A plasmid borne, functionally novel glycoside hydrolase family 30 subfamily 8 endoxylanase from solventogenic *Clostridium*

Franz J. St John¹, Diane Dietrich¹, Casey Crooks¹, Peter Balogun¹, Vesna de Serrano², Edwin Pozharski³, James Kennon Smith⁴, Elizabeth Bales⁴ and Jason Hurlbert⁴

¹Institute for Microbial and Biochemical Technology, Forest Products Laboratory, USDA Forest Service, Madison, WI, U.S.A.; ²Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD, U.S.A.; ³Institute for Bioscience and Biotechnology Research, Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Rockville, MD, U.S.A.; ⁴Department of Chemistry, Physics and Geology, Winthrop University, Rock Hill, SC, U.S.A.

Correspondence: Franz J. St John (fjstjohn@gmail.com)

Glycoside hydrolase family 30 subfamily 8 (GH30-8) β-1,4-endoxylanases are known for their appendage-dependent function requiring recognition of an α-1,2-linked glucuronic acid (GlcA) common to glucuronoxylans for hydrolysis. Structural studies have indicated that the GlcA moiety of glucuronoxylans is coordinated through six hydrogen bonds and a salt bridge. These GlcA-dependent endoxylanases do not have significant activity on xylans that do not bear GlcA substitutions such as unsubstituted linear xylooligosaccharides or cereal bran arabinoxylans. In the present study, we present the structural and biochemical characteristics of xylanase 30A from *Clostridium acetobutylicum* ([*Ca*]Xyn30A) which was originally selected for study due to predicted structural differences within the GlcA coordination loops. Amino acid sequence comparisons indicated that this Gram-positive-derived GH30-8 more closely resembles Gram-negative derived forms of these endoxylanases: a hypothesis borne out in the developed crystallographic structure model of the [*Ca*]Xyn30A catalytic domain ([*Ca*]Xyn30A-CD). [*Ca*]Xyn30A-CD hydrolyzes xylans to linear and substituted oligoxylosides showing the greatest rate with the highly arabinofuranose (Araf)-substituted cereal arabinoxylans. [*Ca*]Xyn30A-CD hydrolyzes xylooligosaccharides larger than xylotriose and shows an increased relative rate of hydrolysis for xylooligosaccharides containing α-1,2-linked arabinofuranose substitutions. Biochemical analysis confirms that [*Ca*]Xyn30A benefits from five xylose-binding subsites which extend from the −3 subsite to the +2 subsite of the binding cleft. These studies indicate that [*Ca*]Xyn30A is a GlcA-independent endoxylanase that may have evolved for the preferential recognition of α-1,2-Araf substitutions on xylan chains.

**Introduction**

Enzymes of glycoside hydrolase (GH) family 30 (GH30, previously classified in GH5 [1]) include several subfamilies known to hydrolyze various different biomass derived carbohydrate polymers. These enzymes consist of a (β/α)₈-barrel with an obligatory side-associated, nine-stranded, aligned β-sandwich [1,2]. This side β-sandwich structure is tightly associated with the (β/α)₈-barrel catalytic core domain through hydrophobic contacts. These separate structural folds are connected together through a dual-linker as the first β-strands of the side β-sandwich derive from the N-terminal sequence and the remaining from the C-terminal sequence therefore establishing two amino acid main-chain tethers to the β-sandwich domain from the (β/α)₈-barrel. Phylogenetic analysis with structure considerations of GH30 enzymes predicted that these two domains may function correctly only when together. The broad substrate specificity observed in the GH30 family, which contains at least eight functionally distinct subfamilies, may be due to the increased plasticity of the catalytic core.
nucleophile region which is buttressed by the additional side β-structure [1]. Support for this consideration was obtained through activity measurements of side β-structure truncation products of the GH30-8 xylanase from *Paenibacillus barcinonensis* [3].

Subfamilies 7 and 8 of GH30 (GH30-7 and GH30-8, respectively) have been functionally characterized to hydrolyze xylan substrates [4, 5]. Members of these two phylogenetically distinct subfamilies typically share less than 25% amino acid sequence identity. While GH30-7 xylanases have been shown to represent diverse xylanolytic function [6, 7] and occur primarily in fungi and some Actinobacteria, GH30-8 xylanases have been shown to be appendage-dependent β-1,4-endoxylanases having a strict requirement for the recognition of the α-1,2-linked gluconic acid (GlcA) substitution of gluconoxylans [8, 9] and occur primarily in bacteria. These unique endoxylanases fix the position of the xylan chain into the subsites of the xylan-binding cleft through specific recognition of a GlcA moiety in the −2b subsite (Figure 1A, B). The resulting hydrolysis limit product consists of a variable set of xylooligosaccharides each substituted with a single GlcA positioned penultimate to the reducing terminal xylose (Supplementary Figure S1A).

Biochemical analyses were performed with xylanase C from the Gram-positive bacterium *Bacillus subtilis* (BbXynC) [8] and xylanase A from the Gram-negative bacterium *Erwinia chrysanthemi* (EcXynA) [9]. These enzymes share only 40% amino acid sequence identity, but functional studies did not distinguish any difference in their mode of action or substrate specificity. Structural studies of these distinct GH30-8 endoxylanases were performed to obtain ligand-bound structures. These separate studies identified the role of the β7–α7 and β8–α8 loop regions in specific coordination of the GlcA appendage [10, 11]. The GlcA substituted on the xylose positioned in the −2a subsite [12] is coordinated with a salt bridge and as many as six hydrogen bonds (Figure 1A, the −2b subsite). These enzymes proved to be highly similar in the regions important for GlcA coordination. This region constitutes the C-terminal face of the negative subsite side of the substrate-binding cleft and is dominated by the β7–α7 and β8–α8 loop regions. The structures also defined distinct features in the N-terminal face of the substrate-binding cleft which correlated with the Gram-type of the bacterium [10]. These differences are primarily localized to the β3–α3 and β4–α4 loop regions and involved how the β3–cα3 loop interacts to stabilize the β4–α4 loop. The significance of this distinction is due to the importance of the β4 strand as it contains the catalytic acid/base catalysts. Structural comparison coupled with extensive phylogenetic analysis all reveals that the β3–α3 and β4–α4 loop regions are distinct between GH30-8 endoxylanases which derive from Gram-positive vs. Gram-negative bacteria.

