doi:10.1016/j.jmb.2011.01.010 J. Mol. Biol. (2011) 407, 92–109

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Ligand Bound Structures of a Glycosyl Hydrolase Family 30 Glucuronoxylan Xylanohydrolase

Franz J. St John[1,2*†], Jason C. Hurlbert[3†], John D. Rice[4], James F. Preston[4] and Edwin Pozharski[1]

[1]*Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD 21201, USA*
[2]*Forest Products Laboratory, United States Forest Service, The United States Department of Agriculture, Madison, WI 53726, USA*
[3]*Department of Chemistry, Physics and Geology, Winthrop University, Rock Hill, SC 29733, USA*
[4]*Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611, USA*

Received 1 December 2010; received in revised form 4 January 2011; accepted 5 January 2011
Available online 19 January 2011

Edited by I. Wilson

Keywords:
glycosyl hydrolase family 30 xylanase;
glycosyl hydrolase family 5 xylanase;
Bacillus subtilis XynC;
glucuronic acid

Xylanases of glycosyl hydrolase family 30 (GH30) have been shown to cleave β-1,4 linkages of 4-*O*-methylglucuronoxylan ($MeGX_n$) as directed by the position along the xylan chain of an α-1,2-linked 4-*O*-methylglucuronate (MeGA) moiety. Complete hydrolysis of $MeGX_n$ by these enzymes results in singly substituted aldouronates having a 4-*O*-methylglucuronate moiety linked to a xylose penultimate from the reducing terminal xylose and some number of xylose residues toward the nonreducing terminus. This novel mode of action distinguishes GH30 xylanases from the more common xylanase families that cleave $MeGX_n$ in accessible regions. To help understand this unique biochemical function, we have determined the structure of XynC in its native and ligand-bound forms. XynC structure models derived from diffraction data of XynC crystal soaks with the simple sugar glucuronate (GA) and the tetrameric sugar 4-*O*-methyl-aldotetrauronate resulted in models containing GA and 4-*O*-methyl-aldotriuronate, respectively. Each is observed in two locations within XynC surface openings. Ligand coordination occurs within the XynC catalytic substrate binding cleft and on the structurally fused side β-domain, demonstrating a substrate targeting role for this putative carbohydrate binding module. Structural data reveal that GA acts as a primary functional appendage for recognition and hydrolysis of the $MeGX_n$ polymer by the protein. This work compares the structure of XynC with a previously reported homologous enzyme, XynA, from *Erwinia chrysanthemi* and analyzes the ligand binding sites. Our results identify the molecular interactions that define the unique function of XynC and homologous GH30 enzymes.

*Corresponding author. Forest Products Laboratory, United States Forest Service, The United States Department of Agriculture, Madison, WI 53726, USA. E-mail address: fjstjohn@gmail.com.

† F.J.S.J. and J.C.H. contributed equally to this work.

Abbreviations used: GH30, glycosyl hydrolase family 30; GH10, glycosyl hydrolase family 10; GH11, glycosyl hydrolase family 11; XynC, GH30 endoxylanase C from *Bacillus subtilis* 168; XynA, GH30 endoxylanase A from *Erwinia chrysanthemi*; CBM, carbohydrate binding module; CD, catalytic domain; $MeGX_n$, 4-*O*-methylglucuronoxylan; $MeGX_3$, 4-*O*-methyl-aldotetrauronic acid; $MeGX_2$, 4-*O*-methyl-aldotriuronic acid; MeGA, 4-*O*-methylglucuronic acid; GA, glucuronate; PDB, Protein Data Bank; PEG, polyethylene glycol; SSRL, Stanford Synchrotron Radiation Lightsource.

0022-2836/$ - see front matter. Published by Elsevier Ltd.

Introduction

Endoxylanases are enzymes involved in the microbial degradation of cellulosic biomass and have important applications in the bioenergy[1] and food products industries.[2] Their role in the complex process of biomass degradation is to cleave the β-1, 4-linked xylan component of hemicellulose into smaller xylooligosaccharides. Due to their widespread distribution in microbial systems, xylanases belonging to glycosyl hydrolase family 10 (GH10) and glycosyl hydrolase family 11 (GH11) represent the vast majority of all xylanases that have been identified and have been the target of the majority of studies characterizing xylanase function. For these xylanases, common appendages along the xylan chain such as 4-*O*-methylglucuronic acid (MeGA), arabinofuranose and acetyl moieties disrupt the β-1,4-xylosyl main chain accessibility and can therefore reduce the ability of these enzymes to degrade substituted xylans. For this reason, ancillary enzymatic activities such as α-1,2-glucuronidase, arabinofuranosidase and acetyl esterase activities are required for complete microbial utilization of this chemically complex polymeric substrate.

A lesser known endoxylanase activity has recently been assigned to a subgroup within glycosyl hydrolase family 30 (GH30; previously, glycosyl hydrolase family 5).[3–6] These endoxylanases cleave the β-1,4-linked xylosyl main chain of xylan as other endoxylanases, with a nucleophilic and an acid/base catalyst pair of glutamate amino acids making possible a double-displacement mechanism with retention of anomeric configuration;[7–10] however, unlike xylanases from GH10 and GH11, this xylanase type has specificity for sites of MeGA substitution. Characterization of GH30 endoxylanase A from *Erwinia chrysanthemi* (XynA) from the Gram-negative phytopathogen *Erwinia chrysanthemi* first identified the unique specificity that these enzymes display in the hydrolysis of 4-*O*-methylglucuronoxylan ($MeGX_n$).[4] Using various $MeGX_n$ substrates differing in their degree of MeGA substitution and xylans having chemically modified MeGA moieties, this work showed that XynA activity correlated with the degree of MeGA substitution along the xylan chain. It was also shown that reduction of the α-1,2-linked MeGA to glucose resulted in an almost complete loss of activity. Early application of ^{13}C NMR and matrix-assisted laser desorption/ionization-time of flight did not unambiguously identify the final limit product of hydrolysis, but supported the conclusion that recognition of MeGA moieties was a requirement for hydrolysis and that all hydrolysis products contained MeGA.

The crystal structure of native XynA [Protein Data Bank (PDB) ID: 1NOF] has been published, revealing an unexpected two-domain protein fold.[11] The N-terminal region consists of the expected $(\beta/\alpha)_8$ barrel classified as belonging to the carbohydrate-active enzymes database clan A (4/7 hydrolases),[7,12–14] but appended to this motif in a structured orientation is a C-terminal β-sheet domain tightly connected through a hinge composed of C-terminal and N-terminal linkers.[3] This side β-structure is associated with the $(\beta/\alpha)_8$ catalytic core through a hydrophobic patch on the outer side of α-helices 7 and 8. Due in part to the inaccurate limit product prediction result from the first XynA biochemical study, structure analysis failed to explain the unique functionality observed by these enzymes from the first available apo-structure.[6,11]

The protein product of the *ynfF* gene from *Bacillus subtilis*, designated GH30 endoxylanase C from *Bacillus subtilis* 168 (XynC), was identified as a putative GH30 xylanase having 40% identity to XynA from *E. chrysanthemi*. Representing the first GH30 xylanase from a Gram-positive bacterium, XynC showed indistinguishable similarities to the catalytic and substrate/product properties of XynA from *E. chrysanthemi* (J. Rice *et al.*, In Abstracts of the 105th General Meeting of the American Society for Microbiology 2005). In that work, matrix-assisted laser desorption/ionization-time of flight analysis of a XynC–$MeGX_n$ limit hydrolysis product preparation provided conclusive evidence that resulting xylooligosaccharides each contained a single MeGA substitution. Application of ^1H NMR directly supported these findings and allowed the further interpretation that the MeGA is substituted penultimate to the reducing terminus. Secondary biochemical evidence in support of this hydrolysis product prediction was obtained from double digestions with XynC and the much better characterized GH11 xylanase from *B. subtilis*.[5,15] An almost simultaneous report striving to resolve the enzymatic function of XynA from *E. chrysanthemi* confirmed these studies using many of the same techniques.[6] In addition, a methodical enzyme mediated degradation of the limit product was performed, which supported their initial limit product characterization. In all reports thus far, no neutral xylooligosaccharides in $MeGX_n$ limit digestions were detected. Together, these publications verify the unique enzymatic functionality of GH30 xylanases.

In this work, we present the first high resolution crystal structure model of XynC, a GH30 glucuronoxylan xylanohydrolase from a Gram-positive bacterium in its native and ligand-bound forms. The native XynC structure is compared to that of XynA of *E. chrysanthemi*; the xylosyl binding subsites are analyzed and compared to the better studied GH10 xylanase family. The study is further extended to address the functional role of the side β-structure as a subfamily-wide carbohydrate binding module (CBM). This work characterizes the XynC–ligand interactions to gain an understanding of the mechanism by which this unique enzyme catalyzes the

