BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 6193–6205 6193
DOI: 10.1021/bi1006139

Signature Active Site Architectures Illuminate the Molecular Basis for Ligand Specificity in Family 35 Carbohydrate Binding Module[†,‡]

Márcia A. S. Correia,[§,○] D. Wade Abbott,[∥,○] Tracey M. Gloster,[⊥,○,●] Vânia O. Fernandes,[§] José A. M. Prates,[§] Cedric Montanier,[@] Claire Dumon,[@] Michael P. Williamson,[#] Richard B. Tunnicliffe,[#] Ziyuan Liu,[@] James E. Flint,[@] Gideon J. Davies,[⊥] Bernard Henrissat,[▽] Pedro M. Coutinho,[▽] Carlos M. G. A. Fontes,*[,§] and Harry J. Gilbert*[,∥]

[§]*CIISA, Faculdade de Medicina Veterinária, Pólo Universitário do Alto da Ajuda, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal,* [∥]*The Complex Carbohydrate Research Center, The University of Georgia, 315 Riverbend Road, Athens, Georgia 30602,* [⊥]*York Structural Biology Laboratory, Department of Chemistry, The University of York, York, YO10 5YW, U.K.,* [@]*Institute of Cellular and Molecular Biosciences, The Medical School, Newcastle University, Newcastle upon Tyne, NE2 4HH, U.K.,* [#]*Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield, S10 2TN, U.K., and* [▽]*Architecture et Fonction des Macromolécules Biologiques, UMR6098, CNRS, Universités Aix-Marseille I & II, Case 932, 163 Avenue de Luminy, 13288 Marseille cedex 9, France* [○]*These authors contributed equally to this work.* [●]*Current address: Department of Chemistry, Simon Fraser University, 8888 University Dr., Burnaby, V5A 1S6, BC, Canada.*

Received April 21, 2010; Revised Manuscript Received May 18, 2010

ABSTRACT: The deconstruction of the plant cell wall is an important biological process that is attracting considerable industrial interest, particularly in the bioenergy sector. Enzymes that attack the plant cell wall generally contain one or more noncatalytic carbohydrate binding modules (CBMs) that play an important targeting function. While CBMs that bind to the backbones of plant structural polysaccharides have been widely described, modules that recognize components of the vast array of decorations displayed on these polymers have been relatively unexplored. Here we show that a family 35 CBM member (CBM35), designated *Ct*CBM35-Gal, binds to α-D-galactose (Gal) and, within the context of the plant cell wall, targets the α-1,6-Gal residues of galactomannan but not the β-D-Gal residues in xyloglucan. The crystal structure of *Ct*CBM35-Gal reveals a canonical β-sandwich fold. Site-directed mutagenesis studies showed that the ligand is accommodated within the loops that connect the two β-sheets. Although the ligand binding site of the CBM displays significant structural similarity with calcium-dependent CBM35s that target uronic acids, subtle differences in the conformation of conserved residues in the ligand binding site lead to the loss of metal binding and uronate recognition. A model is proposed in which the orientation of the pair of aromatic residues that interact with the two faces of the Gal pyranose ring plays a pivotal role in orientating the axial O4 atom of the ligand toward Asn140, which is invariant in CBM35. The ligand recognition site of *exo*-CBM35s (CBM35-Gal and the uronic acid binding CBM35s) appears to overlap with that of CBM35-Man, which binds to the internal regions of mannan, a β-polymer of mannose. Using site-directed mutagenesis, we show that although there is conservation of several functional residues within the binding sites of *endo*- and *exo*-CBM35s, the *endo*-CBM does not utilize Asn113 (equivalent to Asn140 in CBM35-Gal) in mannan binding, despite the importance of the equivalent residue in ligand recognition across the CBM35 and CBM6 landscape. The data presented in this report are placed within a wider phylogenetic context for the CBM35 family.

Quantitatively, the plant cell contains the largest reservoir of organic carbon in the biosphere and, as such, represents an environmentally sustainable industrial substrate that is becoming increasingly important (*1*, *2*). The plant cell wall is a complex macromolecular structure consisting, primarily, of an extensive repertoire of interlocking polysaccharides (*3*). The chemical and physical complexity creates a barrier to rapid biological degradation. Therefore, to enable efficient catalysis, microbial enzymes that attack plant cell walls generally display elaborate molecular architectures comprised of catalytic modules and noncatalytic carbohydrate binding modules [CBMs[1] (reviewed in ref *4*)]. CBMs direct the appended catalytic modules to their target substrate within the wall. Thus, there is generally a close relationship between the substrate specificity of the enzyme and the ligand recognized by the CBM (*5*−*8*). It is likely that CBMs potentiate the activity of enzymes that attack the plant cell wall by proximity effects (*9*), rather than through a disruptive mechanism (*10*), although compelling evidence of noncatalytic CBM-mediated dissociation of crystalline chitin has been reported (*11*).

[†]This work was supported by Grant PTDC/BIAPRO/69732/2006 from the Fundação para a Ciência e a Tecnologia and the BBSRC and by the individual fellowship SFRH/BD/16731/2004 (to M.A.S.C.). T.M.G. is a Sir Henry Wellcome postdoctoral fellow, and G.J.D. holds a Royal Society Wolfson Research Merit Award.

[‡]The structure of *Ct*CBM35-Gal has been added to the Protein Data Bank as entry 2WZ8.

*To whom correspondence should be addressed. H.J.G.: e-mail, h.j.gilbert@newcastle.ac.uk; telephone, +1 706-583-6995; fax, +1 706-542-4412. C.M.G.A.F.: e-mail, cafontes@fmv.utl.pt; telephone, +351 21-3652800; fax, +351 21-3652810.

[1]Abbreviations: CBM, carbohydrate binding module; Gal, galactose; Man, mannose; GH, glycoside hydrolase family; rmsd, root-mean-square deviation.

Published on Web 05/24/2010

CBMs have been grouped into, currently, 59 sequence-based families in the CAZy database (*12*). In some of these families, particularly those that target crystalline structures, ligand specificity is invariant (*13*). By contrast, in CBM families that target single glycan chains, carbohydrate recognition can be highly variable. Such diversity in specificity is particularly evident in CBM families 4 (CBM4) (*5*, *14*), 6 (CBM6) (*15*−*18*), 32 (CBM32) (specificity reviewed in ref *19*), and 35 (CBM35) (*20*, *21*). Indeed, in CBM6, not only are different glycans targeted but the mode of binding (internal regions and nonreducing and reducing termini) and the location of the ligand binding sites vary (*15*, *17*). In general, CBMs display a β-sandwich fold and ligand recognition occurs either on the concave surface presented by one of the β-sheets (site 2 or B) or in the loops (site 1 or A) that connect the two β-sheets (see ref *4* for a review). For CBMs that recognize mono- or disaccharides, calcium can play a direct role in ligand binding (*21*, *22*).

Recent studies have revealed a significant expansion of the group of carbohydrates recognized by CBMs, exemplified by CBM35. Previously, a *Cellvibrio japonicus* mannanase, *Cj*Man5C, was shown to house a module, *Cj*CBM35-Man, which binds to the internal regions of β-mannan (*20*, *23*). NMR titrations suggested that the ligand binding site of *Cj*CBM35-Man partially overlaps with binding site 1 or A, evident in other CBM35 proteins and proteins from the related family, CBM6 (*23*, *24*). It was apparent, however, that *Cj*CBM35-Man underwent significant conformational changes upon mannooligosaccharide binding. Thus, it was difficult to identify the residues that interacted directly with ligand, and those for which their environments were influenced by the observed conformational changes. Recently, a cohort of CBM35s that bind, through a calcium-dependent mechanism, to uronic acids were identified (*21*). These CBM35s display significant sequence similarity with a module (designated *Ct*CBM35-Gal) present in a *Clostridium thermocellum* enzyme that is a component of the bacterium's plant cell wall degrading protein complex termed the cellulosome (*25*).

Here we show that *Ct*CBM35-Gal binds to D-galactose (Gal) and is able to target the terminal α-Gal residues in galactomannan but not the β-Gal epitopes in xyloglucan. The crystal structure of *Ct*CBM35-Gal in conjunction with site-directed mutagenesis shows that the ligand binding site of the protein and that of the uronic acid-specific CBM35s are highly conserved; however, subtle changes in amino acid chemistry and conformation within the binding site have radical impacts on CBM35 specificity. In addition, we also have explored the structure−function relationship of conserved residues within the binding site of *Cj*CBM35-Man. These biochemical studies, in combination with a comprehensive phylogenetic analysis, highlight the distant relationship between CBM6s and CBM35s and have generated a model useful for predicting the specificities of uncharacterized CBM35s.

MATERIALS AND METHODS

Cloning, Expression, and Purification of CtCBM35-Gal and CjCBM35-Man. The open reading frame encoding *Ct*CBM35-Gal (nucleotides 1678−2112 of Cthe_2137) was amplified from *C. thermocellum* YS genomic DNA by polymerase chain reaction (PCR), using appropriate primers. For the purposes of this study and structure deposition, the numbering of the amino acids in *Ct*CBM35-Gal is based on the recombinant protein. Therefore, there is a labeling discrepancy of 556 between the construct and native protein (i.e., Trp40 = Trp596). The resulting PCR product was cloned into pNZY28 (NZYTech Ltd.) and subcloned into NheI- and XhoI-restricted pET28a to generate p*Ct*CBM35Gal. The recombinant plasmid was sequenced to ensure that no mutations had occurred during the polymerase chain reaction that used the thermostable DNA polymerase NZYPremium (NZYTech Ltd.). *Ct*CBM35-Gal contains an N-terminal His_6 tag. To produce *Ct*CBM35-Gal, *Escherichia coli* BL21 DE3 (Novagen) cells, harboring pCBM35Gal, were cultured in Luria-Bertani broth containing kanamycin (50 μg/mL) at 37 °C to midexponential phase (A_{600} of 0.6), at which point isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and the cultures were incubated for a further 16 h at 16 °C. The cells were harvested by centrifugation, and His-tagged recombinant protein was purified from cell-free extracts by immobilized metal ion affinity chromatography (IMAC) using a cobalt-based Talon column deploying the standard methodology (*9*). For crystallization, *Ct*CBM35-Gal was further purified by size exclusion chromatography. Following IMAC, the protein was buffer exchanged into 50 mM Hepes-HCl buffer (pH 7.5) containing 200 mM NaCl (buffer A) and then subjected to gel filtration using a HiLoad 16/60 Superdex 75 column (GE Healthcare); the protein was eluted at 1 mL/min in buffer A. Purified *Ct*CBM35-Gal was concentrated using an Amicon 10 kDa molecular mass centrifugal concentrator and washed three times into 2 mM $CaCl_2$. The cloning of *Cj*CBM35-Man and its purification were described previously (*20*, *23*). All purified proteins were electrophoretically pure as judged by SDS−PAGE.