In this work, we have characterized a novel endoxylanase representing a small subset of GH30-8 xylanases found in solventogenic species of the genus *Clostridium*. We find that xylanase 30A from *Clostridium acetobutylicum* (CaXyn30A) is a GlcA-independent GH30-8 endoxylanase reminiscent of the recently reported xylanase 30A from *Clostridium papyrosolvens* (CpXyn30A). In contrast with this previous enzyme which functions only very poorly [13], CaXyn30A has a much greater rate of hydrolysis on all tested xylan substrates. Functional studies indicate that CaXyn30A has a preference for hydrolysis of arabinoxylans over the less complex xylan forms represented by gluconoxylan and neutral xylooligosaccharides. Results indicate that CaXyn30A does not possess the GlcA-dependent activity displayed by the canonical GH30-8 endoxylanases, but instead, appears to be dependent on the O-2 linked Araf substitutions on the xylose in the −2a subsite. This new endoxylanase type may provide beneficial attributes in industrial applications distinct from the commonly utilized endoxylanases which derive from GH families 10 and 11.

**Materials and methods**

**Amino acid sequence bioinformatics**

Amino acid sequence studies were performed following the collection of diverse homologs of GH30-8 endoxylanases from the UniProt Database [14]. Sequences were trimmed using MEGA 7 [15] and multiple sequence alignments were generated using MAFFT [16]. These alignments were loaded into MEGA 7 and a Boot-Strapped Maximum-Likelihood phylogenetic consensus tree was generated [17, 18]. Levels of amino acid sequence identity were determined using the PRSS sequence shuffling tool (http://fasta.bioch.virginia.edu/fasta_www2/fasta.www.cgi?rm=shuffle) of the FASTA online package available through the University of Virginia (http://fasta.bioch.virginia.edu/fasta/fasta_list.html). Alignment figures were generated using the ESPript web server [19] available at http://espirpt.ibcp.fr/ESPript/ESPript/index.php. The studies involved in this work rely on the continued maintenance of the CAZy Database (http://www.cazy.org/) [20] for analysis and functional correlation of enzyme family members.
Reagents

All reagents and xylanase substrates were of the highest purity available. Low-viscosity wheat arabinoxylan (WAX, Cat. No. P-WAXYL), the xylooligosaccharides X2 through X6, and several specific Araf-oligoxylosides including 23-\(\alpha\)-L-arabinofuranosyl-xylotriose (A2XX, Cat. No. O-A2XX), 33-\(\alpha\)-L-arabinofuranosyl-xylotetraose (XA3XX, Cat. No. O-XA3XX), and 23,33-di-\(\alpha\)-L-arabinofuranosyl-xylotetraose (XA2+3XX, Cat. No. O-XA23XX) were obtained through Megazyme International (Wicklow, Ireland). The hardwood glucuronoxylan from beech weed (BWX, Cat. No. X4252) were obtained from Sigma-Aldrich (St. Louis, MO). The shorthand nomenclature used to describe the complex oligosaccharides has been previously described [21,22]. In a slight deviation, for the naming of structurally defined aldouronates, an \(\alpha\)-1,2-linked GlcA of glucuronoxylans is shown [11]. A schematic of the subsites in GH30-8 endoxylanases to illustrate oligoxyloside–enzyme interaction.

Expression and purification of \(CaXyn30A\)-CD

The codon optimized (for \(Escherichia coli\)) coding sequence for the \(CaXyn30A\)-CD was synthesized by DNA2.0 and subsequently subcloned into pET28 (pCatXyn30A-CD noHis) for expression. The cloned region replaced the N-terminal secretion signal sequence with a methionine and extended to include the full two-motif fold of the GH30 enzyme, but truncating the wild-type \(CaXyn30A\) amino acid sequence by not including the natively
encoded CBM13 domain. Chemically competent E. coli BL21 (DE3) was transformed with this protein expression vector and selected for on LB agar plates containing 50 μg/ml kanamycin. The CaXyn30A-CD protein expression was performed as described previously and as outlined in the pET System Manual, 10th Ed. [13,23].

Cell pellets from 0.5 l of expression culture were thawed partially at room temperature and then on ice. An EDTA-free Mini cOmplete protease inhibitor tablet (Roche) was added to one of these pellets. The cell pellets were suspended in 8 ml of 25 mM Tris–HCl (pH 7.1). Resuspended pellets were combined and lysozyme was added to final concentration of 20 μg/ml. The volume was transferred to a glass rosette sonication vial and allowed to cool on ice for 15 min. A sonic microtip (Misonix Sonicator) was submersed one inch below the surface of the cell suspension. Cell lysis was achieved by the application of 95 W of power in 12–10 s pulses with 50 s rest between pulses while in ice water. Following sonication, 1 M MgCl₂ was added to a final concentration of 2 mM to the lysate and lysozyme was added as above to obtain a final lysozyme amount of 40 μg/ml. A total of 250 U of Benzonase (EMD Millipore, Billerica, MA) was added and the cell lysate was incubated on a rocker at room temperature for 30 min. The lysate was then centrifuged at 15°C for 30 min at 23 500×g. The supernatant was then filtered through a 0.45 μm syringe tip filter to prepare the cell-free extract (CFE).

Purification and processing of CaXyn30A-CD used Tris–HCl-based buffers with a BioLogic Duo-Flow medium pressure chromatography system (Bio-Rad, Hercules, CA). The CFE was fractionated on a 5 ml Econo-Pac CM column (Bio-Rad) equilibrated in 25 mM Tris–HCl, pH 7.1, with a 10 column volume linear gradient from 0 to 500 mM NaCl. The eluted protein peak was collected and concentrated using an Amicon Ultra 15 with the 10 K MWCO membrane (EMD Millipore). The concentrated protein preparations were then desalted using a 5 ml Econo-Pac P-6 desalting column (Bio-Rad) and then again concentrated with another Amicon Ultra 15 10 K MWCO centrifugal concentrator. It was observed that the concentration of CaXyn30A-CD in high salt resulted in a low level of precipitate, so prior to concentrating for subsequent desalting the protein sample was further diluted with 25 mM Tris–HCl (pH 7.1). The preparation was then purified on a Superdex 200PG 16/600 column (GE Healthcare Bio-Sciences, Pittsburgh, PA) equilibrated with 25 mM Tris–HCl (pH 7.5) and 100 mM NaCl. Peak fractions were combined and concentrated as above and subsequently buffer-exchanged using an Amicon Ultra 15 into 25 mM Tris–HCl (pH 7.1). The remaining NaCl is estimated at <5 mM. The protein concentration was determined using the ProtParam predicted extinction coefficient of 87 780 M⁻¹ cm⁻¹, assuming that all cysteines are in the reduced form [24].