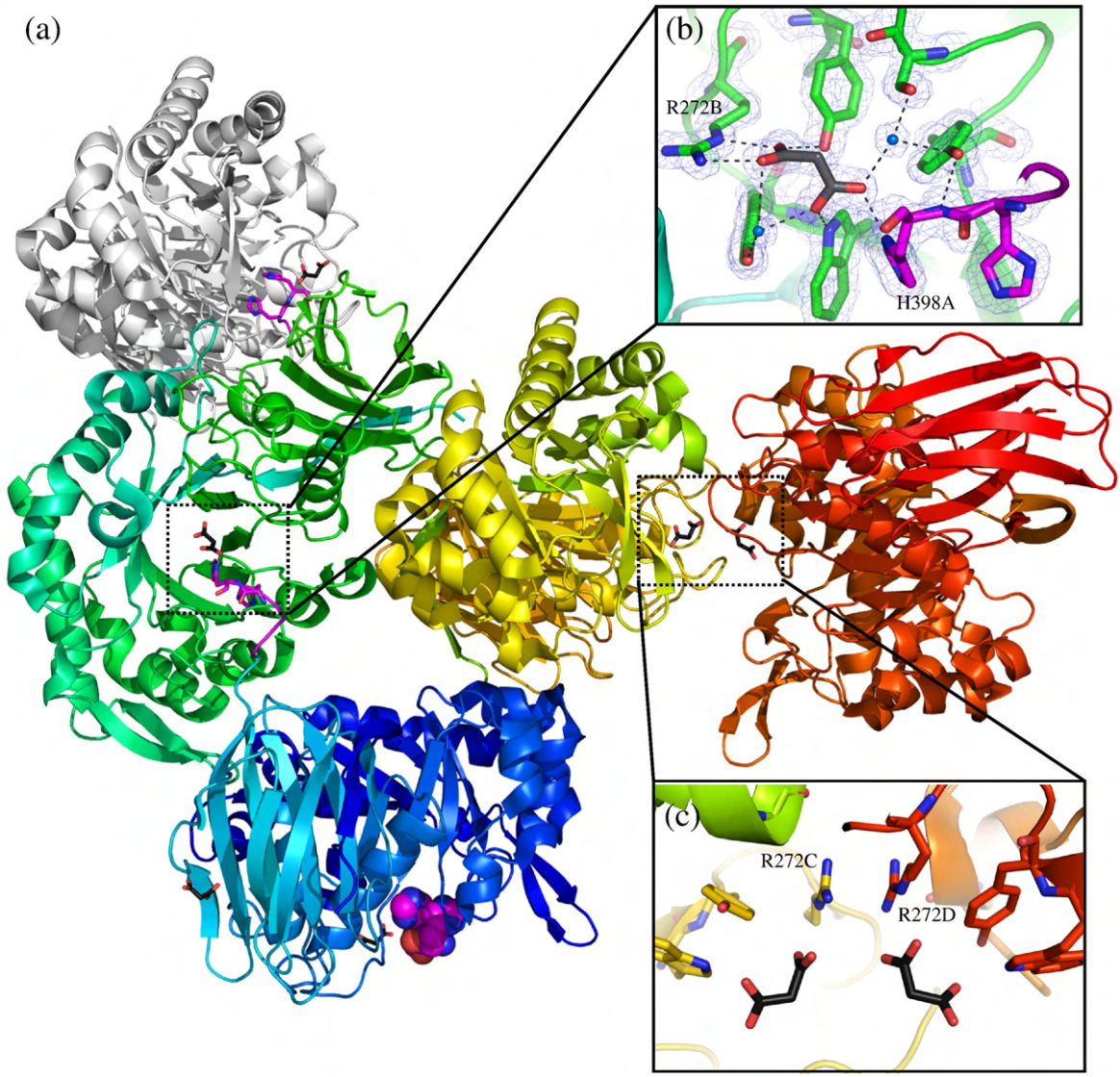

Fig. 1. Representation of (a) unit cell packing of XynC showing chain A (blue), chain B (green), chain C (yellow), chain D (red) and a symmetry related chain C (light gray). A malonate is observed (black) in the active site of each XynC chain, and a coordinated pair of histidine residues (magenta) from outstretched histidine affinity tags is modeled into the active site of chain A, chain B and a symmetry related chain C. $MeGX_2$ was only observed to be associated with the side β-structure of chain B. A close-up of malonate coordination (b) with the associated histidine residues and waters in the chain B active site revealing tight coordination with eight hydrogen bonds. (c) A close-up of the XynC chain C and chain D interface.

endohydrolysis of $MeGX_n$ chains through recognition of the α-1,2-linked MeGA main chain appendage. Application of this enzyme in enzymatic biomass pretreatments may contribute to increase process efficiency and sugar yield for optimized bioconversions.

Results and Discussion

Unit cell contents and protein packing

From the native 1.6 Å structure, C-terminal histidine tags are observed to be coordinated in the catalytic substrate binding cleft of a neighboring chain. That from chain A is the best modeled example and allowed us to extrapolate the position of others based on proximity. The model suggests that this phenomenon involves all the XynC unit cell chains including interactions with symmetry related chains (Fig. 1a). In each case, two histidine residues are coordinated through stacking and hydrogen bonding interactions. Positioning of one of the two histidine residues in the active site is facilitated by a coordinated malonate molecule in the native structure. This interacting malonate is tightly held in place by eight hydrogen bonds: four with amino

Table 1. Data collection statistics and final model quality

	Data sets		
Identification			
Crystal type	Native	GA soak	$MeGX_3$ soak
PDB code	3KL0	3KL3	3KL5
Data collection			
Wavelength	0.976	0.979	0.979
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit cell parameters			
a, *b*, *c* (Å)	138.49, 195.80, 66.25	137.87, 192.73, 65.55	137.72, 194.01, 65.71
Resolution range (Å)	50.0–1.64	50.0–2.33	50.0–2.59
Redundancy	7.2 (6.7)	7.3 (7.2)	6.6 (3.2)
No. of unique reflections	220,570 (21,814)	75,511 (7432)	54,718 (4686)
R_{merge} (%)	12.4	14.5	8.6
Completeness (%)	100 (100)	100 (100)	97.9 (85.4)
Mean (I/σ_I)	15.1 (1.0)	14.1 (2.0)	19.1 (1.0)
Matthews coefficient	2.12	2.43	2.51
Solvent content (%)	41.6	48.9	50.7
Molecules of XynC per asymmetric unit	4	4	4
Refinement statistics			
Resolution range	40.0–1.64	38.2–2.33	44.7–2.59
No. of reflections	208,972	71,579	51,824
R_{work}/R_{free} (%)	16.7/20.1	19.5/24.8	24.1/29.0
No. of protein atoms	12,847	12,280	12,303
No. of waters	2188	458	47
No. of ligand atoms	76	84	127
Average *B*-factors (Å^2)			
Protein	19.5	32.9	58.3
Waters	34.2	37.1	46.9
Ligands[a]		54.4	94.9
Solutes	24.8	39.1	
Disordered[b]		75.3	120.1
Ramachandran statistics			
Most favored (%)	97.4	96.9	94.9
Outliers (%)	0.5	0.2	0.5
RMSD from ideality			
Bond lengths (Å)	0.015	0.010	0.011
Bond angles (°)	1.38	1.21	1.25

Values in parentheses are for the highest-resolution shell.

$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, $R_{work,free} = \sum_{hkl} |F_o(hkl) - F_c(hkl)| / \sum_{hkl} |F_o(hkl)|$.

[a] Carbohydrate ligands.

[b] Refers to chain D disorder in ligand-bound structures

acid side chains lining the putative binding site, three with coordinated waters and one with the Nδ of an imidazole ring of a histidine present in the outstretched histidine tag (Fig. 1b). Unit cell packing of this XynC crystal form seems at least partially dependent upon these stabilizing interactions.

Structure models prepared from data sets collected from the ligand-soaked crystals of XynC were of lower resolution than the native XynC structure (Table 1). Chain D in the ligand bound models was greatly disrupted to the extent that a fourth copy of XynC could not be individually placed using the molecular replacement program *Phaser*.[16] Despite this, restrained refinement with *REFMAC* including only chains A–C resulted in positive density for some secondary structure elements in the region of chain D. We propose that chain D disorder in the ligand bound structures resulted from one primary phenomenon, which is made possible by the limited interactions between chain D and other chains within or between adjacent unit cells. The only crystal contact made by chain D involves an interface between chain C and chain D catalytic active site clefts (Fig. 1a). These two chains have a primary contact such that normal access of ligand would be precluded (Fig. 1c). From this, we conclude that carbohydrate ligand soaks disrupted the only significant stabilizing chain D interface by forcing ligand interaction into the chain C/chain D active site region, primarily resulting in a highly disordered chain D.

Ligand interactions in the XynC catalytic substrate binding cleft

Models resulting from data sets collected from XynC crystals soaked with either 4-*O*-methyl-aldotetrauronic acid ($MeGX_3$) or glucuronate (GA)

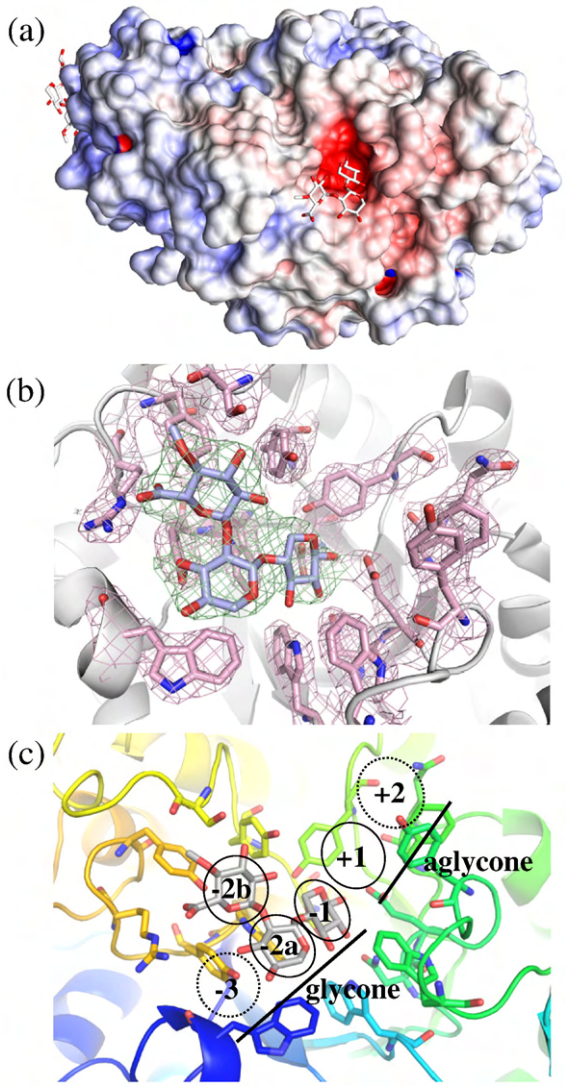

Fig. 2. (a) Electrostatic surface map of the $MeGX_2$-bound model of XynC showing the catalytic active site and the relative position of the side-associated CBM as identified from the distally located ligand and (b) a $2F_o - F_c$ density map at 1 σ differentially highlighting density for $MeGX_2$ (green) and the side chains that interface with the catalytic substrate binding cleft (pink). (c) Positioning of $MeGX_2$ in the XynC catalytic substrate binding site orients the MeGA moiety toward the β7–α7 and β8–α8 loop regions and the xylosyl chain extending in the nonreducing direction toward the β1–α1 loop. Note also the broad β4–α4 loop. Loops β1–α1 (dark blue), β3–α3 (light blue), β4–α4 (dark green), β5–α5 (green), β6–α6 (light green), β7–α7 (yellow) and β8–α8 (orange) all have amino acid residues that interface with $MeGX_2$. Subsites are designated in black text with those having modeled sugars and those assumed to exist shown as continuous lines and other potential subsites shown as broken lines.