Site-Directed Mutagenesis. Site-directed mutagenesis was conducted by employing a PCR-based QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions, using p*Ct*CBM35-Gal as the template and appropriate primers.

Ligand Binding Studies. The capacity of *Ct*CBM35-Gal and *Cj*CBM35-Man to bind to a variety of soluble plant structural polysaccharides was evaluated by affinity gel electrophoresis (AGE). Continuous native polyacrylamide gels that consisted of 7.5% (w/v) acrylamide in 25 mM Tris/250 mM glycine buffer (pH 8.3) were created. 0.1% polysaccharide prior to polymerization was added to the gel. Approximately 5 μg of target proteins and BSA (as a noninteracting negative control) were loaded on the gels and subjected to electrophoresis at 10 mA/gel for 2 h at room temperature. Proteins were visualized when they were stained with Coomassie Blue. Isothermal titration calorimetry (ITC) of *Ct*CBM35-Gal was conducted at 25 °C using a MicroCal Omega titration calorimeter. Titrations were conducted in 50 mM Na-Hepes buffer (pH 7.5) containing 5 mM $CaCl_2$. During a titration, the protein sample (50−150 μM), stirred at 400 rpm in a 1.3586 mL reaction cell, was injected with 25 successive 10 μL aliquots of a 5 mg/mL polysaccharide or 5 mM oligosaccharide solution at 200 s intervals. Raw binding data were corrected for the heat of dilution of both protein and ligand. Integrated heat effects were analyzed by nonlinear regression using a single-site binding model (MicroCal ORIGIN, version 2.9), yielding K_a and $\Delta H°$ values. Other thermodynamic parameters were derived from the equation $-RT \ln K_a = \Delta G° = \Delta H° - T\Delta S°$. NMR spectroscopy was used in conjunction with mannooligosaccharide titrations to determine the affinity of variants of *Cj*CBM35-Man for mannohexaose and mannopentaose. The proteins were concentrated to 700 mM in 50 mM sodium phosphate buffer (pH 7.0) containing 10% (v/v) ^2H_2O.

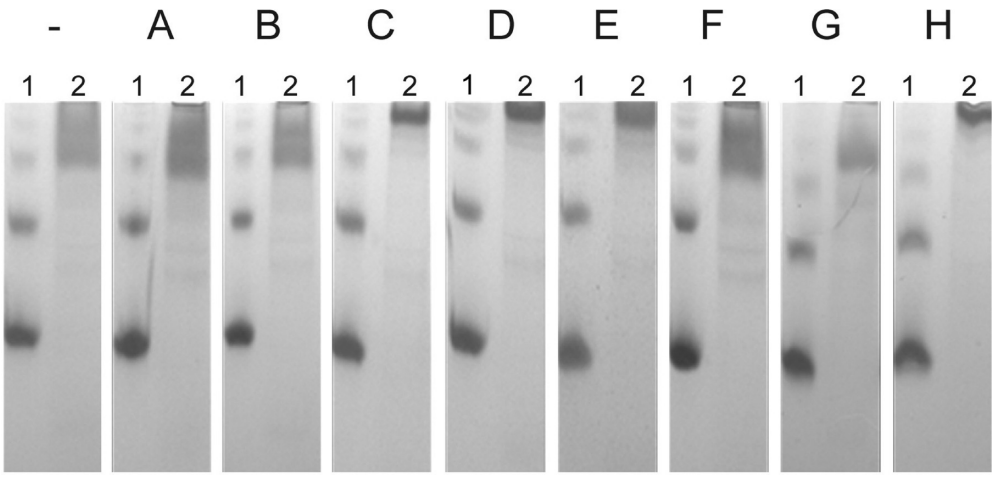

FIGURE 1: Examples of AGE of *Ct*CBM35-Gal against soluble polysaccharides. *Ct*CBM35-Gal in lane 2 was electrophoresed on nondenaturating polyacrylamide gels containing no ligand (−) or 1 mg/mL barley β-glucan (A), lupin galactan (B), carob galactomannan (Gal:Man, 21:79) (C), guar galactomannan (Gal:Man, 38:40) (D), galactomannan (Sigma G-0753) (E), or rye arabinoxylan (F). To evaluate the effect of calcium in ligand recognition, we performed AGE by replacing the divalent ion with EDTA in gels prepared without ligand (G) or guar galactomannan (Gal:Man, 38:40) (H). BSA was used as a nonpolysaccharide binding control (lane 1).

NMR spectra were recorded on a Bruker DRX 500 spectrometer, and the ^{1}H chemical shifts were referenced to an internal standard of 3-trimethylsilyl-2,2,3,3-d_4-propionate at 0.00 ppm. Titrations were acquired at 37 °C with a resolution of 8192 complex points over a sweep width of 12500 Hz. All data were processed with a Gaussian or sine-bell window convolution with FELIX version 2.30 (Felix NMR, Inc., San Diego, CA). The binding of the *Cj*CBM35-Man variants for mannopentaose and mannohexaose was assessed by following the movement of the NHε signal of Trp109 with the addition of increasing concentrations of ligand. For all titrations, dissociation constants were calculated by fitting to the shift changes (*26*). For analysis of chemical shift changes in HSQC spectra, a weighted shift change was calculated as $[(\delta H)^2 + (\delta N/5.7)^2]^{1/2}$.

Crystallization and Structure Solution of CtCBM35-Gal. Crystals of native *Ct*CBM35-Gal were grown by vapor-phase diffusion using the hanging drop method with equal volumes (1 μL) of protein (10 mg/mL in 2 mM $CaCl_2$) and reservoir solution (0.2 M magnesium acetate and 20% PEG 3350). Crystals, which grew over a period of 4−5 days, were stabilized via addition of 1 μL cryoprotectant solutions containing 30% (v/v) glycerol in the crystallization buffer and flash-frozen in liquid nitrogen. Data for a single crystal were collected at 100 K on beamline ID-29 at the European Synchrotron Radiation Facility (Grenoble, France). Data were processed and scaled with HKL2000 (*27*), and all other computing used the CCP4 suite of programs (*28*) unless otherwise stated. Phasing was done using molecular replacement methods; MrBUMP (*29*) was used to test a number of programs and search models. The only successful solution was obtained using MOLREP (*30*) methods for alignment to the target, the MOLREP (*30*) program to find the structural solution, and Protein Data Bank (PDB) entry 2W1W as the search model. The output from MrBUMP, which had automatically been run through REFMAC (*31*), was entered as the starting model in ARP/wARP (*32*) which was used to trace the majority of the protein chain. Subsequent refinement involved alternate rounds of model building in COOT (*33*) and likelihood-based refinement in REFMAC (*31*). The coordinates and structure factors have been deposited in the PDB as entry 2WZ8.

Ligand Model Building. A PDB model of D-galactopyranosyl-α-1,6-D-mannopyranoside was constructed using the online biomolecule building program GLYCAM (*34*). The galactosyl residue was modeled into the binding site of *Ct*CBM35-Gal using the ring atoms of the Chi-CBM35−glucuronic acid complex (PDB entry 2VZR) as a superimposition scaffold in COOT (*33*). The galactosylmannoside ligand was exported from GLYCAM in its lowest-energy state that contained a torsion angle of −178° across the α-1,6-glycosidic linkage. In this conformation, the axial O4 atom of Gal and the equatorial O3 atom of Man are positioned farthest apart (10.3 Å). Small rotations of the disaccharide were introduced to position the carbohydrate favorably (3.3−4.1 Å) within the aromatic sandwich for "stacking" interactions with Trp40 and Trp108. The position of O6 was not altered to compensate for the potential formation of hydrogen bonds with Arg86. The values in the model for distance (4.1 Å) and geometry (83°) may be different in the complex as O6 is free to rotate around its axis, providing a more favorable configuration for interaction.

CBM Family 35 Phylogenetic Analysis. A total of 139 full-length enzyme gene sequences containing CBM35s were extracted from the CAZy database (http://www.cazy.org) (*12*). Boundary sites were determined using in-house systematic manual and semiautomated modular annotation tools (*12*). Edited CBM35 sequences were subsequently aligned with Muscle version 3.52 (http://www.drive5.com/muscle/) (*35*) to generate a distance matrix that was derived with Blosum62 substitution parameters. Secator (*36*) was then used to delineate four subfamilies. The complete phylogenetic tree is represented in wheel format generated with the Dendroscope viewer (*37*).

RESULTS AND DISCUSSION

Ligand Specificity of CtCBM35-Gal. Recent studies have shown that a cohort of four structurally related family 35 CBMs bind to uronic acids; Pel-CBM35 and Rhe-CBM35 recognize Δ4,5-anhydrogalacturonic acid, while Xyl-CBM and Chi-CBM35 display affinity for both D-glucuronic acid and Δ4,5-anhydrogalacturonic acid (*21*). A CBM35 that displays a significant level of sequence identity (29%) to these uronate-specific modules is *Ct*CBM35-Gal. This module is a component of a *C. thermocellum* cellulosomal protein (ABN53341) that, in addition to *Ct*CBM35-Gal, contains a second CBM35, a family 39 glycoside hydrolase

Table 1: Affinity Gel Electrophoresis of *Ct*CBM35-Gal

polysaccharide	binding[a]	polysaccharide	binding[a]
wheat arabinoxylan	−	xyloglucan	−
birchwood xylan	−	glucomannan	−
rye arabinoxylan	−	lupin galactan	−
oat spelt xylan	−	potato galactan	−
hydroxyethylcellulose	−	pectic lupin galactan	−
barley β-glucan	−	pectic potato galactan	−
lichenan	−	sugar beet arabinan	−
laminarin	−	debranched α-1,5-arabinan	−
carob galactomannan (Gal:Man, 21:79)	++	rhamnogalacturonan	−
guar galactomannan (Gal:Man, 38:40)	+++	potato rhamnogalacturonan I	−
galactomannan (Sigma G-0753)	++		

[a]Qualitative binding is indicated as tight (+++), moderate (++), and insignificant (−).

