FEBS Letters

FEBS PRESS science publishing by scientists

RESEARCH ARTICLE

The first crystal structure of a xylobiose-bound xylobiohydrolase with high functional specificity from the bacterial glycoside hydrolase family 30, subfamily 10

Franz J. St John[1], Casey Crooks[1], Youngchang Kim[2], Kemin Tan[2] and Andrzej Joachimiak[2,3]

1 Institute for Microbial and Biochemical Technology, Forest Products Laboratory, USDA Forest Service, Madison, WI, USA
2 Structural Biology Center, X-ray Science Division, Argonne National Laboratory, Lemont, IL, USA
3 Department of Biochemistry and Molecular Biology, University of Chicago, IL, USA

Correspondence
F. J. St John, Forest Products Laboratory, One Gifford Pinchot Dr., Madison, WI 53726, USA
Tel: +1 608 231 9438
E-mail: fjstjohn@gmail.com; franz.j.st.john@usda.gov

(Received 25 May 2022, revised 5 July 2022, accepted 9 July 2022, available online 4 August 2022)

doi:10.1002/1873-3468.14454

Edited by Christian Griesinger

Xylobiose is a prebiotic sugar that has applications in functional foods. This report describes the first X-ray crystallographic structure models of apo and xylobiose-bound forms of a xylobiohydrolase (XBH) from *Acetivibrio clariflavus*. This xylan-active enzyme, a member of the recently described glycoside hydrolase family 30 (GH30), subfamily 10, phylogenetic clade has been shown to strictly release xylobiose as its primary hydrolysis product. Inspection of the apo structure reveals a glycone region X_2-binding slot. When X_2 binds, the non-reducing xylose in the −2 subsite is highly coordinated with numerous hydrogen bond contacts while contacts in the −1 subsite mostly reflect interactions typical for GH30 and enzymes in clan A of the carbohydrate-active enzymes database (CAZy). This structure provides an explanation for the high functional specificity of this new bacterial GH30 XBH subfamily.

Keywords: *Acetivibrio clariflavus*; cellulosome; GH30_10; glycoside hydrolase; prebiotic; xylobiohydrolase

Several enzyme subfamilies categorized as members of glycoside hydrolase (GH) family 30 (GH30), group 2 show specificities in the hydrolysis of hemicellulosic polysaccharides [1,2]. Both GH30 subfamilies 7 and 8 (GH30_7 and GH30_8) hydrolyse the main-chain β-1,4-linkage of xylan but with different modes of action [2–7]. GH30_8 enzymes are almost completely of bacterial origin and are shown to be glucuronoxylanases (GXNs). These enzymes hydrolyse the β-1,4 linkage of glucuronoxylan directed by a α-1,2-linked 4-O-methyl-glucuronic acid (or α-1,2-glucuronic acid, both synonymous in this work and abbreviated GA) xylan main-chain appendage [8,9]. A small subset of enzymes within the GH30_8 subfamily have a high amino acid identity to the canonical functioning GXNs; nevertheless, they lack the GA recognition specificity [10,11]. These unique bacterial GH30_8 endoxylanases are missing functional GA specificity due to an altered amino acid sequence in the typically conserved β8-α8 loop. This loop region is prominently involved in the GA coordination which is required for proper substrate association for subsequent hydrolysis.

In contrast to the GH30_8 GXNs, GH30_7 enzymes are primarily fungal in origin and as not uncommon, this phylogenetic clade also includes members of the bacteria phylum Actinobacteria. Members of subfamily 7 are now shown to represent several functionally distinct modes of action in their hydrolysis of the β-

Abbreviations
CAZy, carbohydrate-active enzymes database; CD, catalytic domain; GA, α-1,2-linked 4-O-methyl-glucuronic acid or α-1,2-glucuronic acid; GH, glycoside hydrolase; GH30, glycoside hydrolase family 30; GH30_X, X = GH30 subfamily designation; GXN, glucuronoxylanase; $U^{4m2}X$, 2^2-(4-*O*-methyl-α-D-glucuronosyl)-xylobiose; X_2, xylobiose; XBH, xylobiohydrolase; $XU^{4m2}X$, 2^2-(4-*O*-methyl-α-D-glucuronosyl)-xylotriose.

1,4-xylan bond. These xylan-centric functions appear remarkably varied for enzymes within the same phylogenetic subfamily. This indicates the ability of an evolved endo-functioning substrate-binding cleft to be repurposed with minor amino acid alterations. These xylan-specific properties include: (a) GXN activity [4] that functions similarly to the canonical bacterial GH30_8 GXNs [3,8]; (b) bifunctional GXNs with an additional secondary non-reducing-terminal xylobiohydrolase (XBH) activity [12–15]; (c) a primary XBH [16], a reducing-end xylose-releasing exoxylanase (REX) [5,17,18]; and (d) several enzymes with activities that have been described broadly as endoxylanases [19–21]. Two other GH30_7s have also been characterized biochemically, but to a lesser extent of detail [6,22]. One of these, xylanase D (XylD) from *Bispora* sp. MEY-1, is described as yielding xylose as its primary product which suggests that it might function as a REX [6]. The other, xylanase C (XynC) from *Penicillium purpurogenum*, makes xylobiose as its major product, suggesting that it might function as either a XBH or endoxylanase [22]. Given that these last two enzymes which are reported to have different functions share ca. 55% sequence identity (ID), the phylogenetic and functional incongruence observed in GH30 subfamily 7 becomes evident. An all-against-all pairwise sequence comparison of all the enzymes described above results in an average ID of ca. 44%, confirming that these enzymes constitute a single-enzyme subfamily with diversified enzymatic functions. The singular unifying aspect of GH30_7s is that they all hydrolyse the β-1,4 linkage of xylan resulting in different products.

The bifunctional GH30_7 GXN/XBH xylanases described above each appears to have a primary GXN function and a secondary XBH activity. These combined functions are complementary and allow a single enzyme to release most of a glucuronoxylan chain as xylobiose (X_2) with lesser amounts of aldouronates. However, unless shown otherwise, it can be anticipated that these multifunctional xylan-active enzymes have reduced functional specificity and hydrolysis rate, which may hinder their use in industrial processes. The strict fungal XBH identified above derives from *Acremonium alcalophilum* (*Aa*Xyn30A) [16] and was also shown to have a minor endoxylanase activity, although a review of the publication indicates high enzyme loading was used in these experiments [23].

Recently, a bacterial GH30 XBH which does not classify as a GH30_7 has been characterized by two separate research groups [23,24]. *Ac*Xbh30A was originally identified as a primary enzyme component of the cellulosome assembly of *Acetivibrio clariflavus* (basonym: *Clostridium clariflavum*) [25]. That original report identified *Ac*Xbh30A as a GH30 enzyme and reported detection of xylanase activity. The first follow-up publication thoroughly detailed enzyme–substrate interactions and identified a secondary endoxylanase activity [24]. The second publication identified *Ac*Xbh30A as having relatively thermostable functional characteristics, described it as being a strict xylobiohydrolase when used under optimal reaction conditions and proposed its classification as a new subfamily of bacterial xylobiohydrolases: GH30_10 [23].

In this study, we report the high-resolution three-dimensional X-ray crystallographic structure of the GH30_10 XBH *Ac*Xbh30A-CD (referred to as *Ac*Xbh30A throughout this work). These findings detail the molecular contacts that establish the glycone region X_2-binding slot. The aglycone +1 subsite is unique relative to the better characterized GH30_7/8 xylan-active enzymes. Conservation in the substrate-binding cleft is analysed broadly with comparisons to structurally characterized, related GH30 subfamily homologs. Understanding the molecular function of this xylobiohydrolase will allow for the rational design of the enzyme to achieve improved or altered functional properties.

Materials and methods

Biochemicals

All biochemicals used in this work were of the necessary grade or better for their intended usage. Xylobiose was acquired from Megazyme (Bray, Ireland).

Amino acid sequence comparisons

Bioinformatic analysis was performed using protein secondary structure influenced primary amino acid sequence alignments generated using MAFFT 7 [26,27] (https://mafft.cbrc.jp/alignment/server/). Structural information was provided to MAFFT from a T-COFFEE EXPRESSO alignment [28] composed of all structurally characterized GH30 sequences including the *Ac*Xbh30A-CD structure reported here. All characterized GH30 enzymes (and others of interest) representing all subfamilies were collected and aligned using MAFFTauto with the generated Expresso alignment (derived from all GH30 structures) used for structure restraints.

Secondary structure restrained MAFFTauto alignments were transferred into MEGAX [29] where evolutionary history was inferred using the maximum-likelihood method and Le_Gascuel_2008 model [30]. The Bootstrap consensus tree inferred from 200 replicates [31] is taken to represent the evolutionary history of the taxa analysed. The percentage of replicate trees in which the associated taxa clustered

together in the Bootstrap test is shown next to the branches [31]. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with a superior log-likelihood value. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.2962)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 0.43% sites). A consensus tree was manually rooted at the branch connecting GH30 group 1 and group 2 sequences. Evolutionary analyses were conducted in MEGA X [29].

DNA cloning

The DNA expression plasmid encoding the catalytic module of *Ac*Xbh30A (*Ac*Xbh30A-CD) was recently described [23] (UniProt Accession: G8LU16). Briefly, the secretion signal peptide was truncated at the Signal-P [32] predicted signal peptidase cleavage site with the inclusion of a N-terminal methionine, and the C-terminal was altered by removal of the C-terminal type I dockerin domain and the inclusion of a His-Tag. Amino acid regions for cloning were determined through alignments with well-characterized GH30 homologs. The final expression construct consists of the amino acid sequence MASTVTVD. . ..DTESASLEHHHHHH.

Protein expression and purification

Protein expression and expression culture processing were performed as recently described [23] using a modified autoinduction method originally described by Studier [33] at 18 °C for over 30 h. Harvested cell pellets resulting from 125 mL expression cultures were processed immediately or frozen at −80 °C. Fresh or thawed cell pellets were resuspended in cell processing buffer and lysed by sonication and a cell-free extract was applied to a 1 mL HisTrap HP column (Cyvita Life Sciences, Marlborough, MA, USA) and eluted with a 20-column volume (CV) 0–500 mM imidazole gradient. The left-most peak, 'peak A', was concentrated and exchanged into 20 mM Tris base pH 7.0 containing 2 mM β-mercaptoethanol. This was then further separated on a Mono-Q column in the same buffer with a 0–500 mM, 20-CV sodium chloride gradient and again following desalting with a 0–125 mM sodium chloride gradient. The collected peak was concentrated and exchanged into 20 mM HEPES pH 8, 250 mM sodium chloride, 2 mM dithiothreitol and lastly purified using size exclusion chromatography using a Superdex 200 pg column in the same buffer. The isolated *Ac*Xbh30A was concentrated using a 10 kDa MWCO Amicon Ultra centrifugal filtration device (Sigma Millipore, Burlington, MA, USA) to 37 $mg{\cdot}mL^{-1}$ and flash cooled by dripping into liquid nitrogen. This preparation was used for kinetic analysis [23] and was sent to Argonne National Laboratory, Structural Biology Center, for crystal screening.

