Crystal Structures of Clostridium thermocellum Xyloglucanase, XGH74A, Reveal the Structural Basis for Xyloglucan Recognition and Degradation*

Received for publication, April 13, 2006, and in revised form, June 12, 2006 Published, JBC Papers in Press, June 13, 2006, DOI 10.1074/jbc.M603583200

Carlos Martinez-Fleites†, Catarina I. P. D. Guerreiro†, Martin J. Baumann§, Edward J. Taylor†, José A. M. Prates§, Luis M. A. Ferreira§, Carlos M. G. A. Fontes§, Harry Brumer¶, and Gideon J. Davies†‡

From the †York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, United Kingdom, the §CIISA, Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal, and the ¶School of Biotechnology, Royal Institute of Technology (KTH), AlbaNova University Centre, 106 91 Stockholm, Sweden

The enzymatic degradation of the plant cell wall is central both to the natural carbon cycle and, increasingly, to environmentally friendly routes to biomass conversion, including the production of biofuels. The plant cell wall is a complex composite of cellulose microfibrils embedded in diverse polysaccharides collectively termed hemicelluloses. Xyloglucan is one such polysaccharide whose hydrolysis is catalyzed by diverse xyloglucanases. Here we present the structure of the Clostridium thermocellum xyloglucanase Xgh74A in both apo and ligand-complexed forms. The structures, in combination with mutagenesis data on the catalytic residues and the kinetics and specificity of xyloglucan hydrolysis reveal a complex subsite specificity accommodating seventeen monosaccharide moieties of the multibranched substrate in an open substrate binding terrain.

Plant cell wall polysaccharides are the most abundant carbohydrate polymers in nature and constitute an important renewable natural source of energy available for conversion to biofuels (1). The plant resource is, however, difficult to exploit primarily because its components are extremely resistant to degradation; plant cell wall polysaccharides are often present as insoluble, often cross-linked, structures in which cellulose is the most abundant component. In the cell wall of flowering plants, cellulose is cross-linked by two major types of glycans: xyloglucans and glucuronoarabinoxylans. The xyloglucans form a complex network of hydrogen-bonded interactions with cellulose microfibrils that confers rigidity and extensibility to the walls of all dicotyledons and about one-half of monocotyledons (2). The xyloglucan polysaccharide consists of a linear chain of β-1,4-D-glucan regularly substituted with α-1,6 D-xylosyl units, which is, in a species-dependent manner, further derivatized with α-L-arabinose or β-D-galactose (2, 3), Fig. 1. In primary cell wall xyloglucans, the first galactose moiety in the oligosaccharide repeat is commonly substituted with α-1,2 L-fucose (4). Considerable interest in the structure, biosynthesis, and enzymatic modification of xyloglucans has been sustained because of the important role these polysaccharides play in plant cell wall morphogenesis (4–8), as well as the emerging technical applications of xyloglucans in food products, pharmaceutical delivery (9, 10), cellulose fiber modification (11–13), and biofuel production (1).

Microorganisms have evolved sophisticated mechanisms to degrade plant cell wall polysaccharides and consequently exploit this rich carbon and energy source. Aerobic bacteria and fungi secrete several individual enzymes that synergistically degrade plant cell walls (14). Some anaerobic microorganisms, notably Clostridia, utilize a large multi-enzymatic complex called the cellulosome (15). The cellulosome displays a consortium of hydrolytic plant cell wall degrading enzymes, which may change with time, including cellulases, hemicellulases, pectinases, and various esterases. The Clostridium thermocellum (Ct) cellulosome is one of the best studied cellulosome systems. This cellulosome is a multiprotein complex of about 3 MDa and displays endoglucanase, cellobiohydrolase (exoglucanase), xylanase, chitinase, and β-glucanase (lichenase) enzymatic activity (16). Cellulosome enzymes are tethered to the scaffolding protein of the complex through the interaction of dockerin domains with one of the nine cohesin platforms of the scaffold (17), Fig. 2A.

There is particular interest in the exploitation of the cellulosome from C. thermocellum, primarily because of the potential it offers for the degradation of lignocellulosic waste and subsequent generation of ethanol (reviewed in Ref. 1). To date, 71 open reading frames have been identified as cellulosomal components (18).

* This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC), the Swedish Foundation for Strategic Research, the KTH Biofibre Materials Centre, and the Fundação para a Ciência e Tecnologia. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (codes 2CN2 and 2CN3) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

† These authors contributed equally to this work.

‡ Fellow (Rådsforskare) of the Swedish Research Council.

§ To whom correspondence should be addressed. E-mail: davies@ysbl.york.ac.uk.

¶1 The abbreviations used are: Ct, Clostridium thermocellum; RMSD, root mean-squared deviation; PDB, Protein Data Bank; PEG, polyethylene glycol; HPAEC-PAD, high performance anion-exchange chromatography with pulsed amperometric detection; CA2Y, carbohydrate active enzyme; GH, glycoside hydrolase; XGO, xylogluco-oligosaccharides; CMC, carboxymethyl cellulose.
XYLOGLUCANASE STRUCTURE AND COMPLEX

The cellulosomal xyloglucanase A from *C. thermocellum* is a bi-modular enzyme containing an N-terminal family 74 glycoside hydrolase (GH) catalytic domain followed by C-terminal dockerin (19). To express the xyloglucanase catalytic module in *Escherichia coli*, Xgh74A hereafter, the DNA fragment encoding the protein domain was amplified by PCR from *C. thermocellum* YS genomic DNA with the thermostable DNA polymerase *Pfu* Turbo (Stratagene). The primers, 5'-CTCGCTAGATTTCGACGCAAGGCTGTA-3' and 5'-CACCTCGAGATCTGAAGCAGGTTGCAG-3', incorporated NheI and XhoI restriction sites, which are depicted in bold. The amplified product was ligated into pMOSBlue (Amersham Biociences) and sequenced to ensure that no mutations had occurred during the polymerase chain reaction. The recombinant pMOSBlue derivative was digested with NheI and Xhol, and the excised Xgh74A encoding gene was cloned into the similarly restricted expression vector pET21a to generate pCG1. CtXgh74A encoded by pCG1 contains a C-terminal His$_6$-tag.

