THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 289, NO. 45, pp. 31624–31637, November 7, 2014

Biochemical Characterization and Crystal Structures of a Fungal Family 3 β-Glucosidase, Cel3A from *Hypocrea jecorina*[*]

Received for publication, June 6, 2014, and in revised form, August 20, 2014 Published, JBC Papers in Press, August 27, 2014, DOI 10.1074/jbc.M114.587766

Saeid Karkehabadi[‡1], Kate E. Helmich[§1], Thijs Kaper[¶], Henrik Hansson[‡], Nils-Egil Mikkelsen[‡], Mikael Gudmundsson[‡], Kathleen Piens[‡], Meredith Fujdala[¶], Goutami Banerjee[‖], John S. Scott-Craig[‖], Jonathan D. Walton[‖], George N. Phillips, Jr.[§2] and Mats Sandgren[‡3]**

*From the [‡]Department of Chemistry and Biotechnology, Swedish University of Agricultural Sciences, P.O. Box 7015, SE-750 07 Uppsala, Sweden, the [§]Department of Energy Great Lakes Bioenergy Research Center and Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, [¶]DuPont Industrial Biosciences, Palo Alto, California 94304, the [‖]Department of Energy Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, Michigan 48824, and the **Department of Biochemistry and Cell Biology and Department of Chemistry, Rice University, Houston, Texas 77251*

Background: β-Glucosidases hydrolyze the β-linkage between two adjacent molecules in oligomers of glucose.
Results: We report the structure and biochemical characterization of Cel3A from *Hypocrea jecorina*.
Conclusion: We determine the structures of Cel3A from protein expressed in two different expression hosts and compare them.
Significance: The structures give new insights into protein glycosylations, stability, and ligand binding in GH3 β-glucosidases.

Cellulase mixtures from *Hypocrea jecorina* are commonly used for the saccharification of cellulose in biotechnical applications. The most abundant β-glucosidase in the mesophilic fungus *Hypocrea jecorina* is *Hj*Cel3A, which hydrolyzes the β-linkage between two adjacent molecules in dimers and short oligomers of glucose. It has been shown that enhanced levels of *Hj*Cel3A in *H. jecorina* cellulase mixtures benefit the conversion of cellulose to glucose. Biochemical characterization of *Hj*Cel3A shows that the enzyme efficiently hydrolyzes (1,4)- as well as (1,2)-, (1,3)-, and (1,6)-β-D-linked disaccharides. For crystallization studies, *Hj*Cel3A was produced in both *H. jecorina* (*Hj*Cel3A) and *Pichia pastoris* (*Pp-Hj*Cel3A). Whereas the thermostabilities of *Hj*Cel3A and *Pp-Hj*Cel3A are the same, *Pp-Hj*Cel3A has a higher degree of *N*-linked glycosylation. Here, we present x-ray structures of *Hj*Cel3A with and without glucose bound in the active site. The structures have a three-domain architecture as observed previously for other glycoside hydrolase family 3 β-glucosidases. Both production hosts resulted in *Hj*Cel3A structures that have *N*-linked glycosylations at Asn^{208} and Asn^{310}. In *H. jecorina*-produced *Hj*Cel3A, a single *N*-acetylglucosamine is present at both sites, whereas in *Pp-Hj*Cel3A, the *P. pastoris*-produced *Hj*Cel3A enzyme, the glycan chains consist of 8 or 4 saccharides. The glycosylations are involved in intermolecular contacts in the structures derived from either host. Due to the different sizes of the glycosylations, the interactions result in different crystal forms for the two protein forms.

Plant cell walls are a rich source of renewable carbon for the production of a wide variety of molecules that could potentially replace petroleum-derived molecules used for biofuels, chemicals, and renewable materials. The cell wall polysaccharides cellulose and hemicellulose, when deconstructed to monomeric sugars, can be fermented to molecules of interest by dedicated microorganisms.

Cellulose is a linear polymer of β-(1,4)-linked glucose residues. It is organized into fibrous microcrystals, which can be highly structured in some regions and less so in others; it therefore contains both crystalline and amorphous regions. The complete decomposition of cellulose to glucose requires a battery of synergistic enzymes. In cellulolytic fungi, these enzymes include 1) endoglucanases (1,4-β-D-glucan-4-glucanohydrolases, E.C. 3.2.1.4), which randomly hydrolyze internal bonds in amorphous regions and release new chain ends; 2) lytic polysaccharide (cellulose) mono-oxygenases that oxidatively break linkages in crystalline regions; 3) processive cellobiohydrolases (1,4-β-D-glucan cellobiohydrolases, E.C. 3.2.1.91), which form cellobiose mainly from the chain ends (1–4); and 4) β-glucosidases, which act on cellobiose and in some cases on cellooligosaccharides to release glucose (4, 5). β-Glucosidases are thought to further enhance cellulose deconstruction by relieving cellobiose inhibition of several cellobiohydrolases and endoglucanases present in the cellulase mixtures (6, 7).

Filamentous fungi are efficient plant biomass degraders. Grown on cellulosic substrates, the ascomycete fungus *Hypocrea jecorina* (teleomorph of *Trichoderma reesei*) produces a set of enzymes that concertedly degrade cellulose to glucose, among which are three β-glucosidases that belong to glycoside

* This work was supported in part by the Faculty for Natural Resources and Agriculture at the Swedish University of Agricultural Sciences through the research program "MicroDrivE." Work in the laboratories of J. D. W. and G. N. P. was supported by the United States Department of Energy Great Lakes Bioenergy Research Center (Department of Energy Office of Science Biological and Environmental Research (BER) Grant DE-FC02-07ER64494).
[1] Both authors contributed equally to this work.
[2] To whom correspondence may be addressed: Dept. of Biochemistry and Cell Biology, Rice University, Houston TX 77251. Tel.: 713-348-6951; E-mail: georgep@rice.edu.
[3] To whom correspondence may be addressed: Dept. of Chemistry and Biotechnology, Swedish University of Agricultural Sciences, P.O. Box 7015, SE-750 07 Uppsala, Sweden. Tel.: 46-18-673179; E-mail: mats.sandgren@slu.se.

hydrolase family 3 (GH3):[4] Cel3A, Cel3B, and Cel3E (8–11). Cel3A makes up ~1% of the total set of enzymes secreted by *H. jecorina*, which is suboptimal for *in vitro* degradation of cellulosic substrates. Therefore, a number of studies that utilize *H. jecorina* whole cellulase mixtures supplement the enzyme mixture with β-glucosidase activity to make degradation more efficient (12–14).

Previously, the *cel3A* gene from *H. jecorina* was cloned, and strains that overexpress *Hj*Cel3A were generated (9). Culture supernatants of these strains were more efficient in the release of glucose from Avicel and phosphoric acid-swollen cellulose (PASC) compared with the culture supernatant from the parent strain. More recently, β-glucosidase-enhanced *H. jecorina* RutC30 strains with additional copies of *Hjcel3A* under the control of the *cbh1* promoter were found to be more efficient in releasing glucose from corncob substrates (15).

Despite the great number of known GH3 sequences in databases, relatively few three-dimensional structures are available from this GH family in the Protein Data Bank. Interestingly, the known structures display a variety of domain organizations from a single domain to four distinct separate domains. The structure of the β-hexosaminidase from *Vibrio cholerae* (16) contains only one domain, whereas the β-D-glucan glucohydrolase Exo1 from *Hordeum vulgare* (barley) (17) has two domains. The β-glucosidases Bgl3B from *Thermotoga neapolitana* (18), AaBGL1 from *Aspergillus aculeatus* (19), and bacterial JMB19063 (20) are three-domain enzymes and the β-glucosidase BglI from *Kluyveromyces marxianus* (21) includes four domains.

Similar to most eukaryotic organisms, fungi and yeast cells attach glycans to secreted proteins. The degree of glycosylation can differ among organisms, and heterologous expression of fungal cellulases can change the native glycosylation patterns, which can, in turn, affect the activities of the enzymes (22).

Here we show the importance of the major GH3 β-glucosidase from *H. jecorina*, *Hj*Cel3A, for the saccharification of a specific pretreated biomass substrate and determine its substrate specificity in detail. We also present the first crystal structures of this enzyme. Two structures solved to 2.1 Å resolution, with and without glucose bound in the active site, were obtained from *H. jecorina*-expressed *Hj*Cel3A, and one structure solved to 2.5 Å with glucose in the active site was obtained with *Pichia pastoris*-expressed *Hj*Cel3A (*Pp-Hj*Cel3A). The structures of the same enzyme produced by different production hosts are compared with each other and with other known GH3 β-glucosidase structures. Special attention is given to the role of *N*-linked glycosylation in crystal formation.

EXPERIMENTAL PROCEDURES

H. jecorina whole cellulase mixture Accellerase® 1000 was a kind gift from Genencor. Corn stover was pretreated with dilute sulfuric acid by the United States Department of Energy National Renewable Energy Laboratory, washed, and adjusted to pH 5. Pretreated corn stover (PCS) contained 56% cellulose, 4% hemicellulose, and 29% lignin. Amorphous PASC was prepared as described (23, 24) and diluted to 0.5% (w/v) in 50 mM sodium acetate, pH 5.0.

Production and Purification of β-Glucosidases from H. jecorina and P. pastoris—The native gene encoding *Hj*Cel3A (UniProt Q12715) was overexpressed in a *H. jecorina* strain lacking four genes coding for cellulases (*cbh1*, *cbh2*, *egl1*, *egl2*). The target genes were cloned into the pTrex3G vector ($amdS^R$, amp^R, P_{cbh1}) (25) and used to transform *H. jecorina*. Transformants were picked from Vogel's minimal medium plates (26) containing acetamide after 7 days of incubation at 37 °C and grown in Vogel's minimal medium with a mixture of glucose and sophorose as carbon sources. The overexpressed protein appeared as a dominant protein in the culture supernatants. *Hj*Cel3A thus produced was about 80% pure as visually determined by SDS-PAGE. Culture filtrate from production of *Hj*Cel3A in *H. jecorina* was diluted 10-fold with 25 mM sodium acetate, pH 4.0 (acetate buffer), and incubated at 37 °C for 30 min. The sample was desalted using a Sephadex G-25M column (GE Healthcare) equilibrated with acetate buffer and concentrated using a centrifugal concentrator with a 10 kDa cut-off (Vivascience, Littleton, MA). Sample was loaded onto a high load 26/60 Superdex 200 column (GE Healthcare) equilibrated with acetate buffer containing 100 mM sodium chloride. Protein was eluted with the same buffer, and protein-containing fractions were checked by SDS-polyacrylamide gel for purity. Fractions with visually pure *Hj*Cel3A were pooled and stored at 4°C. Enzyme purity was also confirmed by isoelectric focusing analysis.