Protein crystallization, data collection and model building

Crystal screening was performed at Argonne National Laboratory, Advanced Protein Characterization Facility. High-throughput crystallization was performed using a Mosquito Nanoliter Liquid Handler (TTP LabTech, Concord, CA, USA). *Ac*Xbh30A was subjected to crystallization screening as protein only and with the hydrolysis product X_2 at a *Ac*Xbh30A:X_2 molar ratio of 1 : 5 using the sitting drop vapour diffusion technique in 96-well CrystalQuick plates (Greiner, Monroe, NC, USA). The crystallization screens, Index (Hampton Research, Aliso Viejo, CA, USA) and PEGs II (Molecular Dimensions, Holland, OH, USA), were used for the screening at a temperature of 16.0 °C. For each condition, 0.4 µL of *Ac*Xbh30A at 37 $mg{\cdot}mL^{-1}$ or *Ac*Xbh30A/X_2 complex and 0.4 µL of precipitant solution were mixed; the mixture was equilibrated against 140 L of the precipitant solution in each reservoir well. Cryoprotectant includes mother liquor with 25% glycerol. They were then flash cooled in LN2 and used to collect X-ray diffraction data at 100 K and a wavelength of 0.9792 A. All data acquisitions were done at the 19ID beamline of the Structural Biology Center at the Advanced Photon Source at Argonne National Laboratory using the program SBCcollect [34]. The intensities of each data set were integrated and scaled with the HKL3000 program suite (Table 1) [35]. Phasing by molecular replacement was performed using the program MOLREP [36] using a SWISS-MODEL-generated homology model [37] of *Ac*Xbh30A for the native crystal data set and then subsequently with a partially refined apo-*Ac*Xbh30A for the X_2 cocrystal dataset. Structure model refinement was accomplished with iterative cycles of restrained refinement using the program REFMAC [38] and real space refinement in COOT [39]. Final model validation was performed using MOLPROBITY [40]. Structure analysis and figure preperation of protein structure models was performed using PYMOL [41]. Two-dimensional contact map was generated using LIGPLOT+ [42].

Results and discussion

The first structure of a new GH30 subfamily

An earlier published work [23] suggested that *Ac*Xbh30A and its closest members constituted a new phylogenetically and functionally distinct GH30 subfamily of bacterial XBHs. A phylogenetic tree (Fig. 1) represents this distinction showing the known GH30 subfamilies (1–10) and highlighting those enzymes used

Table 1. Summary of crystallographic data.

Data collection statistics	*Ac*Xbh30A	*Ac*Xbh30A/ Xylobiose
Space group	$P2_1$	$P2_1$
Unit cell (A, °)	*a* = 59.87	*a* = 59.14
	b = 108.0	*b* = 107.5
	c = 66.84	*c* = 66.86
	b = 90.14	*b* = 89.99
Twinning operator	H, -K, -L	H, -K, -L
MW Da (residue)	48 832 (435)[a]	48 832 (435)[a]
Mol (AU)	2	2
Wavelength (A)	0.9792	0.9792
Resolution (A)	1.28	1.50
Number of unique reflections	211 708	125 373
R_{merge} (%)	11.0 (102.1)[b]	9.9 (94.6)[c]
Completeness (%)	96.9 (95.3)[b]	93.8 (91.6)[c]
Redundancy	3.7 (3.3)[b]	2.7 (2.6)[c]
I/(σ)	20.0 (1.05)[b]	17.7 (1.60)[c]
Solvent content (%)	0.43	0.42
Phasing		
Resolution range (A)	41.2–3.00	34.2–3.00
Correlation coefficient (%)	18.4[d]	64.0[e]
Refinement		
Resolution range (A)	41.2–1.28	34.2–1.50
Number of reflections	211 560	124 905
Completeness (%)	96.3	93.7
R_{work}/R_{free} (%)	20.2/25.4	16.5/21.4
Twinning fraction (%)	48.0	34.0
No. of atoms (protein/HETATM)	6903/752	6883/712
Bond lengths (A)	0.001	0.003
Bond angles (deg)	0.392	0.565
B-factors (A^2) (main/side chain)	14.0/16.3	14.1/15.9
Wilson B-factor (A^2)	11.8	12.30
MOLPROBITY validation		
Ramachandran outliers (%)	0.23	0.23
Ramachandran favoured (%)	96.64	96.28
Rotamer outliers (%)	0.53	0.00
Clashscore	3.84	2.73
MOLPROBITY score	1.40	1.35
PDB ID	7N6H	7N6O

[a]Not including N- and C-terminal vector-derived residues M and LEHHHHHH respectively; [b]Last resolution bin, 1.28–1.30 A; [c]Last resolution bin, 1.50–1.53 A; [d]Molecular replacement method with a model from SWISS-MODEL as a search template; [e]Molecular replacement method with a refined structural model of apo form of *Ac*Xbh30A as a search template.

for structure comparison in this report. This tree distinguishes *Ac*Xbh30A from bacterial GH30_8 endoxylanases and the functionally diverse fungal GH30_7 xylan-active enzymes. The clade in which GH30_10 places contain additional enzymes which are believed to represent other β-1,4-xylan activities. However, several have been checked for obviously related activities, and only one, the enzyme from *Gracilibacillus dipsosauri* (UniProt accession: A0A317L0T6), shows activity in the hydrolysis of the β-1,4-xylan linkage [23]. For comparison, the GH30_7 enzymes are important as having both dual-function XBH activity [5,14] and primary XBH activity (*Aa*Xyn30A) [16] that are categorized in the subfamily. Of all the compared enzymes, only the latter does not have a corresponding X-ray crystallographic protein structure available for comparison. This enzyme also was shown to generate a mixture of endo-hydrolytic products when used at high enzyme levels.

Figure S1 presents a large phylogenetic analysis of GH30 group 2 enzymes. This phylogram analysis represents a state-of-the-art depiction of the subfamily diversity of GH30 group 2 enzymes [1,2,23,43]. Most recently, GH30_10, which is being structurally characterized in this report, was biochemically characterized [23,24], and GH30_11 was identified and shown to have a galactobiohydrolase activity [1]. Importantly, the authors considered this a fungal clade, but as it is observed in numerous other instances, bacteria of the phylum Actinomycetota (basionym: Actinobacteria) [44] also populate this clade (Fig. S1). The GH30 group 2 sequence dataset selection was developed through numerous iterations of MAFFT alignment followed by NJ tree generation in which outliers were selected for subsequent BLASTp search queries, highly homologous sequences were removed and the process repeated. In this manner of manual dataset resampling, a final maximum-likelihood Bootstrap phylogram is presented (Fig. S1). These details provide insight into the diversity of this group of related GH30 enzymes.

The *Ac*Xbh30A protein expression and purification

The *Ac*Xbh30A enzyme utilized in these studies consisted of the main GH30 $(\beta/\alpha)_8 + \beta$ [45] core catalytic domain (CD) having the natively encoded N-terminal secretion signal sequence and C-terminal dockerin domains removed. This $(\beta/\alpha)_8 + \beta$ fold is common to the GH30 enzymes as well as several other glycoside hydrolase families, and for GH30 enzymes, data suggest this fold represents a minimum functional catalytic unit [2,46]. The *Ac*Xbh30A expression product required rigorous purification to resolve two solution state forms. The form used in this work and also for the previous biochemical characterization [23] is consistent with monomeric protein that yielded tight peaks by both size exclusion and Mono-Q anion-exchange chromatography. This form of *Ac*Xbh30A was purified away from an apparent higher-molecular-weight aggregate.

