Aspergillus aculeatus β-1,4-Galactanase: Substrate Recognition and Relations to Other Glycoside Hydrolases in Clan GH-A[†]

Carsten Ryttersgaard,[‡,§] Leila Lo Leggio,[‡] Pedro M. Coutinho,[||,⊥] Bernard Henrissat,[||] and Sine Larsen*[,‡]

Centre for Crystallographic Studies, Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark, and Architecture et Fonction des Macromolécules Biologiques, UMR6098, CNRS and Universités d'Aix-Marseille I and II, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France

Received June 4, 2002; Revised Manuscript Received September 6, 2002

ABSTRACT: The three-dimensional structure of *Aspergillus aculeatus* β-1,4-galactanase (AAGAL), an enzyme involved in pectin degradation, has been determined by multiple isomorphous replacement to 2.3 and 1.8 Å resolution at 293 and 100 K, respectively. It represents the first known structure for a polysaccharidase with this specificity and for a member of glycoside hydrolase family 53 (GH-53). The enzyme folds into a $(\beta/\alpha)_8$ barrel with the active site cleft located at the C-terminal side of the barrel consistent with the classification of GH-53 in clan GH-A, a superfamily of enzymes with common fold and catalytic machinery but diverse specificities. Putative substrate−enzyme interactions were elucidated by modeling of β-1,4-linked galactobioses into the possible substrate binding subsites. The structure and modeling studies identified five potential subsites for the binding of galactans, of which one is a pocket suited for accommodating the arabinan side chain in arabinogalactan, one of the natural substrates. A comparison with the substrate binding grooves of other Clan GH-A enzymes suggests that shape complementarity is crucial in determining the specificity of polysaccharidases.

Pectin is one of the most complex and abundant components of the plant cell wall (*1*). It is a complex polysaccharide featuring "smooth" regions of α-1,4-linked galacturonic acid (homogalacturonan) and "hairy" regions of rhamnogalacturonan (*2*). The most abundant form of rhamnogalacturonan, rhamnogalacturonan I (RG-I), is a polysaccharide built from repeating units of the disaccharide (1,2)-α-L-rhamnopyranose (1,4)-α-D-galacturonic acid (*3*). The C4 position of rhamnose can often serve as an attachment site for carbohydrate side chains such as arabinan, galactan and arabinogalactan. These side chains form long branches, which explains why the regions of pectin containing them are referred to as "hairy" regions (*4*).

Degradation or modification of the galactan and arabinogalactan side chains has many industrial applications in the textile, detergent, and cellulose fiber processing industries and in the improvements of animal feed (*5*). This wide spectrum of potential applications for plant cell wall material fuels the study of enzymes that can assist in its degradation and modification.

Due to its complexity, RG-I requires several different enzymes for its degradation into mono and disaccharide components. As part of our structural characterization of the RG-I degrading system of the fungus *Aspergillus aculeatus*, we have previously determined the structure of two enzymes acting on the backbone, rhamnogalacturonase A (*6*), and rhamnogalacturonan acetylesterase (*7*). Here we describe the structure of an enzyme from the same organism, which acts on RG-I side chains. *A. aculeatus* β-1,4-galactanase (AAGAL, EC 3.2.1.89)[1−3] hydrolyzes β-1,4-linked galactan and type I arabinogalactan to galactose and galactose oligomers. Introduction of the gene for this galactanase in vivo produced potatoes with altered pectin composition (*8*).

From sequence analysis AAGAL was classified into glycoside hydrolase family 53 (*9*), which contains sequences of β-1,4-galactanases as well as sequences of gene products of undetermined function. No three-dimensional structure has been determined for this family to date. Hydrophobic cluster analysis of the amino acid sequence surrounding the catalytic residues (*10*) or inspection of the three-dimensional structures (*11*) revealed that several families of glycoside hydrolases share the same $(\beta/\alpha)_8$ barrel fold, the same catalytic machinery, and the same mechanism and thus form a superfamily or clan, Clan GH-A. Several clans of glycoside hydrolases have since been identified (*12*). Clan GH-A is

[†] This research was funded by the Danish National Research Foundation.

* Corresponding author. Telephone: (+45)35320282. Fax: (+45)-35320299. E-mail: sine@ccs.ki.ku.dk.

[‡] University of Copenhagen.

[§] Present address: UCLA-DOE Laboratory of Structural Biology, Box 951569, Los Angeles, California 90095-1569.

[||] CNRS and Universités d'Aix-Marseille I and II.

[⊥] Present address: Centre for Biological and Chemical Engineering, Department of Chemical Engineering, Instituto Superior Técnico, Technical University of Lisbon, Avenida Rovisco Pais, 1049-001 Lisboa, Portugal.

[1] Since submission of this article, the structure of a family enzyme has been reported: Hidaka, M., Fushinobu, S. Ohtsu, N., Motoshima, H., Matsuzawa, H., Shoun, H., and Wakaji, T. (2002) *J. Mol. Biol. 323*, 79−91.

[2] Abbreviations: AAGAL, *Aspergillus aculeatus* galactanase; AAGAL293, *Aspergillus aculeatus* galactanase structure determined at room temperature; AAGAL100, *Aspergillus aculeatus* galactanase structure determined at 100 K; OrPt, chloro(2,2′:6′,2″-terpyridine)-platinum(II)chloride ("orange platinum"); pcmbs, *p*-chloro-mercuribenzenesulphonate.

[3] Accession numbers: The two structures and structure factors have been deposited in the Protein Data Bank. The accession codes are 1FHL (AAGAL293) and 1FOB (AAGAL100) at RCSB.

10.1021/bi026238c CCC: $22.00
Published on Web 11/27/2002

Table 1: Data Collection Statistics

data set	AAGAL293/AAGAL100	OrPt1[a]/OrPt2/pcmbs
temperature (K)	293/100	293
cell dimensions (Å) *a*, *b*, *c*	60.38, 88.70, 129.24/59.69, 87.33, 127.55	60.33, 88.85, 129.30/60.27, 88.87, 129.35/60.46, 89.14, 129.34
resolution range (Å)	28.90−2.30/28.59−1.80	28.83−2.70/27.90−2.85/28.99−2.55
outermost shell (Å)	2.42−2.30/1.83−1.80	2.84−2.70/3.00−2.85/2.69−2.55
no. of unique reflections	15582/31237	9829/8306/11588
multiplicity	4.6/4.0	4.9/4.4/4.3
completeness overall (%)	97.5/98.7	99.6/98.5/98.7
completeness outermost shell (%)	95.6/97.7	99.2/98.3/97.2
overall $I/(I) > 2$ (%)	82.3/90.4	85.8/82.4/79.7
outermost shell $I/(I) > 2$ (%)	61.4/66.6	70.0/62.1/59.5
R_{merge} overall (%)	9.6/4.1	10.1/13.4/12.6
R_{merge} outermost (%)	39.2/35.4	28.6/35.7/40.2
heavy atom concentration (mM)		2.4/1.9/0.7
soaking time (days)		19/20/16
R_{iso} (%)		14.5/15.5/15.0
statistics from SHARP		
R_{cullis} (centric/acentric) (%)		48/49/51/52/76/77
phasing power (centric/acentric)		2.6/3.3/2.8/3.3/1.3/1.5
heavy atom sites		1/1/1
occupancy (%)		84.7/83.1/49.3
B factor (Å^2)		42.4/39.1/38.6