*Hj*Cel3A was also produced in *P. pastoris* X33 (Invitrogen) (*Pp-Hj*Cel3A) from the native *H. jecorina* cDNA sequence. The *Hjcel3A* gene was cloned in the pPICZ vector (Invitrogen) with the native signal peptide according to the manufacturer's instructions. Positive *P. pastoris* transformants were confirmed by colony PCR using gene-specific primers. Single colonies were grown in 50 ml of BMGY medium (100 mM potassium phosphate, pH 6.0, 1.3% yeast nitrogen base, 2% peptone, 1% yeast extract, 0.4 g/ml biotin, 1% glycerol) in 250-ml flasks. The culture was incubated for 16 h at 30 °C with shaking at 250 rpm and used to inoculate 450 ml of BMGY medium, which was then grown with shaking at 250 rpm for 3 days. On the fourth day, the culture was collected in 500-ml sterile centrifuge bottles and centrifuged for 5 min at 1500 × *g*. The pellet was washed with 100 ml of 1 M sorbitol and recentrifuged for 5 min at 1500 × *g*. The pellet was resuspended in 100 ml of BMM medium (100 mM potassium phosphate, pH 6.0, 1.34% yeast nitrogen base, 0.4 g/ml biotin, 0.5% methanol) and transferred to a 2-liter flask. The flask was incubated at 30°C for 4 days with shaking at 250 rpm. To induce protein expression, 500 μl of 0.5% methanol was added every 24 h starting at time 0.

For crystallization of *Pp-Hj*Cel3A, the enzyme was centrifuged for 5 min at 1500 × *g* and concentrated 3-fold using a tangential flow filtration system with a 10 kDa cut-off membrane (Vivaflow, Sartorius, Bohemia, NY), buffer-exchanged five times with 500 ml of 25 mM sodium acetate, pH 5.0, and concentrated another 4-fold. Sterile glycerol was added from an 80% (v/v) stock to a final concentration of 20% (v/v), and aliquots were stored at −80°C. Final stock enzyme concentration was 10.0 mg/ml. Protein samples were checked on an SDS-

[4] The abbreviations used are: GH, glycoside hydrolase; NAG, *N*-acetylglucosamine; PCS, pretreated corn stover; PASC, phosphoric acid-swollen cellulose; RMSD, root mean square deviation; CNPG, 2-chloro-4-nitrophenyl-β-glucopyranoside; CNP, 2-chloro-4-nitrophenol.

polyacrylamide gel for purity. Protein concentrations were determined by the bicinchoninic acid (BCA) method (Pierce), using bovine serum albumin as standard (27). Background protein concentrations in supernatants of *P. pastoris* transformed with empty vector and grown under identical conditions were subtracted.

Saccharification Assays—Enzymes were dosed based on total protein load, and total protein was measured using either a BCA protein assay kit, or by the biuret method (28). Total enzyme loading was 20 mg of protein/g of cellulose. Several ratios of whole cellulase (Accellerase CB100, DuPont) to β-glucosidase were then used (*e.g.* a 50:50 ratio would be 10 mg/g whole cellulase preparation and 10 mg/g β-glucosidase. Substrate (50 μl/well) was loaded into a flat-bottom 96-well microtiter plate using a repeat pipette. PASC was used at 1% (w/v), and PCS was used at 7% (w/v) cellulose. Appropriately diluted enzyme solution (20 μl) was added to each reaction well. In the case of PASC, the enzyme was added to the plate first. The plates were covered with aluminum plate sealers, placed in incubators at 50°C, and incubated with shaking for 2 h. For PCS, the reaction was terminated after 48 h by adding 100 μl of 100 mM glycine, pH 10. After thorough mixing, the reaction mixtures were filtered through a 96-well filter plate (0.45-mm polyethersulfone membrane, Millipore, Billerica, MA). The filtrate was diluted into a plate containing 100 μl of 10 mM glycine, pH 10, and the amount of soluble sugars produced was measured by HPLC (Agilent 1100, Agilent, Santa Clara, CA) equipped with a de-ashing guard column (Bio-Rad, catalog no. 125-0118) and a lead-based carbohydrate column (Bio-Rad Aminex HPX-87P). The mobile phase was water with a 0.6 ml/min flow rate at 80°C.

Cellobiase Activity Assay—The cellobiase activity of *Hj*Cel3A and variants thereof was measured according to the method prescribed by IUPAC (29) adapted for microplates. Equal volumes (50 μl) of serially diluted *Hj*Cel3A or variants were mixed with 15 mM cellobiose in 50 mM sodium acetate, pH 5.0 (final concentration, 7.5 mM cellobiose) in a 96-well microtiter plate. Wells containing only substrate in buffer or only buffer were used as controls. Sealed plates were incubated at 50 °C with 300-rpm shaking in a Thermomixer (Eppendorf, Hamburg, Germany) for 30 min. Next, 100 μl of 100 mM glycine, pH 10.0, was added, and glucose and cellobiose concentrations were measured via HPLC as described above. The enzyme concentrations (mg/ml) were plotted *versus* produced glucose (mg) on a logarithmic scale. The concentration of enzyme required to turn over 0.1 mg of glucose was determined. Cellobiase units (CB) are defined in the reference as follows: CB = (0.0926/enzyme concentration to release 0.1 mg of glucose) units ml^{-1}.

Measurement of Kinetic Properties of the Enzymes on 2-Chloro-4-nitrophenyl-β-glucopyranoside (CNPG) or 2-Chloro-4-nitrophenyl-β-xylopyranoside—Kinetic studies on the substrates CNPG and 2-chloro-4-nitrophenyl-β-xylopyranoside were carried out in 100 mM phosphate buffer, pH 5.7, in a microtiter plate. A volume of 190 μl of the substrate was added to 10 μl of enzyme at an appropriate concentration, and the release of 2-chloro-4-nitrophenol (CNP) was followed spectrophotometrically every 10 min at 37°C by monitoring the absorbance at 405 nm on a Microplate Reader model EL808 (Bio-Tek Instruments, Inc.). The initial velocity ([CNP] (μM/min)) was calculated using a standard concentration curve of the product CNP in the range of 0–200 μM. The kinetic parameters were calculated by fitting the data to the Michaelis-Menten equation with the software KaleidaGraph™ version 3.0. With *Hj*Cel3A and CNPG as substrate, the kinetic parameters were calculated at low concentrations of the substrate (up to 800 μM).

Measurement of Kinetic Properties of the Enzymes on β-Glucobioses—The substrates were incubated with enzyme at 37 °C at the pH optimum. Every 2 min, an aliquot was taken and heated at 100 °C for 5 min to stop the reaction. The released glucose was determined using a coupled glucose oxidase-peroxidase assay (Sigma-Aldrich), using the protocol provided by the manufacturer.

Kinetic Properties of the Enzymes on Oligosaccharides—Degradation of cellotriose and cellotetraose was followed by high performance anion exchange chromatography with pulsed amperometric detection (Dionex ICS-3000, Sunnyvale, CA). The substrate and the enzyme were incubated at 37 °C at the pH optimum. At fixed intervals of time, an aliquot of 20 μl of the sample was withdrawn and added to 40 μl of 0.1 M sodium hydroxide to stop the reaction. The sample was then loaded onto a CarboPac™ analytical column PA-100 (4 × 250 mm; Dionex) and eluted with 100 mM sodium hydroxide and a gradient of sodium acetate from 10 to 170 mM in 100 mM sodium hydroxide for 27 min at a flow rate of 1 ml/min. The quantification of the hydrolysis products was done using external standards.

Measurement of Protein Thermostability—Thermal stability of *Hj*Cel3A and *Pp-Hj*Cel3A was determined by a fluorescent dye-binding thermal shift assay (30). SyproOrange (Molecular Probes) was diluted 1:1000 in Milli-Q water. In multiplate wells, 8 μl of diluted dye was mixed with 25 μl of 100 mg/liter enzyme in 50 mM sodium acetate, pH 5.0. Sealed plates were subjected to a temperature gradient of 25–95 °C at an approximate rate of 1 °C/min in an ABI 7900HT real-time PCR system (Applied Biosystems, Foster City, CA). The mid-peak temperature of the first derivative of the fluorescence signal was taken as the pseudomelting temperature of the protein sample.

Molecular Weight Determination and Peptide Mapping by Mass Spectrometry, Identification of N-Glycosylation Sites, and N-terminal Sequencing—Molecular weights of *Hj*Cel3A and *Pp-Hj*Cel3A (EndoH- and non-EndoH-treated) were determined by MALDI-TOF mass spectrometry (MS). For analysis, 1 μl of sample was dried with 1 μl of sinapinic acid on a MALDI sample plate (Bruker, Fremont, CA). For peptide mapping, 1 mg of *Hj*Cel3A was treated with 20 μg of EndoH for 2 h at 37 °C. The sample was then precipitated with trichloroacetic acid, reduced, and alkylated and finally digested with various tryptic enzymes for 2 h at 25°C. Two *Hj*Cel3A samples (EndoH- and non-EndoH-treated) were digested under three different conditions. The conditions were as follows: trypsin (30 min), trypsin (30 min)/AspN (90 min), and AspN (90 min). All samples were run on the LCQ Deca ion trap mass spectrometer (Thermo). N-terminal sequencing was performed by Alphalyse (Palo Alto, CA).

Crystallization, Data Collection, Structure Determination, and Refinement—Purified *Hj*Cel3A was concentrated to 3.9 mg/ml in a buffer containing 25 mM sodium acetate, pH 4.0, and 100 mM sodium chloride. *Hj*Cel3A crystals were obtained using the hanging drop vapor diffusion method at 20 °C. The drops were prepared by mixing equal amounts of protein sample and crystallization solution consisting of 0.1 M sodium formate, pH 7.0, and 10–20% PEG 3350. To produce *Hj*Cel3A-glucose complex crystals, *Hj*Cel3A crystals were soaked in crystallization solution containing 50 mM glucose for a period of 10 min before they were frozen. Prior to data collection, crystals were frozen in liquid nitrogen using the crystallization solution with 20% glycerol added as a cryoprotectant. Glucose was also added to the cryoprotectant to a final concentration of 50 mM for the *Hj*Cel3A-glucose crystals. Data for *Hj*Cel3A and *Hj*Cel3A-glucose were collected on beamline I 911-5 at MAX-lab (Lund, Sweden) and at European Synchrotron Radiation Facility beamline BM-14 (Grenoble, France), respectively, from single crystals at 100 K. The x-ray diffraction data were processed using the x-ray data integration program Mosflm (31) and scaled using the scaling program Scala (32) in the CCP4i program package (33, 34). In the case of the *Hj*Cel3A-glucose complex, the data were processed and scaled with the XDS package (35). Details of data collection and processing are presented in Table 1. The wild type *Hj*Cel3A and *Hj*Cel3A-glucose complex crystals were found to belong to the orthorhombic space group $P2_12_12_1$, with approximate unit cell parameters as follows: a = 55.1 Å, b = 82.4 Å, and c = 136.7 Å.