Protein X-ray crystallography

Aliquots of the CaXyn30A-CD protein at 40 mg/ml in 25 mM Tris–HCl (pH 7.1) were used for crystal screening for crystallographic structure determination. Sitting-drop Art Robbins 96-well Intelliplates (three-drop version) were set using an Oryx Nano crystallization robot (Douglas Instruments, Hungerford, U.K.) [25]. Data was collected at the Stanford Synchrotron Radiation Light Source on beamline 7-1 at 105 K with a wavelength of 1 Å. Diffraction images were integrated and scaled using the HKL2000 package [26]. Phasing was performed using the program Phaser [27], with the protein crystal model of BsXynC (PDB code: 3KL3). The structure model was iteratively refined through cycles of maximum-likelihood restrained refinement in Refmac [28] and real-space refinement in Coot [29] all as part of the CCP4 Suite of macromolecular structure processing software [30]. Model quality was accessed using Molprobity [31]. The refined model data were deposited through the Rutgers Center for Structural Biology into the Worldwide Protein Data Bank and was assigned the accession PDB code: 5CXP. Ligand protein contacts were analyzed in LigPlot [32]. Structures were analyzed, compared, and images generated using PyMOL [33].

Preparation of substrates

A soluble fraction of BWX was prepared by dissolving the xylan in water at 30 mg/ml with heating at 60°C with stirring for 5 h. Following room temperature equilibration, the supernatant was collected after centrifugation for 30 min at 34 700×g and 15°C. The sugar concentration was gravimetrically quantified following 2 days of drying at 60°C under vacuum. This clarified xylan was diluted to 20 mg/ml with water and used as the stock for enzyme studies. WAX stocks were prepared at 10 mg/ml as directed by Megazyme product data sheets. Xylooligosaccharide solutions were dissolved in water to an approximate concentration of 50 mM. The sugar concentrations were determined by the phenol–sulfuric acid assay [34] against a xylose standard curve and in the case of Araf-oligoxosylsides of known structure a standard curve composed of the specific ratio of xylose to arabinose.
Analysis of xylans and xylooligosaccharides

Compositional analysis of WAX and BWX was performed by hydrolysis of the polysaccharide for 60 min at 121°C (autoclave) in 1% H2SO4 followed by analysis with a Dionex ICS-3000 DP HPLC system running a PA-100 Carbo Pac column (4 x 250 mm) equipped with a guard column (4 x 50 mm) at a flow rate of 1 ml/min maintained at 22°C with water as eluent and detection with a pulsed amperometric detector, as previously described [35]. The extent of hydrolysis was validated by the negligible detectable levels of furfural and no detectable oligomeric sugars. The resistance of aldobiuronate to acid hydrolysis results in the accumulation of this GlcA-containing sugar in acid-hydrolyzed BWX; however, neither GlcA nor aldouronates are quantified by the described Dionex protocol. The concentration of GlcA was determined using the Blumenkrantz assay as described by Filisetti-Cozzi and Carpita [36]. The degree of substitution (DS) of glucuronoxylans was calculated according to the formula: (mM xylose + mM GlcA)/mM GlcA, which assumes that GlcA is available only as aldouronate.

Biochemical characterization of CaXyn30A-CD

Hydrolysis of polymeric xylan by CaXyn30A-CD was quantified using the Nelson’s test [37,38] for the measurement of reducing end concentration. Enzyme optimization studies of CaXyn30A-CD were performed in a 10 min reaction using BWX and acetate buffer over a range of temperatures from 4 to 60°C. A range of acetate buffers from pH 2.75 to 5.5 were tested using a reaction temperature of 40°C. Following these initial studies, all subsequent reactions contained 100 mM sodium acetate, pH 4.0 and 0.1 mg/ml bovine serum albumin, and were incubated at 40°C. Studies involving preincubations were performed using an MJ Research thermal cycler (Bio-Rad, Hercules, CA) with thin-walled 250 µL PCR tubes. For temperature stability studies, the thermal cycler heating block temperature was confirmed across all 96-well positions using an external digital temperature probe.

Progress curve assays were set up as 5 ml reactions in 13 x 100 mm test tubes. BWX was assayed at 10 mg/ml, while WAX was assayed at 7.5 mg/ml. Reactions were initiated by the addition of 100 µl of appropriately diluted enzyme and were sampled according to a predetermined timing and dilution level to target results to a xylose best-fit line standard. The reaction was mixed by gentle vortexing every few minutes after initiation. Progress curve results represent a minimum of two datasets. Specific activity determinations for WAX and BWX substrates were conducted in three separate 10-min reactions, each performed in triplicate which were then averaged to obtain the final activity values.

For specific activity measurements of xylooligosaccharides, a master reaction mixture was prepared and subsequently subdivided for multiple (25 µl) independent reactions along with a no-enzyme control reaction. The oligoxygenide substrate concentrations varied slightly due to a post assay adjustment for the anhydropentose content. Hexameric oligoxylosides including X₆ and XA²⁺X₃XX and pentameric oligoxylosides including X₅ and XA³XX were assayed at 10.9 mM, and tetrameric oligoxylosides including X₄ and A²XX were assayed at 11.0 mM. Reactions were initiated by the addition of enzyme and killed by incubation at 90°C for 10 min. For these reactions, the enzyme level and reaction time were as follows: X₆, 194.4 pmol/ml for 10 min; X₅, 194.4 pmol/ml for 10 min; X₄, 388.8 pmol/ml for 90 min; X₃, 777.6 pmol/ml for 90 min; XA²⁺X₃XX, 194.4 pmol/ml for 10 min; XA³XX, 388.8 pmol/ml for 65 min, and A²XX, 388.8 pmol/ml for 90 min. HPLC analyses were performed using an Agilent 1260 Infinity with xylooligosaccharides separated on a Phenomenex (Torrance, CA) RNO column using water as eluent at 0.3 ml/min and a column temperature of 80°C, with the refractive index of the column eluate monitored continuously during the separation. All specific activities except for A²XX, see below) were averaged from a minimum of two serial studies, each consisting of at least two separate master reaction mix derived parallel reactions which were analyzed by HPLC in triplicate. Loss of substrate relative to the no-enzyme control reaction was taken as the measure of activity and 10 min reactions were optimized to allow for greater than 25% oligoxyloside conversion. To establish the appropriateness of the chosen method, product inhibition was considered as described above but with the inclusion of 5 mM X₂ and X₃. No significant inhibition by these smaller xylooligosaccharide limit products was detected. Kinetic evaluation of CaXyn30A-CD was performed using a standard Michaelis–Menten kinetic model.

Analysis of the rate of para-nitrophenol (pNP) release from the oligoxylosides pNP-xylotriose (pNP-X₃) and pNP-xyloolriose (pNP-X₄) was performed in 200 µl volumes in the established CaXyn30A-CD reaction conditions. Substrate concentrations of 7.5 mM were used and the reactions were stopped by the addition of 800 µl
Ca2+in preliminary 10 min reactions at different temperatures, 40°C was selected as the optimal reaction temperature. All subsequent reactions were performed in 100 mM sodium acetate, pH 4.0. Based on activity measurements stable at these pH values, with a minimal loss of activity after a 15 h incubation (Supplementary Figure S2C).