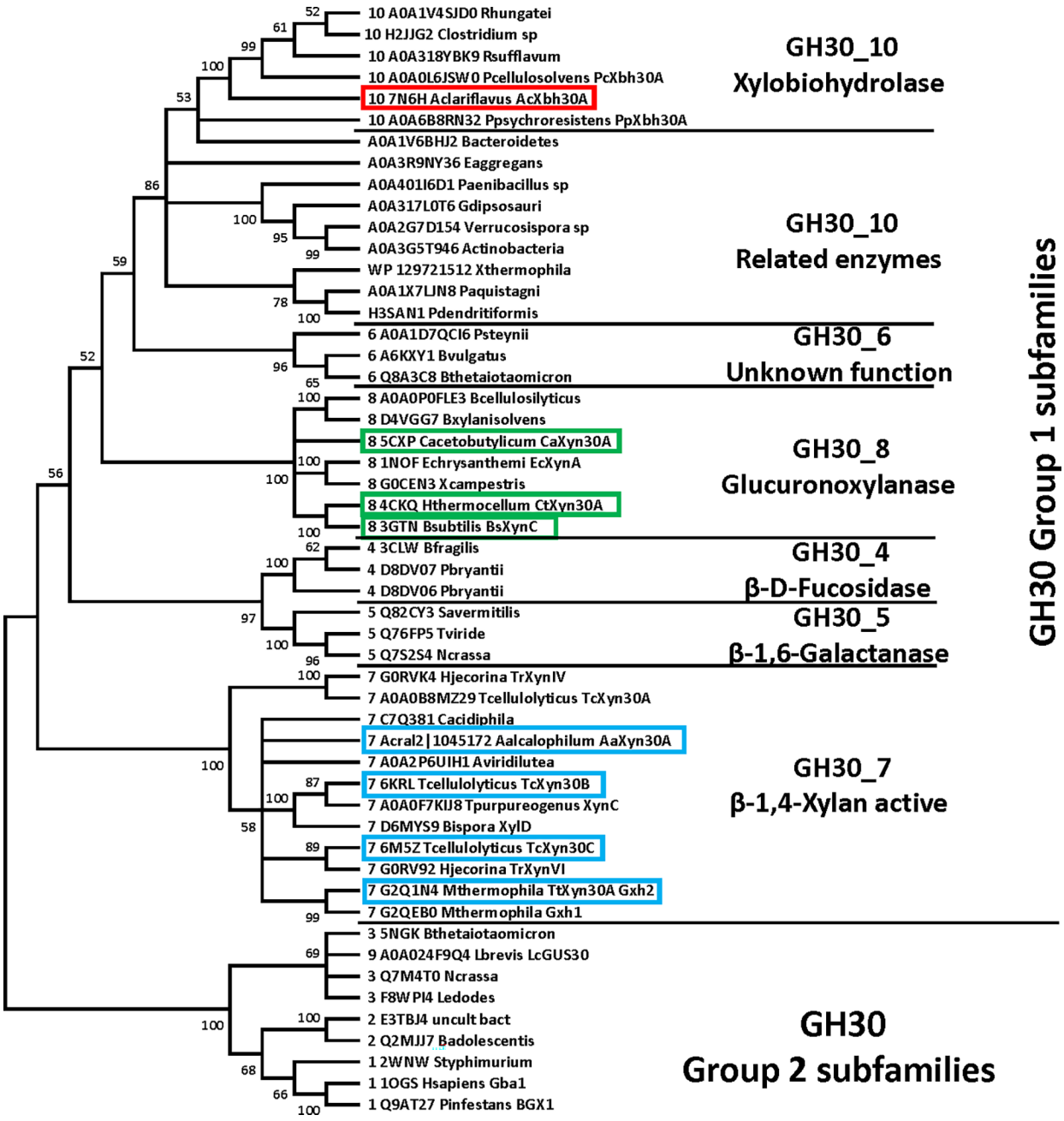

Fig. 1. A phylogram showing GH30 subfamily divisions and highlighting the structurally characterized xylan-active enzymes and others used for comparisons in this research report. For this tree, all biochemically and/or structurally characterized GH30 enzymes (along with a few uncharacterized GH30s) have been included. The alignment was performed using MAFFTauto with structure restrains provided. This Bootstrap maximum-likelihood tree was prepared using MEGAX. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. This analysis involved 52 amino acid sequences. There was a total of 807 positions in the final dataset. This Bootstrap consensus tree was manually rooted at the branch connecting GH30 group 1 and group 2 sequences. Red: GH30_10; green: GH30_8; blue GH30_7. Tip labels consist of the GH30 subfamily if known, PDB or UniProt accession number, organism name (first letter genus, species name), followed by the protein designation if available.

Crystallization, data refinement and model building

*Ac*Xbh30A was stable at high concentrations and crystal screening was performed using a preparation at 37.0 mg·mL^{-1}. *Ac*Xbh30A and *Ac*Xbh30A mixed with a 5× molar excess of X_2 resulted in large crystals grown under several crystallization conditions. Crystals of both *Ac*Xbh30A and the *Ac*Xbh30A/X_2 complex grown in the presence of 0.1 M sodium acetate pH 4.6, 25% (w/v) PEG 4000 (condition #38 of PEGS II Screen), were harvested and cryoprotected in mother liquor containing glycerol and flash cooled in liquid nitrogen. Cryoprotected crystals were used to collect X-ray diffraction data. The apo-*Ac*Xbh30A crystal diffracted to nearly atomic resolution of 1.28 A and the X_2-*Ac*Xbh30A complex diffracted to 1.50 A. Scaling these data sets resulted in overall Rmerge values of 11.0% and 9.90% respectively (Table 1). Both *Ac*Xbh30A and X_2-*Ac*Xbh30A complex crystalized into the $P2_1$ space group and were twinned with the twinning operators H, -K and -L. The asymmetric unit of each crystal form contained two molecules of *Ac*Xbh30A and contained approximately 42% solvent. Refined structure models of the apo- and X_2-bound crystal forms of *Ac*Xbh30A were 96.3% and 93.7% complete, had twinning fractions of 48.0% and 34.0% and had overall B factors of 11.8 A and 12.3 A respectively. Structure model validation using MOLPROBITY [40] showed both structures to have greater than 96% amino acids in favoured Ramachandran positions and yielded MOLPROBITY scores of 1.40 and 1.35 for the apo- and X_2-bound structures respectively (Table 1).

*Ac*Xbh30A X-ray crystallographic model overview

The protein structure model of apo-*Ac*Xbh30A (PDB accession: 7N6H) shows the expected GH30 dual-domain fold common to GH30 enzymes [2] (Fig. 2A). This family of enzymes belongs to Clan A of the Carbohydrate Active Enzymes Database [47,48] (http://www.cazy.org). GH30 enzymes are one of several CAZy Clan A enzyme families often referred to as 4/7 hydrolases because the catalytic residues, in this case, two glutamate side chains, are juxtaposed on the 4 and 7 β-sheets of the barrel structure [45]. In *Ac*Xbh30A, the catalytic nucleophile is Glu261 and the catalytic acid/base catalyst is Glu168, as shown in the centre of the $(\beta/\alpha)_8$ barrel (Fig. 2A). The nucleophile position was originally identified through enzyme mechanistic chemical inactivator studies [49] and the catalytic acid/base catalyst was predicted based on homology [50]. This $(\beta/\alpha)_8$ + β fold consists of the $(\beta/\alpha)_8$ barrel flanked by two sections of a 9-stranded anti-parallel β-sandwich structure previously described as the 'side β-structure'. Two groups of GH30 enzymes have been defined based on how the side β-structure β-sheets are split between the N-terminal and C-terminal flanks of the intervening $(\beta/\alpha)_8$ barrel [2]. *Ac*Xbh30A classifies as a GH30 group 2 enzyme having only a single strand of its side β-structure, β-strand 1 (βs1, Fig. 2B), located on the N-terminal side of the $(\beta/\alpha)_8$ barrel. Because of this secondary structure split, the interface of the side β-structure involves a dual-hinge connection between the β-strand and barrel domains (Fig. 2A,B). In addition, the two domains associate tightly through a hydrophobic interface involving the outer surface of α-helixes 7 and 8, the stability of which is thought to contribute to enzyme function [2,46]. In its overall protein fold, the refined structure model of *Ac*Xbh30A is similar to other structurally characterized GH30 enzymes.

Comparisons to other xylan-active GH30 group 2 subfamilies

*Ac*Xbh30A differs from these related enzymes primarily in the loop regions which define substrate interactions. A comparison of *Ac*Xbh30A to GH30_8 glucuronoxylanases shows that *Ac*Xbh30A is significantly altered in the β2-α2, β7-α7 and β8-α8 loop regions (Fig. 2A). With respect to these GH30 enzymes and how they bind xylan, these loop regions are important to accommodate or block the extending xylan chain in the non-reducing direction (towards the β2-α2 loop), and for offering interaction points for xylan chain substitutions on the xyloses positioned in the +1 and −2 subsites (β7-α7 and β8-α8 loops). Xyloses positioned in these two locations are proximal to the catalytic centre and are oriented to allow for O2 and/or O3 xylan chain substitution interactions to be accommodated. Structural alignment to the functionally unique (as an endoxylanase, instead of a canonical GXN) bacterial GH30_8 endoxylanase *Ca*Xyn30A [10] (Fig. 3A) reveals the openness of the bacterial endoxylanase with its minimized loops. This is anticipated considering its function and it is representative of the GH30_8 GXNs. Structure alignment with the GH30_7 endoxylanase *Tc*Xyn30C (Fig. 3B), which yields the notable limit product 2^2-(4-*O*-methyl-α-D-glucuronosyl)-xylobiose ($U^{4m2}X$) [20,21,51], shows that the β7-α7 loop is similarly sized to that of *Ac*Xbh30A. However, both the β2-α2 and β8-α8 loops are smaller. Similar to the compared *Ca*Xyn30A endoxylanase, the minimal β2-α2 loop also highlights the described endoxylanase function by allowing the extending xylan chain to exit the substrate-binding cleft, and the smaller

