BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2009,** *48,* 10395–10404 10395
DOI: 10.1021/bi9013424

Analysis of the Structural and Functional Diversity of Plant Cell Wall Specific Family 6 Carbohydrate Binding Modules[†,‡]

D. Wade Abbott,[§,@] Elizabeth Ficko-Blean,[§] Alicia Lammerts van Bueren,[§,#] Artur Rogowski,[∥] Alan Cartmell,[∥] Pedro M. Coutinho,[⊥] Bernard Henrissat,[⊥] Harry J. Gilbert,[∥,@] and Alisdair B. Boraston*[,§]

[§]*Department of Biochemistry and Microbiology, University of Victoria, P.O. Box 3055 STN CSC, Victoria, British Columbia V8W 3P6, Canada,* [∥]*School of Biomedical Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, U.K., and* [⊥]*Laboratoire d'Architecture et de Fonction des macromolécules Biologiques, IBSM, CNRS Marseille and University Aix-Marseille I & II, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France.* [@]*Present address: Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Rd., Athens, GA 30602.* [#]*Present address: York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5YW, U.K.*

Received August 3, 2009; Revised Manuscript Received September 29, 2009

ABSTRACT: Carbohydrate binding modules (CBMs) play important biological roles in targeting appended catalytic modules to their dedicated substrate(s) within complex macromolecular structures such as the plant cell wall. Because of the large potential in ligand diversity within nature and our continually expanding knowledge of sequence-based information of carbohydrate-modifying enzymes, empirical determination of CBM binding specificity and identification of novel mechanisms in carbohydrate recognition by these proteins have become time-consuming and complicated processes. To help overcome these experimental hurdles, we present here a predictive model for family 6 CBMs (CBM6) that is based upon several factors, including phylogenetic relatedness, and structural and functional evidence. This analysis has determined that five regions within the binding site, termed A−E, play key roles in ligand selection and affinity. Regions A−C are located in a primary subsite and contribute mainly to binding energy and selection for O2, O3, and O4 equatorial hydroxyls. Region D appears to determine whether the CBM will interact with internal or terminal structures of the carbohydrate ligand. Region E displays the largest degree of variation and is thus predicted to make the most significant contribution to specificity. This model is supported by the biochemical properties and structure of a CBM6 from *Clostridium cellulolyticum* (*Cc*CBM6), which we also report here. The protein bound specifically to xylose and the nonreducing of end of polymers containing this pentose sugar. The crystal structure of *Cc*CBM6 in complex with xylose showed that a tyrosine residue made hydrophobic contacts with the unsubstituted C5 atom of xylose and sterically hindered decorations at this sugar ring position. The mechanism, by which the CBM recognizes xylose but not glucose, a specificity not previously observed in this family, supports our predictive model that holds that variation in region E plays a key role in the diverse ligand selection evident in CBM6.

Carbohydrate binding modules (CBMs) are noncatalytic components of carbohydrate active enzymes, such as glycoside hydrolases and polysaccharide lyases. CBMs fold independently of the parent enzymatic module and retain their carbohydrate binding function when produced independently of the intact polypeptide. CBMs target specific regions of complex macromolecular structures in plant cell walls and exert their influence by holding the appended catalytic module in the proximity of its substrate, thereby enhancing the rate of catalysis (*1*). While the role of CBMs in the biodegradation of plant cell walls and storage polysaccharides is well-established, recent work has highlighted the importance of these modules in directing enzymes derived from human pathogens (i.e., virulence factors) to their target substrates (*2*−*4*). Thus, CBMs are of current and emerging biotechnological importance for both the efficient production of renewable biofuels from plant biomass and the development of therapeutic agents for bacterial infections.

CBMs are grouped into more than 50 sequence-based families in the CAZy database (*5*). The modular property of CBMs has facilitated their study in isolation, an approach that has revealed several examples of CBM families that display diversity in ligand binding (*5*, *6*), a phenomenon that includes flexibility in the recognition of sugar ring configuration and decorations, linkage type, and degree of polymerization. On the basis of the nature of the ligands recognized by CBMs, these modules have been assigned into three functional classes: Types A, B, and C (*7*). Type A CBMs bind to highly crystalline polysaccharides, such as cellulose or chitin; Type B CBMs interact with soluble oligosaccharide chains, while Type C CBMs bind to smaller oligosaccharides, generally recognizing the terminal sugars of these molecules. Although individual CBMs are typically categorized into a single

[†]We declare that we have no competing financial interest or conflict of interest. This work was supported by the National Science and Engineering Research Council of Canada (NSERC). E.F.-B. and A.L.v.B. are funded by a Michael Smith Foundation for Health Research (MSFHR) Doctoral Award and NSERC Doctoral Scholarships, respectively. A.B.B. is a Canadian Research Chair in Molecular Interactions and MSFHR Scholar.

[‡]Coordinates for *Cc*CBM6 have been deposited in the Protein Data Bank as entry 2V4V.