Initial screening for crystallization condition for *Pp-Hj*Cel3A was performed using the high throughput screens UW192 (local), IndexHT, SaltRX, and PegHT (Hampton Research), utilizing a Mosquito® dispenser (TTP Labtech, Ltd., Melbourn, UK) by the vapor diffusion sitting drop method. Crystal growth was monitored by Bruker AXS Crystal Farms at 20 and 4 °C. *Pp-Hj*Cel3A crystals were grown by mixing 1 μl of protein sample solution of 10 mg/ml *Pp-Hj*Cel3A in 25 mM sodium acetate, pH 5.0, with 1 μl of reservoir solution, 4% 2-propanol, 0.1 M BTP, pH 9.0, and 20% methoxy poly(ethylene glycol) (Hampton Research). Crystals were cryoprotected with reservoir solution containing 25% methoxy poly(ethylene glycol) and 10% ethylene glycol. X-ray diffraction data were collected at the 23ID-B beamline (GM/CA@APS) with x-ray wavelength 0.9793 at the Advanced Photon Source at Argonne National Laboratory.

Data sets were indexed and scaled using HKL2000 (36). The *Pp-Hj*Cel3A crystals belong to the space group C2, with unit cell parameters of a = 130.4 Å, b = 107.9 Å, and c = 125.9 Å and with a β angle of 115.6°.

Structure Determination and Refinement—The crystal structures of *Hj*Cel3A and *Hj*Cel3A in complex with glucose were solved by molecular replacement using the program PHASER (37). In the case of *Hj*Cel3A, the search model was a polyalanine model that was generated using two different structures: (i) *H. vulgare* Exo1 (*Hv*Exo1; Protein Data Bank code 1X39 (38)) and (ii) *T. neapolitana* Bgl3B (*Tn*Bgl3B; Protein Data Bank code 2X41 (18)). The three structures were superimposed by the program Coot (39–41), and the elements of the structures that were not superimposable were deleted. The resulting structure was subsequently converted into a polyalanine model. Initial $2F_o - F_c$ and $F_o - F_c$ electron density maps calculated after one round of rigid body refinement of individual molecules using data between 20 and 3 Å showed continuous density for main chain atoms of the protein. The structure of *Hj*Cel3A was subsequently used as the search model for solving the structure of the *Hj*Cel3A-glucose complex. Refinement of the structures was performed using REFMAC5 (42). For cross-validation, 5% of the data were excluded from the refinement for R_{free} calculations (43). The two structures were further refined using a maximum likelihood target function refinement. Solvent molecules were added using ARP/wARP (44). Throughout the refinement, $2mF_o - DF_c$ and $mF_o - DF_c$ σ-A-weighted maps (45) were inspected, and the models manually adjusted in Coot.

Phasing of the *Pp-Hj*Cel3A data were done by molecular replacement using PHENIX-AutoMR with *Tn*Bgl3B (Protein Data Bank code 2X40 (18)) as a search model, and PHENIX-AutoBuild was used for initial model building (46, 47). Structure refinement of *Pp-Hj*Cel3A was performed using PHENIX-Refine (46, 47), and the structure quality was assessed using Procheck (48) and Molprobity (49, 50). Table 1 shows statistics of data collection, processing, and refinement of all three structure models. All figures were created in PyMOL (51).

RESULTS AND DISCUSSION

Effect of HjCel3A on Cellulose Degradation by Cellulase Mixtures—Cellobiose inhibits enzymatic cellulose degradation by inhibiting the activity of cellulases, in particular Cel7A (52, 53). β-Glucosidase hydrolyzes cellobiose to monomeric glucose and, as such, synergistically enhances the degradation of cellulose by reducing cellobiose inhibition. The whole cellulase mixture produced by natural *H. jecorina* isolates contains about 1% *Hj*Cel3A when induced using sophorose (54). When a portion of the whole cellulase mixture was replaced with purified *Hj*Cel3A on an equal protein basis, the conversion of both PASC and PCS increased (Fig. 1). Optimal conversion was observed for mixtures that had enhanced *Hj*Cel3A levels (55). Similar results were obtained when Avicel or acid-pretreated bagasse were used as substrates (data not shown). These results agree with reports that overexpression of the *Hj*Cel3A-encoding gene in *H. jecorina* improves cellulose and lignocellulosic substrate conversion (9, 15). The importance of *Hj*Cel3A for the conversion of cellulose prompted us to study the enzyme in more detail.

HjCel3A Has a Broad Substrate Specificity—It has been shown that *Hj*Cel3A is able to hydrolyze cellooligosaccharides of DP2 to DP6 as well as various β-linked glucobioses (56). In this study, the kinetic parameters of *Hj*Cel3A have been determined in more detail using the chromogenic substrate CNPG, glucosyl disaccharides, and cellooligosaccharides (cellotriose and cellotetraose). The values are summarized in Table 2. The results confirm that *Hj*Cel3A has a broad specificity and is capable of hydrolyzing (1→2)-, (1→3)-, (1→4)-, and (1→6)-β-D-linked disaccharides displaying the highest specificity for (1→3)-β-D-linked laminaribiose. Similar studies have been carried out for the *Hv*Exo1 and *Tn*Bgl3B β-glucosidases, and these enzymes were also found to have the highest specificity for hydrolysis of laminaribiose (57, 58). In the structure of *Hv*Exo1 in complex with thio-linked laminaribiose (58), it was shown

TABLE 1
X-ray data collection, processing, and structure refinement statistics

Structure	*Hj*Cel3A	*Hj*Cel3A-Gluc	*Pp-Hj*Cel3A
Data collection and processing			
PDB code	3ZZ1	3ZYZ	4I8D
Beamline[a]	MAX: I911–5	ESRF: BM14	APS: 23-ID-D
Wavelength (Å)	0.90817	0.95373	0.9793
No. of images	175	120	100
Oscillation range (degrees)	0.6	1.0	2.0
Space group	$P2_12_12_1$	$P2_12_12_1$	$C121$
Cell dimensions: *a*, *b*, *c* (Å)	55.1, 82.4, 136.7	55.1 82.9 136.8	130.4, 107.9, 125.9
Cell angles α, β, χ (degrees)	90, 90, 90	90, 90, 90	90, 115.6, 90
Resolution range (Å)	29.7–2.1	29.7–2.1	50–2.50
Resolution range outer shell	2.21–2.10	2.21–2.10	2.55–2.50
No. of observed reflections	152209	217624	518770
No. of unique reflections	36726	37117	54914
Average multiplicity	4.1 (3.9)	5.9 (5.4)	4.2 (4.2)
Completeness (%)[b]	99.0 (95.0)	99.8 (99.5)	99.8 (99.9)
R_{merge} (%)[c]	14.0 (38.1)	15.0 (49.5)	13.1 (73.5)
$I/\sigma(I)$	8.1 (3.1)	9.8 (3.2)	7.2 (1.2)
Refinement			
Resolution used in refinement (Å)	30.0–2.10	30.0–2.10	44.9–2.50
No. of reflections	34,896	35,114	49,246
R_{work}(%)	17.4	17.2	19.7
R_{free} (%)	22.3	22.2	26.4
No. of residues in protein	713	713	711
No. of residues with alternative conformation	9	12	5
No. of water molecules	690	611	487
Average atomic *B*-factor (Å^2)			
Overall	15.1	14.4	48.4
Protein	14.1	13.6	47.9
N-glycosylations		26.4	78.8
Average real space correlation coefficient			
Bound glucose		0.96	0.76
N-Glycosylation chains		208 = 0.81, 310 = 0.95	A208 = 0.83, A310 = 0.78, B208 = 0.87, B310 = 0.75
RMSD for bond lengths (Å)[d]	0.007	0.009	0.006
RMSD for bond angles (degrees)[d]	1.024	1.190	1.122
Ramachandran outliers (%)			
Favored	94.16	95.41	94.48
Allowed	5.13	3.87	4.96
Outlier	0.71	0.72	0.57

[a] Synchrotron beamlines: MAX, Max-Lab (Lund, Sweden); ESRF, European Synchrotron Radiation Facility (Grenoble, France); APS; Advanced Photon Source (Argonne National Laboratory).
[b] Values in parentheses are those for the highest resolution shell.
[c] $R_{merge} = \Sigma_{h,k,l}\Sigma_i|I - \langle I\rangle|\Sigma_{h,k,l}\Sigma_i|I|$.
[d] From Engh and Huber (68).
[e] Calculated using MOLEMAN2 (69).
[f] Calculated using a strict boundary Ramachandran definition given by Kleywegt and Jones (70).

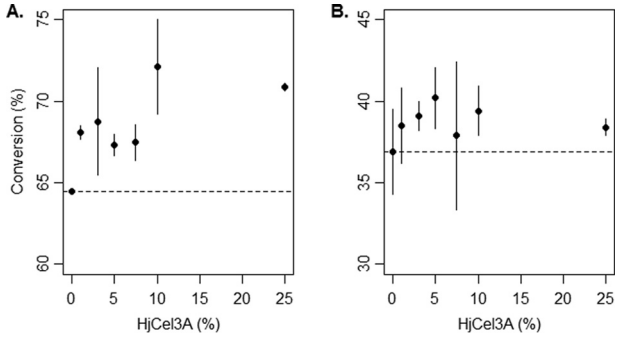


FIGURE 1. **The effect of *Hj*Cel3A produced in *H. jecorina* on saccharification of PASC (*A*) and PCS (*B*) by whole cellulase.** The total enzyme loading was kept constant at 20 mg of enzyme/g of cellulose (*solid symbols*). *x axis*, weight percentage of *Hj*Cel3A; *horizontal lines*, conversion of whole cellulase without added *Hj*Cel3A; *vertical lines*, percentage of *Hj*Cel3A at optimal conversion; *error bars*, S.D. of quadruplicate assays.