[a] With the first heavy-atom derivative, *I*222 was assigned as the correct space group. $R_{merge} = \Sigma_{hkl}\Sigma_i|I(hkl)_i - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I(hkl)_i$, where *i* runs over the number of batches. $R_{iso} = \Sigma_{hkl}||F_{PH}| - |F_P||/\Sigma_{hkl}|F_P|$. $R_{cullis} = \Sigma_{hkl}||F_{PH} \pm F_P| - F_{H(calc)}|/\Sigma_{hkl}|F_{PH} - F_P|$. Phasing power $= \Sigma_{hkl}F_H/\Sigma_{hkl}|F_{PH(obs)} - F_{PH(calc)}|$. The heavy-atom concentration is the actual concentration in the drop. Isomorphous occupancies and *B*-factors correspond to the heavy atom site found in each derivative data set.

the largest of these clans and has been proposed to contain families 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 51, 59, 72, 79, and 86, as well as family 53, on the basis of distant local sequence similarities (*9*). The three-dimensional structures have established the relatedness of families 1, 2, 5, 10, 17, 26, and 39 (the latter as yet unpublished), while the definitive assignment of the other families to the clan still awaits experimental validation.

All established and proposed glycoside hydrolases of clan GH-A, including family 53 enzymes, operate via a mechanism retaining the anomeric configuration at the site of cleavage (*13*, *14*). The mechanism involves initial protonation of the glycosidic oxygen by a catalytic acid/base, while a catalytic nucleophile attacks the anomeric carbon to form a covalent enzyme−substrate intermediate. The intermediate is subsequently released via nucleophilic attack by a water molecule 'activated' by the acid/base residue. In clan GH-A both catalytic residues are glutamates located at the ends of β-strands 4 (acid/base) and 7 (nucleophile). The prediction that family 53 belongs to clan GH-A was used to identify the catalytic residues of the galactanase from *Pseudomonas cellulosa* (*15*).

The members of clan GH-A display a large variety of substrate specificity and cover over 20 EC numbers (*12*)]. The glycosidic substrates are either pyranosides or furanosides, and can differ with respect to the substituents and chirality at the C2, C3, C4 and C5 carbon atoms. The substrates thus include β-D-glucosides, β-D-galactosides, β-D-mannosides, β-D-glucuronides, β-D-xylosides, α-L-arabinofuranosides and α-L-iduronides, either as simple glycosides or as oligo- and polysaccharides of varying connectivity. Clan GH-A of the glycoside hydrolases therefore provides an excellent opportunity to study nature's protein engineering, e.g., the acquisition of different substrate specificity from the same ancestral framework.

We have determined the structure of *Aspergillus aculeatus* galactanase at two different temperatures and, by comparison with the known structures of other clan GH-A glycoside hydrolases, we have been able to propose a binding mode for galactan supported by molecular docking of galactobiosides in the substrate binding groove.

MATERIALS AND METHODS

Crystallization and Data Collection. β-1,4-Galactanase from *Aspergillus aculeatus* was overexpressed in an *Aspergillus oryzae* host system, and the recombinant enzyme was purified as described earlier (*16*). Crystallization of the enzyme was only achieved when divalent cations, preferably Ca^{2+} were present in the crystallization conditions (*17*). The enzyme crystallizes in the orthorhombic space group *I*222 with one molecule in the asymmetric unit.

The structure was determined by MIR. An in-house *R*-axis II image plate system equipped with a Rigaku rotating anode CuKα X-ray source was used for the data collection at room temperature of a native data set reported previously (AAGAL 293, *17*) and of heavy atom derivatives. Preparation of heavy atom derivatives was done by conventional soaking experiments. Two data sets were collected by soaking AAGAL crystals in chloro(2,2′:6′,2″-terpyridine)platinum(II)chloride (OrPt) at two different concentrations, resulting in one platinum site. Another derivative data set was collected on a *p*-chloromercuribenzene sulfonate (pcmbs) soaked crystal, which revealed one mercury site. The integrated intensities were obtained using the HKL package (*18*), and further data processing was done using programs from the CCP4 package (*19*).

After the structure had been determined, a data set was measured on a cryo-cooled crystal (AAGAL100, $T = 100$ K) at beam line BL711 (MAXLAB, Lund, Sweden). The data were collected with a MAR345 imaging plate detector system using a wavelength of 1.087 Å. Details of all data collection statistics are in Table 1.

Structure Determination. The isomorphous difference Patterson maps revealed one heavy atom site in each

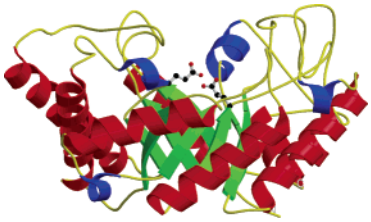

FIGURE 1: The overall fold of the β-1,4-galactanase, with the $(\beta/\alpha)_8$ barrel viewed perpendicular to the barrel axis.The figure was prepared with BOBSCRIPT, an extension of MOLSCRIPT (*73*).

derivative with peak heights of 14 σ, 11 σ, and 12 σ in the two orange platinum and pcmbs derivatives, respectively. The heavy atom parameters were refined with SHARP (*20*) (Table 1). The phases were improved, with the SOLOMON (*21*) solvent-flattening procedure implemented in SHARP (47% solvent content). The resulting electron density map, based on data in the resolution range 10.0−2.55 Å, was of high quality, and a bones skeleton was constructed with the program MAPMAN (*22*). The electron density map and bones skeleton were visualized within the program TURBO−FRODO (*23*), which was used to construct the model.

Electron density corresponding to a cysteine was found adjacent to the mercury site. This cysteine (Cys118) was used as a starting point for the tracing of the molecule. The platinum site was located adjacent to His152 and residues 32−318 could be located. The new structure factors calculated based on this model were used in combination with MIR phases (20−2.55 Å) and followed by refinement in X-PLOR (*24*) based on all data from 10.0 to 2.3 Å. New $2F_o-F_c$ and F_o-F_c electron density maps calculated using these phases enabled identification of additional residues in the structure. After this procedure of phase combination and refinement had been repeated four times, all 334 residues in AAGAL were located. Although successful crystallization required the presence of calcium ions, the difference electron density maps for AAGAL293 did not reveal any calcium ions, nor carbohydrates bound at the potential N-glycosylation sites. Water molecules were selected according to the criterion that they were within hydrogen bonding distance to the protein, had a peak height above 1.5 σ in the $2F_o-F_c$ map, and that they refined with *B* factor less than 60 Å^2. One water molecule (Wat401) was located on a 2-fold axis; the coordinates and occupancy were fixed according to the site symmetry. Ten percent of total reflections were used for calculation of R_{free} (*25*).