*To whom correspondence should be addressed: Department of Biochemistry and Microbiology, University of Victoria, P.O. Box 3055 STN CSC, Victoria, British Columbia V8W 3P6, Canada. E-mail: boraston@uvic.ca. Telephone: (250) 472-4168. Fax: (250) 721-8855.

Published on Web 09/29/2009

functional class, Type B and Type C modules have been observed within the same family.

The majority of CBM families adopt a β-sandwich fold, which consists of two opposing sheets of parallel or antiparallel β-strands connected through various loop topologies. Although this protein fold is commonly observed, the ligand binding site can display variability in amino acid content, metal coordination capabilities, subsite composition, surface topography, and location on the protein surface. Subtle changes to any of these parameters can radically alter the binding specificity of even closely related CBMs (*6*). This phenomenon is responsible for producing one of the greatest challenges in CBM research: predicting target ligands (or at least identifying modules with novel specificities) and subsequently characterizing the molecular determinants involved in carbohydrate recognition.

Structural analysis has revealed that there can be dramatic differences in the locations and topographies of CBM binding sites. For example, binding sites can be located within both the concave face of the β-sandwich, generally modules with Type A or Type B properties, or distally positioned within the loop regions connecting the two β-sheets (*7*), which normally, but not exclusively, exhibit Type C properties. This latter region of the protein appears to be evolving and adopts varied topographies, an observation that partially explains its capacity for accommodating diverse ligands (*6*). CBM6s, the focus of this study, represent a chimera of these two possibilities as there are binding sites described both within the loop regions of the β-sandwich (defined as site I here; previously termed cleft A) and on the concave face (site II; previously termed cleft B) (*8*, *9*).

To help overcome the experimental hurdles that lay between the initial identification of a gene encoding a potential CBM and the characterization of its binding function, different approaches have been developed. These include phylogenetic analysis of individual modules from the same family (*6*), direct biochemical analysis of multiple sequences within the same family (*2*, *10*), and sequence-based creation of subfamilies within the CAZy database (*5*). The power of the phylogenetic approach to predicting target ligands from a diverse set of parent enzyme families is exemplified by a recent analysis of CBM32s, which displays one of the most diverse ligand binding potentials currently described (*6*). This report demonstrated how the relatedness of parent enzyme activities can be harnessed to predict the function of uncharacterized CBMs. Such an approach is useful for CBM families that exhibit diverse binding specificities and contain enzyme modules with characterized activities. This study was followed by a complementary report on CBM51s, which combined bioinformatic and biophysical techniques to establish the presence of six subfamilies (*2*). Importantly, two of these subfamilies (CBM51a and CBM51b) had biologically significant differences in specificity that complemented the activities of their parent catalytic modules, similar to the division of CBM2 into cellulose (CBM2a) and xylan (CBM2b) binding subfamilies (*11*).

CBM6s are found in a range of different enzymes that display activities against xylan, mannan, β-glucans, agarose, and arabinans. There are currently 21 Protein Data Bank (PDB) entries for CBM6s, including many in complex with different ligands, and several complementary biochemical studies demonstrating the ability of these family members to interact with diverse carbohydrate targets. This information provides an appropriate framework with which to evaluate the utilization of bioinformatics to predict novel ligand specificities. Sequence comparisons across the entire family, presented here, identified unique properties in an uncharacterized CBM6 from *Clostridium cellulolyticum* (*Cc*CBM6). Biochemical analysis demonstrated that the protein recognizes xylose either as a monosaccharide, the nonreducing end of xylooligosaccharides, or as decorations on xyloglucan, a general specificity not previously reported for any CBM. The crystal structure of *Cc*CBM6 in complex with xylose corroborates the predictive capabilities of the phylogenic analyses, suggesting that it is a legitimate approach for selecting future CBM6 targets that recognize novel plant cell wall structures.

EXPERIMENTAL PROCEDURES

Modular Determination of CBM6s, Phylogenetic Analysis, and Subfamily Creation. One hundred sixty-seven total sequences were extracted from entries containing CBM6 modules within the CAZy database (http://www.cazy.org), their limits being the result of a systematic manual and semiautomated modular annotation (*5*). The edited CBM6 sequences were first aligned with Muscle version 3.52 (http://www.drive5.com/muscle/) (*12*), and the resulting distance matrix was derived using Blosum62 substitution parameters. The Secator algorithm (*13*) was then used for subfamily definition. The resulting tree was visualized using the Dendroscope viewer (*14*).

