THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 280, NO. 38, pp. 32761–32767, September 23, 2005

How Family 26 Glycoside Hydrolases Orchestrate Catalysis on Different Polysaccharides

STRUCTURE AND ACTIVITY OF A CLOSTRIDIUM THERMOCELLUM LICHENASE, CtLic26A[*S]

Received for publication, June 16, 2005 Published, JBC Papers in Press, June 28, 2005, DOI 10.1074/jbc.M506580200

Edward J. Taylor[‡1], Arun Goyal[§1,2], Catarina I. P. D. Guerreiro[§], José A. M. Prates[§], Victoria A. Money[‡], Natalie Ferry[¶], Carl Morland[¶], Antoni Planas[‖], James A. Macdonald[**], Robert V. Stick[**], Harry J. Gilbert[¶], Carlos M. G. A. Fontes[§], and Gideon J. Davies[‡3]

*From the [‡]York Structural Biology Laboratory, Department of Chemistry, University of York, York, YO10 5YW, United Kingdom, [§]CIISA-Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Rua Prof. Cid dos Santos, 1300-477 Lisboa, Portugal, [¶]Institute for Cell and Molecular Biosciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, United Kingdom, [‖]Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon LIull, 08017 Barcelona, Spain, and **Chemistry, School of Biomedical and Chemical Sciences M313, University of Western Australia, Crawley, Western Australia 6009, Australia*

One of the most intriguing features of the 90 glycoside hydrolase families (GHs) is the range of specificities displayed by different members of the same family, whereas the catalytic apparatus and mechanism are often invariant. Family GH26 predominantly comprises β-1,4 mannanases; however, a bifunctional *Clostridium thermocellum* GH26 member (hereafter *Ct*Lic26A) displays a markedly different specificity. We show that *Ct*Lic26A is a lichenase, specific for mixed (Glcβ1,4Glcβ1,4Glcβ1,3)$_n$ oligo- and polysaccharides, and displays no activity on *manno*-configured substrates or β-1,4-linked homopolymers of glucose or xylose. The three-dimensional structure of the native form of *Ct*Lic26A has been solved at 1.50-Å resolution, revealing a characteristic $(\beta/\alpha)_8$ barrel with Glu-109 and Glu-222 acting as the catalytic acid/base and nucleophile in a double-displacement mechanism. The complex with the competitive inhibitor, Glc-β-1,3-isofagomine (K_i 1 μM), at 1.60 Å sheds light on substrate recognition in the −2 and −1 subsites and illuminates why the enzyme is specific for lichenan-based substrates. Hydrolysis of β-mannosides by GH26 members is thought to proceed through transition states in the $B_{2,5}$ (boat) conformation in which structural distinction of glucosides *versus* mannosides reflects not the configuration at C2 but the recognition of the pseudo-axial O3 of the $B_{2,5}$ conformation. We suggest a different conformational itinerary for the GH26 enzymes active on *gluco*-configured substrates.

Glycoside hydrolases play a critical role in both eukaryotes and prokaryotes, where they fulfill numerous important functions such as substrate acquisition and the remodeling of cell walls and the glycan decorations of glycoproteins. Glycoside hydrolases (GH)[4] are currently grouped into almost 100 sequence-based families (1). A further complexity of these GH families is that many different families, defined by sequence, display similar three-dimensional structures and catalytic apparatus leading to a "clan" classification, analogous to that used for proteases. Thus, clan GH-A unites a large number of diversely related GHs including 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 51, 53, 59, 72, 79, and 86 with a parallel divergence of substrate specificities (2). All these enzymes adopt the same $(\beta/\alpha)_8$ fold and perform catalysis with net retention of the configuration of the anomeric carbon via a double displacement mechanism. Notably, the location of the catalytic apparatus is completely conserved, with the catalytic acid/base and nucleophile glutamates positioned on strands β-4 and β-7, respectively (2, 3). The majority of these enzymes possess a conserved NEP motif (occasionally, such as in GH26, this is replaced by HEP) in which the glutamate is the acid/base and the Asn/His interacts with the O2 of the substrate in the critical −1 subsite, where the glycoside undergoes transition state formation during glycosidic bond hydrolysis. It is, therefore, intriguing how GH-A clan members recognize substrates differing in configuration at the C2 position while maintaining similar active-center environments. This comparison is given particular poignancy by the selective hydrolysis of mannosides, whose ground-state axial O2 renders nonenzymatic mannoside chemistry particularly challenging.

GH26 currently (31 May 2005) contains 63 open reading frames, the vast majority of which are *endo*-β-1,4-mannanases (mannanases), although two members of this family display β-1,3-xylanase activity (4, 5). Currently, the only reported three-dimensional structure for a GH26 member is for the *Cellvibrio japonicus* mannanase Man26A (6). Man26A hydrolyzes *manno*-configured substrates via a $B_{2,5}$ transition state, indicating that the enzyme exploits the pseudo-axial O3 of the transition state as an important specificity determinant and avoids steric clashes of O2 during catalysis (7). Currently, the conformational itinerary of GH26 members that display activity against *xylo*- and *gluco*-configured substrates is unclear. To address this question we have screened family GH26 for enzymes that exhibit β-glucanase activity. Here we report the cloning and overexpression of a GH26 enzyme that displays novel activity for this family; that is, the *Clostridium thermocellum* lichenase, *Ct*Lic26A. Kinetic characterization reveals the enzyme to be a classic mixed linkage β–1,4-β–1,3-glucanse that targets barley β-glucan and lichenan as its primary substrates and which exhibits no activity against β-1,4 or β-1,3 homopolymers of *gluco*- or *manno*-configured polysaccharides. The three-dimensional structure of the enzyme has also been solved in both native and inhibited forms at 1.5 and 1.6 Å, respectively. The structures shed light on the basis for substrate specificity and also hint that although GH26 enzymes active on

* This work was supported by Biotechnology and Biological Sciences Research Council for funding. 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.

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

[1] These authors contributed equally.

[2] Present address: Dept. of Biotechnology, Indian Institute of Technology Guwahati, North Guwahati, Guwahati 781 039, Assam, India.

[3] A Royal Society University Research Fellow. To whom correspondence should be addressed. E-mail: davies@ysbl.york.ac.uk.

[4] The abbreviations used are: GH, glycoside hydrolase; CAPS, 3-(cyclohexylamino)propanesulfonic acid; Glc-isoF, Glc-β1,3-isofagomine; MES, 4-morpholineethanesulfonic acid; MU, methylumbelliferyl; CBM, carbohydrate binding module.

manno-configured substrates use a $B_{2,5}$ transition state (7, 8) this is unlikely to be the case for the GH26 enzymes active on *gluco*-configured substrates. This work adds to the emerging picture (9) that the conformational itinerary of glycosidases is a subtle interplay of substrate configuration with enzyme three-dimensional structure, and thus, the conformational machinations are not conserved in different members of the same family.

