chains of the amino acids in the conserved active site and substrate binding cleft were modelled clos to perfectly (Fig S3B).

Table 1
Data collection and processing statistics. Values in parentheses are for the outermost shell.

Diffraction source	**BESSY II BL 14.1**
Wavelength (Å)	0.91842
Detector	PILATUS 6M
Crystal-to-detector distance (mm)	180.8
Rotation range pr. Image (°)	0.1
Total rotation range (°)	200
Space group	$P2_12_12_1$
a, b, c (Å)	45.04, 68.43, 87.19
a, b, g (°)	90, 90, 90
Mosaicity (°)	0.05
Resolution range (Å)	45.04-1.00 (1.02-1.00)
Total No. of reflections	941741 (15841)
No. of unique reflections	143241 (5490)
Completeness (%)	98.0 (77.7)
Multiplicity	6.6 (2.9)
<I/s(I)>	11.5 (1.5)
R_{merge}	0.077 (0.609)
Overall B factor from Wilson plot ($Å^2$)	6.32

Table 2
Structure determination and refinement statistics. Values in parentheses are for the outermost shell.

PDB code	**8C10**
Resolution range (Å)	45.04-1.00 (1.02-1.00)
Completeness (%)	97.85 (79.7)
s cutoff	None
No. of reflection, working set	140934
No. of reflection, reference set	2221
Final R_{work}	11.42
Final R_{free}	13.39
Cruickshank DPI	0.0178
No. of non-H atoms	
Protein	2427
Water	339
Other*	19
Total	2785
R.m.s. deviations	
Bonds (Å)	0.0291
Angles (°)	2.31
Average B factors ($Å^2$)	
Overall	11.37
Protein	9.71
Water	22.75
Other*	20.70
Ramachandran plot (%)	
Preferred	96.61
Allowed	3.39
Outliers	0

Conclusions

This study describes the detailed biochemical characterization and structural properties of GH5 domain of multi-domain cellulase *Tw*Cel5-6 from the shipworm's endosymbiont *T. waterburyi* residing in the gills of shipworm including *P. megotara*. It is an endo-glucanase enzyme showing activity at broader pH from 5–8 and temperatures 40–50°C, respectively. *Tw*Cel5$_{CAT}$ hydrolytic activity on crystalline Avicel was boosted upon synergistic interaction with cellulose oxidizing CelS2, and in this reaction set-up the presence of a C-terminal CBM10 domain fused to *Tw*Cel5$_{CAT}$ promoted enhanced cellulose saccharification. This endo-glucanase may be a suitable biocatalyst for liberating reducers sugars from pre-treated biomass in a broader pH and temperature range, including alkaline pH, low temperature and high salt tolerance. In summary, our study demonstrates that wood-digesting shipworms are good source of novel enzymes active at alkaline pHs and moderately thermostable cellulases.

Methods

Chemicals and substrates

Analytical grade substrates were used. The pre-packed 5 mL HisTrap™ affinity (HP), PD-10 desalting columns (Sephadex G-25 resin), and size exclusion chromatography column (HiLoad® Superdex, 75 pg) used for protein purification were obtained from GE Healthcare. Amorphous, phosphoric acid swollen cellulose (PASC) was prepared from Avicel according to method described before [53]. High-purity substrates include barley β-glucan, lichenan, konjac glucomannan, wheat arabinoxylan, birchwood xylan, and tamarind xyloglucan purchased from Megazyme. The model crystalline cellulose substrate Avicel PH-101, carboxymethyl cellulose (CMC), gum arabic and standard cello-oligomers purchased from Sigma-Aldrich.

Sample collection and identification of genes

Specimens of adult shipworm *Psiloteredo megotara* collected from Norway spruce (*Picea abies*) wooden panels submerged were for about 8–9 months in the Arctic Sea near Tromsø, Norway (N 69°46'47.515"; E 18°23'53.143"). The sampling was done accordance with the Norwegian Marine Resource Act. The shipworm specimen was initially rinsed with sterile water and dissected on a clean bench to separate the specialised gill tissue containing endosymbionts. Bacterial enrichment performed using crushed gill tissues in a medium supplemented with cellulose as a carbon source that were incubated for several months. DNA was isolated from the enrichment culture using the DNeasy Blood & Tissue Kit (QIAGEN). The metagenome sequencing was carried using Illumina MiSeq 300 paired-end chemistry at the

Norwegian Sequencing Centre (www.sequencing.uio.no). Analysis of contigs were assembled, annotated and uploaded to the GenBank sequence database (accession number grp 8783669). Full details of the metagenomic dataset will be published elsewhere. Genes coding for carbohydrate-active enzymes were mined using the dbCAN meta server [54]. A gene with accession number OP793796 (3312 bp) that located on BankIt2638814 Contig82 was chosen for further studies due to its novel multi-domain architecture.

Gene cloning, expression, and protein purification

A gene (OP793796) encoding multidomain protein *Tw*Cel5-6 has an accession number WAK85940.1 (codon optimized for *E. coli* expression using the OptimumGene™ PSO algorithm) was synthesized by GenScript Biotech (Piscataway, NJ 08854, USA). Gene fragments encoding *Tw*Cel5$_{CAT}$ and *Tw*Cel5$_{CBM}$ were generated using through PCR using Q5 DNA polymerase (New England Biolabs, Ipswich, MA) and the primers described in Table S1. Both the genes were amplified excluding a putative signal peptide (additional file: Fig. S4), as predicted using the SignalP-5.0 prediction tool [55]. PCR products were purified using a PCR clean-up kit as per manufacturer's instructions (Macherey-Nagel, Germany) followed by agarose (1 w/v %) gel electrophoresis. Prior to cloning, the DNA concentration was determined using a nano UV spectrophotometer (Thermo Scientific, San Jose, CA, USA). The DNA fragment encoding *Tw*Cel5$_{CAT}$ was cloned (26–320 amino acid residues) into the pNIC-CH expression vector (AddGene, Cambridge, MA), which adds a C-terminal polyhistidine-tag (6xHis-tag) to the protein as per manufactures instructions. Similarly, DNA fragment encoding *Tw*Cel5$_{CBM}$ was cloned (30–412 amino acid residues) into the directional champion pET151/D-TOPO™ expression vector (Invitrogen, Carlsbad, CA), which adds a cleavable N-terminal V5-6xHis-tag to the protein as per manufactures instructions.

The recombinant vectors were transformed into chemically competent OneShot™ *E. coli* TOP10 (Invitrogen, Carlsbad, CA) cells using a heat shock transformation method. Cells were grown in SOC medium for 60 minutes prior to plating on lysogenic broth (LB) agar plates supplemented with antibiotics; 50 µg/mL kanamycin (for *Tw*Cel5$_{CAT}$) and 100 mg/mL ampicillin (for *Tw*Cel5$_{CBM}$) depending on the vector, followed by overnight incubation at 37°C. Colonies on the LB plates were picked and screened by colony PCR using the pair of T7 primers: T7, 5´-TAATACGACTCACTATAGGG-3´ and T7 reverse 5´-TAGTTATTGCTCAGCGGTGG-3´. Positive clones were picked and inoculated in liquid LB containing appropriate antibiotics (as mentioned above) and the cultures were incubated overnight at 37°C with shaking at 200 rpm. The recombinant plasmids were isolated from the *E. coli* cells using the Zymo MiniPrep Kit (Zymo Research). Prior to transformation to expression cells, the correct integration and DNA sequence of the genes was confirmed using Sanger sequencing (GATC Biotech, Constance, Germany). The expression vectors were transformed into chemically competent OneShot BL-21 Star™ (DE3) and ArcticExpress (DE3) *E. coli* expression cells, for *Tw*Cel5$_{CAT}$ and *Tw*Cel5$_{CBM}$, respectively, using heat shock method as described above.

To produce recombinant proteins, *E. coli* transformants were inoculated and grown in terrific broth medium supplemented with appropriate antibiotics and cells were incubated at 37°C with horizontal shaking (200 rpm) until the optical density (OD_{600nm}) reached between 0.6 to 0.8 followed by induction by adding 0.3 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 24 h incubation at 15°C with horizontal shaking (200 rpm). Cultures were harvested by centrifugation at 10,000 x g for 15 min at 8°C, using a Beckman coulter centrifuge (Brea, CA, USA). Cells were stored frozen at -20°C until further use. For protein purification, a cell-free extract was prepared by re-suspending about 5 g wet cell biomass in 50 mL of 50 mM Tris-HCl buffer, pH 7.4, supplemented with 200 mM NaCl, 10% glycerol and 30 mM imidazole (lysis buffer). Prior to cell disruption, the suspension was supplied with 0.5–1 mg of cOmplete™, EDTA-free protease inhibitor cocktail (Roche) and lysozyme (0.5 mg/mL). Cells were disrupted using a cooled high-pressure homogenizer (LM20 Microfluidizer, Microfluidics). Cell debris was removed by centrifugation at 27,000 x g for 30 min at 4°C. The resulting cell-free extracts, containing cytosolic soluble proteins, were filtered using a sterile 0.45 µm filter (Sarstedt, Nümbrecht, Germany).