Results and discussion

Amino acid sequence-based characteristics of CaXyn30A

The CaXyn30A xylanase (UniProt ID: Q97TI2) is encoded within the pSOL1 megaplasmid of C. acetobutylicum. Previous studies of this bacterium have attributed its industrially important solvent production capabilities to this megaplasmid [40]. Comparative amino acid sequence analysis confidently classifies CaXyn30A as an enzyme in subfamily 8 of the GH30 enzymes (Figure 2A). The DNA sequence for the xyn30A gene encodes the predicted GH30-8 CD along with a C-terminal family 13 carbohydrate-binding domain. CaXyn30A was originally selected for study based on differences in the typically conserved β7–α7 and β8–α8 loop regions which are known to be important for the specificity of these GlcA-dependent endoxylanases [8–11]. From sequence comparisons, it can be seen that CaXyn30A has a gap in place of the partially conserved β7–α7 loop region and a novel, divergent sequence in place of the highly conserved 'RR-motif' of the β8–α8 loop region (Figure 2B). A recently characterized GH30-8 xylanase from C. papyrosolvens, CpXyn30A, also showed similar sequence alignment characteristics and was selected for study for the same reason [13]; however, the specific sequence of the loop regions in CaXyn30A and CpXyn30A are dissimilar (Figure 2B) and their overall functional characteristics are not similar [13], with the altered loop region found in CpXyn30A being predicted to have disrupted efficient hydrolysis of xylan substrates by this endoxylanase.

CaXyn30A-CD expression, purification, and reaction conditions

Expression of CaXyn30A-CD from the codon-optimized expression vector pCaXyn30A-CD yielded soluble protein. Taking advantage of the high pl, purification was achieved using a carboxymethyl-cellulose cation-exchange column. This approach yielded a single well-resolved elution peak (Supplementary Figure S2A) and was followed by gel filtration chromatography to prepare the protein sample for crystal screening studies. From SDS–PAGE analysis (Supplementary Figure S2B), CaXyn30A-CD was judged to be >95% pure following the cation-exchange purification and this improved through Superdex gel filtration.

Basic biochemical analyses of CaXyn30A-CD were performed to identify the optimal reaction conditions for the enzyme. Initial analysis showed that the enzyme benefits from the inclusion of bovine serum albumin, and based on titration studies (data not shown), the optimal concentration of BSA was found to be 0.1 mg/ml. The optimal reaction pH was determined using sodium acetate buffers ranging from a pH of 2.75 through a pH of 5.5. These data indicate that CaXyn30A-CD has an optimal reaction pH of ~3.75. Interestingly, CaXyn30A-CD maintains ~92% and 68% of its maximum activity at pH 2.75 and 4.75, respectively. It was also found to be stable at these pH values, with a minimal loss of activity after a 15 h incubation (Supplementary Figure S2C). All subsequent reactions were performed in 100 mM sodium acetate, pH 4.0. Based on activity measurements in preliminary 10 min reactions at different temperatures, 40°C was selected as the optimal reaction temperature (data not shown). Thermal stability studies indicate that for a 1.5 h incubation at increased temperatures, CaXyn30A-CD is stable up to 50°C and maintains 85% of its activity over this time period while kept at 55°C (Supplementary Figure S2D).
Hydrolysis of polymeric xylans by CaXyn30A-CD

Initial biochemical analyses of CaXyn30A-CD focused on hydrolysis of glucuronoxylan and identified that unlike the canonical GlcA-dependent GH30-8 endoxylanases, CaXyn30A-CD produced neutral sugar hydrolysis products such as xylobiose and xylotriose. It was later found for CaXyn30A-CD that WAX, with its high degree of Araf substitution, also served as a substrate. As shown in Supplementary Figure S3, hydrolysis of WAX appears to proceed at least as efficiently as the BWX. For the hydrolysis of WAX, Araf-substituted xyloo-ligosaccharides are detected at high enzyme dilutions, and low relative levels of xylobiose and xylotriose were detected at the higher concentrations of enzyme. A likely explanation for this finding is that due to the substantially greater substitution level of WAX relative to a typical glucuronoxylan, there are fewer linear oligosaccharides that can be generated. Of the available Araf-oligosaccharide standards, only the hydrolysis product corresponding...
to the standard XA^3XX appears to be produced by CaXyn30A-CD, while the standards XA^2+3XX and A^2XX do not definitively result from the action of CaXyn30A-CD. Interestingly, two seemingly smaller unknown Arf-oligoxylosides appear even at the lower enzyme loads and remain present over the course of hydrolysis (Supplementary Figure S3). Hydrolysis of BWX by high enzyme dilutions of CaXyn30A-CD yielded detectable levels of neutral xylooligosaccharides distinguishable up to xylotetraose (but no xylose) with no small aldouronates identified. For BWX, we found that neutral oligoxylosides are produced early and aldouronates accumulate later with the eventual formation of those equivalent in size to aldotriuranone, aldottedrauranone, and aldopentauronate.

Progress curves were performed to compare hydrolysis of the model arabinoxylan and glucuronoxylan substrates by CaXyn30A-CD (Figure 3A). CaXyn30A-CD appears to hydrolyze the more complex WAX (Supplementary Figure S1C) at a higher rate than the less complex BWX (Supplementary Figure S1D). These analyses suggest a role for recognition of arabinofuranose groups along the xylan chain for optimal function. To provide perspective, we compared these results with hydrolysis by a generic GH10 xylanase. Xyn10B from C. mixtus (CmXyn10B) was used as a model GH10 and is expected to hydrolyze the xylan chain in unsubstituted regions and, as such, should have a lower activity when presented the more highly substituted WAX substrate. Our hydrolysis results support this expected function showing approximately a two-fold greater activity on BWX than on WAX (Figure 3B). Kinetic analysis of CaXyn30A-CD with WAX and BWX confirmed that WAX was a preferred substrate (Table 1). While the apparent K_m of CaXyn30A-CD for the WAX substrate was only slightly lower than the K_m of BWX, the apparent V_max for the WAX substrate was 287 U/mg, roughly two-fold greater than the V_max measured for BWX. This corresponds to greater than a two-fold increased turn-over number. The apparent specificity constant for WAX is 4180 compared with 1660 for BWX verifying the CaXyn30A-CD’s preference for WAX over BWX (Table 1). Importantly, while most of these Arf substitutions are substituted α-1,3, it is believed that a significant portion is double-substituted bearing Arf substitutions on both the α-1,3 and α-1,2 positions [22]. As verified in this study, the difference in the DS between WAX and BWX is large with WAX having ~1 Arf for every 1.5 xylose residues (DS of 0.66) and BWX having 1 GlcA for every 12 xyloses (DS of 0.083) [8,22]. This difference in the DS translates to a large increase in the level of complexity for a endoxylanase which cleaves the main-chain β-1,4-xylan backbone. However, the measured specific activity for these two polymers (Table 2), the determined calculated kinetic constants (Table 1), and the progress curves (Figure 3A) all indicate that CaXyn30A prefers the complex cereal arabinoxylan over the less complex hardwood glucuronoxylan.