EXPERIMENTAL PROCEDURES

Bacterial Strains, Plasmids, and Culture Conditions—*Escherichia coli* strains XL1-Blue (Stratagene) and BL21 (Novagen) were cultured at 37 °C in Luria broth (LB) unless otherwise stated. Media were supplemented with 100 mg/liter ampicillin or 2 mg/liter 5-bromo-4-chloro-3-indolyl β-D-galactoside to select for *E. coli* transformants and recombinants, respectively. *E. coli* cells used to propagate bacteriophage were grown on LB supplemented with 10 mM $MgSO_4$ and 0.2% maltose and were plated out on NZYM top agar (0.7%). The phage library was screened for glycoside hydrolase activity as described previously (10). The bacteriophage and plasmids employed in this work were λZAPII (Stratagene) and pET21a (Novagen).

Expression and Purification of CtLic26A-Cel5E Recombinant Derivatives—DNA encoding *Ct*Lic26A, *Ct*Cel5, *Ct*Lic26-Cel5, and *Ct*Lic26-Cel5E-CBM11 were amplified by PCR and cloned into NheI- and XhoI-restricted pET21a to generate pCF1, pCF2, pCF3, and pCF4, respectively. The recombinant proteins encoded by these four plasmids contain a C-terminal His_6 tag. To generate pCF1, QuikChange site-directed mutagenesis (primers are described in supplemental Table 1S) of pCF1 was used to introduce the authentic stop codon into *lic26A*, and thus, the encoded protein (*Ct*Lic26As) does not contain a His tag. QuikChange site-directed mutagenesis was also employed to generate E109A and E222A mutants of *Ct*Lic26A. To express the clostridial proteins, *E. coli* BL21 cells harboring the appropriate recombinant plasmid were cultured in LB containing ampicillin at 37 °C to mid-exponential phase (A_{550} 0.6), at which point isopropyl 1-thio-β-D-galactopyranoside was added to a final concentration of 1 mM, and the cultures were incubated for a further 5 h. The His_6-tagged recombinant proteins were purified from cell-free extracts by immobilized metal ion affinity chromatography as described previously (11). For crystallization, *Ct*Lic26A was further purified by size exclusion chromatography following the method of Dias *et al.* (11). *Ct*Lic26As was purified from osmotic shock fractions by anion exchange chromatography (12) followed by size exclusion chromatography (11). Preparation of selenomethionine *Ct*Lic26A followed standard procedures (11), and the protein was purified using the same procedure employed for the native enzyme except that 10 mM mercaptoethanol was included in all buffers and the purified protein was exchanged into water containing 5 mM dithiothreitol.

Enzyme Assays—The activity of *Ct*Lic26A-Cel5E-truncated derivatives against various polysaccharides was determined as described previously (11) by detecting the release of reducing sugars using the Somogyi-Nelson reagent (13). Standard assays for *Ct*Lic26A were carried out in 50 mM sodium phosphate buffer, pH 7.0, containing 1 mg/ml bovine serum albumin at 37 °C, whereas the kinetic parameters of Cel5E were determined in 50 mM sodium acetate buffer, pH 4.5, containing 1 mg/ml bovine serum albumin also at 37 °C. To explore the pH profile of both Cel5E and *Ct*Lic26A 50 mM sodium acetate, pH 4–6, sodium phosphate/citrate, pH 6–7.5, Tris/HCl, pH 7.5–8.5, and CAPS, pH 8.5–10, buffers were used in enzyme assays that employed 0.2% lichenan as the substrate. Kinetic parameters were determined using six substrate concentrations that straddled the K_m value. 4-Methylumbelliferyl glucosides were prepared and utilized as lichenase substrates as described previously (14). In the case of *Ct*Lic26A, to determine activity against 4-methylumbelliferyl glucosides, product formation was monitored at 365 nm, and the concentration of 4-methylumbelliferone was calculated using a molar extinction coefficient at 365 nm and pH 7.0 of 3520 M^{-1} cm^{-1}. The K_i value for Glc-β1,3-isofagomine (Glc-isoF) was determined by measuring the apparent K_m for 4-methylumbelliferyl Glc-β1,4-Glc-β1,3-Glc-MU using inhibitor concentrations ranging from 200 nM to 6 μM.

Crystallization and Data Statistics—Pure proteins as judged by SDS-PAGE were concentrated to 10–20 mg ml^{-1}, and buffer was exchanged into water (Sigma) using a Vivaspin 10-kDa cut-off concentrator. *Ct*Lic26A was screened using the hanging drop vapor diffusion method together with the Hampton Crystal screen, Crystal screen 2, and Hampton PEG/Ion screen (Hampton research, Alison Viejo, CA). Drops containing 1 μl of protein were mixed together with 1 or 2 μl of the mother liquor. Initial crystals were found to grow in the CSS Crystal screen I, condition number 12. This condition was optimized further to improve crystal quality and the cryogenic properties, resulting in conditions of 0.2–1.2 M sodium formate, 0.1 M sodium cacodylate buffered at pH 6.5 and 5–20% polyethylene glycol 4000. A cryo-protectant solution was produced by supplementing the mother liquor with an additional 20% glycerol. Crystals were harvested in rayon fiber loops then bathed in cryoprotectant solution before flash-freezing in liquid N_2. Selenomethionine crystals were produced in cacodylate buffer and back-soaked into MES buffer at the same pH to remove the arsenic from the crystals (and hence prevent the masking of the anomalous signal; arsenic and selenium have similar absorption spectra at energies at the Se K edge 11.8667 and 12.6578 keV, respectively). Data were collected at the European Synchrotron Radiation Facility from single crystals at 120 K with a $\Delta\varphi$ of 0.5°. Non-derivative native data were collected at a wavelength of 0.9340 Å over an oscillation range of 135° on ID14–1 using an ADSC Q4R CCD charged-coupled device detector. Selenium derivative data were collected at a wavelength of 0.9791 Å over an oscillation range of 400° on ID29 using an ADSC Q210 2D charged-coupled device detector

Structure Solution and Refinement—All data were processed and reduced with the HKL suite (15) or MOSFLM (16). All other computing was undertaken using the CCP4 suite (17) unless otherwise stated. Native *Ct*Lic26A crystals were found to belong to the space group $P2_12_12_1$ with the approximate cell dimensions of $a = 49.8$ Å, $b = 63.6$ Å, and $c = 100.0$ Å. The structure solution of *Ct*Lic26A was solved using SAD. Heavy atom positions were determined using SHELXD, initial phases were calculated using MLPHLARE, and phase improvement was performed with DM. 5% of the total reflections were flagged for cross-validation before refinement, and the behavior of these data was used to monitor the model at various stages of refinement for the weighting of geometrical and temperature factor restraints. REFMAC (18), in conjunction with ARP/wARP (19), was used to build the sequence into the electron density automatically. QUANTA and X-FIT (Accelrys, San Diego, CA) were used to make manual corrections to the model. Solvent molecules were added using X-SOLVATE and checked manually. The structure was validated using PROCHECK (20) before deposition.