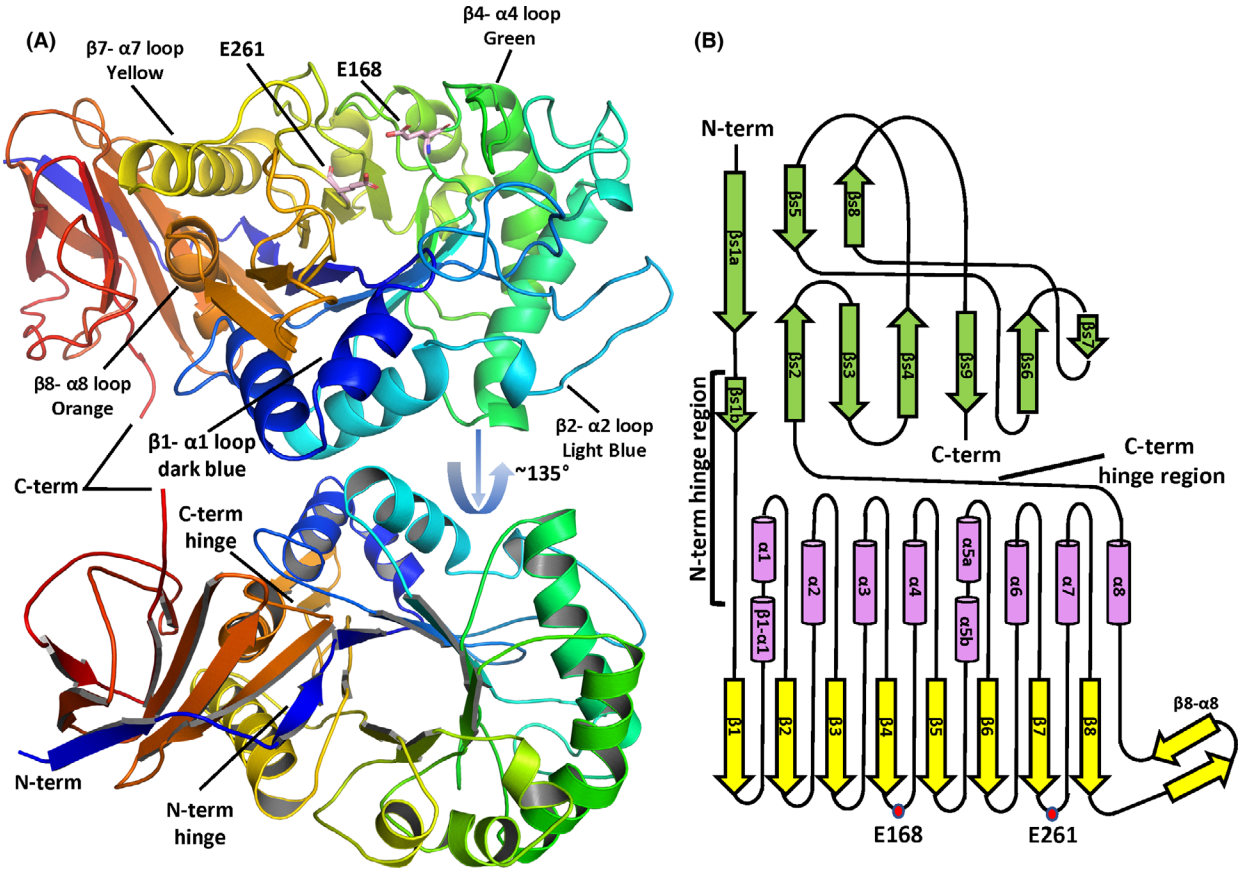

Fig. 2. General protein fold (A) of *Ac*Xbh30A (PDB accession: 7N6H) showing the top side of the catalytic centre indicated by the juxtaposed glutamic acid amino acid side chains (upper image). The amino acids Glu168 and Glu261 (shown in pink) are known to serve as the catalytic acid–base catalyst and catalytic nucleophile, respectively [49]. Several of the barrel loop regions are indicated given their significant role in enzyme function or to guide structure visualization. A rotation of ca. ~ 135° reveals (lower image) the bottom side of the $(\beta/\alpha)8 + \beta$ fold and shows the double-hinge region which connects the central catalytic $(\beta/\alpha)_8$ barrel with the side β-structure. The distribution of β-sheets in this motif is represented by the N-terminal (blue) to C-terminal (red) rainbow coloration. The double-hinge structure feature of this type occurs in several glycoside hydrolase families [2]. A secondary structure map (B) provides a two-dimensional representation of this characteristic and labels secondary structure elements important to the *Ac*Xbh30A and general GH30 protein fold. Compared to other GH30 group 2 protein structures which have a single βs1, this part of *Ac*Xbh30A is interrupted with a non-β-strand section so the naming accommodates this by designated βs1a and βs1b. The α-helical section labelled β1-α1 to indicate a α-helix-structured loop region at first glance may be considered an extension of the α1-helix. However, this specific secondary structure element is conserved in the GH30_7 but not the GH30_8 xylanases. In contrast, the two parts of the α5-helix, α5a and α5b, are observed in all GH30 group 2 structures. The α6 region is similar to that found in GH30_8 xylanases, but this same α-helix does not form in GH30_7 xylanases. Lastly, the β-hairpin structure within the β8-α8 loop is considered as such as it is not conserved in the GH30_8 xylanases. Overall, the secondary structure elements of *Ac*Xbh30A appear more like GH30_7 than GH30_8 xylanases.

β8-α8 loop makes space available for main-chain appendages. Both *Tc*Xyn30B (Fig. 3C) and *Tt*Xyn30A (Fig. 3D) are characterized as dual-function GXN/XBH enzymes and, given the XBH functional overlap, are thought to have structural aspects possibly similar to *Ac*Xbh30A. Upon inspection of the aligned structures, both *Tc*Xyn30B and *Tt*Xyn30A are clearly very similar (Fig. 4A) to each other, and both are more similar to *Ac*Xbh30A than the endoxylanases used for comparison (Fig. 3A,B). The most significant difference between these two GH30_7 dual-function xylan-active enzymes and *Ac*Xbh30A is in the β2-α2 loop. While this loop is larger than that observed in the exampled endoxylanases (Fig. 3A,B), it is still not as large when compared with *Ac*Xbh30A (Fig. 3C,D).

Glycone region X_2 coordination

From the X_2-bound structure model, xylobiose is observed within an X_2-sized slot consisting of the −1

and −2 subsites. The protein surface view (Fig. 5A) provides a visual of the walls formed from the amino acids of the surrounding loop regions in which X_2 binds. In this slot, X_2 was confidently modelled given the clear electron density as shown in Fig. 5B. The −1 subsite hydrogen bonds are common to GH30 enzymes. The xylose is positioned in the correct orientation within this subsite by the C2 and C3 hydroxyls that hydrogen bond with the δ-N of Asn167 and ε-N of Trp116, respectively (Fig. 5C,D). Also supporting the proper orientation of xylose in this position, the ε-N of Trp297 establishes a hydrogen bond with the O-linkage connecting the −1 and −2 subsite xyloses. These three hydrogen bond contacts are often observed in the GH30 family and other endo-acting CAZy Clan A 4/7 hydrolase families [52,53]. The catalytic nucleophile Glu261 ε1-O is just 2.6 A from the −1 subsite, xylose C2 hydroxyl position, and the acid/base catalyst Glu168 ε2-O is 2.2 A from the anomeric (C1) hydroxyl indicating their direct roles in hydrolysis (Fig. 5C,D).

The xylose in the −1 subsite is further positioned through several hydrophobic or van der Waals interactions. These amino acids include Tyr175, Tyr240 (Fig. 5D, not represented in Fig. 5C for clarity), Met265 and Trp297. Only Met265 which extends down into the substrate-binding cleft from the β7-α7 loop (four amino acids past the catalytic nucleophile Glu261, Fig. 5C) is not conserved among the GH30 group 2 xylan-active enzymes. In the bacterial GH30_8 endoxylanases, a side-chain hydroxyl of a conserved tyrosine occupies the same area as the δ-S of Met265. This tyrosine has been predicted to establish stabilizing hydrogen bond interactions with both the C2 hydroxyl of the GA positioned within the -2b GA-coordinating subsite [8] and/or with the C3

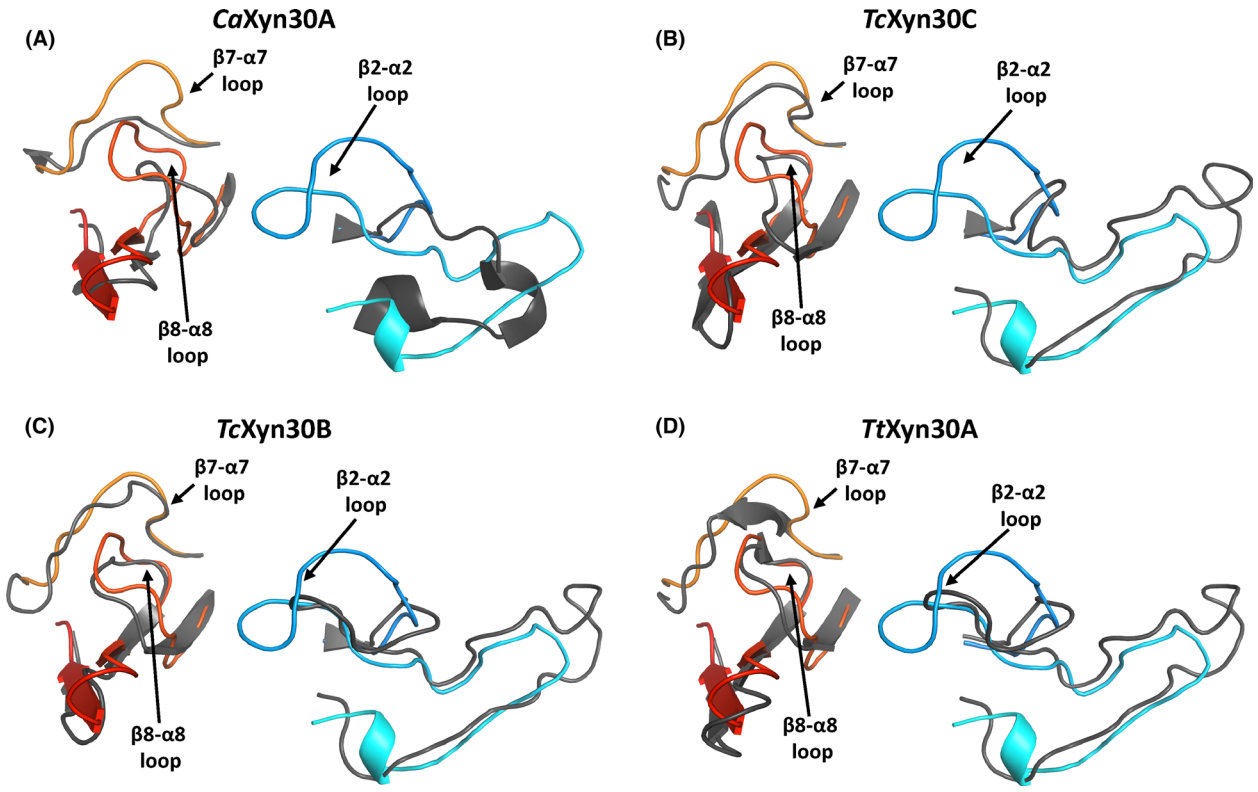

Fig. 3. Macro-structure comparison of *Ac*Xbh30A with structurally and biochemically characterized GH30_7 and 8 enzymes. A visual of just the β2-α2, β7-α7 and β8-α8 loops of *Ac*Xbh30A shown in blue, yellow and red coloration, respectively. This was overlaid with related structurally characterized GH30 xylan-active enzymes from GH30_8 (A) and GH30_7 (B–D) (all in grey). In each case, the β2-α2, β7-α7 and β8-α8 loop regions were compared to identify differential features. Comparison to the unique GH30_8 endoxylanase *Ca*Xyn30A (A, PDB accession: 5CXP) from *Clostridium acetobutylicum* shows a diminutive form of both of these loop regions. Overlay comparison with the GH30_7 endoxylanase *Tc*Xyn30C (B, PDB accession: 6M5Z) from *Talaromyces cellulolyticus* shows a still minor β2-α2 loop with a slightly larger β8-α8 loop. Most comparable the two dual-function GH30_7 xylobiohydrolase/endoxylanase functioning enzymes *Tc*Xyn30B (C, PDB accession: 6KRN) from *T. cellulolyticus* and *Tt*Xyn30A (D, PDB accession: 7O0E) from *Thermothelomyces thermophila* shows only a slightly smaller β2-α2 loop and a β8-α8 loop which, while still smaller, is more significant than the GH30_8 endoxylanase and the GH30_7 *Tc*Xyn30C.