resulted in clear density within the substrate binding cleft (Fig. 2a and b). Refinement of the resulting data sets revealed that the active site cleft of these XynC models contained either a well-coordinated 4-*O*-methyl-aldotriuronic acid ($MeGX_2$) (Fig. 3c) or GA (Fig. 3e), respectively. Both GA and the MeGA moiety of $MeGX_2$ coordinated alike. Further discussion will focus primarily upon the more complex $MeGX_2$ ligand bound structure but is intended to represent both ligand bound structures of XynC with respect to protein interactions involving the GA sugar in the catalytic substrate binding cleft and the side β-domain. Finding $MeGX_2$ in the catalytic site was expected from biochemical characterization that showed that XynC limit hydrolysis resulted in singly substituted aldouronates substituted penultimate to the reducing terminus (Fig. 3b).[5,6] $MeGX_3$, as a limit product of a GH10 xylanase, consists of xylotriose with a MeGA moiety substituted on the nonreducing xylose (Fig. 3d).[17,18] From these previous biochemical studies, it was anticipated that the $MeGX_3$ interaction with the catalytic substrate binding cleft of XynC would result in the hydrolysis of the reducing terminal xylose, if indeed a ligand bound structure were obtained. The resulting $MeGX_2$ (Fig. 3c) was positioned in the substrate binding cleft of XynC such that the reducing terminal xylose was in hydrogen bonding distance to the catalytic residues (catalytic nucleophile Glu229 and the acid/base catalyst Glu140) and that the nonreducing end xylose and its α-1,2-linked MeGA extend toward the β1–α1 and β7–α7/β8–α8 loop regions, respectively (Fig. 2c). Visual inspection of the $MeGX_2$-bound structure identifies three glycone region xylosyl or glucuronosyl binding subsites corresponding to positions −1, −2a and −2b. The −2a and −2b designations refer to the xylose and MeGA subsites, respectively. This subsite assignment allows for the possibility of a −3 glycone subsite. From the positioning of the substrate in the binding site, it is assumed that at least a single xylan binding subsite exists in the aglycone region as the polymeric sugar would extend across the catalytic residues into a stabilized position for hydrolysis (Fig. 2c).

Ligand coordination within the −1 and −2a subsites

A xylose moiety in the −1 subsite is positioned through several interactions (Fig. 4a and b). Side chains of amino acid residues Trp147, Tyr204 and Trp268 form a physical interface with the −1 subsite xylosyl ring from three sides and act to guide the sugar into position for hydrolysis. Two hydrogen bonds stabilize these nonspecific interactions. The N^{ε} of Trp85 is oriented upward toward the bound xylose, allowing the formation of a hydrogen bond between the C-3 hydroxyl oxygen of the xylose ring and the tryptophan nitrogen. The Nδ of neighboring Asn139 forms a similar hydrogen bond with the C-2 hydroxyl oxygen. These last two interactions effectively orient the xylosyl ring into position proximal to the catalytic center for hydrolysis (Fig. 4a and b).

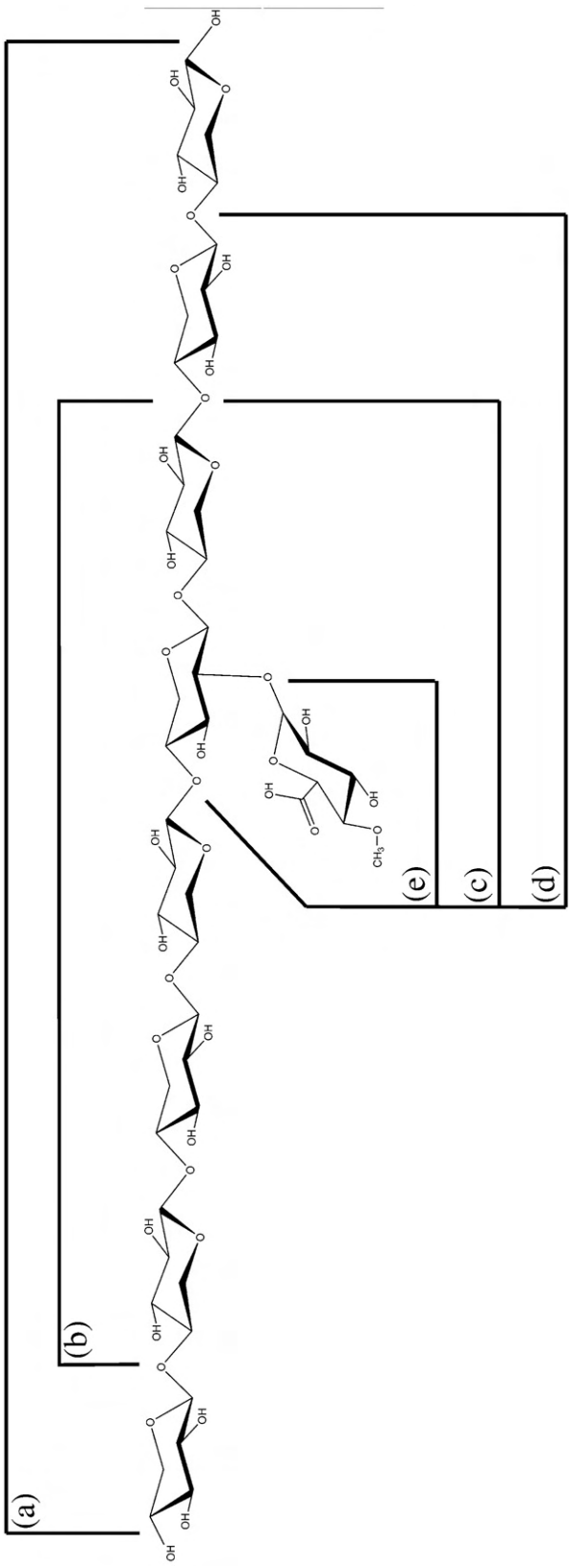

Fig. 3. Idealized $MeGX_n$ polymer and selected xylanase hydrolysis products. (a) Polymeric $MeGX_n$ consists of a β-1,4-linked xylosyl chain periodically substituted with acid-resistant α-1,2-linked MeGA moieties. (b) Upon hydrolysis of $MeGX_n$ by XynC, the limit product is an aldouronate substituted with a single MeGA moiety penultimate to the reducing terminus. (c) $MeGX_2$ was modeled into the active site of XynC. (d) $MeGX_3$ was purified as the primary aldouronic acid limit product of a GH10-mediated hydrolysis of $MeGX_n$. (e) The monomeric sugar MeGA was substituted with commercially available GA to obtain the XynC–ligand complex.

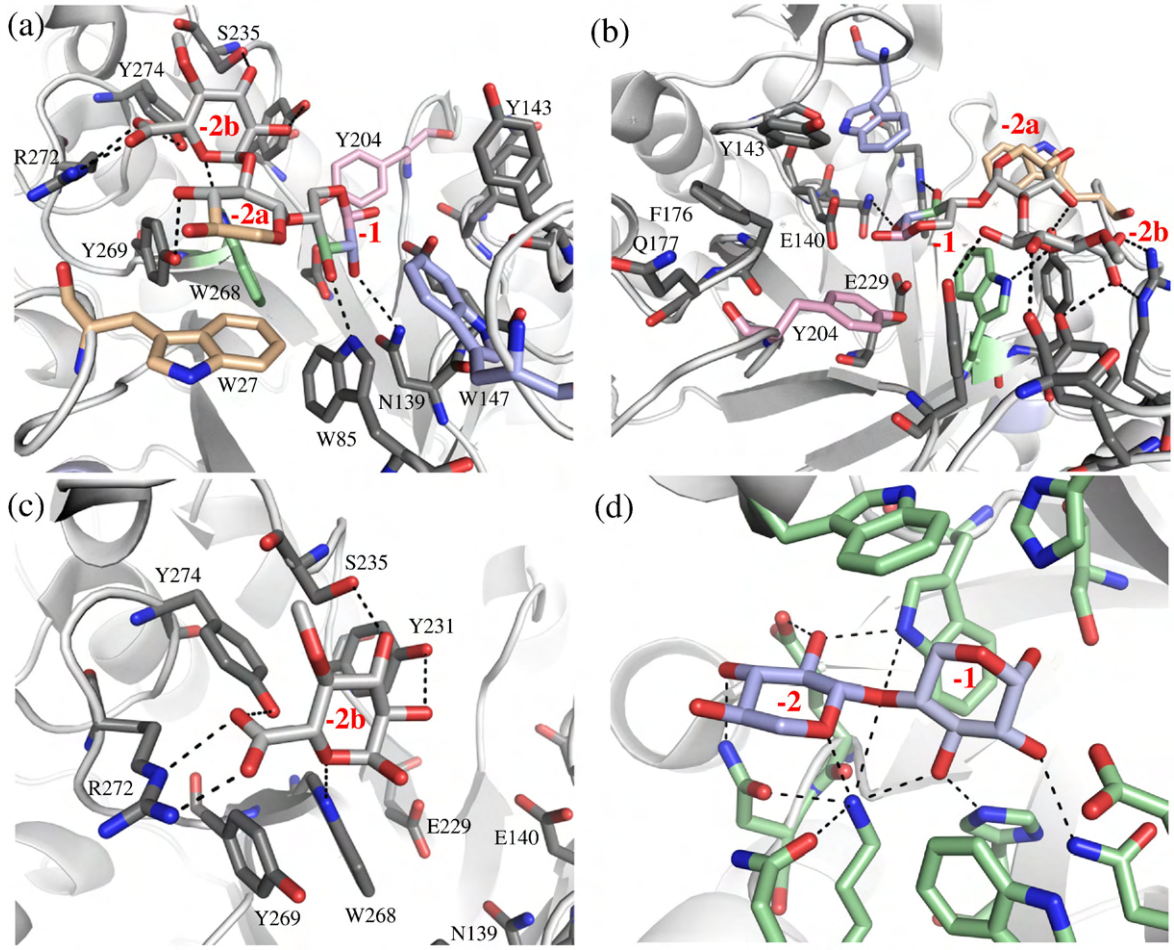

Fig. 4. Coordination of $MeGX_2$ in the glycone region of the XynC substrate binding cleft. (a) View from the nonreducing terminus indicating the subsites (red text) and showing the interactions that are thought to stabilize the sugar rings in the −1 and −2 subsites. Predicted hydrophobic interactions are color coded, matching sugar atoms with interacting amino acid side chains. (b) View from the reducing terminus showing the same as for (a) but showing also the proximity of the catalytic glutamate residues (E140 and E229) and the amino acid side chains that are thought to coordinate the extended xylan chain in the aglycone portion of the catalytic substrate binding cleft. (c) Enlargement of subsite −2b revealing the details of MeGA coordination that determines the unique catalytic specificity of XynC. (d) Xylobiose coordination in the −1 and −2 glycone subsites of the GH10 xylanase Xyn10B of *Cellvibrio mixtus* revealing the extensive ligand coordination typical of these endoxylanases which do not require a xylosyl chain appendage for hydrolysis.