Hydrolysis of xylooligosaccharides by CaXyn30A-CD

Hydrolysis of xylooligosaccharides shows that CaXyn30A-CD has its highest rate on the largest available oligoside, X_4, with a 17% decline observed for X_3, a further 93% decline to 2.79 U/mg on X_4, and no detectable activity on X_5 (Table 2). This pattern indicates that CaXyn30A-CD benefits from at least five distinct xylan-binding subsites [12]. TLC hydrolysis product analysis of these oligosides showed that hydrolysis of X_6 resulted in the products X_2, X_3, and X_4, a result anticipated for a typical endoxylanase (Figure 4). Hydrolysis of X_5 resulted in the release of X_2 and X_3 and hydrolysis of X_4 resulted in xylose, X_2 and X_3 (Figure 4). The results for X_4 hydrolysis indicate that it is hydrolyzed in two different ways. Hydrolysis of X_3 was not detected, indicating that this sugar either does not stably associate into the substrate-binding cleft or that its binding forms a nonproductive complex. Nonproductive complex formation could result with binding from the −1 through +2 subsites or through binding in the negative subsite region only, from subsite −3 through −1 (Figure 1B).

To obtain more specific information regarding the hydrolysis of X_4 and the balance of affinity within the xylan-binding subsites, we monitored the hydrolysis of X_4 over time. In the study depicted in Figure 5, it has shown that X_3 is produced more often than X_2. From the slope of the best-fit lines, it can be calculated that ∼65% of the X_4 hydrolytic events produce X_3. This would most probably be due to X_4 binding from subsites −3 through +1. Conversely, binding from the −2 subsite through the +2 subsite which would yield just X_2 occurs in ∼35% of the hydrolytic events.

For the hydrolysis of X_4 to X_3 and xylose, an unlikely alternative to the proposed binding from the −3 through +1 subsites is binding from the −1 through the +2 subsite. This would only accommodate three xylose moieties of X_4 because a +3 subsite is not thought to exist in CaXyn30A. Owing to this, the total binding energy for this scenario would probably be less than the stable association of all four xylose moieties either binding the −3 through +1, to yield the X_3 and xylose or −2 through +2 to yield X_2. Furthermore, GH30

© 2018 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY-NC-ND).
endoxylanases are similar to GH10 endoxylanases both belonging to CAZy Clan A (4/7 hydrolases), consisting of (β/α)_{8}-barrel structures, displaying endo-function, and having the same double-displacement catalytic mechanism which retains the anomeric configuration of the newly generated reducing terminus [20, 41–44]. GH10 endoxylanases are well characterized and are known to not hydrolyze a single nonreducing xylose occupying the −1 subsite extending from the positive subsites, supporting the likelihood that this also would not be likely for GH30 endoxylanases [45,46].

Based on the X4 hydrolysis results, the −3 subsite is thought to contribute significant stabilizing interactions. To consider this further, we analyzed the hydrolysis of ρNP-X2 and ρNP-X3 as analogs to xylotriose and xylo-tetraose, respectively (Table 2). It was anticipated that the artificial nitrophenol moiety would stabilize the interaction in the +1 subsite through tighter hydrophobic interactions. The measured activity showed a 10-fold increase for ρNP-X3 over ρNP-X2, confirming the importance of the −3 subsite in endo-hydrolysis of xylan and xylooligosaccharides (Table 2).

Based on the increased rate of hydrolysis of WAX over BWX, we obtained the specialty AraF-substituted xylooligosaccharides XA^{2+3}XX, XA^{1}XX, and A^{2}XX (Supplementary Figure S1E–G, respectively) to better dissect the contribution of the AraF substitutions in binding the xylan chain [21,22]. It is anticipated, based on the expected orientation, that xylan takes in the substrate-binding cleft of GH30-8 endoxylanases that for these defined AraF-oligoxyllosides, the substituted xylose would position into the −2 subsite allowing its C2 and/or C3 AraF-substituted hydroxyls to extend upward (Figure 1A) out of the enzyme and thus position the reducing terminal xylose within the +1 subsite (Figure 1B). Confirmation of this result is obtained by the specific release of xylose as shown by TLC (Figure 4). The doubly substituted xylotetraose analog (XA^{2+3}XX) resulted in a specific activity of ∼50 U/mg CaXyn30A-CD, an increase of ∼18-fold over the rate of unsubstituted X4. This suggests that its aggregate binding energy was similar to X6 (Table 2), a substrate having a hydrolysis rate of ∼48 U/mg. This result identified a clear role for the AraF substitutions in arabinoxylan hydrolysis at least as it

### Table 1 Lineweaver–Burk kinetic analysis for CaXyn30A-CD

<table>
<thead>
<tr>
<th>Apparent kinetic constant</th>
<th>Wheat arabinoxylan (WAX)</th>
<th>Beech wood glucuronoxylan (BWX)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K_m (mg/ml)</td>
<td>2.93 ± 0.35</td>
<td>3.12 ± 0.15</td>
</tr>
<tr>
<td>V_max (U/mg)</td>
<td>287 ± 23</td>
<td>122 ± 11</td>
</tr>
<tr>
<td>k_cat (s^{-1})</td>
<td>203 ± 16</td>
<td>86 ± 8</td>
</tr>
<tr>
<td>k_cat/K_m (min^{-1} mg^{-1} ml)</td>
<td>4180 ± 170</td>
<td>1660 ± 80</td>
</tr>
</tbody>
</table>

1Substrate concentration-dependent kinetic studies were performed following several trial assays to determine the optimal substrate range. For each substrate, final values were obtained from the average of two independent studies.
maps to the negative subsite region of the substrate-binding cleft. To determine which Ara\textsubscript{f} contributed most to the overall binding energy, we measured hydrolysis of XA\textsuperscript{3XX}, which yielded a specific activity of only 3.43 U/mg, a 14.4-fold decrease from the rate of hydrolysis of XA\textsuperscript{2+3XX}. By deduction, these findings indicate that the \(\alpha\)-1,2-linked Ara\textsubscript{f} on the xylose in the \(-2\) subsite contributes to tighter binding of X4 for hydrolysis of xylose, while the \(\alpha\)-1,3-linked Ara\textsubscript{f} accommodated within this region of the enzyme active site does not lend stabilizing

