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.

Crystal structure of the GH30_12 endoxylanase AcXyn30B in complex with xylotriose reveals an extended xylan binding cleft

Panagiota Karampa
University of Patras
Christina Pentari
National Technical University of Athens
Evangelos Topakas
vtopakas@chemeng.ntua.gr

National Technical University of Athens
Maria Dimarogona
University of Patras

Research Article

Keywords: Acetivibrio clariflavus, AcXyn30B, GH30_12, non-specific xylanase, cellulosome, xylotriose

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11418445/v1

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

Additional Declarations: Competing interest reported. Prof Evangelos Topakas is an Editor-in-Chief of the journal Biotechnology for Biofuels and Bioproducts. The peer-review of this paper was therefore overseen by an independent member of the editorial board while the submission was subject to the exact same review process as applied to any other manuscript.

Crystal structure of the GH30_12 endoxylanase *Ac*Xyn30B in complex with xylotriose reveals an extended xylan binding cleft

Panagiota Karampa[1+], Christina Pentari[2+], Evangelos Topakas[2*], Maria Dimarogona[1*]

1. Laboratory of Structural Biology and Biotechnology, Department of Chemical Engineering, University of Patras, Patras, Greece

2. Industrial Biotechnology & Biocatalysis Group, Biotechnology Laboratory, School of Chemical Engineering, National Technical University of Athens, Athens, Greece

*Correspondence to: E. Topakas vtopakas@chemeng.ntua.gr; Iroon Polytechniou 9, Zografou Campus, 15780, Athens, Greece, Tel: +302107723264, Fax: +302107723163

M. Dimarogona mdimarog@chemeng.upatras.gr; Department of Chemical Engineering, University of Patras, University Campus, Patras 26504, Greece

+These authors contributed equally to the present work

Abstract

Background Glycoside hydrolase 30 (GH30) xylanases are key enzymes for the degradation of substituted xylan. *Ac*Xyn30B from *Acetivibrio clariflavus* is a modular enzyme, consisting of a GH30 catalytic domain that belongs to subfamily 12 (GH30_12), a carbohydrate binding module (CBM) and a dockerin module, active on glucuronoxylan, arabinoxylan and non-substituted xylooligosaccharides.

Results We examine the effect of the CBM and dockerin modules on enzyme activity. The first crystal structure of the catalytic domain of *Ac*Xyn30B is determined in apo- and xylotriose-bound state at resolutions of 1.14 and 1.55 Å, respectively. We provide a detailed

description of -4 to +4 xylose binding subsites and identify a calcium binding site, which does not seem to be directly implicated in enzyme function.

Conclusions The present study demonstrates for the first time the effect of the CBM and dockerin modules on the catalytic activity of *Ac*Xyn30B, shifting the preference from beechwood glucuronoxylan to rye arabinoxylan. In contrast to the other structurally characterized GH30_12 enzyme (PDB ID: 9YWN), *Ac*Xyn30B possesses a xylan-binding cleft that is open at both ends, enabling the accommodation of longer part of the hemicellulosic polymer. Furthermore, the experimentally determined structure revealed a *trans* peptide bond at a position where AlphaFold predicted a *cis* conformation, most likely because the prediction was biased by the *cis* bond present in the only available experimental structure of homologous enzyme (PDB ID: 9YWN), underscoring the need for experimental validation of computational models. Overall, this work provides new insight into the hydrolysis pattern of *Ac*Xyn30B as a member of the GH30_12 subfamily.

Keywords

*Acetivibrio clariflavus; Ac*Xyn30B*;* GH30_12; non-specific xylanase; cellulosome; xylotriose

1. Background

In the frame of a circular bio-based economy approach, significant effort has been dedicated to the exploitation of lignocellulosic waste for the production of valuable chemicals. The second most abundant polysaccharide found in lignocellulosic substrates is xylan, which is composed of β-1,4-xylose residues bearing various decorations, depending on the plant, tissue, and season. Glucuronoyl and arabinofuranosyl side-decorations of xylan may be associated with aromatic compounds of lignin, forming lignin-carbohydrate complexes (1–3). Moreover, xylan may adopt a 2-fold flat ribbon conformation (4), which allows binding to cellulose fibrils. The

sugar content of xylan holds high biotechnological potential, as its oligosaccharides can serve as prebiotics, while its monosaccharides can be used for fermentation to bioethanol or as precursors for the environmentally friendly production of value-added materials and chemicals (5–7). Due to the heterogeneous nature of xylan, many enzymes are involved in its dissociation from lignin and cellulose (8), as well as in its decomposition to monosaccharides and acids.

Glycoside hydrolase 30 (GH30, CAZy database (9)) xylanases in particular, distributed in subfamilies 7, 8, 10 and recently, 12, are of increased biotechnological interest due to their implication in the degradation of highly substituted xylans. Several xylanolytic activities have been identified within the GH30 family, including non-specific endoxylanases (10,11), arabinoxylanases (12), glucuronoxylanases (13) and xylobiohydrolases (14,15), many of them combining endo- and exo-specificities. The identification of catalysts that tolerate xylan substituents may hold a higher biotechnological advantage for the efficient production of oligomers from complex lignocellulosic structures (16). Elucidating the structural determinants that contribute to the distinct specificities of xylan-targeting enzymes (17–20) and understanding the nature of the substrate-enzyme interactions that eventually lead to catalysis may enable protein engineering to generate tailor-made biocatalysts suitable for any type of carbohydrate modification- or degradation-related application.

Acetivibrio clariflavus is an anaerobic thermophilic lignocellulose degrading bacterium, and most of its strains have been isolated from anaerobic sludges of methanogenic reactors (21). *A. clariflavus* DSM19732 in particular has been shown to exhibit higher xylanolytic activity compared to *A. thermocellus* (22), which could be attributed to the abundant occurrence of GH30 enzymes in all *A. clariflavus* cellulosome fractions, demonstrated by proteomic analysis (23). The two GH30 xylanase genes, Clocl_1795 and Clocl_2746, present in *A. clariflavus* genome, were subsequently cloned and characterized. The former, named *Ac*Xbh30A, was shown to be a xylobiohydrolase (24,25) and the latter, named *Ac*Xyn30B, was characterized as

a non-specific endoxylanase (11). *Ac*Xyn30B is active on glucuronoxylan, arabinoxylan and aryl glycosides, contrary to GH30_8 xylanases that are specific to either glucuronoxylan or arabinoxylan, and it is the first characterized member of the GH30_12 subfamily. It is a modular enzyme, comprising a carbohydrate binding module 6 (CBM6) and a dockerin (DK) domain sequentially appended on the C-terminal of the GH30 catalytic domain (CD). Currently, there is no crystal structure of *Ac*Xyn30B. Recently, the structure of full-length *Ac*Xyn30B was modelled using computational analysis and SAXS (26). Molecular docking with xylooligosaccharides (XOs) revealed preferable binding for X3, followed by X2. The only GH30_12 family member whose structure has been crystallographically resolved is that of the catalytic domain of *Ach*Xyn30A from *Anaerobacterium chartisolvens* (27). Similarly to all GH30 members, the CD of *Ach*Xyn30A consists of a $(\beta/\alpha)_8$ barrel and a side-β-domain, while the xylan binding cleft is more open compared to other GH30s, in agreement with the promiscuous GH30_12 activity. A sodium ion was found at the interface of the side- and $(\beta/\alpha)_8$-barrel domains, and an Ala-*cis*-Ala peptide, located near the catalytic amino acids, with potential implication in the function of GH30_12 enzymes.

Here we examine the effect of CBM6 and DK domains on enzyme function, by characterizing and comparing full length *Ac*Xyn30B, and its truncated variants, namely *Ac*Xyn30B_CD+CBM, *Ac*Xyn30B_CD. Additionally, we report the first crystallographic structure of a GH30_12 subfamily member in ligand-bound form, providing detailed analysis of the xylose binding subsites.

2. Materials and Methods

2.1 Substrates and chemicals

Acetylated xylan, beechwood xylan, rye arabinoxylan, insoluble wheat arabinoxylan (IWA), and XOs with a degree of polymerization (DP) 2-6 (X2-X6) were purchased from Megazyme

(Bray, Co. Wicklow, Ireland). Other chemicals were from Sigma-Aldrich and Applichem and were of the highest purity available. Enzymatic activities were also examined on pretreated lignocellulosic materials. Beechwood (BW) was delignified via Milox pulping (28) using acetic acid at 80 ◦C for 3 h (composition: 52 % w/w cellulose, 29 % w/w xylan, 5 % w/w acid insoluble lignin, 3 % w/w acid soluble lignin). Corn bran (CB) was enzymatically de-starched and autoclaved (29) in the presence of 3 % v/v acetic acid under 20 bars at 100 ◦C for 1 h (composition: 22 % w/w glucose, 20 % w/w xylose, 9 % w/w arabinose, 3.5 % w/w galactose, 6.5 % w/w acid-insoluble lignin, 13.5 % w/w acetic acid). Wheat bran (WB) was pretreated via the OxiOrganosolv (30) method, using ethanol at 130 ◦C under 16 bars for 4 h (composition: 30.4 % w/w cellulose, 10 % w/w lignin, 17.4 % w/w arabinose, 35.5 % w/w xylose, 1.7 % w/w galactose). A mixture of acetylated xylo-oligosaccharides was prepared for the soaking experiments, by digestion of acetylated xylan (0.5 % w/v) with a GH10 xylanase (E-XYLATM, Megazyme, 20 $\mu g \cdot mL^{-1}$) in 20 mM sodium acetate-acetic acid buffer, pH 5, at 45 °C and 950 rpm for 24 h, in 5 mL reaction volumes. The endoxylanase was removed by ultrafiltration through a 10kDa Amicon spin filter, and the filtrate was dried by lyophilization.

2.2 Cloning, expression and purification of *Ac*Xyn30B and its variants

The non-specific GH30_12 endoxylanase gene originates from *Acetivibrio clariflavus* DSM 19732 (chromosome 3188531-3190552; GenBank: CP003065.1, Uniprot: G8M2Z1). The gene encodes a protein which is part of a cellulosome, composed of an N-terminal catalytic domain, a CBM6 domain and a C-terminal DK domain interconnected via linkers. In order to determine the crystal structure of *Ac*Xyn30B, three different genes were designed in the pET-26b(+) vector system for heterologous expression in *Escherichia coli* and synthesized by GenScript Biotech (USA). The first gene contained only the catalytic domain (*Ac*Xyn30B_CD, residues Ala28 to Asp446), which was intended for structural studies due to the limited mobility of the

molecule, the second contained the CD and the CBM domains (*Ac*Xyn30B_CD+CBM, residues Ala28 to Gly595), while the third represented the full-length enzyme with the CD, the CBM and the DK domains (*Ac*Xyn30B, Ala28 to Glu673). For the plasmid design, the putative signal peptide was removed from all three genes, which were inserted into the vector, in line with the ClaI-SalI restriction sites. The physicochemical properties of the putative enzymes were predicted with the Expasy-Protparam (31) online tool.

Gene constructs were transformed into chemically competent *E. coli* One Shot Top10 cells (Invitrogen) and cultivated overnight at 37 °C and 180 rpm in 5 mL low-salt Luria-Bertani (LB) medium, supplemented with 25 $\mu g \cdot mL^{-1}$ kanamycin. The multiplied genes were extracted using the plasmid extraction mini prep kit (Macherey-Nagel, Germany). The extracted genes were then transformed into chemically competent *E. coli* BL21 cells. Precultures of 5 mL LB medium were incubated at 37 °C and 180 rpm for 16 h and were then inoculated in 500 mL LB containing 100 $\mu g \cdot mL^{-1}$ kanamycin. After 3-4 h of incubation, when the cell density reached an OD of 1.3-1.5 at 600 nm, protein expression was induced by 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Following a 16-h incubation at 37 °C and 180 rpm, cells were collected by centrifugation for 10 min at 18038 × g and were resuspended in 50 mM Tris base and 300 mM NaCl buffer, pH 8. Cell lysis took place by sonication (40 % amplitude, 8 sec ON & 5 sec OFF for 90 sec), and debris was removed by centrifugation for 10 min at 18038 × g. The crude protein solution was filtered through 0.45 μm PES filters, and protein purification was achieved via cobalt ion metal affinity chromatography. The purified proteins were dialyzed in 20 mM Tris-HCl buffer of pH 8. The homogeneity of the protein products was verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (32). The *Ac*Xyn30B_CD protein intended for crystallization studies was further filtered through 0.45 μm syringe filters and concentrated to 10 $mg \cdot mL^{-1}$.