The filtered cell-free extract was subjected to immobilized metal affinity chromatography using an Äkta pure chromatography system equipped with a 5-mL HisTrap HP column (GE HealthCare) equilibrated with lysis buffer (see above). After sample loading, the HisTrap column was washed extensively using 50 mM Tris-HCl, pH 7.4, supplemented with 200 mM NaCl, 10% glycerol containing 70 mM imidazole (wash buffer) until UV absorbance dropped and stabilised at baseline level. Bound proteins were eluted using the same buffer supplemented with 500 mM imidazole (elution buffer). Eluted proteins were first analysed using SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using TGX Stain-Free precast gels (Bio-Rad, Hercules, Ca, USA). The molecular weight of the recombinant proteins was estimated using Invitrogen™ BenchMark™ Pre-stained Protein Ladder (Fisher Scientific, Waltham, Massachusetts, USA). Eluted proteins were concentrated using Vivaspin® 10,000 MWCO centrifugal filter units (Sartorius, Göttingen, Germany). Proteins were purified to homogeneity using a size-exclusion chromatography column (HiLoad® Superdex, 75 pg) pre equilibrated with 50 mM sodium phosphate buffer (pH 7.4) containing 150 mM NaCl. Finally, buffer exchange was performed to 50 mM sodium citrate (pH 5.6) using a PD-10 desalting column. The protein concentration was determined by UV absorbance at 280 nm (A_{280}) using theoretical molar extinction coefficients (*Tw*Cel5$_{CAT}$: 71390 $M^{-1}\cdot cm^{-1}$ and *Tw*Cel5$_{CBM}$: 96745 $M^{-1}\cdot cm^{-1}$) estimated using the ProtParam tool [56]. Purified enzymes were stored at 4°C.

Biochemical characterization of recombinant TwCel5

The standard assays were performed in 300 µL reaction volume containing 50 mM sodium phosphate buffer (pH 7.5), 0.5 M NaCl, 0.5% (w/v) β-glucan and 50 nM of the purified enzyme. The reaction was started either by the addition of enzyme or substrate followed by incubation at 30 ºC using a thermomixer with horizontal agitation (500 rpm; Eppendorf, Hamburg, Germany) for 60 minutes (unless stated otherwise). To determine the pH stability and pH optima, reactions were carried out in 50 mM buffer systems (sodium citrate, pH 3.0–6.0; potassium phosphate, pH 6.5–8.0; and glycine-NaOH, 9.6–10.6) containing 0.5 M NaCl, 0.5% (w/v) β-glucan and 50 nM of the purified enzyme. To determine the thermal

stability and optimal temperature, reactions were performed in 50 mM sodium phosphate buffer (pH 7.5), 0.5 M NaCl, 0.5% (w/v) β-glucan and 50 nM of the purified enzyme at various temperatures ranging from 5–70°C. The effect of the NaCl concentration on activity was determined using 50 mM sodium phosphate buffer (pH 7.5), 0.5% (w/v) β-glucan, 50 nM of the purified enzyme and various concentrations of NaCl ranging from 0 to 1.5 M. Aliquots were collected at different time intervals in a period of 60 minutes; reactions were stopped by mixing the samples immediately with DNSA reagent. Product formation was determined by quantifying the amount of reducing end sugars using the 3,5-dinitrosalicylic acid (DNSA) assay method [57] using glucose as a standard. The absorbance (A_{540nm}) was recorded using Varioskan™ LUX multimode microplate reader (Thermo Scientific, San Jose, CA, USA). All the assays were performed in triplicate.

Lignocellulose substrate specificity

The substrate specificity of purified *Tw*Cel5$_{CAT}$ and *Tw*Cel5$_{CBM}$ was evaluated using a wide variety of complex lignocellulosic substrates, both soluble and crystalline. The insoluble and model crystalline polysaccharides include Avicel PH-101, Whatman® cellulose filter paper (0.5 μm particle size) and phosphoric-acid swollen cellulose (PASC) was used as amorphous substrate. The soluble lignocellulosic substrates included β-glucan, birchwood xylan, carboxymethyl cellulose (CMC), wheat arabinoxylan, konjac glucomannan, xyloglucan, and lichenan. Soluble substrates were dissolved according to the supplier's protocol. The standard reactions were carried out in 300 μL reaction volume using 50 mM sodium phosphate buffer, pH 7.5, 0.5 M NaCl, using 50 nM of enzyme for soluble substrates (0.5% w/v) whereas 1 μM for insoluble substrates (1% w/v) that were incubated at 30°C for 60 minutes, for soluble substrates, or 24 hours, for insoluble substrates, with horizontal agitation (500 rpm). Aliquots were taken at different intervals; reactions were stopped by mixing the samples immediately with DSNA reagent. Release of reducing end sugars was measured using DNSA assay, using glucose as a standard, as described (see above). When using insoluble substrates samples were filtered before measurement.

Cellulase-LPMO synergy experiment

Purified CelS2 from *Streptomyces coelicolor* was a kind gift from Dr. Zarah Forsberg [20]. The cellulase-LPMO synergy was assessed by performing reactions with crystalline Avicel (1% w/v) in 50 mM sodium phosphate buffer (pH 6.0) using a thermomixer (Eppendorf, Hamburg, Germany) incubated at 30°C with horizontal agitation (1000 rpm). Experiments to determine the synergy were conducted at a fixed total enzyme concentration of, such as 1 μM copper saturated CelS2 and/or 1 μM of one of the *Tw*Cel5 cellulase variants. The reactions were started by firstly supplying 1 mM ascorbic acid (final concentration) to all reaction mixtures, immediately followed by addition of the enzyme. Reactions were incubated for 48 hours, and aliquots were taken at different time intervals followed by filtration using 0.45 μm filter for removing insoluble substrate and to stop the reaction. To check for generation of reducing ends due to the action of the cellulase variants of oligomeric products solubilized by the LPMO,

which could lead to a false impression of synergy, control reactions were performed in which Avicel was first incubated with the LPMO for 48 hours, after which the products were treated with the *Tw*Cel5 variants. Cellulose saccharification was assessed using the reducing end assay described above and all the experiments were performed in triplicate.

Product analysis by HPAEC-PAD (ICS-6000)

Hydrolytic products generated from PASC were detected by a Dionex ICS6000 system (Thermo Scientific, San Jose, CA, USA) using high performance anion exchange chromatography connected to pulsed amperometric detector using CarboPac™ PA200 IC analytical column. The eluent B (0.1 M NaOH and 1 M sodium acetate) and eluent A (0.1 M NaOH) was applied using following gradient program: 0–5.5% B for 3 min, 5.5–15% B for 6 min, 15–100% B for 11 min, 100–0% B for 6 s, 0% B for 6 min. The eluent flow rate was set to 0.5 mL/min. The cello-oligosaccharide with a degree of polymerization from one to five (DP1 - DP5), was used as standards for product identification. The data was analysed using Chromeleon 7.2.9 software.

Product analysis by MALDI-TOF MS

Hydrolytic products generated from β-glucan and konjac glucomannan were identified using a UltrafleXtreme matrix assisted laser desorption ionization time of flight mass spectrometer (Bruker Daltonics GmbH, Bremen, Germany) equipped with a Nitrogen 337-nm laser. Samples were prepared by mixing one microliter of sample with two microliter of 2,5-dihydrooxybenzoic acid (DHB) solution (9 mg/mL) that directly mounted onto a MTP 384 ground steel target plate (Bruker Daltonics). Sample spots were allowed to dry on the plate using a flow of dry hot air. Data was acquired using Bruker flexControl and flexAnalysis software. Products were identified based on *m/z* values.