TABLE 2
***Hj*Cel3A enzyme kinetics**

Substrate	K_m	k_{cat}	k_{cat}/K_m
	mM	s^{-1}	$M^{-1}\,s^{-1}$
Cellobiose	0.35 ± 0.04	16.0 ± 0.48	0.5×10^5
Cellotriose	0.036 ± 0.006	31 ± 0.80	8.5×10^5
Cellotetraose	0.036 ± 0.006	24 ± 0.11	8.0×10^5
Gentiobiose	0.53 ± 0.08	8.0 ± 0.40	0.2×10^5
Laminaribiose	0.25 ± 0.03	28.0 ± 0.84	1.1×10^5
Sophorose	0.45 ± 0.03	23.0 ± 0.46	0.5×10^5
CNPG	0.087 ± 0.01	28.0 ± 1.12	3.2×10^5

that the glucopyranose residue in the +1 subsite bound with the β (apolar) face toward Trp^{286}, whereas in the structure complex with thio-linked cellobiose, the glucopyranose residue bound with the less hydrophobic α face toward Trp^{286}. In *Hj*Cel3A, the corresponding tryptophan residue, Trp^{237}, has swung inward to the −1 subsite, whereas Trp^{37}, with the side chain oriented 90° in relation to Trp^{237}, together with Phe^{260}, forms a hydrophobic patch at the approximate location of Trp286 of *Hv*Exo1. On the other side of the +1 subsite, the phenyl ring of Tyr^{443} is likely to play the same role as the benzene ring of the Trp^{434} in *Hv*Exo1. The hydrophobic patch may better complement the β face of a predicted bound laminaribiose than the assumed α face that would be the case for cellobiose and sophorose. This may contribute to the preference of *Hj*Cel3A for (1→3)- over (1→4)- and (1→2)-β-D-linked disaccharides seen in our experiments. The C6 hydroxyl group of a glucopyranosyl residue also fits slightly better with a β-(1,3)-linked glycan bound than with a β-(1,4)-linked glycan. Our data also show that *Hj*Cel3A prefers hydrolysis of slightly longer

oligosaccharides (*i.e.* cellotriose and cellotetraose) to that of disaccharides. For longer substrates, subsites further away from the active site may be important for specificity (59, 60). The dramatic increase in specificity for the hydrolysis of cellotriose *versus* cellobiose suggests that the +2 subsite should be important for activity on β-(1,4)-linked oligosaccharides. A putative +2 subsite of *Hj*Cel3A has only two residues, Asp^{370} and Phe^{260}, that are in position to interact with a glucopyranose. None of these are conserved among GH3 enzymes, although they commonly occur in sequences of fungal β-glucosidases. Given the importance of the +2 subsite in *Hj*Cel3A, it is surprising that the assumed site cannot show more potential interactions with a glycan residue.

Pp-HjCel3A Is More Glycosylated Than HjCel3A—For crystallization, *Hj*Cel3A was produced and purified from both *H. jecorina* (*Hj*Cel3A) and *P. pastoris* (*Pp-Hj*Cel3A). In SDS-PAGE analysis, the *Pp-Hj*Cel3A sample ran at a significantly higher apparent molecular mass than *Hj*Cel3A. Previous studies on fungal GH3 β-glucosidases that were heterologously produced in *P. pastoris* have found no significant differences in activity and stability between homologously and heterologously produced forms of the enzymes (61, 62). Limited characterization of the activity of the two different forms of the enzyme on short cellooligosaccharides (cellobiose to cellotetraose) found that this applied to *Hj*Cel3A as well (data not shown).

The effect of *N*-glycosylation on stability was evaluated in more detail. After overnight incubation with EndoH, the migration pattern of *Hj*Cel3A on a SDS-polyacrylamide gel had not changed, whereas *Pp-Hj*Cel3A ran at a lower apparent molecular mass equal to that of *Hj*Cel3A (data not shown). The results were confirmed by molecular weight determination by MALDI-TOF/MS. EndoH treatment reduced the molecular mass of *Pp-Hj*Cel3A by about 7 kDa, corresponding to about 40 saccharide units. This difference in glycosylation did not affect the stability of the proteins; the apparent melting temperatures for the two samples were 74.0 ± 0.2 °C and not significantly different from each other. Thus, the main difference between *Hj*Cel3A and *Pp-Hj*Cel3A is the more extensive *N*-linked glycosylation of the latter, which does not appear to change the properties of the enzyme and agrees with previous analyses of fungal GH3 enzymes produced in *P. pastoris* (61, 62).

*Hj*Cel3A was analyzed for glycosylation sites by peptide analysis using MS as described under "Experimental Procedures." For six of the seven *N*-glycosylation sequons that are present in *Hj*Cel3A, the glycosylated form was identified (positions 45, 208, 310, 417, 566, and 613). However, for three of those residues (Asn^{417}, Asn^{566}, and Asn^{613}), the non-glycosylated form was also detected in the MS analysis, which indicated the presence of various glycoforms. Endo-β-*N*-acetylglucosaminidase EndoH hydrolyzes the bond between two *N*-acetylglucosamine (NAG) subunits directly proximal to the asparagine residue, generating a truncated sugar molecule with one *N*-acetylglucosamine residue remaining on the asparagine. The additional mass of *Hj*Cel3A after EndoH treatment (668 Da) indicated that, on average, three sites are glycosylated in *Hj*Cel3A. In contrast, the additional mass of *Pp-Hj*Cel3A (1573 Da) corresponds to all seven sites being glycosylated. No *O*-linked glycosylated peptides were detected. To obtain homogeneous *Hj*Cel3A samples for crystallization, proteins with various glycoforms were separated by gel filtration.

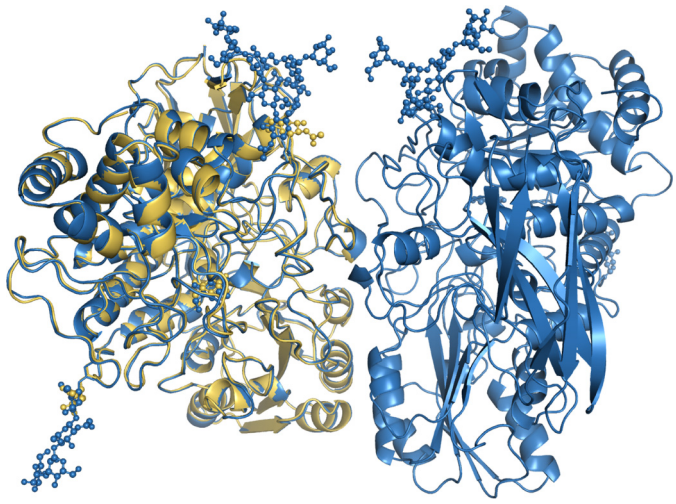


FIGURE 2. **Superposition of the *Hj*Cel3A and the *Pp-Hj*Cel3A structures.** The *schematic representation* of the *Hj*Cel3A crystal structure, with a single protein molecule in the asymmetric unit, is *colored yellow*, whereas the *schematic representation* of the *Pp-Hj*Cel3A crystal structure is *colored blue* for both of the protein molecules of the asymmetric unit. The glycosylations of *Hj*Cel3A and *Pp-Hj*Cel3A are depicted using a *ball-and-stick representation*.

HjCel3A Crystal Structures—*H. jecorina*-expressed *Hj*Cel3A crystallized with one molecule in the asymmetric unit in space group $P2_12_12_1$ for both the apo and glucose-complexed forms and the *P. pastoris*-expressed *Hj*Cel3A (*Pp-Hj*Cel3A) crystallized with two molecules in the asymmetric unit in space group *C*2 (Fig. 2). The structures were solved to 2.1 Å (*Hj*Cel3A, *Hj*Cel3A + glucose) and 2.5 Å (*Pp-Hj*Cel3A + glucose). The crystallographic *R*-factors for the final structure models of the *Hj*Cel3A, *Hj*Cel3A-glucose complex, and *Pp-Hj*Cel3A are 17.5, 18.3, and 20.1%, respectively, whereas the *R*-free values are 22.2, 22.8, and 27.0%, respectively. Other refinement statistics are provided in Table 1.

The overall fold of *Hj*Cel3A strongly resembles that of *Tn*Bgl3B (18) and is composed of three distinct domains (Fig. 3). Superposition of the two structures gives an RMSD of 1.63 Å for 713 equivalent Cα positions, using the SSM algorithm (41).

Domain one encompasses residues 7–300. This domain is joined to domain two with a 16-residue linker (residues 301–316). Domain two, a five-stranded α/β sandwich, comprises residues 317–522 and is followed by a third domain, domain three, which is composed of residues 580–714 and has an immunoglobulin type topology. The folds represented by domains one and two together are present in many GH3 β-glucosidases, and the fold was first described for *Hv*Exo1 (17). Whereas domain one of Exo1 has a canonical TIM barrel fold, with an alternating repeat of eight α-helices and eight parallel β-strands in an α/β barrel, domain one of *Hj*Cel3A lacks three of the parallel β-strands and the two intervening α-helices. Similarly to what was reported for *Tn*Bgl3B, domain one has instead three short antiparallel β-strands, which, together with five parallel β-strands and six α-helices, form an incomplete or collapsed α/β barrel.

The structure of domain three of *Hj*Cel3A is almost identical to that of *Tn*Bgl3B. This is evidenced by the low RMSD value

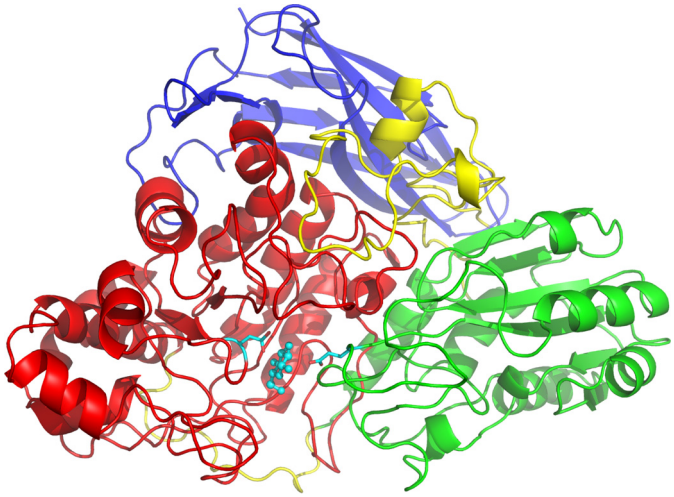


FIGURE 3. ***Schematic representation* of overall structure of *Hj*Cel3A.** The three domains are *colored red* (domain one), *green* (domain two), and *blue* (domain three). The two domain linker regions are shown in *yellow*. The glucose bound in the active site (depicted using a *ball-and-stick representation*) and the two catalytic residues (depicted in *sticks* only) are *colored cyan*.