**Site-directed Mutagenesis**

Mutants of Xgh74A were generated using a PCR-based QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer’s instructions and using pCG1 as the template DNA. The sequences of the primers used to generate the protein mutants were as follows: 5'-GATTATGCCACGTCGCCTATGAGAGGCAGCTTACC-3' and 5'-GTTAGGCTCTTCCAGTACCGGGA-3', D480A. The mutated DNA sequences were sequenced to ensure that only the appropriate mutations had been incorporated into the nucleic acid.

**Production of Recombinant Xgh74A and Mutants**

*E. coli* Tuner cells (Novagen) harboring the pET21a-Xgh74A plasmid were cultured in LB containing ampicillin to mid-exponential phase (A$_{600}$ 0.6) at which point, cultures were transferred to 20 °C and induced by addition of 1 mM isopropyl 1-thio-β-d-galactopyranoside (IPTG) whereupon they were grown for a further 20 h. SeMet-labeled Xgh74A was produced in *E. coli* B834 (DE3) containing the pET21a-Xgh74A plasmid with recombinant protein expression specificity determinants responsible for xyloglucan recognition and provide insight into the hydrolysis of this important plant cell wall polysaccharide.
induced by 1 mM IPTG and incubation at 20 °C for 20 h. Cells were harvested by centrifugation and disrupted by sonication in 20 mM HEPES-NaOH, 400 mM NaCl, pH 7.5 buffer. The cell-free extract was incubated at 65 °C for 15 min and centrifuged to remove insoluble material. Samples were further purified by Ni²⁺ affinity chromatography and buffer exchanged to 10 mM HEPES-NaOH pH 7.5. Xgh74A samples thus purified were assessed pure by SDS-PAGE and were used for crystallization experiments. Xgh74A-D70A and D480A mutants were produced and purified following the native Xgh74A expression and purification protocols.

**High Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD)**

Oligosaccharides were analyzed on a Waters HPLC system with a Dionex Carapoc PA100 column. A Waters Concorde electrochemical detector was used in PAD mode with a 3-mm gold electrode and a HyREF platinum reference electrode. Two optimized gradients were used for different sizes of xyloglucan-oligosaccharides.

Gradient A (for Analysis of Glc₄-based Xyloglucan-oligosaccharides)—Solvent A: 100 mM NaOH, solvent B: 100 mM NaOH 200 mM NaOAc. Gradient program: 0–4 min, 100 mM NaOH; 60 mM NaOAc; 4–10 min linear gradient of 60 mM NaOAc to 117 mM NaOAc; 10–11 min 200 mM NaOAc. The system was then re-equilibrated for 4 min with the initial conditions prior to the next injection.

Gradient B (for Analysis of Higher Order Xyloglucan-oligosaccharides)—Solvent A: 100 mM NaOH, solvent B: 100 mM NaOH 500 mM NaOAc. Gradient program: 0–3 min, 100 mM NaOH; 40 mM NaOAc; 3–11 min linear gradient of 40 mM NaOAc to 170 mM NaOAc; 11–18 min. linear gradient 170 mM NaOAc to 200 mM NaOAc; 18–19 min 500 mM NaOAc. The system was then re-equilibrated for 4 min with the initial conditions prior to the next injection.

**Mass Spectrometry of Xyloglucan-oligosaccharides**

Mass spectrometric analysis was performed with a Q-Tof™ 2 mass spectrometer fitted with a nanoflow ion source (Waters Corporation, Micromass MS Technologies, Manchester, United Kingdom). External calibration of the TOF analyzer (single-reflection mode, resolution 10000 FWHM) was obtained over the m/z range 50–1000 using a solution of NaI (1.5 g/liter) in 1:1 2-propyl alcohol/water. Solutions of xyloglucan-oligosaccharides (typical concentration 0.01–0.1 g/liter in 1:1 MeOH/water containing 0.5 mM NaCl) were infused into the ion source (3 kV) at 200 nl/min (syringe pump). The cone voltage was varied between at 35 V and 130 V to optimize the intensity of [M+Na]⁺ and [M+2Na]⁺⁺ ions. Argon was present in the collision cell at all times, and the collision energy was 10 V. A scan time of 2.5 s with an interscan delay of 0.1 s was used, and continuum data were collected until an acceptable signal-to-noise ratio was achieved after the combination of individual spectra (typically 1–30 spectra).

