
Article:
Ariza, A. orcid.org/0000-0003-4364-823X, Eklöf, J.M., Spadiut, O. et al. (8 more authors) (2011) Structure and activity of Paenibacillus polymyxa Xyloglucanase from glycoside hydrolase family 44. Journal of Biological Chemistry, 286 (39). pp. 33890-33900. ISSN 0021-9258

https://doi.org/10.1074/jbc.m111.262345

⌘ *Author's Choice*

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 39, pp. 33890–33900, September 30, 2011

Structure and Activity of *Paenibacillus polymyxa* Xyloglucanase from Glycoside Hydrolase Family 44*[S]

Received for publication, May 19, 2011, and in revised form, July 12, 2011 Published, JBC Papers in Press, July 27, 2011, DOI 10.1074/jbc.M111.262345

Antonio Ariza[‡1], Jens M. Eklöf[§1], Oliver Spadiut[§¶1], Wendy A. Offen[‡], Shirley M. Roberts[‡], Werner Besenmatter[‖], Esben P. Friis[‖], Michael Skjøt[‖], Keith S. Wilson[‡2], Harry Brumer[§¶3], and Gideon Davies[‡4]

From the [‡]Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, United Kingdom, the [§]Division of Glycoscience, School of Biotechnology, and [¶]Wallenberg Wood Science Center, Royal Institute of Technology, SE-106 91 Stockholm, Sweden, and [‖]Novozymes A/S, DK-2880 Bagsværd, Denmark

The enzymatic degradation of plant polysaccharides is emerging as one of the key environmental goals of the early 21st century, impacting on many processes in the textile and detergent industries as well as biomass conversion to biofuels. One of the well known problems with the use of nonstarch (nonfood)-based substrates such as the plant cell wall is that the cellulose fibers are embedded in a network of diverse polysaccharides, including xyloglucan, that renders access difficult. There is therefore increasing interest in the "accessory enzymes," including xyloglucanases, that may aid biomass degradation through removal of "hemicellulose" polysaccharides. Here, we report the biochemical characterization of the *endo*-β-1,4-(xylo)glucan hydrolase from *Paenibacillus polymyxa* with polymeric, oligomeric, and defined chromogenic aryl-oligosaccharide substrates. The enzyme displays an unusual specificity on defined xyloglucan oligosaccharides, cleaving the XXXG-XXXG repeat into XXX and GXXXG. Kinetic analysis on defined oligosaccharides and on aryl-glycosides suggests that both the −4 and +1 subsites show discrimination against xylose-appended glucosides. The three-dimensional structures of *Pp*XG44 have been solved both in apo-form and as a series of ligand complexes that map the −3 to −1 and +1 to +5 subsites of the extended ligand binding cleft. Complex structures are consistent with partial intolerance of xylosides in the −4′ subsites. The atypical specificity of *Pp*XG44 may thus find use in industrial processes involving xyloglucan degradation, such as biomass conversion, or in the emerging exciting applications of defined xyloglucans in food, pharmaceuticals, and cellulose fiber modification.

The xyloglucans comprise a family of cell wall polysaccharides found in many land plants, including both monocots and dicots, where they function as structural and/or storage glycans. Whereas grasses can have as little as 1–5% xyloglucan (dry weight) in their cell walls, the cell walls of diverse monocots and dicots may contain up to 25% xyloglucan as a matrix glycan that is intimately associated with cellulose (1–5). In the seeds of plants that have recruited xyloglucan as a storage polysaccharide, xyloglucan levels can approach half of the dry weight (6, 7). Estimates place the annual terrestrial net primary production of plant biomass on the tens of gigatonnes scale (as elemental carbon), with total a terrestrial biomass pool of ~2000 Gt (8, 9). The near-ubiquity of xyloglucans in the terrestrial biosphere thus provides ample fodder for saprophagic micro-organisms, as well as a renewable commodity for potential valorization.

Xyloglucans can also be considered as Nature's water-soluble (or, at least water-dispersible) cellulose derivatives. The hallmark of xyloglucans is a linear β-1,4-glucan main chain, identical to that of cellulose, which is substituted with α-1,6-linked xylopyranosyl units at regular repeating intervals. Common xyloglucan repeats are built on a cellotetraosyl core, including the XXXG motif, in which one unbranched glucopyranosyl unit (denoted "G" (10)) is followed by three xylosylated Glc*p* units (each denoted "X"), and the XXGG motif, in which two unbranched Glc*p* units are followed by two xylosylated Glc*p* units (Fig. 1). A limited number of more heavily and more sparsely branched xylogluco-oligosaccharides are also known. Further elaboration of these motifs with combinations of galactopyranosyl, arabinofuranosyl, L-fucopyranosyl, xylopyranosyl, and *O*-acetyl groups gives rise to a diversity of species- and tissue-specific xyloglucans (see Refs. 2, 3, 11, and 12 for a comprehensive overview of known structures).

It has been long recognized that certain cellulases, more formally endo-β-(1–4)-glucanases (EC 3.2.1.4), can also catalyze endo-hydrolysis of XGs (EC 3.2.1.151), as a direct consequence of the backbone homology of these two polysaccharides. Moreover, limit-digestion of xyloglucan with endo-glucanases typically produces well defined oligosaccharide mixtures arising from cleavage at specific, usually unbranched, glucosyl residues. Initial speculation about the specificity determinants of endo-glucanases (13) was contemporary with the first three-dimensional structural studies on these enzymes in the mid-1990s (14), but only recently have combined structure-function analyses of endo-(xylo)glucanases been given focused attention (reviewed in Ref. 15). Presently, tertiary structures of endo-glucanases with some degree of specificity for the highly branched xyloglucan chain have been obtained for glycoside hydrolase

* This work was supported in part by the Biotechnology and Biological Sciences Research Council, the Swedish Research Council, the Swedish Center for Biomimetic Fiber Engineering (Biomime), and by Novozymes A/S. W. B., E. P. S., and M. S. are employed by Novozymes A/S.

⌘ *Author's Choice*—Final version full access.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. 1 and 2 and Table 1.

The atomic coordinates and structure factors (codes 2ykk, 3zq9, 2yih, and 2yjq) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] These authors contributed equally to this work.

[2] To whom correspondence may be addressed. E-mail: keith@ysbl.york.ac.uk.

[3] To whom correspondence may be addressed. E-mail: harry@biotech.kth.se.

[4] Recipient of a Royal Society/Wolfson research merit award. To whom correspondence may be addressed. E-mail: davies@ysbl.york.ac.uk.

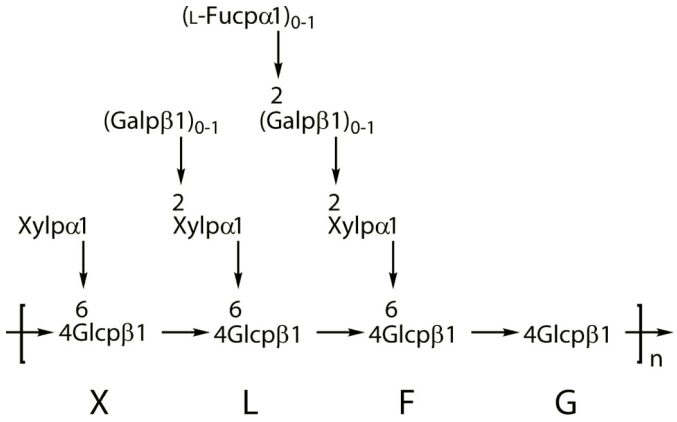

FIGURE 1. **Common side-chain decorations of xyloglucans.** Xyloglucan fragments are named according to Ref. 10.

families (16) GH5 (17), GH12 (17), GH16 (18, 19), GH44 (20–22), and GH74 (23–26). In a number of these cases, crystallographic complexes with xylogluco-oligosaccharides have illuminated key discriminating enzyme-substrate interactions (17, 19, 23, 24).

Presently, the discovery of new xyloglucanases continues (27–31), stimulated in part by a significant contemporary interest in plant cell wall saccharification (32). Indeed, the potential of endo-xyloglucanases to synergize with cellulases in the degradation of the recalcitrant plant cell wall was highlighted over 15 years previously (33). Specific xyloglucanases may also be useful in tailoring the polysaccharide for diverse applications (34). The crystal structures of GH44 enzymes belonging to three different organisms have been reported to date as follows: six structures of an endo-glucanase from *Clostridium thermocellum* (*Ct*Cel44A) (20), one structure of an endo-glucanase from *Clostridium acetobutylicum* ATCC 824 (*Ca*XG44) (22), and two structures of a "bifunctional glucanase-xylanase" isolated from a metagenome library (*uc*CelM2) (35). In this context, we present here the structural and functional characterization of a new mixed-function *Paenibacillus polymyxa* endo-β-(1–4)-xyloglucanase, (*Pp*XG44) from glycoside hydrolase family 44 (GH44). It exhibits broad activity toward tamarind xyloglucan, barley β-glucan, and synthetic soluble cellulose derivatives. Notably, the enzyme was able to accommodate both G and X units in the −1 subsite, in contrast to many known endo-xyloglucanases from other GH families. Fine grain kinetic and protein structural analysis using a range of specific oligosaccharide substrates and inhibitors served to further illuminate active-site features responsible for this unusual specificity.

MATERIALS AND METHODS

Protein Production and Purification—The *Pp*GH44 enzyme is derived from a naturally occurring parent enzyme, which is a multifunctional assembly organized "GH44-FN3-(GH26(mannanase))-FN3-CBM3 (36, 37). The GH44 module for kinetic and structural studies was thus cloned with a C-terminal truncation (corresponding to residues 1–559 of sequence 2 in United States patent 6815192 (38)).