Crystallization, data collection and analysis

Crystallization experiments were performed with a stock solution of the purified protein at 12.4 mg/mL, as estimated by A_{280}, in 6 mM NaCl, 20 mM Tris-HCl at pH 8.0. Initial crystallization experiments were performed using the vapour diffusion sitting drop method set up by a Phoenix crystallization robot (Art Robbins Instruments). The crystallisation experiments were set up with sixty µl reservoir solutions and sitting drops with equal amounts of reservoir solution mixed with protein stock solution in a total drop volume of one microliter. The plates were incubated at 20°C. Crystals appeared in 1–2 weeks in conditions containing 0.2 M $MgCl_2$, 0.1 M Tris-HCl, pH 8.5, and 25% w/v of polyethylene glycol 3350 (PEG 3350). Crystals were harvested, transferred to a cryoprotectant solution consisting of the reservoir solution supplemented with 15% ethylene glycol and flash cooled in liquid N_2. X-ray diffraction data were collected at BEAMLINE 14.1 at BESSY II (Berlin, Germany). Data collection and processing statistics are presented in Table 1, and structure determination and refinement statistics are presented in Table 2. The

crystal structure was solved by molecular replacement using MolRep in the CCP4 program package [58] with 1egz.pdb as search model [45]. The initial refinement was executed in Refmac [59] followed by automated model improvement in Buccaneer [60]. The manual model building was done in Coot [61] interspersed by cycles of refinement in Refmac and resulted in a final R_{work}/R_{free} of 11.42/13.39. The atomic coordinates and structural details have been deposited in the Protein Data Bank with the accession code 8C10. Figures presented in the results section were generated using Pymol v4.60 (www.pymol.org).

Declarations

Ethics approval and consent to participate:

Not applicable

Consent for publication:

Not applicable

Availability of data and materials:

All the data is available for publication and information used from online resources has been cited properly. All data generated or analysed during this study are included in this published article [and its supplementary information files].

Competing interests:

Not applicable

Funding:

ERA-NET MarineBiotech (ERA-MBT), provided by the Research Council of Norway (grant number 283647), and by the NorZymeD and OXYMOD projects financed by Research Council of Norway (grant numbers: 221568 and 269408), respectively.

Authors contributions

MJ contributed majorly to perform the experiments, data analysis and writing of the first draft of the manuscript. TV helped to perform cellulase-LPMO synergy experiment. LF, BA and IL helped to perform crystallization and analysis of data. VE helped to revise the final version of the manuscript. GVK supervised the project and revised the final manuscript. All authors read and approved the final manuscript.

Acknowledgements

This research work received the financial support to Dr. Madan Junghare, position by a grant from ERA-NET MarineBiotech, provided by the Research Council of Norway (grant number 283647), and by the NorZymeD and OXYMOD projects financed by Research Council of Norway (grant numbers: 221568 and 269408), respectively.

References

1. Klemm D, Heublein B, Fink HP, Bohn A. Cellulose: fascinating biopolymer and sustainable raw material. Angew Chem Int Ed Engl. 2005;44(22):3358-93; doi: 10.1002/anie.200460587.
2. Mariano M, El Kissi N, Dufresne A. Cellulose nanocrystals and related nanocomposites: Review of some properties and challenges. Journal of Polymer Science Part B: Polymer Physics. 2014;52(12):791-806; doi: https://doi.org/10.1002/polb.23490.
3. Zhang Z, Donaldson AA, Ma X. Advancements and future directions in enzyme technology for biomass conversion. Biotechnol Adv. 2012;30(4):913-9; doi: 10.1016/j.biotechadv.2012.01.020.
4. Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, et al. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. Science. 2007;315(5813):804-7; doi: doi:10.1126/science.1137016.
5. Østby H, Hansen LD, Horn SJ, Eijsink VGH, Várnai A. Enzymatic processing of lignocellulosic biomass: principles, recent advances and perspectives. Journal of Industrial Microbiology and Biotechnology. 2020;47(9-10):623-57; doi: 10.1007/s10295-020-02301-8.
6. Andlar M, Rezić T, Marđetko N, Kracher D, Ludwig R, Šantek B. Lignocellulose degradation: An overview of fungi and fungal enzymes involved in lignocellulose degradation. Eng Life Sci. 2018;18(11):768-78; doi: 10.1002/elsc.201800039.
7. Cragg SM, Beckham GT, Bruce NC, Bugg TD, Distel DL, Dupree P, et al. Lignocellulose degradation mechanisms across the Tree of Life. Curr Opin Chem Biol. 2015;29:108-19; doi: 10.1016/j.cbpa.2015.10.018.
8. Berlemont R, Martiny AC. Genomic potential for polysaccharide deconstruction in bacteria. Appl Environ Microbiol. 2015;81(4):1513-19; doi: 10.1128/aem.03718-14.
9. Bhardwaj N, Kumar B, Agrawal K, Verma P. Current perspective on production and applications of microbial cellulases: a review. Bioresources and Bioprocessing. 2021;8(1):95; doi: 10.1186/s40643-021-00447-6.
10. Sabbadin F, Pesante G, Elias L, Besser K, Li Y, Steele-King C, et al. Uncovering the molecular mechanisms of lignocellulose digestion in shipworms. Biotechnology for Biofuels. 2018;11(1):59; doi: 10.1186/s13068-018-1058-3.
11. Stravoravdis S, Shipway JR, Goodell B. How Do Shipworms Eat Wood? Screening Shipworm Gill Symbiont Genomes for Lignin-Modifying Enzymes. Frontiers in Microbiology. 2021;12; doi: 10.3389/fmicb.2021.665001.

12. Distel DL. The Biology of Marine Wood Boring Bivalves and Their Bacterial Endosymbionts. In: Wood Deterioration and Preservation. American Chemical Society; 2003. p. 253-71.
13. Nair NB, Saraswathy M. The Biology of Wood-Boring Teredinid Molluscs. In: Russell FS, Yonge M, editors. Advances in Marine Biology. Academic Press; 1971. p. 335-509.
14. Distel DL, Roberts SJ. Bacterial Endosymbionts in the Gills of the Deep-Sea Wood-Boring Bivalves Xylophaga atlantica and Xylophaga washingtona. The Biological Bulletin. 1997;192(2):253-61; doi: 10.2307/1542719.
15. Lechene CP, Luyten Y, McMahon G, Distel DL. Quantitative Imaging of Nitrogen Fixation by Individual Bacteria Within Animal Cells. Science. 2007;317(5844):1563-6; doi: 10.1126/science.1145557.
16. O'Connor RM, Fung JM, Sharp KH, Benner JS, McClung C, Cushing S, et al. Gill bacteria enable a novel digestive strategy in a wood-feeding mollusk. Proceedings of the National Academy of Sciences. 2014;111(47):E5096-E104; doi: doi:10.1073/pnas.1413110111.
17. Brito TL, Campos AB, von Meijenfeldt FAB, Daniel JP, Ribeiro GB, Silva GGZ, et al. The gill-associated symbiont microbiome is a main source of woody-plant polysaccharide hydrolase genes and secondary metabolite gene clusters in Neoteredo reynei, a unique shipworm from south Atlantic mangroves. bioRxiv. 2018:357616; doi: 10.1101/357616.
18. Pesante G, Sabbadin F, Elias L, Steele-King C, Shipway JR, Dowle AA, et al. Characterisation of the enzyme transport path between shipworms and their bacterial symbionts. BMC Biology. 2021;19(1):233; doi: 10.1186/s12915-021-01162-6.
19. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 2022;50(D1):D571-d7; doi: 10.1093/nar/gkab1045.
20. Forsberg Z, Vaaje-Kolstad G, Westereng B, Bunæs AC, Stenstrøm Y, MacKenzie A, et al. Cleavage of cellulose by a CBM33 protein. Protein Sci. 2011;20(9):1479-83; doi: 10.1002/pro.689.
21. Howard MB, Ekborg NA, Taylor LE, Hutcheson SW, Weiner RM. Identification and analysis of polyserine linker domains in prokaryotic proteins with emphasis on the marine bacterium Microbulbifer degradans. Protein Sci. 2004;13(5):1422-5; doi: 10.1110/ps.03511604.
22. Altamia MA, Shipway JR, Stein D, Betcher MA, Fung JM, Jospin G, et al. Teredinibacter waterburyi sp. nov., a marine, cellulolytic endosymbiotic bacterium isolated from the gills of the wood-boring mollusc Bankia setacea (Bivalvia: Teredinidae) and emended description of the genus Teredinibacter. Int J Syst Evol Microbiol. 2020;70(4):2388-94; doi: 10.1099/ijsem.0.004049.
23. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9; doi: 10.1038/s41586-021-03819-2.
24. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022;19(6):679-82; doi: 10.1038/s41592-022-01488-1.
25. Song Y-H, Lee K-T, Baek J-Y, Kim M-J, Kwon M-R, Kim Y-J, et al. Isolation and characterization of a novel endo-β-1,4-glucanase from a metagenomic library of the black-goat rumen. Brazilian Journal

of Microbiology. 2017;48(4):801-8; doi: https://doi.org/10.1016/j.bjm.2017.03.006.