**Preparation of Xyloglucan-oligosaccharides (XGOs) from Deoiled Tamarind Kernel Powder**

**Mixture of Xyloglucan-oligosaccharides Based on a Glc₄ Backbone (XLLG, XLXG, XXLG, XXXG)—**Deoiled tamarind kernel powder (Saiguru Food Gum Manufacturer, Mumbai, India) (20 g) was suspended in ammonium acetate buffer (1 liter; 10 mM; pH 4.5) at 60 °C and was vigorously stirred until homogenous. Then the suspension was cooled to 30 °C and crude cellulase from *Trichoderma reesei* (Fluka) was added (100 mg, 500 units). The resulting solution was incubated at 30 °C during 18 h under gentle stirring. The progress of the digestion was monitored by HPAEC-PAD (gradient B). The solution was filtered on a glass fiber filter, 1 ml of NH₃ (37% in H₂O) was added, and the basic solution was pumped over a Q-Sepharose (GE Healthcare) column (10 cm high, 2.6 cm diameter) to remove the cellulase. The resulting solution was freeze-dried to yield a mixture of XGOs as a white powder (typical yield 8–9 g). The following oligosaccharide composition was obtained by HPAEC-PAD (gradient A): XXXG, XLXG, XXLG, XLLG (2:1:3:3).

**Mixture of Higher Order Xyloglucan Oligosaccharides (Glc₄–Glc₁₆ Backbone)—**A modified protocol based on that described by Vincken et al. (26) was used. One gram of deoiled tamarind kernel powder was dissolved in ammonium acetate buffer (50 ml; 10 mM; pH 4.5) at 60 °C for 1 h. The solution was cooled to 30 °C and 1 mg (1 unit) of crude cellulase (*T. reesei*, Fluka) was added. The progress of the digestion was monitored by HPAEC-PAD (gradient B). When HPAEC-PAD analysis indicated the presence of predominantly Glc₄ and Glc₆ oligosaccharides, the digestion was stopped (16 h) by boiling for 30 min. The solution was cooled to room temperature, filtered on a glass fiber filter, and the filtrate concentrated in vacuo to a volume of 10 ml. The oligosaccharides were separated by size exclusion chromatography on two Bio-Gel P6 (Bio-Rad) columns (2 × 90 cm, 2.6 cm diameter) connected in series and maintained at 60 °C. The products were eluted with a flow rate of 0.5 ml/min with ultrapure water. Fractions (5 ml) were analyzed by HPAEC-PAD (gradient B), pooled; concentrated in vacuo, and finally freeze-dried (typical yield 130 mg of Glc₄ oligosaccharides, 190 mg of Glc₆ oligosaccharides, 98 mg of Glc₁₂ oligosaccharides, 20 mg of Glc₁₆ oligosaccharides and 8 mg of Glc₂₀ oligosaccharides).

**XXXGXXGX and Partially Degalactosylated Glc₁₂-based Xyloglucan Oligosaccharides—**Degalactosylated higher order oligosaccharides were produced exactly as described in the preceding paragraph, except that 4 units of β-galactosidase (*Aspergillus niger*, Megazyme, Eire) were added immediately after boiling and cooling the crude cellulase digestion mixture. The degalactosylation reaction was carefully controlled by HPAEC-PAD analysis (gradient B) because of contaminating isopimemerase activity in the commercial β-galactosidase (27).

**Crystallization, Data Collection, Structure Solution, and Refinement**

Crystals of SeMet-Xgh74A and Xgh74A-D70A in complex with xyloglucan-derived oligosaccharides were grown by the hanging-drop method. SeMet-Xgh74A (10 mg/ml) was crystallized in 5 mM CdCl₂, 12% PEG 4000, 100 mM sodium cacodylate,
TABLE 1  
Kinetics of Xgh74A on different polysaccharides  

| Polysaccharide          | \( k_{cat}^{a} \) \( \text{min}^{-1} \) | \( K_{m} \) \( \text{g/liter} \) | \( k_{cat}/K_{m}^{b} \) |
|------------------------|----------|---------|--------------|
| Xyloglucan             | 47.54    | 0.76    | 62.55        |
| CMC-4M                 | 43.19    | 4.61    | 9.37         |
| Lichenan               | 35.69    | 6.57    | 5.43         |
| Hydroxyethylcellulose  | 13.05    | 5.10    | 2.55         |

\( ^{a} k_{cat} \) are in molecules of product per molecule of enzyme per minute. 
\( ^{b} k_{cat}/K_{m} \) values are in (g/liter) \( ^{1} \text{min}^{-1} \).
(19), we saw no evidence for the production of XXG, XXX, or XXGG by mass spectrometry (Fig. 4). We speculate that the acidic ESI conditions and/or different ion optics employed by Zverlov et al. (19) may have contributed to the formation of fragment ions in the ESI source. Here, our use of MeOH/H$_2$O/NaCl as an ESI solvent minimized oligosaccharide fragmentation and yielded exclusively [M+Na]$^+$ or [M+2Na]$^{2+}$ adducts, depending upon the applied cone voltage (Fig. 4).