The program CNS (*26*) was used for refinement of AAGAL100. The AAGAL293 structure was used as an initial model in a rigid body refinement, and the refinement followed the same procedure as described above. The high-resolution limit for the AAGAL100 data was 1.8 Å. A calcium ion was identified in the active site, coordinated to seven water molecules (Figure 2). During the final refinement, three surface residues (Asn51, Asp91, and Asn112) were modeled in two conformations of equal occupancy. A summary of the statistics for the refinements of AAGAL293 and AAGAL100 is shown in Table 2. The Ramachandran plot calculated by PROCHECK (*27*) had all residues in the allowed regions except for Leu293 in AAGAL293 and Ser219 in AAGAL293 and AAGAL100. These residues were in the generously allowed regions.

Table 2: Refinement Summary

data set	AAGAL293	AAGAL100
temperature (K)	293	100
resolution range (Å)	28.90−2.30	28.59−1.80
R-factor	0.166	0.212
R_{free}	0.245	0.252
protein atoms[a]	2598	2607
calcium ion	0	1
water molecules	59	193
RMSD[b]		
bond lengths (Å)	0.008	0.006
bond angles (deg)	1.4	1.3
average *B*-factor (Å^2)		
main chain	26.1	27.0
side chains	28.2	28.8
water molecules	27.2	35.3

[a] Nonhydrogen protein atoms inserted in the model. [b] Root-mean-square deviation from ideal values.

Modeling. Single α-D-galactopyranose rings were initially constructed in the $^4C_1$ and $^1S_3$ conformations and were subsequently minimized using the molecular mechanics program MM3(92) (*28*−*30*) employing a dielectric constant of 4. The $^1S_3$ conformer corresponds to a local minimum with an energy about 6.8 kcal/mol higher than the $^4C_1$ global minimum (*31*). A methyl β-galactobioside with a $^4C_1$ conformation in both rings (β-D-Galp-β (1−4)-D-Galp-O-Me, $^4C_1-^4C_1$) was constructed. The orientation of the two pyranose rings relative to each other can be described by the two torsion angles, $\phi_H = \Theta(H1_A-C1_A-O4_B-C4_B)$, $\psi_H = \Theta(C1_A-O4_B-C4_B-H4_B)$. Each galactopyranose ring can have the exocyclic hydroxyl groups simultaneously oriented clockwise or anticlockwise with respect to the carbohydrate ring, and the hydroxymethyl group can adopt three orientations (*gg*, *gt*, *tg*) relative to the ring (*32*). The six exocyclic combinations for each pyranose ring led to 36 different

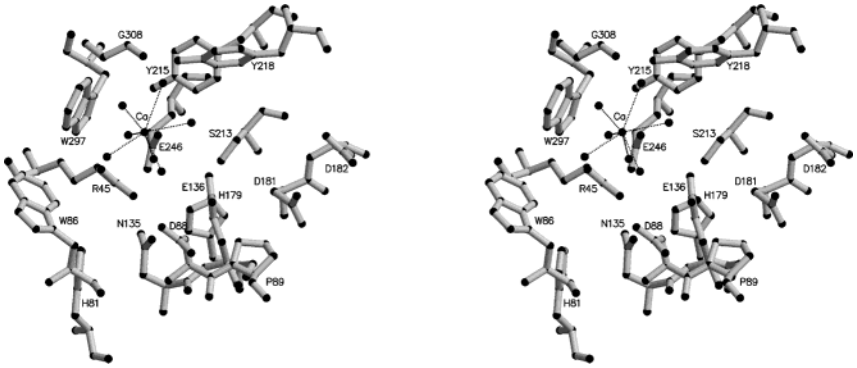

FIGURE 2: Stereoview of the cleft surrounding the active site of AAGAL. The calcium ion coordinated to seven water molecules is shown. Figure prepared with MOLSCRIPT (*73*).

conformations for which ϕ_H,ψ_H maps were computed. Each of the two torsion angles was varied by 20°, giving rise to a total of $18 \times 18 \times 36 = 11\,664$ minimized conformations. The 36 maps were combined so that the conformation with the lowest energy at each point was kept in the final map. The lowest-energy conformation in the final map was used as a starting conformation in the subsequent docking simulations.

AUTODOCK (version 2.4) was used for the docking simulations (*33*−*35*). The AAGAL293 structure was stripped of water molecules and hydrogen atoms; partial charges were added using X-PLOR (*24*). Energy grids were calculated for a cubic box with dimension 30 Å centered on the active site. The docking simulation was performed in two steps following described procedures (*36*, *37*), whereby the structures from an initial search are clustered with a tolerance of 1 Å RMSD relative to the lowest-energy result that defines the cluster. The global minimum structure along with the lowest energy structures of relevant clusters were then redocked and clustered using the same criteria. With this method a larger conformational space is searched in the first step, and only the relevant low-energy structures are optimized in the second.

The galactosyl−enzyme and galactobiosyl−enzyme intermediates were constructed by linking an α-D-galactose or α-D-galactobiose to the catalytic nucleophile (Glu246). Two initial MM3(92)-optimized alternative conformations for the resulting linkage were used for the galactosyl−enzyme intermediate. These in combination with the set of five global and local low energy minima issued from the conformational study of methyl β-galactobioside were optimized in MM3-(92), to give a set of 10 different initial conformers for the galactobiosyl−enzyme intermediate. The positions of the Cα and Cβ atoms of the nucleophile were kept fixed while the remaining side chain atoms along with the galactose or galactobiose were docked on a grid calculated on the protein without the side chain (including Cα) of the nucleophilic Glu246, by a technique that we refer to as anchored docking. After two rounds of anchored docking and cluster analysis, the resulting structures clustered all around one conformation in the −1 subsite for the galactosyl−enzyme intermediate. For the galactobiosyl−enzyme intermediate, the results were consistent with the galactosyl−enzyme intermediate for subsite −1, with some variation in conformations obtained for subsite −2.

A $^4C_1-^1S_3$ methyl β-galactobioside was constructed by adopting the inter-ring dihedral angles assigned from the $^4C_1-^4C_1$ conformational analysis. This $^4C_1-^1S_3$ methyl β-galactobioside was positioned in the −2 and −1 subsites so that the position of the galactose in subsite −2 overlapped with the best conformation found during the anchored docking procedure. To explore the possible location of the putative +2 subsite, a galactobioside in $^4C_1-^4C_1$ conformation was placed with the nonreducing end galactose in the +1 subsite. The two methyl-β-galactobiosides were docked on the intact enzyme grid, subjected to cluster analysis, and redocked as described above.

RESULTS AND DISCUSSION

Overall Structure. The three-dimensional structure of β-1,4-galactanase from *Aspergillus aculeatus* was determined by multiple isomorphous replacement using data collected to 2.3 Å resolution at 293 K (AAGAL293). Later the model was refined against 1.8 Å data collected from a crystal cryo-cooled at 100 K (AAGAL100). The final model of both structures comprises all 334 amino acid residues. The statistics for the final models are listed in Table 2. Like other enzymes belonging to clan GH-A, AAGAL folds into a $(\beta/\alpha)_8$ barrel as depicted in Figure 1. In addition to the core comprising the eight alternating parallel β-strands and α-helices, AAGAL contains one additional α-helix and two short 3_{10} helices (five residues each) in the βα−loop regions. A third short 3_{10} helix is located in a αβ-loop prior to β-strand 6. The disulfide bridge between Cys253 and Cys311 located after β-strands 7 and 8, respectively, connects the two strands and could contribute to the overall stability of the $(\beta/\alpha)_8$-barrel.