2.3 Biochemical characterization of *Ac*Xyn30B and its variants

The catalytic activities of the purified proteins were compared against 100 µM xylopentaose (X5) and 0.25 % w/v polymeric substrates (acetylated xylan, BW xylan, rye arabinoxylan, and IWA). In a 400 µL reaction mixture, 0.2 µM of each protein was used, except for IWA, for which 0.4 µM enzyme was added. Reactions were incubated in 50 mM NaH_2PO_4/Na_2HPO_4 buffer, pH 6, at 40 °C and 950 rpm for 15 min when using X5 or 1 h for polymeric substrates (except IWA for which a 24 h incubation was performed). Enzymatic and control reactions (without enzyme addition) were performed in duplicate. Reactions were terminated by 10-min boiling and the hydrolysates were analyzed by High-Performance Anion Exchange Chromatography (HPAEC). The system was equipped with an ED40 pulsed amperometric detector (PAD) (Dionex, Sunnyvale, CA, USA), a gold electrode, a CarboPac PA- 1 (4 × 250 mm) column with a CarboPac PA1 IC (4 × 50 mm) standard bore guard column, while the mobile phase involved three eluents of 10 mM NaOH, 100 mM NaOH, and 100 mM NaOH with 500 mM sodium acetate, as described elsewhere (33). The reducing sugars released from polymeric substrates were quantified by the DNS assay (34), using a Spectra Max 250 microplate reader (Molecular Devices, CA, USA) at 540 nm. The amount of released reducing sugars was quantified as xylose equivalents, based on standard xylose concentrations. One Unit (U) of activity is defined as the amount of enzyme releasing 1 µmol of reducing sugars expressed as xylose equivalent per min of reaction. An alternative activity assay was also performed under the conditions indicated in the work of Choudhury et al. (2025) (16), in order to obtain comparable results (4-min assay at 70 °C using 1 % w/v BW xylan,rye arabinoxylan or acetylated xylan and 5 $\mu g \cdot mL^{-1}$ protein in 100 µL 50 mM citrate-phosphate buffer, pH 5.5). The activity of the three proteins was also investigated on pretreated lignocellulosic substrates (BW, CB, WB). Each substrate (50 $mg \cdot mL^{-1}$) was incubated with 1 µM of each protein in 400 µL reaction mixtures of 50 mM NaH_2PO_4/Na_2HPO_4 buffer pH 6, at 40 °C and 950 rpm for 6

176 h. Reactions were performed in duplicate, along with the corresponding control reactions, and
177 terminated by boiling for 10 min. Hydrolysis products were analyzed by HPAEC.
178 The catalytic activity of *Ac*Xyn30B_CD was examined against XOs (X4-X6), aiming to
179 identify the hydrolysis pattern of the enzyme on different chain lengths. Each substrate (50
180 μM) was incubated with 0.25 μg·mL^{-1} protein in 50 mM NaH_2PO_4/Na_2HPO_4, pH 6, in reaction
181 volumes of 200 μL at 40 °C and 950 rpm. Duplicate samples were collected at different time
182 intervals between 5-30 min, inactivated via 10-min boiling, and analyzed by HPAEC.
183 Finally, the activity of *Ac*Xyn30B_CD was measured in the presence of Ca^{2+} and EDTA. More
184 specifically, 30 μg·mL^{-1} protein were incubated with 0.25 % w/v rye arabinoxylan in the
185 presence of $CaCl_2$ solution or EDTA solvent in varying concentrations (0.5-5 mM). Reactions
186 were performed in triplicate, in 100 μL of 50 mM NaH_2PO_4/Na_2HPO_4, pH 6, at 40 °C and 950
187 rpm for 60 min, while appropriate control reactions were also prepared. Afterwards, reactions
188 were terminated through boiling for 10 min and mixtures were analyzed using HPAEC.
189
190 **2.4 Investigation of protein polymerization and CBM binding affinity**
191 Size exclusion chromatography (SEC) was performed on Agilent 1260 Infinity II LC system
192 (Agilent, USA) equipped with a 1260 Infinity III Variable Wavelength Detector (Agilent,
193 USA), using a Zorbax GF-250 column (9.4 mm ID x 25 cm), 130 mM NaCl – 20 mM Na_2HPO_4
194 buffer, pH 7, as mobile phase at 1 mL/min flow rate and 23 °C, and a 100 μL injection volume,
195 while monitoring absorption at 210 and 280 nm. Protein samples were analyzed at a
196 concentration of 0.5 m g·mL^{-1}, while a standard curve was obtained using a kit for molecular
197 weights (12-200 kDa; Sigma-Aldrich).

198 Non-denaturing affinity gel electrophoresis (19) was performed to investigate the mobility of
199 *Ac*Xyn30B and its truncates in their native form, in the presence or absence of rye arabinoxylan.
200 Gels were prepared using 6 % acrylamide in 60 mM Tris - 40 mM CAPS buffer, pH 9.4, with

and without 0.5 % w/v polysaccharide, and electrophoresis was performed concurrently for both gels at 100 V with a Mini-PROTEAN System (Bio-Rad), which was placed in a cold bath (~10 °C) to avoid protein denaturation. Gels were visualized by staining with Coomassie brilliant blue G-250.

The identification of key amino acids that are suggested to attribute carbohydrate affinity to the CBM6 of *Ac*Xyn30B (*Ac*CBM6) was performed via sequence alignment with structurally characterized CBM6 domains (PDB entries: 1W9S, 1UY1, 1O8P, 1GMM, 1UXZ, 2CDO) using the Clustal Omega (35) multiple sequence alignment tool and visualized with ESPript 3.0 (36). UCSF Chimera (V:1.17.1) (37) was used for the visualization of the AlphaFold (38) predicted structure of *Ac*CBM6 (AF-G8M2Z1).

The binding of CBM6 on insoluble substrates (19) (0.5 % w/v avicel or 50 $mg \cdot mL^{-1}$ pretreated BW) was investigated by incubating 200 $\mu g \cdot mL^{-1}$ protein (*Ac*Xyn30B_CD or *Ac*Xyn30B_CD+CBM) with the substrate in a 50 mM NaH_2PO_4/Na_2HPO_4 buffer, pH 6, at 4 °C for 16 h, under mild agitation. After incubation, samples were centrifuged to collect the supernatant. The precipitate was washed once with water and once with SDS sample buffer, following centrifugation and collection of the supernatant after each step. All incubation and wash supernatants were analyzed by SDS-PAGE electrophoresis (Section 2.2).

2.5 Crystallization – Structure determination

*Ac*Xyn30B_CD was submitted to high-throughput crystallization screening with the sitting drop vapor diffusion technique using an OryxNano crystallization robot (Douglas Instrument Ltd, East Garston, Berkshire, UK) and commercially available 96-well kits of Molecular Dimensions (Morpheus HT and JCSG plus). Equal volumes (0.2 μL) of enzyme and reservoir solution were mixed in each well and crystals appeared in various PEG-based conditions.

Crystals of *Ac*Xyn30B_CD grew within 1 week in the presence of 0.1 M carboxylic acids (sodium formate, ammonium acetate, sodium citrate tribasic dihydrate, potassium sodium tartrate tetrahydrate and sodium oxamate), 0.1 M buffer solution (Tris base and BICINE) pH 8.5, 12.5 % v/v MPD, 12.5 % w/v PEG 1000 and 12.5 % w/v PEG 3350. Crystals were flash cooled in liquid nitrogen, and a complete dataset was collected to 1.14 Å at beamline BioMAX (39), operated by MAX IV Laboratory (Lund, Sweden). *Ac*Xyn30B_CD crystals appearing in 0.1 M amino acids (DL-glutamic acid monohydrate, DL-alanine, glycine, DL-lysine monohydrochloride and DL-serine), 0.1 M buffer solution (Tris base and BICINE) pH 8.5, 20 % v/v PEG 500 MME and 20 % w/v PEG 20000 were used to create a seed stock for microseed matrix seeding (MMS) (40) in the JCSG plus crystallization screen. Each drop consisted of 0.2 µL protein, 0.2 µL reservoir solution and 0.1 µL seed stock. Cubic crystals appeared after 4 days in the presence of 0.2 M sodium malonate, 20 % w/v PEG 3350. The complex structure of *Ac*Xyn30B_CD-X3 was obtained by soaking *Ac*Xyn30B_CD crystals from the sodium malonate condition in 5 mM X5 dissolved in mother liquor supplemented with 20 % v/v glycerol, for 30 minutes. Crystals were then flash cooled in liquid nitrogen and X-ray data for the *Ac*Xyn30B_CD-X3 were collected to 1.55 Å resolution at BioMAX beamline (39). In parallel, crystals from the same crystallization condition were soaked in mother liquor supplemented with 30 mM 2^3-(4-*O*-Methyl-α-D-glucuronosyl)-xylotetraose (XUXX) and 20 % v/v glycerol for 30 min, resulting in a structure diffracting to 1.79 Å resolution with a poorly defined electron density in *Ac*Xyn30B binding pocket.

X-ray datasets were integrated with XDS (41), while scaling was performed with AIMLESS from the CCP4i2 suite (42). An *in silico* AlphaFold model (38) was generated as an initial structure for molecular replacement performed by PHASER (43). Iterative cycles of manual model building and refinement of the structure were carried out using COOT (44) and REFMAC5 (45), respectively. The same procedure was followed for the determination of the

250 *Ac*Xyn30B_CD-X3 complex structure. The apo crystal structure of *Ac*Xyn30B was used as the
251 initial model for molecular replacement and the carbohydrate moiety of the structure was
252 validated using PRIVATEER (46).
253 X-ray data collection and refinement statistics are presented in Table 1. The structures have
254 been deposited to the PDB with accession codes 9QY5, 9SLE and 34AO, respectively.
255
256 Table 1. Diffraction data and refinement statistics for *Ac*Xyn30B_CD and *Ac*Xyn30B_CD-X3
257 structures. Values in parentheses are for the outer shell.

	*Ac*Xyn30B_CD	*Ac*Xyn30B_CD-X3	*Ac*Xyn30B_CD soaked with XUXX
PDB ID	9QY5	9SLE	34AO
Data collection			
Diffraction source	BioMAX, MAX IV		
Wavelength (Å)	0.7293		
Temperature (K)	100		
Detector	Dectris Eiger2 16M Cd Te area detector		
Crystal-detector distance (mm)	165.2	331.1	308.6
Rotation range per image (°)	0.1	0.1	0.1
Total rotation range (°)	360	360	360
Exposure time per image (s)	0.01	0.01	0.01

Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a, *b*, *c* (Å)	64.60, 72.95, 90.92	64.05, 72.67, 90.92	64.44, 74.48, 90.92
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Mosaicity (°)	0.10	0.17	0.15
Resolution range (Å)	48.36-1.14 (1.16-1.14)	48.14-1.55 (1.58-1.55)	48.73-1.79 (1.83-1.79)
Total No. of reflections	2097552 (103082)	775997 (24270)	564877 (33168)
No. of unique reflections	156623 (7658)	61724 (2748)	41896 (2424)
Completeness (%)	100.0 (99.9)	99.1 (91.4)	100.0 (99.6)
Half-set correlation CC (1/2)	0.997 (0.934)	0.999 (0.884)	0.999 (0.920)
Redundancy	13.4 (13.5)	12.6 (8.8)	13.5 (13.7)
$\langle I/\sigma(I)\rangle$	15.1 (4.0)	21.6 (2.9)	16.9 (3.3)
R_{merge}	0.091 (0.555)	0.063 (0.650)	0.097 (0.732)
$R_{p.i.m.}$	0.037 (0.226)	0.026 (0.325)	0.039 (0.212)
Overall *B* factor from Wilson plot	11.60	16.31	20.22
	Refinement statistics		
Resolution range (Å)	48.41-1.14	48.18-1.55	48.78-1.79
No. of reflections	156457	61612	41829
working set	148688	58525	39701

free set	7769	3087	2128
Final R_{cryst}	0.154	0.141	0.141
Final R_{free}	0.169	0.167	0.186
No. of non-H atoms			
Protein	3543	3513	3477
Ions	1	1	1
Ligands	8	101	38
Solvent molecules	540	454	400
Total	4092	4069	3916
R.m.s. deviations			
Bonds (Å)	0.013	0.011	0.009
Angles (°)	1.968	1.855	1.723
Average *B* factors (Å^2)			
Protein	15.43	19.79	23.81
Ions	13.09	16.56	19.83
Ligands	33.01	37.06	51.64
Solvent molecules	27.66	33.61	36.11
Ramachandran plot			
Most favoured (%)	96.14	96.14	96.38

Outlies (%)	0.48	0.49	0.48

3. Results

3.1 Heterologous expression of *Ac*Xyn30B and its variants and investigation of multimer formation

The full-length *Ac*Xyn30B protein expressed in this study (Supplementary table 1) consists of 677 amino acids with a predicted molecular weight of 75.9 kDa, a theoretical pI of 4.78 and an extinction coefficient of 124125 M^{-1} cm^{-1}. The corresponding parameters for the 599 amino acid-long *Ac*Xyn30B_CD+CBM protein were predicted to be 67.3 kDa, pI 4.89 and 115185 M^{-1} cm^{-1}, while for *Ac*Xyn30B_CD, consisting of 450 amino acids, the predicted values were 51.1 kDa, pI 5.44 and 90650 M^{-1} cm^{-1}. The molecular weight of the proteins was verified by SDS-PAGE, demonstrating values of approximately 75 kDa for *Ac*Xyn30B, 65 kDa for *Ac*Xyn30B_CD+CBM and 50 kDa for *Ac*Xyn30B_CD (Supplementary figure 1A).