Structure Determination of a CtLic26A Complex with Glc-isoF—*Ct*Lic26As (wild type without His tag) was concentrated to ~22 mg ml^{-1} in a Vivaspin 10-kDa cut-off concentrator and washed into water. The protein was screened using the sitting drop vapor diffusion method together with the Hampton Crystal screen, Crystal screen 2 and Hampton PEG/Ion (Hampton research, Alison Viejo, CA), and the CSS Crystal screen I. 100-nl droplets were formed by the mosquito® (TTP LabTech Ltd, Royston, Herts, UK) liquid-handling robot, mixing 50 nl of protein solution with 50 nl of well solution. Initial crystals were formed in Hampton crystal screen 2, condition 26. Further optimization of this condition in the hanging drop vapor diffusion method using 1 μl

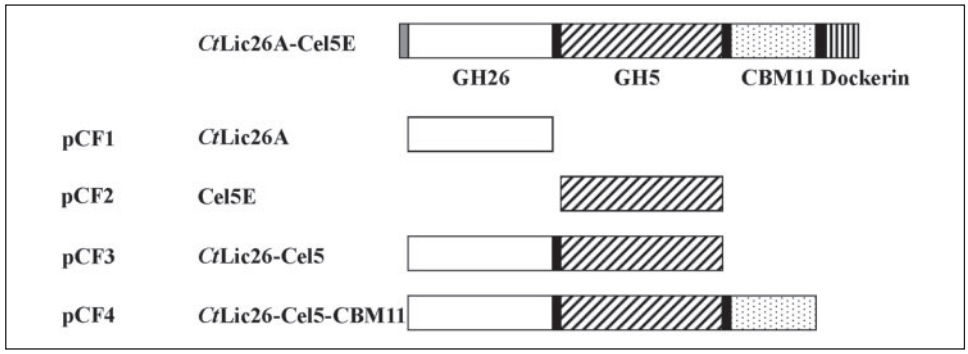

FIGURE 1. **Schematic of the molecular architecture of *Ct*Lic26A-Cel5E.** The modules encoded by the defined recombinant plasmids are indicated with the *gray* and *black boxes* representing the signal peptide and linker sequences, respectively.

of protein mixed with 1 μl of well solution combined with serial streak seeding of the resultant crystalline precipitate led to the formation of diffraction quality, prismatic crystals in a mother liquor of 0.15 M ammonium sulfate and 30% polyethylene glycol 5000 monomethylether buffered to pH 6.5 with MES. The crystals were harvested in rayon fiber loops with the mother liquor providing cryoprotection before being flash-frozen in liquid N_2. Data were collected on station ID29 at the European Synchrotron Radiation Facility from single crystals at 120 K at a wavelength of 0.9792 Å; a high resolution sweep was collected with $\Delta\varphi$ of 0.5° over an oscillation range of 134.5°. Low resolution data were collected with $\Delta\varphi$ of 1.5° over an oscillation range of 135°.

The data were indexed and integrated with MOSFLM. The protein was found to crystallize in space group $P2_12_12_1$ with cell dimensions of $a = 49.27$ Å, $b = 63.01$ Å, and $c = 78.18$ Å. Data reduction and merging were performed in SCALA. Starting phases were obtained by molecular replacement using AMoRe (21) and the previous native *Ct*Lic26A structure as a starting model. 5% of the data were set aside for determination of R_{free} to monitor the model during refinement. Refinement was undertaken in REFMAC, and solvent molecules were added using ARP/wARP and checked manually.

RESULTS AND DISCUSSION

Cloning and Expression of CtLic26A-Cel5E—In a previous study (22) it was shown that a *C. thermocellum* cellulosomal enzyme, EGH (now designated *Ct*Lic26A-Cel5E to reflect current nomenclature for carbohydrate-active enzymes (23)), contained a GH5 and a GH26 catalytic module and a C-terminal type I dockerin, the latter domain targeting the enzyme to the Clostridial cellulosome. More recently a family-11 carbohydrate binding module (CBM11) in *Ct*Lic26A-Cel5E that is located between the GH5 domain and the dockerin, which binds to β1,4- and mixed linked β1,3-β1,4-glucans, was also identified (24) (Fig. 1). The GH5 module (Cel5E) was in preliminary data shown to display endoglucanase activity. To characterize the properties of Cel5E in more detail and to determine the substrate specificity of *Ct*Lic26A, the gene, designated *celH*, encoding *Ct*Lic26A-Cel5E was isolated from a genomic library of *C. thermocellum* constructed in λZAPII by screening for clones that displayed glycoside hydrolase activity and sequencing the clostridial insert. The clone encoding *Ct*Lic26A-Cel5E also contained a second putative glycoside hydrolase gene, *cel5M*, which is located 269 bp upstream to and transcribed in the same direction as *celH.* Cel5M, encoded by *cel5M*, is 561 amino acids. Although the protein lacks an obvious signal peptide and a dockerin sequence and is, thus, not a component of the *C. thermocellum* cellulosome, it contains a transmembrane domain that extends from Leu-96 to Phe-114. The N-terminal 113 residues of Cel5M exhibit no homology to sequences in the SWISS-PROT data base, whereas the C-terminal ~450 residues displays sequence similarity with GH5 endoglucanases, exhibiting 31% identity with a *Cryptococcus neoformans* protein (accession number AAW44253). Interestingly, the catalytic acid-base glutamate is replaced by glutamine (Gln-318), although the catalytic nucleophile (Glu-443) is retained, and thus, we predict that Cel5M is unable to hydrolyze β-glucans. It is possible that the GH5 protein displays activity against *S*-linked glucosides (in GH1 myrosinases the catalytic acid/base glutamate is similarly replaced by glutamine (25)) or perhaps its true role in the bacterium is noncatalytic as is the case with many glycoside hydrolase family members.