Three protein variants of this GH44 module, wild type, an increased stability mutant (K129A and R156Y), and a nucleophile mutant (E358S; containing additional stabilizing mutations Q68H, T92V, K118A, K129A, R156Y, G200P, and N331F), were kind gifts from Novozymes A/S (Bagsvaerd, Denmark) where they had previously been expressed in *Bacillus subtilis* as follows. The xyloglucanase genes were genomically integrated at the *pel* locus together with three promoters in tandem and with a fragment that complemented the *cat* gene for chloramphenicol resistance and expressed in strain SHa273, which is a *trp amyE aprE nprE*-inactivated descendant of *B. subtilis* 168 (39). The protein was secreted into the growth medium, which contained, in tap water, 100 g/liter sucrose, 40 g/liter soy meal, 10 g/liter $Na_2HPO_4 \cdot 12H_2O$, 5 g/liter $CaCO_3$, chloramphenicol, and an antifoaming chemical for the shake flask fermentation at 37 °C for 4 days. The supernatant was sterile-filtered, adjusted to pH 5 with acetic acid, and applied to an XpressLine ProA column (UpFront chromatography A/S, Germany) equilibrated in 50 mM succinic acid/NaOH, 1 mM CaCl, pH 5. Proteins were eluted by a step elution with 50 mM Tris-HCl, pH 9. Fractions with xyloglucanase activity (measured as described below) were pooled and adjusted to pH 9 with 3 M Tris base. The solution was diluted to the same (or lower) conductivity as 50 mM Tris-HCl, pH 9, and applied to a SOURCE Q column (GE Healthcare) equilibrated in 50 mM Tris-HCl, pH 9. Proteins were eluted with a linear NaCl gradient (from 0 to 500 mM) in the same buffer over 5 column volumes. Fractions with xyloglucanase activity and only one band on an SDS-polyacrylamide gel were pooled. The buffer of the samples was exchanged to 10 mM CHES,[5] pH 9, and 50 mM NaCl.

Xyloglucanase Screening Assay—Xyloglucanase activity was measured with the substrate azurine cross-linked xyloglucan (Megazyme) as follows. A 500-μl substrate suspension (4 mg/ml azurine cross-linked xyloglucan substrate homogeneously suspended in 0.01% Triton X-100 by stirring) and 500 μl of assay buffer (50 mM succinic acid/NaOH, 0.01% Triton X-100, pH 5) were pipetted into a microcentrifuge tube on ice. 20 μl of enzyme sample (diluted in 0.01% Triton X-100) were added to the ice-cold mixture. The assay was initiated by transferring the tube to a thermomixer at 37 °C and shaking at 1400 rpm. After 15 min, the tube was put back into the ice bath. To remove unreacted substrate, the mixture was centrifuged, and the absorbance of the supernatant at 650 nm was measured. A sample with 20 μl of 0.01% Triton X-100 instead of enzyme was assayed in parallel, and its value was subtracted from the enzyme sample measurement.

Substrate Specificity Assays—The substrate specificity of *Pp*XG44 on tamarind xyloglucan, carboxymethylcellulose, medium viscosity barley β-glucan, wheat arabinoxylan, carob galactomannan (all from Megazyme), and hydroxyethylcellulose (Fluka) was determined using the BCA-reducing sugar assay (40, 41). In general, a total assay volume of 250 μl (1 g/liter polysaccharide; 50 mM sodium citrate; 3.1 or 4.2 nM *Pp*XG44)

[5] The abbreviations used are: CHES, 2-(cyclohexylamino)ethanesulfonic acid; PDB, Protein Data Bank; r.m.s.d., root mean square deviation; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; CNP, 2-chloro-4-nitrophenyl; HPAEC-PAD, high performance anion-exchange chromatography with pulsed amperometric detection.

was incubated at 30 °C for 10 or 25 min. The reaction was stopped by addition of 250 μl of BCA solution and color developed at 80 °C for 20 min. Reducing sugars were quantified *versus* a linear glucose standard (1–75 μM) by absorption at 560 nm with a Cary 50 UV-visible spectrophotometer (Varian, Darmstadt, Germany).

pH Rate Profile—The pH rate profile of *Pp*XG44 was determined using tamarind xyloglucan (1 g/liter) and the BCA assay as described above with 3.1 nM *Pp*XG44. The buffers used were sodium citrate between pH 2.1 and 6.4 and sodium phosphate between pH 6.2 and 8.0. The experimental data were fitted to Equation 1 using Origin 8 (OriginLab Corp.).

$$k_{obs} = \frac{k_{max}}{1 + 10^{pK_{a1} - pH} + 10^{pH - pK_{a2}} + 10^{pK_{a1} - pK_{a2}}} \quad \text{(Eq. 1)}$$

Product Analysis by High Performance Anion-exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD)—For product analysis, a Dionex ICS-3000 HPAEC-PAD system and a Dionex CarboPac PA-200 column were used. Two different programs were run depending on the sample. Program 1 was used for cello-oligosaccharides and consisted of an isocratic elution with 100 mM NaOH and 30 mM NaOAc at a flow rate of 0.5 ml/min. Program 2, as described previously (42), was used for polymeric substrates and xylogluco-oligosaccharides.

Substrate Specificity on Cello- and Xylogluco-oligosaccharides—Cello- and xylogluco-oligosaccharides were incubated with *Pp*XG44 for different time periods at 30 °C in 50 mM sodium citrate buffer, pH 6.0. An overview of the different assay conditions used is given in supplemental Table 1. Reaction products were analyzed by HPAEC-PAD (program 1 for cello-oligosaccharides and program 2 for xylogluco-oligosaccharides). Observed peaks were quantified in comparison with linear standards of glucose, GG, GGG, GGGG, and XXXG from 1 to 50 μM.

Time-dependent Depolymerization of Xyloglucan—Samples (145 μl) containing xyloglucan (1.7 g/liter) and *Pp*XG44 (7.1 nM) in 70 mM NH_4OAc, pH 5.5, were incubated for 0, 10, 30, 60, and 180 min and overnight at 30 °C. The reactions were stopped by heating at 95 °C for 20 min. The samples were subsequently lyophilized and dissolved in 350 μl of DMSO and left at 80 °C for 5 min before gel permeation chromatography (GPC). GPC analysis was performed using the method described in Ref. 43, with the exception that two PLgel 10-μm mixed-B columns and a guard column (Agilent) were used in series. The molecular mass was estimated using an 11-point pullulan standard curve (180–1,660,000 Da).

Product Analysis by Mass Spectrometry—Analysis of reaction products of different β-aryl glycosides as well as end products of XXXGXXXG digestion were analyzed by mass spectrometry using a Q-Tof2 mass spectrometer fitted with a nano-flow ion source (Waters Corp.) as described previously (23).

Initial Rate Enzyme Kinetics with β-Aryl Oligosaccharides—The hydrolysis of the 2-chloro-4-nitrophenyl (CNP)-β-glycosides of GG, GGG, GGGG, XXXG, and XLLG (where G = Glcβ(1→4); X = Xylα(1→6) Glcβ(1→4); and L = Galβ(1→2) Xylα(1→6) Glcβ(1→4)) was followed by assessing the release of 2-chloro-4-nitrophenolate at 405 nm (ϵ = 12,936 M^{-1} cm^{-1}, 50 mM sodium citrate buffer, pH 6.0) using a Cary300 Bio UV-visible spectrophotometer and a Cary Peltier temperature controller (both Varian). The synthesis of these chromogenic substrates has been described earlier (44, 45). Substrate (1.0 μM to 5.0 mM), buffer (sodium citrate 50 mM, pH 6.0), and H_2O were equilibrated at 30 °C in a 1-cm path quartz cuvette before the reaction was started by addition of 5 μl of diluted enzyme, yielding a total assay volume of 100 μl. Initial rates were determined from the linear slope of the reaction time course, corresponding to less than 10% conversion of substrates. Kinetic constants were determined by fitting the Michaelis-Menten equation to the data using Origin 8.0 (OriginLab). To analyze the influence of the single active-site subsites on catalysis, the transition state stabilization energies ($\Delta\Delta G^{\ddagger}$) were determined by comparing k_{cat}/K_m values for the different β-aryl oligosaccharide substrates using Equation 2.

$$\Delta\Delta G^{\ddagger} = \Delta G_{n+1} - \Delta G_n = -RT\ln\frac{(k_{cat}/K_m)_{n+1}}{(k_{cat}/K_m)_n} \quad \text{(Eq. 2)}$$

Inhibition by β-1,4-Glucosyl-noeuromycin and β-1,4-Cellobiosyloxazine—To determine the inhibiting effect of β-1,4-glucosyl noeuromycin (Glc-noeuromycin (46)) and β-1,4-cellobiosyloxazine (Glc-Glc-oxazine (47)) on *Pp*XG44, the IC_{50} value at a substrate concentration of 10 μM GGGG-CNP was determined. The release of the 2-chloro-4-nitrophenolate from GGGG-CNP was followed in the presence of varying concentrations of the competitive inhibitors Glc-noeuromycin (0.037–180 μM) or Glc-Glc-oxazine (0.75–100 μM) at 405 nm (ϵ = 12 936 $M^{-1}cm^{-1}$, 50 mM sodium citrate buffer, pH 6.0) in triplicate using a Cary300 Bio UV-visible spectrophotometer with a Cary temperature control (both Varian). Substrate, buffer, inhibitor, and H_2O were equilibrated at 30 °C in a 1-cm path quartz cuvette before the reaction was started by the addition of 5 μl of diluted enzyme (0.96 μM), yielding a total assay volume of 100 μl. The IC_{50} value was determined by plotting the relative activity against the concentration of inhibitor and fitting Equation 3 by nonlinear regression. From the IC_{50} value, the K_i value was approximated employing Equation 4, (48).