Based on the SEC analysis, full-length *Ac*Xyn30B elutes at 11.1 min, corresponding to an 180 kDa complex, implying the multimerization of the protein. Similarly, *Ac*Xyn30B_CD+CBM demonstrates 2 main peaks, corresponding to 150 and 250 kDa, while analysis of the *Ac*Xyn30B_CD exhibits a complex of approximately 180 kDa. Non-denaturing gel analysis reveals distinct migration patterns among the proteins, which could be justified by their distinct sizes and net charges (Supplementary figure 1B; no polysaccharide containing gel). The full-length enzyme and *Ac*Xyn30B_CD demonstrate two distinct bands with different migration distances, while *Ac*Xyn30B_CD+CBM appears as one smeared band. The presence of distinct bands in the *Ac*Xyn30B_CD and *Ac*Xyn30B samples implies the formation of multimeric structures. The dimerization of *Ac*Xyn30B has also been suggested to take place at higher concentrations (e.g., 5 $mg \cdot mL^{-1}$), as indicated by small angle X-ray scattering (26), while conserving its catalytic activity.

3.2 Enzymatic activity against oligomers, polymers and pretreated lignocellulosic substrates

The truncated enzymes exhibit altered activities against oligomeric and polymeric substrates

The full-length xylanase, the CBM-containing enzyme and the CD of the protein were examined against a variety of oligosaccharides and polymeric substrates, in order to investigate the effect of the CBM and the DK domains on the catalytic activity. The substrates used for this cause demonstrated different structural features. In detail, X5 is a linear oligomer, acetylated xylan is a mixture of acetylated linear XOs and polymeric chains (24 % degree of acetylation), BW xylan is polymeric glucuronoxylan (12.8 % glucuronic acid content), rye arabinoxylan has a 40:60 arabinose to xylose ratio, while IWA (34:65 arabinose to xylose ratio) contains in addition to arabinoxylan, cellulose, arabinogalactan and glucomannan traces.

Regarding the enzymatic activities, two different assays were applied for each variant (Table 2; 1-h and 4-min assays), in order to obtain comparable results with previous publications (11,16). Focusing on the 1-h assay, *Ac*Xyn30B_CD demonstrated the highest activity on BW xylan, followed by rye arabinoxylan, acetylated xylan and IWA. *Ac*Xyn30B_CD+CBM generally demonstrated lower activities than *Ac*Xyn30B_CD, with the exception of rye arabinoxylan, on which both enzymes had similar activities. Similar patterns for the CBM carrying enzyme were also reported by Suchova et al. (11), where commercial *Ac*Xyn30B enzyme was used (Product number CZ0917 from NZYTech). Full length *Ac*Xyn30B demonstrated the poorest activity on all substrates compared to the other two variants, except for rye arabinoxylan, for which the highest activity was detected, almost 3 times that of the truncated enzymes. Summarizing the above, the two domains appended to the GH30_12 CD seem to shift the preference from beechwood glucuronoxylan to rye arabinoxylan, based on the 1-h assay conditions (11,16).

On the contrary, in the work of Choudhury et al. (2025) (16), where a 4-min assay was used, the highest activity of *Ac*Xyn30B_12 (full-length enzyme) was detected on acetylated xylan, followed by glucuronoxylan and arabinoxylan. For comparison, the activity of *Ac*Xyn30B was also measured using the same 4-min assay. The results agreed with those of the 1-h assay in that the full-length enzyme exhibited higher activity on rye arabinoxylan than on BW glucuronoxylan (Table 2; 4-min assay). In contrast, the omission of CBM and DK domains resulted in higher activity than the full-length enzyme on both xylans, more pronounced on glucuronoxylan. Activity on acetylated xylan could not be measured by this assay as the DNS signal of the hydrolysates was similar to the blank reaction. In summary, the higher activity on arabinosylated or glucuronylated xylans is in accordance with previous observations (11,16), that such sugar decorations could improve substrate recognition and enzyme specificity of *Ac*Xyn30B.

Table 2. Enzymatic activity of *Ac*Xyn30B variants (U/mg protein) on various polymeric substrates. Errors represent the standard deviations among replicates.

	*Ac*Xyn30B_CD	*Ac*Xyn30B_CD+ CBM	*Ac*Xyn30B
		1-h assay	
Beechwood glucuronoxylan	16.4±1.8	8.7±0.7	1.9±0.2
Rye arabinoxylan	7.3±1.2	6.8±1.2	18.6±0.5
Acetylated xylan	6.6±0.5	3.8±1.3	0.0±0.0
Insoluble wheat arabinoxylan[1]	0.24±0.00	0.14±0.01	0.11±0.00

	4-min assay		
Beechwood glucuronoxylan	103.0±3.7	27.7±2.1	10.4±1.0
Rye arabinoxylan	69.3±6.6	57.9±0.8	15.0±2.5

[1]A 24-h incubation was performed for this substrate, since no activity was detected within 1 h.

Focusing on the product profile on the various substrates (hydrolysates obtained from the 1-h assay), several differences can be spotted among the three variants. When tested against X5, all three enzymes released X1, X2 and X3 as main hydrolysis products (Figure 1). Only *Ac*Xyn30B_CD completely hydrolyzed X5 and X4 secondary product, while a minor amount of X5 (approximately 10 %) remained intact only in the hydrolysate of full-length *Ac*Xyn30B. All three enzymes, however, preferentially hydrolyze X5 towards X2 and X3 rather than X1 and X4, as can be observed by the final products X2-to-X1 ratio (approximately 1.6 for *Ac*Xyn30_CD, 2.3 for *Ac*Xyn30_CD+CBM and 2.8 for *Ac*Xyn30). This preference towards the release of X2 and X3 was also reported by Suchova et al. (11), suggesting the occupation of at least −2 to +2 subsites for efficient hydrolysis. Regarding the polymeric substrates (acetylated xylan, BW xylan, rye arabinoxylan, IWA), *Ac*Xyn30B_CD demonstrated the highest X1-X6 release, followed by *Ac*Xyn30B_CD+CBM and *Ac*Xyn30B that showed progressively lower neutral XOs release (Figure 1). However, the product profile of *Ac*Xyn30B_CD+CBM shows more diverse and intense peaks on HPAEC graphs, when acting on BW xylan (Supplementary Figure 2) and arabinoxylan (Supplementary Figure 3), corresponding most probably to complex acidic or arabinosylated XOs, respectively, compared to *Ac*Xyn30B_CD. These data corroborate the fact that *Ac*Xyn30B_CD is more efficient on rather simple substrates, releasing a higher amount of neutral XOs of lower DP, while the

CBM-containing proteins can target more complex structures leading to products with a higher DP and complexity.

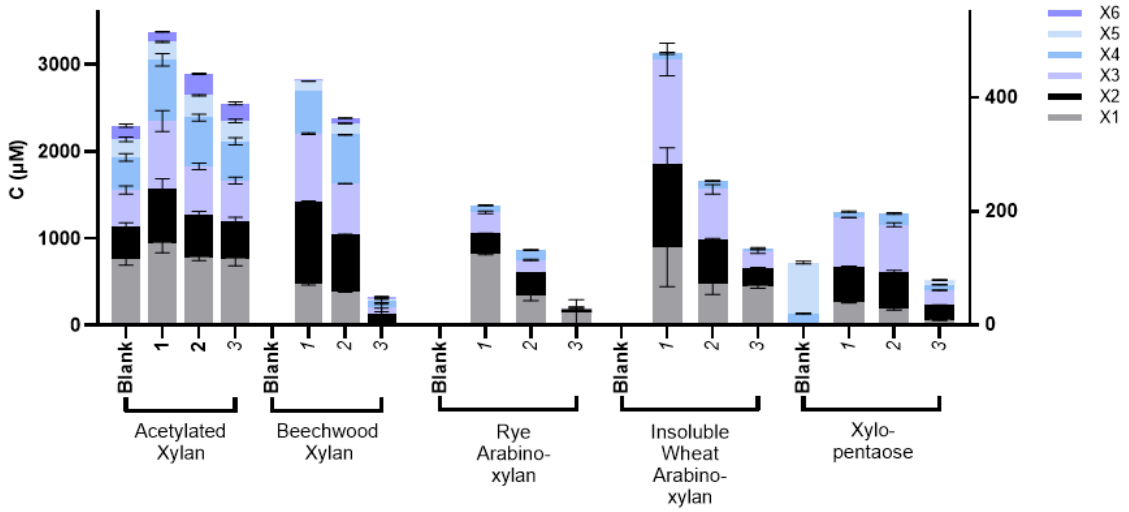


Figure 1. *Product profile of the AcXyn30B enzymes on different substrates. Blank reactions: soluble sugars of the substrates without the addition of the enzymes; 1: AcXyn30B_CD; 2: AcXyn30B_CD+CBM; 3: full-length AcXyn30B. The left Y-axis represents bars of the hydrolysis products from acetylated and BW xylans, while bars corresponding to the hydrolysis products from rye arabinoxylan, insoluble wheat arabinoxylan and xylopentaose refer to the right Y-axis. X1-X6: xylose and XOs of DP 2-6. Error bars represent the standard deviations among replicates.*

Differences in the performance of the xylanases on pretreated lignocellulosic biomasses

The three *Ac*Xyn30B constructs were tested against pretreated lignocellulosic substrates (BW, WB and CB). Even though the three variants exhibited the same product profile, *Ac*Xyn30B_CD alone seems to be more efficient on BW and WB, based on the concentration of the released sugars, compared to the xylanases with the CBM and the DK domains (Figure

2). The product profile of *Ac*Xyn30B_CD on pretreated BW was diverse, including a variety of neutral and acidic oligosaccharides (Supplementary figures 4 and 5). *Ac*Xyn30B_CD released a higher amount of X1 and X2, while X3-X6 sugars were released in similar concentrations compared to those released by *Ac*Xyn30B and *Ac*Xyn30B_CD+CBM. When acting on pretreated WB, the only sugars that could be identified and quantified were X1 and X2. The complex nature of this substrate could be responsible for the lower release of XOs, while it is also possible that more complex oligosaccharides were released as well, which could not be identified by HPAEC analysis. None of the tested variants was able to release a significant amount of sugars from the heavily substituted arabinoxylan of CB, suggesting that there are particular patterns of inhibition, despite the tolerance of *Ac*Xyn30B for both MeGlcA and Ara*f* substitutions.

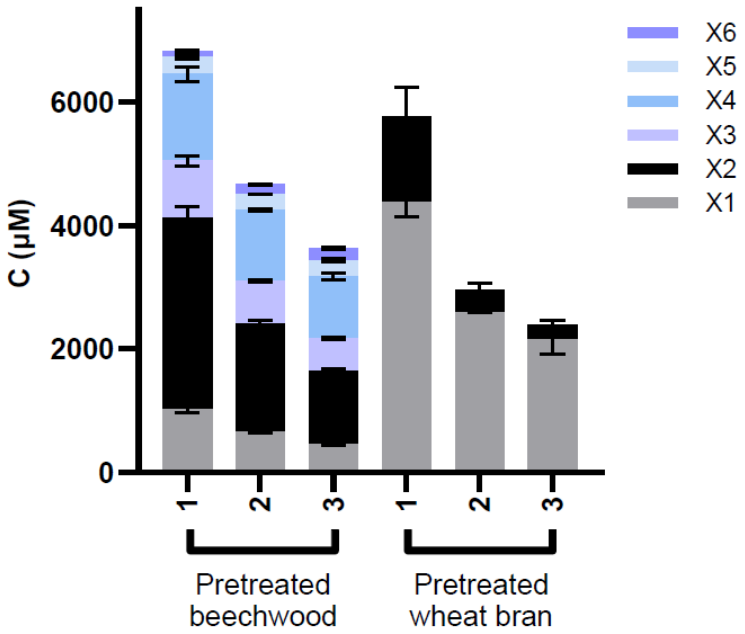


Figure 2. *Neutral oligosaccharide release of the AcXyn30B variants from pretreated lignocellulosic substrates. 1: AcXyn30B_CD; 2: AcXyn30B_CD+CBM; 3: full-length*

AcXyn30B. Blank reactions (not containing the enzyme) exhibited no release of the corresponding sugars. X1-X6: xylose and XOs of DP 2-6. Error bars represent the standard deviations among replicates.