To investigate whether Xgh74A cleaves tamarind xyloglucan at a position other than the anomeric carbon of the unbranched glucosyl moiety (Fig. 1), limit digest experiments were performed on xylo- glucan-oligosaccharides based on Glc$_4$-, Glc$_8$-, and Glc$_{12}$ backbones. Cleavage of the $\beta$ (1→4) glycosidic bond between two branched Glc units bearing $\alpha$ (1→6) Xyl residues has been previously reported in the related GH74 oligoxyloglucan reducing end-specific cellobiohydrolase (OXG-RCBH) from Geotrichum sp (24). Incubation of a high concentration of Xgh74A with a 2:1:3:3 mixture of XXXG/XLXG/XXLG/XLLG showed no formation of smaller oligosaccharides by HPAEC-PAD (data not shown) demonstrating that Xgh74A, in contrast to OXG-RCBH, cannot cut at substituted glucosyl moieties, at least on these short substrates. Similarly, incubation of Xgh74A with XXXGXXXG or a variably galactosylated Glc$_8$ XGO mixture, yielded only XXXG or the expected XXXG/XLXG/XXLG/XLLG mixture, respectively, as determined by HPAEC-PAD and/or MS analysis. Likewise, Xgh74A digestion of a partially degalactosylated Glc$_{12}$ XGO mixture yielded predominantly XXXG and minor amounts of other Glc$_8$-based XGOs (Fig. 3). In all cases the action of Xgh74A, as determined by HPAEC-PAD and/or MS analysis, was indistinguishable from that of commercially available T. longibrachiatum endoglucanase (Fig. 3).

Structure of C. thermocellum Xyloglucanase Xgh74A—The structure of the catalytic module of Xgh74A was solved in a P2$_1$ crystal form by molecular replacement using the homologous Geotrichum sp. M128 oligoxyloglucan reducing end-specific cellobiohydrolase (OXG-RCBH, PDB ID: 1SQJ) as a search model. The final Xgh74A model, an apoenzyme incorporating
Xgh74A consists of two seven-bladed β-propeller domains (Fig. 5A), as expected from the sequence homology with OXG-RCBH. The N-terminal domain of Xgh74A comprises residues 63 to 459 while the C-terminal domain involves residues 33–62 and 460–760. Similarly to OXG-RCBH, the Xgh74A N-terminal domain is orientated at angle of ~90 degrees relative to the C-terminal domain and interactions between these domains occur primarily through H-bonding and hydrophobic interactions over a shared contact area of about 7530 Å². The N- and C-terminal domains, which can be superimposed with an RMSD of 3.1 Å, exhibit 19% sequence identity, which most likely reflects an ancient gene duplication event. The Xgh74A N- and C-terminal domains are connected by two loop segments, one located in the N terminus and the other in the middle of the sequence. In the OXG-RCBH structure, the N- and C-terminal domains are linked by three segments; one in the N terminus, the second in the middle and the third in the C terminus of the sequence (this C-terminal segment adds a fifth strand to the second blade of the N-terminal propeller, which is absent in Xgh74A).

The overall topology of Xgh74A is thus very similar to OXG-RCBH with all the secondary structure elements linked through identical connectivity. The Ca traces superimpose with a RMSD of 1.8 Å for 664 equivalent residues, reflecting 39% sequence identity (calculation performed with DALI, Ref. 34). Not surprisingly, greater structural divergence is found in the loops connecting the blades of the β-propeller architecture. The most important of these differences is the different conformation adopted by the loop Thr397—Pro406 in Xgh74A compared with its structural equivalent Asn374—Thr391 in OXG-RCBH, which may contribute to the significantly different substrate specificity of the two enzymes. These structural details are discussed, below, in the light of the ligand complexes of Xgh74A.

Active Site Structure—It is immediately apparent (Fig. 5) that the substrate binding region of Xgh74A lies in an open cleft. This groove is formed at the intersection of the N- and C-terminal domains. The surface of this cleft is formed by the loops connecting the β-propeller blades in both domains. In the apo
Xyloglucanase Structure and Complex

TABLE 2
Data collection and refinement statistics

|                          | Xgh74A           | Xgh74A-D70A-XLLG-XXLG complex |
|--------------------------|------------------|-------------------------------|
| **Data collection**      |                  |                               |
| Space Group              | P2₁              | P4₂,2₂                        |
| Cell dimensions          |                  |                               |
| a, b, c (Å)              | 100.4, 97.9, 199.1 | 141.4, 141.4, 193.4           |
| α, β, γ (°)              | 90.0, 97.6, 90.0  | 90.0, 90.0, 90.0              |
| Wavelength (Å)           | 0.93100          | 0.97565                       |
| Resolution               | 45.79-2.10 (2.21-2.10) | 54.15-1.95 (2.06-1.95)         |
| Rmerge                   | 0.121 (0.515)    | 0.148 (0.533)                 |
| Complementary (°)        | 9.9 (2.5)        | 14.3 (4.6)                    |
| Completeness (%)         | 99.0 (98.3)      | 100.0 (100.0)                 |
| Redundancy               | 3.9 (3.9)        | 11.5 (11.5)                   |
| **Refinement**           |                  |                               |
| Resolution (Å)           | 19.90-2.10       | 54.15-1.95                    |
| Number of reflections    | 208,964          | 135,354                       |
| Rcryst/Rfree             | 0.18/0.21        | 0.18/0.20                     |
| **Number of atoms**      |                  |                               |
| Protein                  | 21792            | 11294                         |
| Ions (Cd²⁺)              | 7                | NA*                           |
| Ligand                   | NA*              | 354                           |
| Water                    | 2138             | 1277                          |
| **B-factors**            |                  |                               |
| Protein                  | 19.7             | 15.0                          |
| Ions (Cd²⁺)              | 25.8             | NA*                           |
| Ligand                   | NA*              | 16.0                          |
| XLLG (minus subites)     | NA*              | 19.0                          |
| XXLG (positive subites)  | NA*              | 19.0                          |
| Water                    | 26.4             | 25.5                          |
| **RMSD**                 |                  |                               |
| Bond lengths (Å)         | 0.012            | 0.010                         |
| Bond angles (°)          | 1.175            | 1.251                         |
| Ramachandran statistics (%) | 97.3/0.07     | 96.5/3.5/0                    |
| PDB code                 | 2CN2             | 2CN3                          |

a Ramachandran statistics indicate the fraction of residues in the most favored, additionally allowed and disallowed regions of the Ramachandran diagram, as defined by the program MOLPROBITY (39).
b NA, not applicable.