$$k_{obs} = \frac{k_{max}}{1 + \left(\frac{[I]}{IC_{50}}\right)} \quad \text{(Eq. 3)}$$

$$K_i = \frac{IC_{50}}{1 + \left(\frac{[S]}{K_m}\right)} \quad \text{(Eq. 4)}$$

Crystallization and Data Collection—Crystallization screening was performed at 291 K with commercially available crystal screens using the sitting-drop vapor diffusion method. Drops were set up employing a Mosquito Crystal liquid handling robot (TTP LabTech) with 150 nl of protein solution plus 150 nl of reservoir solution in 96-well format plates (MRC 2-well crystallization microplate, Swiss Sci) equilibrated against 54 μl of reservoir solution. Optimization screening was performed using the hanging-drop vapor diffusion method with 0.5 μl of

TABLE 1
Data collection and refinement statistics for *Pp*XG44

Data collection	Native	Glc-noeuromycin	GXGGG	Glc-Glc-oxazine
Resolution range	1.79 Å (1.89–1.79 Å)	1.86 Å (1.96–1.86 Å)	1.70 Å (1.79–1.70 Å)	2.25 Å (2.37–2.25 Å)
Space group	$P3_121$	$P3_121$	$P3_121$	$P3_221$
Unit cell dimensions				
a, b, c	84.17, 84.17, 157.57 Å	83.61, 83.61, 157.53 Å	83.53, 83.53, 157.14 Å	84.20, 84.20, 320.05 Å
α, β, γ	90, 90, 120°	90, 90, 120°	90, 90, 120°	90, 90, 120°
Completeness	98.8% (91.6%)	99.5% (96.6%)	95.8% (93.2%)	98.5% (92.5%)
R_{merge}	6.0 (15.5)	5.5 (13.1)	8.8 (41.8)	9.7 (56.0)
Redundancy	6.9 (6.2)	7.0 (6.5)	5.3 (4.6)	8.1 (5.4)
$I/\sigma(I)$	23.9 (9.6)	26.3 (11.3)	12.0 (3.1)	14.9 (2.7)
Refinement				
Resolution range	1.79 Å (1.84 Å)	1.86 Å (1.91 Å)	1.70 Å (1.74 Å)	2.25 Å (2.31 Å)
R_{work}/R_{free}	14.4%/17.7%	13.7%/17.1%	15.7%/18.4%	19.7%/26.0%
r.m.s.d. bond lengths (Å)	0.019 Å	0.018 Å	0.014 Å	0.020 Å
r.m.s.d. bond angles°	1.6°	1°	1.3°	1.8°
Average *B* factors (Å^2)	13	14	14	30
Protein	11	12	13	30
Ligand/ion	25	17	26	29
Solvent	25	25	23	29
Ramachandran plot				
Most favored	97.6%	97.6%	97.4%	95.8%
Allowed	2.4	2.4	2.6	4.1
Outliers	0.0	0.0	0.0	0.1
PDB codes	2ykk	3zq9	2yih	2yjq

protein solution plus 0.5 μl of reservoir solution in Cellstar 24-well cell culture plates (Greiner Bio-One) equilibrated against 500 μl of reservoir solution.

Native crystals were grown from protein in 10 mM Tris-HCl, pH 9.0, 50 mM NaCl, and reservoir solution containing 0.1 M BisTris, pH 6.5, 0.2 M Li_2SO_4, 25% (w/v) PEG 3350. The crystals were cryoprotected by dipping them into a reservoir solution containing 15% xylitol for 5 s before vitrifying them in liquid nitrogen. A complex with Glc-noeuromycin (46) (kindly supplied by Professor R. V. Stick, University of Western Australia) was obtained from a crystal grown from protein in 10 mM CHES, pH 9.0, 50 mM NaCl and reservoir solution containing 0.1 M HEPES, pH 6.5, 0.2 M Li_2SO_4, 25% (w/v) PEG 3350. The crystal was soaked in reservoir solution with 5 mM Glc-noeuromycin for 15 min and then cryoprotected by dipping into the same solution with 15% ethylene glycol for 5 s before vitrifying in liquid nitrogen. X-ray data were collected using a MicroMax-007 (Rigaku) x-ray generator (CuK_α, $\lambda = 1.54179$ Å) equipped with a MAR345 image plate detector (MARresearch).

A further crystal of the nucleophile mutant E358S, grown over 25% (w/v) PEG 3350, 0.2 M Li_2SO_4, 0.1 M BisTris, pH 6.5 (1:1 protein/well solution), was soaked in mother liquor with a few specks of solid powder of Glc-Glc-Oxazine (47) (kindly supplied by Professor R V Stick, University of Western Australia) for about 50 min. The crystal was transferred by stepwise replacement of the buffer with well solution with 15% ethylene glycol, over a total period of about 30 min. Data were collected on beamline I02 at Diamond.

Complexes with the nucleophile mutant E358S were also obtained with a xyloglucan oligosaccharide ligand (XXXGXXXG) as follows. Protein at a concentration of 9.5 mg/ml in 10 mM MOPS, pH 7.0, was crystallized using the hanging drop method, with 27% (w/v) PEG 3350, 0.1 M BisTris, pH 6.5, 0.2 M Li_2SO_4 in the well, mixed in a ratio of 1:0.8 μl with the well solution in the hanging drop. The crystal was transferred to a drop consisting of the well solution with 10 mM XXXGXXXG and left to soak for 40 min. It was gently transferred to a cryo solution consisting of the well solution with 15% (v/v) ethylene glycol, by stepwise replacement of the soak solution with increasing proportions of cryo-protectant solution. Data were collected at Diamond on beamline I04.

Data Processing, Structure Solution, and Refinement—X-ray data were reduced and processed using programs from the *CCP*4 suite (49). The images were integrated with *MOSFLM* (50) and then scaled with SCALA. A molecular replacement solution for the native data set was obtained using the CCP4 implementation of MOLREP with a GH family 44 endoglucanase from *C. thermocellum* (PDB code 2E0P (20)), with which it shares 59% identity, as the search model. The residues of the MR solution model were mutated to the correct sequence for *Pp*XG44 using BUCCANEER (51) followed by manual building in COOT (52) and automated refinement using REFMAC5 (53). The data from the complexes were solved using the refined native structure as the search model. Model validation was performed with MOLPROBITY (54) prior to deposition at the Protein Data Bank. Data and structure refinement quality statistics are given in Table 1. Structure similarity searches of *Pp*XG44 were performed with SSM (55).

RESULTS AND DISCUSSION

Polysaccharide Substrate Specificity and pH Optimum—*Pp*XG44 demonstrated a broad substrate specificity for polysaccharides containing β-1,4 linkages (Table 2). The highest activity was observed with tamarind xyloglucan ($v_0/[E]_t = 763 \pm 17\ min^{-1}$), and a similar rate was observed for carboxymethylcellulose (Table 2). Hydrolytic activity toward mixed-linked β-glucan from barley and hydroxyethyl cellulose was also comparably high, with specific activities of 87 and 65% relative to the xyloglucanase activity, respectively. The activities toward wheat arabinoxylan and carob galactomannan were significantly lower than the xyloglucanase activity, 15 and 4%, respectively.

The pH-rate profile for the *Pp*XG44-catalyzed hydrolysis of tamarind xyloglucan exhibited an optimum at pH 6.0. The pro-

TABLE 2
Substrate specificity of *Pp*XG44 on different polysaccharides at 1 g/liter

Substrate	$v_0/[E]_t$
	min^{-1}
Tamarind xyloglucan	760 ± 17
Carboxymethylcellulose	720 ± 6
β-Glucan from barley	660 ± 22
Hydroxyethylcellulose	500 ± 28
Arabinoxylan from wheat	120 ± 20
Carob galactomannan	34 ± 1.2

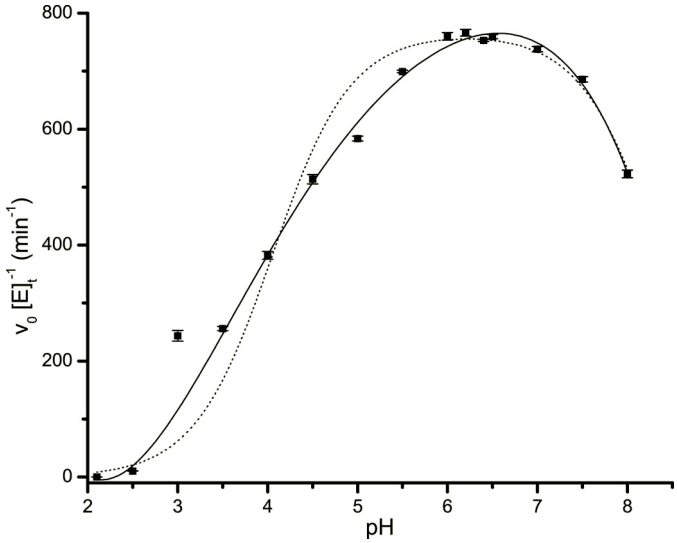

FIGURE 2. **pH rate profile of *Pp*XG44.** The *dotted line* represents the fit of Equation 1 to the experimental data, and the *solid line* represents a smooth Bezier curve through the data.