Biochemical Properties of CtLic26A-Cel5E Catalytic Derivatives—Regions of *celH* encoding the various modules of *Ct*Lic26A-Cel5E (Fig. 1) were amplified by PCR and cloned into *E. coli* expression vectors. The encoded recombinant proteins, which were all expressed at high levels, were purified by immobilized metal ion affinity chromatography to electrophoretic homogeneity. Interrogation of the biochemical properties of the *Ct*Lic26A-Cel5E derivatives shows that Cel5E displays typical endo-β1,4-glucanase activity, hydrolyzing cellulosic substrates such as carboxymethyl cellulose, hydroxyethyl cellulose, and β-glucans while exhibiting very low levels of xylanase activity and no activity against mannans and the β-1,3-glucan, laminarin (data not shown). Typical of endoglucanases, Cel5E was considerably more active against lichenan, a highly accessible β-glucan, than the crystalline (and, thus, inaccessible) substrate Avicel (TABLE ONE). CBM11 potentiated the activity of Cel5E against the insoluble substrate Avicel but not against the soluble polysaccharide lichenan, a feature that is common to many cellulose binding CBMs (26–29).

*Ct*Lic26A, in contrast to Cel5E, does not exhibit any activity against galactomannans, glucomannans, laminarin, Avicel, or soluble cellulosic derivatives but appears to be specific for β-1,3–1,4-glucans, such as lichenan and barley β-glucan (TABLE ONE). Although CBM11 interacts with lichenan, it does not enhance the activity of *Ct*Lic26A against this substrate, and similarly, there is no obvious synergy between the two catalytic modules. Although the temperature optimum of both catalytic modules in *Ct*Lic26A-Cel5E is ~60 °C, *Ct*Lic26A and Cel5E display maximal activity at pH 7.0 and 4.5, respectively. It is unusual for two catalytic modules in the same enzyme to display significantly different pH optima, and the biological significance of this phenomenon is unclear. The relative activity of *Ct*Lic26A against aryl β-glucooligosaccharides is as follows: Glc-β1,3-Glc-MU $\ll$ Glc-β1,4-Glc-β1,3-Glc-MU $<$ Glc-β1,4-Glc-β1,4-Glc-β1,3-Glc-MU (TABLE ONE). Because the lichenase displays no activity against laminaritriose (Glc-β1,3-Glc-β1,3-Glc) the natural bond hydrolyzed by the enzyme is Glc-β1,4-Glc. Thus, the glycone region of the substrate binding cleft of *Ct*Lic26A accommodates a substrate that comprises (−4 to −1) Glc-β1,4-Glc-β1,4-Glc-β1,3-Glc; Fig. 2. Such a subsite specificity reflects the structures of lichenan and barley β-glucan whose dominant repeating unit (along with Glc-β1,4-Glc-β1,3) is Glc-β1,4-Glc-β1,4-Glc-β1,3 (30). The relative increase in k_{cat}/K_m values from the MU disaccharide to the tetrasaccharide suggest that the −3 subsite contributes significantly to catalysis, as was observed previously with the structurally unrelated lichenases from family GH16 (14). Furthermore, the higher k_{cat} of

TABLE ONE

Kinetic parameters of CtLic26A-Cel5E and its derivatives

Enzymes	Substrate	k_{cat}	K_m	k_{cat}/K_m
		min^{-1}		
*Ct*Lic26A	Lichenan	776	0.020 mg ml^{-1}	3.9×10^4 min^{-1} mg^{-1} ml
Cel5E	Lichenan	496	0.032 mg ml^{-1}	1.6×10^4 min^{-1} mg^{-1} ml
*Ct*Lic26-Cel5	Lichenan	700	0.011 mg ml^{-1}	6.4×10^4 min^{-1} mg^{-1} ml
*Ct*Lic26-Cel5-CBM11	Lichenan	679	0.009 mg ml^{-1}	7.5×10^4 min^{-1} mg^{-1} ml
E109A *Ct*Lic26A	Lichenan	NA[a]	NA	NA
E222A *Ct*Lic26A	Lichenan	NA	NA	NA
*Ct*Lic26A	Avicel	NA	NA	NA
Cel5E	Avicel	1.38	0.190 mg ml^{-1}	7.3 min^{-1} mg^{-1} ml
*Ct*Lic26-Cel5	Avicel	1.64	0.187 mg ml^{-1}	8.7 min^{-1} mg^{-1} ml
*Ct*Lic26-Cel5-CBM11	Avicel	20.5	0.405 mg ml^{-1}	5.1×10^1 min^{-1} mg^{-1} ml
*Ct*Lic26A	Glcβ3GlcβMU	2.8	5.3 mM	5.0×10^{-1} min^{-1} mM^{-1}
*Ct*Lic26A	Glcβ4Glcβ3GlcβMU	32	87 μM	3.7×10^2 min^{-1} mM^{-1}
*Ct*Lic26A	Glcβ4Glcβ4Glcβ3GlcβMU	67	42 μM	1.6×10^3 min^{-1} mM^{-1}
E109A *Ct*Lic26A	Glcβ4Glcβ3GlcβMU	0.7	3.2 μM	2.2×10^2 min^{-1} mM^{-1}
E222A *Ct*Lic26A	Glcβ4Glcβ3GlcβMU	NA	NA	NA

[a] NA, no activity detected. The assay can detect activity that is $>10^{-5}$ that of the wild type enzyme.

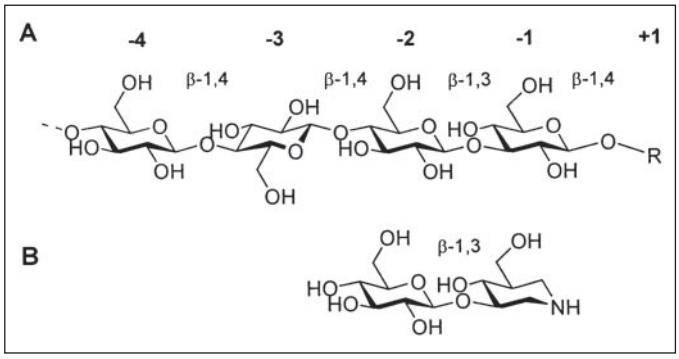

FIGURE 2. **Specificity of *C. thermocellum* *Ct*Lic26A.** Shown is a schematic diagram indicating the subsite specificity (*A*) and the disaccharide inhibitor laminaribiose-derived isofagomine (K_i 1.2 μM at pH 7.0) (*B*), which is observed binding to the −2 and −1 subsites.

*Ct*Lic26A against lichenan compared with an oligosaccharide that occupies subsites −4 to +1 indicates that the enzyme contains additional aglycone subsites, which only the polysaccharide can harness.