Table 2 CaXyn30A-CD specific activity

<table>
<thead>
<tr>
<th>Xylan</th>
<th>Specific activity(^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat arabinoxylan</td>
<td>113.0 ± 8.3</td>
</tr>
<tr>
<td>Beech wood glucuronoxylan</td>
<td>90.9 ± 7.5</td>
</tr>
<tr>
<td>Oligoxyloside</td>
<td></td>
</tr>
<tr>
<td>Xylohexaose (X\textsubscript{6})</td>
<td>47.94 ± 1.74</td>
</tr>
<tr>
<td>Xylopentaose (X\textsubscript{5})</td>
<td>39.96 ± 2.44</td>
</tr>
<tr>
<td>Xylotetraose (X\textsubscript{4})</td>
<td>2.79 ± 0.16</td>
</tr>
<tr>
<td>Xylotriose (X\textsubscript{3})</td>
<td>None detected</td>
</tr>
<tr>
<td>(\rho)NP-(X\textsubscript{3})</td>
<td>248.7 ± 6.27</td>
</tr>
<tr>
<td>(\rho)NP-(X\textsubscript{2})</td>
<td>21.61 ± 1.18</td>
</tr>
<tr>
<td>XA\textsuperscript{2+3XX}</td>
<td>49.52 ± 3.66</td>
</tr>
<tr>
<td>XA\textsuperscript{3XX}</td>
<td>3.43 ± 0.51</td>
</tr>
<tr>
<td>A\textsuperscript{2XX}</td>
<td>1.98 ± 0.17</td>
</tr>
</tbody>
</table>

\(^1\)The measure of specific activity is provided as U/mg CaXyn30A-CD, where one Unit is equal to one \(\mu\)mole of reducing terminus generated per minute.

Figure 4. TLC analysis of the processing of xylooligosaccharides and Ara\textsubscript{f}-substituted xylooligosaccharides by CaXyn30A-CD.

Release of xylose from the Ara\textsubscript{f}-oligoxylloside indicates the association of the Ara\textsubscript{f} substitutions into the \(-2\) subsite region positioned above the \(-2\)a subsite bound xylose.
contacts, as it yields an approximately equivalent rate of hydrolysis as X4. From these observations, we hypothesize that the α-1,3-Araf substitution does not sterically interfere in X4 binding for hydrolysis in the described manner. To further investigate the role of the Araf in CaXyn30A function, hydrolysis of A2XX was studied. This small Araf-oligoxyloside yielded a low rate of ~2 U/mg (Table 2). Although low, this is a significant rate compared with the lack of detectable activity for X3 and indicates that the α-1,2-linked Araf on the nonreducing terminus favored the orientation of the underlying X4 from the −2 through the +1 subsite for the observed hydrolytic release of xylose (Figure 4). Figure 6 provides a visual synopsis of the oligoxyloside hydrolysis results.

To confirm our findings regarding the lower activity on the hardwood glucuronoxylan BWX relative to the more complex WAX arabinoxylan and also to confirm that the GlcA specificity of the canonical GH30-8 endoxylanases is not observed by CaXyn30A, we measured the hydrolysis rate of aldopentauronate (XGXX, Supplementary Figure S1H) derived from the specific hydrolysis of glucuronoxylan with a GH11 endoxylanase [13,47]. We anticipated that this aldopentauronate, X4 analog, with its penultimate nonreducing terminal α-1,2-GlcA substitution would occupy the substrate-binding cleft in a similar manner as the Araf-substituted xyloooligosaccharides described above. It is important to consider that for canonical GH30-8 GlcA-dependent xylanases, the optimum reaction pH (~6.5) indicates a requirement for the C6 carboxylate moiety of glucuronoxylan to be preferentially ionized. This is considered to be due to the specificity determining salt-bridge interaction between the conserved arginine and the C6 carboxylate [10,11]. Even so, given the low pKₐ of GlcA moieties [48], the selected optimum reaction pH for CaXyn30A-CD (pH 4.5) would very likely maintain the major portion of C6 carboxylate in the anionic form. Furthermore, CaXyn30A-CD was originally pH-optimized for these studies using BWX as a substrate. Preliminary results obtained using TLC indicated

Figure 5. Hydrolysis of X4 indicates substrate binding preference.
From the yield of X2 and X3 over time, it is calculated that ~64% of the time X4 is hydrolyzed following binding from the −3 through the +1 subsites in favor of binding from the −2 through the +2 subsites.

Figure 6. A visual summary of the oligoxyloside hydrolysis results.
Xylose is represented with a “X” and arabinose with an “A”. For the binding of X4 and X3, xylose binding subsites containing a grey “X” indicate hypothesized shared occupancy.
that the rate of hydrolysis of this specific aldouronate is much lower than that observed for X4 (Figure 7, see the Materials and methods section for reaction details). This would indicate that the altered binding region of CaXyn30A does not accommodate the \( \alpha-1,2 \) GlcA, as would be expected for a canonical GH30-8 endoxylanase. Compared with an Araf substitution, a GlcA is larger and contains a negative charge on the C6 carboxylate. Importantly, we predict that the binding position of an \( \alpha-1,2 \)-L-Araf would be shifted slightly in the direction of the catalytic center compared with an \( \alpha-1,2 \)-D-GlcA. The changes within the \( \beta8-\alpha8 \) and \( \beta7-\alpha7 \) loop regions, of CaXyn30A relative to canonical GH30-8 endoxylanases, seem likely to result in a preference for \( \alpha-1,2 \)-Araf and may prevent the association of the GlcA at this position.

**CaXyn30A-CD crystallization, X-ray crystallographic data analysis, and model building**

High-throughput crystal screening was performed to obtain crystals of CaXyn30A-CD. The original form of expressed protein containing a C-terminal hexahistidine tag failed to crystallize despite extensive screening efforts. Subsequently, a form of CaXyn30A-CD having no affinity tag crystallized in 4.3 M NaCl and 100 mM HEPES (pH 7.5). Only a single protein crystal was recovered from the drop among many salt crystals. Efforts to reproduce this crystallization condition were not successful. Diffraction data refinement and analysis identified a crystal in the \( \text{C}2\text{2}2\text{1} \) space group with a solvent content of 60% containing only a single CaXyn30A-CD molecule in the unit cell. The data were nine-fold redundant and were 99.8% complete. The CaXyn30A-CD model was refined to a final \( R_{work}/R_{free} \) (16.9/21.0) difference of 4.11%. The final resolution was 1.77 Å, and the model containing 469 waters, 102 ligand atoms, and 96.4% of the amino acid side chains present were in the most favored Ramachandran orientation (Table 3). The resulting model of CaXyn30A-CD contained numerous molecules of PEG which could not be readily explained. Upon analysis, we concluded that use of the Oryx Nano Robot with its consecutive drop setting approach likely leads to carryover from the upstream well precipitants. In the case of the Nextal JCSG Core Suite IV, the condition for the drop set just before the successful
precipitant condition (position D6) contains 40% PEG400 (position E6). The molecules of PEG present in the crystal are likely derived from this upstream precipitant condition.