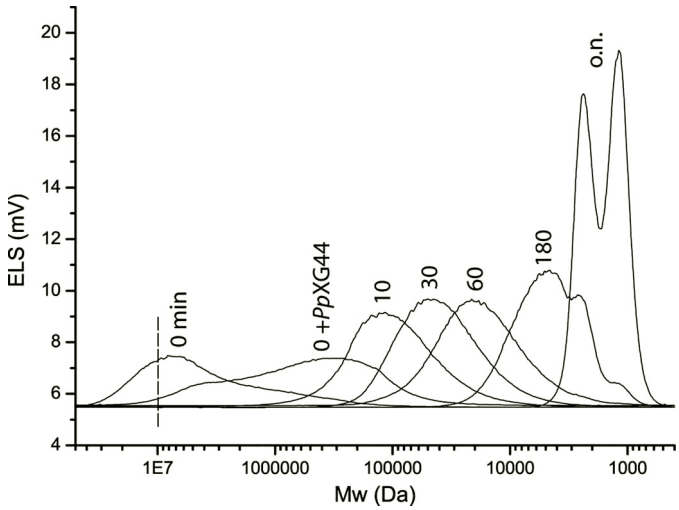

FIGURE 3. **Time-dependent depolymerization of xyloglucan.** The incubation time of *Pp*XG44 with xyloglucan is indicated *above* each chromatogram. The "*0-min*" time point is for native tamarind xyloglucan without enzyme, and in the "*0 + PpXG44*" enzyme is added just before heating the sample at 95 °C. The *dotted line* at 10^7 Da indicates the cutoff value of the column.

TABLE 3
Substrate specificity of *Pp*XG44 on oligosaccharides based on initial rates measured at less than 10% substrate conversion

Oligosaccharide	Substrate concentration	$v_0/[E]_t$
	mM	min^{-1}
GG	1	
GGG	1	
GGGG[a]	1	0.15 ± 0.002
GGGGG[b]	1	5.3 ± 0.1
GGGGGG[c]	0.175[d]	150 ± 4
XXXGXXXG[e]	1	140 ± 9
GXXGXXXG[f]	1	120 ± 10

[a] Hydrolysis rate ratios are from − to + subsites, G3 + G1, and 2G2 (18:1).
[b] Hydrolysis rate ratios are G4 + G1 and G3 + G2 (3.6:1).
[c] Hydrolysis rate ratios are G5 + G1, G4 + G2, and 2G3 (2:4.5:1).
[d] Cellohexaose was assayed at 175 μM due to poor solubility.
[e] Hydrolysis rate ratios are XXX + GXXXG and XXXG + XXXG (8.7:1).
[f] Hydrolysis rate ratios are GXX + GXXXG and GXXG + XXXG (1:1.2).

file was approximately bell-shaped over the measured range pH 2–8, although the data were not precisely fit by Equation 1 describing an active site with two ionizable groups (Fig. 2). Apparent kinetic pK_a values of 4 and 8.5 can be approximated from the acidic and basic limbs of the pH-rate profile, respectively.

A time course of xyloglucan degradation by *Pp*XG44 was monitored by gel permeation chromatography to discern the mode of cleavage on xyloglucan. The gel permeation chromatography data strongly support an endo-type of action of *Pp*XG44 on xyloglucan, with low mass xyloglucan oligosaccharides appearing only at the very last stages of xyloglucan depolymerization (Fig. 3).

Most characterized retaining endo-xyloglucanases, such as those found in GH5, GH7, GH12, and GH16 (15, 56), preferably cleave xyloglucan at unsubstituted glucosyl units. The limit digestion product profile of *Pp*XG44 on tamarind xyloglucan, however, differed from that of "classical" retaining endo-xyloglucanases. Therefore, the products resulting from the hydrolysis of a well defined Glc_8-based xylogluco-oligosaccharide (XXXGXXXG) were identified by MS and HPAEC-PAD. The most abundant masses observed corresponded to XXX and GXXXG indicating a preferential binding of XXXGXXXG in −3 to +5 subsites with an X motif in the −1 subsite and a G motif in the +1 subsite. XXXG products from a less favored −4 to +4 binding mode were also detected (supplemental Fig. 1). The cleavage pattern on XG observed with *Pp*XG44 is reminiscent of that of an inverting GH74 endo-xyloglucanase (57). However, this GH74 enzyme cleaves XX**X**GXXXG at **X** with a limited frequency, and it preferentially cleaves this minimal substrate at the internal unbranched glucose.

Specificity toward Cello- and Xylogluco-oligosaccharides—To understand better the substrate specificity of *Pp*XG44, the activity on cello- and xylogluco-oligosaccharides was determined. The minimal substrate for which hydrolysis could be detected was cellotetraose (GGGG), although the activity was very low (Table 3). The longer cellopentaose was 35 times more potent as a substrate, and the hydrolysis rate of cellohexaose was again almost 30 times faster than that of cellopentaose ($v_0/[E]_t$ 150 ± 4 min^{-1}), clearly showing that *Pp*XG44 has at least six subsites (Table 3). Several cleavage modes were observed for cellohexaose cleavage, but there was a preference for producing cellobiose and cellotetraose products. The longer xylogluco-oligosaccharides XXXGXXXG and GXXGXXXG were hydrolyzed at similar rates to cellohexaose (albeit at a higher substrate concentration), thus supporting the observed broad substrate specificity of *Pp*XG44 on polysaccharides. Interestingly, removal of the α-1,6-linked xylosyl unit on the nonreducing

TABLE 4
Kinetic parameters and transition state stabilization energies of *Pp*XG44 with β-aryl oligosaccharides

Substrate	$K_{m(app)}$	$k_{cat\ (app)}$	$k_{cat}/K_{m(app)}$	Relative activity	$\Delta\Delta G^{\ddagger}$
	mM	min^{-1}	$mM^{-1}\ min^{-1}$	%	*kcal/mol*
GG-CNP	Catalytic activity not detected[a]				
GGG-CNP	0.361 ± 0.097	440.4 ± 40.4	1220	3.8	2
GGGG-CNP	0.029 ± 0.001	955.9 ± 7.01	32,500	100	0
XXXG-CNP	0.134 ± 0.008	842.8 ± 9.44	6280	19.4	1
XLLG-CNP	0.998 ± 0.150	417.8 ± 22.1	419	1.3	2.6

[a] Activity assays were performed with up to 5 mM substrate and up to 2.4 μM *Pp*XG44 at 30 °C for up to 30 min.

end of the xylogluco-oligosaccharide XXXGXXXG does not significantly affect the rate of hydrolysis. Instead, the subsite binding, with strong preference for −3 to +5 over −4 to +4 (8.7:1 in XXXGXXXG) shifts to roughly equal preference (1:1.2 in GXXGXXXG) (Table 3). It is noteworthy that in interaction with the natural substrate, when a G motif is bound in the +1 subsite (X in subsite −1), the regular repeating pattern of xyloglucan also places G motifs in register in the −4 and +5 subsites, which likely makes this binding mode even more favorable.

Initial Rate Enzyme Kinetics with β-Aryl Oligosaccharides—To dissect the requirements for catalysis and the contributions of negative subsites, k_{cat} and K_m values for CNP-β-glycosides GG, GGG, GGGG, XXXG, and XLLG-CNPs were determined assaying aglycone release.

The minimal CNP substrate on which *Pp*XG44 was active was GGG-CNP. This nicely corresponds to the minimal oligosaccharide substrate for which hydrolysis could be observed, cellotetraose. GGGG-CNP is ~25-fold "better" as a substrate than GGG-CNP, as judged by the k_{cat}/K_m values. The addition of xyloside or galactoside substituent to GGGG-CNP penalizes catalysis, to some extent, discussed later in light of the three-dimensional structure (Table 4 and Fig. 4).

Mass spectrometry analysis of the reaction products of different β-aryl oligosaccharides with *Pp*XG44 only showed release of the "hot" leaving group 2-chloro-4-nitrophenolate; no internal glucosidic bond cleavage events could be detected (data not shown).

Inhibition by β-1,4-Glucosyl Noeuromycin and Glc-Glc-oxazine—The IC_{50} values (inhibitor concentration causing 50% inhibition, Equation 3) of Glc-noeuromycin and Glc-Glc-oxazine were determined to be 1.3 and 7 μM, respectively, using GGGG-CNP as a substrate. In general, the IC_{50} value reflects K_i when substrate concentration is much lower than K_m (58). However, because of assay limitations, the lowest feasible concentration of GGGG-CNP was 10 μM, ~⅓ of K_m (29 μM, Table 4). K_i values were therefore calculated from Equation 4, which gave a K_i of 0.95 μM for Glc-noeuromycin and 5.2 μM for Glc-Glc-oxazine (supplemental Fig. 2). Intriguingly, the K_i value of the trisaccharide analog, Glc-Glc-oxazine, is more than 5-fold higher than the K_i value for the disaccharide analog Glc-noeuromycin. Apart from the cello-oligosaccharide chain length, the inhibitors also differ in the region where they interact with the catalytic residues (see below). Glc-noeuromycin has a C2 hydroxyl group that Glc-Glc-oxazine lacks (Fig. 5) and a carbon where natural sugars and Glc-Glc-oxazine have oxygen in the sugar ring.

Three-dimensional Structure of Apo-PpXG44—The structure of apo-*Pp*XG44 was refined to a maximum resolution of 1.8 Å, and the final model contains a single protein molecule in the asymmetric unit, with 512 amino acids well defined, residues 7–518 (equivalent to 42–553 in Ref. 38). The numbering for the structures reported here does not include the 35-residue signal peptide. The electron density for all 512 residues is well defined with the exception of the side chain of Gln-495. The enzyme is composed of 27 β-strands and 18 α-helices that fold into two distinct domains, the catalytic $(\beta/\alpha)_8$ triose-phosphate isomerase barrel-like domain with a small insertion and a β-sandwich domain (Fig. 6*a*).