(1.04 Å) after superposition of the two domains over 113 equivalent Cα positions. The major structural difference between the two domains is observed in the region where the β-strands Lys^{581}–Thr^{592} and Val^{614}–Ser^{624} of *Hj*Cel3A are connected. The two corresponding β-strands in *Tn*Bgl3B are connected with a short loop, whereas a larger structured insertion, Ala^{593}–Asn^{613}, is present at this position in *Hj*Cel3A.

Mass spectrometry and N-terminal sequencing results both showed that the N terminus of *Hj*Cel3A starts with VVPP, which is consistent with the sequence observed in the crystal structure of *Hj*Cel3A. These observations agree with the signal peptide and mature protein start listed for *Hj*Cel3A in the UniProt database (63). In the *Pp-Hj*Cel3A structure, no electron density is observed for the first two valine residues and one proline residue. Except for the number of modeled glycans in the glycosylations, the structure models of *Hj*Cel3A and *Pp-Hj*Cel3A are nearly identical with only minor differences, and the active sites superpose well (Fig. 4*A*). Superposition of the two structures gives an RMSD of 0.382 for 711 equivalent Cα positions, using the SSM algorithm (41). The ligand conformations of the glucose complex forms of *Hj*Cel3A and *Pp-Hj*Cel3A are nearly if not completely identical as well. Fig. 4*B* shows the electron density for the glucose in the −1 subsite of the active site of the *Hj*Cel3A-glucose complex. The final σ-A-weighted $2mF_o - DF_c$ electron density map was continuous for all main chain atoms of the protein. There are three disulfide bonds in the protein that are formed between Cys^{42} and Cys^{58}, Cys^{202} and Cys^{213}, and Cys^{368} and Cys^{373}.

Active Site Geometry and Substrate Specificity—The active site of *Hj*Cel3A is situated, as in other β-glucosidases, at the interface of domain one and domain two at the surface of the molecule. The nucleophile and acid/base active site residues of *Hj*Cel3A (Asp^{236} and Glu^{441}) have their side chains pointing toward the active site and are present at almost identical positions as the catalytic residues of *Hv*ExoI (Asp^{285} and Glu^{491}) and *Tn*Bgl3B (Asp^{242} and Glu^{458}) (Fig. 4*C*).

A comparison of the structures of *Hj*Cel3A and *Tn*Bgl3B shows that the variations between the loops define a shallower active site pocket in the structure of *Hj*Cel3A. One of the reasons for the shallower active site pocket in *Hj*Cel3A is that two loops that connect α-helix Asn^{24}–Ser^{33} and β-strand Gly^{44}–Thr^{46} and β-strand Leu^{57}–Gln^{60} and α-helix Pro^{77}–Thr^{84} are both shorter in the structure of *Hj*Cel3A than the corresponding loops of the *Tn*Bgl3B structure.

Another region that causes the active site pocket of *Hj*Cel3A to be shallower than that of the *Tn*Bgl3B is the small loop that connects β-strand Gln^{409}–Ser^{413} and the α-helix Asn^{417}–Arg^{426}. The corresponding loop of *Tn*Bgl3B is situated much closer to the active site, because both the loop (Glu^{411}–Leu^{433}, *Tn*Bgl3B numbering) as well as the secondary elements that are connected by this loop (α-helices Asp^{393}–Thr^{410} and Ser^{434}–Asn^{445}) are longer than those of *Hj*Cel3A. This loop is flanked over the active site loop in which the acid/base amino acid residue, Glu^{458} in *Tn*Bgl3B, resides and causes the active site pocket to become deeper.

Interestingly, the active site loop of *Hj*Cel3A (residues 438–460) in which the acid/base amino acid Glu^{441} is situated is 8 residues longer (Tyr^{443}–Ala^{450}) in *Hj*Cel3A than the corresponding loops in *Hv*ExoI and *Tn*Bgl3B. There are 2 glycine residues (positions 442 and 451) at each side of the extra section of this loop. Generally, promotion of structural flexibility is a well known feature of glycine residues (64, 65). The two glycine residues at these positions could therefore potentially serve as two hinges for opening and closing of this loop and could potentially play a role in the specificity and/or activity of the enzyme for the substrate. The presence of a longer loop that hovers on the top of this active site loop, as observed in the structure of *Tn*Bgl3B, would not support the flexibility of this extra section of the active site loop of *Hj*Cel3A. In the structure of *Hj*Cel3A, this loop is bent away from the active site, and as a result, the side chain of Tyr^{443} is brought to a position that has become an integral part of the +1 subsite.

The structures of apo-*Hj*Cel3A and *Hj*Cel3A-glucose are essentially identical except for the presence of a glucose molecule bound at the −1 subsite in the complex structure. Clear density is observed for this glucose accommodated at the −1 subsite. The −1 subsite seems to have a highly conserved composition among the structures of *Hj*Cel3A, *Tn*Bgl3B, and *Hv*ExoI (17, 18, 38, 66) with a tight network of hydrogen bonding involving residues Asp^{62}, Arg^{125}, Lys^{158}, His^{159}, Tyr^{204}, Asp^{236}, and Glu^{441} of *Hj*Cel3A.

Superposition of the structure of *Hj*Cel3A with those of *Tn*Bgl3B and *Hv*ExoI show that most of the loops forming the active site of *Tn*Bgl3B are present in *Hj*Cel3A and also that the +1 subsite is situated in a similar position to the corresponding site in *Tn*Bgl3B and *Hv*ExoI. In the structure of *Hv*ExoI, two tryptophan side chains define two sides of the +1 subsite. The structure of *Hj*Cel3A shows that the C-α of Trp^{237} is present at an identical position, but the side chain is swung away relative to the side chain of the corresponding tryptophan, Trp^{286}, in *Hv*ExoI (Fig. 4*C*). On the other side of the *Hj*Cel3A +1 subsite, the side chain of Tyr^{443} is turned away from the surface, pointing toward the core of the molecule. The side chain of this tyrosine replaces the side chain of Trp^{434} of *Hv*ExoI. In the