Three-dimensional Structure of CtLic26A—The structure of *Ct*Lic26A was solved using single-wavelength anomalous dispersion techniques, optimized for the *f″* signal of Se from a selenomethionyl-derived form of His-tagged *Ct*Lic26A. Subsequently, an isomorphous native form of *Ct*Lic26A (details of data and structure quality are shown in TABLE TWO) was refined at 1.5-Å resolution. The structure of the native form of the enzyme could be traced with no breaks from Ser-7 to His-290 with the final five traceable residues from the metal affinity tag. *Ct*Lic26A forms a $(\beta/\alpha)_8$ barrel structure, with the catalytic acid/base (Glu-109) and nucleophile (Glu-222) on strands β-4 and β-7, respectively (Fig. 3*a*), as is typical for enzymes classified into clan GH-A (2). The observation that E222A was inactive ($<10^{-5}$ of wild type *Ct*Lic26A activity) whereas E109A displayed no activity against lichenan and a significantly reduced k_{cat} and K_m (compared with wild type *Ct*Lic26A) against Glcβ1,4Glcβ1,3GlcβMU (TABLE ONE) is entirely consistent with Glu-222 and Glu-109 comprising the catalytic nucleophile and acid-base residues, respectively. Another notable feature of the catalytic apparatus of *Ct*Lic26A is that it bares a GH26 "fingerprint" in that the residue preceding the acid/base, which interacts with the 2-position of the substrate, is histidine (unlike other non-family 26 members of Clan GH-A, where the residue preceding the acid is asparagine). A DALI (31) search reveals that *Ct*Lic26A is similar to other clan GH-A glycoside hydrolases with the closest similarity, not surprisingly, with the family GH26 mannanase, Man26A, with which *Ct*Lic26A shares ~17% sequence identity (DALI Z score 21.6 with a root mean square deviation of 2.7 Å over 239 common Cα positions).

Structure of a CtLic26A-Inhibitor Complex; Molecular Basis for Substrate Specificity in Family GH26—The structure of the C-terminal His-tagged *Ct*Lic26A revealed that the affinity tag of one molecule occluded the catalytic and substrate binding center of a symmetry-equivalent molecule. This crystal form could not be used for crystal soaking experiments. Subsequently, "native" *Ct*Lic26A (*Ct*Lic26As), lacking the C-terminal tag, through the re-introduction of authentic stop codon was prepared for ligand binding studies. Crystals were grown in the presence of the disaccharide competitive inhibitor Glc-isoF (32, 33) (Fig. 2). The disaccharide competitively inhibits the lichenase at pH 7.0, with a K_i of 1.2 μM, which is >1000-fold lower than the K_m for the corresponding substrate Glc-β1,3-Glc-β-MU. Isofagomine analogues generally bind much more tightly to glycoside hydrolases than the corresponding substrate, reflecting a positively charged "anomeric" nitrogen and a strongly favorable entropy of binding (34). The structure of the inhibited enzyme could be traced with no breaks from Ser-7 to Leu-283. Comparison of the complex structure with that of the native enzyme reveals the only significant main-chain difference occurs in the loop between Glu-258 and Trp-264, which is raised by ~3.5 Å at Glu-261 to allow binding to the affinity tag in the C-terminal-tagged protein used for the native, but not complex structure determination. Indeed, in a subsequent structure of non-complexed, but His-tag-less protein (not shown), the mobile loop is indeed found to coincide with that of the inhibited enzyme.

Glc-isoF binds as expected in the −2 and −1 subsites. Extremely low level residual electron density in the +1 subsite suggested very weak binding of a second Glc-isoF molecule, but this could not be modeled appropriately. In the −2 and −1 subsites, however, the electron density at 1.6-Å resolution is unambiguous (Fig. 3*b*), and thus, the complex sheds light on the crucial interactions that define substrate specificity in GH26 (Fig. 4). In both subsites the rings lie in $^4C_1$ (chair) conformations. Consistent with other work (34–37), the endocyclic nitrogen of the isofagomine moiety makes favorable interactions with both the nucleophile, Glu-222, and the acid-base Glu-109. Other interactions of the −1 subsite include recognition of both O4 and O6 by Glu-261, with Tyr-185 interacting with the nucleophile (and most-likely O5). Phe-256 provides

TABLE TWO

X-ray data and structure quality for *C. thermocellum* CtLic26A

Outer resolution shell statistics are given in parentheses. r.m.s.d., root mean square deviation.

	*Ct*Lic26A SeMet	*Ct*Lic26A native	*Ct*Lic26A Glc-IsoF
Beam line	ID29	ID14.1	ID29
Wavelength (Å)	0.9791	0.9340	0.9792
Resolution of data (Å)	25-1.45 (1.5-1.45)	36-1.50 (1.5 -1.50)	40-1.60 (1.64-1.60)
R_{merge}	0.054 (0.11)	0.048 (0.103)	0.060 (0.087)
Mean I/σI	27 (14)	37 (12)	26 (17)
Completeness %	96 (80)	96 (94)	99 (98)
Multiplicity	5.4 (5.0)	5.4 (5.3)	6.5 (5.3)
R_{cryst}		0.131	0.15
R_{free}		0.154	0.18
r.m.s.d. 1–2 bonds (Å)		0.014	0.008
r.m.s.d. 1–3 bonds (°)		1.446	1.149
Average main chain B (Å^2)		9	4
Average side chain B (Å^2)		15	6
Average substrate B (Å^2)		NA	3
Average solvent B (Å^2)		26	19

[a] NA, no activity detected.