A search of the PDB using the secondary structure matching algorithm (see under "Materials and Methods") revealed three GH44 enzymes displaying high structural homology: two endoglucanases (*Ct*Cel44A and *Ca*XG44) and a proposed bifunctional glucanase-xylanase protein from a metagenome library (*uc*CelM2). *Pp*XG44, *Ct*Cel44A, and *Ca*XG44 have highly similar folds. There are only minor differences between the apo-structure of *Pp*XG44 and those of apo-wild type *Ct*Cel44A (r.m.s.d. = 0.97 Å over 505 equivalent Cα atoms, sequence identity ~60%) and *Ca*XG44 (r.m.s.d. = 0.90 Å over 506 Cα, identity 54%). There are, however, more significant differences between these three enzymes and *uc*CelM2 (r.m.s.d. = 1.46 Å over 443 Cα, sequence identity = 33.0% compared with *Pp*XG44) (values calculated with COOT). The fold of *Pp*XG44 will therefore be only briefly summarized here.

The β-sandwich domain of *Pp*XG44 is formed by 11 β-strands in two twisted β-sheets. These strands possess a hydrophobic center, composed of residues 7–23 from the N-terminal region and 419–518 from the C-terminal region, which are linked to the first and last strands of the $(\beta/\alpha)_8$ TIM-like barrel, respectively. The major domain is composed of 17 α-helices and 12 β-strands and is composed of residues 24–112 and 159–418, which form the catalytic $(\beta/\alpha)_8$ TIM barrel-like fold. A small insertion that includes residues 113–158 forms a ψ-loop motif (described as "the small additional domain" for the *Ct*Cel44A structure (20)), which sits on the opposite site of the $(\beta/\alpha)_8$ TIM barrel-like fold from the β-sandwich domain. The ψ-loop is formed by a single α-helix and four β-strands, and its structure is stabilized by a Ca^{2+} ion that coordinates with two water molecules and ligates the side chains of Glu-56 and Asp-154 and the main-chain carbonyl oxygens of Asp-151 and Tyr-156. The overall structures of the ψ-loop and the Ca^{2+} ion are both conserved between *Pp*XG44, *Ct*Cel44A, and *Ca*XG44, as is the identity of the residues involved in coordinating the Ca^{2+} ion. In contrast, *uc*CelM2 differs from the other three enzymes in that the ψ-loop is substituted by a small domain formed by three α-helices, three β-strands, and a twisted β-strand. The ψ-loop in the three enzymes is attached to the side of the barrel. In contrast, in *uc*CelM2 the small domain

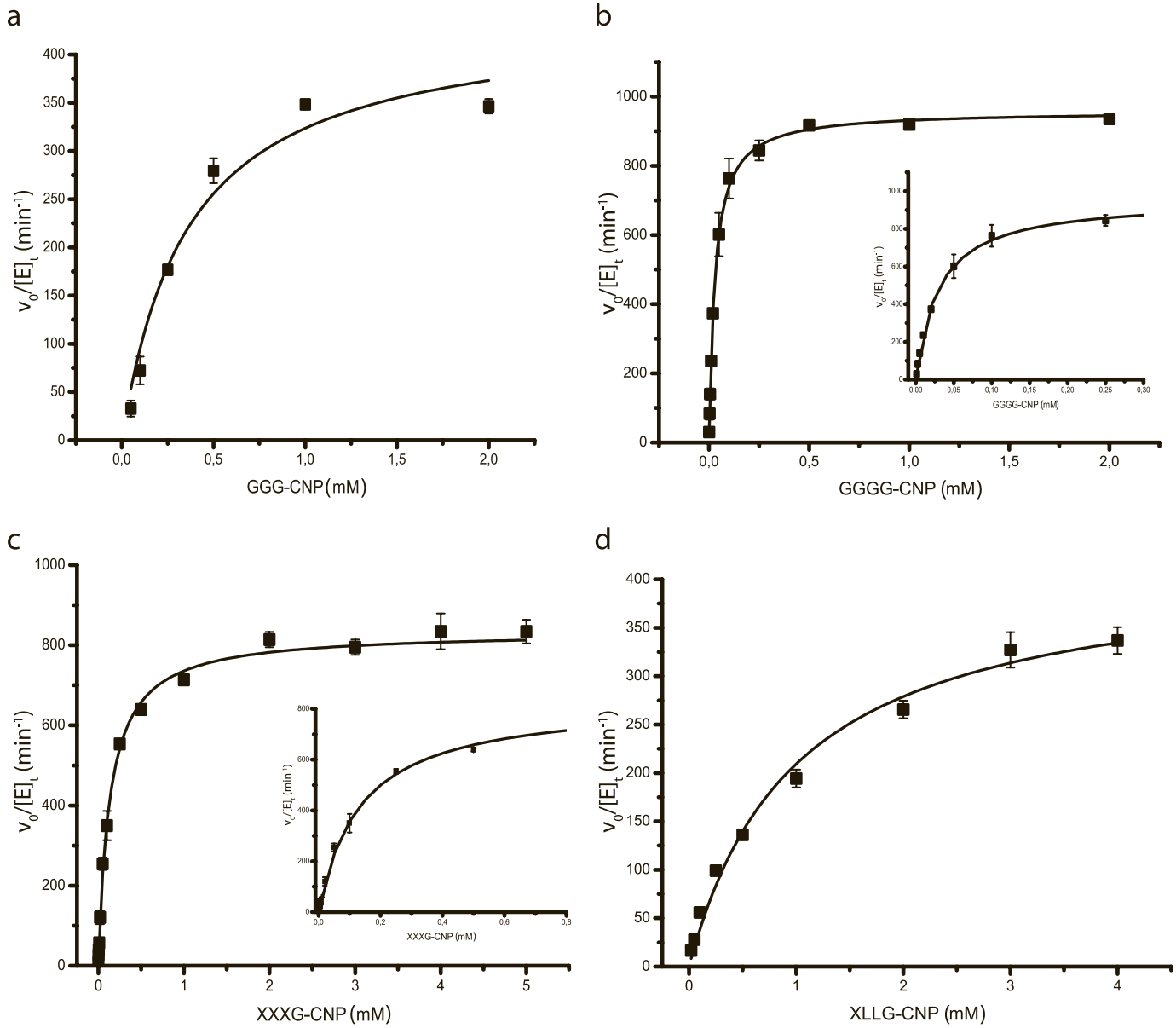

FIGURE 4. **Initial rate kinetics of *Pp*XG44 hydrolysis of β-aryl oligosaccharides.** *a,* GGG-CNP; *b,* GGGG-CNP (*inset,* low substrate concentration regime); *c,* XXXG-CNP (*inset,* low substrate concentration regime); *d,* XLLG-CNP.

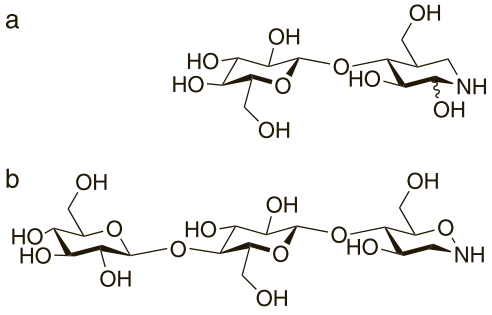

FIGURE 5. **Structures of inhibitors.** *a,* β-1,4-glucosyl-noeuromycin (Glc-noeuromycin); *b,* β-1,4-cellobiosyl-oxazine (Glc-Glc-oxazine).

extends beyond the top of the catalytic barrel and, together with an additional insertion to the barrel, including residues 244–257, deepens the binding pocket of the *uc*CelM2 active site.

The active site of *Pp*XG44 is a large cleft located in the $(\beta/\alpha)_8$ barrel. The two catalytic residues, the proton donor/acceptor Glu-187 and the catalytic nucleophile Glu-358, are both well ordered and superimpose closely to the catalytic residues of other GH44 three-dimensional structures. The nucleophile and the proton donor/acceptor are part of the NEP motif also found to be conserved in clan GH-A members (59). The γ-carbon atoms are separated by 5.3 Å, which is consistent with a retaining catalytic mechanism, as was demonstrated for *Ct*Cel44A (20).