AAGAL293 and AAGAL100 are very similar (RMSD 0.25 Å for 334 Cα-atoms) with only Trp108 showing a different rotamer conformation. Mass spectroscopy showed a higher molecular mass than the one calculated for the protein (*17*), and the sequence contains a potential N-glycosylation site at Asn112. However, neither in AAGAL293 nor in the higher-resolution AAGAL100 structure was there any indication of O- or N-linked carbohydrate in the difference electron density. The AAGAL100 structure revealed a calcium ion from the mother liquor in the active site. The calcium ion has seven water molecules as ligands (Figure 2), and its shortest distances to the enzyme are 4.7 and 4.8 Å to the aromatic ring of Trp297 and the hydroxyl group of Tyr218, respectively. The *B*-factors for the calcium ion (43.8 Å^2) and coordinated water molecules (average 39.1 Å^2) are significantly higher than those of the surrounding protein residues (average 20.8 Å^2 within a 8 Å radius). The calcium ion was not identifiable with certainty in the AAGAL293 structure, presumably because of lower resolution and higher temperature. With a pI of 2.85 (*16*), the protein is negatively charged at the crystallization pH (7.5), and calcium ions at the surface can balance this charge, which might explain why divalent metal ions are necessary for crystallization.

The Active Site of the β-1,4-Galactanase. The catalytic residues of AAGAL have been identified as Glu136 at the end of β strand 4 and Glu246 at the end of β strand 7 from sequence similarity with the galactanase of *P. cellulosa*, whose catalytic machinery and retaining mechanism have been characterized previously (*15*). The distance between the two catalytic residues is 4.8 Å (AAGAL100), in agreement with the distance observed in other retaining glycoside hydrolases (*13*, *14*). Figure 3a illustrates a close up of the active site. The residues surrounding the active site of AAGAL create a cleft with two aromatic residues forming the steepest side of the cleft (Tyr215 and Tyr218), while polar and charged residues mark the other side and bottom of the cleft (Figure 2). Trp297 functions as a saddle point located at the top of the cleft with Trp86 and Gly308 in uphill directions.

A hydrogen bond from Arg45 to Glu246 (Figure 3a), the catalytic nucleophile, ensures that this residue remains deprotonated even at relatively low pH values (AAGAL has a pH optimum of 4−4.5 (*16*)). The other oxygen atom of the carboxylate group is hydrogen bonded to Tyr215. In contrast, the other catalytic glutamic acid residue must be

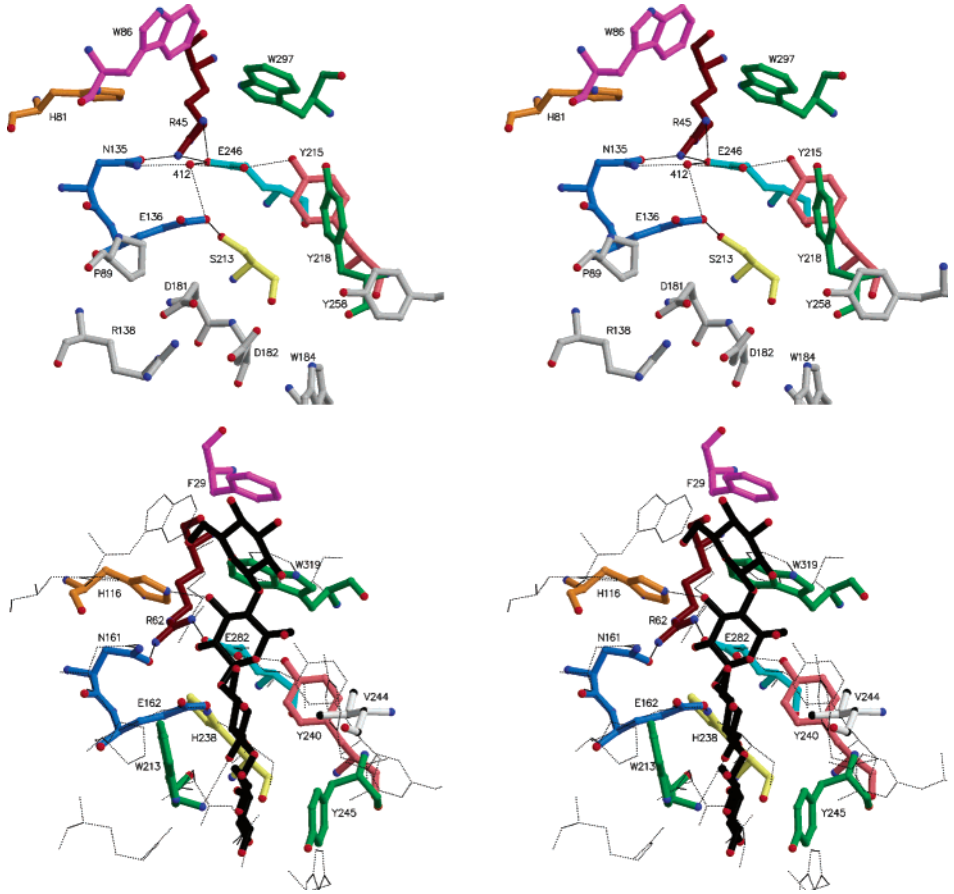

FIGURE 3: Stereoviews of the active site in the family 53 β-1,4-galactanase AAGAL (top) and family 5 endoglucanase from *A. cellulolyticus* with bound cellotetraose (bottom) [38, PDB code 1ECE] superposed on the AAGAL structure (thin line). The water bridging the catalytic carboxylates and selected hydrogen bonds are shown. The catalytic acid/base and nucleophile are marked with blue and cyan colors respectively, and the aromatic platforms shown in green for subsites −1, +1, and +2 and magenta for subsite −2. Figure prepared with MOLSCRIPT (*73*).

in an environment where it can function both as an acid and a base. The catalytic acid/base Glu136 makes hydrogen bonds to Ser213 in the bottom of the cleft and to water412 that connects the two catalytic glutamic acid residues (Figure 3a). The water is positioned so it can readily donate a proton to the deprotonated Glu136 and subsequently make a nucleophilic attack on the covalent glycosyl−enzyme intermediate.

Galactanase Residues Corresponding to Functionally Conserved Residues in Clan GH-A. Eight key residues have been identified previously to be functionally conserved (*38*) or have functional counterparts (*39*) in Clan GH-A and are also found in AAGAL. Figure 3 shows the location of these residues in the active sites of AAGAL and the family 5 *A. cellulolyticus* endoglucanase (PDB code 1ECE (*38*)) for comparison. Reference structures for the other families in clan GH-A used here have PDB codes 2MYR (*40*) (family 1), 1BHG (*41*) (family 2), 1EXP (*42*) (family 10), 1GHR (*43*) (family 17), and 1J9Y (*44*) (family 26). The eight residues are either involved directly in catalysis, or they interact with the catalytic residues and/or the substrate at subsite −1.