(GH39), and a type I dockerin, which mediates integration into the multienzyme complex (*38*). To explore the ligand specificity of *Ct*CBM35-Gal, the protein was produced in *E. coli* and purified to electrophoretic homogeneity by IMAC.

Affinity gel electrophoresis (Figure 1 and Table 1) showed that *Ct*CBM35-Gal bound to galactomannans but did not recognize the other major polysaccharides present in the plant cell wall. The observation that the protein does not bind to β-mannan (not shown), a homopolymer of β-1,4-Man, suggests that the α-1,6-Gal side chain, present in galactomannan, is a component of the ligand recognized by *Ct*CBM35-Gal. To explore the specificity of *Ct*CBM35-Gal in more detail, ITC was used to quantify the binding of the protein to oligosaccharides and polysaccharides. Examples of the titrations are shown in Figure 2, and the full data set is reported in Table 2. In addition to binding galactomannan, *Ct*CBM35-Gal also exhibited affinity for singly and doubly substituted oligosaccharides, derived from the polysaccharide, and the monosaccharide D-galactose (Gal). Indeed, the affinity of *Ct*CBM35-Gal for the monosaccharide is similar to that for the oligosaccharides, indicating that Gal is the major specificity determinant of galactomannan. The CBM also bound to other α-D-Gal-containing oligosaccharides such as raffinose (Figure 2). The inability of the module to bind arabinose indicates that O6 of Gal contributes to ligand recognition. The observation that *Ct*CBM35-Gal does not bind to xyloglucan, which contains terminal β-D-Gal residues, may suggest that the protein is able to distinguish between the two anomers of the hexose sugar. To test this hypothesis, we assessed the capacity of the CBM to bind to 4-nitrophenyl-β-D-Gal and 4-nitrophenyl-α-D-Gal. The data showed that the CBM only bound to the α anomer of the two aryl galactosides. The affinity of *Ct*CBM35-Gal for galactomannan was ~8-fold higher than for the oligosaccharides. Although tighter binding of CBMs to multivalent, compared to monovalent, ligands is normally associated with avidity effects, mediated through protein oligomerization (*39*, *40*), the difference in the affinities observed here is not sufficient to indicate such a mechanism. Rather, we propose that the conformation of galactomannan is more rigid than that of oligosaccharides, and thus, there is a smaller entropic penalty when the polysaccharide binds to *Ct*CBM35-Gal, which is consistent with the $T\Delta S$ values. Nevertheless, typical of carbohydrate−protein recognition, binding to all the ligands was driven by changes in enthalpy, while the reduction in entropy had a negative impact on overall affinity (*5*, *16*, *17*, *20*, *22*, *39*). To explore the possible role of metal ions in the function of *Ct*CBM35-Gal, we assessed the capacity of the protein to bind galactomannan in the presence of EDTA (Figure 1). Binding was observed in the presence of the chelating agent. These data indicate that calcium does not contribute to the binding of *Ct*CBM35-Gal to its carbohydrate ligands.

Structure of CtCBM35-Gal. The crystal structure of *Ct*CBM35-Gal was determined using molecular replacement and refined to a crystallographic *R* factor of 0.13 (R_{free} = 0.18) using data extending to 1.5 Å resolution. The *Ct*CBM35-Gal crystals belong to space group $P2_12_12_1$ with one molecule per asymmetric unit and the following cell dimensions: a = 40.8 Å, b = 52.9 Å, and c = 58.7 Å. Data statistics are listed in Table 3. In general, the electron density is of high quality, although residues 75−81 were highly disordered and the protein chain could not be built within this region. The final model contains 135 amino acid residues, 191 water molecules, and two metal ions that, on the basis of their coordination spheres, ligand compositions, and *B* factors, have been modeled as a calcium and a magnesium atom.

*Ct*CBM35-Gal adopts a β-sandwich fold in which the two β-sheets containing four and five antiparallel β-strands, respectively, are connected entirely by loops (Figure 3). The two β-sheets are twisted, and it could be argued that the protein displays an elongated β-barrel-like structure. The first β-sheet (β-sheet 1) includes β-strands β-1 (Ile8−Glu12), β-9 (Asn140−Ala147), β-4 (Gly60−Ala69), and β-7 (Thr113−Leu120), whereas the second β-sheet (β-sheet 2) consists of strands β-2 (Gly21−Ala24), β-3 (Tyr48−Tyr55), β-8 (Asn124−Tyr130), β-5 (Arg86−Val92), and β-6 (Lys98−Phe102). The hydrophobic core of the protein comprises four leucines, seven isoleucines, five phenylalanines, four valines, and one methionine. The calcium ion in *Ct*CBM35-Gal is positioned between the loops linking β-1 and β-2 and the loop that joins β-2 and β-3, while β-9 also contributes to the metal binding site. This cofactor is highly conserved in CBMs that display a β-sandwich fold and, in the case of Xyl-CBM4 from the thermophile *Rhodothermus marinus*, and likely other CBMs that display this fold, confers significant thermostability (*41*). This metal has pentagonal bipyramidal coordination and is completely dehydrated, with each ligand being donated by the protein: Glu12 Oε1, Glu14 Oε1 and Oε2, Asp142 Oε1, and the backbone carbonyls of Asp142, Ser33, and Gln36. The second metal, modeled as magnesium because of the ligand content, coordination number, and lack of anomalous signal, displays octahedral coordination, although it forms only one interaction with the protein (an oxygen from the carboxylate group of Asp114). The rest of the five coordination sites are occupied with water molecules.

Identification of the Ligand Binding Site of CtCBM35-Gal. The ligand binding site of CBMs that display a β-sandwich fold where known is either on the concave surface presented by

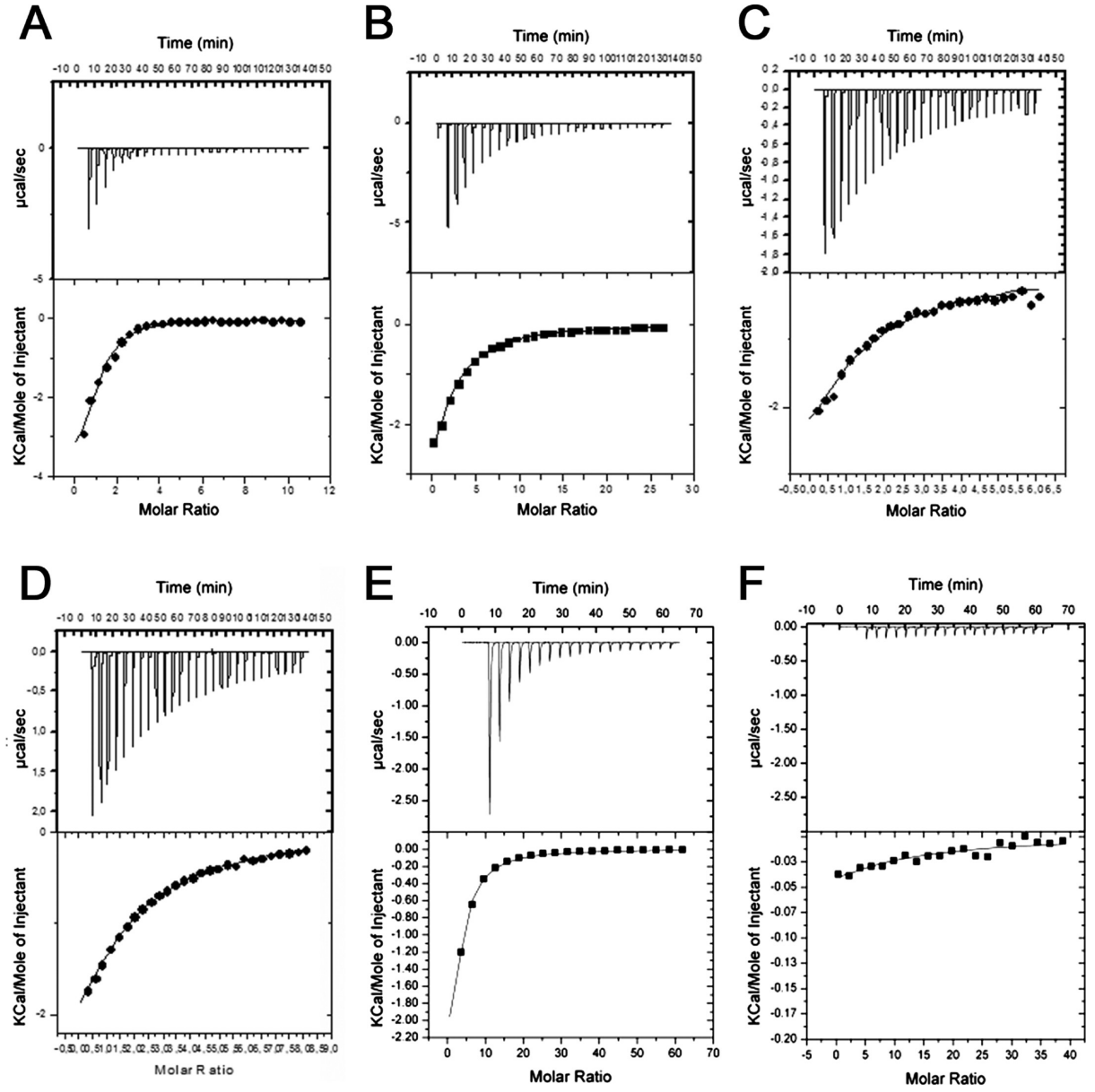

FIGURE 2: Representative ITC data of *Ct*CBM35-Gal titrated with carbohydrates The various panels represent different ligand−*Ct*CBM35-Gal reactions: (A) galactomannan, (B) raffinose, (C) 6^1-α-D-galactosyl-β-1,4-mannotriose, (D) 6^3-6^4-di-α-D-galactosyl-β-1,4-mannopentaose, (E) PNP-α-galactose, and (F) PNP-β-galactose. Ligands were titrated into *Ct*CBM35-Gal (100 μM) in the cell with continuous stirring. The top half of each panel shows the raw ITC heats; the bottom half of each panel shows the integrated peak areas fitted using a one-binding site model with MicroCal Origin.