26. Várnai A, Siika-Aho M, Viikari L. Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. Biotechnol Biofuels. 2013;6(1):30; doi: 10.1186/1754-6834-6-30.
27. Bernardes A, Pellegrini VOA, Curtolo F, Camilo CM, Mello BL, Johns MA, et al. Carbohydrate binding modules enhance cellulose enzymatic hydrolysis by increasing access of cellulases to the substrate. Carbohydr Polym. 2019;211:57-68; doi: 10.1016/j.carbpol.2019.01.108.
28. Chalak A, Villares A, Moreau C, Haon M, Grisel S, d'Orlando A, et al. Influence of the carbohydrate-binding module on the activity of a fungal AA9 lytic polysaccharide monooxygenase on cellulosic substrates. Biotechnol Biofuels. 2019;12:206; doi: 10.1186/s13068-019-1548-y.
29. Liberato MV, Silveira RL, Prates ÉT, de Araujo EA, Pellegrini VOA, Camilo CM, et al. Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism. Scientific Reports. 2016;6(1):23473; doi: 10.1038/srep23473.
30. Alvarez TM, Paiva JH, Ruiz DM, Cairo JP, Pereira IO, Paixão DA, et al. Structure and function of a novel cellulase 5 from sugarcane soil metagenome. PLoS One. 2013;8(12):e83635; doi: 10.1371/journal.pone.0083635.
31. Naas AE, MacKenzie AK, Dalhus B, Eijsink VGH, Pope PB. Structural Features of a Bacteroidetes-Affiliated Cellulase Linked with a Polysaccharide Utilization Locus. Sci Rep. 2015;5:11666; doi: 10.1038/srep11666.
32. Lazaridou A, Biliaderis CG. Molecular aspects of cereal β-glucan functionality: Physical properties, technological applications and physiological effects. Journal of Cereal Science. 2007;46(2):101-18; doi: https://doi.org/10.1016/j.jcs.2007.05.003.
33. Takigami S. Chapter 18 - Konjac glucomannan. In: Phillips GO, Williams PA, editors. Handbook of Hydrocolloids (Third Edition). Woodhead Publishing; 2021. p. 563-77.
34. Witkamp RF. 3.15 - Biologically Active Compounds in Food Products and Their Effects on Obesity and Diabetes. In: Liu H-W, Mander L, editors. Comprehensive Natural Products II. Oxford: Elsevier; 2010. p. 509-45.
35. Barras F, Bortoli-German I, Bauzan M, Rouvier J, Gey C, Heyraud A, et al. Stereochemistry of the hydrolysis reaction catalyzed by endoglucanase Z from Erwinia chrysanthemi. FEBS Letters. 1992;300(2):145-8; doi: https://doi.org/10.1016/0014-5793(92)80183-H.
36. Bischoff KM, Rooney AP, Li XL, Liu S, Hughes SR. Purification and characterization of a family 5 endoglucanase from a moderately thermophilic strain of Bacillus licheniformis. Biotechnol Lett. 2006;28(21):1761-5; doi: 10.1007/s10529-006-9153-0.
37. Xu P-N, Distel DL. Purification and characterization of an endo-1, 4-β-D glucanase from the cellulolytic system of the wood-boring marine mollusk Lyrodus pedicellatus (Bivalvia: Teredinidae). Marine Biology. 2004;144(5):947-53; doi: 10.1007/s00227-003-1251-0.
38. Santos CR, Paiva JH, Sforça ML, Neves JL, Navarro RZ, Cota J, et al. Dissecting structure-function-stability relationships of a thermostable GH5-CBM3 cellulase from Bacillus subtilis 168. Biochem J.

2012;441(1):95-104; doi: 10.1042/bj20110869.

39. Liu J, Liu WD, Zhao XL, Shen WJ, Cao H, Cui ZL. Cloning and functional characterization of a novel endo-β-1,4-glucanase gene from a soil-derived metagenomic library. Appl Microbiol Biotechnol. 2011;89(4):1083-92; doi: 10.1007/s00253-010-2828-4.
40. Annamalai N, Thavasi R, Vijayalakshmi S, Balasubramanian T. A novel thermostable and halostable carboxymethylcellulase from marine bacterium Bacillus licheniformisAU01. World Journal of Microbiology and Biotechnology. 2011;27(9):2111-5; doi: 10.1007/s11274-011-0674-x.
41. Hakamada Y, Koike K, Yoshimatsu T, Mori H, Kobayashi T, Ito S. Thermostable alkaline cellulase from an alkaliphilic isolate, Bacillus sp. KSM-S237. Extremophiles. 1997;1(3):151-6; doi: 10.1007/s007920050028.
42. Yin YR, Zhang F, Hu QW, Xian WD, Hozzein WN, Zhou EM, et al. Heterologous expression and characterization of a novel halotolerant, thermostable, and alkali-stable GH6 endoglucanase from Thermobifida halotolerans. Biotechnol Lett. 2015;37(4):857-62; doi: 10.1007/s10529-014-1742-8.
43. Fowler CA, Sabbadin F, Ciano L, Hemsworth GR, Elias L, Bruce N, et al. Discovery, activity and characterisation of an AA10 lytic polysaccharide oxygenase from the shipworm symbiont Teredinibacter turnerae. Biotechnology for Biofuels. 2019;12(1):232; doi: 10.1186/s13068-019-1573-x.
44. Forsberg Z, Mackenzie AK, Sørlie M, Røhr ÅK, Helland R, Arvai AS, et al. Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. Proceedings of the National Academy of Sciences. 2014;111(23):8446-51; doi: doi:10.1073/pnas.1402771111.
45. Chapon V, Czjzek M, El Hassouni M, Py B, Juy M, Barras F. Type II protein secretion in gram-negative pathogenic bacteria: the study of the structure/secretion relationships of the cellulase cel5 (formerly EGZ) from Erwinia chrysanthemi11Edited by I. B. Holland. Journal of Molecular Biology. 2001;310(5):1055-66; doi: https://doi.org/10.1006/jmbi.2001.4787.
46. Holm L, Laakso LM. Dali server update. Nucleic Acids Res. 2016;44(W1):W351-5; doi: 10.1093/nar/gkw357.
47. Py B, Bortoli-German I, Haiech J, Chippaux M, Barras F. Cellulase EGZ of Erwinia chrysanthemi: structural organization and importance of His98 and Glu133 residues for catalysis. Protein Eng. 1991;4(3):325-33; doi: 10.1093/protein/4.3.325.
48. Shaw A, Bott R, Vonrhein C, Bricogne G, Power S, Day AG. A Novel Combination of Two Classic Catalytic Schemes. Journal of Molecular Biology. 2002;320(2):303-9; doi: https://doi.org/10.1016/S0022-2836(02)00387-X.
49. KOSHLAND Jr. DE. STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS. Biological Reviews. 1953;28(4):416-36; doi: https://doi.org/10.1111/j.1469-185X.1953.tb01386.x.
50. Chang CJ, Lee CC, Chan YT, Trudeau DL, Wu MH, Tsai CH, et al. Exploring the Mechanism Responsible for Cellulase Thermostability by Structure-Guided Recombination. PLoS One. 2016;11(3):e0147485; doi: 10.1371/journal.pone.0147485.

51. Garg R, Srivastava R, Brahma V, Verma L, Karthikeyan S, Sahni G. Biochemical and structural characterization of a novel halotolerant cellulase from soil metagenome. Scientific Reports. 2016;6(1):39634; doi: 10.1038/srep39634.
52. Zhu Y, Han L, Hefferon KL, Silvaggi NR, Wilson DB, McBride MJ. Periplasmic Cytophaga hutchinsonii Endoglucanases Are Required for Use of Crystalline Cellulose as the Sole Source of Carbon and Energy. Appl Environ Microbiol. 2016;82(15):4835-45; doi: doi:10.1128/AEM.01298-16.
53. Wood TM. Preparation of crystalline, amorphous, and dyed cellulase substrates. In: Methods in Enzymology. Academic Press; 1988. p. 19-25.
54. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46(W1):W95-W101; doi: 10.1093/nar/gky418.
55. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnology. 2019;37(4):420-3; doi: 10.1038/s41587-019-0036-z.
56. Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. The Proteomics Protocols Handbook. Totowa, NJ: Humana Press; 2005. p. 571-607.
57. Miller GL. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Analytical Chemistry. 1959;31(3):426-8; doi: 10.1021/ac60147a030.
58. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. 2011;67(Pt 4):235-42; doi: 10.1107/S0907444910045749.
59. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr. 1997;53(Pt 3):240-55; doi: 10.1107/s0907444996012255.
60. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallographica Section D. 2006;62(9):1002-11; doi: doi:10.1107/S0907444906022116.
61. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004;60(Pt 12 Pt 1):2126-32; doi: 10.1107/s0907444904019158.