Glu136 and Glu246, the catalytic residues, are almost strictly conserved in clan GH-A, with only one enzyme using an ascorbate cofactor instead of a glutamate acid−base (*45*). Like in the other clan GH-A structures, the catalytic acid/base in AAGAL (Glu136) interacts with a hydrogen bond donor (Ser213) at the end of β-strand 6, which can be an asparagine (family 1, 2, and 17), an aspartate (family 26), a glutamine (family 10), and a tyrosine or histidine (family 5).

An aromatic residue at the end of β-strand 8 (Trp297) is absolutely conserved in the clan members, where it is involved in substrate binding at subsite −1. Unlike in most other structures from clan GH-A, Trp297 in AAGAL does not participate in a cis-peptide bond.

His81 is located at the end of β-strand 3 and is almost completely conserved in families 1, 5 (missing in subfamily 8 *T. reesei* mannanase (*46*, *47*), 10, and 26. As shown by crystallographic studies (*38*, *40*, *42*, *48*−*60*) or suggested by mutagenesis (*44*), this histidine makes a hydrogen bond to the C3 hydroxyl from the glycosyl moiety at subsite −1. Though the histidine position shows some variation between the different families, this does not seem to be correlated to substrate specificity, as the side chain position in family 5 *T. fusca* mannanase is more like the one in family 5 endoglucanases than in the family 26 mannanase.

Asn135, preceding the catalytic acid−base, is conserved in all clan GH-A families except family 26 mannanase, where the corresponding residue is a histidine. This Asn (or His) appears to fulfill two functions, one in maintaining the orientation and protonation state of active site residues, as supported by site directed mutagenesis (*44*, *61*−*62*), and the other as a hydrogen bond partner to the C2 hydroxyl of the

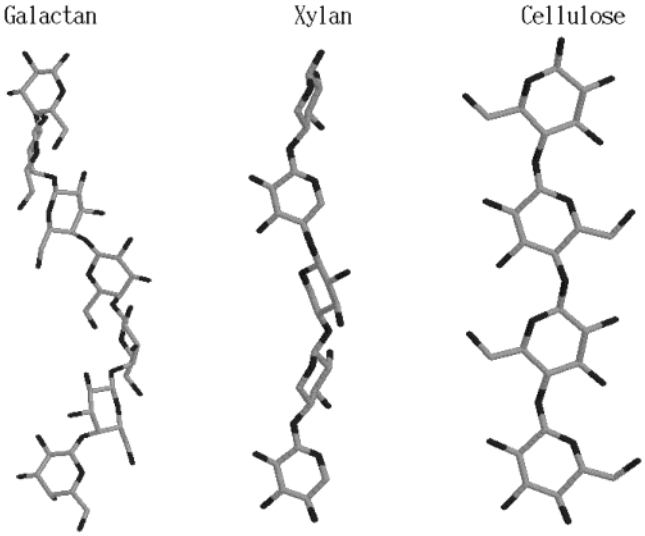

FIGURE 4: Three-dimensional models for galactan (this work, consistent with ref *65*), xylan (*70*), and cellulose (*68*). Figure prepared with RASMOL (*74*).

substrate at subsite −1, as shown by crystallographic studies (*38*, *40*, *42*, *48*−*53*, *55*−*60*).

Arg45 at the end of β-strand 2 is likely to maintain the catalytic nucleophile deprotonated. It is conserved in families 1, 2, 5, and 17, while in families 10 and 26, this role is taken over by a histidine on β-strand 6 and an arginine on β-strand 4, respectively. Recently the presence of the Arg at the end of β-strand 2 in the clan has been correlated with the presence of a type IV capping motif (*63*) which includes a conserved glycine at the C-terminus of the preceding helix (*47*), thought to maintain the orientation of the arginine. This motif is indeed also present in the galactanase where the conserved glycine is Gly40.

Crystallographic studies of 2-deoxy-2-fluoro-glycosyl-enzyme intermediates (*40*, *42*, *48*, *49*, *53*) in the clan show that the ring oxygen (O5) of the carbohydrate in the −1 subsite is within hydrogen bonding distance of a tyrosine residue (Tyr 215) at the end of β-strand 6 (a histidine in family 10). This side chain also serves as a hydrogen bond donor to the catalytic nucleophile, an interaction that must be important for keeping the substrate and nucleophile in suitable positions for the nucleophilic attack (*38*).

Modeling of Galactobiose in the Active Site. The detailed structural and mechanistic knowledge derived from the structures of covalent glycosyl−enzyme intermediates and noncovalent complexes in the GH-A families together with the location of conserved residues at the active site of AAGAL enabled us to identify putative substrate binding subsites −2 to +2, employing the standard subsite nomenclature (*64*). The energetically favored conformations of galactobiosides have dihedral angles consistent with those derived earlier from CD-spectroscopic measurements on galactans (*65*). If the favored conformations are extended to model a longer polysaccharide chain they suggest that galactan forms a right-handed helix with approximately 6 galactose residues/turn (Figure 4).

The trapped glycosyl−enzyme intermediates observed in the crystal structures of family 1 *S. alba* myrosinase (*40*), family 10 *C. fimi* xylanase Cex (*42*), and family 5 *B. agaradhaerens* endoglucanase (*48*) were used as a basis for the anchored docking into subsites −1 and −2 of a regular

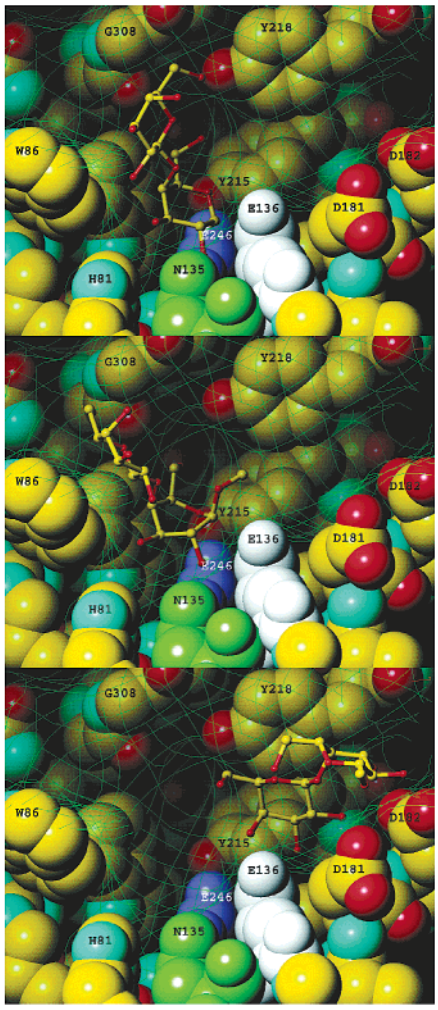

FIGURE 5: Galactobiosides docked into the active site of AAGAL293. The galactobiosides in their lowest energy conformations are shown as ball-and-stick models. Three conserved residues in clan GH-A are shown in green (Asn135), white (Glu136), and dark blue (Glu246). The surface of the protein is outlined by a green Connolly surface. Top: the anchored $^4C_1-^4C_1$ conformation of the α-galactobiosyl glutamate in subsite −2 and −1. Middle: the $^4C_1-^1S_3$ conformation of the methyl-β-galactobioside in subsite −2 and −1. Bottom: the $^4C_1-^4C_1$ conformation in subsite +1 and +2.