In addition to these substrate interactions, the catalytic glutamate residues are within hydrogen bonding distance, but the specific manner of coordination is not clear.

The xylosyl ring in the −2a subsite (Fig. 4a) engages in a single specific interaction only but benefits from a nonspecific stacking interface. The single hydrogen bond is formed between the terminal hydroxyl oxygen of Tyr269 and the O-3 of the xylosyl ring in this subsite. The horizontally oriented Trp27 is positioned to form a nonspecific stacking interaction with the xylose in this subsite. This tryptophan side chain forms a shelf at the bottom of the xylan binding cleft and interacts with C-4 and C-5 of the xylosyl ring (Fig. 4a).

Ligand coordination within the −2b subsite

The limited interactions of the xylosyl chain in subsites −1 and −2a of XynC highlight the necessity of an alternative route of substrate recognition for catalysis. The unique specificity of XynC can be understood from the selective coordination of the MeGA within the −2b subsite (Fig. 4c). Connected through an α-1,2-linkage from the xylose in the weakly coordinated −2a subsite, the MeGA residue is tightly coordinated through as many as seven hydrogen bonds. Four of these specific contacts involve side chains of amino acids Trp268, Arg272 and Tyr274, all located within the β8–α8 loop, and the C-6 carboxylate and endocyclic oxygen of the

MeGA (Fig. 4a and c). A pair of hydrogen bonds is formed with the extended side chain of Arg272. Both N^{ε} and N^{η} groups of Arg272 hydrogen bond to the O-6A and O-6B carboxylate oxygen acceptors of the MeGA. The O-6A oxygen is also likely to accept a proton from the terminal hydroxyl of Tyr274. The N^{ε} of Trp268 is also positioned within an acceptable range to form a hydrogen bond with the endocyclic oxygen of the MeGA. Movement of the sugar toward this amino acid may be important for proper orientation of the oligomeric ligand into position for hydrolysis. On the solvent exposed side of the bound MeGA pyranose ring, the O-3 oxygen forms a hydrogen bond with the side chain hydroxyl of Ser235 that is positioned by the β7–α7 loop and extends above the ligand. The side chain of Tyr231 is located on this same β7–α7 loop, and the O-2 oxygen forms a hydrogen bond involving the phenolic hydrogen of Tyr231. An alternative scenario from our analysis revealed a possibility of three hydrogen bonds between these same four atoms provided that both the O-2 and O-3 hydroxyls of the MeGA act as acceptors. In this case, the angle and distance are suitable for the O-3 hydroxyl to act as a donor to the phenolic oxygen of Tyr231. The C-4 *O*-methyl substitution of the MeGA makes no apparent specific or nonspecific contact with surrounding amino acid side chains, nor does the O-1 oxygen of the α-1,2-glycosidic linkage.

A possible −3 glycone subsite and interactions in the aglycone region

While no ligand is observed in what would be a −3 subsite, any potential interactions that might occur would likely involve Trp27 as the side chain ring structure extends further out toward the nonreducing sugar terminus from the −2 subsite position (Figs. 4a and 2c). It is possible that the endocyclic oxygen of a xylose in this position may stabilize with hydrogen bond formation with the Arg272 guanidinium group that also serves as a primary specificity factor for the MeGA xylosyl chain appendage.

Although no ligand was apparent for modeling into the aglycone region of the catalytic substrate binding cleft, structure analysis reveals a channel formed by Tyr143 and Tyr204 that would serve to guide the extending xylosyl chain as it moves beyond the glycone into the aglycone region through the catalytic center (Fig. 4b). The catalytic proton donor, Glu140, could possibly hydrogen bond with a xylose in this position. As the xylosyl chain extends into a hypothetical +2 aglycone position, the vertically oriented side chain of Phe176 lining the bottom of the Tyr143–Tyr204 channel may provide stacking interactions, and Gln177 may be in position to hydrogen bond with the endocyclic oxygen of this xylose.

Endoxylanase comparison: Ligand coordination for catalysis

Comparison of the crucial −1 and −2 subsites with the equivalent subsites from the well characterized GH10 endoxylanase Xyn10B of *Cellvibrio mixtus* (PDB ID: 1UR1)[19] reveals large differences in the degree of xylose ring interaction for sugar positioning and catalysis. The −1 and −2 subsites of the glycone region in GH10 xylanases are considered to be generally conserved throughout the family. In the −1 subsite of these xylanases, the xylose residue is positioned through a large interface with two tryptophan rings that sit above the pyranose ring (Fig. 4d). The −1 subsite is thought to be stabilized with up to three additional hydrogen bonds that involve the O-2 and O-3 hydroxyl positions, not considering the hydrogen bonding contacts of the catalytic glutamate residues. This is one more hydrogen bond and what would seem to be more significant stacking contacts compared to XynC. The −2 subsite also has increased interactions relative to XynC. In this position, only minor nonspecific contacts occur, but every hydroxyl (except O-4) of the xylosyl ring, the endocyclic oxygen and the oxygen of the glycosidic linkage coordinate through a network of hydrogen bonds (Fig. 4d).[19] These observations support the conclusion that while these two endoxylanase families are expected to perform hydrolysis of the β-1,4-xylosyl linkage by the same catalytic mechanism, major differences exist in the way these xylanases recognize, bind and stabilize substrate sugars for efficient catalytic cycling. The low number of stabilizing contacts in the −1 and −2 glycone subsites that are located proximal to the catalytic center identifies these sites as having a limited role in determining specificity for XynC (Fig. 4a–c).

A notable difference in xylose coordination between GH10 xylanases and XynC is that the xylosyl chain symmetry while bound into the −1 and −2 glycone subsites appears to be different. Xylan has been reported to have a 3-fold helical symmetry.[20–22] This conformation is observed throughout the catalytic cleft of GH10 xylanases.[19] Between the adjacent −1 and −2 subsites, xylosyl rings are rotated ca 120°, as can be observed in the structure of Xyn10B from *C. mixtus* (Fig. 4d).[19] In the ligand-bound structure of XynC, the xylan is observed in a near 2-fold symmetry with the pyranose ring in subsite −1 and subsite −2, being flipped approximately 180° with respect to the other. Coordination of the xylosyl chain in this orientation may induce ring strain that may facilitate catalysis by XynC. Based on the reduced number of protein–ligand coordinating contacts in the −1 and −2 subsites of XynC and the precise coordination of the distal-oriented MeGA, a possible sugar binding mechanism may involve a lever

action where the MeGA binding forces the xylosyl chain into proximity for hydrolysis.

Another interesting difference between XynC and GH10 xylanases is found in the aglycone region of the substrate binding cleft (Fig. 4b). In this region of XynC, a channel with sides formed by Tyr143 and Tyr204 and a bottom formed by Phe176 defines the +1 subsite. While it is possible that a single hydrogen bond contact may facilitate stabilization of xylose in this position, overall binding appears to be relatively nonspecific. However, the three aromatic side chains that form the guiding channel in this subsite constitute a much more significant contact than what is observed in the typical +1 subsite of GH10 xylanases.[23] GH10 xylanases have been shown to require at least two xylose residues in the glycone side of the catalytic center as xylose does not bind tightly in the −1 subsite.[17] From biochemical and structural data, it is also known that the α-1,2-linked MeGA moiety can be accommodated in the +1 and −3 xylosyl binding subsites.[19] Previous hydrolysis product studies of XynC[5] showed that the smallest MeGA-substituted aldouronate generated by $MeGX_n$ hydrolysis was $MeGX_2$. This suggests that XynC is able to accept MeGA substitutions in the +1 aglycone subsite, just two positions away from the −2b MeGA-coordinating glycone subsite. The near 2-fold symmetry of the xylosyl chain bound in the substrate binding cleft of XynC helps explain this observation and also suggests that the 2-fold symmetry of the bound sugar continues into the aglycone region where the O-2 hydroxyl of the sugar in this position would be solvent accessible.

Glucuronoxylan interactions with the putative CBM

The side β-domain of XynC consists of a nine-stranded aligned β-sandwich. This structure is tightly associated with the $(\beta/\alpha)_8$ barrel catalytic domain (CD) through a pair of hinge regions and a hydrophobic interface.[3] This motif, originally described in the structure report on XynA from *E. chrysanthemi*, was suggested to have xylan binding function due to its close proximity to the catalytic $(\beta/\alpha)_8$ barrel.[11] The ligand bound structure of XynC identifies an unambiguous role for this motif of XynC in targeting $MeGX_n$ as substrate (Fig. 5a and b). However, MeGA coordination within the CBM cleft of XynC is not an obviously conserved feature in XynA and its homologs (discussed below). CBM xylosyl binding subsites are assigned in a similar fashion as those of the aglycone region of the catalytic xylan binding cleft. In this arrangement, the 1 subsite is occupied by the xylose residue toward the reducing terminus; the 2a subsite is occupied by the adjacent residue toward the nonreducing terminus and is the xylosyl ring