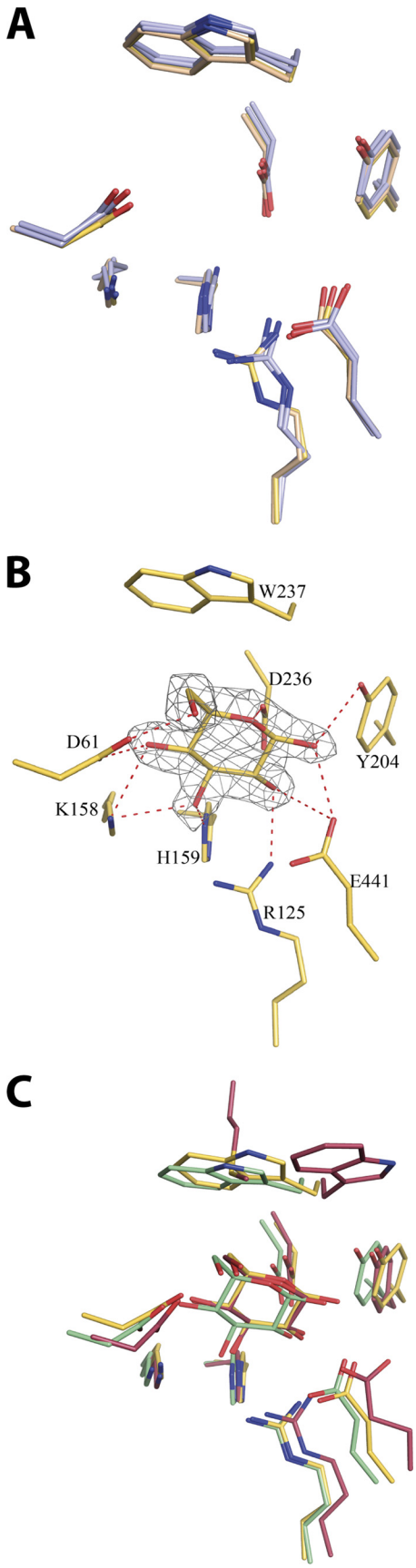


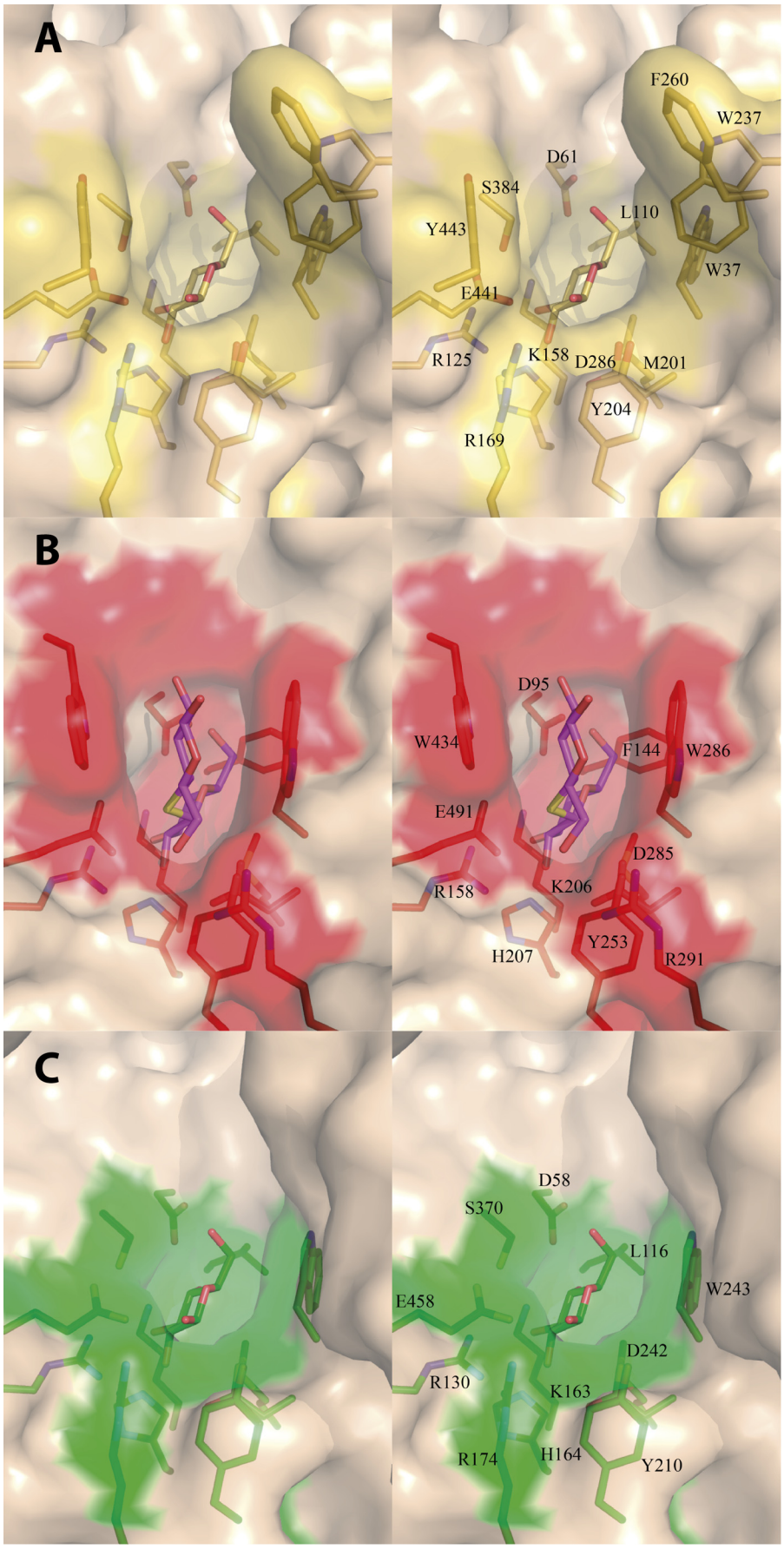
FIGURE 4. *A*, overlay of −1 subsite of the four *Hj*-Cel3A structure models (not showing the glucose): *Hj*Cel3A apo (*light yellow*), *Hj*Cel3A + glucose (*gold*), and both of the *Pp-Hj*Cel3A protein molecules of the asymmetric unit (*blue*). The depicted side chains are the same as in *B*. *B*, the −1 subsite of *Hj*Cel3A with important residues *numbered* and depicted using *sticks*. The bound glucose is shown with electron density contoured at σ level 1. *C*, a comparison of the −1 subsites of *Hj*Cel3A (*gold*), *Tn*Bgl3B (*pale green*), and *Hv*Exo1 (*pale red*). The *Hj*Cel3A −1 subsite is represented by the same residues as shown in *B*. For *Tn*Bgl3B and *Hv*Exo1, the corresponding residues are shown.

structure of *Hv*ExoI, another tyrosine (Tyr^{253}), has a hydrogen bond to the O_2-hydroxyl group of the glucopyranosyl bound in the +1 subsite of the enzyme (Protein Data Bank code 1J8V). The corresponding residue in *Hj*Cel3A (Tyr^{204}) should have the same role in the +1 subsite of *Hj*Cel3A.

Comparison of the structures of *Hj*Cel3A and *Hv*ExoI hints that Tyr^{443} and Trp^{237} may be important for the architecture of +1 subsite of *Hj*Cel3A (Fig. 5, *A–C*). The loop containing Tyr^{443} is located on the surface of the molecule and interacts with the rest of the molecule via a single hydrogen bond between Tyr^{443} and Asp^{370}. This suggests that this loop may be flexible and capable of changing conformation for accommodation of different substrates. The biochemical analysis of *Hj*Cel3A showed that the active site can indeed bind a range of β-linked glucobioses. A big portion of this loop is involved in crystal contact in the *Hj*Cel3A structure, and as a result of this, the temperature factors for this loop are inconclusive as a predictor of its flexibility. Trp^{237} is situated on the other side of the +1 subsite relative to Tyr^{443} and is located close to Trp^{37}. The side chain of Trp^{37} is stacked against the side chain of Phe^{260}. These three hydrophobic residues at the entrance of the substrate tunnel constitute a long hydrophobic patch that could accommodate a longer substrate. This would agree well with the determined specificity of *Hj*Cel3A for cellotriose and cellotetraose.

HjCel3A N-Glycosylation Sites—The electron density maps for the structures of *Hj*Cel3A, *Hj*Cel3A-glucose complex, and *Pp-Hj*Cel3A support *N*-glycosylations at the same two sites, namely asparagines 208 and 310 (Fig. 6, *A–D*). There is no density for *N*-glycosylation at Asn^{45} in either of the maps. In *Hj*Cel3A, which crystallized in space group $P2_12_12_1$ with one molecule in the asymmetric unit, there is no density for glycosylations at asparagines 417, 566, and 583. Furthermore, the presence of glycans at any of these locations would most likely interfere with the crystal packing in $P2_1$. The corresponding asparagines in *Pp-Hj*Cel3A are not in regions of crystal packing, but there is still no density for any attached glycans at Asn^{417} and Asn^{566}. For Asn^{583}, there is weak extra density that could indicate a glycosylation but at a very low frequency.

The two monomers in the asymmetric unit of *Pp-Hj*Cel3A appear to have an orientation with an axis of pseudosymmetry between the Asn^{208} *N*-glycosylations, for which continuous electron density supports eight glycans that have been modeled as two NAGs and six mannose residues (Fig. 6, *C* and *E*). The interactions between the two monomers are mainly of the protein-protein type, and the glycosylations on Asn^{208} in the two monomers are not observed to interact directly. The conformation of the glycan polymer might instead be stabilized by van der Waals interactions with a symmetry-related A molecule (Fig. 7*A*). At the corresponding Asn^{208} sequon in *Hj*Cel3A, a single NAG is present and seems to be important for the crystal


A
F260
W237
D61
S384
L110
Y443
W37
E441
R125
K158
D286
M201
Y204
R169
B
D95
W434
F144
W286
E491
D285
R158
K206
Y253
H207
R291
C
D58
S370
L116
W243
E458
D242
R130
K163
R174
H164
Y210

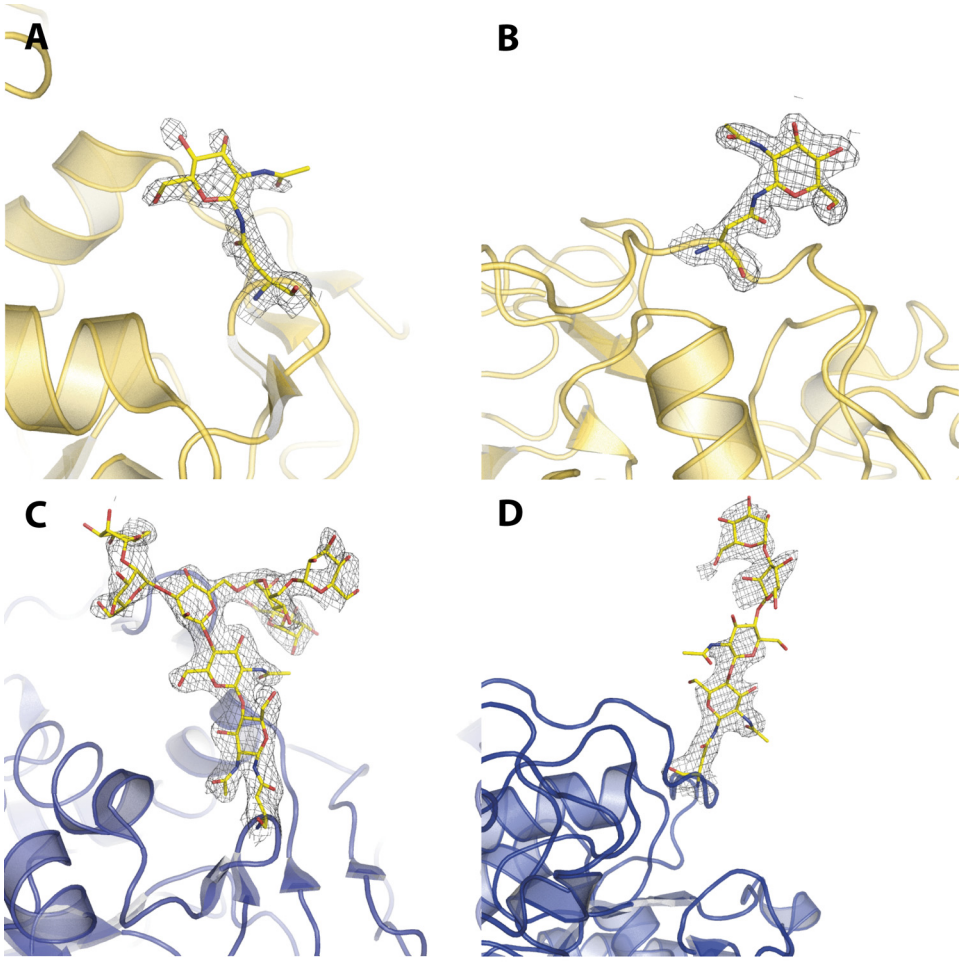


FIGURE 6. *A–E*, the data quality and modeling of glycosylations sites in *Hj*Cel3A structural models. All maps are $2mF_o - DF_c$ omit maps contoured to 1.2σ. *A*, electron density (*gray mesh*) for the single modeled NAG (*yellow sticks*) at Asn^{208} of *Hj*Cel3A (*yellow schematic*). *B*, electron density (*gray mesh*) for the single modeled NAG (*yellow sticks*) at Asn^{310} of *Hj*Cel3A (*yellow schematic*). *C*, electron density (*gray mesh*) of the modeled glycans (*yellow sticks*) of the glycosylation at Asn^{208} of *Pp-Hj*Cel3A (*blue schematic*). *D*, electron density (*gray mesh*) of the modeled glycans (*yellow sticks*) of the glycosylation at Asn^{310} of *Pp-Hj*Cel3A (*blue schematic*). *E*, *schematic drawing* of the Asn^{208} glycosylation of *Pp-Hj*Cel3A.

packing with hydrogen bonds to the side chains of Asp^{86}, Arg^{328}, and Asp^{329} of a symmetry-related molecule (Fig. 7*B*).

The glycosylations attached to Asn^{310} were modeled differently in the *Hj*Cel3A and *Pp-Hj*Cel3A structure models (Fig. 6, *B* and *D*) and also differently between the two molecules of *Pp-Hj*Cel3A. In the *Hj*Cel3A and in the A molecule of *Pp-Hj*Cel3A, only one NAG could be modeled. In the B molecule of *Pp-Hj*Cel3A, the modeled *N*-glycosylation contains four glycans that stretch toward the Asn^{208} of the symmetry-related B molecule, to which it has water-mediated hydrogen bonds.