$^4C_1-^4C_1$ form of β-galactobioside covalently bound to Glu246 by an α-ester linkage. Furthermore, docking was conducted for methyl β-galactobioside in conformation $^4C_1-^1S_3$ into subsites −2 and −1 and in conformation $^4C_1-^4C_1$ into the subsites +1 and +2. Distortion at the −1 subsite was dictated by the difficulty of fitting a $^4C_1$ conformation due to steric clashes and by the crystallographic evidence of ring distortion in other retaining β-glycosidases (*48*, *66*). In the structure of the Michaelis−Menten complex with *B. agaradhaerens* endoglucanase (*48*) belonging to family 5 and thus Clan GH-A, the distortion is in the form of a skew-boat $^1S_3$ conformation, and this was chosen for modeling in AAGAL. Nonetheless, it cannot be excluded that galactose might be distorted to a different conformation than glucose.

Information about the carbohydrate interactions with subsites −2 and −1 can be obtained from the modeled covalently bound α-galactobiosyl glutamate in $^4C_1-^4C_1$ conformation and the modeled Michaelis complex of methyl-β-galactobioside in conformation $^4C_1-^1S_3$ (Figures 5a and 5b). A comparison of the two models reveals similar

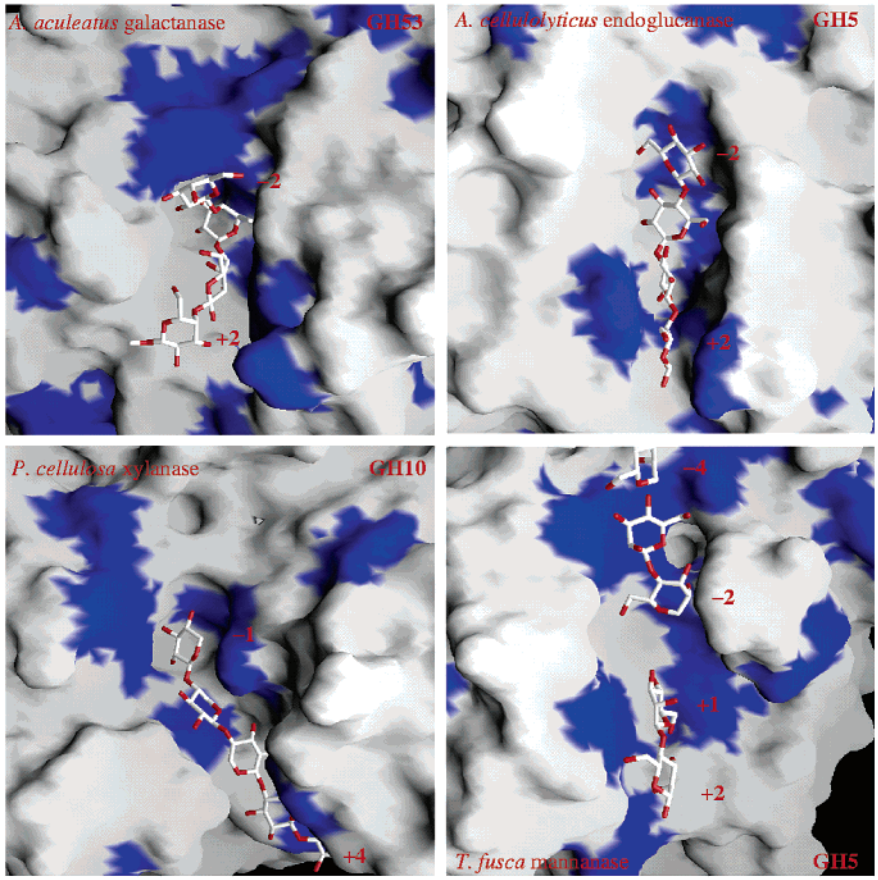

FIGURE 6: Surface representations of the substrate binding groove of selected members of Clan GH-A. Aromatic (Tyr, Phe, Trp) side chains are mapped to the surface in blue. The structures are shown in the same orientation. Molecular surfaces were calculated and displayed using default parameters in GRASP (*76*) after removal of water molecules and other ligands. The structures shown are AAGAL100 with the methylgalactobioside at subsites −2 and −1 and galactobiose at subsites +1 and +2 (this work); *A. cellulolyticus* endoglucanase (PDB code 1ECE) (*38*) in complex with cellotetraose; *T. fusca* mannanase (PDB code 3MAN) (*75*) in complex with mannotriose at subsites −4 to −2 with overlay of mannobiose at subsites +1 and +2 from the *T. reeseii* mannanase structure (PDB code 1QNR) (*46*); *P. cellulosa* xylanase active site mutant (*54*) in complex with xylopentaose (PDB code 1E5N). The shape of the substrate binding groove of AAGAL is consistent with galactan binding as a large helix. In the endoglucanase and mannanase, oligosaccharides bind each half of the substrate binding groove at either side of the catalytic residues in flat cellulose/mannan-like conformations, but mannanases necessitate a larger groove to accommodate the axial C2 hydroxyl groups. Xylopentaose bound to the xylanase has a xylan-like 3-fold helical conformation for the carbohydrate moieties occupying subsites −1 to +3.

positions of the atoms adjacent to the C4 atom of the galactose in subsite −1. The change in position of C1 from the skew-boat conformation to the covalently bound intermediate is 2.0 Å, close to the value of 1.7 Å observed in the corresponding structures of the *Bacillus agaradhaerens* endoglucanase (*48*). The aromatic residue at the end of β-strand 6 (Tyr215) is hydrogen bonded (2.8 Å) to O5 of the galactose at subsite −1, while Asn135 Nδ1 is within hydrogen bonding distance of the C2 hydroxyl, as predicted from their function in other members of clan GH-A where they are conserved. Conserved Trp297 provides the aromatic platform at subsite −1.

The distance of about 4 Å from O3 of the carbohydrate docked in subsite −1 to His81 is longer than a normal hydrogen bond. Interestingly, the natural substrates for the galactanase, β-1,4-galactan and arabinogalactan differ by the presence of an α-L-arabinofuranosyl side chain on the O3 atom of the galactan backbone of the latter. A small pocket is located between His81 and Arg45 and is occupied by water molecules in AAGAL293 and AAGAL100. Four of the residues delineating this pocket, Arg45, His81, Glu246, and Trp297, are conserved in the superfamily, but Trp86, Asp8, Arg47, and Asp79 are only strictly conserved in family 53. These types of residues are typically involved in protein−carbohydrate interactions, and this leads us to suggest that the pocket could interact with an arabinofuranosyl side chain bound to the galactosyl at subsite −1.

Subsite −2 could be localized to a saddle between Trp86 and Gly308, two residues which are totally and partially conserved in family 53, respectively. Trp86 might serve as aromatic platform for subsite −2. Although differences exist in the orientations of galactose units modeled at the −2 subsite, all of them enable the formation of hydrogen bonds between the O2 and O3 atoms in the carbohydrate at subsite −2 and Asp88, another residue that is conserved in family 53 but not in clan GH-A.