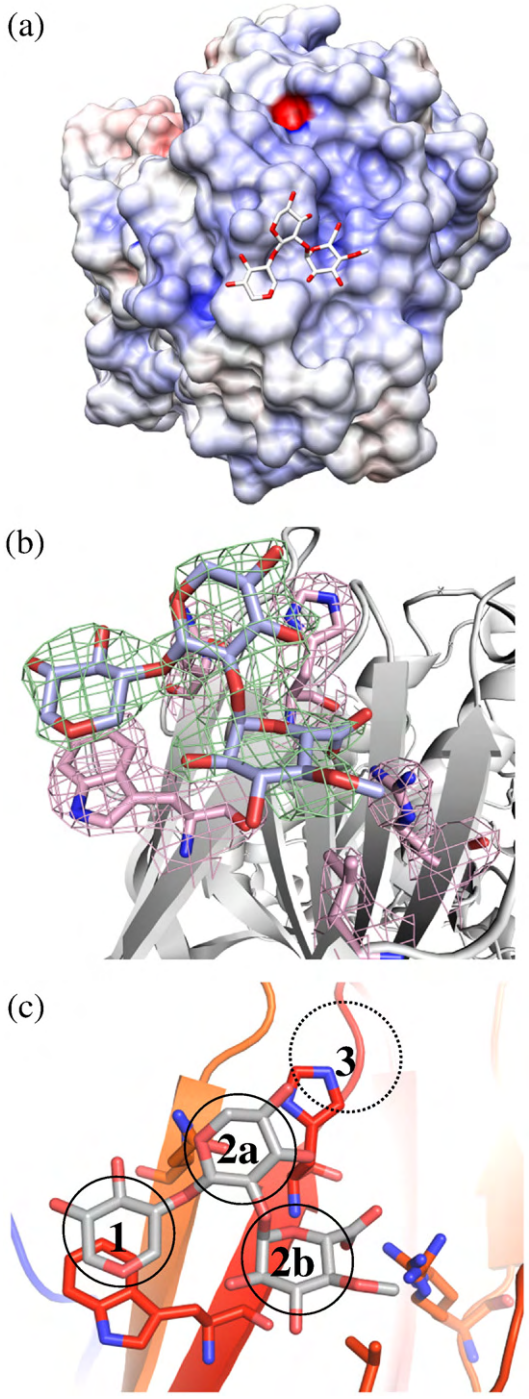

Fig. 5. Ligand binding to the side β-domain. (a) Electrostatic surface map showing $MeGX_2$ coordinated into the CBM motif. See Fig. 2a for relative position from ligand bound in the catalytic substrate binding cleft. (b) The $2F_o - F_c$ density map showing differentially colored density for the ligand and coordinating side chains. (c) Subsite assignment is similar to the catalytic substrate binding cleft. Subsites are designated in black text with those having modeled sugars and those assumed to exist shown as continuous lines and broken lines, respectively.

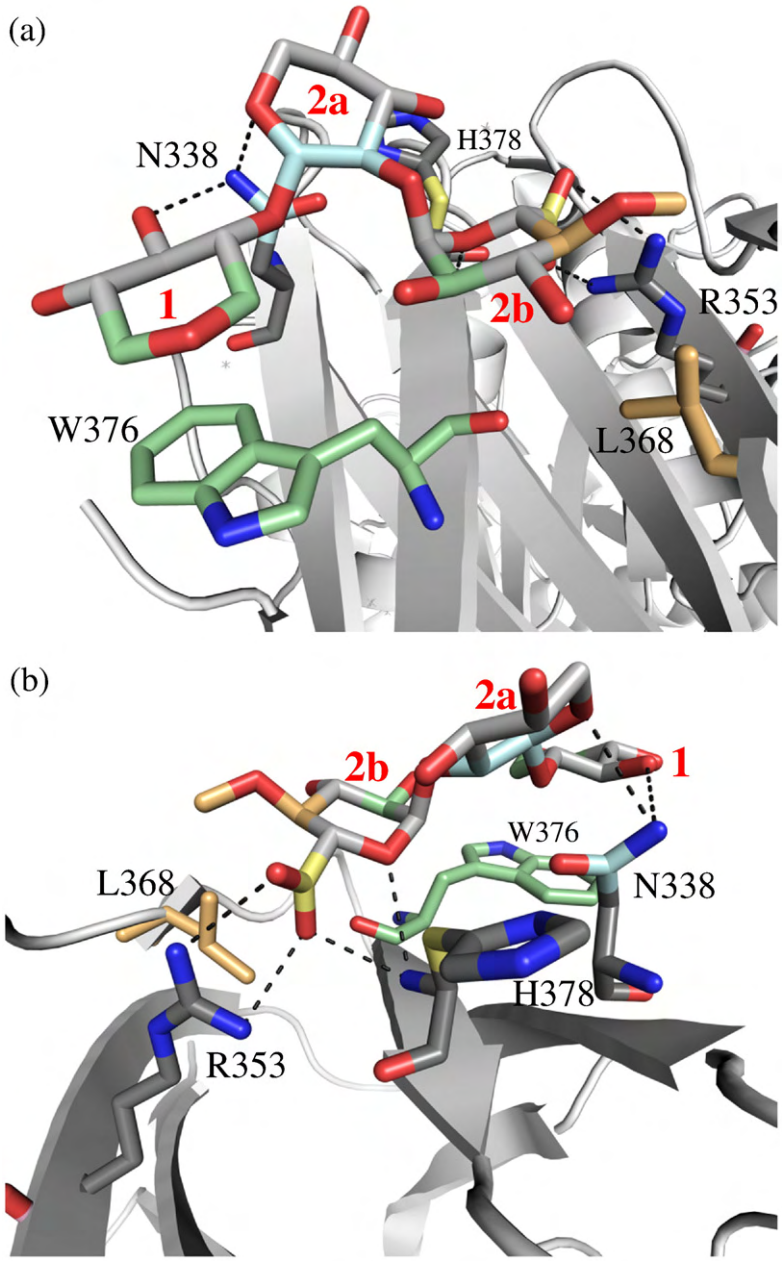

Fig. 6. Coordination of $MeGX_2$ into the side β-domain. (a) Viewed from the reducing terminal sugar in the 1 subsite. (b) Imaged from the 2b subsite coordinating the MeGA residue. Hydrogen bonds and color-matched hydrophobic contacts are detailed.

substituted with an α-1,2-linked MeGA. Subsite 2b is occupied by MeGA (Fig. 5c). In this ligand bound structure, no sugar is additionally appended to account for a third subsite, but from inspection of the binding site, it seems possible that this site may exist.

Stabilizing xylosyl ring contacts in the 1, 2a and predicted 3 subsites of the CBM

The xylosyl ring in the 1 subsite has a primary stacking interface with surface localized Trp376, and the Nδ protons of Asn338 coordinate between the O-3 hydroxyl of this xylose and the endocyclic oxygen of the xylose in the 2a subsite (Fig. 6a). Analysis suggests that C^{γ} of Asn338 interacts with the C-2 and C-1 carbons of the xylose in the 2a subsite, stabilizing this position through hydrophobic contacts. There is no observed sugar in the ligand that might identify a 3 subsite, but potential interactions may occur in this position with the imidazole ring of His378 (Fig. 6b).

Coordination of the methylglucuronosyl moiety in the 2b subsite

In a manner analogous to the observed MeGA specificity in the XynC catalytic site, the CBM also

has the majority of stabilizing substrate interactions with this xylosyl chain substitution. The oxygens of the C-6 carboxylate engage in several hydrogen bonds with the amino acids lining the region (Fig. 6b). Specificity seems to be endowed through hydrogen bonds formed with the outstretched side chain Arg353, which originates from the βs6 strand and projects through the βs7–βs8 surface cleft, forming hydrogen bonds with the MeGA. Although not conclusively identified from the model, it seems likely that the carboxyl oxygens of MeGA form two hydrogen bonds with the $N^{\eta 1}$ and $N^{\eta 2}$ atoms of Arg353 in a way similar to the contacts observed with Arg272 in the catalytic −2b subsite. Furthermore, the peptide nitrogen of His378 has its hydrogen available for bonding with the C-6 carboxylate or, alternatively, the O-5 endocyclic oxygen of the MeGA (Fig. 6b).

In addition to these very specific hydrogen bonding interactions, several hydrophobic associations between the CBM and substrate are observed. Analysis identified the C-6 carboxylate carbon and the C-2 carbon of the MeGA pyranose ring interacting with C^{β} of His378 and Trp376, respectively. The most specific hydrophobic association occurs between Leu368 and the C-7 *O*-methyl and C-4 pyranose ring carbons of the MeGA. This interaction suggests that the CBM may preferentially interact with the 4-*O*-methyl derivative of $MeGX_n$, the form observed in mature woody biomass. It is interesting to note that no such interaction between the protein and the 4-*O*-methyl substitution in the −2b MeGA-coordinating subsite of the catalytic substrate binding cleft is observed.

Structure comparison and conservation in GH30 glucuronoxylan xylanohydrolases

The catalytic substrate binding region of XynC is located on the C-terminal side of the $(\beta/\alpha)_8$ barrel. Seven of the eight β–α loop regions interact with the xylosyl chain (Fig. 2c). Only the β2–α2 loop does not make stabilizing contacts with bound ligand. Loops 1, 3 and 6 each possess a single amino acid side chain that interacts with bound substrate, while loop 5 presents two amino acids that are predicted to form the putative +2 subsite. Beyond these contacts, the majority of the amino acid side chains that interact with bound substrate reside on loops 4, 7 and 8. This is of notable interest since two of these loops emerge from the juxtaposed β4 and β7 catalytic glutamate positions and the third is the primary factor in MeGA coordination.

The β4–α4 loop is a broad inward-leaning loop (Figs. 2c and 4a) that extends from the catalytic proton donor (Glu140) and provides coordination for xylose residues in the +1 and −1 subsites. The intermolecular contacts defining these subsites involve stacking interactions, which serve to define the β3–α3/β4–α4 side of the substrate binding cleft. Its large interface and role in coordinating the xylosyl chain identify this loop as being important to the function of these enzymes. The β7–α7 and β8–α8 loops also have a large role in substrate coordination (Fig. 4c). However, these loops primarily serve to coordinate the MeGA appendage through multiple hydrogen bonding interactions as previously described. Combined, the MeGA contacts established with these loops form the −2b subsite. The β7–α7 loop extends from the position of the catalytic nucleophile (Glu229) and forms up to three hydrogen bonds with one side of the MeGA ligand, while amino acid side chains in the β8–α8 loop form four hydrogen bonds. The novel substrate specificity of these enzymes is no doubt directed by the extensive MeGA coordination observed in these two later loop regions.