Table 2: Binding of the Wild Type and Mutants of *Ct*CBM35-Gal to Galactose-Containing Ligands Determined by ITC

*Ct*CBM35-Gal	ligand	K_a ($\times 10^3$ M^{-1})	ΔG (kcal/mol)	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol)	N^a
wild type	galactomannan (carob)	22.7 ± 3.3	−4.15 ± 0.08	−5.41 ± 0.38	−1.26 ± 0.30	1.00 ± 0.07
wild type	6^4-di-α-D-galactosylmannopentaose	3.25 ± 0.18	−4.37 ± 0.03	−7.66 ± 1.15	−3.29 ± 1.12	1.03 ± 0.14
wild type	6^1-α-D-galactosylmannotriose	4.30 ± 1.1	−4.54 ± 0.15	−7.68 ± 2.09	−3.14 ± 1.94	0.92 ± 0.43
wild type	galactose	3.84 ± 0.60	−4.43 ± 0.04	−6.96 ± 1.99	−2.53 ± 1.95	0.99 ± 0.23
Q27A	galactomannan (carob)	binding[b]				
Y37F, Y37A, W40A, R86A, W108A, Y137A, N140A	galactomannan (carob)	nb[c]	–	–	–	–

[a]All interactions were fitted to a single-binding site model by nonlinear regression. For interactions between the protein and polysaccharide, *N* was fixed to 1. [b]Binding detected but too weak for accurate quantification. [c]No binding detected.

one of the β-sheets (site 2) (*5*) or in the extended loops that connect to two β-sheets (site 1) (*15*, *21*, *22*), although both sites are functional in *Cm*CBM6 (*17*). The elongated barrel-like structure of *Ct*CBM35-Gal does not present an obvious concave surface, suggesting that site 1 comprises the ligand binding site. Closer inspection of site 1 reveals the presence of two surface-exposed tryptophans poised ~8.6 Å apart in an appropriate orientation to sandwich a potential ligand. To explore the

Table 3: X-ray Data and Structure Quality Statistics for *Ct*CBM35-Gal

resolution of data (Å) (outer shell)	20−1.50 (1.55−1.50)
R_{merge} (outer shell)	0.046 (0.119)
mean $I/\sigma I$ (outer shell)	37.3 (12.1)
completeness (%) (outer shell)	97.1 (95.8)
multiplicity (outer shell)	6.9 (5.5)
R_{cryst}/R_{free}	0.13/0.18
rmsd for 1−2 bonds (Å)	0.015
rmsd for 1−3 bonds (deg)	1.46
rmsd for chiral volume (Å^3)	0.097
avg main chain B (Å^2)	14
avg side chain B (Å^2)	17
avg solvent B (Å^2)	30
PDB entry	2WZ8

functionality of this binding site, therefore, various polar and aromatic residues in the surface-exposed loop regions were substituted with alanine and the capacity of the CBM variants to recognize galactomannan was assessed by ITC. The data, presented in Table 2, show that all the mutations prevented ligand binding, demonstrating that site 1 comprises the ligand binding site.

Structural Similarity of CtCBM35-Gal and Other Proteins. The closest three-dimensional structural homologues to *Ct*CBM35-Gal are the cohort of CBM35s that bind to D-glucuronic acid and/or Δ-4,5-anhydrogalacturonic acid (5-keto-4-deoxyuronate) (*21*). The DaliLite version 3 (*42*) server revealed the following structural similarities: Chi-CBM35 (PDB entry 2VZP, Z_{value} = 19.3, rmsd = 1.6 Å^2 over 122 aligned C_α atoms), Rhe-CBM35 (PDB entry 2W1W, Z_{value} = 18.6, rmsd = 1.6 Å^2 over 123 aligned C_α atoms), Xyl-CBM35 (PDB entry 2W46, Z_{value} = 17.0, rmsd = 1.9 Å^2 over 122 aligned C_α atoms), and Pel-CBM35 (PDB entry 2W3J, Z_{value} = 16.9, rmsd = 1.9 Å^2 over 123 aligned C_α atoms). This high degree of structural similarity is somewhat surprising when one considers the different ligand specificities displayed by *Ct*CBM35-Gal and the four CBM35s that target uronic acids. Inspection of the ligand binding sites of Chi-CBM35 and *Ct*CBM35-Gal, however, provides insight into the structural basis for these differences in ligand recognition. A key element of uronic acid recognition is a protein-bound calcium that makes a polar contact with the carboxylic acid of the ligand and, in the case of GlcA, O4. Calcium recognition is conferred, in part, by a stretch of amino acids at the end of β-strand 4, consisting of an Asp/Asn-Tyr/Thr-X-Asn consensus sequence (Figure 3). The flanking residues Asp/Asn and Asn coordinate the metal directly through side chain oxygen atoms. The Tyr/Thr motif presents its peptidyl backbone carbonyl oxygen, regardless of functional group chemistry, and the X residue operates as a spacer to create appropriate spatial geometry for the C-terminal asparagine to be positioned for metal coordination.

While *Ct*CBM35-Gal displays some components of the calcium binding pocket of Chi-CBM35 in its structure, there are two key structural differences that explain its inability to utilize the metal ion directly in ligand recognition. First, the "Asp/Asn" residue in Chi-CBM35 (Asn33) is replaced with Gly39 in *Ct*CBM35-Gal. This amino acid substitution not only removes one of the calcium ligands (Oδ1 of Asn33 in Chi-CBM35) but also reduces the torsional constraints on the next residue, Trp40, allowing the side chain to rotate 180° relative to Tyr34, the equivalent residue in Chi-CBM35. This structural transition has two effects. First, the bulky side chain occupies the equivalent space utilized by the coordination pocket within the uronic acid binding CBM35s. Second, while the carbonyl of Tyr34 in Chi-

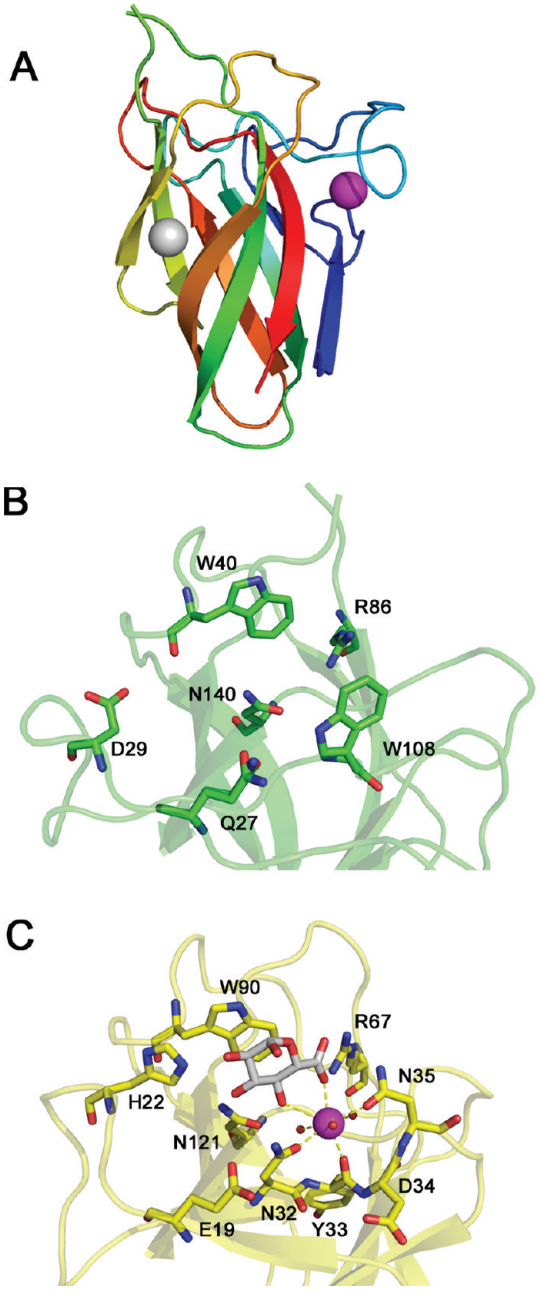

FIGURE 3: Three-dimensional crystallographic structure of *Ct*CBM35-Gal. (A) Cartoon representation of *Ct*CBM35-Gal. The color scheme is ramped from the N-terminus (blue) to the C-terminus (red). The structural calcium is shown as a magenta sphere and the bound magnesium as a silver sphere. Structural alignment of *Ct*CBM35-Gal (green) (B) and Chi-CBM35 (yellow) (C). Amino acids involved in binding uronic acids in Chi-CBM35 are labeled and aligned with binding site residues in *Ct*CBM35-Gal. The coordinated calcium in Chi-CBM35 is represented as a magenta sphere.