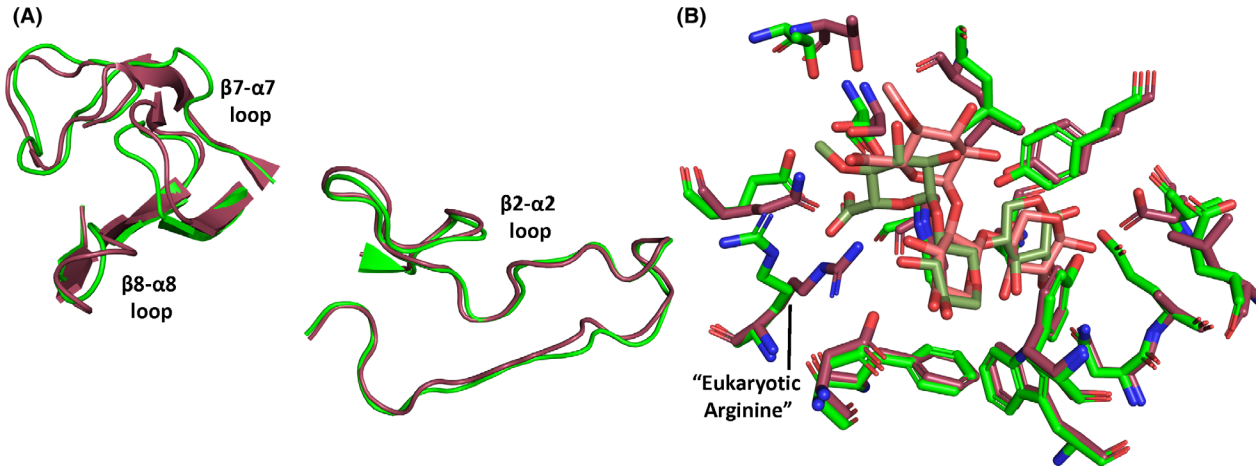

Fig. 4. Inconsistent coordination observed for the two 2^2-(4-O-methyl-α-D-glucuronosyl)-xylobiose ($U^{4m2}X$) bound GH30_7 GXN/XBH dual-function enzymes. *Tc*Xyn30B (green, PDB accession: 6KRN) and *Tt*Xyn30A (red-brown, PDB accession: 7O0E) share nearly 45% amino acid sequence identity. The β2-α2, β7-α7 and β8-α8 loop regions (A) primarily involved in ligand coordination align well for these two similarly functioning enzymes. Further analysis of the protein structure superposition of these two GH30_7 enzymes shows a number of conserved amino acid side chains in alternative positions including the 'eukaryotic arginine' (B) thought to have a role in GA coordination. Although these enzymes are thought to function similarly, the hydrogen bonding differences between these proteins and their crystallographically modelled ligand clearly shows differences in their 'specific' contacts.

hydroxyl of the +1 subsite xylose in these GXNs [10,54] (Fig. S2). This position in the GH30_7 xylan-active enzymes is more similar to *Ac*Xbh30A than the GH30_8 enzymes. In place of the methionine found in *Ac*Xbh30A, the GH30_7s have a conserved leucine (Leu288 in *Tc*Xyn30C, PDB accession: 6M5Z). The δ1-C-terminal methyl group of this leucine is in the near exact position as the δ-S of Met265 in *Ac*Xbh30A (Fig. 6A). In all three compared GH30 subfamilies, an amino acid side-chain atom occupies this position indicating that it is an important location for substrate–protein interaction.

The ε-C of this methionine in *Ac*Xbh30A is the only differently positioned atom in the substrate-binding cleft between the apo and X_2-bound structures (Fig. 6B). In the apo form, the ε-C methyl points towards the −1 subsite xylose and at just 1.9 A away would be expected to clash with the C5 position of a bound xylose. In the X_2-bound structure, the side-chain terminal methyl of Met265 flips to point towards the +1 subsite position. While in this position, the δ-S is also at a near-optimum distance of 3.7 A from the endocyclic oxygen and has an approximate θ angle of 45° to the perpendicular of the C_γ-S_δ-C_ε bond line. This arrangement may indicate an evolved specific electrostatic interaction between these two positions [55]. Structure model alignment comparisons with the aglycone-bound X_2 from the *Acetivibrio thermocellus* (basonym: *Clostridium thermocellum*) GH30_8 *Ct*Xyn30A glucuronoxylanase (see Fig. S2, PBD accession: 5A6M) [54] indicate that in this flipped position, the Met265 δ-S and ε-C of *Ac*Xbh30A may also directly interact with the aglycone +1 subsite xylose (Fig. 6C, see Aglycone region below).

Coordination in the −2 subsite defines the function of this high-specificity xylobiohydrolase. In contrast to the GXN/XBH dual-function GH30_7 and GXN GH30_8 xylan-active enzymes with less prominent β8-α8 loops that facilitate coordination of GA in a -2b subsite, the larger β8-α8 loop in *Ac*Xbh30A allows for amino acid side chains to reach the top side of the −2 subsite xylose and establish stabilizing interactions. Both Lys303 and Glu308 extend down from the β8-α8 loop and hydrogen bond with the C2 hydroxyl position (Fig. 5C,D), while Lys303 also establishes a hydrogen bond with the C3 hydroxyl position. The existence of these two amino acids undoubtedly supports higher functional specificity through their specific contacts, but also sterically blocks accommodation of substitutions on the C2 and C3 hydroxyl positions of the −2 subsite xylose. Lastly, Asp73 established two hydrogen bonds involving both its carboxyl oxygens with the C3 and C4 hydroxyl groups of the xylose. The C4 hydroxyl also establishes a hydrogen bond with coordinated water (W218, Fig. 5D). This coordinated water further hydrogen bonds with a second water (W375) which is coordinated from the main-chain peptide bond carbonyl oxygen of Asp73.

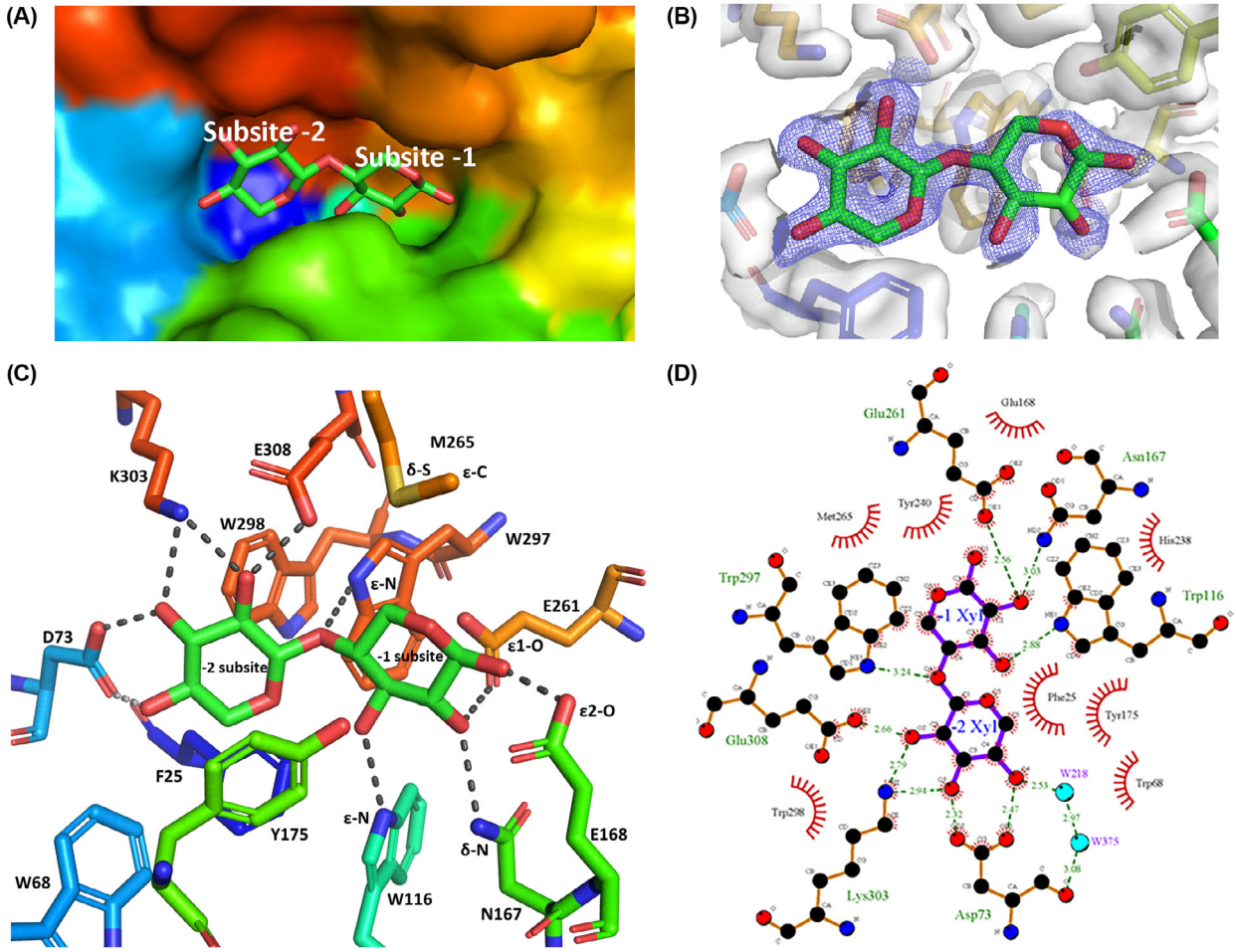

Fig. 5. Structure analysis of the glycone xylobiose-binding slot. Surface representation (A) of the *Ac*Xbh30A glycone region depicting the crystallographically determined xylobiose positioned in a slot that is formed from the β2-α2 (blue), β4-α4 (green), β7-α7 (orange) and β8-α8 (red) loops. An electron density map visual (B) of the xylobiose-binding region intended to represent the crystallographic data quality. The map around the X_2 is from a CCP4 Omit program map and the density around the amino acid side chains is from the refinement-generated 2Fo-fc map. A representation (C) of the amino acid side chains establishing polar or van der Waals contacts with the X_2 and therefore shaping the glycone xylobiose-binding slot and revealing the specificity for the non-reducing terminal xylobiose activity. Lastly, a LIGPLOT+ plot (D) [42] showing all the predicted molecular contacts involved in X_2 positioning into the glycone region.

Underscoring the functional specificity of *Ac*Xbh30A, all hydroxyls of the −2 subsite xylose benefit from two hydrogen bonds. Both of the GXN/XBH dual-function GH30_7 enzymes, *Tc*Xyn30B and *Tt*Xyn30A, contain an amino acid homolog of Asp73 located also in their respective β2-α2 loop. The α-C positions of Asn93 in *Tc*Xyn30B and Asp78 in *Tt*Xyn30A are positioned 2.9 A and 2.6 A further from the same in *Ac*Xbh30A Asp73 (Fig. 6D). This shift in position somewhat matches a shift in the modelled ligand and both establish a unique network of hydrogen bonds different from that predicted for Asp73 in *Ac*Xbh30A.