XynC is the second GH30 glucuronoxylan xylanohydrolase that has been structurally studied. While this enzyme comes from the Gram-positive organism *B. subtilis*, the homologous enzyme with known structure, XynA, comes from the Gram-negative phytopathogen *E. chrysanthemi*.[11] XynC shares 40% sequence identity with XynA, and structural superposition results in a percentile-based spread of just 0.87 Å.[24] Compared to larger families of xylanases such as GH10 and GH11 in which all members have xylanase activity, XynA and XynC are only a subgroup (subgroup H) of enzymes classified belonging to the larger, enzymatically diverse GH30 family.[3] A UniProt *BLASTp* (release 2010_08) search with the XynC query sequence reveal only 43 proteins with more than 35% sequence identity, a value below confident protein function prediction.[25–27] Comparison of 30 amino acid sequences with levels of identity to XynC or XynA of approximately 40% or greater reveals a high level of amino acid conservation in the catalytic xylan binding subsites (Table 2). This analysis reveals that these 30 homologous proteins can be grouped based upon their similarity to either XynC or XynA or those from Gram-positive or Gram-negative bacteria, respectively.

The catalytic xylan binding subsites

As briefly described above, the β4–α4 loop in both XynC and XynA extends from the topside of the β4-strand-positioned catalytic proton donor to a proximal aromatic residue, which in part forms the +1 subsite aglycone channel and, after another three amino acids, stabilizes the −1 subsite or the −2 glycone subsite in XynC and XynA, respectively. The extensive substrate interaction provided by this singular loop region as it extends from the catalytic proton donor accounts for a significant proportion of substrate coordination that is not MeGA based

Table 2. List of unique XynC and XynA, GH30 homologs, showing their expected Gram-stain type, level of identity and modular composition

No.	Organism	Gram	UniProt	Identity (%, XynC/XynA)	Module
1	*Bacillus subtilis* (XynC)	+	Q45070	100/40	CD[a]
2	*Bacillus subtilis*	+	D4FXC2	99/40	CD[a]
3	*Bacillus amyloliquefaciens*	+	A7Z5A1	91/40	CD[a]
4	*Bacillus* sp.	+	D1MEP8	91/40	CD[a]
5	*Bacillus amyloliquefaciens*	+	Q70K02	91/40	CD[a]
6	*Bacillus subtilis*	+	Q6YK37	91/39	CD[a]
7	*Geobacillus* sp.	+	D3EH02	87/42	CD[a]
8	*Bacillus pumilus*	+	A8FDV2	84/39	CD[a]
9	*Bacillus pumilus*	+	B4ADC5	86/39	CD[a]
10	*Aeromonas punctata*	–	P70733	81/40	CD/CBM6
11	*Aeromonas punctata*	–	O24852	78/41	CD/CBM6
12	*Clostridium lentocellum*	+	D5R2K1	75/40	CD[C]/CBM13/CBM2
13	*Cellulosilyticum ruminicola*	+	D2KFJ5	73/42	CD[C]/CBM13/CBM2
14	*Clostridium cellulolyticum*	+	B8I0M3	72/42	CD/CBM6
15	*Clostridium papyrosolvens*	+	C7IGY2	72/42	CD/CBM6/Dck
16	*Clostridium thermocellum*	+	A3DJS9	71/41	CD/CBM6/2Dck
17	*Clostridium acetobutylicum*	+	Q97TI3	74/40	CD/CBM13
18	*Streptomyces sviceus*	+	B5HS07	68/41	CD
19	*Streptomyces scabies*	+	C9YUC5	65/41	CD
20	*Catenulispora acidiphila*	+	C7Q130	65/42	CD/CBM2
21	*Ruminococcus* sp.	+	D4LE58	54/40	CD/CBM4_9/Est
22	*Ruminococcus albus*	+	Q9AJB3	51/37	CD[C]/CBM4_9
23	*Cytophaga hutchinsonii*	–	Q11TF9	51/41	CD/CBM9
24	*Cellvibrio japonicus*	–	B3PEH7	43/57	CD[A]/CBM2
25	*Xanthomonas campestris*	–	Q3BX17	40/59	CD[b]
26	*Xanthomonas oryzae*	–	B2SQM5	40/59	CD[b]
27	*Xanthomonas campestris*	–	B0RUB6	40/57	CD[b]
28	*Dickeya dadantii*	–	C6C727	40/82	CD[b]
29	*Erwinia chrysanthemi*	–	Q938A4	40/83	CD[b]
30	*Erwinia chrysanthemi* (XynA)	–	Q46961	40/100	CD[b]

Sequences were numbered according to an alignment output.
[a] Sequences that have a conserved XynC-like MeGXn binding site.
[b] Sequences that have a conserved XynA-like predicted MeGXn binding site.

and highlights its importance in xylosyl chain coordination.

With this consideration, a significant difference in the tertiary fold of XynC and XynA with implications in substrate coordination is found in the β3–α3 loop of the $(\beta/\alpha)_8$ barrel between these homologs. In XynC, this loop adopts a structured two-stranded β-sheet conformation that extends upward and over the adjacent β4–α4 loop. The β3–α3 loop of XynC is stabilized through a multitude of intra-strand hydrogen bonds, which serve to provide rigidity to the structure (Fig. 7a and b). In XynC, positioning above the β4–α4 loop region is stabilized by the guanidinium group of Arg97 extending down and stacking in a parallel fashion with the side chain of Trp149. The side chain of this arginine residue appears to be positioned via stacking contacts with the side chain of Phe95. Such side chain stacking continues along the β-strand involving the side chains of Arg105 and, finally, Tyr108 positioned at the top of the α3 helix (Fig. 7b). In addition to the support provided by this array of bulky side chains, the main chain carbonyl of Lys104 positioned in the β3–α3 loop hydrogen bonds with the peptide nitrogen of Trp149 in the β4–α4 loop. The sum of these contacts results in the preservation of a close interface between the β-structure of the β3–α3 loop and the β4–α4 loop region in XynC. These interactions reflect a role for the XynC β3–α3 loop internal β-sheet structure in stabilizing the β4–α4 loop region and highlight the importance of this latter loop region in the functioning of the enzyme.

Inspection of the same β3–α3 loop in XynA shows it to be rather minimized compared to XynC, extending slightly inward toward the region of substrate binding. However, this loop forms up to eight hydrogen bonds with the main chain and the numerous interfacing side chains of the β4–α4 loop (Fig. 7c). These hydrogen bonds are involved with a much larger network of hydrogen bonds that could be attributed to stabilizing this region together with the β4–α4 loop. Additionally, the β4–α4 loop of XynA is predicted to interact with the −2 glycone subsite and is held in proper orientation via a hydrogen bond between Glu173 and Arg83. These amino acid residues are conserved in homologs of XynA. The sum of these contacts seems to be equivalent to the highly structured β3–α3 region in XynC and likely serves the same role in the stabilization of

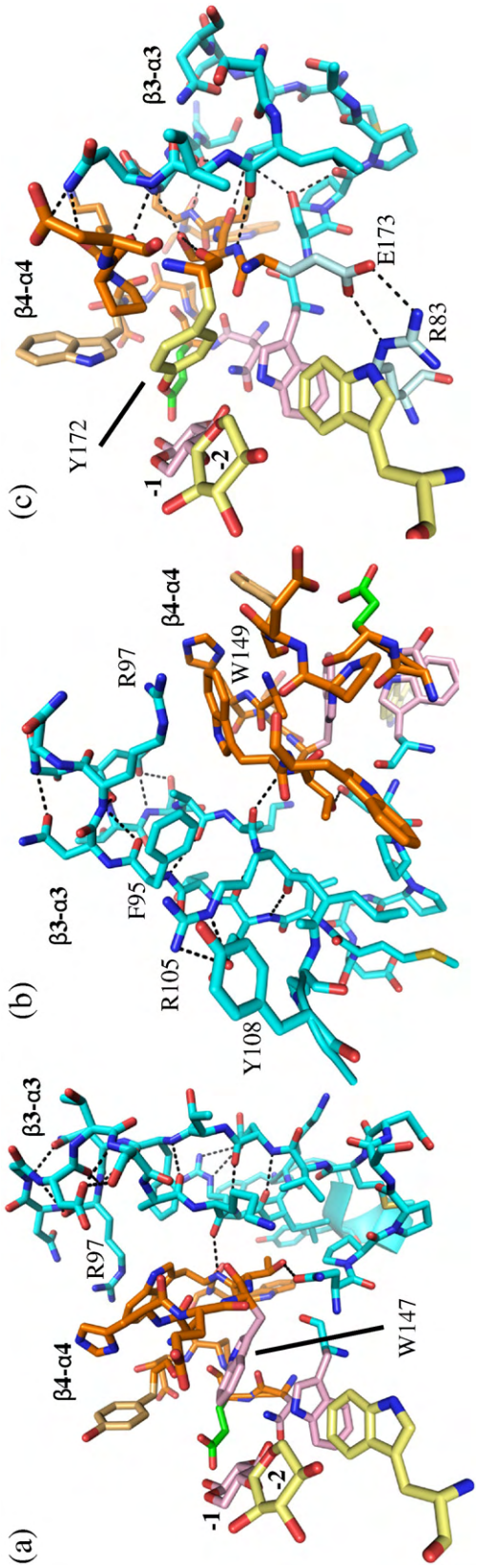

Fig. 7. The differential β3–α3 loop region of XynC and XynA. (a) Contacts between the β3–α3 loop (cyan) and the β4–α4 loop (orange) of XynC revealing the limited interactions between these two adjacent loops. (b) The same region of XynC viewed from the opposite perspective showing the stacking interface between R97 in the β3–α3 loop and W149 in the β4–α4 loop and also the side chain stacking involving F95, R105 and Y108 that are thought to stabilize this extended loop structure. (c) The role of hydrogen bond stabilization of the β4–α4 loop in the XynC homolog, XynA, of *E. chrysanthemi*. Subsite –2, yellow; subsite –1, pink; subsite 1, light orange; catalytic acid/base catalyst, green.

the functionally important β4–α4 substrate coordination loop region.

This difference in tertiary structure between XynC and XynA with respect to the β3–α3/β4–α4 loop region may also account for the observed difference in the generally conserved xylosyl subsite coordinating amino acids. Comparison of the catalytic clefts of XynC and XynA reveals that just above the catalytic acid/base catalyst (Glu140 in XynC), the tyrosine (Tyr143) forming one side of the putative +1 subsite of XynC (Fig. 4b) is replaced with a tryptophan (Trp168) in XynA. This seemingly small conservation of function substitution may compensate for a difference in additional amino acid sequence found further along the β4–α4 loop of XynA. A two amino acid insertion shifts the position occupied by Tyr172 by 5.3 Å relative to the equivalent amino acid (Trp147) of XynC. This shift in Cα position replaces Trp147, which is thought to stabilize substrate binding in the −1 subsite with Tyr172 (of XynA) that would interface primarily with the xylose in the −2a subsite (Fig. 7a and c). These small differences may lead to slight variations in substrate recognition and mechanisms of stable complex formation.