FIGURE 5. Three-dimensional structure of C. thermocellum xyloglucanase Xgh74A. A, divergent (wall-eyed) stereo cartoon of the structure, color-ramped from N terminus (blue) to C terminus (red) with the ligands in ball-and-stick representation. B, mono view of the structure showing the N- and C-terminal domains, presumably resulting from an ancestral gene duplication event, colored khaki and pale blue, respectively. This figure was drawn with MOLSCRIPT (40).

Xyloglucanase Structure and Complex

Xgh74A structure some of these loops (Tyr²⁰⁶–Asp²¹⁷, Thr²⁹¹–Asn²⁹⁸, and Asp⁵²⁴–Asp⁵²⁷) are disordered, whereas in the ligand complexed forms they become ordered and participate in substrate binding (described below).

Catalysis by family GH74 enzymes occurs with inversion of anomeric configuration; i.e. the stereochemistry of the product is inverted with respect to the β-linkage of the substrate. A classical interpretation of glycoside hydrolysis with inversion of anomeric configuration implicates two key residues; a catalytic acid, to facilitate leaving group departure by protonation and a catalytic base, to activate the incoming water molecule for nucleophilic attack by deprotonation (glycosidase catalytic mechanisms are reviewed in Ref. 25). In GH74 enzymes, it is believed (22) that two aspartate residues play the role of Brøn-
sted acid and base; in Xgh74A these are believed to be Asp<sup>480</sup> and Asp<sup>70</sup>, respectively. Asp<sup>70</sup> and Asp<sup>480</sup> are located in the middle of the active center cleft, lying on opposite sides, and deep within the cavity with their carboxylate groups separated by 10 Å apart. Site-directed mutagenesis of either of these residues, to alanine, results in an inactive enzyme (which within the sensitivity of the assay suggests at least 1000–5000 times less activity than wild-type enzyme). Asp<sup>70</sup> is located in the middle of the loop connecting the second and third strand of the first propeller blade of the N-terminal domain. The peptide sequence in this region is strictly conserved among the members of the GH74 family. In the apoenzyme, Asp<sup>70</sup> forms a hydrogen bonding interaction with the side chain of Glu<sup>459</sup> and two molecules of water. Asp<sup>480</sup> is located in the C-terminal domain in an equivalent position to Asp<sup>70</sup>. However, in contrast to Asp<sup>70</sup>, Asp<sup>480</sup> does not form H-bonds with protein atoms and points directly into the cleft. It is not immediately apparent what might contribute to an elevated p<sub>K<sub>a</sub></sub> for this catalytic acid in the absence of bound substrate.

**Structure of the Xgh74A-D70A Mutant in Complex with Glc<sub>4</sub>-based Xyloglucosidase**—To probe the structural determinants of xyloglucan recognition we attempted to co-crystallize Xgh74A and two inactive Xgh74A variants (D70A and D480A) with preparations of xyloglucosidase based on Glc<sub>4</sub> backbones. Crystals of the Xgh74A-D70A mutant, complexed with a mixture of Glc<sub>4</sub>-based oligosaccharides, were obtained by co-crystallization and diffracted to 1.95-Å resolution. Crystals were indexed in the P<sub>4</sub>2<sub>1</sub>2<sub>1</sub> space group and contained two molecules in the asymmetric unit. The complex structure is essentially identical to the ordered parts of the apo structure resulting in an RMSD value of 0.3 Å for the Cα atoms. In the ligand complex structure, however, it is also possible to build the previously disordered loops, which all interact directly with the bound oligosaccharide.