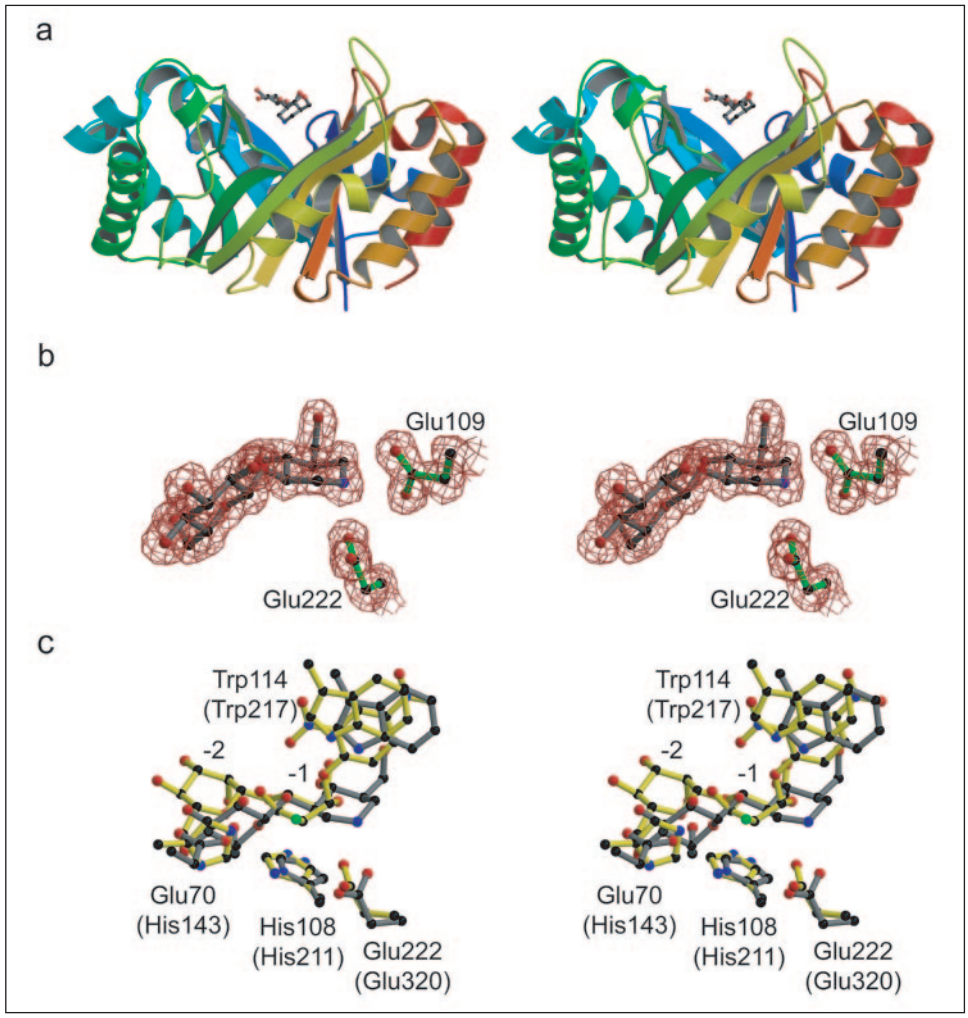

FIGURE 3. **Three-dimensional structure of *C. thermocellum* Lic26A and its complex with the laminaribio-derived isofagomine.** *a*, protein schematic, color-ramped from the N terminus (*blue*) to the C terminus (*red*) with the Glc-isof ligand in *ball-and-stick*. *b*, observed 2 $F_{obs} - F_{calc}$ electron density (contoured at 1σ) for Glc-isof and the catalytic acid/base and nucleophile, Glu-109 and Glu-222, respectively. *c*, the overlap of the active centers of the Glc-IsoF complex of *Ct*Lic26 (this study, *gray*) and the Michaelis complex of unhydrolyzed DNP 2Fmannobioside complex of *Cj*Man26 (7) (*yellow*). *Ct*Lic26 residues discussed in the text are labeled with the equivalent residue of *Cj*Man26 in *parentheses*. This figure, in divergent (wall-eyed) stereo was drawn with MOLSCRIPT (46) and BOBSCRIPT (47).

an aromatic platform which, although it does not "stack" with a ring in $^4C_1$ conformation, may well favor binding of a distorted sugar at the transition state (see below). Isofagomine lacks a hydroxyl at C2, but simple modeling studies suggest that both His-108 and Tyr-115 would interact with the O2 of a glucoside in this subsite, although the relative contribution of these two side chains will very much reflect the type and extent of substrate distortion during catalysis.

In the −2 subsite there is "tight" recognition of the sugar hydroxyls with Glu-258 and Lys-260 interacting with O6, Glu-70 with O3 and O2, and Tyr-115 with O2. Residues 258–261 are also positioned to prevent

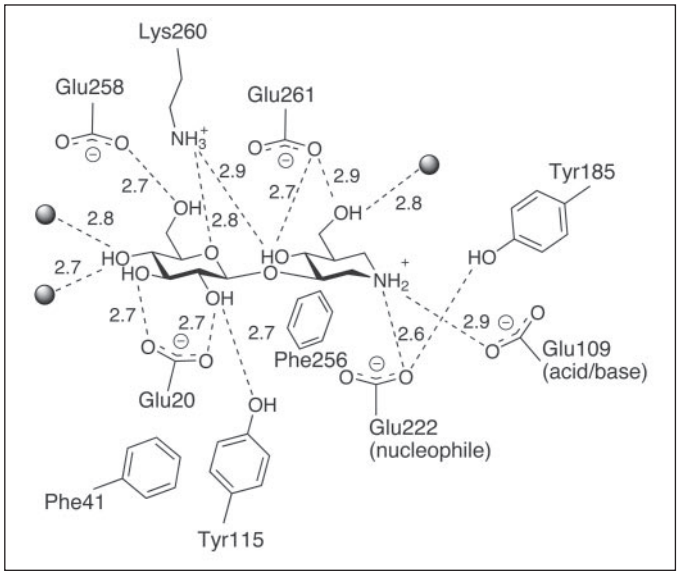

FIGURE 4. **Schematic diagram of the interactions of the *C. thermocellum* *Ct*Lic26A with Glc-IsoF.**

binding of a β-1,4-linked sugar in the -2 subsite. Phe-41 provides an aromatic platform making van der Waals contacts with the α face of the glucoside. This latter interaction presumably also contributes significantly to the recognition of laminaribiosyl moieties in subsites -2 and -1 as the relative twist of the two sugars is very different in 1,3- compared with 1,4-linked disaccharides. The -2 to -3 subsite boundary most likely selects for a β-1,4 linkage as no binding of the Glc-Glc-isoF trisaccharide was observed in crystal (data not shown). The structure of *Ct*Lic26A becomes more open in this region, but the steric clashes provided by Glu-70, Trp-72, and the loop from Asn-42 to Trp-44 would most likely all disfavor binding of a β-1,3-linked glucoside in subsite -3. In contrast, simple docking of a β-1,4-linked glucoside suggests that it would make no steric clashes.

Catalysis and Substrate Distortion—Although GH26 is dominated by *endo*-β-1,4 mannanases, two other activities have recently been described, lichenase (this work) and β-1,3-xylanase (4, 5). The comparison of *Ct*Lic26A with β-mannanases from this family is, therefore, particularly important from a mechanistic standpoint. The enzymatic hydrolysis of glycosides features oxocarbenium ion-like transition states in which the anomeric center becomes sp^2-hybridized, and partial positive charge accumulates primarily across the endocyclic O5-C1 bond (38, 39). For pyranosides such a species demands planarity of C5, O5, C1, and C2 at or near the transition state, a situation found only in $^4H_3$ and $^3H_4$ (half-chair) conformations (or their closely related envelope forms) and $^{2,5}B$ and $B_{2,5}$ (boat conformations) (38). In the context of these conformational criteria, β-mannanases, exemplified by Man26A, are a chemically fascinating group of enzymes. The Michaelis complex of unhydrolyzed substrate was trapped on Man26A (7). It revealed an unusual $^1S_5$ conformation which together with the $^OS_2$ conformation for the covalent intermediate, strongly suggested a novel conformational itinerary for these enzymes through a $B_{2,5}$ transition state. Such a transition state would have the crucial facet that it would alleviate 1,2-*cis*-di-axial repulsions at the transition state by placing a *manno*-configured O2 pseudo-equatorial while still placing the glycosidic oxygen axial, allowing in-line nucleophilic attack. Support for these proposals comes both from the demonstration that retaining α-mannosidases use the reverse ($^OS_2$-$B_{2,5}$) conformational pathway (40) and by otherwise counterintuitive observations that inhibitors with in-plane substituents at O2 are potent mannosidase inhibitors (8, 41).