The *N*-glycosylations were modeled using previously described patterns for chain formation pathways in yeast (67). The NAGs attached to asparagines 208 and 310 were connected by a β-1,*N* bond between the Nδ_2 nitrogen of the asparagine side chain and the C_1 carbon of the NAG and, in *Pp-Hj*Cel3A, a β-1,4 linkage between the two *N*-acetylglucosamines. The glycan polymer at

FIGURE 5. *A–C*, *stereoviews* of the −1 and +1 subsites with a van der Waals *surface representation* of *Hj*Cel3A + glucose (*yellow*) (*A*), *Hv*Exo1 (*red*) + thiocellobiose (*violet*) (*B*), and *Tn*Bgl3B + glucose (*green*) (*C*).

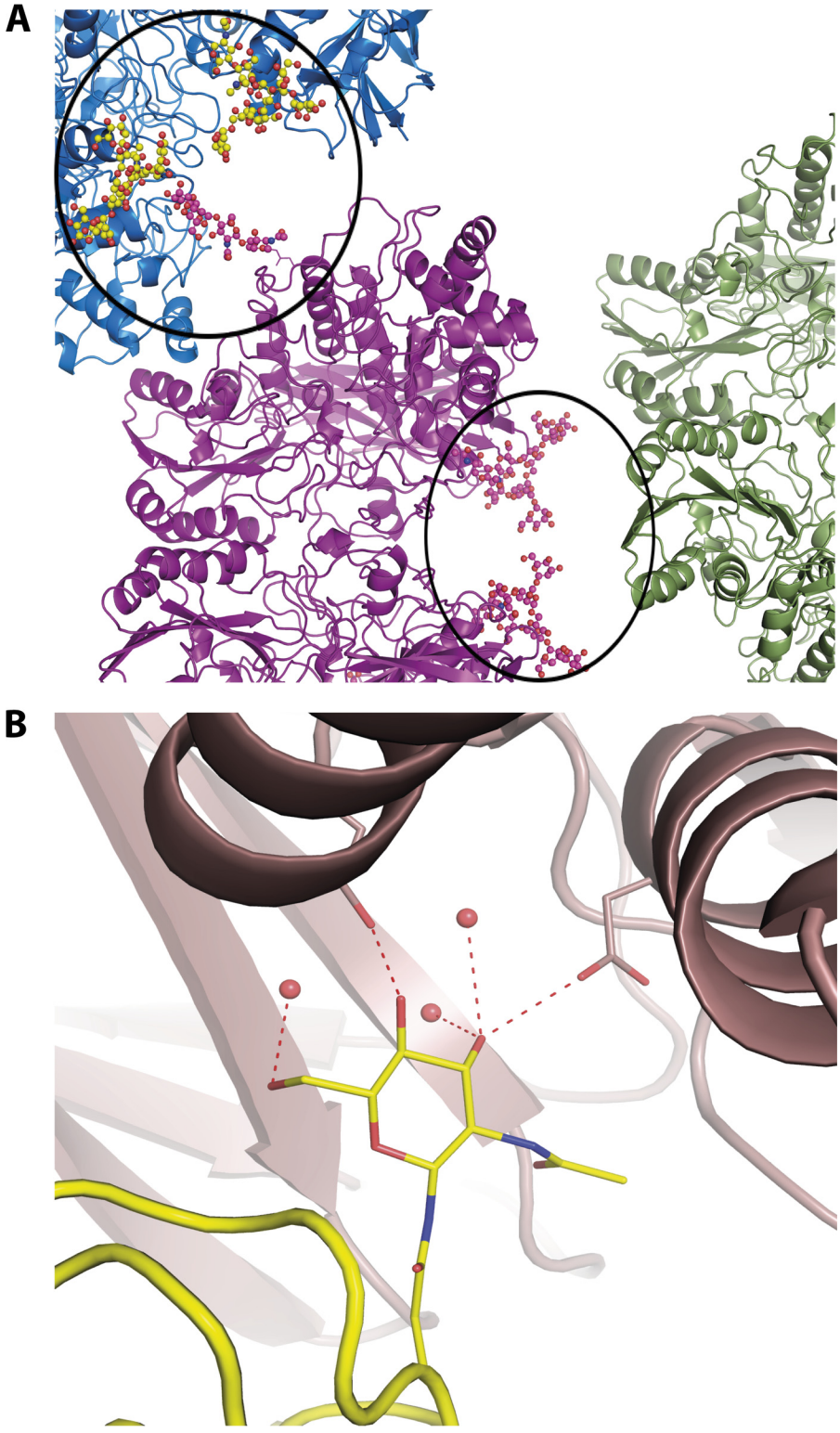


FIGURE 7. **Glycosylations that are involved in intermolecular interactions.** *A*, three asymmetric units of *Pp-Hj*Cel3A, each containing two protein molecules, are shown in *blue*, *purple*, and *green*. The crystal contacts with van der Waals interaction between glycan molecules are indicated by *black circles*. The intermolecular interactions involve both of the glycosylation sites at Asn^{208}, located at the dimer interface of the non-crystallographic symmetry-related molecules, and at Asn^{310}. *B*, interaction of the NAG molecule at Asn^{310} at one of the crystal contacts in the *Hj*Cel3A structure. This NAG glycan makes both direct interaction and water-mediated interactions with residues of a neighboring molecule.

208 continues with a β-1,4 bond to the first mannose residue. This mannose residue is a branch point with α-1,3 and α-1,6 linkages to mannose residues 5 and 2, respectively. The chain continues with another branch point at mannose 2 with α-1,2 and α-1,6 linkages and mannose 5 with an α-1,2 linkage to mannose 6 (Fig. 6*E*).

For both *Hj*Cel3A and *Pp-Hj*Cel3A, glycosylation plays a role in crystal formation through effects on intermolecular contacts.

Crystallization of *H. jecorina*-produced *Hj*Cel3A was relatively fast; crystals formed within a few days. The determined molecular weight of this sample was in agreement with that of the obtained *Hj*Cel3A molecular model, indicating that the *Hj*Cel3A species that crystallized was abundant in the sample. On the other hand, *Pp-Hj*Cel3A crystals took more than 6 months to form. The determined molecular weight of the *Pp-Hj*Cel3A was much larger than that of the obtained *Pp-Hj*Cel3A model, indicating that the *Pp-Hj*Cel3A species that crystallized represents only a subfraction of the sample. In addition, it is likely that the conformation of the *N*-glycosylation chains stabilizes associations that lead to crystal formation. The flexible nature of the longer glycosylations in *Pp-Hj*Cel3A would also contribute to the crystallization time.

Conclusions—The β-glucosidase *Hj*Cel3A from *H. jecorina* is an enzyme of industrial relevance and, as part of cellulase mixtures, contributes to the efficient production of fermentable sugars from lignocellulosic substrates. Enhanced levels of *Hj*Cel3A in *H. jecorina* cellulase mixtures benefit cellulose conversion. *Hj*Cel3A contributes to cellulose degradation by converting cellobiose to glucose and is especially efficient in converting longer cellooligosaccharides to glucose.

*Hj*Cel3A was produced in *H. jecorina* as well as in *P. pastoris*. Both samples crystallized, and the determined protein structures are essentially identical. The −1 subsite of *Hj*Cel3A is well conserved and displays the same geometry as determined for other GH3 β-glucosidases. The +1 subsite of *Hj*Cel3A appears to be narrower than those of the other known GH3 β-glucosidases for which the three-dimensional structure has been determined. Features were identified that are in agreement with the determined substrate specificity of *Hj*Cel3A.

Both crystallized *Hj*Cel3A samples have *N*-linked glycosylations attached to the enzyme, and these are more extensive in the *P. pastoris*-produced sample. The extended glycosylation of the *P. pastoris*-produced sample changes neither the thermal stability nor the activity of the enzyme. It appears to be a coincidence that in both structures, the same two asparagine residues are glycosylated, and in both structures, the glycosylations are involved in crystal contacts and as such contribute to the formation of the crystallographic space group. The protein structures obtained from the two samples are identical; thus, the main effect of the differences in glycosylation appears to be crystallization in either the $P2_12_12_1$ or $C2$ space groups.

*Acknowledgment—We thank Lucigen, Inc. (Madison, WI) for producing Pp-Hj*Cel3A.