An important difference between *uc*CelM2 and the other enzymes is that *uc*CelM2 lacks an analog to Tyr-73 (*Pp*XG44). This residue lies at one end of the binding cleft opposite Trp-66 (*Pp*XG44), which has Trp-94 as the equivalent residue in *uc*CelM2. These two residues are involved in substrate binding, as can be seen in the *Ca*XG44 E186Q mutant complexed with a

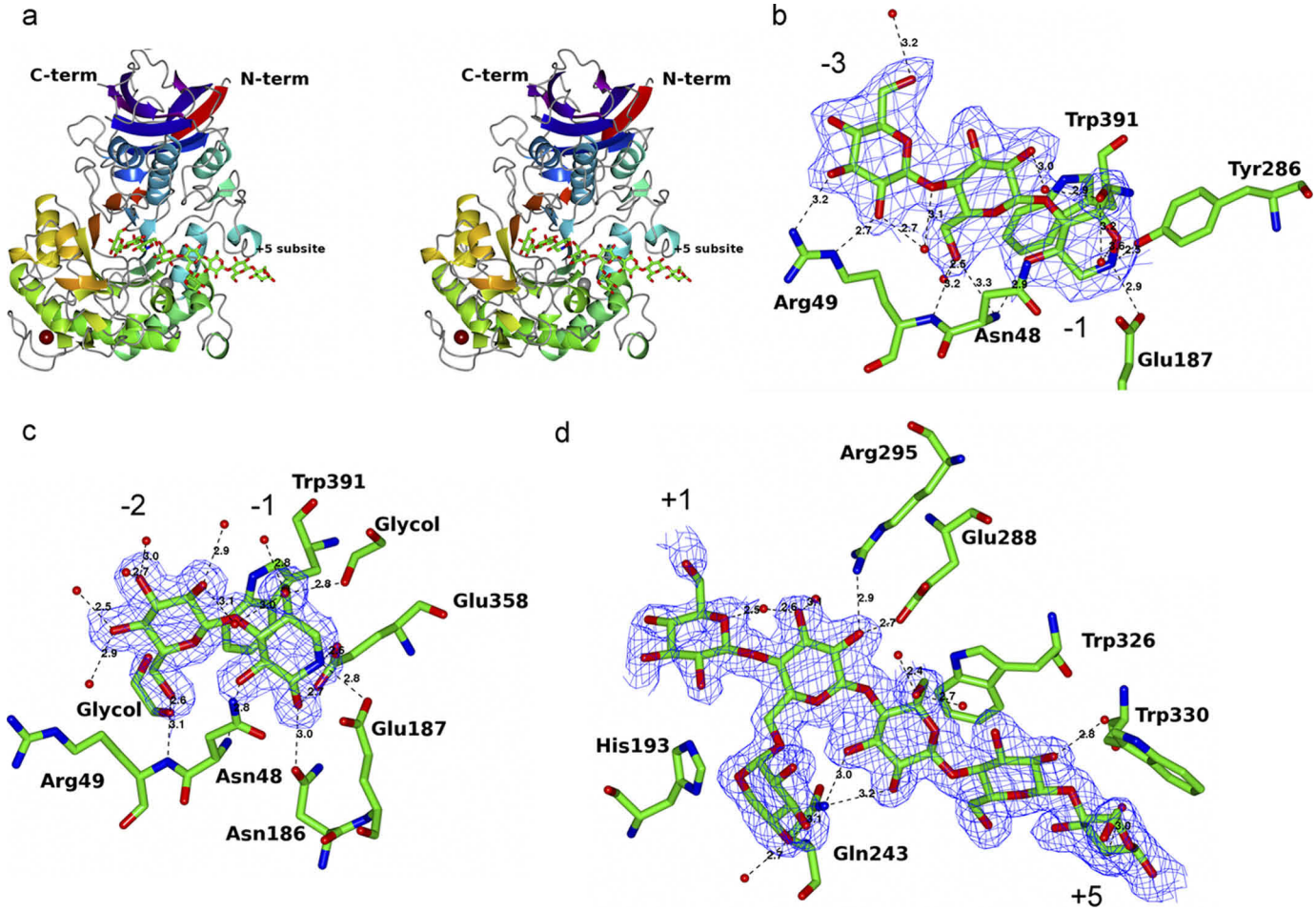

FIGURE 6. **Three-dimensional structure of *Pp*XG44 and its ligand complexes.** *a,* stereo view of a ribbon diagram of *Pp*XG44, color-ramped from the N terminus (*red*) to the C terminus (*magenta*), showing the positions of Glc-Glc-oxazine and GXGGG from their respective complexes with *Pp*XG44 as *cylinders*. A calcium ion is shown in *tan*, and chloride ion is shown in *gray*. The +5 subsite position on GXGGG is labeled. Stereo views are shown of the $2F_o - F_c$ electron density maps, contoured at 1.0 σ, around the ligand in complexes of *Pp*XG44 with Glc-Glc-oxazine (*b*), Glc-noeuromycin (*c*), and GXGGG (*d*) (for the latter complex, atoms in the subsite +1 glucose are modeled with an occupancy of 0.5; all other atoms are fully occupied). Hydrogen bonds are shown as *dashed lines*, and their approximate length in Ångstroms is indicated. This figure was drawn using CCP4*mg* (65).

high concentration of cellohexaose (PDB code 2EQD), but the α-helix on which Tyr-73 lies is substituted by a shorter loop in *uc*CelM2, which increases the size of the cleft at this position. In contrast, Trp-162 from the small domain of *uc*CelM2 that overhangs the active site is positioned beside Trp-94 and extends the length of the catalytic cleft. One may speculate that the position of this residue may suggest involvement in binding longer substrate molecules than those accepted by the other three enzymes.

There is a Tris molecule in the active site of the apo *Pp*XG44. It is located in subsite −1 of the binding cleft, where it coordinates with the side chains of Glu-187, Asp-65, and Asn-48, the main-chain nitrogen of Asn-48, and four water molecules. Tris (or BisTris) molecules have the ability to bind to carbohydrate-binding sites and often appear in crystal structures where it is used as the protein buffer or is part of the crystallization solution (examples include, but are not limited to, Refs. 60–62). Crystals of the ligand complexes were therefore grown in the absence of Tris.

Ligand binding to *Pp*XG44 was probed through a series of ligand complexes with a Glc-noeuromycin (synthesis described in Ref. 46) bound in the −1 and −2 subsites, a Glc-Glc-oxazine (synthesis described in Ref. 47, −3 to −1 subsites), and a complex with a xyloglucan oligosaccharide that occupies sites +2 to +1, with a clear α-1,6-xyloside in the +2′ subsite (Fig. 6). The resolutions of the complexes are 1.9, 2.3, and 1.7 Å for Glc-noeuromycin, Glc-Glc-oxazine, and GXGGG, respectively (Table 1).

Glc-noeuromycin-PpXG44—Glc-noeuromycin was soaked into a crystal of a *Pp*XG44 mutant with increased stability. Except for the presence of a Glc-noeuromycin molecule instead of BisTris, four ethylene glycol molecules, and the absence of a polyethylene glycol molecule, there are no other noticeable differences between the structure of the Glc-noeuromycin-*Pp*XG44 complex and that of the apo-structure of *Pp*XG44 (r.m.s.d. = 0.15 Å over 512 Cα). The nitrogen atom of noeuromycin in the −1 subsite lies within hydrogen bonding distance of the catalytic glutamates 358 and 187 (2.6 and 2.8 Å, respectively) and the ligand is in a relaxed $^4C_1$ chair conformation. The Glc moiety, in the −2 subsite, adopts the position seen for Glc(−2) in cello-oligosaccharide complexes with *Ca*XG44 (22).

Furthermore, the noeuromycin O2 forms hydrogen bonds with OE2 Glu-358 (2.6 Å) and OD1 Asn-186 (3.0 Å). Both the −1 noeuromycin and the Glc(−2) moieties interact with main-chain amide groups (O3 to Asn-48 and O6 to Arg-49, respec-

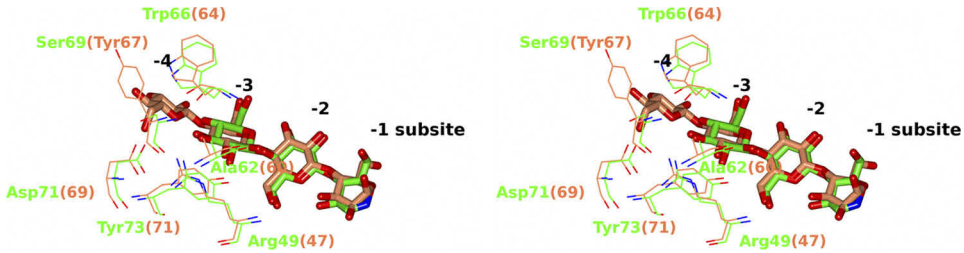

FIGURE 7. **Overlay of *Pp*GH44 enzyme complexes.** Stereo view of overlay of cellotetraose from *Ct*GH44A (PDB code 2E0P) (in *coral*) with Glc-Glc-oxazine from the *Pp*XG44 complex (in *green*) shows residues from both structures neighboring the −4 subsite region (which is occupied in PDB code 2E0P but unoccupied in the Glc-Glc-oxazine complex). This figure was drawn using CCP4*mg* (65).

tively). O2, O3, and O4 of the Glc(−2) and O6 of noeuromycin form hydrogen bonds to water molecules at the surface of the ligand-binding cleft. In addition, O6 of the noeuromycin is within hydrogen bonding distance of an ethylene glycol molecule bound within the active site (2.9 Å). This ethylene glycol molecule might reflect the position of a xyloside in xylogluco-oligosaccharides in the −1 subsite. The noeuromycin moiety also has hydrophobic interactions with Trp-391.

Glc-Glc-oxazine-PpXG44—Glc-Glc-oxazine (and XXXGXXXG, below) were soaked into crystals of the *Pp*XG44 nucleophile mutant E358S. The −2 subsite glucosyl moiety as well as the −1 subsite oxazine of the Glc-Glc-oxazine ligand exhibit very similar interactions to Glc-noeuromycin with the oxazine taking a relaxed $^4C_1$ chair conformation. The active site in the nucleophile *Pp*XG44 mutant E358S was slightly altered. Whereas the noeuromycin complex features a hydrogen bond from its nitrogen to the catalytic nucleophile, the N5 nitrogen of the oxazine interacts with the hydroxyl of Tyr-286 (at a distance 2.5 Å; Fig. 6). The noeuromycin O2 interaction with Asn-186 is also lost in the oxazine, because it lacks the O2 group, which could explain why Glc-noeuromycin is a better inhibitor than Glc-Glc-oxazine (Fig. 5 and supplemental Fig. 2).

As in the Glc-noeuromycin structure, O3 of the −1 sugar analog interacts with the main-chain amide nitrogen of Asn-48, but the O6 glucose of Glc(−2) hydrogen bonds to a water molecule only and thus not to the main-chain amide of Arg-49. The Glc(−2) O3 makes no hydrogen bonds with water molecules, whereas the equivalent atom in Glc-noeuromycin forms two, but the latter structure is at higher resolution, and therefore, the water molecules are likely to be more ordered.