The higher catalytic activity of *Ac*Xyn30B_CD on pretreated biomass compared with the other two variants (containing either a CBM or a CBM and DK modules) may be attributed to steric hindrance effects. The presence of these additional domains could restrict the accessibility of the catalytic domain to polysaccharide chains embedded within the complex lignocellulosic matrix, thereby reducing catalytic efficiency. A decrease in binding ability and catalytic activity has been previously observed for a GH5 endoglucanase from *Trichoderma viride* upon activity on filter paper (47), after having its native CBM removed and a CBM3 attached to the CD. On the other hand, in the case of the AA10 LPMO (48) of the modular virulence factor GbpA from *Vibrio cholerae*, the auxiliary GbpA3 domain is associated with the catalytic cleft, in addition to a CBM37 domain, preventing off-pathway reactions in the absence of substrate. When CBM73 binds to the substrate, the aforementioned interaction is weakened, enabling the activation of the LPMO. In analogy, the CBM6 or DK domains of the full-length *Ac*Xyn30B could interact with the CD, while allowing activity in the presence of the substrate. It should be noted, though, that *Ac*Xyn30B is a modular protein natively encountered in a cellulosome, meaning it could engage different dynamics in the whole system, compared to the modular xylanase alone. For instance, Hammel et al. (2004) (49) observed higher flexibility for the linker connecting the CD with the DK domains in a free enzyme, compared to when the DK was bound to the cognate cohesin. On the contrary, extended accordion-like structures have also been identified for modular proteins as part of a cellulosome (50,51), which could interfere with the catalytic activity of a CD domain.

In addition, the presence of a CBM has the capability to alter the adsorption/desorption patterns (52) of an enzyme, which could subsequently affect the catalysis and product release rate. Non-productive binding of the CBM or its competition with the CD for productive substrate positioning could also explain the persistence of longer oligosaccharides in the reaction mixture of the CBM-containing enzymes. For instance, the catalytic domain of an α-amylase from *Lactobacillus amylovorus* demonstrates activity on substrates with DP 3 or higher (53), while the corresponding CBM-containing proteins cannot hydrolyze maltotriose molecules. Similarly, *Ac*Xyn30B_CD releases X3 and X4 as final hydrolysis products, whereas longer hydrolysis products can also be detected for the CBM-containing enzymes. Thus, the CBM domain could interact with longer XOs, impeding their complete hydrolysis. Moreover, a CBM domain can contribute distinct specificities, such, for example, a preference for soluble or insoluble xylan substrates by the truncated enzymes Xyl10AdC and Xyl10AdLC from *Talaromyces cellulolyticus* (54), in contrast to the full-length xylanase which was more effective on wet disc-mill treated rice straw.

The non-denaturating affinity gel electrophoresis showed binding affinity of *Ac*Xyn30B and its truncates on rye arabinoxylan (Supplementary figure 1B). This suggests that both the CBM-like domain of the GH30 CD and the CBM6 could contribute to the arabinoxylan affinity. Moreover, the multimerization of the proteins seems to be disrupted in the presence of the polysaccharide, as all three proteins appear as one band. According to the CAZy database (9), CBM6 family involves binding affinity for amorphous cellulose, xylan and in some cases β-1,3-, β-1,4-or mixed-linked-glucan. According to the study of Abbott et al. (2009) (55), five features (A-E) of CBM6s have been suggested to attribute affinity for xylose residues and sugar oligosaccharides. Based on the sequence alignment of the CBM6 from *Ac*Xyn30B with structurally characterized CBM6 domains, the A, B and C features are identified in *Ac*Xyn30B (comprising W550, Y493 and N582, respectively), which are generally correlated with

carbohydrate recognition (Figure 3). Moreover, the CBM6 domain of *Ac*Xyn30B demonstrates two additional residues of the D domain which indicate interaction with xylan (e.g., isoleucine I483 is suggested to interact with internal regions of xylan, while glutamate E481 is regarded to interact with the non-reducing end of the chain). Finally, an extended E-loop (E region; G576-F581) involving an aromatic amino acid (Y579), is suggested to enlarge the binding site and allow the polymer to wrap around the surface of the CBM. Nevertheless, the pull-down assays of the present work demonstrated only one single protein band on SDS electrophoresis gels (data not shown), specifically in the supernatant of the incubation mixtures, either using pretreated BW or Avicel, suggesting no binding of the CBM-containing proteins to these particular insoluble substrates.

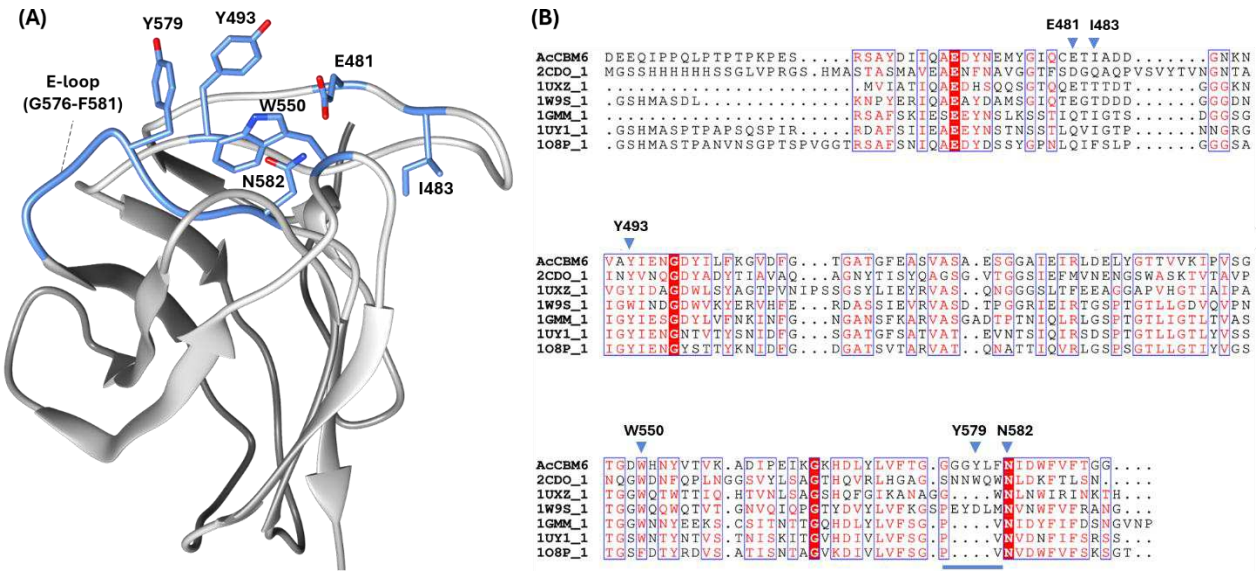


Figure 3. *(A) Structural features of the CBM6 domain of AcXyn30B suggested to contribute to carbohydrate binding. The E-loop is indicated in blue color and key amino acids are colored in blue and represented as sticks. (B) Sequence alignment of the CBM6 of AcXyn30B (AcCBM6) with structurally characterized CBM6 domains. AcCBM6 includes all five features (A-E), suggested by Abbott et al. (2009)* (55) *to be involved in carbohydrate binding. Key amino*

acids are indicated by blue arrows and the E-loop by blue line. Feature A: W550; Feature B: Y493; Feature C: N582; Feature D: I483 and E481; Feature E (E-loop): G576-F581.

Hydrolysis patterns of the catalytic domain on XOs

In order to further examine the hydrolysis pattern of the catalytic domain, *Ac*Xyn30B_CD was tested against XOs of varying DPs. X4 is slowly degraded to X3 and X1 (approximately only 17% of the substrate is hydrolyzed within 30 min). On the contrary, 30% of X5 is degraded even within 10 min, mostly towards X2 and X3 (82±3 % of the hydrolyzed X5 towards this direction) and to a lesser extent towards X1 and X4 (11±4 % of the hydrolyzed X5 towards this direction). After 30 min of hydrolysis, approximately 60 % of X5 has been hydrolyzed, maintaining, however, the same ratio between the products (89±2 % of the substrate is hydrolyzed towards X2 and X3 while 12±4 % is hydrolyzed towards X1 and X4). In the case of X6, the interpretation of the results is rather complicated, as the residual oligosaccharides present in the commercial substrate (e.g., X4, X5 or XOs with DP>6), although minor, are hydrolyzed concurrently, contributing to smaller XOs release. According to the product certificate of analysis, a 95 % purity is stated for this compound based on the HPLC peak area relative to other oligosaccharide compounds visible in the chromatograph. Nevertheless, X3 release is lower than the corresponding X2-X4 concentrations, suggesting a preference for the occupation of -4 to +2 or –2 to +4 subsites (rather than -3 to +3). For instance, after 10 min of incubation approximately 10±1 μM of X3 and 20±1 μM of X4 have been synthesized. Similarly, after 30 min of hydrolysis, when approximately 60 % of the substrate has been hydrolyzed (20±1 μM of X6 remaining), 41±3 μM of X4 and 22±2 μM of X3 can be observed.

Effect of cations and chelating agents on AcXyn30B_CD

No significant boosting effect in the enzymatic activity was observed in the presence of Ca^{2+} ions in the reaction mixture. However, X2 and X3 release was reduced approximately by 40 % and 20 %, respectively, in the presence of EDTA at different concentrations. These data, along with those previously published (16), suggest that *Ac*Xyn30B is metal-independent. Similar to that, other metal-independent enzymes have been reported to exhibit decreased activity in the presence of chelating agents (i.e., EDTA), such as the bacterial hydrolase from *Penicillium oxalicum* (56) or xylan-targeting *Tt*Xyn30A (15) from the fungus *Thermothelomyces thermophila*. The inhibitory effect posed by EDTA could be caused by the chelation of other metals or buffer components, indirectly affecting enzyme activity. In the reaction mixtures of *Ac*Xyn30B, there are Na^{+} cations from the sodium acetate/acetic acid buffering system that could interfere with the chelating agent. On the contrary, metal-dependent enzymes tend to be strongly activated in the presence of exogenously supplied cations (57), which seems not to be the case for *Ac*Xyn30B_CD. Nevertheless, the cation identified in the interface of the $(\beta/\alpha)_8$ barrel and the side-β-domain of the CD (e.g., Ca^{2+} as discussed in Section 3.3) may play other roles in the enzyme's performance besides catalytic stimulation, such as in structural integrity or stability of the protein, which could explain the minor inhibitory effect posed by the chelating agent.

3.3 Structural analysis of *Ac*Xyn30B

Crystal structure of apo-*Ac*Xyn30B and in complex with X3

The crystal structure of *Ac*Xyn30B_CD in apo state was determined at 1.14 Å resolution, while its complex with xylotriose (X3) was solved at 1.55 Å. Both structures belong to space group $P2_12_12_1$ and contain one molecule in the asymmetric unit (a.u.). PISA analysis suggests that the monomer is the biologically relevant conformation, contrary to what was observed in the SEC analysis (Section 3.1). The first four N-terminal residues (Met1-Ser4) and the last eight

C-terminal residues (Asp421-His428) of recombinant *Ac*Xyn30B_CD (Supplementary table 1) were not modelled in the final structures, as no electron density was observed for these regions. *Ac*Xyn30B_CD exhibits a classical GH30 fold, consisting of a $(\beta/\alpha)_8$ barrel catalytic domain and a side domain comprising 9 β strands. The first strand (sβ1) is located at the N-terminal (residues Tyr5-Leu18), while the other 8 (sβ2 to 9, residues Lys325-Val420) are located at the C-terminal of the trimmed variant (Figure 4A). Based on structure-based sequence alignment with *Ach*Xyn30A (PDB ID: 9YWN) (27) and structural evidence, the catalytic amino acids are Glu254 (nucleophile) and Glu146 (acid/base) (Figure 4A, Supplementary Figure 6).