**CaXyn30A-CD structure model analysis**

CaXyn30A-CD presents an overall structure of that recognized as a GH30-8 xylanase and has an all atom percentile based spread of ~1 Å with the canonical GH30-8 xylanases BsXynC (PCB code: 3KL5) and EcXynA (PDB code: 2Y24) [49]. Superposing CaXyn30A-CD (Figure 8A, cyan, PDB code: 5CXP) with the canonical GH30-8 GlcA-dependent endoxylanases BsXynC (green) and EcXynA (magenta) along with the recently described CpXyn30A (gray, PDB code: 4FMV) shows the overall structural similarity of these diverse GH30-8
the observed novel functional characteristics of Ca machinery (Figure 8B). This observation is of significance, differences between these enzymes since the GlcA-dependent endoxylanases such as amino acid residues of this loop region are completely unique compared with those of the canonical GH30-8 GlcA-coordination motif. Visual inspection of the structural superposition (Figure 8A) reveals qualitative distinctions within the β3–α3 loop region of CaXyn30A, showing that it appears more like the Gram-negative EcXynA when compared with the two Gram-positive structures. This difference observed between the Gram-types is predicted to alter the structural mechanism of how the β4–α4 loop interact with and are stabilized by the β3–α3 loop. This structural change may influence small functional differences between these enzymes since the β4-strand contains the acid–base catalyst portion of the catalytic machinery (Figure 8B). This observation is of significance, given the finding that the Gram-positive sourced CaXyn30A enzyme is of the Gram-negative form and it is not known what if any this difference may play in the observed novel functional characteristics of CaXyn30A.

Analysis of the CaXyn30A substrate-binding cleft

Because CaXyn30A bears a strong resemblance to the Gram-negative GH30-8 endoxylanases in the N-terminal face of the substrate-binding cleft, it can best be analyzed in comparison with the ligand-bound structure of EcXynA [11] (PDB code: 2Y24, Figures 1A and 9A). In EcXynA, aldotetrauronate (XGX, Supplementary Figure S1B) is bound in the negative subsite region, and superposition of the CaXyn30A-CD structure model helps to identify subsites −1 through −3 (Figure 9A). This comparison shows that the two enzymes are identical, except in the β7–α7 and β8–α8 GlcA-coordinating loop regions. CaXyn30A lacks four amino acids found in the EcXynA β7–α7 loop and has a two residue insertion (Tyr273 and Asn274) in addition to a significant loss of sequence identity with the EcXynA sequence in the β8–α8 loop (Ser266-Tyr281) (Figures 2B and 9A).

In the canonical GH30-8 endoxylanases, the C-terminal face of the negative subsite region is involved in GlcA coordination, with five amino acid residues interacting with the GlcA moiety (EcXynA amino acids, Tyr255, Ser258, Trp289, Arg293, and Tyr295; Figure 1A). In CaXyn30A, immediately following Trp263, the canonical conserved tyrosine is replaced with Trp264 which forms a bulky addition to the substrate-binding cleft at the point of the −2 subsite xylose (Figure 9A). Presumably due to the close stacking interaction which occurs between these consecutive tryptophan residues, the remaining loop sequence turns sharply inward.

![Figure 8. GH30-8 structure comparison and the substrate binding cleft of CaXyn30A.](image-url)
toward the enzyme center when compared with the canonical GH30-8 endoxylanases (Figures 8A and 9A, $\beta_8$–$\alpha_8$ loop). The change in this region results in a $C_\alpha$ position difference of 5.2 Å between Gly267 in CaXyn30A and Arg293 in EcXynA. In the EcXynA structure, the guanidinium group of Arg293 establishes the crucial salt-bridge interaction with the C6 carboxylate group of GlcA. In CaXyn30A, the side chain carboxylate of Asp268 extends into the typical GlcA-coordination site and occupies nearly the same space as the C6 carboxyl group of the GlcA observed in the ligand-bound crystal structures (Figures 1A and 9A) [11]. The lack of a positively charged moiety and the presence of aspartate (Asp268) in the would-be GlcA-binding position both greatly decrease the possibility of significant interaction with GlcA substitutions on the $-2$ subsite xylose (Figure 8A). The $\beta_8$–$\alpha_8$ loop of CaXyn30A then extends upward where Gln269 reaches out over Asp268 and is potentially available for interaction with sugar substitutions substituted on the $-2b$ subsite. Asn270 is similarly positioned for access to this region and is also the amino acid side chain in closest proximity to the position occupied by Tyr295 in EcXynA (Figures 1A and 9A), an amino acid known to establish a hydrogen bond with the GlcA C6 carboxylate moiety. The differences in the $\beta_8$–$\alpha_8$ loop amino acid sequence positions, chemical character, and the effects these changes may have on the binding of xylan chain substitutions are thought to be the determinants of the altered binding specificity of CaXyn30A relative to the GlcA-dependent GH30-8 endoxylanases.

In the $\beta_7$–$\alpha_7$ loop, Tyr232 is conserved from the GlcA-binding motif of canonical GH30-8 endoxylanases and should still be available to hydrogen bond with sugar substitutions substituted on the C2 hydroxyl of the $-2a$ subsite xylose (Figure 9A). However, further along the $\beta_7$–$\alpha_7$ loop, the serine residue typically found in GH30-8 enzymes, which provides a hydrogen-bonding contact with the C3 hydroxyl of GlcA, is replaced by a glycine in CaXyn30A and the loop is truncated in size by four amino acids (Figures 1A and 9A).