REFERENCES

1. Horn, S. J., Vaaje-Kolstad, G., Westereng, B., and Eijsink, V. G. (2012) Novel enzymes for the degradation of cellulose. *Biotechnol. Biofuels* **5,** 45
2. Saloheimo, M., Nakari-Setälä, T., Tenkanen, M., and Penttilä, M. (1997) cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. *Eur. J. Biochem.* **249,** 584–591
3. Teeri, T. T., Lehtovaara, P., Kauppinen, S., Salovuori, I., and Knowles, J. (1987) Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. *Gene* **51,** 43–52
4. Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* **66,** 506–577, table of contents
5. Béguin, P., and Aubert, J. P. (1994) The biological degradation of cellulose. *FEMS Microbiol. Rev.* **13,** 25–58
6. Tengborg, C., Galbe, M., and Zacchi, G. (2001) Influence of enzyme loading and physical parameters on the enzymatic hydrolysis of steam-pretreated softwood. *Biotechnol. Prog.* **17,** 110–117
7. Krogh, K. B., Mørkeberg, A., Jørgensen, H., Frisvad, J. C., and Olsson, L. (2004) Screening genus *Penicillium* for producers of cellulolytic and xylanolytic enzymes. *Appl. Biochem. Biotechnol.* **113,** 389–401
8. Foreman, P. K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N. S., Goedegebuur, F., Houfek, T. D., England, G. J., Kelley, A. S., Meerman, H. J., Mitchell, T., Mitchinson, C., Olivares, H. A., Teunissen, P. J., Yao, J., and Ward, M. (2003) Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *J. Biol. Chem.* **278,** 31988–31997
9. Barnett, C. C., Berka, R. M., and Fowler, T. (1991) Cloning and amplification of the gene encoding an extracellular β-glucosidase from *Trichoderma reesei*: evidence for improved rates of saccharification of cellulosic substrates. *Biotechnology* **9,** 562–567
10. Chirico, W. J., and Brown, R. D., Jr. (1987) β-glucosidase from *Trichoderma reesei*. Substrate-binding region and mode of action on [1-^{3}H]cello-oligosaccharides. *Eur. J. Biochem.* **165,** 343–351
11. Chirico, W. J., and Brown, R. D., Jr. (1987) Purification and characterization of a β-glucosidase from *Trichoderma reesei*. *Eur. J. Biochem.* **165,** 333–341
12. Kumar, R., and Wyman, C. E. (2009) Effect of additives on the digestibility of corn stover solids following pretreatment by leading technologies. *Biotechnol. Bioeng.* **102,** 1544–1557
13. Banerjee, G., Car, S., Scott-Craig, J. S., Borrusch, M. S., Aslam, N., and Walton, J. D. (2010) Synthetic enzyme mixtures for biomass deconstruction: production and optimization of a core set. *Biotechnol. Bioeng.* **106,** 707–720
14. Salvi, D. A., Aita, G. M., Robert, D., and Bazan, V. (2010) Dilute ammonia pretreatment of sorghum and its effectiveness on enzyme hydrolysis and ethanol fermentation. *Appl. Biochem. Biotechnol.* **161,** 67–74
15. Zhang, J., Zhong, Y., Zhao, X., and Wang, T. (2010) Development of the cellulolytic fungus *Trichoderma reesei* strain with enhanced β-glucosidase and filter paper activity using strong artificial cellobiohydrolase 1 promoter. *Bioresour. Technol.* **101,** 9815–9818
16. Stubbs, K. A., Balcewich, M., Mark, B. L., and Vocadlo, D. J. (2007) Small molecule inhibitors of a glycoside hydrolase attenuate inducible AmpC-mediated β-lactam resistance. *J. Biol. Chem.* **282,** 21382–21391
17. Varghese, J. N., Hrmova, M., and Fincher, G. B. (1999) Three-dimensional structure of a barley β-D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure* **7,** 179–190
18. Pozzo, T., Pasten, J. L., Karlsson, E. N., and Logan, D. T. (2010) Structural and functional analyses of β-glucosidase 3B from *Thermotoga neapolitana*: A thermostable three-domain representative of glycoside hydrolase 3. *J. Mol. Biol.* **397,** 724–739
19. Suzuki, K., Sumitani, J. I., Nam, Y. W., Nishimaki, T., Tani, S., Wakagi, T., Kawaguchi, T., and Fushinobu, S. (2013) Crystal structures of glycoside hydrolase family 3 β-glucosidase 1 from *Aspergillus aculeatus*. *Biochem. J.* **452,** 211–221
20. McAndrew, R. P., Park, J. I., Heins, R. A., Reindl, W., Friedland, G. D., D'haeseleer, P., Northen, T., Sale, K. L., Simmons, B. A., and Adams, P. D. (2013) From soil to structure: a novel dimeric β-glucosidase belonging to glycoside hydrolase family 3 isolated from compost using metagenomic analysis. *J. Biol. Chem.* **288,** 14985–14992
21. Yoshida, E., Hidaka, M., Fushinobu, S., Koyanagi, T., Minami, H., Tamaki, H., Kitaoka, M., Katayama, T., and Kumagai, H. (2010) Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β-glucosidase from *Kluyveromyces marxianus*. *Biochem. J.* **431,**

39–49
22. Jeoh, T., Michener, W., Himmel, M. E., Decker, S. R., and Adney, W. S. (2008) Implications of cellobiohydrolase glycosylation for use in biomass conversion. *Biotechnol. Biofuels* **1,** 10
23. Walseth, C. S. (1952) Occurrence of cellulases in enzyme preparations from microorganisms. *TAPPI J.* **35,** 228–233
24. Wood, T. M. (1971) The cellulase of *Fusarium solani*: purification and specificity of the β-(1–4)-glucanase and the β-(1–4)-glucosidase components. *Biochem. J.* **121,** 353–362
25. Foreman, P., Goedegebuur, F., Van Solingen, P., and Ward, M. (January 6, 2005) Novel *Trichoderma* genes. International Patent WO/2005/001036
26. Vogel, H. J. (1956) A convenient growth medium for Neurospora (medium N). *Microbial Genet. Bull.* **13,** 42–43
27. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150,** 76–85
28. Lowry, O. H., Rosebrough, N. J., and Farr, A. L. (1951) Protein measurement with the folin phenol reagent. *J. Biol. Chem.* **193,** 265–275
29. Ghose, T. K. (1987) Measurement of cellulase activities. *Pure Appl. Chem.* **59,** 257–268
30. Lavinder, J. J., Hari, S. B., Sullivan, B. J., and Magliery, T. J. (2009) High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *J. Am. Chem. Soc.* **131,** 3794–3795
31. Leslie, A. G. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62,** 48–57
32. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D* **62,** 72–82
33. Collaborative Comuptational Project 4 (2002) High-throughput structure determination. Proceedings of the 2002 CCP4 (Collaborative Computational Project in Macromolecular Crystallography) study weekend. January, 2002. York, United Kingdom. *Acta Crystallogr. D* **58,** 1897–1970
34. Dodson, E. J., Winn, M., and Ralph, A. (1997) Collaborative Computational Project, number 4: providing programs for protein crystallography. *Methods Enzymol.* **277,** 620–633
35. Kabsch, W. (2010) Xds. *Acta Crystallogr. D* **66,** 125–132
36. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326
37. McCoy, A. J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63,** 32–41
38. Hrmova, M., Streltsov, V. A., Smith, B. J., Vasella, A., Varghese, J. N., and Fincher, G. B. (2005) Structural rationale for low-nanomolar binding of transition state mimics to a family GH3 β-D-glucan glucohydrolase from barley. *Biochemistry* **44,** 16529–16539
39. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132
40. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501
41. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60,** 2256–2268
42. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255
43. Brünger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355,** 472–475
44. Perrakis, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997) wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr. D* **53,** 448–455
45. Pannu, N. S., and Read, R. J. (1996) Improved structure refinement through maximum likelihood. *Acta Crystallogr. A* **52,** 659–668
46. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Echols, N., Headd, J. J., Hung, L. W., Jain, S., Kapral, G. J., Grosse Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2011) The Phenix software for automated determination of macromolecular structures. *Methods* **55,** 94–106
47. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954
48. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8,** 477–486
49. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21
50. 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) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–W383
51. DeLano, W. L. (2012) *The PyMOL Molecular Graphics System*, version 1.5.0.1, Schroedinger, LLC, New York
52. Bezerra, R. M., Dias, A. A., Fraga, I., and Pereira, A. N. (2006) Simultaneous ethanol and cellobiose inhibition of cellulose hydrolysis studied with integrated equations assuming constant or variable substrate concentration. *Appl. Biochem. Biotechnol.* **134,** 27–38
53. Gruno, M., Väljamäe, P., Pettersson, G., and Johansson, G. (2004) Inhibition of the *Trichoderma reesei* cellulases by cellobiose is strongly dependent on the nature of the substrate. *Biotechnol. Bioeng.* **86,** 503–511
54. Gritzali M., J., and D., B. R. (1979) The cellulase system of *Trichoderma.* The relationship between purified extracellular enzymes from induced or cellulose grown cells. *Adv. Chem. Ser.* **181,** 237–260
55. Fujdala, M. K., and Larenas, E. A. (September 2, 2010) United States Patent US2010/0221784A1
56. Korotkova, O. G., Semenova, M. V., Morozova, V. V., Zorov, I. N., Sokolova, L. M., Bubnova, T. M., Okunev, O. N., and Sinitsyn, A. P. (2009) Isolation and properties of fungal β-glucosidases. *Biochemistry* **74,** 569–577
57. Zverlov, V. V., Volkov, I. Y., Velikodvorskaya, T. V., and Schwarz, W. H. (1997) *Thermotoga neapolitana* bglB gene, upstream of lamA, encodes a highly thermostable β-glucosidase that is a laminaribiase. *Microbiology* **143,** 3537–3542
58. Hrmova, M., De Gori, R., Smith, B. J., Fairweather, J. K., Driguez, H., Varghese, J. N., and Fincher, G. B. (2002) Structural basis for broad substrate specificity in higher plant β-D-glucan glucohydrolases. *Plant Cell* **14,** 1033–1052
59. Hrmova, M., Garrett, T. P., and Fincher, G. B. (1995) Subsite affinities and disposition of catalytic amino acids in the substrate-binding region of barley 1,3-β-glucanases: implications in plant-pathogen interactions. *J. Biol. Chem.* **270,** 14556–14563
60. Hrmova, M., MacGregor, E. A., Biely, P., Stewart, R. J., and Fincher, G. B. (1998) Substrate binding and catalytic mechanism of a barley β-D-glucosidase/(1,4)-β-D-glucan exohydrolase. *J. Biol. Chem.* **273,** 11134–11143
61. Hong, J., Tamaki, H., and Kumagai, H. (2007) Cloning and functional expression of thermostable β-glucosidase gene from *Thermoascus aurantiacus*. *Appl. Microbiol. Biotechnol.* **73,** 1331–1339
62. Liu, D., Zhang, R., Yang, X., Zhang, Z., Song, S., Miao, Y., and Shen, Q. (2012) Characterization of a thermostable β-glucosidase from *Aspergillus fumigatus* Z5, and its functional expression in *Pichia pastoris* X33. *Microb. Cell Fact.* **11,** 25
63. Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L. S. (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **33,** D154–D159
64. Schwarzinger, S., Wright, P. E., and Dyson, H. J. (2002) Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. *Biochemistry* **41,** 12681–12686
65. Callebaut, I., Tasso, A., Brasseur, R., Burny, A., Portetelle, D., and Mornon, J. P. (1994) Common prevalence of alanine and glycine in mobile reactive centre loops of serpins and viral fusion peptides: do prions possess a fusion peptide? *J. Comput. Aided Mol. Des.* **8,** 175–191

66. Hrmova, M., De Gori, R., Smith, B. J., Vasella, A., Varghese, J. N., and Fincher, G. B. (2004) Three-dimensional structure of the barley β-D-glucan glucohydrolase in complex with a transition state mimic. *J. Biol. Chem.* **279,** 4970–4980
67. Wildt, S., and Gerngross, T. U. (2005) The humanization of *N*-glycosylation pathways in yeast. *Nat. Rev. Microbiol.* **3,** 119–128
68. Engh, R. A., and Huber, R. (1991) Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr. A* **47,** 392–400
69. Kleywegt, G. J. (1997) Validation of protein models from Ca coordinates alone. *J. Mol. Biol.* **273,** 371–376
70. Kleywegt, G. J., and Jones, T. A. (1996) Phi/Psi-cology: Ramachandran revisited. *Structure* **4,** 1395–1400