The Glc(−3) forms a close interaction with the side chain of Arg-49 (O2 to NE and O3 to NH_2, 2.7 and 3.1 Å respectively), and O1 and O2 of Glc(−3) form hydrogen bonds with a water molecule that interacts with OD2 Asp-65 and O Ala-62. The Glc-Glc-oxazine structure has two protein molecules (A and B) in the unit cell, and the active-site clefts are closely associated with the other molecule. The O4 Glc(−3) of molecule A is just over 5 Å from ND1 and OD1 Asn-399 of molecule B. In the Glc-Glc-oxazine of molecule B, O4 Glc(−3) is 3.1 Å from OD1 N399A, and thus the ligand is within hydrogen bonding distance of a residue on the other protein chain.

Complex with an XXXGXXXG-derived Oligosaccharide—Although crystals were soaked with the ligand XXXGXXXG, unambiguous electron density was only observed for a XGGG xylogluco-oligosaccharide, occupying the +2 to +5 subsites (and thus with a Xyl(+2′); Fig. 6). There is partial, low occupancy, density for a Glc(+1) that has been modeled with an occupancy of 0.5.

The partially degraded/mobile ligand fragment (hereafter GXGGG) shows few hydrogen bonds with the protein at subsites +3 to +5; it lies on an exposed surface of the protein. There is a hydrogen bond from the O2 of Glc(+5) to a water molecule (3 Å) that is 2.9 Å from the main-chain carbonyl oxygen of Trp-330, and for Glc(+4) between O2 and a single water molecule (2.8 Å). Glc(+3) is hydrogen-bonded to NE2 Gln-243 via O3 and to two water molecules via O6. Two tryptophan residues, 326 and 330, form a hydrophobic lining for the +3 to +5 subsites. The Glc(+2) interacts via O2, with OE2 Glu-288 (2.7 Å) and NH1 Arg-295 (2.9 Å). The O4 of the Xyl(+2′) forms hydrogen bonds to NE2 Gln-243 (3.1 Å) and also to a water molecule that is hydrogen-bonded to NZ Lys-202. The Xyl(+2′) also makes stacking interactions with the imidazole ring of His-193 (Fig. 6).

Unidentified densities were observed at O6 of Glc(+1), where a disordered +1′ Xyl may be partially bound, and at O1 of Glc(+5) where Glc(+6) would lie. These residues could, however, not be modeled, reflecting disorder and/or extremely low occupancy.

What Determines Xyloglucan Specificity? Kinetics in Light of Three-dimensional Structure—*Pp*XG44 cleaves XXXGXXXG-based xyloglucan units at a substituted glycoside, with X in the −1 subsite, *i.e.* XXX ↓ GXXXG. Given the importance of the −4 subsite to catalysis, as reflected in the kinetic data for cello-oligosaccharide hydrolysis (Tables 3 and 4), the observation that the XXXGXXXG substrate is bound in a manner that does not harness the −4 subsite is particularly intriguing. This preference has two potential origins, a discrimination against a xylose-substituted glucoside in the −4 subsite and/or a similar legislation against a substituted glucoside in the +1 subsite. There is certainly strong kinetic support for a partial occlusion of xylose from the −4′ subsite as follows: (*a*) GXX ↓ G ↓ XXXG oligosaccharide is cleaved roughly equally at both X and G (as shown), and (*b*) chromogenic substrate GGGG-CNP is a 25-fold better substrate than XXXG-CNP (based on k_{cat}/K_m), although here the contributions of the −2′- and −3′-xylosides partially mask interpretation.

Inspection of the three-dimensional structure certainly supports the importance of the −4 subsite. Comparison of the *C. thermocellum* endoglucanase *Ct*Cel44A (20) complexes with the −1 to −3 Glc-Glc-oxazine complex of *Pp*XG44 allows us to speculate on the potential steric barriers to complex formation in the −4′ subsite (Fig. 7). It can be seen from the structures that

the −1 and −3 subsites bind a glucosyl moiety in which the O6 group points out into solvent and that there would likely be neither barrier to nor obvious benefit from 6-linked xylosides in these positions. The −2 sugar points the O6 atom into the protein interior, and here it is clearly important that *Pp*XG44 has a pocket large enough to accommodate (at least) xylose in this region (indeed a glycerol moiety is found in this pocket). As with the −2 subsite, a glucoside in the critical −4 subsite points its O6 toward the protein, but here in neither *Ct*Cel44A nor *Pp*XG44 is there an equivalent pocket. Instead, although the −4 glucoside would interact with Trp-66 (*Pp*XG44 numbering) via stacking and through H-bonds of O6 to Asp-71, a xyloside attached to this sugar would likely clash into the protein (Fig. 7). Thus, the −4 subsite preferably binds Glc and, to accommodate Xyl, the glucan chain must take on a different conformation (to that seen for the −4 glucosyl unit in *Ct*Cel44A, PDB code 2E0P) with the subsequent loss in binding energy reflected in the kinetics (Table 4).

So although the role of the −4 subsite in contributing to specificity is clear, the preferences of the +1 subsite are much harder to assess because information into this site is not given by the chromogenic substrates. Soaking of crystals of *Pp*XG44, in a crystal form in which the −3 to +5 subsites are accessible, resulted in a "GXGGG" oligosaccharide complex occupying subsites +1 to +5 with an ordered xyloside in +2′. However, occlusion of the −4 subsite through packing interactions likely discriminates against binding of the XXXGXXXG across the active center in this form. There is low level density for a glucose in +1 (modeled at an occupancy 0.5), and it there may be sufficient space for a +1′ xyloside linked to the O6. Such a structural accommodation would seem to resonate with the kinetic observation that GXXGXXXG is hydrolyzed roughly equally in a "−3 to +5" *versus* "−4 to +4" binding mode, whereas XXXGXXXG has a marked preference for subsites −3 to +5 (Table 3). Thus, when the nonreducing end xyloside is removed, the enzyme shows no apparent discrimination against an X in the +1 subsite).

Summary—The controlled enzymatic degradation of plant cell walls is of increasing economic importance to today's society. The ubiquity of xyloglucans in the plant cell walls makes them important polysaccharides to harness. The specificity of glycoside hydrolases for these highly decorated polysaccharides is complex; xyloglucans are hydrolyzed by classical "cellulases" (endoglucanases), which in some cases merely tolerate branching substitutions, as well as by obligate xyloglucanases that harness certain substitutions for activity (17, 19, 23, 44, 63). Elucidation of the true specificities of endo-glucanases is further complicated by the reality that undecorated β-1,4-glucans are insoluble in water and hence difficult to assay quantitatively, whereas xyloglucans are far more soluble. Given appropriate accommodation of substituents in enzyme active sites, they may be generally better substrates in a similar manner to the artificial substrates carboxymethylcellulose and hydroxyethylcellulose. Here, we show that *Pp*XG44 uses tamarind xyloglucan as its "best" substrate; hence, it is classified here as a "xyloglucanase." Similar results have very recently been obtained on a GH44 enzyme from the ruminant bacterium *Ruminococcus flavefaciens* FD-1 (64), which has also been shown to hydrolyze tamarind xyloglucan, thus suggesting that xyloglucanase activity may be a general feature of the GH44 family.

The *Pp*GH44 enzyme has an extended V-shaped substrate-binding site with a linear interaction length of ~50 Å, comprising nine primary subsites (−4 to +5). Particularly notable is that *Pp*XG44 appears to have evolved a substrate-binding cleft in which the (potentially substituted) O6 atoms of the glucan chain point outward into the solvent in the −3, −1, and +3 to +5 subsites. This suggests an "xyloglucan-tolerant" mode of action. Yet α-1–6 linked xyloside moieties have the potential to interact constructively with the protein in the −2 and +2 subsites, as would be expected of a true xyloglucanase. Furthermore, *Pp*XG44 displays a relatively rare mode of action, in which the classical XXX ↓ GXXXG repeating unit of xyloglucans is preferentially cut at a branched glucosyl unit (indicated). This cleavage mode may in fact be encouraged by steric and electrostatic considerations in the −4 (potential clashes with pendant Xyl units, H-bond donation from O6 on main-chain Glc) and the +5 subsites (positive stacking interactions with main-chain Glc). Such unusual specificity may well provide a complementary synergistic tool in large scale applications with plant polymers.

Acknowledgments—We thank Novozymes A/S for supplying the enzyme variants. The staff of the Diamond Light Source are thanked for provision of data collection facilities.