We have previously proposed that a crucial structural feature that helps discriminate between these two conformational pathways is the recognition of O3, an atom whose position changes markedly between $^4H_3$ and $B_{2,5}$ transition states (7, 8). It may be significant that the two enzyme activities specific for *xylo*- and *gluco*-configured substrates in family GH26 are both β-1,3 glycosidases in which O3 is tethered to the adjacent sugar which, in tandem with active-site topography, may legislate against the sugar attaining a $B_{2,5}$ transition state.

Previous comparison of family GH1 and GH5 enzymes specific for glucosides, with closely related enzymes active on mannose-derived substrates, suggested that changes in loop conformations around the -1 subsite resulted in a different position for the O3-interacting side chains and contributed to the *gluco versus manno* specificity. In GH26, such loop conformational changes are not obvious, but such a structural difference may not be as necessary because the 3-position in a β-1,3-linked polymer is necessarily tethered. Such a hypothesis is certainly consistent with the overlap of the catalytic centers of Man26A with *Ct*Lic26A (Fig. 3*c*). In the -1 subsite, interactions around O2, C1, and O5 are invariant; in particular, the positions of catalytic acid (Glu-109) and nucleophile (Glu-222), the O2-interacting His-108 and Tyr-185 are all near identical between the *manno*- and *gluco*-specific enzymes. In the $+1$ subsite, Trp-114 of *Ct*Lic26A lies in the same position as Trp-217 of Man26A, the latter known to be involved in "stacking" against the leaving group moiety. What are indeed demonstrably different are the interactions around O3. In Man26A, O3 is coordinated by His-143. In *Ct*Lic26A O3 is involved in the glycosidic linkage to the -2 subsite glucoside and, thus, makes no hydrogen bonds ($<$3.1 Å) to protein. Instead, the *Ct*Lic26A -2 sugar lies both displaced 4–6 Å and rotated almost exactly 90° relative to the -2 sugar of Man26A. Glu-70, the residue most structurally analogous to His-153 of Man26A is then able to coordinate the O2 and O3 of the -2 subsite sugar.

Summary—The biochemical properties of *Ct*Lic26A, thus, indicate that the enzyme is a typical lichenase. The 63 members of GH26 predominantly display endo-β1,4-mannanase activity, whereas two enzymes in this family hydrolyze β-1,3-xylan. *Ct*Lic26A is, thus, the first example of a GH26 enzyme that displays lichenase activity. Inspection of the SwissProt data base reveals a GH26 *Clostridium acetobutylicum* protein (accession number Q97G16) that displays 52% sequence identity with *Ct*Lic26A. Significantly, Glu-70, Phe-41, Tyr-115, Glu-258, and Lys260, which all play key roles in recognition of the β-1,3-linked *gluco*-configured sugar in the -2 subsite of Lic26A (see below), are conserved in the *C. acetobutylicum* protein, indicating that this GH26 member is likely to display lichenase activity.

The 90 families of glycoside hydrolases provide an excellent system for studying enzyme specificity and particularly the conformation of sugars at and near the transition state. A defining feature of recent years has been the realization that different enzymes may use distinct transition-state structures, all governed by the requirements for an oxocarbenium ion-like species. It is clear that transition-state structure is dictated both by the chemical structure of the glycon moiety itself and the three-dimensional environment of the active center; neither facet is dominant. For example, xyloside hydrolysis in family GH11 is believed to go via $^{2,5}B$ transition states (42), whereas in the structurally dissimilar family GH10, a $^4H_3$ transition state, seems more likely. Glucoside hydrolysis likely also takes these two different transition states exemplified by families GH5 (43) and GH6 (44), respectively. Similarly, mannoside hydrolysis most likely occurs via a $B_{2,5}$ transition state in β-mannosidase families 5 (11) and 26 (7) and α-mannosidase family GH38 (40) but via a $^3H_4$ conformation in family GH47 (45). Furthermore, although many of these families are structurally quite distinct, others display a high degree of sequence and structural similarity. β-Glycosidases from

GH5 and GH26 are particularly fascinating since these families contain enzymes specific for both *manno*-and *gluco*-configured substrates, yet they appear to have structural invariance around the 2-position of the glycone in the −1 subsite. So, although in peripheral subsites substrate specificity simply reflects recognition of the axial *versus* equatorial nature of $^4C_1$ (chair)-conformed glycoside moieties, in the catalytic −1 subsite similar recognition apparatus around O2 favors binding of transition states in which O2 lies pseudoequatorial, $^4H_3$ for glucosides and $B_{2,5}$ for mannosides. GH5 and GH26 present different evolutionary routes to solve the conformational aspects of catalysis. In mannosidases in these two families the O3-interacting residue favors binding of a pseudoaxial O3 at the $B_{2,5}$ transition state, and this residue is indeed displaced for the glucosidases from this family. Such features need not apply to the GH26 enzymes specific for *gluco*-configured substrates, exemplified by *Ct*Lic26A here, since all non-mannosidases in this family reported thus far are active on β-1,3-linked substrates in which O3 is bonded to the adjacent sugar. The different yet complementary strategies adopted throughout evolution by glycosidases not only reveal how three-dimensional structure and substrate reactivity combine to allow catalysis on different substrates but also strongly point to new and powerful strategies for enzyme inhibition through conformational mimicry with ramifications far beyond plant cell wall degradation.

Acknowledgments—We thank the European Synchrotron Radiation Facility for provision of data collection facilities and Professor Eleanor Dodson for advice with phasing strategies.