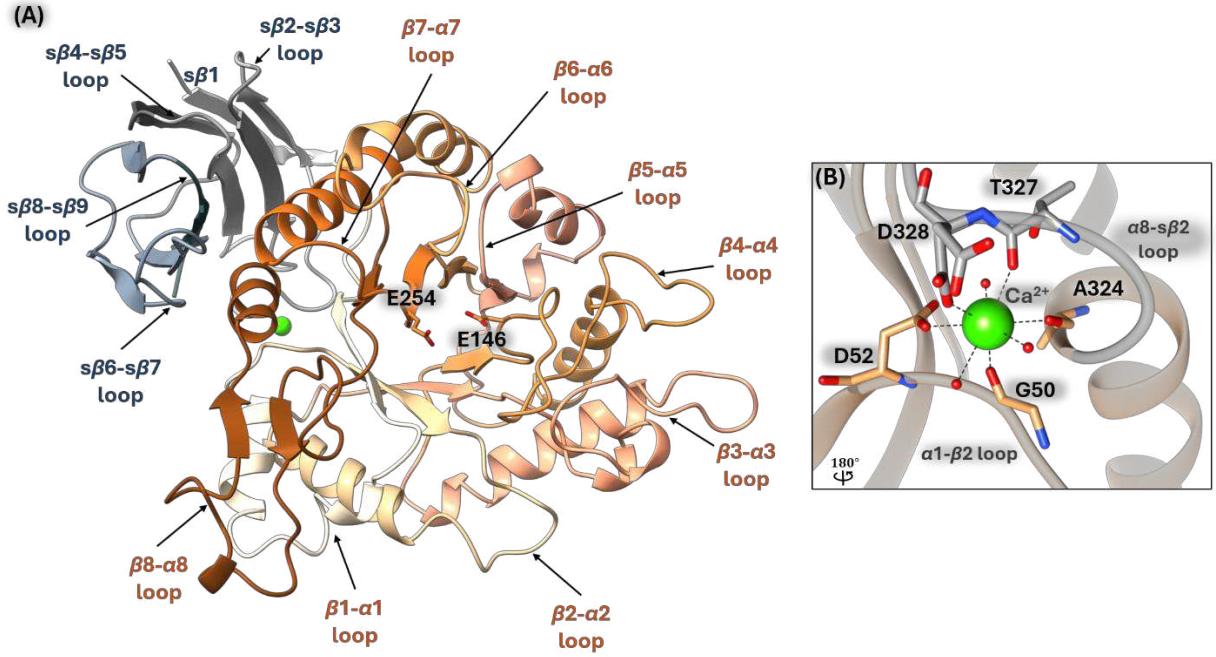


Figure 4. *(A) Overall structure of AcXyn30B_CD, where the catalytic domain is shown in orange/beige and the side domain in gray. The calcium ion at the interface of the two domains is shown as green sphere. (B) Zoomed up view of calcium binding site, depicting the amino acids and water molecules implicated in its coordination. Water molecules are shown as red spheres, and the bonds are depicted in dashed black lines.*

507 Electron density indicated the occurrence of a bound metal ion, located at the interface of the
508 two domains, between the loops connecting α1 helix with β2 strand and α8 with sβ2 (Figure
509 4A). Based on the difference map after refinement and coordination geometry, a calcium ion
510 was modelled at that site. It is coordinated to the backbone carbonyl oxygens of Gly50, Ala324
511 and Thr327, the side chain carboxyl groups of Asp52 and Asp328, as well as 3 water molecules
512 (Figure 4B, Supplementary Figure 7). Validation using CheckMyMetal server (58,59)
513 confirmed that the metal-oxygen distances and coordination geometry are consistent with Ca^{2+}
514 binding. Although the presence of calcium ion has previously been reported to enhance the
515 activity of *Ac*Xyn30B (16), its crystal structure revealed that the ion is located away from the
516 catalytic site of the enzyme (Figure 3A) and is therefore unlikely to be directly involved in
517 catalysis, as previously discussed in Section 3.2. Instead, it may contribute to the structural
518 integrity or stability of the protein.
519 Soaking crystals of *Ac*Xyn30B_CD with X5 yielded a complex structure with the hydrolytic
520 product X3. The overall fold of the enzyme remains almost identical between its apo and
521 product-bound states, with a global root mean square deviation (RMSD) of 0.097 Å per C_α
522 atoms. Three xylose moieties could be unambiguously modelled, spanning subsites –4 to –2
523 (Figure 5A and B). Additional electron density was observed at the reducing end of the bound
524 trisaccharide, at the position corresponding to subsite –1. However, this density was
525 insufficient to support reliable modelling of a sugar molecule and was therefore left unmodelled
526 (Figure 5B). An imidazole molecule, likely originating from the protein purification process,
527 was also observed occupying the potential +1 subsite (60). The presence of this molecule at
528 subsite +1 is consistent with the ability of planar heteroatomic compounds to bind tightly in
529 the active site of retaining glycosidases, as these molecules can be protonated by the catalytic
530 acid residue (60,61). In addition, two glycerol molecules bound in the vicinity may indicate the
531 location of the +2 and +4 subsites (Figure 5A). A well-defined water molecule was identified

between the catalytic glutamates (Glu146 and Glu254), forming hydrogen bonds with both catalytic residues and the side chain amide nitrogen of Asn145 (Figure 5C). This water molecule is positioned at the location corresponding to the cleavable glycosidic bond, suggesting potential involvement in the catalytic mechanism.

The residues implicated in substrate recognition via hydrogen-bonding and aromatic stacking (CH-π) interactions (62), unless otherwise stated, are shown in Figure 5A. Starting from the non-reducing end of the bound ligand, the –4 subsite is defined by Asp30 and Tyr28 which are located in the middle of loop β1-α1, with Tyr28 forming CH-π stacking interactions with the xylose occupying this subsite. The –3 subsite is formed through aromatic stacking interactions with Trp27 and Trp286, as well as hydrogen-bonding interactions with the main chain carbonyl group of Asp153. The xylose at subsite –2 is mainly stabilized by Gln290 and Trp154, with the side chain amide of Gln290 (loop β8-α8) forming a hydrogen bond with the O3 hydroxyl group of the –2 xylose. Comparison with the apo *Ac*Xyn30B_CD (PDB ID: 9QY5) suggests that xylose binding seems to stabilize Gln290, as it adopts a single well-defined side chain conformation in the complex, whereas two alternative conformations are observed in the apo form. The –1 subsite is located around the unmodelled electron density, suggesting that the conserved residues Trp91 and Trp282 contribute to substrate binding, in agreement with their proposed role in other GH30 xylanases.

At the positive subsites, the pocket is narrowed by the aromatic side chains of Tyr149 and Tyr223, potentially contributing to +1 subsite. Although no xylose ligand was observed in this region, an imidazole molecule was identified in the electron density map, positioned between these aromatic residues and forming stacking interactions with both tyrosines. Tyr257, located on the loop β7-α7, has its hydroxyl oxygen positioned 4.5 Å from the nearest imidazole carbon atom and, could also be implicated in hydrogen bond formation with a xylose residue bound at subsite +1. At the potential +2 subsite, Asn224 appears to contribute to substrate binding,

whereas the potential +4 subsite, defined by the bound glycerol molecule, is likely defined by Glu161, Asn200 and Asn205 at the interface between the loop β4-α4 and the α5 helix.

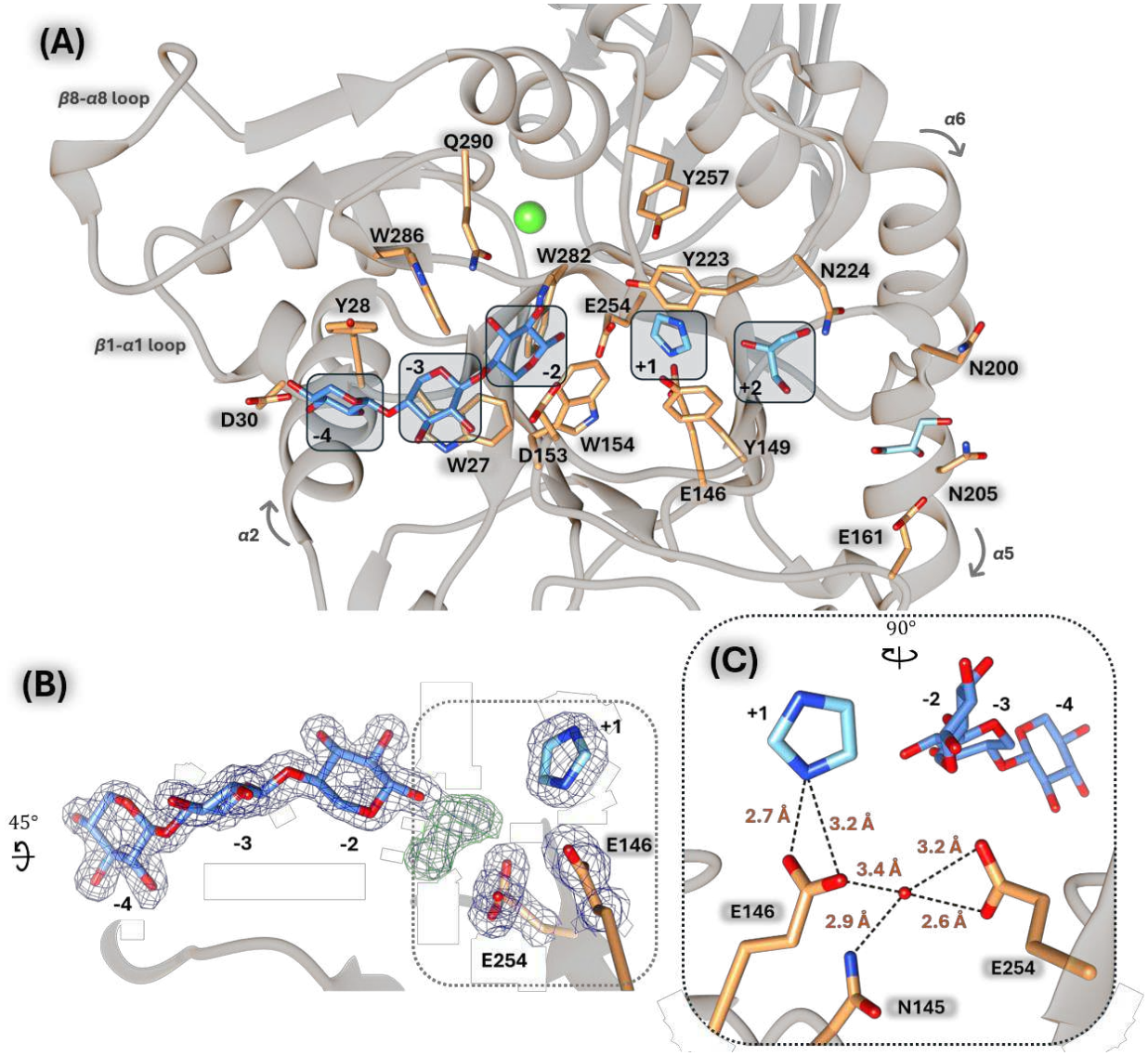


Figure 5. *(A) Xylan binding cleft, indicating subsites –4 to +2. Residues implicated in substrate binding are shown as orange sticks, while X3 and imidazole/glycerol molecules are presented in cornflower and light blue sticks, respectively. (B) Electron density map of X3-bound structure, focusing on X3 and highlighting the unmodelled density at the reducing end of the bound ligand. The $2mF_o - DF_c$ electron density map is contoured at 1σ displayed as blue mesh, whereas the $mF_o - DF_c$ electron density map is contoured at 3σ and is indicated as green mesh.*

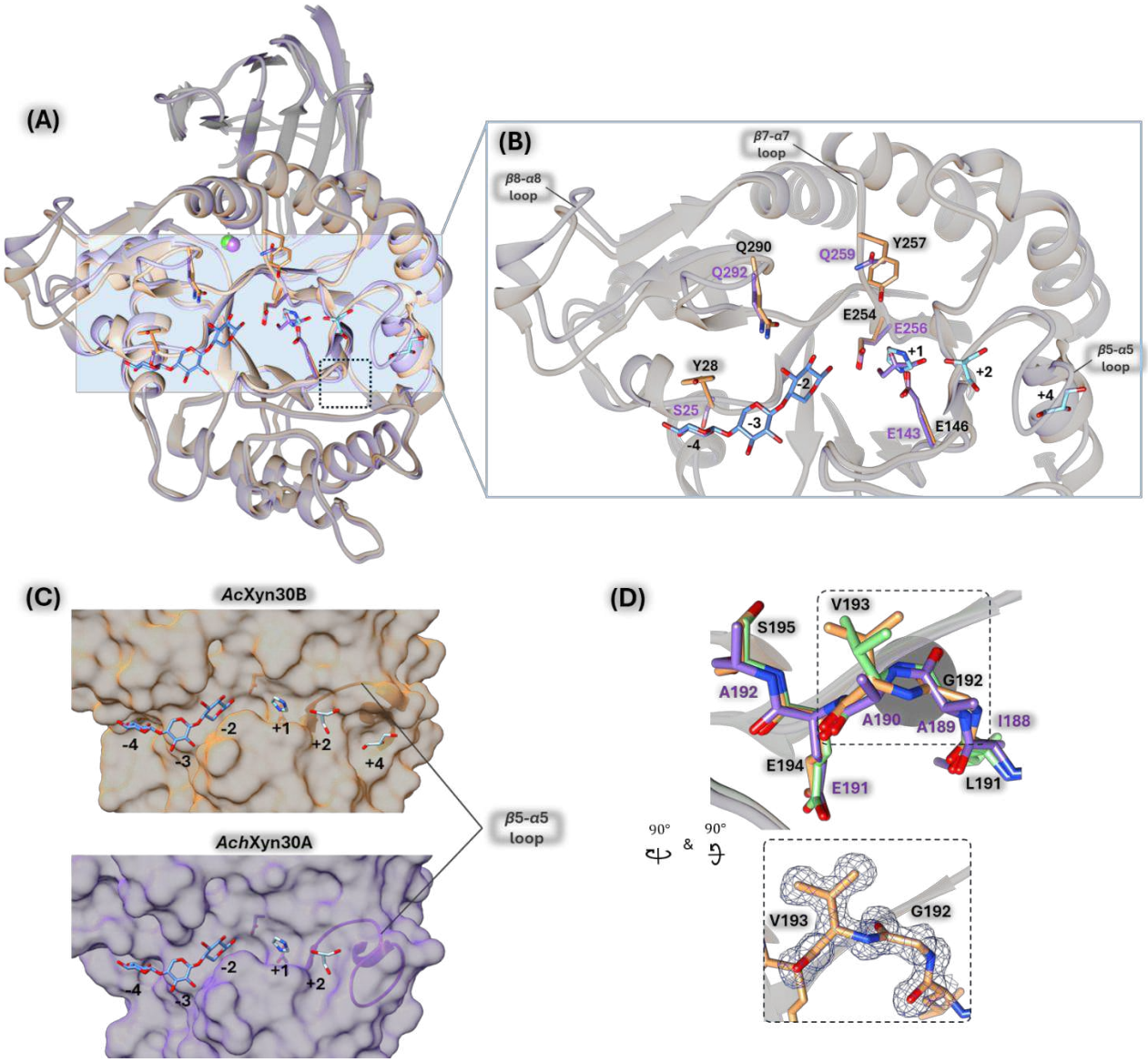
(C) Close-up view of H-bond network formed by the catalytic glutamates. Black dashed lines depict hydrogen bonds and red sphere presents the water molecule between the two glutamates.