The electron density map displays well defined density for seventeen sugar rings, corresponding to a molecule of XLLG and another of XXLG, either side of the catalytic center (Fig. 6). The two molecules sit in an extended conformation at the bottom of the cleft on both sides of the catalytic residue Asp<sup>480</sup>. It is possible that the desolation afforded by ligand binding contributes to the p<sub>K<sub>a</sub></sub> elevation of the catalytic acid. The glucosyl backbones extend about 20 Å in opposite directions from the center point described by the Asp<sup>480</sup> residue. The interaction of the two ligand molecules extends over an area of ~316 Å<sup>2</sup>. Despite co-crystallization with a mixture of oligosaccharides (in relative proportions XXXG 2 : XLXG 1 : XXLG 3 : XLLG 3), the species that we observe bound to the enzyme corresponds to XLLG in the “minus” subsites (subsite nomenclature according to Ref 25) and XXLG in the “positive” leaving group subsites. Thus, four β-1,4 glucosyl moieties of one XLLG molecule are located in the negative binding sites –1 (Glc<sup>−1</sup>), –2 (Glc<sup>−2</sup>), –3 (Glc<sup>−3</sup>), and –4 (Glc<sup>−4</sup>) in an extended confor-
tion. Three $\alpha$-1,6-linked xylose residues branch from the $-2$ (Xyl$^{-2}$), $-3$ (Xyl$^{-3}$), and $-4$ (Xyl$^{-4}$) glucosyl units, with both Xyl$^{-2}$ and Xyl$^{-3}$ also bearing a $\beta$-1,2-linked galactosyl (Gal$^{-2}$) unit, the latter partially disordered. The mean temperature factor for this oligosaccharide is 16 Å$^2$ and its interaction area with the enzyme is 162 Å$^2$. Glc$^{-1}$ is positioned in the middle of the diagonal line connecting the C$_\text{O}$ atoms of the catalytic residues Asp$^{70}$ and Asp$^{480}$ (Ala$^{70}$ in the complex structure). Ala$^{70}$ in the Xgh74A-D70A mutant lies below the plane of the Glc$^{-1}$ ring at a distance of about 5.8Å between Glc$^{-1}$ C1 and Ala$^{70}$ C$_\text{O}$, consistent with the position demanded for the catalytic base in an inverting mechanism. Asp$^{480}$ is located above this plane at a distance of 4.3 Å. Glc$^{-1}$ forms a network of H-bonding interactions where all its oxygen atoms except O4 are involved (Figs. 6 and 7). Glc$^{-1}$ O1, O2, and O3 atoms interact with the nitrogen main chain of Phe$^{51}$, Arg$^{158}$ side chain and Asn$^{154}$ side chain respectively, whereas the putative catalytic acid, Asp$^{480}$, forms H-bonds with Glc$^{-1}$ O5 and O6. All the side chains involved in the recognition of Glc$^{-1}$ appear conserved in the multiple sequence alignment of GH74 family, Fig. 8. Superposition of the Xgh74A and OXG-RCBH structures shows similar environments around the Glc$^{-1}$ O6 position ($-1'$ subsite). In both structures, the loops enclosing this region adopt similar conformations and some of the side chains lining the cavity are conserved, leaving room to accommodate a xylose residue at the $-1'$ position. Accordingly, it is not evident what constitutes the structural basis supporting the fact that endoxyloglucanases prefer to cleave the xyloglucan chain at unbranched glucosyl positions ($-1$) in contrast with the ability of exo-xyloglucanases to process substrates with a xylose ramification at the $-1'$ position (24, 35).

At subsite $-2$, the plane of the Glc$^{-2}$ ring is rotated 180 degrees relative to the plane of Glc$^{-1}$ contacting the enzyme through two H-bonding interactions mediated by Glc$^{-2}$ O2 and O3 with the side chain of Asn$^{249}$. Xyl$^{-2}$ ring is located almost parallel to the plane of Tyr$^{214}$ aromatic ring and its interaction with the protein is mediated by two H-bonds between Xyl$^{-2}$ O4 and the side chain oxygen of Tyr$^{295}$ and side chain NH1 of Arg$^{158}$. Gal$^{-2}$ does not contact the protein directly but forms an H-bond interaction with a water molecule which in turn interacts with the main chain nitrogen atom of Tyr$^{214}$. The Gal$^{-2}$ moiety does not appear to be specifically recognized by the enzyme; instead it is located close to the border of the cleft and exposed to the solvent.

The $-3$ subsite glucoside is “stacked” on the side chain of Trp$^{125}$ and does not form any hydrogen-bonding interactions with the enzyme. The Xyl$^{-3}$ ring in Xgh74A describes an angle of $-90$ degrees with respect to the plane of Glc$^{-3}$ ring pointing directly to the middle of the cleft in the N-terminal to C-terminal direction. This residue forms two H-bonds between its O3 and O4 atoms and the side chain of Asp$^{731}$. An additional partially occupied Gal (Gal$^{-3}$) residue was modeled in the electron density adjacent to Xyl$^{-3}$ O2 atom. This Gal$^{-3}$ residue is located at the entrance of the cleft and exposed to solvent. Gal$^{-3}$ does not interact directly with any protein atom but forms one H-bond between its O4 and the O3 of Glc$^{-2}$. This structural feature is in agreement with the described absence of specificity toward Gal residues on positions $-2$ and $-3$. The
XYLOGLUCANASE STRUCTURE AND COMPLEX

The glucosyl backbone starts to emerge from the binding cleft from site +2. None of the sugar residues at sites +2, +3, and +4 makes direct H-bonding contacts with protein atoms but are found to be involved in complex networks of interactions mediated by water molecules as is represented in Fig. 6B. In the Xgh74A complex described here there is no density, not even at very low levels, indicative of a galactosyl residue attached to the +2' xyloside. Indeed, inspection of the structure would suggest that there are steric blockages for the accommodation of a +2' galactosyl moiety. The interpretation of limit digest patterns, in particular the observation of XLLG, demands that the +2' region must be able to accommodate a galactosyl moiety during catalysis. One possibility is that the binding mode for XLLG is more flexible, at either protein or ligand levels, than that observed here for XXLG.

**Conservation of Xyloglucan Recognition Sites in Members of GH74 Family—GH74 family groups enzymes that are able to hydrolyze xyloglucan oligosaccharides but also are active on non-branched substrates like barley β-glucans (β-1,3/1,4 glucan), carboxymethyl cellulose (CMC), Avicel (microcrystalline cellulose) or galactomannan (β-1,4-mannose). A wide spectrum of activities is observed among members of this family. This spectrum covers enzymes that can, apparently, only process xyloglucan up to enzymes that actually prefer non-branched substrates such as barley β-glucan. For example, *Paenibacillus sp. KM21* (36) and *Thermobifida fusca YZ* (37) xyloglucanases are only active on xyloglucans from various sources but not on non-branched polymers such as barley or xylan, *Geotrichum sp. M128* displays its highest activity on the more branched xyloglucans from tamarind or pea, but it is less efficient on barley xyloglucan that contains fewer xylose decorations and the enzyme shows no activity on non-branched substrates (44).