REFERENCES

1. Pauly, M., Albersheim, P., Darvill, A., and York, W. S. (1999) *Plant J.* **20,** 629–639
2. Hsieh, Y. S., and Harris, P. J. (2009) *Mol. Plant* **2,** 943–965
3. Hoffman, M., Jia, Z., Peña, M. J., Cash, M., Harper, A., Blackburn, A. R., 2nd, Darvill, A., and York, W. S. (2005) *Carbohydr. Res.* **340,** 1826–1840
4. Vogel, J. (2008) *Curr. Opin. Plant Biol.* **11,** 301–307
5. Nishikubo, N., Awano, T., Banasiak, A., Bourquin, V., Ibatullin, F., Funada, R., Brumer, H., Teeri, T. T., Hayashi, T., Sundberg, B., and Mellerowicz, E. J. (2007) *Plant Cell Physiol.* **48,** 843–855
6. Buckeridge, M. S., dos Santos, H. P., and Tine, M. A. (2000) *Plant Physiol. Biochem.* **38,** 141–156
7. Kumar, C. S., and Bhattacharya, S. (2008) *Crit. Rev. Food Sci. Nutr.* **48,** 1–20
8. Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Högberg, P., Linder, S., Mackenzie, F. T., Moore, B., 3rd, Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V., and Steffen, W. (2000) *Science* **290,** 291–296
9. Zhao, M., and Running, S. W. (2010) *Science* **329,** 940–943
10. Fry, S. C., York, W. S., Albersheim, P., Darvill, A., Hayashi, T., Joseleau, J. P., Kato, Y., Lorences, E. P., Maclachlan, G. A., Mcneil, M., Mort, A. J., Reid, J. S. G., Seitz, H. U., Selvendran, R. R., Voragen, A. G., and White, A. R. (1993) *Physiol. Plant.* **89,** 1–3
11. Hilz, H., de Jong, L. E., Kabel, M. A., Verhoef, R., Schols, H. A., and Voragen, A. G. (2007) *Carbohydr. Res.* **342,** 170–181
12. Peña, M. J., Darvill, A. G., Eberhard, S., York, W. S., and O'Neill, M. A. (2008) *Glycobiology* **18,** 891–904
13. Vincken, J. P., Beldman, G., and Voragen, A. G. (1997) *Carbohydr. Res.* **298,** 299–310
14. Davies, G., and Henrissat, B. (1995) *Structure* **3,** 853–859
15. Gilbert, H. J., Stålbrand, H., and Brumer, H. (2008) *Curr. Opin. Plant Biol.* **11,** 338–348
16. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) *Nucleic Acids Res.* **37,** D233–D238
17. Gloster, T. M., Ibatullin, F. M., Macauley, K., Eklöf, J. M., Roberts, S.,

Turkenburg, J. P., Bjørnvad, M. E., Jørgensen, P. L., Danielsen, S., Johansen, K. S., Borchert, T. V., Wilson, K. S., Brumer, H., and Davies, G. J. (2007) *J. Biol. Chem.* **282,** 19177–19189
18. Baumann, M. J., Eklöf, J. M., Michel, G., Kallas, A. M., Teeri, T. T., Czjzek, M., and Brumer, H., 3rd (2007) *Plant Cell* **19,** 1947–1963
19. Mark, P., Baumann, M. J., Eklöf, J. M., Gullfot, F., Michel, G., Kallas, A. M., Teeri, T. T., Brumer, H., and Czjzek, M. (2009) *Proteins Struct. Funct. Bioinformat.* **75,** 820–836
20. Kitago, Y., Karita, S., Watanabe, N., Kamiya, M., Aizawa, T., Sakka, K., and Tanaka, I. (2007) *J. Biol. Chem.* **282,** 35703–35711
21. Najmudin, S., Guerreiro, C. I., Carvalho, A. L., Prates, J. A., Correia, M. A., Alves, V. D., Ferreira, L. M., Romão, M. J., Gilbert, H. J., Bolam, D. N., and Fontes, C. M. (2006) *J. Biol. Chem.* **281,** 8815–8828
22. Warner, C. D., Hoy, J. A., Shilling, T. C., Linnen, M. J., Ginder, N. D., Ford, C. F., Honzatko, R. B., and Reilly, P. J. (2010) *Appl. Environ. Microbiol.* **76,** 338–346
23. Martinez-Fleites, C., Guerreiro, C. I., Baumann, M. J., Taylor, E. J., Prates, J. A., Ferreira, L. M., Fontes, C. M., Brumer, H., and Davies, G. J. (2006) *J. Biol. Chem.* **281,** 24922–24933
24. Yaoi, K., Kondo, H., Hiyoshi, A., Noro, N., Sugimoto, H., Tsuda, S., Mitsuishi, Y., and Miyazaki, K. (2007) *J. Mol. Biol.* **370,** 53–62
25. Yaoi, K., Kondo, H., Hiyoshi, A., Noro, N., Sugimoto, H., Tsuda, S., and Miyazaki, K. (2009) *FEBS J.* **276,** 5094–5100
26. Yaoi, K., Kondo, H., Noro, N., Suzuki, M., Tsuda, S., and Mitsuishi, Y. (2004) *Structure* **12,** 1209–1217
27. Vlasenko, E., Schülein, M., Cherry, J., and Xu, F. (2010) *Bioresour. Technol.* **101,** 2405–2411
28. Sinitsyna, O. A., Fedorova, E. A., Pravilnikov, A. G., Rozhkova, A. M., Skomarovsky, A. A., Matys, V. Y., Bubnova, T. M., Okunev, O. N., Vinetsky, Y. P., and Sinitsyn, A. P. (2010) *Biochemistry* **75,** 41–49
29. Wong, D. D., Chan, V. J., McCormack, A. A., and Batt, S. B. (2010) *Appl. Microbiol. Biotechnol.* **86,** 1463–1471
30. Wong, D. D., Chan, V. J., McCormack, A. A., and Batt, S. B. (2010) *Protein Pept. Lett.* **17,** 803–808
31. Powlowski, J., Mahajan, S., Schapira, M., and Master, E. R. (2009) *Carbohydr. Res.* **344,** 1175–1179
32. Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., and Foust, T. D. (2007) *Science* **315,** 804–807
33. Vincken, J. P., Beldman, G., and Voragen, A. (1994) *Plant Physiol.* **104,** 99–107
34. Mishra, A., and Malhotra, A. V. (2009) *J. Materials Chem.* **19,** 8528–8536
35. Nam, K. H., Kim, S. J., and Hwang, K. Y. (2009) *Biochem. Biophys. Res. Commun.* **383,** 183–186
36. Cho, K. M., Hong, S. Y., Lee, S. M., Kim, Y. H., Kahng, G. G., Kim, H., and Yun, H. D. (2006) *Appl. Microbiol. Biotechnol.* **73,** 618–630
37. Cho, K. M., Math, R. K., Hong, S. Y., Asraful Islam, S. M., Kim, J. O., Hong, S. J., Kim, H., and Yun, H. D. (2008) *Biotechnol. Lett.* **30,** 1061–1068
38. Schnorr, K., Jørgensen, P. L., and Schulein, M. (November 9, 2004) U. S. Patent 6815192
39. Spizizen, J. (1958) *Proc. Natl. Acad. Sci. U.S.A.* **44,** 1072–1078
40. McFeeters, R. F. (1980) *Anal. Biochem.* **103,** 302–306
41. Mopper, K., and Gindler, E. M. (1973) *Anal. Biochem.* **56,** 440–442
42. Maris, A., Kaewthai, N., Eklöf, J. M., Miller, J. G., Brumer, H., Fry, S. C., Verbelen, J. P., and Vissenberg, K. (2011) *J. Exp. Bot.* **62,** 261–271
43. Brumer, H., 3rd, Zhou, Q., Baumann, M. J., Carlsson, K., and Teeri, T. T. (2004) *J. Am. Chem. Soc.* **126,** 5715–5721
44. Ibatullin, F. M., Baumann, M. J., Greffe, L., and Brumer, H. (2008) *Biochemistry* **47,** 7762–7769
45. Claeyssens, M., and Henrissat, B. (1992) *Protein Sci.* **1,** 1293–1297
46. Meloncelli, P. J., Gloster, T. M., Money, V. A., Tarling, C. A., Davies, G. J., Withers, S. G., and Stick, R. V. (2007) *Aust. J. Chem.* **60,** 549–565
47. Macdonald, J. M., Stick, R. V., Tilbrook, D. M., and Withers, S. G. (2002) *Aust. J. Chem.* **55,** 747–752
48. Cer, R. Z., Mudunuri, U., Stephens, R., and Lebeda, F. J. (2009) *Nucleic Acids Res.* **37,** W441–W445
49. Collaborative Computational Project No. 4 (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763
50. Leslie, A. G. (1992) *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, No. 26, Daresbury Laboratory, Warrington, UK
51. Cowtan, K. (2006) *Acta Crystallogr. D Biol. Crystallogr.* **62,** 1002–1011
52. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
53. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
54. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) *Nucleic Acids Res.* **35,** W375–W383
55. Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2256–2268
56. Juhász, T., Szengyel, Z., Réczey, K., Siika-Aho, M., and Viikari, L. (2005) *Process Biochem.* **40,** 3519–3525
57. Yaoi, K., Nakai, T., Kameda, Y., Hiyoshi, A., and Mitsuishi, Y. (2005) *Appl. Environ. Microbiol.* **71,** 7670–7678
58. Burlingham, B. T., and Widlanski, T. S. (2003) *J. Chem. Educ.* **80,** 214–218
59. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92,** 7090–7094
60. Brzozowski, A. M., Lawson, D. M., Turkenburg, J. P., Bisgaard-Frantzen, H., Svendsen, A., Borchert, T. V., Dauter, Z., Wilson, K. S., and Davies, G. J. (2000) *Biochemistry* **39,** 9099–9107
61. Sevcík, J., Solovicová, A., Hostinová, E., Gasperík, J., Wilson, K. S., and Dauter, Z. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54,** 854–866
62. Tailford, L. E., Money, V. A., Smith, N. L., Dumon, C., Davies, G. J., and Gilbert, H. J. (2007) *J. Biol. Chem.* **282,** 11291–11299
63. Saura-Valls, M., Fauré, R., Brumer, H., Teeri, T. T., Cottaz, S., Driguez, H., and Planas, A. (2008) *J. Biol. Chem.* **283,** 21853–21863
64. Warner, C. D., Go, R. M., Garcia-Salinas, C., Ford, C., and Reilly, P. J. (2010) *Enzyme Microb. Technol.* **48,** 27–32
65. Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A., and Noble, M. (2011) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2288–2294