Comparison with crystal structure of *Ach*Xyn30A

Focusing on the recently established GH30_12 family (11), the *Ac*Xyn30B_CD-X3 structure was compared with the first structurally determined GH30_12 family member from *A. chartisolvens*, *Ach*Xyn30A (PDB ID: 9YWN) (27). Superposition of the two enzyme structures over C_α atoms yielded an RMSD of 1.04 Å, revealing a highly conserved overall fold despite several notable structural differences (Figure 6A). Interestingly, the two structures share a glutamine residue at the loop β8-α8 (Gln290 in *Ac*Xyn30B and Gln292 in *Ach*Xyn30A), which contributes to the formation of –2 subsite in the glycone region (Figure 6B). This residue appears to be a distinctive feature of these enzymes, as the corresponding glutamine is not conserved among other GH30_12 members (Supplementary Figure 8). Globally, the residues involved in substrate binding are also highly conserved between the two enzymes, with only two differences observed in subsites –4 and +1. Specifically, Tyr28 and Tyr257 in *Ac*Xyn30B and Ser25 and Gln259 in *Ach*Xyn30A, respectively (Figure 6B). Notably, Tyr257 is located in the relatively short loop β7-α7, which also contains the catalytic nucleophile residue (Glu254 in *Ac*Xyn30B and Glu256 in *Ach*Xyn30A) and is shorter in both enzymes than the corresponding loops in other GH30 xylanases (Supplementary Figure 9). It should be noted that in *Ach*Xyn30A, a glycerol molecule occupies the position corresponding to the imidazole molecule observed at +1 subsite of *Ac*Xyn30B_CD-X3 (Figure 6B).

Despite these conserved characteristics, several notable structural differences were observed between the two enzymes. First, while a metal-binding site was observed in both crystal structures, occupying the same position between the loops α1-β2 and α8-sβ2, it was modelled as sodium ion in *Ach*Xyn30A, and a calcium ion in *Ac*Xyn30B_CD (Figure 6A). The sodium

592 ion in the structure of *Ach*Xyn30A is coordinated by residues Gly47, Asp49, Ala326, Thr329
593 and Glu330, corresponding to Gly50, Asp52, Ala324, Thr327 and Asp328, respectively, in
594 *Ac*Xyn30B_CD. The number of water molecules participating in the coordination also differs
595 between the two structures. In *Ach*Xyn30A, one water molecule participates in the metal
596 coordination, whereas three water molecules are involved in the coordination of the calcium
597 ion in *Ac*Xyn30B_CD.
598 Second, another major structural difference is observed in loop β5-α5, which contributes to
599 substrate binding at the aglycone subsites. An insertion of five amino acids occurring at
600 *Ach*Xyn30A (Supplementary Figure 6) results in steric inhibition at subsites beyond +3, at a
601 location where a glycerol molecule is bound in *Ac*Xyn30B_CD-X3 structure, potentially
602 indicating the +4 subsite (Figure 6B and C). Consequently, the substrate-binding channel in
603 *Ach*Xyn30A is more restricted in this region, potentially affecting the preferred xylan cleavage
604 mode. The β5 strand of *Ach*Xyn30A harbors a conserved *cis*-peptide bond between Ala189 and
605 Ala190, a structural feature observed in several GH30 xylanases (17,20,27,63). In contrast, the
606 corresponding peptide bond between Gly192 and Val193 in *Ac*Xyn30B_CD adopts a *trans*
607 conformation in both apo and X3-bound structures (Figure 6D), consistent with that observed
608 in *Ac*Xbh30A, a GH30_10 xylobiohydrolase from *A. clariflavus* (18). Notably, the initial
609 AlphaFold-predicted structure of *Ac*Xyn30B_CD used for molecular replacement adopts a *cis*
610 conformation at this peptide bond, whereas the experimentally determined structures
611 unambiguously indicate a *trans* conformation (Figure 6D).
612



613

614 ***Figure 6.*** *(A) AcXyn30B_CD-X3 (in orange) structure superimposed on AchXyn30A (in*
615 *purple; PDB ID: 9YWN). The ligands bound in AcXyn30B are shown in stick representation,*
616 *indicating the xylan binding cleft. The calcium and sodium ions are shown as green and purple*
617 *spheres, respectively. The dashed square indicates the region of trans peptide bond described*
618 *in (D). (B) Close-up view of the binding cleft. (C) Surface representation of the two structures,*
619 *with the ligands bound in AcXyn30B in stick representation. The loop β5-α5* is shown in
620 cartoon, highlighting its extended conformation in *AchXyn30A*. *(D) Superposition of*
621 *AcXyn30B_CD-X3 (in orange), AchXyn30A (in purple) and AlphaFold-predicted AcXyn30B*
622 *(in green), highlighting the peptide bond that adopts a trans conformation in AcXyn30B. The*

$2mF_o – DF_c$ electron density map is contoured at 1σ displayed as blue mesh, confirming the observed trans peptide bond conformation.

Accommodation of side decorations based on comparison with other GH30 enzymes

According to previous studies by Suchova *et al.* (11), cleavage of substituted substrates by *Ac*Xyn30B is favored when at least four catalytic subsites are occupied. The substituted xylose residue can be accommodated at several subsites, with subsite –2 appearing to be most favorable, whereas glucuronic acid substitution at subsite –1 is not tolerated. Soaking *Ac*Xyn30B_CD crystals with XUXX resulted in no well-defined electron density in the substrate binding cleft, with the structure being determined at 1.79 Å resolution (Figure 7A). Superposition with *Ac*Xyn30B_CD-X3 (Figure 7B), indicates that the main part of the observed density is consistent with three xylose residues occupying subsites –4 to –2, while an imidazole molecule is also present at subsite +1. At the reducing end of the putative substrate, additional electron density is observed that resembles the density found at the reducing end of the bound X3 (Figure 4B).

To identify putative binding mode of acidic XOs within the active site of *Ac*Xyn30B, a comparison was made with GH30 structures in complex with acidic XOs, available in the Protein Data Bank (PDB). Specifically, two structures containing uronic xylotriose were identified, one belonging to the GH30_7 subfamily, corresponding to the inactive double mutant *Tt*Xyn30AEE from the thermophilic fungus *Thermothelomyces thermophila* (20), and the second to the GH30_8 subfamily, corresponding to *Ec*Xyn30A from the phytopathogenic bacterium *Erwinia chrysanthemi* (64). Based on structural comparison of *Ac*Xyn30B_CD-X3 with *Tt*Xyn30A-UXX and *Ec*Xyn30A-XUX over all atoms (Figure 7C, D and Supplementary Figure 9), accommodation of an uronic acid decoration at subsite –2 in *Ac*Xyn30B is less favored due to the presence of Gln290, unless a side chain shift is induced upon substrate

binding. Comparison with *Tt*Xyn30A-UXX (Figure 7C) further reveals that the side chain of Trp286 occupies a position corresponding to the side chain of Arg34, a residue conserved in many GH30 xylanases known to contribute to glucuronic acid recognition. The presence of an aromatic side chain at this position may therefore provide an alternative mode of substrate recognition, with Trp286 contributing to xylan binding via CH-π stacking interactions. Additionally, the α2-β2 loop of *Tt*Xyn30A, which is known to append xylobiohydrolase activity in GH30 xylanases, is shorter in *Ac*Xyn30B (Figure 7C), consistent with its apparent endo mode of action. Superimposition of *Ac*Xyn30B_CD-X3 on *Tt*Xyn30A shows that the presence of a more elongated loop at the corresponding position (loop α2-β2) of *Ac*Xyn30B would prevent binding of xylose residues beyond subsite –2.

Regarding the superposition of *Ac*Xyn30B_CD-X3 with *Ec*Xyn30A-XUX (Figure 7D), which contains the acidic xylooligosaccharide expected as the reaction product of *Ac*Xyn30B acting on XUXX (11), the position corresponding to Trp286 is occupied by Arg293 in *Ec*Xyn30A, mediating substrate recognition through an ionic interaction with the carboxylate group of the uronic acid. Additionally, an imidazole molecule is also present at subsite +1, as observed in the *Ac*Xyn30B_CD-X3 structure, where it similarly interacts electrostatically with the acid/base glutamate (Glu165). Finally, the α8-β8 loop of *Ec*Xyn30A is significantly shorter than that of *Ac*Xyn30B, while the α2-β2 loops are similar in length between the two structures (Figure 7D).

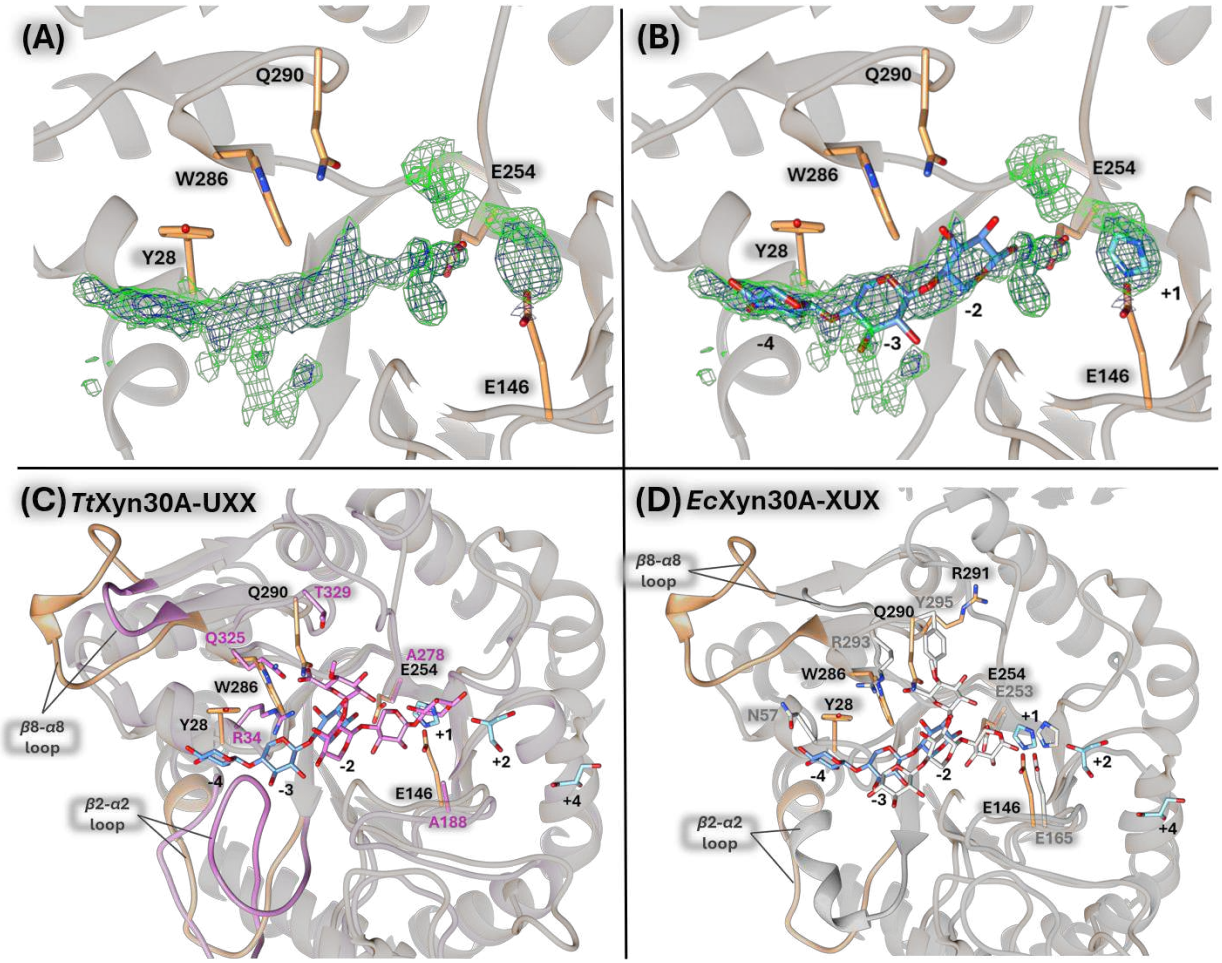


Figure 7. *(A) Electron density observed in AcXyn30B_CD crystal structure obtained from crystals soaked in mother liquor supplemented with 30 mM XUXX. The $2mF_o - DF_c$ and $mF_o - DF_c$ electron density maps are contoured at 1σ (blue mesh) and 3σ (green mesh), respectively. (B) Putative accommodation of X3 and imidazole molecule from AcXyn30B_CD-X3 (PDB ID: 9SLE) into the electron density observed in XUXX-soaked structure after superposition of the two structures. (C) Superposition of AcXyn30B_CD-X3 with TtXyn30AEE-UXX (in pink; PDB ID: 8P67) and (D) EcXyn30A-XUX (light gray; PDB ID: 2Y24).*

4. Discussion

The recently established GH30_12 subfamily of non-specific endoxylanases exhibits considerable biotechnological potential for applications involving hemicellulose valorization. Their tolerance towards substituted xylans represents a particularly appealing feature for

681 industrial biocatalysis, since it enables the conversion of a wide range of structurally diverse
682 substituted xylan polymers and oligomers. The biochemical and structural characteristics of
683 *Ac*Xyn30B, described in the present study, provide new insight into the mode of action and the
684 molecular basis underlying the non-specific functionality of GH30_12 endoxylanases.