This possibility seems likely following a recent publication describing a low-level hydrolytic activity of neutral xylooligosaccharides from the close XynC homolog Xyn5B (92% identity) from the Gram-positive *Bacillus* sp. BP-7.[28] XynA was previously verified not to cleave such neutral xylooligosaccharides,[6] but XynC was never specifically analyzed in the absence of aldouronate sugars.[5] From the structure analysis, Trp147 of XynC appears to stabilize the −1 subsite, while the two amino acid insertion in the β4–α4 loop moves the equivalent Tyr172 of XynA further out, interfacing primarily at the −2 subsite. The increased xylose coordination in the −1 subsite of XynC may allow for the low-level hydrolytic activity observed on neutral xylooligosaccharide by the XynC homolog.[28]

Few other slight differences are observed in the active site xylosyl binding cleft. Phe176 positioned at the end of β-strand 5 in XynC is predicted to facilitate substrate association into a hypothetical +2 subsite along with Gln177. In XynA, the equivalent amino acid residue to Phe176 is Leu204, and there is no structural equivalent for Gln177. All other residues in the catalytic xylan binding cleft are conserved between the two enzyme groups, although with minor positional deviations in the MeGA coordination between the two.

The side β-structure as a CBM: An evolving GH30 subfamily H function

Ligand binding by the XynC side β-domain CBM reveals an evolving function, the structural details of which do not appear to be directly conserved in XynA and its closest homologs. Comparison of the structures of XynC and XynA with the sequence alignment of GH30 subfamily H homologs reveals the potential for convergent evolution of carbohydrate recognition in the side β-structure. While XynA does not have the precise motif as found in XynC, a structurally similar region, shifted just two amino acids in one direction, supports the likelihood that XynA may also have a mechanism for recognition of the $MeGX_n$ substrate with similar, if not the same, specificity for the MeGA moiety. From the comparison in Fig. 8a and b, it can be seen just how similar the putative CBM ligand binding site appears. The same three β-strands of the side β-structure, which form the ligand binding region of XynC, are also involved in XynA. Most notably, positioned identically to Arg353 in XynC, Lys375 in XynA, reaches out between strands β7 and β8. Along with a closely positioned Asn403, the MeGA moiety C-6 carboxylate may form multiple hydrogen bonds (Fig. 8b) in a similar way as observed for the analogous interaction with Arg353 in XynC (Figs. 8a and 7b). Shifted up just two amino acid residues, the XynC Trp376 equivalent in XynA is Trp401 (Fig. 8b). This shift greatly reduces the size of the putative XynA CBM site by bringing the tryptophan ring in close proximity to Lys375 where the MeGA C-6 is thought to coordinate. Asp358 would then be in position to hydrogen bond with the xylose in the site equivalent to the 2a subsite of XynC. Such a binding site would primarily coordinate just the MeGA appendage with its attached xylose (4-*O*-methyl-aldobiuronic acid). This would result in a smaller binding site than observed in XynC, but one that is just as specific for the intended substrate.

Most of the GH30 subgroup H enzymes can be assigned into one of three categories based upon carbohydrate binding (Table 2). XynC represents a closely related group of GH30 subgroup H sequences from Gram-positive bacteria with a CBM located within the side β-domain (Table 2, denoted as CD^C). While the crystallographic evidence is compelling, the findings would benefit from biochemical verification. Similarly, XynA represents a group of closely related sequences from Gram-negative bacteria that have a conserved motif that appears divergent from the XynC motif but still likely to facilitate substrate association (Table 2, denoted as CD^A). This is based solely on XynA structure analysis and needs to be confirmed biochemically and/or structurally. The last group consists of those enzymes with appended, separately folding, characterized CBMs. Together, these three groups account for 28 of the 30 sequences in Table 2 and suggest that evolution of a self-contained CBM is displacing the need for a separately appended CBM function.

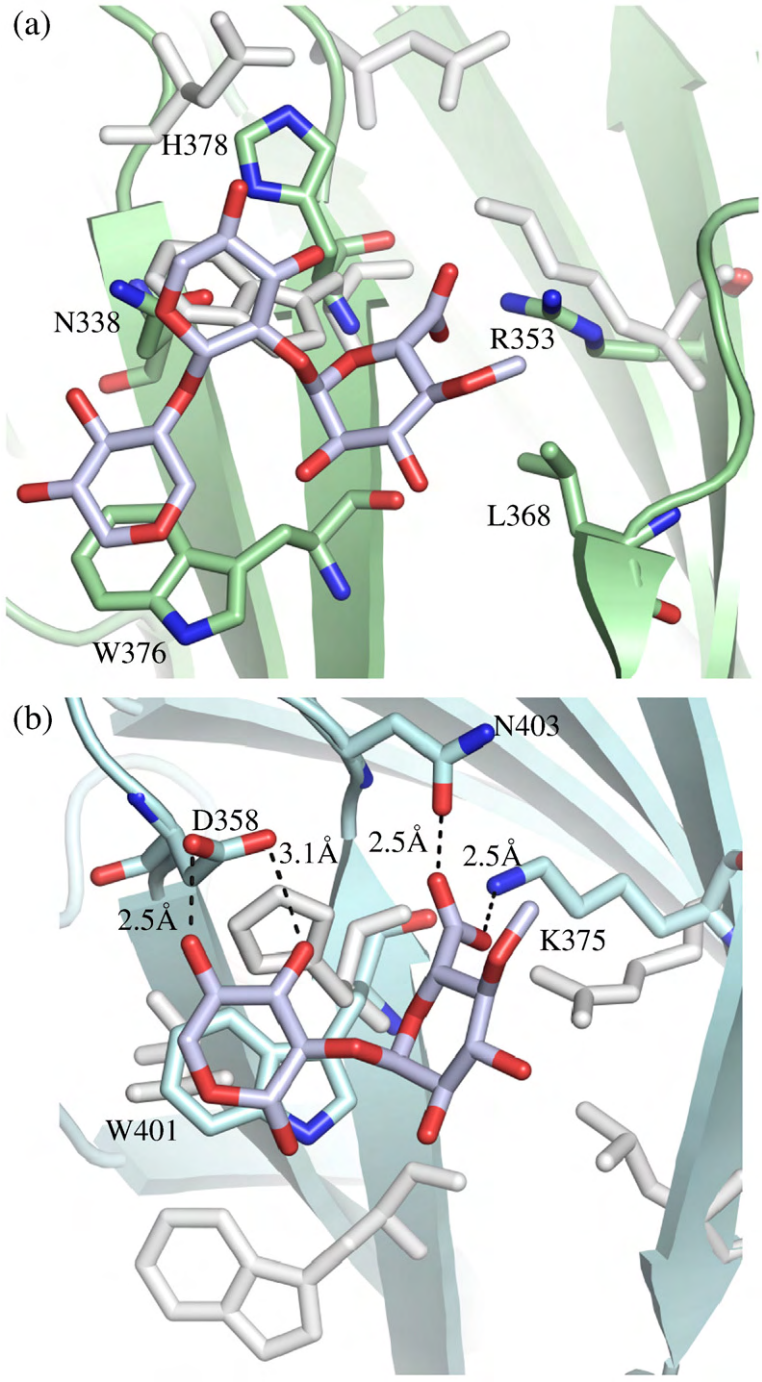

Fig. 8. A potentially similar CBM role for the side β-domain of the XynC homolog, XynA from the Gram-negative bacteria *E. chrysanthemi*. (a) Ligand coordination on the side β-structure of XynC as determined through structural studies. (b) 4-*O*-methyl-aldobiuronic acid oriented to the region of XynA that appears, upon close inspection, to have structural similarity to that observed in XynC of *B. subtilis*. Residues colored white represent XynA overlaid in the XynC picture (a) and XynC inset into the XynA picture (b).

Summary

The ligand-bound structure models of XynC from *B. subtilis* 168 have provided details that support the previous biochemical characterization. These endo-β-1,4-xylanases are directed to hydrolyze the complex $MeGX_n$ substrate by the positioning of an α-1,2-MeGA moiety along the xylan chain. Limit hydrolysis by these enzymes results in aldouronates with single MeGA substitution on the penultimate xylose toward the reducing terminus. Ligand–XynC interactions highlight a significant role for MeGA substitution in contrast to the two preceding xylose positions leading into the catalytic center for hydrolysis. Studies currently underway are focused on understanding how these enzymes efficiently turn over substrate, considering that the primary sugar recognition is a xylosyl main chain substitution located distally from the catalytic center. The findings of this work also support previous

assertions that the side β-domain characteristic of GH30 enzymes serves a CBM role in XynC, not just coordinating a xylan chain but rather showing specificity for MeGA substitution in a way similar to that observed in the catalytic substrate binding cleft. From our structure analysis, it is proposed that a similar MeGA-specific binding motif exists in the structurally homologous enzymes from Gram-negative bacteria. The unique function of these enzymes may offer a benefit in the enzymatic pretreatment of complex lignocellulosic substrates or in the development of second-generation bacterial biocatalysts for the production of green, value-added products and fuels.