CBM35 can coordinate with the calcium, the backbone oxygen of Trp40 in *Ct*CBM35-Gal is angled in the opposite direction, toward the core of the protein, and is therefore unable to interact with a potential metal ion (see Figure 3). The second feature of *Ct*CBM35-Gal that leads to the loss of calcium binding is the insertion of two residues (Ile41 and Gly42) between the conserved Tyr/Thr and Asn residues (Asn43 in *Ct*CBM35-Gal and Asn36 in Chi-CBM35). The insertion of an additional residue results in the translation of Asn43 out of proximity to where the metal binding site would lie. Thus, despite the high degree of sequence identity, the only components of the calcium binding site that are

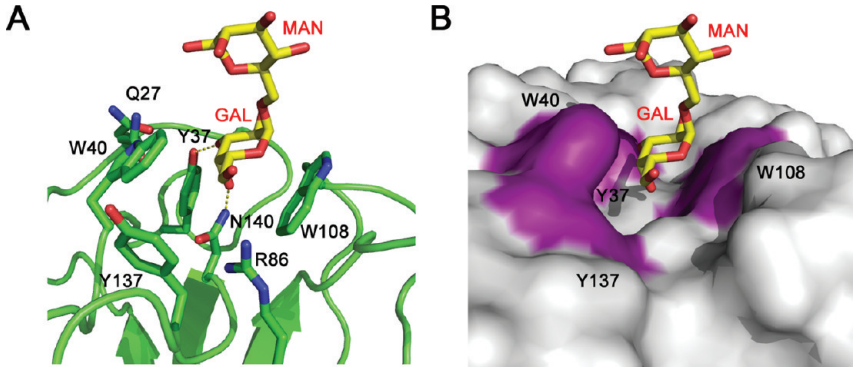

FIGURE 4: Model of the α-galactosyl 1,6-mannoside complex displayed in a cartoon (A) and solvent accessible surface (B) format. The putative hydrogen bonds selected on the basis of atomic distances (2.5−3.5 Å) and appropriate geometries are represented with yellow dashed lines.

structurally conserved in *Ct*CBM35-Gal and Chi-CBM35 are Asn140 and Asn122, respectively.

Despite differences in calcium coordination at the ligand binding site, several of the residues in Chi-CBM35, which make polar contacts with the uronic acids, are conserved in *Ct*CBM35-Gal (Asn122/Asn140 and Arg68/Arg86), while the hydrophobic platform provided by Trp108 in Chi-CBM35 is structurally equivalent to Trp92 in *Ct*CBM35-Gal (Figure 3). It would appear, therefore, that the inability of *Ct*CBM35-Gal to recognize uronic acids exists primarily because of the loss of calcium coordination in the ligand binding site. However, many of the remaining features of the ligand binding site of *Ct*CBM35-Gal, which are conserved in the uronate-specific CBM35s, are harnessed by the clostridial protein in Gal recognition, as discussed below.

Model of Gal Recognition by CtCBM35-Gal. The specificity determinants that distinguish the target ligand of *Ct*CBM35-Gal from other neutral sugars that interact with similar modules are the axial orientation of O4 and the α-anomer of the D-sugar. Attempts to cocrystallize *Ct*CBM35-Gal with target ligands were unsuccessful. Therefore, a model of an α-1,6-galactosyl mannoside complex was generated using GlcA from the Chi-CBM35 structure as a superimposition scaffold (Figure 4). Although this type of analysis is useful for identifying candidate amino acids within the binding site of the apo structure that interact with ligand, the results must be interpreted conservatively as the position of the galactosyl ligand may result in subtle rearrangements of amino acids following the binding event. Such conformational changes are rarely observed in CBMs, although using NMR, the binding site of *Cj*CBM35-Man was observed to be significantly remodeled following mannooligosaccharide binding (*20*, *23*). Selection for the nonreducing end of α-galactosyl residues appears to involve key interactions with the axial O4 atom and the δN atom of Asn140 at the binding site floor. The sugar is oriented primarily through hydrophobic interactions between Trp40 and Trp108, and the planar faces of the C3−C4−C5 and O5−C1−C2 regions of Gal in the chair conformer. The role of these two residues in directing the axial O4 atom of Gal toward Asn140, which is conserved across both the CBM35 and CBM6 landscapes, is intriguing. In several CBM6 and CBM35 proteins, the structurally equivalent Asn (Asn140 in *Ct*CBM35-Gal) makes polar contacts with the equatorial O3 and/or O4 atom of D-*xylo*-, D-*gluco*-, and L-*gal*-configured ligands (*15*−*18*, *21*). While the two aromatic residues that sandwich the sugar are opposite each other in the CBM6s (*15*−*17*), in *Ct*CBM35-Gal they are staggered, and the hydrophobic contact with the ligand may force Gal into an orientation that is optimal for the formation of a hydrogen bond between Asn140 and the axial O4 atom. In addition, potential hydrogen bonds between the Tyr37 hydroxyl group and O3 and between one of the amino groups of Arg86 and O6 of Gal would assist in orienting the axial O4 atom to make a polar contact with Asn140, while projecting the α-6-linked mannosyl substituent into solvent, to thus position the galactoside subunit such that it would be displayed on a mannan polymer. In this model, Asn140 and Arg86 play the same role as their structurally equivalent residues in Chi-CBM35, while Tyr137 seals the base of the binding pocket, which likely contributes to the selectivity for the terminal subunits of complex carbohydrates. This residue may also make weak apolar interactions with C6 of Gal, while also contributing to the conformation adopted by Trp40 through extensive hydrophobic interactions. The lack of binding of the Y137A mutant to Gal illustrates the importance of Tyr137 to the function of *Ct*CBM35-Gal. In this configuration, selectivity for α-linked galactosyl residues likely arises from the proximity of Trp108 to the anomeric hydroxyl group as a β-configured O1 glycosidic bond would clash with the surface of the protein (Figure 4). Attempts to rotate the ligand to accommodate a β-linkage result in the axial O4 hydroxyl encroaching upon the hydrophobic face of Trp108. When viewed as a solvent surface accessible model (Figure 4B), the core binding site outlets onto a planar face of CBM35. This structural feature would accommodate, but not interact with, the highly polymerized β-mannosyl backbone of galactomannan. Support for this model is provided by the observation that the individual substitution of all residues that are proposed to interact with the ligand results in a substantial reduction in affinity for Gal (Table 2).

Phylogenetic Analysis of CBM35s. CBM35s are found in a variety of modular enzymes active on the hemicellulosic and pectic components of plant cell walls (*20*, *21*). Consistent with the varied specificity displayed by their cognate catalytic modules, CBM35s have been shown to bind diverse ligands, including uronic acids (*21*), β-1,3-galactan (*43*), the backbone of β-1,4-mannan (*23*), and, in this report, the α-1,6-Gal side chains of galactomannan. It should be emphasized, however, that glucomannans, present in softwoods, are considerably more abundant in nature than galactomannan, which is restricted to specialized plant tissues. It is likely, therefore, that the natural ligand for *Ct*CBM35-Gal is galactoglucomannan and not galactomannan.

The CBM35s deploy diverse mechanisms of ligand recognition exemplified by calcium-dependent (*21*) and calcium-independent binding (*23*, *43*). Phylogenetic−functional analysis is described here to provide a platform for predicting the specificities of uncharacterized CBM35 targets, thereby helping to streamline