Figures

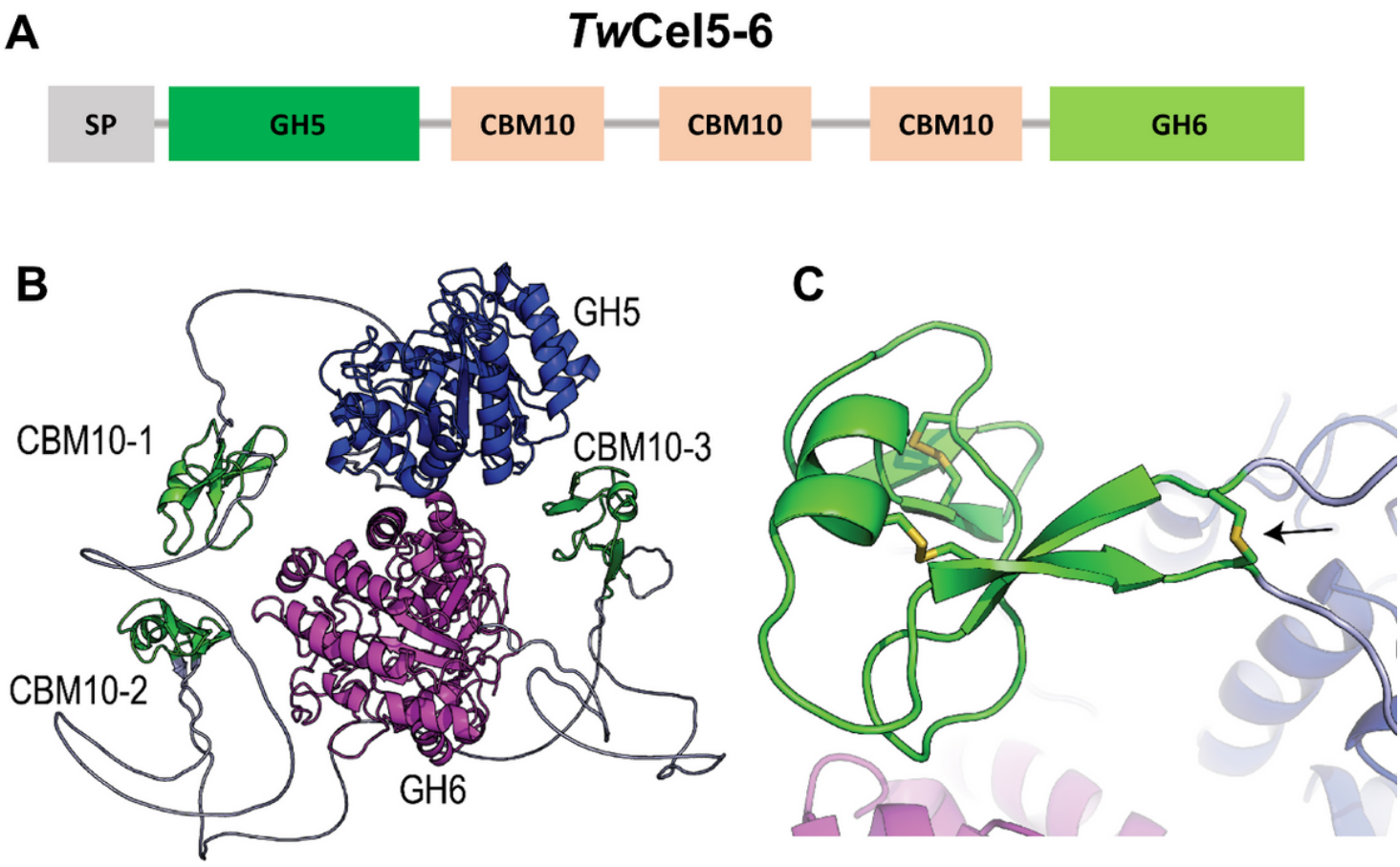


Figure 1

Multi-modular architecture of the *Tw*Cel5-6. (A) Display of the full-length gene encoding protein *Tw*Cel5-6 is composed of a signal peptide (SP), a glycosyl hydrolase 5 (GH5), three family 10 cellulose-binding modules (CBMs), and a glycosyl hydrolase 6 (GH6) catalytic domains, respectively. (B) Alphafold2 predicted structure of *Tw*Cel5-6; the unstructured linkers, CBM10s and the GH5 and GH6 catalytic modules are colored grey, green, blue, and magenta, respectively. (C) Details of the first CBM10 module showing the disulphide bridge (black arrow) connecting the linkers that enter and exit the module.

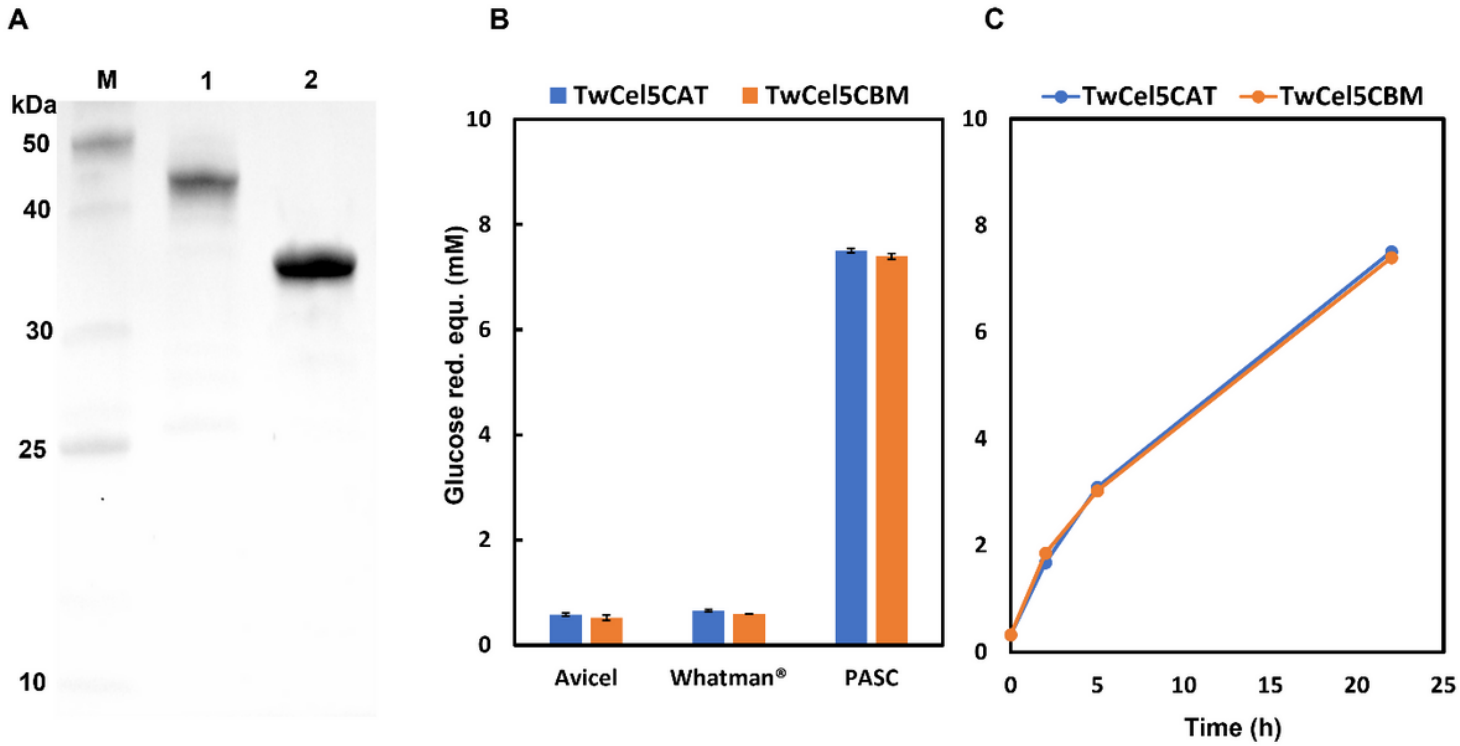


Figure 2

A) SDS-PAGE analysis of purified *Tw*Cel5. Lane 1, standard molecular weight markers (kDa); lane 2, *Tw*Cel5$_{CBM}$; lane 3, *Tw*Cel5$_{CAT}$. B). Screening of hydrolytic activity on three cellulosic 0.5 % (w/v) substrates. C). Comparative progress curves for the degradation of 0.5 % (/w/v) PASC. Reactions were performed in 50 mM sodium phosphate buffer (pH 7.5) containing 0.5 M NaCl and 1 µM enzyme incubated at 30 ºC. Hydrolytic activities were determined by measuring the release of reducing sugars using DNSA assay.