Docking of a $^4C_1{-}^4C_1$ conformer in the +1 and +2 subsites suggests that Tyr218 serves as the aromatic platform at subsite +1, while Asp182 interacts with galactose moieties at both +1 and +2 subsites (Figure 5c). Asp 181 and Arg138 might also interact with the moiety at subsite +2. Only the aromatic residue is conserved in family 53, but it is not unusual that conservation of residues involved in enzyme−polysaccharide interaction breaks down at distal subsites (*54*).

Structural Differences and Substrate Specificity within Clan GH-A. Despite the common catalytic machinery,

enzymes in clan GH-A display a large diversity in substrate specificity since the equatorially linked glycosidic substrates can be either pyranosides or furanosides, can differ in the chirality at the C2, C3, C4, and C5 carbon atoms, and be β-1,4- or β-1,3-linked. Family 1 and 2 enzymes are exo-acting, i.e., can only liberate a nonreducing terminal monosaccharide, and have an active site pocket instead of an open groove, as the -2 subsite is blocked by extra domains or extra loops (*67*).

The differences between the building blocks of different polysaccharides result in differences in their overall three-dimensional shape. For example, cellulose and mannan (β-1,4-linked polymers of glucose and mannose, respectively) are expected to form flat ribbon structures (*68*−*69*), while xylan (β-1,4-linked polymer of xylose) has a 3-fold left-handed helical structure (*70*) (Figure 4). The axial versus equatorial orientation of the glycosidic bond at C4 in β-1,4-galactan and cellulose, respectively, generates a more twisted structure for the former, as shown by our modeling and consistent with earlier chirooptical studies (*65*) suggesting a right-handed helix with approximately six galactose units per turn as the conformation of the polymer (Figure 4).

Shape complementarity is increasingly recognized as a determinant of specificity, as for example suggested for barley glucanases (*71*). Structural evidence showing that bound polysaccharide conformations are similar to the unbound conformations is growing. For example, the crystal complexes of xylopentaose with a family 15 carbohydrate binding module (*72*) and an inactive mutant of *P. cellulosa* xylanase 10A (*54*) (Figure 6) show that the ligand adopts a conformation close to the 3-fold helix found in fiber diffraction studies. The shape of the substrate binding groove might therefore dictate the specificity by excluding binding of substrates with inappropriate shape, although in the case of catalytic domains the polysaccharide chain is then distorted at the site of hydrolysis.

β-1,4-Endoglucanases and mannanases have more linear binding grooves than either xylanases or AAGAL, reflecting the linear structure of cellulose and mannan. The grooves of mannanases are somewhat wider than in β-1,4-endoglucanases, presumably because the axial C2-hydroxyl needs to be accommodated. Because of the linear nature of the substrate, both β-1,4-endoglucanases and mannanases can sandwich the polysaccharide chain between parallel aromatic side chains on opposite sides of the substrate binding groove. This arrangement of aromatic side chains is neither found in AAGAL, nor in family 10 xylanases.

Both xylanases and AAGAL contain a twisted substrate-binding groove, but whereas the xylanase groove is shallow, the galactanase groove is deeper and wider. This is consistent with its complementarity to a helix larger than that found in xylan, which is in tune with our modeling results and previous experimental work.

ACKNOWLEDGMENT

We are grateful to our collaborating partners at Novozymes A/S, Bagsværd (Denmark), who have provided the enzyme and purification procedure. We thank MAX-lab, Lund University, for the synchrotron beamtime at BL711 and Jens-Christian Navarro Poulsen and Flemming Hansen for excellent technical assistance.

REFERENCES

1. Carpita, N. C., and Gibeaut, D. M. (1993) *Plant J. 3*, 1−30
2. O'Neill, M., Albersheim, P., and Darvill, A. G. (1990) *The Pectic Polysaccharides of Primary Cell Walls*, pp 415−422, Academic Press Limited, London.
3. Lau, J. M., McNeil, M., Darvill, A. G., and Albersheim, P. (1985) *Carbohydr. Res. 137*, 111−125.
4. Bacic, A., Harris, P. J., and Stone, B. (1988) *The Biochemistry of Plants*, pp 297−371, Academic Press, New York.
5. Heldt-Hansen, H. P., Kofod, L. V., Budolfsen, G., Nielsen, P. M., Hüttel, S. and Bladt, T. (1996) In *Progress in Biotechnology Volume 14*: *Pectins and Pectinases*, pp 463−474, Elsevier Science B. V., New York.
6. Petersen, T. N., Kauppinen, S., and Larsen, S. (1997) *Structure 5*, 533−544.
7. Mølgaard, A., Kauppinen, S., and Larsen, S. (2000) *Struct. Fold. Des. 8*, 373−383.
8. Sørensen, S. O., Pauly, M., Bush, M., Skjot, M., McCann, M. C., Borkhardt, B., and Ulvskov, P. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 7639−7644.
9. Coutinho, P. M., and Henrissat, B. (1999) *Recent Advances in Carbohydrate Bioengineering*, pp 3−12, The Royal Society of Chemistry, London (http://afmb.cnrs-mrs.fr/CAZY).
10. 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.
11. Jenkins, J., Lo Leggio, L., Harris, G., and Pickersgill, R. (1995) *FEBS Lett. 362*, 281−285.
12. Henrissat, B., and Davies, G. J. (1997) *Curr. Opin. Struct. Biol. 7*, 637−644.
13. Davies, G., and Henrissat, B. (1995) *Structure 3*, 853−859.
14. McCarter, J. D., and Withers, S. G. (1994) *Curr. Opin. Struct. Biol. 4*, 885−892.
15. Braithwaite, K. L., Barna, T., Spurway, T. D., Charnock, S. J., Black, G. W., Hughes, N., Lakey, J. H., Virden, R., Hazlewood, G. P., Henrissat, B., and Gilbert, H. J. (1997) *Biochemistry 36*, 15489−15500.
16. Christgau, S., Sandal, T., Kofod, L. V., and Dalbøge, H. (1995) *Curr. Genet. 27*, 135−141.
17. Ryttersgaard, C., Poulsen, J.-C. N., Christgau, S., Sandal, T., Dalbøge, H., and Larsen, S. (1999) *Acta Crystallogr., Sect. D 55*, 929−930.
18. Gewirth, D. (1994) *The HKL Manual: An Oscillation Data Processing Suite for Macromolecular Crystallography*, Yale University, New Haven, CT.
19. Collaborative Computational Project Number 4. (1994) *Acta Crystallogr., Sect. D 50*, 760−763.
20. de La Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol. 276*, 472−494.
21. Abrahams, J. P., and Leslie, A. G. W. (1996) *Acta Crystallogr., Sect. D 52*, 30−42.
22. Kleywegt, G. J., and Jones, T. A. (1996) *Acta Crystallogr., Sect. D 52*, 826−828.
23. Roussel, A., and Cambillau, C. (1992). *TURBO*−*FRODO*, Biographics and AFMB (Architecture et Fonction des Macromolécules Biologiques), Marseille, France.
24. Brünger, A. T. (1992). X-PLOR version 3.1: A system for X-ray crystallography and NMR. Yale University Press, New Haven, CT.
25. Brünger, A. T. (1992) *Nature 355*, 472−474.
26. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr., Sect. D 54*, 905−921.
27. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
28. Allinger, N. L., Yuh, Y. H., and Lii, J. H. (1989) *J. Am. Chem. Soc. 111*, 8551−8566.
29. Lii, J. H., and Allinger, N. L. (1989) *J. Am. Chem. Soc. 111*, 8566−8575.
30. Allinger, N. L., Rahman, M., and Lii, J. H. (1990) *J. Am. Chem. Soc. 112*, 8293−8307.
31. Dowd, M. K., French, A. D., and Reilly, P. J. (1994) *Carbohydr. Res. 264*, 1−19.
32. Marchessault, R. H., and Pérez, S. (1979) *Biopolymers 18*, 2369−2374.
33. Goodsell, D. S., and Olson, A. J. (1990) *Proteins 8*, 195−202.