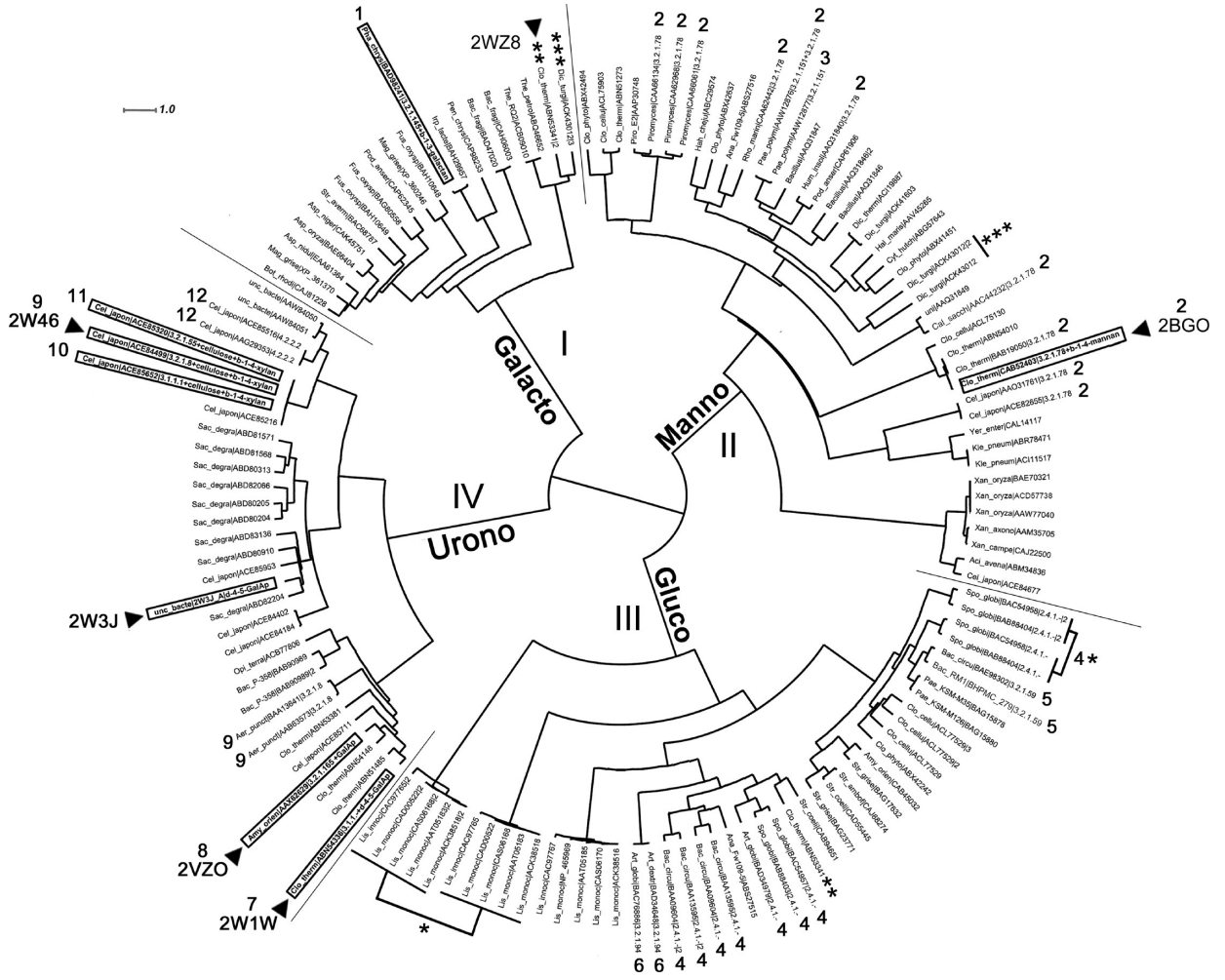

FIGURE 5: Phylogenetic wheel diagram of family 35 CBMs. The family partitions into four major subfamilies (I−IV). General ligand specificities for each CBM35 subfamily are labeled: *galacto*, *manno*, *gluco*, and *urono* configurations. CBM35s with known specificities are boxed, and determined three-dimensional structures are shown with a black triangle and labeled with their corresponding PDB entry. CBM35s appended to enzymes with characterized specificities are indicated with numbers ranging from 1 to 12. EC numbers correlate to the following enzyme activities: (1) 3.2.1.145, galactan β-1,3-galactosidase; (2) 3.2.1.78, mannan *endo*-β-1,4-mannosidase; (3) 3.2.1.151, xyloglucan-specific *endo*-β-1,4-glucanase; (4) 4.2.1, broad specificities, isomaltosyl- and glucosyltransferases predicted for CBM35s; (5) 3.2.1.59, glucan *endo*-α-1,3-glucosidase; (6) 3.2.1.94, glucan α-1,6-isomaltosidase; (7) 3.1.1, carboxyesterase, pectin acetylesterase predicted for ABN54336.1; (8) 3.2.1.165, *exo*-β-1,4-D-glucosaminidase; (9) 3.2.1.8, *endo*-β-1,4-xylanase; (10) 3.1.1.1, carboxyesterase; (11) 3.2.1.55, α-*N*-arabinofuranosidase; (12) 4.2.2.2, pectate lyase. Homogeneous clustering patterns of *S. globis* and *Listeria* enzymes are indicated with one asterisk; heterogeneous clustering of *C. thermocellum* ABN53341 CBM35s is indicated with two asterisks, and the blended pattern of *D. turgidum* ACK42012 is indicated with three asterisks.

the future analysis of new CBM35s of biological and industrial significance.

In total, 139 CBM35 primary structures were retrieved from the CAZy database (*12*) and analyzed using MUSCLE version 3.52 (*35*), a program used for large-scale multiple-sequence alignments. On the basis of protein sequence relatedness, family 35 partitions into four major subfamilies (I−IV). Functional predictions for each subfamily are based upon the distributions of CBM35s with known specificities, including carbohydrates with *galacto* (I), *manno* (II), *gluco* (III), and uronate (IV) configurations (Figure 5). The analysis reveals several enzymes that contain multiple copies of CBM35s that cluster into different subfamilies (Figure 5), suggesting that these modules display diverse specificities. For example, ABN53341, the enzyme encoding *Ct*CBM35-Gal, is a predicted xylosidase from *C. thermocellum* ATCC 27405. Besides *Ct*CBM35-Gal, which falls into the *galacto* cluster (subfamily I), the protein also contains a second CBM35 that is positioned within the *gluco* cluster [subfamily III]. This type of sequence pattern [previously termed heterogeneous clustering (*19*, *24*)] suggests a more complex mode of substrate targeting as the diverse combination of CBM specificities would be selective for regions of heterogeneity in carbohydrate structure. Heterogeneous clustering represents a significantly different binding mechanism from "homogeneous clustering" (*19*, *24*) in which multiple CBM35s, such as those appended to GH31 catalytic modules from *Listeria* sp. and *Sporosarcina globispora*, cluster as tandem pairs within the tree, thus displaying the most similarity to each other (*19*, *24*) (Figure 5). Homogeneous clustering is indicative of a composite CBM assembly tailored to bind a structurally similar multivalent ligand and enhance overall affinities of the modular enzyme for substrate regions through avidity effects (*39*, *40*). There is also an enzyme which possesses CBM35s that deploys a blended pattern of clustering (Figure 5). ACK43012, a potential mannanse from *Dictyoglomus turgidum* DSM 6724, contains three CBM35s, two modules, which fall into the *manno* cluster (subfamily II, homogeneous clustering), and one that falls into the *galacto* cluster (subfamily I, heterogeneous clustering). This observation indicates that the

Table 4: Interactions of *Cj*CBM35-Man with Mannosyl-Containing Oligo- and Polysaccharides

	relative affinity as determined by AGE		K_a ($\times 10^3$ M^{-1}) determined by NMR	
*Cj*CBM35-Man	galactomannan	glucomannan	mannohexaose	mannopentaose
wild type	1	1	11.1	10.0
Q32A	0.95	0.80	nd[b]	10.0
Y60A	nb[a]	nb[a]	0.33	0.33
K63A	nb[a]	nb[a]	nd[b]	0.014
D108A	0.081	0.06	1.7	0.67
W109A	nb[a]	nb[a]	nb[a]	nb[a]
Y111A	0.017	0.04	nd[b]	0.13
N113A	0.761	0.51	2.0	2.5

[a]No binding detected. [b]Not determined.

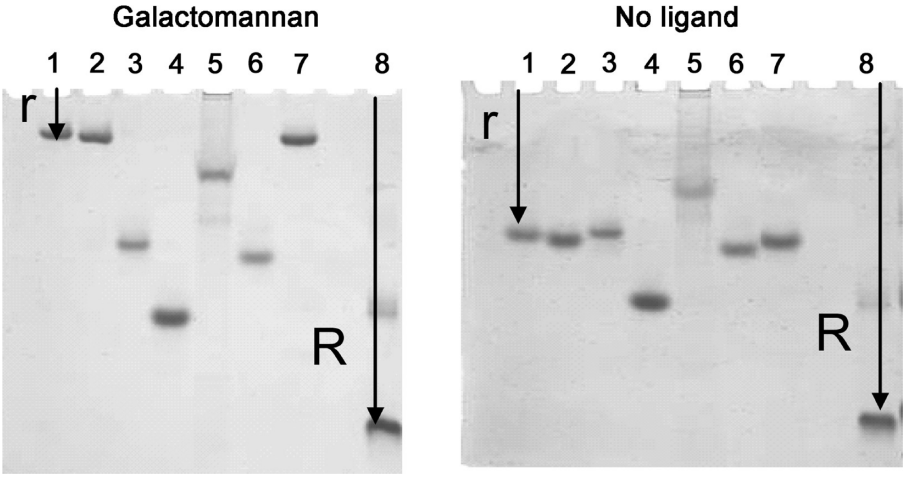

FIGURE 6: Affinity gel electrophoresis of the wild type and mutants of *Cj*CBM35-Man against 0.8% galactomannan. Lanes 1−8 contained wild type, Q32A, Y60A, K63A, D108A, W109A, N113A of *Cj*CBM35-Man, and BSA as a negative control, respectively. The distances migrated by the CBMs (*r*) relative to the reference noncarbohydrate binding protein (BSA) (*R*) were measured on the no ligand gel and on gels containing a range of different ligand concentrations. These data were used to determine the relative affinity of each mutant for galactomannan compared to that of the wild type.

CBMs of this enzyme likely target distinct carbohydrate structures and enhance binding affinity through avidity effects. Demonstration of these possibilites still awaits further investigation.

Insights into Mannosyl Recognition by CjCBM35-Man. One of the phylogenetic subfamilies of CBM35 targets β-1,4-mannan, and the NMR solution structure and binding properties of one of its members, *Cj*CBM35-Man, have been reported previously (*20*, *23*). NMR chemical shifts identified amino acids within the potential binding site of *Cj*CBM35-Man that were likely candidates in mannopentaose recognition (*23*). The environment of these residues, however, may have been influenced by conformational changes associated with ligand binding and not through direct interactions with the mannose polymer. To further characterize the underlying mechanism of ligand recognition by *Cj*CBM35-Man, site-directed mutagenesis was used to generate the following mutants of the protein: Q32A, Y60A, K63A, D108A, W109A, Y111A, and N113A (Table 4). The binding properties of these mutants were characterized by NMR shifts and affinity gel electrophoresis (Figure 6). The data (Table 4) showed that only Gln32 and Asn113 did not appear to play an important role in ligand recognition. Gln32 aligns with polar amino acids from other CBM35 structures that are involved in binding target ligands; however, in *Cj*CBM35-Man, this residue is distal to the modeled core binding site and most likely does not interact directly with the ligand. In contrast, the retention of mannan and mannohexaose recognition by the N113A mutant is surprising. This amino acid aligns with the highly conserved asparagines (equivalent to Asn140 in *Ct*CBM35-Gal) found within the core binding site of virtually all other CBM35s and in site 1 of the related CBM6 family (*15*−*18*, *21*, *24*). In known structures, as discussed above, the asparagine makes direct hydrogen bonds with O3 and/or O4 of the bound ligand and is a major component of the core binding site. Furthermore, the chemical shift of the asparagine changes significantly when the protein is bound to mannohexaose. It is hard to determine, therefore, why mutating this amino acid would have such a minimal effect on the binding affinity for galactomannan and mannooligosaccharides. Possibly, the contribution to binding energy by Asn113 may be compensated by multiple subsites present within this module. Indeed, favorable increases in affinity were observed for mannohexaose compared to mannopentaose, suggesting a minimum of six subsites. In contrast, each of the other mutations had noticeable effects on the binding properties of the protein. Replacing Tyr60 with alanine eliminated binding of gluco- and galactomannan in affinity gels and resulted in an ~30-fold loss of affinity for mannopentaose in solution. Mutation of amino acids within the *Cj*CBM35-Man C-terminus