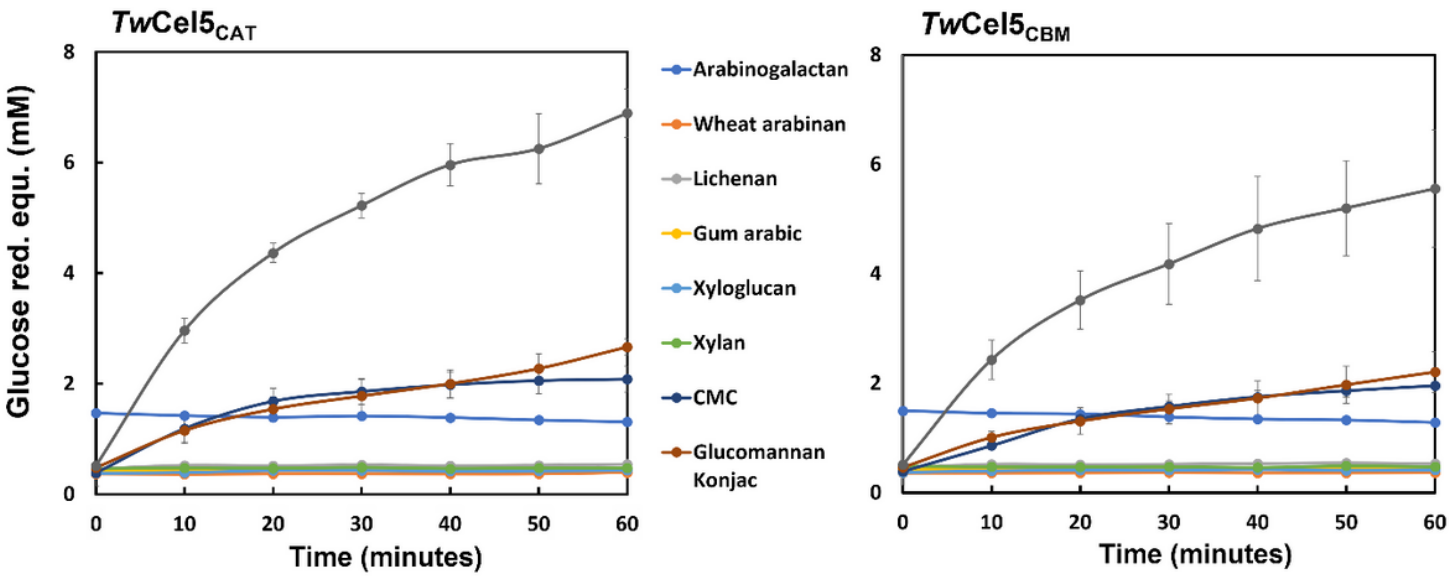


Figure 3

Hydrolysis of cellulosic and hemicellulose substrates by *Tw*Cel5$_{CAT}$ and *Tw*Cel5$_{CBM}$ as a function of time. Reactions were performed in 50 mM sodium phosphate buffer (pH 7.5) containing 0.5 M NaCl and 50 nM enzyme incubated at 30 ºC for 1 h in triplicates. Reducing sugar equivalents were quantified using glucose as a standard. The results correspond to mean and standard deviations of triplicates.

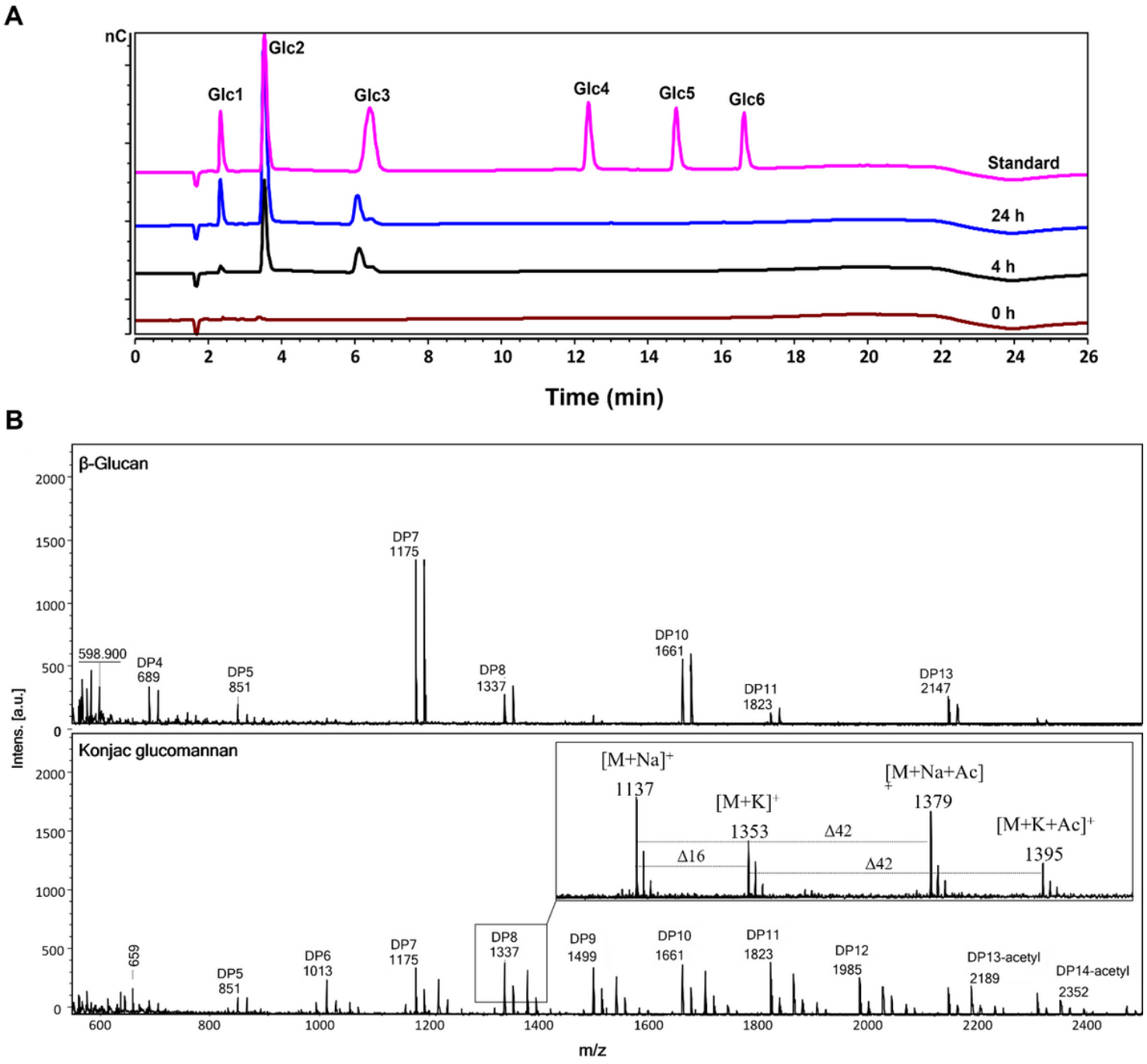


Figure 4

Analysis of hydrolytic products generated by *Tw*Cel5$_{CAT}$ on different polysaccharides. A). HPAEC-PAD chromatogram showing the soluble products generated upon 0.5 % (w/v) PASC incubation of 1 µM enzyme in 50 mM sodium phosphate buffer (pH 7.5) at 30 °C for 24 h. B) MALDI-TOF MS analysis of products generated from β-glucan (upper) and konjac glucomannan (lower) 50 nM enzyme incubation in 50 mM sodium phosphate, pH 7.5 and 0.5 M NaCl for 60 min at 30 °C. The lower panel shows the Na^+ and K^+ adducts of oligosaccharide, and 'DP' stands for degree of polymerization or 'Ac' for acetylation. None of the labelled peaks were observed in the negative control (i.e., a reaction without enzyme). Note

the *m/z* difference of 162 observed between the peaks suggest difference of a hexose residue. The products were identified based on cello-oligo standards, as shown.