The specificity of the interaction between xylo-oligosaccharide ligands and Xgh74A interpreted in light of direct H-bonding contacts with the enzyme is reduced to positions −3', −2', −1, −1/−2'.
Family GH74—In both Xgh74A and OXG-RCBH the substrate binding cavities are open grooves well exposed to solvent (Fig. 9). OXG-RCBH is an exoglucanase that releases two glucosyl residues from the reducing end of the xyloglucan polymer, suggesting the presence of at least two negative reducing end subsites and two positive leaving group subsites (22, 24) and demanding the ability to cleave at xylose-substituted glucosyl moieties. In contrast, Xgh74A processes the xyloglucan chain in an endo fashion releasing four glucosyl residue segments (Ref. 19 and this work).

Xgh74A residues Trp<sup>125</sup> and Asp<sup>731</sup> appear to contribute to specificity for the sugars located at subsite −3 where they interact directly with Glc<sup>−3</sup> and Xyl<sup>−3</sup> (Figs. 5 and 6). In the OXG-RCBH structure, the equivalent structural determinants for sugar recognition at the −3 site are absent. The OXG-RCBH equivalent to Trp<sup>125</sup> of Xgh74A is Asp<sup>89</sup>, which is situated in a loop two residues shorter than in Xgh74A and is consequently too distant to interact with the Glc<sup>−3</sup> residue. Similarly, while Xgh74A Asp<sup>731</sup> interacts directly Xyl<sup>−3</sup> O2 and O3 atoms, the

-2, −1, +1, and +1’ (Fig. 8). Apart from the strict conservation of the catalytic residues Asp<sup>70</sup> and Asp<sup>180</sup> and their sequence equivalents in the GH74 family, the residues responsible for the recognition of the glucosyl units at positions −2, −1, and +1 appear highly conserved in the multiple sequence alignment of GH74 members (Fig. 8). Two substitutions are only observed at the glucosyl recognition site −1 and +1 in the multiple sequence alignment. The first one is in the equivalent position to Xgh74A Phe<sup>51</sup> where a Tyr residue appears in Thermobifida fusca and Hypocreajecorina respectively. At the −3 subsite, Xgh74 Asp<sup>731</sup> is found not conserved in any of the family members, the loop in which this residue is located displays variations in length and overall amino acid composition making difficult to assess the presence or not of an equivalent interaction in the absence of structural data on these other family members.

**Exo versus Endo Specificity in**

**Family GH74**

| Xyloglucanase Structure and Complex |
|-------------------------------------|
| **A** | van der Waals’ surface for Xgh74A (yellow) with the ligands in blue showing the open cleft of this endoenzyme. B, van der Waals’ surface for OXG-RCBH (pale green) with the ligands from Xgh74A superimposed in yellow ball-and-stick. The surface of the OXG-RCBH loop Gly<sup>272</sup>–His<sup>285</sup> that closes off the substrate binding surface and is responsible for the exo mode of action is shaded in red. The figures are shown in divergent (‘wall-eyed’) stereo. |

**FIGURE 9. Comparison of endoxyloglucanase Xgh74A and the reducing end-specific GH74 enzyme OXG-RCBH.** A, van der Waals’ surface for Xgh74A (yellow) with the ligands in blue showing the open cleft of this endoenzyme. B, van der Waals’ surface for OXG-RCBH (pale green) with the ligands from Xgh74A superimposed in yellow ball-and-stick. The surface of the OXG-RCBH loop Gly<sup>272</sup>–His<sup>285</sup> that closes off the substrate binding surface and is responsible for the exo mode of action is shaded in red. The figures are shown in divergent (‘wall-eyed’) stereo.
structural equivalent of Xgh74 Asp$^{731}$ in OXG-RCBH is Thr$^{736}$, but its position is again distant from Xyl$^{-3}$ most likely as a result of conformational constraints imposed in the loop by the presence of residues Gly$^{734}$–Pro$^{735}$ in OXG-RCBH. These features likely contribute to the reported differences between the two enzymes with respect to the number of reducing end subsites.

Differences between the two enzymes at the +3 subsite (Fig. 9) are more pronounced. The conformations of the (structurally equivalent) loops Xgh74A Thr$^{397}$–Pro$^{406}$ and OXG-RCBH Asn$^{374}$–Thr$^{391}$ are dramatically different; the latter closing the binding cleft immediately after the subsite +2 and presumably restricting OXG-RCBH to exo-hydrolysis. The cleft in Xgh74A is thus open in both extremes and differs from OXG-RCBH in which the loop Gly$^{375}$–His$^{383}$ blocks one-half of the substrate binding landscape.

The Xgh74A structure reveals both a complex binding architecture in which subsites accommodate seventeen distinct sugar moieties accounting for the pattern of xyloglucan recognition observed by limit hydrolysis of tamarind xyloglucan. Catalysis occurs with inversion of anomeric configuration in a mechanism in which (kinetically essential) aspartates 480 and 70 likely play the role of catalytic acid and base, respectively. Given the growing importance of plant biomass conversion, especially in the context of the demand for clean energy sources, the Xgh74A structure provides the first insights into the recognition and hydrolysis of this crucial component of the plant cell wall.

Acknowledgment—We thank Gustav Sundqvist (KTH Biotechnology) for MS analysis.