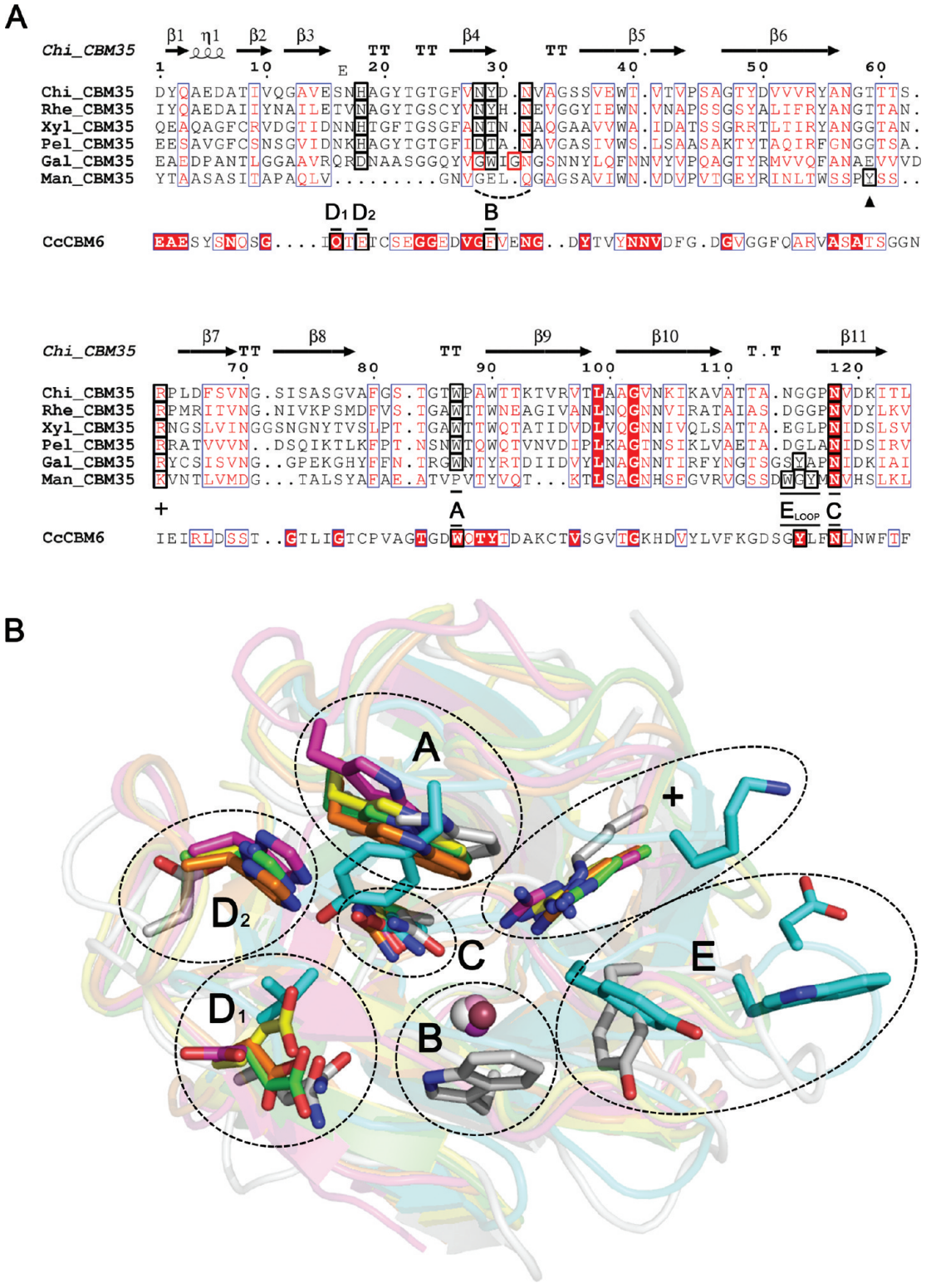

FIGURE 7: Structural conservation within the binding site of family 35 CBMs. (A) Primary structure alignments of CBM35s with known three-dimensional structures. Amino acids that operate as functional hot spots conserved within family 6 CBMs (A−E) and the basic (+) and metal binding sites (dashed line) unique to CBM35s are labeled below the sequences. *Cc*CBM6 is provided as a reference alignment for family 6 CBMs and is aligned with Gal-CBM35. The functional Tyr37 unique to *Cj*CBM35-Man is indicated with a black triangle. (B) Structural superimposition of CBM35s displayed in a cartoon format in which the amino acids involved in binding site formation are displayed as sticks and metal ions as spheres. Regions of functional residues are circled with dashed lines and labeled. Color scheme: green for Chi-CBM35, yellow for Rhe-CBM35, orange for Xyl-CBM35, purple for Pel-CBM35, silver for *Ct*CBM35-Gal, and cyan for *Cj*CBM35-Man.

(Asp108, Trp109, and Tyr111) and Lys63, which aligns with the functional arginine found within the uronate binding CBM35s, compromised ligand recognition and in some cases completely ablated binding affinity. While it is tempting to predict that Lys63 will play a role equivalent to that of the conserved arginine present in *Ct*CBM35-Gal and the uronate binding CBM35s, where it targets O6, the lysine adopts a conformation different from that of the conserved basic residue. Thus, in *Cj*CBM35-Man,

Lys63 extends into solvent away from the core binding site. The role of this amino acid in binding, confirmed by site-directed mutagenesis, may result from the aliphatic portion of the side chain helping to nestle against the hydrophobic planar face of the sugar rings. This observation underlines the biological significance of the location and chemistry, in addition to the functional plasticity, of the basic amino acids at this position.

Conserved Mechanisms of Ligand Selection within CBM35 and CBM6. Primary structure analysis and conservation of the overall fold suggest that CBM35 is distantly related to CBM6. CBM35s, however, lack key aromatic residues that contribute to ligand specificity and binding energy, and while two ligand binding sites have been identified in CBM6s, to date, a single binding site has been observed in CBM35 proteins (equivalent to site I in CBM6s) (*20*). Within CBM6, site I has been dissected into five key molecular regions (A−E) that contribute to affinity and specificity (*24*).

Comparison of the five characterized regions (A−E) in CBM6s with the three-dimensional structures of CBM35 demonstrates that there are indeed similarities between the two families (Figure 7). Region A (aromatic residues that stack against the two faces of a pyranose ring) is highly conserved, as *Cj*CBM35-Man is the only sequence without an aromatic residue in the primary structure aligning with residues from region A in other CBMs. Interestingly, the lack of this residue appears to be offset by an evolutionary convergence as a structurally equivalent aromatic residue, Tyr60, is provided from the loop connecting β-strands 6 and 7 (the aromatic residue that comprises site A is derived from the loops that join β-strands 8 and 9) (Figure 7). To accommodate this rearrangement in the β-sandwich scaffold, the trajectory of this loop in *Cj*CBM35-Man shifts ~3.8 Å from the analogous loop in Chi-CBM35. The presence of two different topologies to position an aromatic residue at this position underlines the structural plasticity of CBMs in carbohydrate recognition and the significance of region A in the binding mechanism. Region B displays variability within CBM35s, ranging from a single tryptophan in *Ct*CBM35-Gal, to a metal coordination site in the uronate binding CBMs, to being completely absent in *Cj*CBM35-Man. Such structural diversity suggests that this region is a functional hot spot that significantly contributes to ligand selectivity in this family. By contrast, region B in CBM6 displays less plasticity comprising a highly conserved aromatic residue. Indeed, the aromatic residues in regions A and B in CBM6 contribute to a conserved mechanism for ligand recognition in which the aromatic residues stack against the two planar faces of the bound sugar. In CBM35s, region B is more diverse and contributes directly to ligand specificity. Region C, comprising an asparagine at the base of the binding site, is invariant in both CBM6 and CBM35. Depending on the orientation of the ligand, the residue makes critical hydrogen bonds with O2, O3, or O4 of the central sugar. Region D partitions into two substructures defined as D_1 and D_2. Region D_1 contains a well-conserved collection of polar residues except for *Cj*CBM35-Man, which displays a valine. When compared to the proposed model for CBM6 binding site prediction (*24*), these CBM35 amino acid signatures suggest the polar amino acids contribute to a pocketlike structure and the mannan-specific CBM should display an open cleft, an observation that is confirmed by its three-dimensional structure and biochemical properties (*23*). In addition, region D is expanded in the three uronic acid binding CBM35s to include a histidine (defined as D_2 in Figure 7) that forms a hydrogen bond to O2 of the ligand. Loops are present in the sequence of region E in each CBM35 with a known structure; however, only *Ct*CBM35-Gal and *Cj*CBM35-Man display hallmark aromatic residues, hinting at a functional specialization in these regions. *Ct*CBM35-Gal contains a tyrosine residue (Tyr137) that is conserved within the binding site of CBM6 from *Clostridium cellulolyticum* (*Cc*CBM6). In both structures, these tyrosines block off the binding site cleft and contribute to specificity for the termini of complex carbohydrates (discussed above). The loop between β-strands 10 and 11, or the "E_{loop}", of *Cj*CBM35-Man displays a functional aspartate (Asp108) and two aromatic residues (Trp109 and Tyr111) that make a contribution to mannooligosaccharide binding. These amino acids significantly extend the wall of the binding cleft away from its core toward the concave β-sheet. This signature motif forms novel subsites that equip the CBM with the capacity to bind oligosaccharides and further strengthens the correlation between aromatic residues within the E_{loop} and ligand specificity.

Conclusions. This report provides evidence that CBMs can target the terminal α-D-Gal residues of complex polymers, expanding further the repertoire of carbohydrates recognized by CBMs. The likely biological targets for *Ct*CBM35-Gal are structural polysaccharides decorated with α-linked galactopyranose residues. While *Ct*CBM35-Gal was shown to bind to galactomannan in this study, the natural ligand for the protein is likely to be galactoglucomannan. The major hemicellulose in softwoods is galactoglucomannan, while galactomannan is restricted to highly specialized cell types in a restricted number of plants and is not abundant. *Ct*CBM35-Gal is a component of a multimodular enzyme directed, by its dockerin module, to the cellulosome, an enzyme complex that deconstructs the plant cell wall (reviewed in ref *25*). In this regard, *Ct*CBM35-Gal likely functions to localize the enzymes tethered within the cellulosome to regions of galactomannan (or other components of the plant cell wall that contain terminal α-D-Gal residues). The coordinated functions of CBMs within the context of a multienzyme assembly remain poorly understood, and exploring the mechanistic determinants of how individual modules specifically recognize novel carbohydrate stereochemistries, linkages, and decorations is a valuable exercise in unlocking the coordinated process of plant cell wall turnover. An intriguing feature of *Ct*CBM35-Gal is that, despite retaining many features of the ligand binding site of the cohort of uronate binding CBM35s, its specificity is very different, targeting a neutral α-D-hexose sugar in which O4 is axial. In conjunction with mutagenesis studies, this report reveals how subtle changes in protein structure, even to the extent that key residues are conserved but their orientation is modified through flanking glycine residues, confer radical changes in ligand specificity. Within a wider context, the structure of *Ct*CBM35-Gal, the related module *Cj*CBM35-Man, which targets the mannan backbone of galactomannan, and the uronate-specific CBM35s provide a framework for identifying generic specificity determinants that extend to the related CBM6 family of modules. Due to the discovery of novel specificities within closely related families of CBMs and in combination with the expansion of new CBMs identified through accelerated genomic sequencing efforts, there is an urgent need to deploy bioinformatic tools, based on known biochemical and structural information, for predicting the specificity of CBMs and streamlining the characterization of novel binding mechanisms. The detection of novel CBM specificities contributes to this iterative process and expands the available repertoire of targeting molecules that can be harnessed in the deconstruction of the plant cell wall. Ultimately, this will

contribute to the exploitation of an environmentally sustainable substrate for generation of molecules previously produced by the petrochemical industry.