685 Specific hydrolysis patterns of *Ac*Xyn30B_CD can be observed upon activity on linear XOs.
686 While X4 is slowly hydrolyzed towards X1 & X3, X5 and X6 substrates are more readily
687 hydrolyzed towards X2 & X3 and X2 & X4, respectively. This pattern suggests that the CD
688 preferentially accommodates substrates occupying more than 2 subsites on one side of the
689 cleavage site and at least 2 subsites on the other, without being able to distinguish which side
690 refers to the (+) and which to the (-) subsites of the active site.

691 The appended CBM and DK domains alter enzyme efficiency, substrate preference and product
692 profiles. *Ac*Xyn30B_CD preferably acts on BW xylan, while when CBM6 or CBM6-DK
693 domains are present, the enzyme has higher specificity for rye arabinoxylan and lower
694 specificity for BW xylan. Taking a closer look at the product profile on polymeric substrates,
695 *Ac*Xyn30B_CD readily hydrolyzes neutral XOs of lower DP, compared to the CBM-containing
696 enzymes, which release lower concentrations of xylose and neutral XOs. On the other hand,
697 *Ac*Xyn30B_CD+CBM shows a more diverse product profile of acidic or arabinosylated XOs.
698 On more complex lignocellulosic substrates (e.g., pretreated BW and WB biomass), the
699 contribution of the CBM6 and DK domains is rather inhibitory.

700 The decreased hydrolysability of XOs or the increased specificity for rye arabinoxylan by the
701 CBM6-containing variants could be justified by specific features of the CBM6 domain
702 correlated with the recognition and binding of xylan and XOs. These structural traits, for
703 instance, could lead to non-productive binding of linear XOs or increased affinity for
704 arabinoxylan polymeric chains, which is yet to be verified.

*Ac*Xyn30B_CD crystal structure was determined in apo and product-bound forms, shedding light on the architecture of the catalytic cleft and the xylose-binding subsites. Additionally, a calcium ion was identified between the catalytic and side domain of the protein. Although *Ac*Xyn30B_CD is metal-independent, the calcium ion may contribute to the structural integrity or stability of the protein. In the ligand-bound structure, X3 was bound at subsites –4 to –2 after soaking crystals with X5, while subsites +1, +2 and +4 could be traced due to the presence of an imidazole and two glycerol molecules, respectively. Additionally, weak electron density at subsite –1 indicates the presence of an additional xylose molecule reducing end of X3, though density was insufficient for reliable modelling.

Regarding the loops surrounding the cleft and determining specificity and mode of action, loop *β*5-α5 is shorter in *Ac*Xyn30B_CD than in *Ach*Xyn30A, the other structurally characterized GH30_12 enzyme, giving a more open cleft toward the positive subsites. Conversely, both enzymes share shortened *β*7-*α*7 (above the catalytic glutamates, on one side of the cleft) and *β*2-*α*2 (at the cleft edge toward the glycone subsites) loops relative to other GH30 xylanases, which is consistent with endo-activity on decorated polymers. While other endo-xylanases from other GH30 families are known to accommodate an uronic acid decoration at subsite –2, in *Ac*Xyn30B_CD this is less favored due to the presence of Gln290, unless a side chain shift is induced upon substrate binding. Last, the findings regarding the *trans* peptide bond between residues Gly192 and Val19 in *Ac*Xyn30B_C, modeled as *cis*-peptide bond of by AlphaFold, provide strong evidence of the limitations of computational tools and the need to combine both experimental and computational approaches to achieve detailed biochemical insights

Overall, these findings advance our understanding on the structure–function determinants of GH30_12 non-specific endoxylanases and create a foundation for the engineering of

biocatalysts with tailored properties for the selective deconstruction and valorization of structurally diverse xylan-containing feedstocks.

5. Conclusions

This study provides new structural and functional insights into the recently identified GH30_12 subfamily of non-specific endo-xylanases, highlighting the contribution of individual protein domains to substrate recognition and hydrolysis. The presence of the CBM and DK modules appears to shift the preference of *Ac*Xyn30B_CD from beechwood glucuronoxylan towards rye arabinoxylan, while affinity electrophoresis further supports the ability of *Ac*Xyn30B and its truncated variants to interact with arabinoxylan. Regarding the CD module, an open binding cleft allows the accommodation of a decorated xylan chain. At least 4 subsites are required for efficient hydrolysis, with a preference for the occupation of -4 to +2 or –2 to +4 subsites. Collectively, these findings advance our understanding on the functinal role of the CD and its adjancent modules, as well as on the hydrolysis traits of *Ac*Xyn30B, providing a foundation for further investigation and potential exploitation of GH30_12 enzymes.

6. Acknowledgements

We acknowledge the MAX IV Laboratory for beamtime on the BioMAX beamline under proposal 20240403. Research conducted at MAX IV, a Swedish national user facility, is supported by Vetenskapsrådet (Swedish Research Council, VR) under contract 2018-07152, Vinnova (Swedish Governmental Agency for Innovation Systems) under contract 2018-04969 and Formas under contract 2019-02496. We would like to thank Dr. Ana Gonzalez for the assistance in using the beamline. The research leading to this result has been co-funded by the

project NEPHEWS under Grant Agreement No 101131414 from the EU Framework Programme for Research and Innovation Horizon Europe.

Author contributions

Panagiota Karampa: Investigation; data analysis; writing original draft; review and editing.

Christina Pentari: Investigation; data analysis; writing original draft; review and editing.

Evangelos Topakas: Conceptualization, supervision, writing – review and editing, funding acquisition.

Maria Dimarogona: Conceptualization, supervision, writing – review and editing, funding acquisition.

Funding

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101221903. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Innovation Council and SMEs Executive Agency (EISMEA). Neither the European Union nor the granting authority can be held responsible for them.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Declarations

Prof Evangelos Topakas is an Editor-in-Chief of the journal Biotechnology for Biofuels and Bioproducts. The peer-review of this paper was therefore overseen by an independent member of the editorial board while the submission was subject to the exact same review process as applied to any other manuscript.

8. References

1. d'Errico C, Jørgensen JO, Krogh KBRM, Spodsberg N, Madsen R, Monrad RN. Enzymatic degradation of lignin-carbohydrate complexes (LCCs): Model studies using a fungal glucuronoyl esterase from Cerrena unicolor. Biotechnology and Bioengineering. 2015;112(5):914–22. doi:10.1002/bit.25508

2. Huang C, He J, Li X, Min D, Yong Q. Facilitating the enzymatic saccharification of pulped bamboo residues by degrading the remained xylan and lignin–carbohydrates complexes. Bioresource Technology. 2015 Sep 1;192:471–7. doi:10.1016/j.biortech.2015.06.008

3. Zhao Y, Shakeel U, Saif Ur Rehman M, Li H, Xu X, Xu J. Lignin-carbohydrate complexes (LCCs) and its role in biorefinery. Journal of Cleaner Production. 2020 Apr 20;253:120076. doi:10.1016/j.jclepro.2020.120076

4. Busse-Wicher M, Gomes TCF, Tryfona T, Nikolovski N, Stott K, Grantham NJ, et al. The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of Arabidopsis thaliana. The Plant Journal. 2014;79(3):492–506. doi:10.1111/tpj.12575

5. Huang LZ, Ma MG, Ji XX, Choi SE, Si C. Recent Developments and Applications of Hemicellulose From Wheat Straw: A Review. Front Bioeng Biotechnol. 2021 Jun 22;9. doi:10.3389/fbioe.2021.690773

6. Martins M, Sganzerla WG, Forster-Carneiro T, Goldbeck R. Recent advances in xylo-oligosaccharides production and applications: A comprehensive review and bibliometric analysis. Biocatalysis and Agricultural Biotechnology. 2023 Jan 1;47:102608. doi:10.1016/j.bcab.2023.102608

7. Naidu DS, Hlangothi SP, John MJ. Bio-based products from xylan: A review. Carbohydrate Polymers. 2018 Jan 1;179:28–41. doi:10.1016/j.carbpol.2017.09.064

8. Zerva A, Pentari C, Ferousi C, Nikolaivits E, Karnaouri A, Topakas E. Recent advances on key enzymatic activities for the utilisation of lignocellulosic biomass. Bioresource Technology. 2021 Dec 1;342:126058. doi:10.1016/j.biortech.2021.126058

9. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 2022 Jan 7;50(D1):D571–7. doi:10.1093/nar/gkab1045

10. Šuchová K, Spodsberg N, Mørkeberg Krogh KBR, Biely P, Puchart V. Non-Specific GH30_7 Endo-β-1,4-xylanase from Talaromyces leycettanus. Molecules. 2021 Jan;26(15):4614. doi:10.3390/molecules26154614

11. Šuchová K, Fathallah W, Puchart V. Characterization of a novel GH30 non-specific endoxylanase AcXyn30B from Acetivibrio clariflavus. Appl Microbiol Biotechnol. 2024 Apr 29;108(1):312. doi:10.1007/s00253-024-13155-w

12. Verma AK, Goyal A. A novel member of family 30 glycoside hydrolase subfamily 8 glucuronoxylan *endo*-β-1,4-xylanase (*Ct*XynGH30) from *Clostridium thermocellum* orchestrates catalysis on arabinose decorated xylans. Journal of Molecular Catalysis B: Enzymatic. 2016 Jul 1;129:6–14. doi:10.1016/j.molcatb.2016.04.001

13. Hurlbert JC, Preston JF. Functional Characterization of a Novel Xylanase from a Corn Strain of Erwinia chrysanthemi. Journal of Bacteriology. 2001 Mar 15;183(6):2093–100. doi:10.1128/jb.183.6.2093-2100.2001

14. Šuchová K, Puchart V, Spodsberg N, Mørkeberg Krogh KBR, Biely P. A novel GH30 xylobiohydrolase from *Acremonium alcalophilum* releasing xylobiose from the non-reducing end. Enzyme and Microbial Technology. 2020 Mar 1;134:109484. doi:10.1016/j.enzmictec.2019.109484

15. Katsimpouras C, Dedes G, Thomaidis NS, Topakas E. A novel fungal GH30 xylanase with xylobiohydrolase auxiliary activity. Biotechnol Biofuels. 2019 May 11;12(1):120. doi:10.1186/s13068-019-1455-2

16. Choudhury B, Singh YR, Khaire KC, Ahmed N, Sharma K, Fontes CMGA, et al. Cellulosomal *endo*-1,4-*β*-D-xylanase (*Ac*Xyn30B_12) from *Acetivibrio clariflavus* acts synergistically with xylobiohydrolase (*Ac*GH30A) upon the hydrolysis of complex carbohydrates. International Journal of Biological Macromolecules. 2025 May 1;306:141620. doi:10.1016/j.ijbiomac.2025.141620

17. Urbániková Ľ, Vršanská M, Mørkeberg Krogh KBR, Hoff T, Biely P. Structural basis for substrate recognition by Erwinia chrysanthemi GH30 glucuronoxylanase. The FEBS Journal. 2011;278(12):2105–16. doi:10.1111/j.1742-4658.2011.08127.x