REFERENCES

1. Coutinho, P. M., and Henrissat, B. (1999) in *Recent Advances in Carbohydrate Engineering* (Gilbert, H. J., Davies, G. J., Svensson, B., and Henrissat, B., eds) pp. 3–12, Royal Society of Chemistry, Cambridge, UK
2. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 7090–7094
3. Jenkins, J., Leggio, L. L., Harris, G., and Pickersgill, R. (1995) *FEBS Lett.* **362,** 281–285
4. Araki, T., Hashikawa, S., and Morishita, T. (2000) *Appl. Environ. Microbiol.* **66,** 1741–1743
5. Okazaki, F., Tamaru, Y., Hashikawa, S., Li, Y. T., and Araki, T. (2002) *J. Bacteriol.* **184,** 2399–2403
6. Hogg, D., Woo, E.-J., Bolam, D. N., McKie, V. A., Gilbert, H. J., and Pickersgill, R. W. (2001) *J. Biol. Chem.* **276,** 31186–31192
7. Ducros, V., Zechel, D. L., Murshudov, G. N., Gilbert, H. J., Szabo, L., Stoll, D., Withers, S. G., and Davies, G. J. (2002) *Angew. Chem. Int. Ed. Engl.* **41,** 2824–2827
8. Vincent, F., Gloster, T. M., Macdonald, J. M., Morland, C., Stick, R. V., Dias, F. M. V., Prates, J. A. M., Fontes, C. M. G. A., Gilbert, H. J., and Davies, G. J. (2004) *Chembiochem* **5,** 1596–1599
9. Davies, G. J., Ducros, V. M.-A., Varrot, A., and Zechel., D. L. (2003) *Biochem. Soc. Trans.* **31,** 523–527
10. Fontes, C. M., Hazlewood, G. P., Morag, E., Hall, J., Hirst, B. H., and Gilbert, H. J. (1995) *Biochem. J.* **307,** 151–158
11. Dias, F. M. V., Vincent, F., Pell, G., Prates, J. A. M., Centeno, M. S. J., Tailford, L. E., Ferreira, L. M. A., Fontes, C. M. G. A., Davies, G. J., and Gilbert, H. J. (2004) *J. Biol. Chem.* **279,** 25517–25526
12. Bolam, D. N., Hughes, N., Virden, R., Lakey, J. H., Hazlewood, G. P., Henrissat, B., Braithwaite, K. L., and Gilbert, H. J. (1996) *Biochemistry* **35,** 16195–16204
13. Somogyi, M. (1952) *J. Biol. Chem.* **195,** 19–23
14. Malet, C., and Planas, A. (1997) *Biochemistry* **36,** 13838–13848
15. Otwinowski, Z., and Minor, W. (1997) *Macromolecular Crystallography, Pt A*, Vol. 276, pp. 307–326, Academic Press Inc., San Diego, CA
16. Leslie, A. G. W. (1992) in *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, Vol. 26, Daresbury Laboratory, Warrington, UK
17. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. Sect. D* **50,** 760–763
18. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D* **53,** 240–255
19. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6,** 458–463
20. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
21. Navaza, J., and Saludjian, P. (1997) *Methods Enzymol.* **276,** 581–594
22. Yague, E., Beguin, P., and Aubert, J. P. (1990) *Gene* (*Amst.*) **89,** 61–67
23. Henrissat, B., Teeri, T. T., and Warren, R. A. (1998) *FEBS Lett.* **425,** 352–354
24. Carvalho, A. L., Goyal, A., Prates, J. A., Bolam, D. N., Gilbert, H. J., Pires, V. M., Ferreira, L. M., Planas, A., Romao, M. J., and Fontes, C. M. (2004) *J. Biol. Chem.* **279,** 34785–34793
25. Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., and Henrissat, B. (1997) *Structure* **5,** 663–675
26. Boraston, A. B., Kwan, E., Chiu, P., Warren, R. A., and Kilburn, D. G. (2003) *J. Biol. Chem.* **278,** 6120–6127
27. Hall, J., Black, G. W., Ferreira, L. M., Millward-Sadler, S. J., Ali, B. R., Hazlewood, G. P., and Gilbert, H. J. (1995) *Biochem. J.* **309,** 749–756
28. Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., and Claeyssens, M. (1988) *Eur. J. Biochem.* **170,** 575–581
29. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) *Biochem. J.* **382,** 769–781
30. Brett, C. T., and Waldren, K. (1996) *Physiology and Biochemistry of Plant Cell Walls*, Chapman & Hall, London
31. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
32. Macdonald, J. M., Stick, R. V., Tilbrook, D. M. G., and Withers, S. G. (2002) *Aust. J. Chem.* **55,** 747–752
33. Macdonald, J. M., Hrmova, M., Fincher, G. B., and Stick, R. V. (2004) *Aust. J. Chem.* **57,** 187–191
34. Varrot, A., Tarling, C. A., Macdonald, J. M., Stick, R. V., Zechel, D., Withers, S. G., and Davies, G. J. (2003) *J. Am. Chem. Soc.* **125,** 7496–7497
35. Notenboom, V., Williams, S. J., Hoos, R., Withers, S. G., and Rose, D. R. (2000) *Biochemistry* **39,** 11553–11563
36. Mark, B. L., Vocadlo, D. J., Zhao, D. L., Knapp, S., Withers, S. G., and James, M. N. G. (2001) *J. Biol. Chem.* **276,** 42131–42137
37. Zechel, D. L., Boraston, A. B., Gloster, T. M., Boraston, C. M., Macdonald, J. M., Tilbrook, D. M. G., Stick, R. V., and Davies, G. J. (2003) *J. Am. Chem. Soc.* **125,** 14313–14323
38. Davies, G. J., Sinnott, M. L., and Withers, S. G. (1997) in *Comprehensive Biological Catalysis* (Sinnott, M. L., ed) Vol. 1, pp. 119–209, Academic Press, Ltd., London
39. Vasella, A., Davies, G., and Böhm, M. (2002) *Curr. Opin. Chem. Biol.* **6,** 619–629
40. Numao, S., Kuntz, D. A., Withers, S. G., and Rose, D. R. (2003) *J. Biol. Chem.* **278,** 48074–48083
41. Lillelund, V. H., Liu, H. Z., Liang, X. F., Søhoel, H., and Bols, M. (2003) *Org. Biomol. Chem.* **1,** 282–287
42. Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jorgensen, P. L., Schulein, M., Dupont, C., Davies, G. J., and Wilson, K. S. (1999) *Chem. Biol.* **6,** 483–492
43. Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schulein, M., and Withers, S. G. (1998) *Biochemistry* **37,** 11707–11713
44. Varrot, A., Leydier, S., Pell, G., Macdonald, J. M., Stick, R. V., Gilbert, H. J., and Davies, G. J. (2005) *J. Biol. Chem.* **280,** 20181–20184
45. Vallée, F., Karaveg, K., Herscovics, A., Moremen, K. W., and Howell, P. L. (2000) *J. Biol. Chem.* **275,** 41287–41298
46. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950
47. Esnouf, R. M. (1997) *J. Mol. Graph. Model.* **15,** 132–134