REFERENCES

1. Demain, A. L., Newcomb, M., and Wu, J. H. (2005) Microbiol. Mol. Biol. Rev. 69, 124–154
2. Carpita, N., and McCann, M. (2000) in Biochemistry and Molecular Biology of Plants (Buchanan, B., Gruissem, W., and Jones, R. eds) pp. 52–108, John Wiley & Sons, Inc., Somerset, NJ.
3. Vincken, J. P., York, W. S., Beldman, G., and Voragen, A. G. J. (1997) Plant Physiol. 114, 9–13
4. Reiter, W. D. (2002) Curr. Opin. Plant Biol. 5, 536–542
5. Thompson, D. S. (2005) J. Exp. Bot. 56, 2275–2285
6. Cosgrove, D. J. (2005) Nat. Rev. Mol. Cell. Biol. 6, 850–861
7. Yokoyama, R., Rose, J. K. C., and Nishitani, K. (2004) Plant Physiol. 134, 1088–1099
8. Scheible, W. R., and Pauly, M. (2004) Curr. Opin. Plant Biol. 7, 285–295
9. Yamatoaya, K., and Shirakawa, M. (2003) Curr. Trends Polym. Sci. 8, 27–72
10. Miyazaki, S., Suisha, F., Kawasaki, N., Shirakawa, M., Yamatoaya, K., and Attwood, D. (1998) J. Control. Release 56, 75–83
11. Zhou, Q., Baumann, M. J., Pispansen, P. S., Teeri, T. T., and Brumer, H. (2006) Biocatal. Biotransform. 24, 107–120
12. Zhou, Q., Greffe, L., Baumann, M. J., Malmström, E., Teeri, T. T., and Brumer III, H. (2005) Macromolecules 38, 3547–3549
13. Brumer, H., Zhou, Q., Baumann, M. J., Carlsson, K., and Teeri, T. T. (2004) J. Am. Chem. Soc. 126, 5715–5721
14. Milkowski, K., Clark, J. H., and Doi, S. (2004) Green Chem. 6, 189–190
15. Bayer, E. A., Setter, E., and Lamed, R. (1985) J. Bacteriol. 163, 552–559
16. Bayer, E. A., Shimon, L. J., Shoham, Y., and Lamed, R. (1998) J. Struct. Biol. 124, 221–234
17. Carvalho, A. L., Dias, F. M. V., Prates, J. A. M., Nagy, T., Gilbert, H. J., Davies, G. J., Ferreira, L. M. A., Romão, M. J., and Fontes, C. M. G. A. (2003) Proc. Natl. Acad. Sci. (U. S. A.) 100, 13809–13814
18. Zverlov, V. V., Kellermann, J., and Schwarz, W. H. (2005) Proteins 5, 3646–3653
19. Zverlov, V. V., Schantz, N., Schmitt-Kopplin, P., and Schwarz, W. H. (2005) Microbiology 151, 3395–3401
20. Henrisatt, B. (1991) Biochem. J. 280, 309–316
21. Davies, G. J., Glesler, T. M., and Henrisatt, B. (2005) Curr. Op. Struct. Biol. 15, 637–645
22. Yaoi, K., Kondo, H., Noro, N., Suzuki, M., Tsuda, S., and Mitsuishi, Y. (2004) Structure 12, 1209–1217
23. 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. J., and White, A. R. (1993) Physiol. Plant. 89, 1–3
24. Yaoi, K., and Mitsuishi, Y. (2002) J. Biol. Chem. 277, 48276–48281
25. Davies, G. J., Wilson, K. S., and Henrisatt, B. (1997) Biochem. J. 321, 557–559
26. Vincken, J. P., Yeake, A., Beldman, G., and Voragen, A. G. J. (1995) Plant Physiol. 108, 1579–1585
27. Kato, Y., Matsushita, J., Kubodera, T., and Matsuoka, K. (1985) J. Biochem. 97, 801–810
28. Collaborative Computational Project Number 4. (1994) Acta Crystallogr. Sect. D. 50, 760–763
29. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) Acta Crystallogr. Sect. D. 61, 458–464
30. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. Sect. D. 60, 2126–2132
31. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. Sect. D. 53, 240–255
32. Somogyi, M. (1945) J. Biol. Chem. 160, 69–73
33. Nelson, N. (1944) J. Biol. Chem. 153, 375–380
34. Holm, L., and Sander, C. (1993) J. Mol. Biol. 233, 123–138
35. Bauer, S., Vasu, P., Mort, A. J., and Somerville, C. R. (2005) Carbohydr. Res. 340, 2590–2597
36. Yaoi, K., Nakai, T., Kameda, Y., Hiyoshi, A., and Mitsuishi, Y. (2005) Appl. Environ. Microbiol. 71, 7670–7678
37. Irwin, D. C., Cheng, M., Xiang, B., Rose, J. K., and Wilson, D. B. (2003) Eur. J. Biochem. 270, 3083–3091
38. Chhabra, S. R., and Kelly, R. M. (2002) FEBS Lett. 531, 375–380
39. Lovell, C. S., Davis, I. W., Adreindall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Prot. Struct. Funct. Genet. 50, 437–450
40. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
41. Emsley, R. M. (1997) J. Mol. Graphics Model 15, 132–134
42. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nucleic Acids Res. 22, 4673–4680
43. Gouet, P., Courcelle, E., Stuurt, D. I., and Metoz, F. (1999) Bioinformatics 15, 305–308
44. Yaoi, K., and Mitsuishi, Y. (2004) FEBS Lett. 560, 45–50