A recent publication of the ligand-bound structure of the canonical GH30-8 endoxylanase Xyn30A from Clostridium thermocellum (CtXyn30A) [50] has revealed how xylobiose interacts within the positive subsite region of a Gram-positive form of GH30-8 enzymes (Figure 9B) [51]. Generally, all GH30-8 enzymes appear to benefit from a similar positive subsite region hydrophobic cleft as formed in CaXyn30A by Trp144 on the
N-terminal face and by Tyr209 on the C-terminal face, with Leu181 forming the cleft floor. Tyr227 in the β7–α7 loop of CtXyn30A (Tyr232, CaXyn30A equivalent) is within hydrogen-bonding distance to the C3 (and C2) hydroxyl groups of the +1 subsite xylose. In other GlcA-dependent canonical GH30-8 enzymes, this conserved tyrosine has been shown to hydrogen bond with the −2b subsite GlcA through its C2 hydroxyl. This suggests that this single tyrosine side chain may be positioned to facilitate substrate recruitment for both the negative and positive subsite sides of the substrate-binding cleft (Figure 9A,B). In CtXyn30A, the +2 subsite xylose also benefits from a hydrogen bond with Gln173 which extends from the bottom of the cleft toward what would be the glycosidic bond oxygen between the +2 subsite xylose with the next xylose toward the reducing terminus (a putative +3 subsite). However, in the CaXyn30A structure, no such interaction is possible due to the presence of a glycine (Gly182) in the equivalent position. Based upon comparison of xylobiose coordination in the positive subsite region of the CtXyn30A enzyme, no hydrogen bonds are expected to occur in the +2 xylose-binding subsite of CaXyn30A, but the analysis indicates that a substrate-stabilizing hydrogen bond contact between Tyr232 (in CaXyn30A) and the xylose in the +1 subsite may be a normal substrate-stabilizing interaction for these enzymes.

Functional comparison of CAZy Clan A endoxylanases

Structural analysis of CaXyn30A-CD in conjunction with the biochemical data supports the possibility that this unique GH30-8 endoxylanase interacts with substrate in a manner that sets it apart from the canonical GH30-8 endoxylanases as well as those belonging to more generic endo-functioning xylanases such as those from GH family 10. Both GH10 and GH30 xylanases are CAZy Clan A (β/α)8 barrels (4/7 hydrolases) having Clan-conserved features, which, for these enzymes, includes proper positioning of the xylose in the −1 subsite with two hydrogen bonds between conserved amino acid side chains and the C2 and C3 hydroxyls of the xylose occupying the subsite. In addition to these two specific contacts, both of these Clan A-type endoxylanases employ a multitude of other specificity determining hydrogen bond contacts in addition to less specific hydrophobic interactions which stabilize the xylan chain for hydrolysis. In the case of the GH10 endoxylanases, there are often more than 10 hydrogen bonds between the four xylose-binding subsites most proximal to the catalytic center subsites and the xylan chain. For the canonical GlcA-dependent GH30-8 endoxylanases, the primary determinant of substrate specificity is the recognition of the GlcA appendage in the −2b subsite. In the Gram-negative derived EcXynA (PDB code: 2Y24), the GlcA moiety alone is predicted to interact with the enzyme through six hydrogen bonds and a salt bridge (Figure 1), whereas the xylan-binding subsites for this enzyme only benefit from, at most, three additional hydrogen bonds, two of which are the Clan A specific, −1 subsite contacts mentioned above. Because CaXyn30A differs almost completely in the amino acid residues responsible for the specific tight coordination of GlcA and given its canonical GH30-8 characteristic of lacking amino acid residues in the substrate-binding cleft capable of hydrogen bonding, the mechanism of xylan interaction for efficient endoxylanase function is unique in that it is largely based on hydrophobic interactions.

Protein structure/function results synopsis

In the studies presented here, we have shown that CaXyn30A is a GlcA-independent GH30-8 endoxylanase yielding xylobiose and xylotriose as neutral xylooligosaccharide limit products upon hydrolysis of xylans and larger oligoxyllosides. It is not known from the protein structure analysis what aspects of the minimally altered core xylan-binding cleft of CaXyn30A allows for such efficient hydrolysis of linear xylooligosaccharides. The relatively low rate of hydrolysis measured for X4 and the subsequent leap in activity for hydrolysis of X5 and X6 indicate that CaXyn30A benefits from a minimum of five xylose-binding subsites. CaXyn30A also showed increased activity on arabinoxylan relative to glucuronoxylan. Analysis of hydrolysis rate of Araf-oligoxyllosides indicated that CaXyn30A may have a preference for α-1,2-Araf substitutions, indicating that these common arabinoxylan substitutions may act as specificity factors to enhance the rate of hydrolysis of arabinoxylans. A test to determine if an α-1,2-GlcA appendage is accommodated in a similar manner as in canonical GH30-8 endoxylanase was negative, confirming that CaXyn30A has no specificity for GlcA substitutions.

Abbreviations

A²XX, 2°-α-L-arabinofuranosyl-xylotriose; Araf, arabinofuranose; BsXynC, xylanase C from the Gram-positive bacterium Bacillus subtilis; BWX, beech wood glucuronoxylan; CaXyn30A, xylanase 30A from Clostridium acetobutylicum; CaXyn30A-CD, CaXyn30A catalytic domain; CFE, cell-free extract; CmXyn10B, Xyn10B from Clostridium papyrosolvens; DS, degree of substitution; EcXynA,
xylanase A from the Gram-negative bacterium *Erwinia chrysanthemi*; GH, glycoside hydrolase; GH30, glycoside hydrolase family 30; GH30-8, glycoside hydrolase family 30 subfamily 8; GlcA, α-1,2-linked 4-O-methylglucuronic acid; TLC, thin layer chromatography; WAX, wheat arabinoxylan; XA2+3XX, 2,3,3-di-α-L-arabinofuranosyl-xylotetraose; XA3XX, 3,3-α-L-arabinofuranosyl-xylotetraose; pNP, para-nitrophenol.

**Author Contribution**
F.J.S.J. conceived of the study. F.J.S.J., C.C., and D.D. planned and coordinated the research with methodological contributions from J.H. and E.P. All authors contributed to research data acquisition and analysis. The manuscript was written by F.J.S.J. with input from all authors. All authors have approved the final form of the manuscript.

**Funding**
J.H. is supported by a grant from the National Center for Research Resources [5 P20 RR016461].

**Acknowledgements**
We thank the Institute for Microbial and Biochemical Technology of the Forest Products laboratory (FPL), a National Laboratory of the USDA Forest Service for ongoing financial support. We also thank Fred Matt in the FPL Analytical Chemistry Laboratory for sugar quantification with the Dionex HPLC. Also, we thank Dr. James F. Preston at the University of Florida, Department of Microbiology and Cell Science for general advice regarding this manuscript and Dr. Zui Fujimoto from the National Agriculture and Food Research Organization in Tsukuba, Japan for thoughtful discussion regarding the manuscript and protein structure figures. 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.

**Competing Interests**
The Authors declare that there are no competing interests associated with the manuscript.

**References**