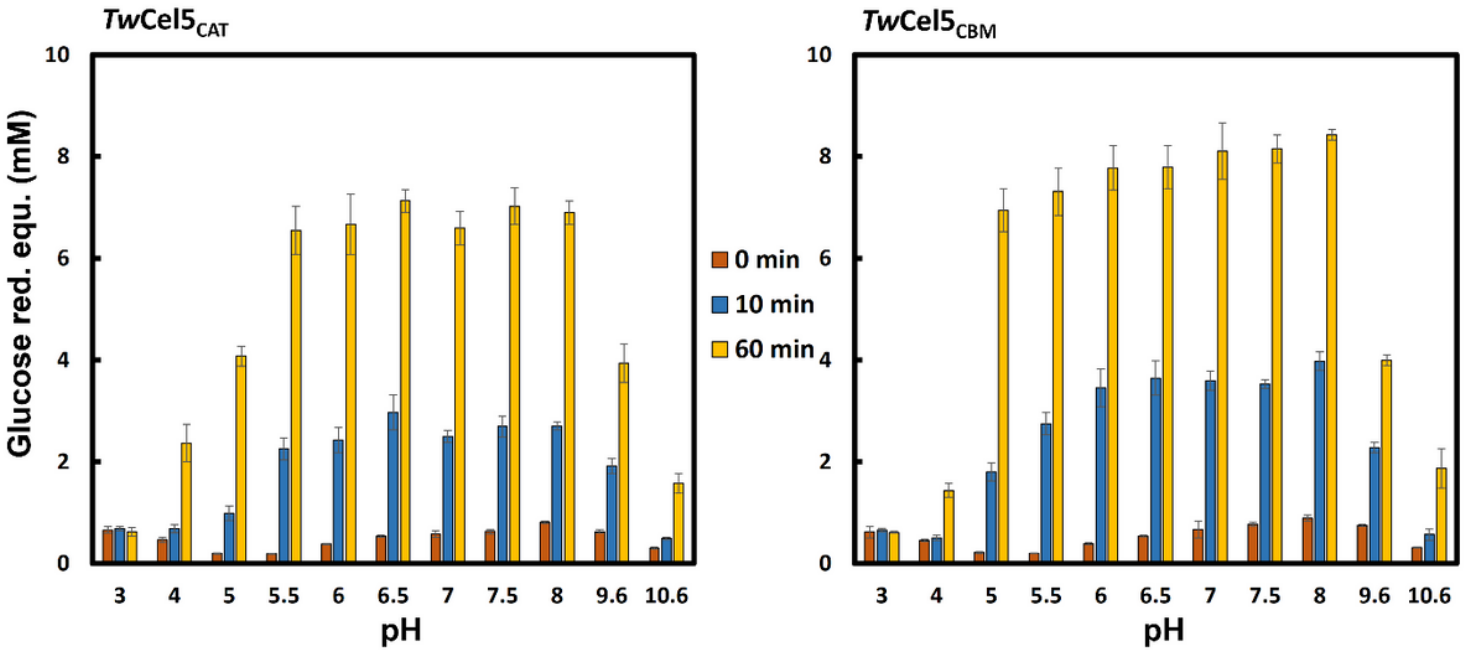


Figure 5

Influence of pH on the hydrolytic activity and stability of *Tw*Cel5 displaying the pH dependency of the hydrolytic activity performed in 50 mM different buffer systems using 0.5 % (w/v) β-glucan, 0.5 M NaCl and 50 nM enzyme variants, incubated at 30 ºC for 60 min. The reducing sugars values are mean and standard deviation derived from data obtained from triplicates.

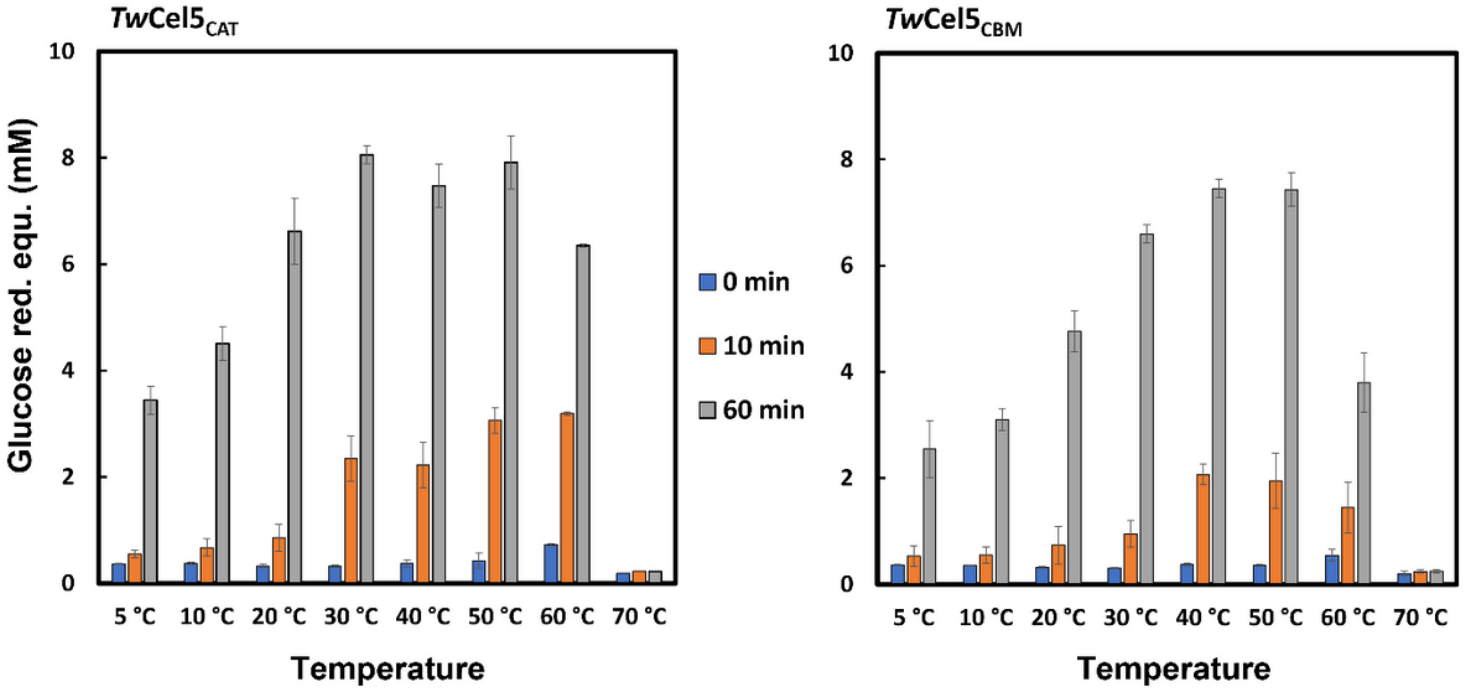


Figure 6

Influence of temperatures on the hydrolytic activity and stability of *Tw*Cel5 displaying the temperature dependency on the hydrolysis of β-glucan performed in 50 mM sodium phosphate buffer, pH 7.5, using 0.5 % (w/v) β-glucan, 0.5 M NaCl and 50 nM enzyme variants, incubated at different temperatures ranging from 5 ºC - 30 ºC for 60 min. The reducing sugars values are mean and standard deviation derived from triplicates.