The GH30_7 XBH *Aa*Xyn30A is the enzyme that has been shown to function most similar to *Ac*Xbh30A [16]. A SWISS-MODEL [37] generated protein structure homology model of this enzyme (generated using *Tc*Xyn30B, PDB accession: 6KRN as a template and reproduced for this work) indicates that similar molecular contacts as that observed in *Ac*Xbh30A may be present. As previously speculated [16], Asp100 of this model is in the near exact position as Asp78 of *Tt*Xyn30A and therefore may be reasonably expected to serve a similar purpose (Fig. 6D). Although the *Aa*Xyn30A homology model was prepared using the dual-function GXN/XBH *Tc*Xyn30B as a template, two amino acids are seen extending down towards the X_2 and may serve a role similar to Glu308 and Lys303 in *Ac*Xbh30A. In the model of *Aa*Xyn30A, Asn356

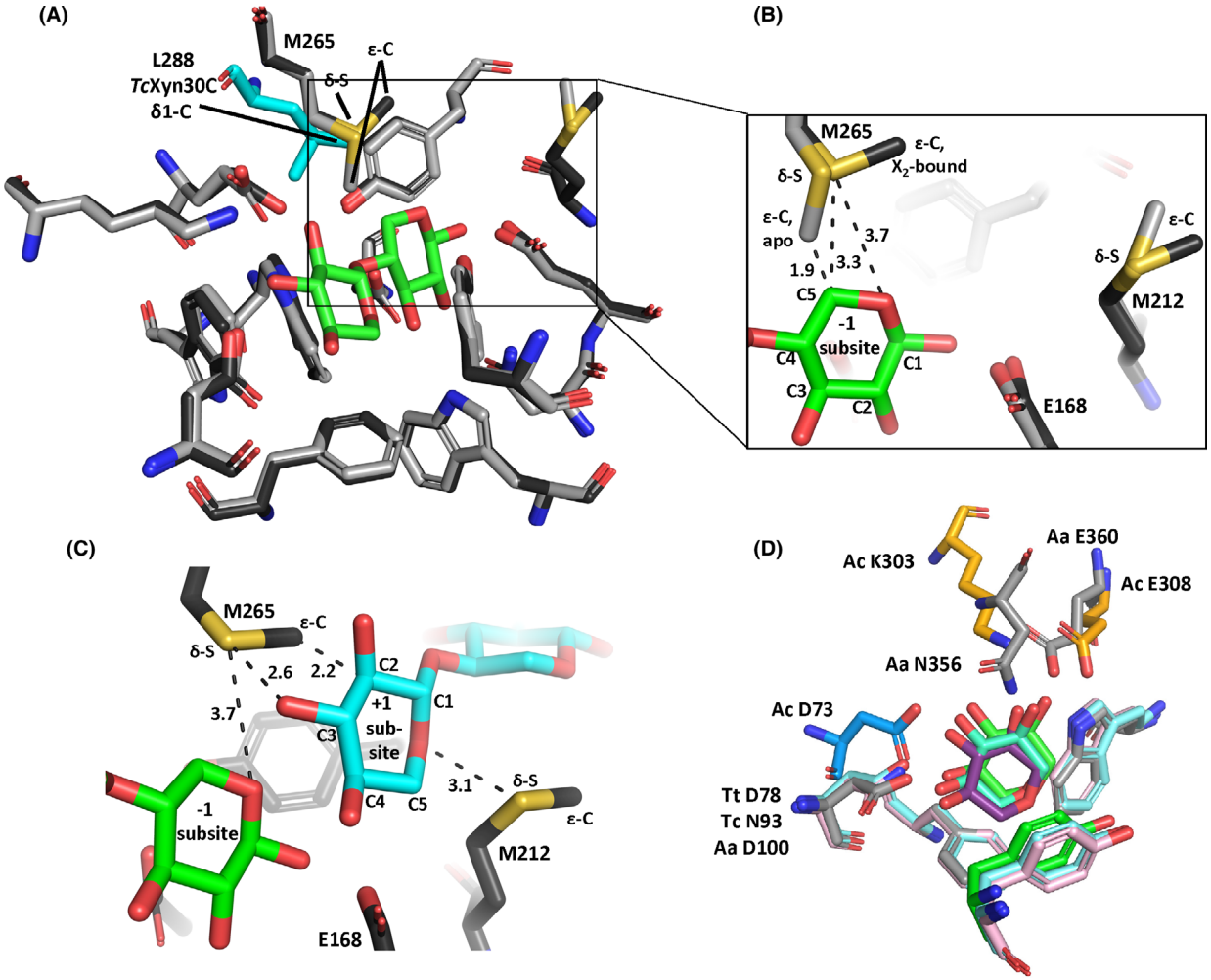

Fig. 6. Analysis of the Met265 position in *Ac*Xbh30A and comparison with GH30_7 homologs. Comparison of the position of Met265 (A) in the apo form (light grey, PDB accession: 7N6H) and X_2-bound form (dark grey, PDB accession: 7N6O). It is seen that the ε-C repositions upon substrate binding and result in an orientation that may allow interaction with the +1 subsite xylose. Also shown in the conserved leucine from GH30_7 xylan-active enzymes represented here by Leu288 from *Tc*Xyn30C (cyan, PDB accession: 6M5Z). In this, the δ1-C resides in nearly the same position as the Met265 δ-S from *Ac*Xbh30A. In the apo form, the position of ε-C of Met265 would clash with a −1 subsite xylose being just 1.9 A away from the C5 of the xylose ring (B), and while oriented towards the +1 subsite following ligand binding, it would seem likely to engage through van der Waals interactions. More importantly, δ-S in the X_2-bound form, at 3.7 A away, looks likely to interact specifically [55] with the endocyclic oxygen of the −1 subsite xylose. Further consideration of this region of *Ac*Xbh30A (C) with the aglycone X_2 modelled from a simple structure alignment with the GH30-8 glucuronoxylanase *Ct*Xyn30A (PDB accession: 5A6M, see Fig. S2) shows that a second methionine, Met212, may interact with the endocyclic oxygen of the +1 subsite xylose. Comparisons of *Ac*Xbh30A (D) (ac, rainbow) with the dual-function GH30_7 GXN/XBH, *Tc*Xyn30B (Tc, pink) and *Tt*Xyn30A (Tt, cyan) and the homology model of the GH30_7 XBH *Aa*Xyn30A (aa, grey). The position of the −2 subsite xylose is shown (*Ac*Xbh30A, green; *Tc*Xyn30B, purple, *Tt*Xyn30A, aquamarine).

and Glu360 (Fig. 6D) are in proximity to potentially establish hydrogen bonds with the C2 and C3 hydroxyls of the −2 subsite xylose. In summary, the homology model, originally analysed by Suchova et al. [16] may indicate these two XBH enzymes, from different GH30 subfamilies display similar molecular substrate interactions which govern their function. Importantly, a SWISS-MODEL of *Aa*Xyn30A using *Ac*Xbh30A as a template yielded less convincing molecular explanations.

Hydrophobic and/or van der Waals interactions that stabilize the −2 subsite xylose, thereby forming the X_2-binding slot, include Phe25, Trp68, Tyr175 and Trp298. In GH30_7 homologs, Phe25 is conserved,

but in GH30_8s, this position is represented by a conserved tryptophan. Whether a phenylalanine or tryptophan, this amino acid side chain forms the floor of the substrate-binding cleft in these enzymes. Trp68 in *Ac*Xbh30A fills a gap and provides van der Waals contact only with the C4 hydroxyl position. As described in part above, Tyr175, which also interacts with the −1-subsite xylose, is highly conserved in several related GH30 subfamilies. Its aromatic side chain establishes the outer wall of the N-terminal side of the glycone region. Lastly, Trp298 in *Ac*Xbh30A is involved in van der Waals interactions with the C3 hydroxyl. This tryptophan is unique among these related GH30 subfamilies and is also observed in GH30_7 REX enzymes [5,17]. In GH30_8 GXNs, this position is a conserved tyrosine, but in both non-canonical functioning GH30_8 endoxylanases [10,11], this position is also a tryptophan. However, the tryptophan in *Ac*Xbh30A models as a flipped rotamer with the ε1-N pointing towards the cleft bottom, opposite of that observed in the two unconventional GH30_8 endoxylanases. Quite distinctly, GH30_7 enzymes have a conserved Leu/Ile which is significantly smaller than the large aromatic represented by the side chain of tryptophan. The resulting void is filled by an adjacent amino acid from the β1-α1 loop. In the available GH30_7 X-ray crystallographic structure models, this position is represented by an arginine which has been dubbed the 'eukaryotic arginine' [56] to identify it as having an important role in GXN activity by these xylan-active enzymes. Not all GH30_7s have this arginine, and alternative interpretations have been explored [17,20,21]. Stranger yet, the two $U^{4m2}X$ GA substituted xylooligosaccharide-bound GH30_7 structures *Tc*Xyn30B (PDB accession: 6KRN) and *Tt*Xyn30A (PDB accession: 7O0E) although significantly conserved, reveal two distinct hydrogen-bonding networks (Fig. 4B) involving the eukaryotic arginine coordinating the GA.

Aglycone region

The +1 aglycone subsite in *Ac*Xbh30A is considerably different from other closely related GH30 subfamilies. Only H238 and Y240 are conserved in all the compared enzymes (Fig. 7). As alluded to above, the Met265 side-chain flip that occurs upon X_2 binding into the glycone side, as a result, likely then interacts with the +1 subsite xylose. Considering again as a visual guide, the position of the X_2 bound into the aglycone region of the GH30_8 enzyme *Ct*Xyn30A (see Figs S2 and 6C): given the static three-dimensional alignment, while the measured intermolecular distances are too small (Fig. 6C), the orientation of the xylose indicates that the δ-S of Met265 may possibly hydrogen bond with the C3 hydroxyl position of the +1 subsite xylose [57]. The ε-C may further interact through van der Waals with the C2 position of this xylose. Another methionine, Met212, located in the +1/+2 subsite region extends upwards and could specifically interact with the endocyclic oxygen of the +1 subsite xylose through a specific electrostatic interaction [55] in a manner similar to that considered above involving Met265 and the endocyclic oxygen of the −1 subsite xylose. Met265 and Met 212 are comparably unique not just among the homologous GH30 xylan-active enzymes, but also it is not common to find a methionine side chain in the active site of related glycoside hydrolases. This may be most possible in an anaerobic microbial niche which may help explain this observation.