18. St John FJ, Crooks C, Kim Y, Tan K, Joachimiak A. The first crystal structure of a xylobiose-bound xylobiohydrolase with high functional specificity from the bacterial glycoside hydrolase family 30, subfamily 10. FEBS Letters. 2022;596(18):2449–64. doi:10.1002/1873-3468.14454

19. Nikolaivits E, Pentari C, Kosinas C, Feiler CG, Spiliopoulou M, Weiss MS, et al. Unique features of the bifunctional GH30 from *Thermothelomyces thermophila* revealed by structural and mutational studies. Carbohydrate Polymers. 2021 Dec 1;273:118553. doi:10.1016/j.carbpol.2021.118553

20. Pentari C, Kosinas C, Nikolaivits E, Dimarogona M, Topakas E. Structural and molecular insights into a bifunctional glycoside hydrolase 30 xylanase specific to glucuronoxylan. Biotechnology and Bioengineering. 2024;121(7):2067–78. doi:10.1002/bit.28731

21. Šuchová K, Puchart V. Cellulolytic and hemicellulolytic capacity of Acetivibrio clariflavus. Appl Microbiol Biotechnol. 2025 Apr 28;109(1):105. doi:10.1007/s00253-025-13471-9

22. Shinoda S, Kurosaki M, Kokuzawa T, Hirano K, Takano H, Ueda K, et al. Comparative Biochemical Analysis of Cellulosomes Isolated from Clostridium clariflavum DSM 19732 and Clostridium thermocellum ATCC 27405 Grown on Plant Biomass. Appl Biochem Biotechnol. 2019 Mar 1;187(3):994–1010. doi:10.1007/s12010-018-2864-6

23. Artzi L, Morag E, Barak Y, Lamed R, Bayer EA. Clostridium clariflavum: Key Cellulosome Players Are Revealed by Proteomic Analysis. mBio. 2015 May 19;6(3):10.1128/mbio.00411-15. doi:10.1128/mbio.00411-15

24. Šuchová K, Puchart V, Biely P. A novel bacterial GH30 xylobiohydrolase from Hungateiclostridium clariflavum. Appl Microbiol Biotechnol. 2021 Jan 1;105(1):185–95. doi:10.1007/s00253-020-11023-x

25. Crooks C, Bechle NJ, St John FJ. A New Subfamily of Glycoside Hydrolase Family 30 with Strict Xylobiohydrolase Function. Front Mol Biosci. 2021 Sep 7;8. doi:10.3389/fmolb.2021.714238

26. Choudhury B, Goyal A. Deciphering the structural insights and concentration-dependent dimerisation of endo-β-1,4-xylanase (AcXyn30B_12) from Acetivibrio clariflavus using SAXS and computational methods. The FEBS Journal. 2026;293(12):3640–62. doi:10.1111/febs.70432

27. St John FJ, Crooks C, Endres M, Pakdaman L, Koch L, Bynum L, et al. Protein structure of a glycoside hydrolase family 30, subfamily 12 endo-1,4-β-xylanase. Acta Cryst D. 2026 Apr 1;82(4). doi:10.1107/S2059798326002160

28. Katsimpouras C, Kalogiannis KG, Kalogianni A, Lappas AA, Topakas E. Production of high concentrated cellulosic ethanol by acetone/water oxidized pretreated beech wood. Biotechnol Biofuels. 2017 Feb 28;10(1):54. doi:10.1186/s13068-017-0737-9

29. Pentari C, Mylona EP, Zerva A, Topakas E. Synergistic effects of distinct arabinofuranosidase specificities in lignocellulose degradation by different hemicellulases. International Journal of Biological Macromolecules. 2025 Apr 1;302:140575. doi:10.1016/j.ijbiomac.2025.140575

30. Kalogiannis KG, Karnaouri A, Michailof C, Tzika AM, Asimakopoulou G, Topakas E, et al. OxiOrganosolv: A novel acid free oxidative organosolv fractionation for lignocellulose fine sugar streams. Bioresource Technology. 2020 Oct 1;313:123599. doi:10.1016/j.biortech.2020.123599

31. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. The Proteomics Protocols Handbook [Internet]. Totowa, NJ: Humana Press; 2005 [cited 2026 Mar 23]. p. 571–607. Available from: https://doi.org/10.1385/1-59259-890-0:571 doi:10.1385/1-59259-890-0:571

32. Laemmli UK. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature. 1970 Aug;227(5259):680–5. doi:10.1038/227680a0

33. Pentari C, Zerva A, Dimarogona M, Topakas E. The xylobiohydrolase activity of a GH30 xylanase on natively acetylated xylan may hold the key for the degradation of recalcitrant xylan. Carbohydrate Polymers. 2023 Apr 1;305:120527. doi:10.1016/j.carbpol.2022.120527

34. Miller GL. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal Chem. 1959 Mar 1;31(3):426–8. doi:10.1021/ac60147a030

35. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2022 Jul 5;50(W1):W276–9. doi:10.1093/nar/gkac240

36. Robert X, Guillon C, Gouet P. FoldScript: a web server for the efficient analysis of AI-generated 3D protein models. Nucleic Acids Res. 2025 Jul 7;53(W1):W277–82. doi:10.1093/nar/gkaf326

37. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. Journal of Computational Chemistry. 2004;25(13):1605–12. doi:10.1002/jcc.20084

38. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9. doi:10.1038/s41586-021-03819-2

39. Ursby T, Åhnberg K, Appio R, Aurelius O, Barczyk A, Bartalesi A, et al. BioMAX – the first macromolecular crystallography beamline at MAX IV Laboratory. J Synchrotron Rad. 2020 Sep 1;27(5):1415–29. doi:10.1107/S1600577520008723

40. D'Arcy A, Bergfors T, Cowan-Jacob SW, Marsh M. Microseed matrix screening for optimization in protein crystallization: what have we learned? Acta Crystallogr F Struct Biol Commun. 2014 Aug 29;70(Pt 9):1117–26. doi:10.1107/S2053230X14015507 PubMed PMID: 25195878; PubMed Central PMCID: PMC4157405.

41. Kabsch W. XDS. Acta Cryst D. 2010 Feb 1;66(2):125–32. doi:10.1107/S0907444909047337

42. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr. 2013 Jul 1;69(Pt 7):1204–14. doi:10.1107/S0907444913000061 PubMed PMID: 23793146; PubMed Central PMCID: PMC3689523.

43. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Cryst. 2007 Aug 1;40(4):658–74. doi:10.1107/S0021889807021206

44. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Cryst D. 2010 Apr 1;66(4):486–501. doi:10.1107/S0907444910007493

45. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Cryst D. 2011 Apr 1;67(4):355–67. doi:10.1107/S0907444911001314

46. Dialpuri JS, Bagdonas H, Atanasova M, Schofield LC, Hekkelman ML, Joosten RP, et al. Analysis and validation of overall N-glycan conformation in Privateer. Acta Cryst D. 2023 Jun 1;79(6):462–72. doi:10.1107/S2059798323003510

47. Ran Q, Zhang D, Jiang W, Zhang H, Cheng W, Li H, et al. How carbohydrate-binding module affects the catalytic properties of endoglucanase. International Journal of Biological Macromolecules. 2024 Oct 1;278:134653. doi:10.1016/j.ijbiomac.2024.134653

48. Zhou Y, Kommedal EG, Forsberg Z, Vaaje-Kolstad G, Suginta W, Eijsink VGH. Functional Characterization of Multidomain LPMOs from Marine Vibrio Species Reveals Modulation of Enzyme Activity by Domain–Domain Interactions. Biochemistry. 2026 Jan 6;65(1):90–103. doi:10.1021/acs.biochem.5c00529

49. Hammel M, Fierobe HP, Czjzek M, Finet S, Receveur-Bréchot V. Structural Insights into the Mechanism of Formation of Cellulosomes Probed by Small Angle X-ray Scattering*. Journal of Biological Chemistry. 2004 Dec 31;279(53):55985–94. doi:10.1074/jbc.M408979200

50. Currie MA, Cameron K, Dias FMV, Spencer HL, Bayer EA, Fontes CMGA, et al. Small angle X-ray scattering analysis of Clostridium thermocellum cellulosome N-terminal complexes reveals a highly dynamic structure. J Biol Chem. 2013 Mar 15;288(11):7978–85. doi:10.1074/jbc.M112.408757 PubMed PMID: 23341454; PubMed Central PMCID: PMC3597834.

51. Hammel M, Fierobe HP, Czjzek M, Kurkal V, Smith JC, Bayer EA, et al. Structural Basis of Cellulosome Efficiency Explored by Small Angle X-ray Scattering*. Journal of Biological Chemistry. 2005 Nov 18;280(46):38562–8. doi:10.1074/jbc.M503168200

52. Zheng F, Ding S. Processivity and Enzymatic Mode of a Glycoside Hydrolase Family 5 Endoglucanase from Volvariella volvacea | Applied and Environmental Microbiology. Applied and Environmental Microbiology [Internet]. 2013 [cited 2026 Aug 30];79(3). Available from: https://journals.asm.org/doi/10.1128/aem.02725-12

53. Herrera C, Guillén D, Manterola AM, Rizo J, Sanchez S, Rodríguez-Sanoja R. More is not always better. Multiple carbohydrate-binding domains for efficient starch hydrolysis. Food Bioscience. 2024 Dec 1;62:105306. doi:10.1016/j.fbio.2024.105306

54. Inoue H, Kishishita S, Kumagai A, Kataoka M, Fujii T, Ishikawa K. Contribution of a family 1 carbohydrate-binding module in thermostable glycoside hydrolase 10 xylanase from Talaromyces cellulolyticus toward synergistic enzymatic hydrolysis of lignocellulose. Biotechnol Biofuels. 2015 May 13;8(1):77. doi:10.1186/s13068-015-0259-2

55. Abbott DW, Ficko-Blean E, van Bueren AL, Rogowski A, Cartmell A, Coutinho PM, et al. Analysis of the Structural and Functional Diversity of Plant Cell Wall Specific Family 6 Carbohydrate Binding Modules. Biochemistry. 2009 Sep 29;48(43):10395–404. doi:10.1021/bi9013424

56. Klimek-Ochab M, Lejczak B, Forlani G. A metal-independent hydrolase from a Penicillium oxalicum strain able to use phosphonoacetic acid as the only phosphorus source. FEMS Microbiol Lett. 2003 May 1;222(2):205–9. doi:10.1016/S0378-1097(03)00301-X

57. Jordan DB, Lee CC, Wagschal K, Braker JD. Activation of a GH43 β-xylosidase by divalent metal cations: Slow binding of divalent metal and high substrate specificity. Archives of Biochemistry and Biophysics. 2013 May 1;533(1):79–87. doi:10.1016/j.abb.2013.02.020

58. Gucwa M, Lenkiewicz J, Zheng H, Cymborowski M, Cooper DR, Murzyn K, et al. CMM—An enhanced platform for interactive validation of metal binding sites. Protein Science. 2023;32(1):e4525. doi:10.1002/pro.4525

59. Zheng H, Chordia MD, Cooper DR, Chruszcz M, Müller P, Sheldrick GM, et al. Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. Nat Protoc. 2014 Jan;9(1):156–70. doi:10.1038/nprot.2013.172

60. Chagas RS, Otsuka FAM, Pineda MAR, Salinas RK, Marana SR. Mechanism of imidazole inhibition of a GH1 β-glucosidase. FEBS Open Bio. 2023;13(5):912–25. doi:10.1002/2211-5463.13595

61. Pengthaisong S, Piniello B, Davies GJ, Rovira C, Ketudat Cairns JR. Reaction Mechanism of Glycoside Hydrolase Family 116 Utilizes Perpendicular Protonation. ACS Catal. 2023 Apr 14;13(9):5850–63. doi:10.1021/acscatal.3c00620 PubMed PMID: 37180965; PubMed Central PMCID: PMC10167657.

62. Keys AM, Kastner DW, Kiessling LL, Kulik HJ. CH−π interactions confer orientational flexibility in protein–carbohydrate binding sites. J Biol Chem. 2025 Jun 14;301(8):110379. doi:10.1016/j.jbc.2025.110379 PubMed PMID: 40523618; PubMed Central PMCID: PMC12309604.

63. Nakamichi Y, Watanabe M, Fujii T, Inoue H, Morita T. Crystal structure of reducing-end xylose-releasing exoxylanase in subfamily 7 of glycoside hydrolase family 30. Proteins: Structure, Function, and Bioinformatics. 2023;91(9):1341–50. doi:10.1002/prot.26505

64. Urbániková Ľ, Vršanská M, Mørkeberg Krogh KBR, Hoff T, Biely P. Structural basis for substrate recognition by Erwinia chrysanthemi GH30 glucuronoxylanase. The FEBS Journal. 2011;278(12):2105–16. doi:10.1111/j.1742-4658.2011.08127.x

Supplementary Files

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

- SupplementaryInformation.docx