34. Goodsell, D. S., Lauble, H., Stout, C. D., and Olson, A. J. (1993) *Proteins 17*, 1−10.
35. Goodsell, D. S., Morris, G. M., and Olson, A. J. (1996) *J. Mol. Recognit. 9*,1−5.
36. Coutinho, P. M., Dowd, M. K., and Reilly, P. J. (1997) *Proteins 27*, 235−248.
37. Coutinho, P. M., Dowd, M. K., and Reilly, P. J. (1997) *Proteins 28*, 162−173.
38. Sakon, J., Adney, W. S., Himmel, M. E., Thomas, S. R., and Karplus, P. A. (1996) *Biochemistry 35*, 10648−10660.
39. Durand, P., Lehn, P., Callebaut, I., Fabrega, S., Henrissat, B., and Mornon, J. P. (1997) *Glycobiology 7*, 277−284.
40. Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., and Henrissat, B. (1997) *Structure 5*, 663−675.
41. Jain, S., Drendel, W. B., Chen, Z., Mathews, F. S., Sly, W. S., and Grubb, J. H. (1996) *Nat. Struct. Biol. 3*, 375−380.
42. White, A., Tull, D., Johns, K. L., Withers, S. G., and Rose, D. R. (1996) *Nat. Struct. Biol. 3*, 149−154.
43. Varghese, J. N., Garret, T. P. J., Colman, P. M., Chen, L., Høj, P. B., and Fincher, G. B. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2785−2789.
44. Hogg, D., Woo, E. J., Bolam, D. N., McKie, V. A., Gilbert, H. J., Pickersgill, R. W. (2001) *J. Biol. Chem. 276*, 31186−31192.
45. Burmeister, W. P., Cottaz, S., Rollin, P., Vasella, A., Henrissat, B. (2000) *J. Biol. Chem. 275*, 39385−39393.
46. Sabini, E., Schubert, H., Murshudov, G., Wilson, K. S., Siika-Aho, M. and Penttilä, M. (2000) *Acta Crystallogr., Sect. D 56*, 3−13.
47. Lo Leggio, L. and Larsen, S. *FEBS Lett. 523*, 103−108.
48. Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schülein, M., and Withers, S. G. (1998) *Biochemistry 37*, 11707−11713.
49. Varrot, A., Schülein, M., and Davies, G. J. (2000) *J. Mol. Biol. 297*, 819−828.
50. Dominguez, R., Souchon, H., Lascombe, M.-B., and Alzari, P. M. (1996) *J. Mol. Biol. 257*, 1041−1051.
51. Cutfield, S. M., Davies, G. J., Murshudov, G., Anderson, B. F., Moody, P. C. E., Sullivan, P. A., and Cutfield, J. F. (1999) *J. Mol. Biol. 294*, 771−783.
52. Varrot, A., Schülein, M., Pipelier, M., Vasella, A., and Davies, G. J. (1999) *J. Am. Chem. Soc. 121*, 2621−2622.
53. Notenboom, V., Birsan, C., Warren, R. A. J., Withers, S. G., and Rose, D. R. (1998) *Biochemistry 37*, 4751−4758.
54. Lo Leggio, L., Jenkins, J., Harris, G. W., and Pickersgill, R. W. (2000). X-ray crystallographic study of xylopentaose binding to *Pseudomonas fluorescens* xylanase A, *Proteins 41*, 362−373.
55. Notenboom, V., Birsan, C., Nitz, M., Rose, D. R., Warren, R. A. J., and Withers, S. G. (1998) *Nat. Struct. Biol. 5*, 812−818.
56. Schmidt, A., Gübitz, G. M., and Kratky, C. (1999) *Biochemistry 38*, 2403−2412.
57. Ducros., V., Charnock, S., Derewenda, U., Derewenda, Z., Dauter, Z., Dupont, C., Shareck, F., Morosoli, R., Kluepfel, D., and Davies, G. (2000) *J. Biol. Chem. 275*, 23020−23026.
58. Notenboom, V., Williams, S. J., Hoos, R., Withers, S. G., and Rose, D. R. (2000) *Biochemistry 39*, 11553−11563.
59. Fujimoto, Z., Kuno, A., Kaneko, S., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2002) *J. Mol. Biol. 316*, 65−78.
60. Lo Leggio, L., Kalogiannis, S., Eckert, K. Teixeira, S. C. M., Bhat, M. K., Andrei, C., Pickersgill, R. W., and Larsen, S. (2001) *FEBS Lett. 509*, 303−308.
61. Charnock, S. J., Lakey, J. H., Virden, R., Hughes, N., Sinnott, M. L., Hazlewood, G. P., Pickersgill, R. and Gilbert, H. J. (1997) *J. Biol. Chem. 272*, 2942−2951.
62. Roberge, M., Dupont, C., Morosoli, R., Shareck, F., and Kluepfel, D. (1997) *Protein Eng. 10*, 399−403.
63. Aurora, R., and Rose, G. D. (1998) *Protein Sci. 7*, 21−38.
64. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) *Biochem. J. 321*, 557−559.
65. Duda, C. A., Stevens, E. S., and Reid, J. S. G. (1991) *Macromolecules 24*, 431−435.
66. Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M., and Davies, G. J. (1996) *Biochemistry 35*, 15280−15287.
67. Juers, D. H., Huber, R. E., and Matthews, B. W. (1999) *Protein Sci. 8*, 122−136.
68. Kroon-Batenburg, L. M. J., and Kroon J. (1995) *Carbohydr. Eur. 12*, 15−19.
69. Dorset, D. L., and McCourt, M. P. (1993) *J. Struct. Biol. 111*, 118−124.
70. Atkins, E. D. T. (1992) in *Xylans and Xylanases* (Visser, J., Ed.) pp 349−360, Elsevier, Amsterdam.
71. Hrmova, M., and Fincher, G. B. (2001) *Plant Mol. Biol. 47*, 73−91.
72. Szabó, L., Jamal, S., Xie, H., Charnock, S. J., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2001) *J. Biol. Chem. 276*, 49061−49065.
73. Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946−950.
74. Sayle, R. A., and Milner-White, E. J. (1995) *TIBS 20*, 374−376.
75. Hilge, M., Gloor, S. M., Rypniewski, W., Sauer, O., Heightman, T. D., Zimmermann, W., Winterhalter, K., and Piontek, K. (1998) *Structure 6*, 1433−1444.
76. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins 11*, 281−296.

BI026238C