Variations evident from multiple-sequence alignments

Overall, the phylogenetic clade defined by the confirmed XBH enzymes *Ac*Xbh30A (this work) and *Pc*Xbh30A constitutes a set of highly homologous enzymes that derive strictly from the bacterial order and family Eubacteriales *Oscillospiraceae* [23]. The boundary of this GH30_10 clade is established by the outlier *Pp*Xbh30A from the Gram-positive psychrophilic bacterium *Paenibacillus psychroresistens*, which also has been shown to have XBH activity [23]. Sequence comparisons using the statistical alignment tool PRSS show that the Eubacteriales-derived enzymes share greater than 75% amino acid sequence ID, while comparison with *Pp*Xbh30A provides an average ID of 47.8% (SD 0.80). Inspection of a multiple-sequence alignment (Fig. S3) comparing this group of sequences shows that *Pp*Xbh30A differs significantly in the active-site cleft. The amino acid sequence in the β2-α2 loop responsible for two hydrogen bonds between the −2 subsite xylose and Asp73 is a completely different sequence in *Pp*Xbh30A. Although this enzyme is known to have XBH activity, it was not thoroughly characterized so it is not clear how strict this function is relative to other XBH enzymes. While the pair of substrate-binding cleft methionines are perfectly conserved in most of the Eubacteriales enzyme and only one is missing in *Pp*Xbh30A, the highly conserved XBH with UniProt Accession: A0A4U7JJW1 has both methionine residues changed to alternative amino acids. Fig. S3 highlights key amino acid positions that from this report are considered important for XBH activity.

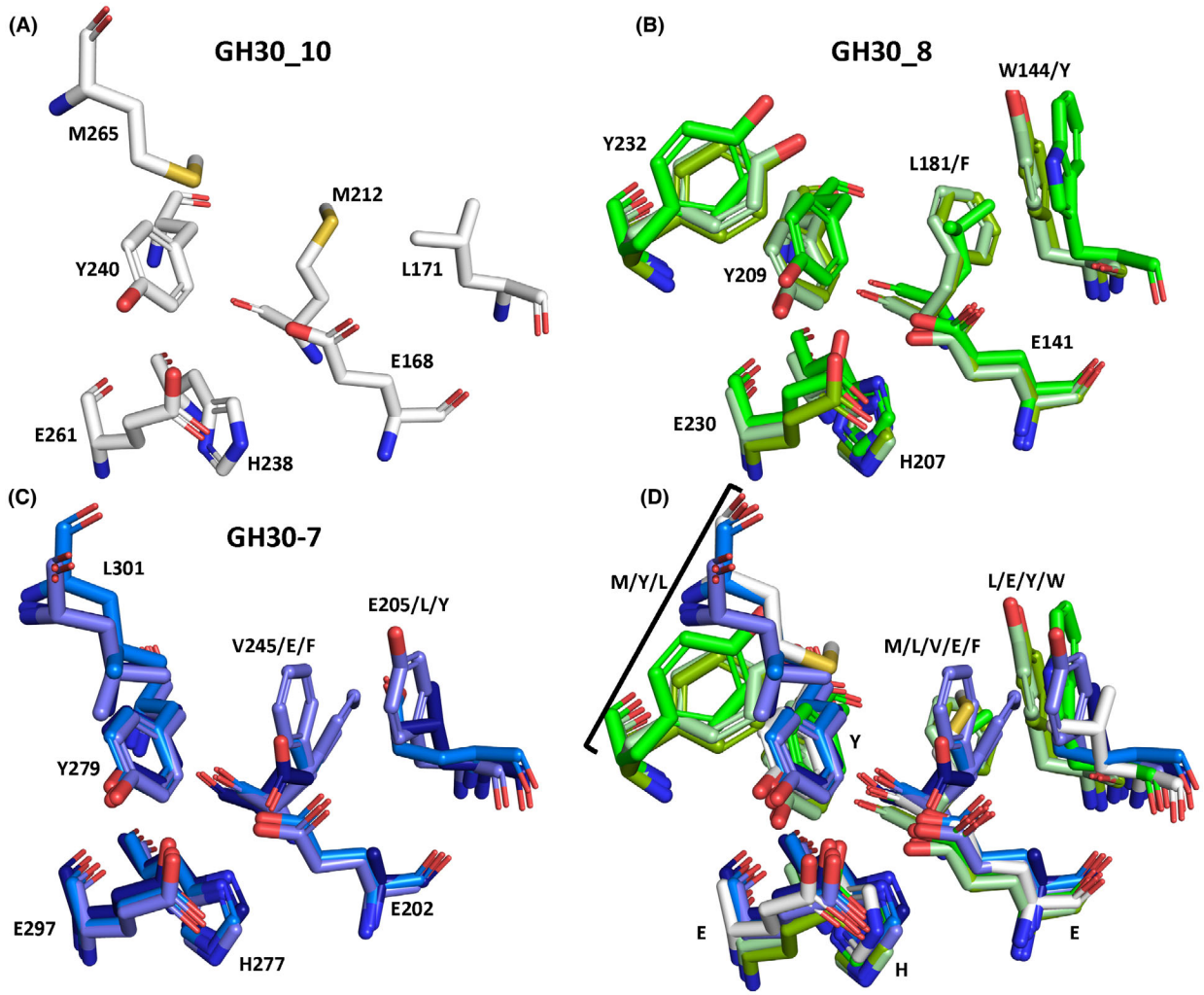

Fig. 7. Comparison of GH30 subfamily 7, 8 and 10 substrate-binding cleft aglycone regions. The aglycone region of *Ac*Xbh30A (A; PDB accession 7N6O, white) showing for perspective, the two catalytic glutamates and Met265 which are thought to interact with both the −1 (glycone) and +1 (aglycone) subsite xylose residues. For comparison, the aglycone region of three GH30_8 and three GH30_7 structurally characterized enzymes is shown (B, C). The GH30_8 enzymes (B) include the non-canonical functioning, GA-independent endoxylanase *Ca*Xyn30A (PDB accession: 5CXP; green) and the two canonical functioning GA appendage-dependent endoxylanases *Bs*XynC (PDB accession: 3KL5; olive) and *Ct*Xyn30A (PDB accession: 5A6M; light green). Amino acid labels correspond to the *Ca*Xyn30A structure and where not fully conserved, following a forward slash other representative amino acids are listed. For the GH30_7 enzymes (C), *Tc*Xyn30B (PDB accession: 6KRN; blue), *Tc*Xyn30C (PDB accession: 6M5Z; purple) and *Tt*Xyn30A (PDB accession: 7O0E; dark blue) are shown. Amino acid labels correspond to *Tc*Xyn30B and where not fully conserved, following a forward slash other representative amino acids are listed. Overlaying all structures from A to C (D), the positions are labelled with the list of amino acids represented in that group. Significant diversity exists in the aglycone region of these GH30 group 2 xylan-active enzymes.

Conclusion

This research report details the structural attributes of the first GH30_10 'strict' XBH enzyme from the bacterial Family *Oscillospiraceae*. The high-resolution structure of *Ac*Xbh30A is compared with the other described GH30 xylan-active enzymes from subfamilies 7 and 8 and the report further investigates the phylogeny of GH30 enzymes. Only one other enzyme, a GH30-7 from the fungus *A. alcalophilum* (*Aa*Xyn30A) [16] may have a similar function. No crystal structure has yet been reported for this GH30_7 enzyme. Other GH30_7 dual-function GXN/XBH enzymes have been structurally characterized [12–15] and these enzymes may have unique applications, but the multiple functions indicate a potential lack of

specificity. This may make their use in industrial applications inferior to an XBH which shows high specificity. Given the XBH enzymatic properties and thermostability demonstrated by this enzyme [23], *Ac*Xbh30A may be useful in a biorefinery-based process for the production of X_2. Inclusion of the high specificity, appendage-dependent GH30-8 glucuronoxylanases in the hydrolysis of a glucuronoxylan would additionally yield the two aldouronates, $U^{4m2}X$ and $XU^{4m2}X$ [2^2-(4-*O*-methyl-α-D-glucuronosyl)-xylotriose]. When coupled to other high-specificity appendage-dependent endoxylanases such as the GH5 arabinoxylanases [58,59] in the hydrolysis of arabinofuranose-substituted xylans, a number of other novel oligomeric xylan-derived sugars of value may be produced.

Acknowledgements

We would like to thank the Institute for Microbial and Biochemical Technology (IMBT) of the Forest Products Laboratory (FPL), a National Laboratory of the USDA Forest Service for ongoing financial support, and USDA research scientists Dan Cullen and J. Y. Zhu for providing valuable internal review and comments for this manuscript prior to submission. The Argonne National Laboratory, Structural Biology Center, beamlines are supported by U.S. Department of Energy, Office of Biological and Environmental Research, under contract DE-AC02-06CH11357.

Author contributions

FSJ and CC conceived of the research project. AJ, FSJ and CC provided support and funding for research. CC performed gene cloning and protein expression work while FSJ purified the enzyme for crystallization. KT crystalized the protein, collected x-ray diffraction data and built and refined the protein structure model with some data processing and phasing contribution from YK. FSJ analyzed the protein structure, prepared the primary manuscript, and made all correction. FSJ, CC, KT and AJ reviewed and edited the manuscript.

Data accessibility

The structural data that support these findings are openly available through the Research Collaboratory for Structural Bioinformatics Protein Data Bank at https://www.rcsb.org searchable using their accession numbers 7N6H and 7N6O for the apo- and xylobiose-bound structures respectively.