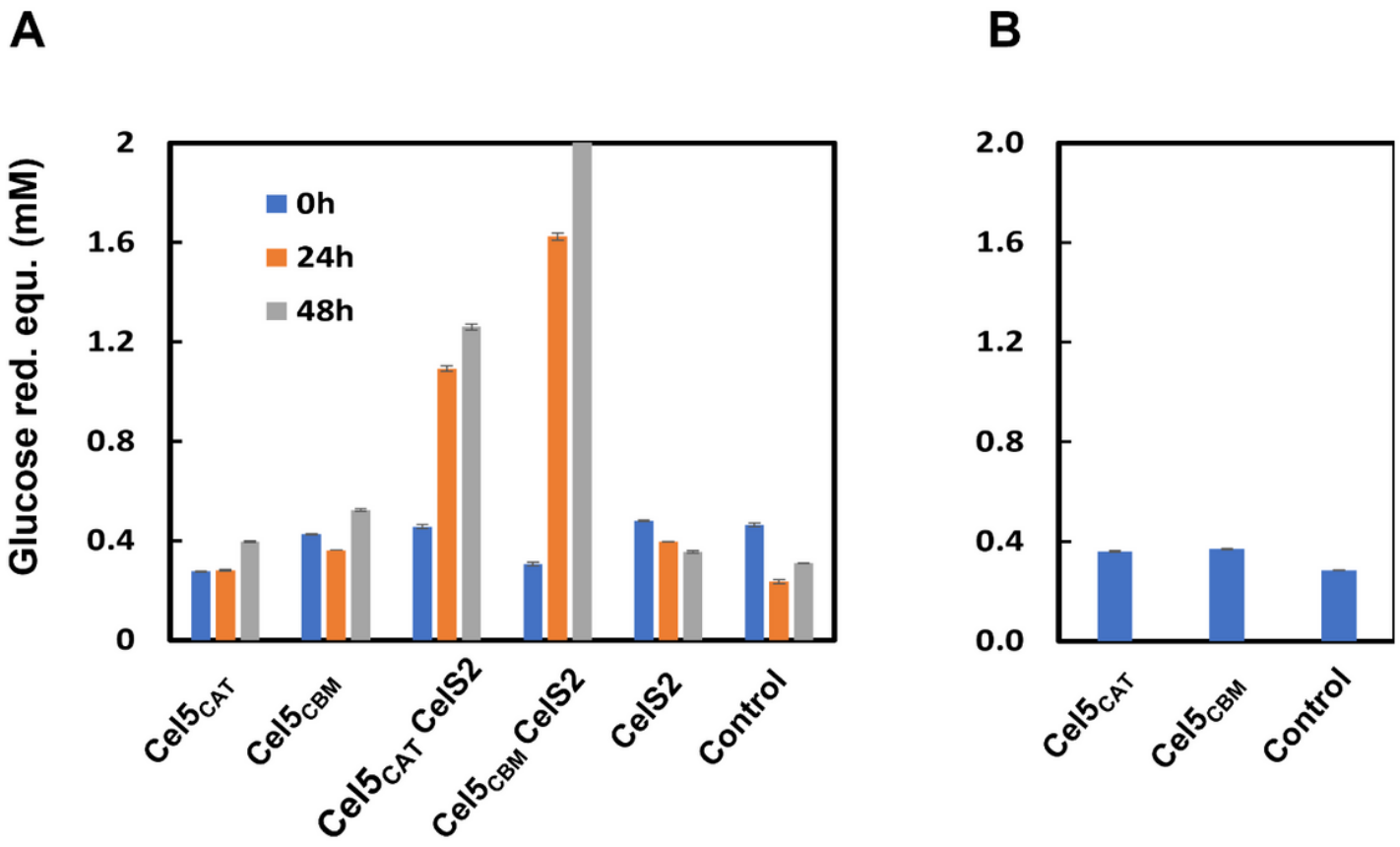


Figure 7

Synergy experiment between cellulase active LPMO and cellulase *Tw*Cel5 showing the boost in hydrolytic activity against Avicel. A) *Tw*Cel5$_{CAT}$, *Tw*Cel5$_{CBM}$ or CelS2 each 1 µM were incubated alone or in combination for different time durations, all with 1 mM ascorbate included in the reaction mixture. B) Reactions conducted from the oxidised products of avicel by CelS2 after 48 h treatment were incubated with the *Tw*Cel5 variants for 18 h; “control” refers to a reaction in which no enzyme was added. Values obtained are means and standard deviation derived from triplicate.

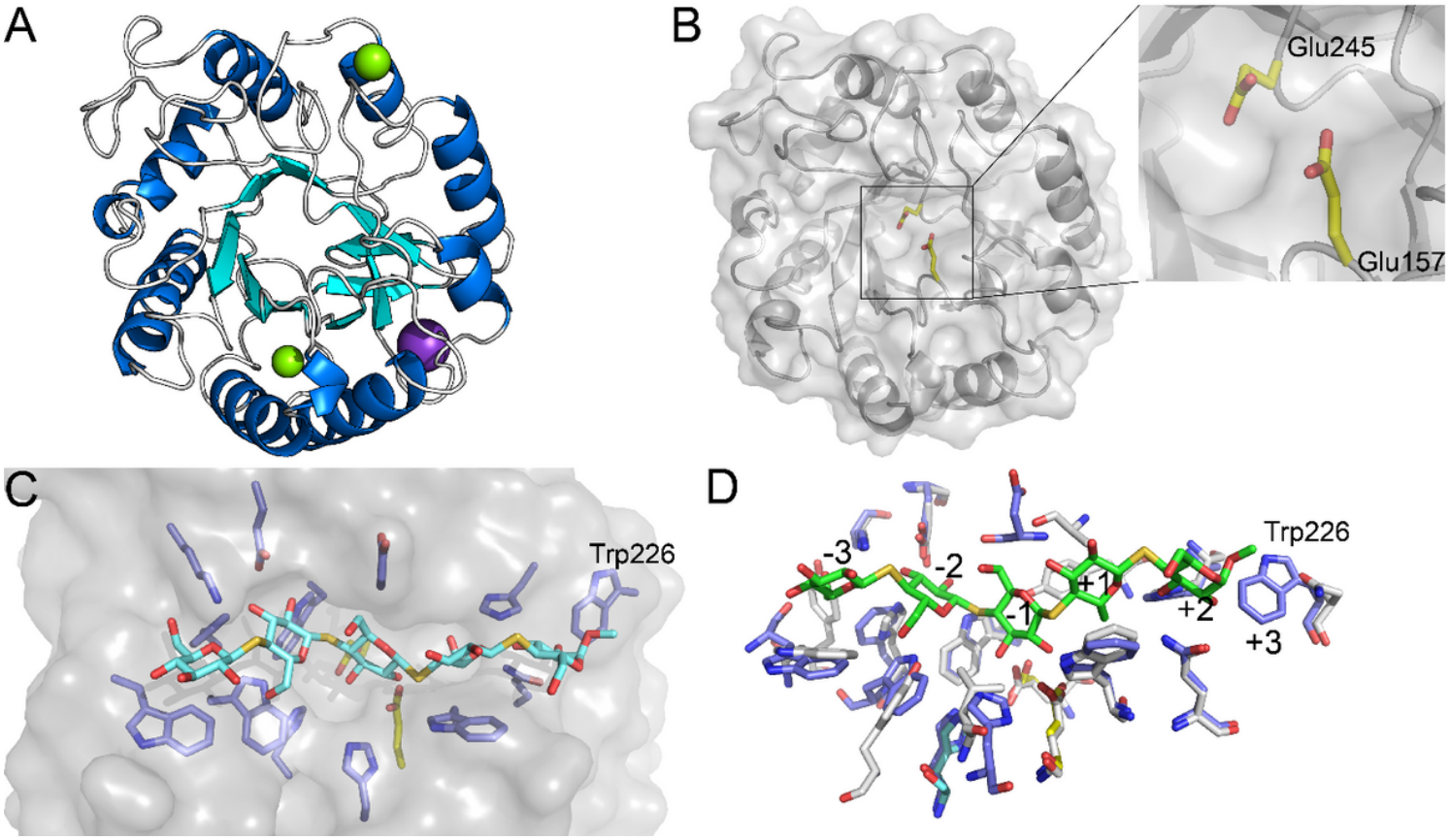


Figure 8

Crystal structure analysis of *Tw*Cel5$_{CAT}$. (A). Orthogonal view of the *Tw*Cel5$_{CAT}$ tertiary structure showing the characteristic $(\beta/a)_8$-barrel fold. Helices are blue, β-strands cyan, and loops white are shown. Mg^{++} and K^{+} ions are shown as green- and purple-colored spheres, respectively. (B) The putative active site of *Tw*Cel5$_{CAT}$, with the side chains of Glu157 and Glu245 shown with yellow carbons. (C) Illustration of substrate binding groove with the side chains of amino acids potentially interacting with substrate shown with blue carbons. For illustration purposes, a thiocellopentaose molecule has been placed in substrate binding site by structural superposition of *Tw*Cel5$_{CAT}$ structure with the structure of ligand containing Cel5A from *Bacillus agaradharens* (PDB identifier: 1H5V). (D) Structural superposition of subsite defining amino acids of *Tw*Cel5$_{CAT}$ and B. *agaradharens* Cel5A, highlighting the putative +3 subsite in *Tw*Cel5$_{CAT}$. *Tw*Cel5$_{CAT}$, blue carbons; Cel5A, grey carbons; substrate, green carbons.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.pdf