REFERENCES

1. Himmel, M. E., and Bayer, E. A. (2009) Lignocellulose conversion to biofuels: Current challenges, global perspectives. *Curr. Opin. Biotechnol. 20*, 316–317.
2. Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., Frederick, W. J., Jr., Hallett, J. P., Leak, D. J., Liotta, C. L., Mielenz, J. R., Murphy, R., Templer, R., and Tschaplinski, T. (2006) The path forward for biofuels and biomaterials. *Science 311*, 484–489.
3. Brett, C. T., and Waldren, K. (1996) Physiology and Biochemistry of Plant Cell Walls. Topics in Plant Functional Biology, Vol. *1*, Chapman and Hall, London.
4. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochem. J. 382*, 769–781.
5. Boraston, A. B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D. R., Kilburn, D. G., and Davies, G. J. (2002) Differential oligosaccharide recognition by evolutionarily-related β-1,4 and β-1,3 glucan-binding modules. *J. Mol. Biol. 319*, 1143–1156.
6. 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) The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates β-1,4- and β-1,3−1,4-mixed linked glucans at a single binding site. *J. Biol. Chem. 279*, 34785–34793.
7. Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J., and Fontes, C. M. (2000) The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: Structure and biochemistry of the *Clostridium thermocellum* X6b domain. *Biochemistry 39*, 5013–5021.
8. Szabo, L., Jamal, S., Xie, H., Charnock, S. J., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2001) Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. *J. Biol. Chem. 276*, 49061–49065.
9. Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P., and Gilbert, H. J. (1998) *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem. J. 331*, 775–781.
10. Din, N., Gilkes, N. R., Tekant, B., Miller, R. C., Warren, A. J., and Kilburn, D. G. (1991) Non-hydrolytic disruption of cellulose fibers by the binding domain of a bacterial cellulase. *Bio/Technology 9*, 1096–1099.
11. Vaaje-Kolstad, G., Horn, S. J., van Aalten, D. M. F., Synstad, B., and Eijsink, V. G. H. (2005) The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *J. Biol. Chem. 280*, 28492–28497.
12. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res. 37*, D233–D238.
13. Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., and Teeri, T. T. (2003) The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proc. Natl. Acad. Sci. U.S.A. 100*, 484–489.
14. Abou Hachem, M., Nordberg Karlsson, E., Bartonek-Roxa, E., Raghothama, S., Simpson, P. J., Gilbert, H. J., Williamson, M. P., and Holst, O. (2000) Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: Cloning, expression and binding studies. *Biochem. J. 345*, 53–60.
15. Czjzek, M., Bolam, D. N., Mosbah, A., Allouch, J., Fontes, C. M., Ferreira, L. M., Bornet, O., Zamboni, V., Darbon, H., Smith, N. L., Black, G. W., Henrissat, B., and Gilbert, H. J. (2001) The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved. *J. Biol. Chem. 276*, 48580–48587.
16. Henshaw, J., Horne-Bitschy, A., van Bueren, A. L., Money, V. A., Bolam, D. N., Czjzek, M., Ekborg, N. A., Weiner, R. M., Hutcheson, S. W., Davies, G. J., Boraston, A. B., and Gilbert, H. J. (2006) Family 6 carbohydrate binding modules in β-agarases display exquisite selectivity for the non-reducing termini of agarose chains. *J. Biol. Chem. 281*, 17099–17107.
17. Henshaw, J. L., Bolam, D. N., Pires, V. M., Czjzek, M., Henrissat, B., Ferreira, L. M., Fontes, C. M., and Gilbert, H. J. (2004) The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. *J. Biol. Chem. 279*, 21552–21559.
18. van Bueren, A. L., Morland, C., Gilbert, H. J., and Boraston, A. B. (2005) Family 6 carbohydrate binding modules recognize the non-reducing end of β-1,3-linked glucans by presenting a unique ligand binding surface. *J. Biol. Chem. 280*, 530–537.
19. Abbott, D. W., Eirin-Lopez, J. M., and Boraston, A. B. (2008) Insight into ligand diversity and novel biological roles for family 32 carbohydrate-binding modules. *Mol. Biol. Evol. 25*, 155–167.
20. Bolam, D. N., Xie, H., Pell, G., Hogg, D., Galbraith, G., Henrissat, B., and Gilbert, H. J. (2004) X4 modules represent a new family of carbohydrate-binding modules that display novel properties. *J. Biol. Chem. 279*, 22953–22963.
21. Montanier, C., van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., Firbank, S. J., Lewis, R. J., Grondin, G. G., Ghinet, M. G., Gloster, T. M., Herve, C., Knox, J. P., Talbot, B. G., Turkenburg, J. P., Kerovuo, J., Brzezinski, R., Fontes, C. M., Davies, G. J., Boraston, A. B., and Gilbert, H. J. (2009) Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proc. Natl. Acad. Sci. U.S.A. 106*, 3065–3070.
22. Jamal-Talabani, S., Boraston, A. B., Turkenburg, J. P., Tarbouriech, N., Ducros, V. M., and Davies, G. J. (2004) Ab initio structure determination and functional characterization of CBM36: A new family of calcium-dependent carbohydrate binding modules. *Structure 12*, 1177–1187.
23. Tunnicliffe, R. B., Bolam, D. N., Pell, G., Gilbert, H. J., and Williamson, M. P. (2005) Structure of a mannan-specific family 35 carbohydrate-binding module: Evidence for significant conformational changes upon ligand binding. *J. Mol. Biol. 347*, 287–296.
24. Abbott, D. W., Ficko-Blean, E., van Bueren, A. L., Rogowski, A., Cartmell, A., Coutinho, P. M., Henrissat, B., Gilbert, H. J., and Boraston, A. B. (2009) Analysis of the structural and functional diversity of plant cell wall specific family 6 carbohydrate binding modules. *Biochemistry 48*, 10395–10404.
25. Bayer, E. A., Belaich, J. P., Shoham, Y., and Lamed, R. (2004) The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol. 58*, 521–554.
26. Charlton, A. J., Baxter, N. J., Khan, M. L., Moir, A. J., Haslam, E., Davies, A. P., and Williamson, M. P. (2002) Polyphenol/peptide binding and precipitation. *J. Agric. Food Chem. 50*, 1593–1601.
27. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307–326.
28. Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D50*, 760–763.
29. Keegan, R. M., and Winn, M. D. (2007) Automated search-model discovery and preparation for structure solution by molecular replacement. *Acta Crystallogr. D63*, 447–457.
30. Vagin, A., and Teplyakov, A. (1997) MOLREP: An automated program for molecular replacement. *J. Appl. Crystallogr. 30*, 1022–1025.
31. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D53*, 240–255.
32. Lamzin, V. S., and Wilson, K. S. (1993) Automated refinement of protein models. *Acta Crystallogr. D49*, 129–147.
33. Emsley, P., and Cowtan, K. (2004) *Coot*: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.
34. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem. 29*, 622−655.
35. Edgar, R. C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res. 32*, 1792–1797.
36. Wicker, N., Perrin, G. R., Thierry, J. C., and Poch, O. (2001) Secator: A program for inferring protein subfamilies from phylogenetic trees. *Mol. Biol. Evol. 18*, 1435–1441.
37. Huson, D. H., Richter, D. C., Rausch, C., Dezulian, T., Franz, M., and Rupp, R. (2007) Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinf. 8*, 460.
38. Carvalho, A. L., Dias, F. M., Prates, J. A., Nagy, T., Gilbert, H. J., Davies, G. J., Ferreira, L. M., Romao, M. J., and Fontes, C. M. (2003) Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proc. Natl. Acad. Sci. U.S.A. 100*, 13809–13814.
39. Boraston, A. B., McLean, B. W., Chen, G., Li, A., Warren, R. A., and Kilburn, D. G. (2002) Co-operative binding of triplicate carbohydrate-binding modules from a thermophilic xylanase. *Mol. Microbiol. 43*, 187–194.
40. Freelove, A. C., Bolam, D. N., White, P., Hazlewood, G. P., and Gilbert, H. J. (2001) A novel carbohydrate-binding protein is a

component of the plant cell wall-degrading complex of *Piromyces equi*. *J. Biol. Chem. 276*, 43010–43017.
41. Abou-Hachem, M., Karlsson, E. N., Simpson, P. J., Linse, S., Sellers, P., Williamson, M. P., Jamieson, S. J., Gilbert, H. J., Bolam, D. N., and Holst, O. (2002) Calcium binding and thermostability of carbohydrate binding module CBM4-2 of Xyn10A from *Rhodothermus marinus*. *Biochemistry 41*, 5720–5729.
42. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol. 233*, 123–138.
43. Ichinose, H., Yoshida, M., Kotake, T., Kuno, A., Igarashi, K., Tsumuraya, Y., Samejima, M., Hirabayashi, J., Kobayashi, H., and Kaneko, S. (2005) An exo-β-1,3-galactanase having a novel β-1,3-galactan-binding module from *Phanerochaete chrysosporium*. *J. Biol. Chem. 280*, 25820–25829.