References

1 Li X, Kouzounis D, Kabel MA, de Vries RP, Dilokpimol A. Glycoside hydrolase family 30 harbors fungal subfamilies with distinct polysaccharide specificities. *New Biotechnol*. 2022;**67**:32–41.
2 St John FJ, Gonzalez JM, Pozharski E. Consolidation of glycosyl hydrolase family 30: a dual domain 4/7 hydrolase family consisting of two structurally distinct groups. *FEBS Lett*. 2010;**584**:4435–41.
3 St John FJ, Rice JD, Preston JF. Characterization of XynC from *Bacillus subtilis* subsp. subtilis strain 168 and analysis of its role in depolymerization of glucuronoxylan. *J Bacteriol*. 2006;**188**:8617–26.
4 Biely P, Puchart V, Stringer MA, Krogh KB. *Trichoderma reesei* XYN VI–a novel appendage-dependent eukaryotic glucuronoxylan hydrolase. *FEBS J*. 2014;**281**:3894–903.
5 Tenkanen M, Vrsanska M, Siika-aho M, Wong DW, Puchart V, Penttilä M, et al. Xylanase XYN IV from *Trichoderma reesei* showing exo-and endo-xylanase activity. *FEBS J*. 2013;**280**:285–301.
6 Luo HY, Yang J, Li J, Shi PJ, Huang HQ, Bai YG, et al. Molecular cloning and characterization of the novel acidic xylanase XYLD from Bispora sp MEY-1 that is homologous to family 30 glycosyl hydrolases. *Appl Microbiol Biotechnol*. 2010;**86**:1829–39.
7 Maehara T, Yagi H, Sato T, Ohnishi-Kameyama M, Fujimoto Z, Kamino K, et al. GH30 glucuronoxylan-specific xylanase from *Streptomyces turgidiscabies* C56. *Appl Environ Microbiol*. 2018;**84**:e01850-17.
8 St John FJ, Hurlbert JC, Rice JD, Preston JF, Pozharski E. Ligand bound structures of a glycosyl hydrolase family 30 glucuronoxylan xylanohydrolase. *J Mol Biol*. 2011;**407**:92–109.
9 Urbanikova L, Vrsanska M, Morkeberg Krogh KB, Hoff T, Biely P. Structural basis for substrate recognition by *Erwinia chrysanthemi* GH30 glucuronoxylanase. *FEBS J*. 2011;**278**:2105–16.
10 St John FJ, Dietrich D, Crooks C, Balogun P, de Serrano V, Pozharski E, et al. A plasmid borne, functionally novel glycoside hydrolase family 30 subfamily 8 endoxylanase from solventogenic clostridium. *Biochem J*. 2018;**475**:1533–51.
11 St John FJ, Dietrich D, Crooks C, Pozharski E, Gonzalez JM, Bales E, et al. Novel member of glycoside hydrolase family 30 subfamily 8 with altered substrate specificity. *Acta Crystallogr D Biol Crystallogr*. 2014;**70**:2950–8.
12 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. *Carbohydr Polym*. 2021;**273**:118553.

13 Nakamichi Y, Watanabe M, Matsushika A, Inoue H. Substrate recognition by a bifunctional GH30-7 xylanase B from *Talaromyces cellulolyticus*. *FEBS Open Bio*. 2020;**10**:1180–9.
14 Nakamichi Y, Fouquet T, Ito S, Watanabe M, Matsushika A, Inoue H. Structural and functional characterization of a bifunctional GH30-7 xylanase B from the filamentous fungus *Talaromyces cellulolyticus*. *J Biol Chem*. 2019;**294**:4065–78.
15 Katsimpouras C, Dedes G, Thomaidis NS, Topakas E. A novel fungal GH30 xylanase with xylobiohydrolase auxiliary activity. *Biotechnol Biofuels*. 2019;**12**:120.
16 Suchova K, Puchart V, Spodsberg N, Krogh KBM, Biely P. A novel GH30 xylobiohydrolase from *Acremonium alcalophilum* releasing xylobiose from the non-reducing end. *Enzyme Microb Technol*. 2020;**134**:109484.
17 Nakamichi Y, Fouquet T, Ito S, Matsushika A, Inoue H. Mode of action of GH30-7 reducing-end xylose-releasing exoxylanase a (Xyn30A) from the filamentous fungus *Talaromyces cellulolyticus*. *Appl Environ Microbiol*. 2019;**85**:e00552-19.
18 Tenkanen M, Saloheimo M, Biely P, Siika-aho M, Vrsanska M, Keskuslaboratorio O. A novel xylanase XYN IV from trichoderma reesei rut C30 showing unique exo-action. *Abstr Pap Am Chem Soc*. 2002;**223**: U125.
19 Suchova K, Spodsberg N, Mørkeberg Krogh KB, Biely P, Puchart V. Non-specific GH30_7 Endo-β-1, 4-xylanase from *Talaromyces leycettanus*. *Molecules*. 2021;**26**:4614.
20 Nakamichi Y, Fujii T, Watanabe M, Matsushika A, Inoue H. Crystal structure of GH30-7 endoxylanase C from the filamentous fungus *Talaromyces cellulolyticus*. *Acta Crystallogr F Struct Biol Commun*. 2020;**76**:341–9.
21 Nakamichi Y, Fujii T, Fouquet T, Matsushika A, Inoue H. GH30-7 endoxylanase C from the filamentous fungus *Talaromyces cellulolyticus*. *Appl Environ Microbiol*. 2019;**85**:e01442-19.
22 Espinoza K, Eyzaguirre J. Identification, heterologous expression and characterization of a novel glycoside hydrolase family 30 xylanase from the fungus *Penicillium purpurogenum*. *Carbohydr Res*. 2018;**468**: 45–50.
23 Crooks C, Bechle NJ, St John FJ. A new subfamily of glycoside hydrolase family 30 with strict Xylobiohydrolase function. *Front Mol Biosci*. 2021;**8**:714238.
24 Suchova K, Puchart V, Biely P. A novel bacterial GH30 xylobiohydrolase from *Hungateiclostridium clariflavum*. *Appl Microbiol Biotechnol*. 2020;**1-11**:185–95.
25 Artzi L, Morag E, Barak Y, Lamed R, Bayer EA. *Clostridium clariflavum*: key cellulosome players are revealed by proteomic analysis. *MBio*. 2015;**6**: e00411-15.
26 Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;**30**: 772–80.
27 Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform*. 2008;**9**:286–98.
28 Armougom F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, et al. Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-coffee. *Nucleic Acids Res*. 2006;**34**:W604–8.
29 Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;**35**:1547–9.
30 Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol*. 2008;**25**:1307–20.
31 Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985;**39**: 783–91.
32 Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019;**37**:420–3.
33 Studier FW. Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif*. 2005;**41**:207–34.
34 Rosenbaum G, Alkire RW, Evans G, Rotella FJ, Lazarski K, Zhang RG, et al. The structural biology center 19ID undulator beamline: facility specifications and protein crystallographic results. *J Synchrotron Radiat*. 2006;**13**:30–45.
35 Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr*. 2006;**62**:859–66.
36 Vagin A, Teplyakov A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr*. 2010;**66**:22–5.
37 Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;**46**:W296–303.
38 Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr*. 2011;**67**:355–67.
39 Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 2010;**66**:486–501.

40 Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010;**66**:12–21.
41 DeLano WL. The PyMOL User's manual DeLano scientific. Santa Clara, CA: Palo Alto; 2002.
42 Laskowski RA, Swindells MB. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. *J Chem Inf Model*. 2011;**51**:2778–86.
43 Sakurama H, Kishino S, Uchibori Y, Yonejima Y, Ashida H, Kita K, et al. β-Glucuronidase from lactobacillus brevis useful for baicalin hydrolysis belongs to glycoside hydrolase family 30. *Appl Microbiol Biotechnol*. 2013;**1-12**:4021–32.
44 Oren A, Garrity GM. Valid publication of the names of forty-two phyla of prokaryotes. *Int J Syst Evol Microbiol*. 2021;**71**:005056.
45 Nagano N, Porter CT, Thornton JM. The (βα) 8 glycosidases: sequence and structure analyses suggest distant evolutionary relationships. *Protein Eng*. 2001;**14**:845–55.
46 Valenzuela SV, Diaz P, Pastor FJ. Modular glucuronoxylan-specific xylanase with a family CBM35 carbohydrate-binding module. *Appl Environ Microbiol*. 2012;**78**:3923–31.
47 Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon JP, Davies G. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci USA*. 1995;**92**:7090–4.
48 Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 2014;**42**:D490–5.
49 Miao S, McCarter JD, Grace ME, Grabowski GA, Aebersold R, Withers SG. Identification of Glu340 as the active-site nucleophile in human glucocerebrosidase by use of electrospray tandem mass spectrometry. *J Biol Chem*. 1994;**269**:10975–8.
50 Durand P, Lehn P, Callebaut I, Fabrega S, Henrissat B, Mornon J-P. Active-site motifs of lysosomal acid hydrolases: invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. *Glycobiology*. 1997;**7**:277–84.
51 Faure R, Courtin CM, Delcour JA, Dumon C, Faulds CB, Fincher GB, et al. A brief and informationally rich naming system for oligosaccharide motifs of heteroxylans found in plant cell walls. *Aust J Chem*. 2009;**62**:533–7.
52 St John FJ, Preston JF, Pozharski E. Novel structural features of xylanase A1 from *Paenibacillus* sp. JDR-2. *J Struct Biol*. 2012;**180**:303–11.
53 Kim H-W, Ishikawa K. Functional analysis of hyperthermophilic endocellulase from *Pyrococcus horikoshii* by crystallographic snapshots. *Biochem J*. 2011;**437**:223–30.
54 Freire F, Verma A, Bule P, Alves VD, Fontes CM, Goyal A, et al. Conservation in the mechanism of glucuronoxylan hydrolysis revealed by the structure of glucuronoxylan xylanohydrolase (CtXyn30A) from *Clostridium thermocellum*. *Acta Crystallogr D Biol Crystallogr*. 2016;**72**:1162–73.
55 Pal D, Chakrabarti P. Non-hydrogen bond interactions involving the methionine sulfur atom. *J Biomol Struct Dyn*. 2001;**19**:115–28.
56 Puchart V, Suchova K, Biely P. Xylanases of glycoside hydrolase family 30–an overview. *Biotechnol Adv*. 2021;**47**:107704.
57 Gregoret LM, Rader SD, Fletterick RJ, Cohen FE. Hydrogen bonds involving sulfur atoms in proteins. *Proteins*. 1991;**9**:99–107.
58 Labourel A, Crouch LI, Bras JLA, Jackson A, Rogowski A, Gray J, et al. The mechanism by which Arabinoxylanases can recognize highly decorated Xylans. *J Biol Chem*. 2016;**291**:22149–59.
59 Correia MA, Mazumder K, Bras JLA, Firbank SJ, Zhu Y, Lewis RJ, et al. Structure and function of an arabinoxylan-specific xylanase. *J Biol Chem*. 2011;**286**:22510–20.

Supporting information

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

Fig. S1. Presents a large-scale, current GH30, group 2 subfamily phylogenetic analysis annotated with known subfamily designations and contains all biochemically and/or structurally characterized GH30 group 2 enzyme examples.

Fig. S2. Shows a structure overlay of two GH30_8 glucuronoxylanases (GXNs) providing a visual perspective for comparisons made in the main body of the corresponding publication.

Fig. S3. Presents a GH30_10 alignment, indicating with arrows important amino acid locations, which define the GH30_10 functional specificity.