Materials and Methods

XynC cloning, protein expression and purification

Construction of the XynC protein expression vector for XynC crystal studies, high-level protein expression and XynC purification was previously described.[5,29] In short, the gene encoding XynC (*ynfF*) of *B. subtilis* subsp. *subtilis* str. 168 was cloned by PCR using a previous expression construct as template.[5] XynC was cloned truncating the N-terminal secretion signal sequence and creating an in-frame fusion with the C-terminal-localized 8× His tag. The new construct was verified by sequencing. XynC protein expression and purification were performed as originally described.[5,29]

Crystallization and ligand soaks

Crystallization of XynC was described previously.[29] The optimized condition contained 200 mM sodium tartrate and 200 mM sodium malonate (pH 7.0) in 19% polyethylene glycol (PEG) 3350. High-quality crystals were obtained under this condition only when crystallization drops were set at 500 nl volume (50–66.6% XynC) using the OryxNano crystallization robot from Douglas Instruments (Berkshire, UK). Methods that were originally described[29] to obtain XynC–ligand cocrystals by cocrystallization or ligand soaking did not succeed with this crystal form and will be discussed below. Crystal structures of XynC containing sugar ligands, as presented in this work, were obtained by soaking the mechanically sturdy XynC crystals in consecutive PEG-based conditions to first remove the malonate and tartrate mother liquor components and then to introduce the putative sugar ligands. Briefly, crystals were transferred into 35% PEG 3350 and then allowed to equilibrate overnight. To promote removal of a tightly coordinated malonate found in the binding site in the initial XynC structure, we transferred the crystals into 40% PEG 3350 containing 60 mM glucuronic acid and allowed to equilibrate overnight. The glucuronic acid was removed by exchange with 40% PEG 3350 followed by a similar equilibration. This was then exchanged with a PEG mixture containing 35% PEG 3350 and 30% PEG 400, which was subsequently diluted with a third of the approximate volume of carbohydrate stock solutions. These soaks were allowed to equilibrate for 48 h. The PEG concentration in the final carbohydrate soak was 23% PEG 3350 and 20% PEG 400. This was determined to be a suitable cryoprotectant for flash cooling in liquid nitrogen. The concentration of carbohydrate in the soaks was 33 mM, 28.6 mM, 8.3 mM and 15.3 mM for GA, xylobiose, xylotriose and 4-*O*-methyl-aldotetrauronate, respectively. At the time the crystal was flash cooled for data collection, most large crystals looked visually deteriorated. Small crystals were selected for data collection. Although similar quality data sets were collected from crystals derived from each of the different ligand soaks, after data reduction and preliminary refinement, only data sets resulting from the GA-based substrate soaks had evidence of ligand interaction and were subsequently processed to obtain the final structure model.

Data collection, processing and model refinement

Data collection and scaling of the native XynC data set was described previously.[29] Phases were determined by molecular replacement with the program *Phaser*,[16] using the original low-quality 2.7-Å model (PDB ID: 3GTN).[29] Model building was performed iteratively using *Coot* as part of the CCP4 suite (version 6.1.1) of programs for manual real-space refinement followed by intermittent rounds of restrained refinement in *REFMAC*.[30,31]

For ligand-soaked crystals, all data sets were collected at the Stanford Synchrotron Radiation Lightsource (SSRL, beamline 12-2) at 105 K. Reflections were integrated and scaled in HKL-2000.[32] Structures were solved by molecular replacement using *Phaser*[16] and the well-refined structure model of native XynC reported in this work and refined as described above. Ligand-bound structures suffered from having a highly disordered chain D with respect to the native structure, with the structure model derived from 4-*O*-methyl-aldotetrauronate soaking being worse than the GA soak. Having prior knowledge from the native structure as to the existence of chain D, we placed it using the *Superpose* tool[33] in *Coot* and began refinement. Side chains were removed for which there was no corresponding Cα trace at low (<0.8) map σ levels ($2F_o - F_c$) but were preserved as long as the Cα trace was intact. Several small loop regions were removed based on the lack of density. For both ligand-bound structures, the overall chain D occupancy of 0.8 was found to minimize the R_{free}. In the final rounds of refinement, applying tight non-crystallographic symmetry restraints between sections of chains A, B and C and to a limited portion of chain D resulted in an $R - R_{free}$ difference of less than 5%. Models of XynC were analyzed using a variety of software tools. These include *HBPlus*,[34] LIGPLOT,[35] MolProbity[36] and PISA[37] to identify protein–ligand contacts. Models were visualized and studied, and images were prepared using PyMOL[38] and Chimera.[39]

Amino acid sequence comparison and phylogenetic analysis

Sequences with similarity to XynC were obtained using the *BLASTp* tool of the UniProt database.[40] Sequences were trimmed to match their *BLASTp* alignment with XynC, thereby removing N-terminal secretion signal sequences

and additional separately folding domains. In all cases, the length of the homologous region was very similar to the length of mature secreted XynC. Several sequences were rejected based on either high level of sequence identity with other collected sequences or sequence irregularities such as truncations resulting in partial sequences. In all, 30 sequences were collected for analysis.

Accession numbers

For all models reported in this work, coordinates and structure factors have been deposited in the PDB with the accession numbers 3KL0, 3KL3 and 3KL5 for the native, GA-bound and $MeGX_2$-bound models, respectively.

Acknowledgements

We would like to thank the University of Maryland X-ray Crystallography Shared Service for crystallographic resources. J.C.H. is supported by the National Science Foundation EPSCoR under grant no. P20 RR-016461. Portions of this research were carried out at the SSRL, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences.

References

1. Banerjee, G., Scott-Craig, J. S. & Walton, J. D. (2010). Improving enzymes for biomass conversion: a basic research perspective. *Bioenergy Res.* **3**, 82–92.
2. Beg, Q. K., Kapoor, M., Mahajan, L. & Hoondal, G. S. (2001). Microbial xylanases and their industrial applications: a review. *Appl. Microbiol. Biotechnol.* **56**, 326–338.
3. St John, F. J., González, J. M. & Pozharski, E. (2010). Consolidation of glycosyl hydrolase family 30: a dual domain 4/7 hydrolase family consisting of two structurally distinct groups. *FEBS Lett.* **584**, 4435–4441.
4. Hurlbert, J. C. & Preston, J. F., III (2001). Functional characterization of a novel xylanase from a corn strain of *Erwinia chrysanthemi*. *J. Bacteriol.* **183**, 2093–2100.
5. St John, F. J., Rice, J. D. & Preston, J. F. (2006). Characterization of XynC from *Bacillus subtilis* subsp. *subtilis* strain 168 and analysis of its role in depolymerization of glucuronoxylan. *J. Bacteriol.* **188**, 8617–8626.
6. Vrsanská, M., Kolenová, K., Puchart, V. & Biely, P. (2007). Mode of action of glycoside hydrolase family 5 glucuronoxylan xylanohydrolase from *Erwinia chrysanthemi*. *FEBS J.* **274**, 1666–1677.
7. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P. & Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl Acad. Sci. USA*, **92**, 7090–7094.
8. Vocadlo, D. J., Davies, G. J., Laine, R. & Withers, S. G. (2001). Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature*, **412**, 835–838.
9. Koshland, D. E. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. Cambridge Philos. Soc.* **28**, 416–436.
10. McCarter, J. D. & Withers, S. G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **4**, 885–892.
11. Larson, S. B., Day, J., Barba dela Rosa, A. P., Keen, N. T. & McPherson, A. (2003). First crystallographic structure of a xylanase from glycoside hydrolase family 5: implications for catalysis. *Biochemistry*, **42**, 8411–8422.
12. Henrissat, B. & Romeu, A. (1995). Families, superfamilies and subfamilies of glycosyl hydrolases. *Biochem. J.* **311**, 350–351.
13. Jenkins, J., Lo Leggio, L., Harris, G. & Pickersgill, R. (1995). β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. *FEBS Lett.* **362**, 281–285.
14. Nagano, N., Porter, C. T. & Thornton, J. M. (2001). The $(\beta\alpha)_8$ glycosidases: sequence and structure analyses suggest distant evolutionary relationships. *Protein Eng.* **14**, 845–855.
15. Lindner, C., Stulke, J. & Hecker, M. (1994). Regulation of xylanolytic enzymes in *Bacillus subtilis*. *Microbiology*, **140**, 753–757.
16. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *Phaser* crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
17. Kolenová, K., Vrsanská, M. & Biely, P. (2006). Mode of action of endo-β-1,4-xylanases of families 10 and 11 on acidic xylooligosaccharides. *J. Biotechnol.* **121**, 338–345.
18. Biely, P., Vrsanská, M., Tenkanen, M. & Kluepfel, D. (1997). Endo-β-1,4-xylanase families: differences in catalytic properties. *J. Biotechnol.* **57**, 151–166.
19. Pell, G., Taylor, E. J., Gloster, T. M., Turkenburg, J. P., Fontes, C. M. G. A., Ferreira, L. M. A. *et al.* (2004). The mechanisms by which family 10 glycoside hydrolases bind decorated substrates. *J. Biol. Chem.* **279**, 9597–9605.
20. Settiner, W. J. & Marchess, R. H. (1965). Derivation of possible chain conformations for poly-β-1,4-anhydroxylose. *J. Polym. Sci., Part C: Polym. Symp.* **11**, 253–264.
21. Sundararajan, P. R. & Rao, V. S. R. (1969). Conformational studies of β-D-1,4-xylan. *Biopolymers*, **8**, 305–312.
22. Marchessault, R. H. & Liang, C. Y. (1962). The infrared spectra of crystalline polysaccharides. 8. Xylans. *J. Polym. Sci.* **59**, 357–378.
23. Charnock, S. J., Spurway, T. D., Xie, H., Beylot, M. H., Virden, R., Warren, R. A. J. *et al.* (1998). The topology of the substrate binding clefts of glycosyl hydrolase family 10 xylanases are not conserved. *J. Biol. Chem.* **273**, 32187–32199.

24. Pozharski, E. (2010). Percentile-based spread: a more accurate way to compare crystallographic models. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 970–978.
25. Devos, D. & Valencia, A. (2000). Practical limits of function prediction. *Proteins*, **41**, 98–107.
26. Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Eng.* **12**, 85–94.
27. Whisstock, J. C. & Lesk, A. M. (2003). Prediction of protein function from protein sequence and structure. *Q. Rev. Biophys.* **36**, 307–340.
28. Gallardo, O., Fernandez-Fernandez, M., Valls, C., Valenzuela, S. V., Roncero, M. B., Vidal, T. *et al.* (2010). Characterization of a family GH5 xylanase with activity on neutral oligosaccharides and evaluation as a pulp bleaching aid. *Appl. Environ. Microbiol.* **76**, 6290–6294.
29. St John, F. J., Godwin, D. K., Preston, J. F., Pozharski, E. & Hurlbert, J. C. (2009). Crystallization and crystallographic analysis of *Bacillus subtilis* xylanase C. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **65**, 499–503.
30. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.
31. Emsley, P. & Cowtan, K. (2004). *Coot*: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
32. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
33. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2256–2268.
34. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.
35. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134.
36. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X. *et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383.
37. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
38. DeLano, W. L. (2002). *The PyMOL User's Manual.* DeLano Scientific, Palo Alto, CA.
39. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
40. Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, N., Suzek, B. E. *et al.* (2009). Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics*, **10**, 136.