Construction of Plasmids and Protein Production. The gene fragment encoding the family 6 CBM, here called *Cc*CBM6, from GH59 (base pair 2751 to 3153), and the highly similar CBM6 from GH27 were cloned from *C. cellulolyticum* genomic DNA into expression vector pET28b (Novagen, Gibbstown, NJ) using standard PCR cloning procedures (*15*). Primers contained 5′ *Nhe*I and 3′ *Xho*I restriction sites to enable directional cloning (see the Supporting Information for oligonucleotide sequences). The coding sequences of both *Cc*CBM6s were fused in frame with an upstream sequence encoding six histidine residues (His tag) and a thrombin cleavage site. The resulting plasmids, named pA59 and pA27, were used to transform *Escherichia coli* BL21-(DE3) cells for production of the encoded polypeptide. Mutagenesis of *Cc*CBM6 (GH59) was performed using QuikChange (Stratagene, La Jolla, CA) with the primers listed in the Supporting Information.

We produced soluble wild-type and native proteins by inoculating transformed *E. coli* into LB medium supplemented with kanamycin (50 μg/mL) and shaking the cultures at 37 °C until the optical density at 600 nm reached 0.4. Isopropyl β-D-thiogalactopyranoside (IPTG) was added to final concentration of 1 mM to induce protein production, and the culture was incubated with shaking for an additional 5 h at 37 °C. Cells were harvested by centrifugation at 4 °C and resuspended in 10 mL of 20 mM Tris-HCl (pH 8.0) containing 300 mM NaCl. The cells were disrupted by either sonication (for functional studies) or chemical lysis (structural studies). Following clarification of the cell lysate by centrifugation, the polypeptide was purified by immobilized metal affinity chromatography.

Isothermal Titration Calorimetry of CcCBM6. To prepare the protein for calorimetry, *Cc*CBM6 was exhaustively dialyzed against 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl. Dialysate was collected and used to dissolve ligands at a concentration of 10 mM for the xylooligosaccharides and 1% (w/v) for xyloglucan. All calorimetry runs were performed using 100 μM protein and a temperature of 25 °C. Reagents were filtered and extensively degassed prior to use. Each run consisted of 27 injections. Because of the relative low affinities observed in these experiments [C values of < 1 (*16*)], the stoichiometries for the interactions could not be experimentally determined. Thus, to

provide the most conservative estimates of the association constants, data were twice fit using Origin scientific plotting software (version 7) to a single-binding site model, first with the stoichiometry constrained to 0.5 and then with the stoichiometry constrained to 1.5. It was reasoned that this would provide conservative upper and lower bounds for the association constant assuming that the stoichiometry was indeed near 1 as indicated by the crystallographic data. Reported values indicate the middle of the range produced by these two fits, while the error gives the upper and lower bounds of the range.

CcCBM6 Crystallization and Structure Determination. Prior to crystallization, the six-histidine tag was removed from *Cc*CBM6 by overnight thrombin treatment at room temperature. The protein was then repurified by size exclusion chromatography using a Sephacryl S-200 column (GE Biosciences). Following multiple rounds of optimization, *Cc*CBM6 was ultimately crystallized by the hanging drop vapor diffusion method at 18 °C in 20% polyethylene glycol 3350, 0.1 M sodium acetate, 0.1 M sodium citrate dihydrate (pH 5.8), and 10 mM xylose. The crystals were cyroprotected in the crystallization solution supplemented with 20% ethylene glycol and cryocooled at 113 K directly in a nitrogen stream. Diffraction data were collected on a Rigaku R-AXIS4++ area detector and a MM-002 X-ray generator with Osmic "blue" optics and an Oxford Cryostream 700 instrument. A molecular replacement solution using the *Clostridium thermocellum* xylanase CBM6 coordinates, *Ct*CBM6, as a search model [PDB entry 1GMM (*17*)] was found for the single molecule in the asymmetric unit using Phaser (*18*). This model was manually corrected using COOT (*19*) and refined with REFMAC5 (*20*). Water molecules were added using the ARP/wARP (*21*) option within REFMAC and manually inspected. Five percent of the reflections were flagged as "free" (*22*) and used to monitor refinement procedures. Model validation was performed with SFCHECK (*23*) and PROCHECK (*24*).

Enzyme Assays. Kinetic runs were conducted in triplicate for statistical relevance with seven different PNP-α-Gal concentrations straddling the K_m (62.5 μM, 125 μM, 250 μM, 500 μM, 1 mM, 2 mM, and 4 mM) with 1 nM *Cc*GH27 in 50 mM sodium phosphate (pH 7.0) at 37 °C. The production of PNP was monitored by UV absorbance at a wavelength of 400 nM using an extinction coefficient of 10000 M^{-1} cm^{-1} to determine its concentration. Data fitting was done by nonlinear regression using Graphpad Prism version 4.0.

RESULTS AND DISCUSSION

Description of the CBM6 Phylogenetic Tree. At the time of analysis, CBM family 6 consists of 167 total entries, including three from archaea (*Haloarcula marismortui*), 158 from bacteria, five from eukayota, and three unclassified sequences. Each entry was extracted from the modular parent enzyme and treated independently using the MUSCLE analysis (*12*) to identify the sequence relationships between the CBM6 members. This process resulted in the creation of four major subfamilies (CBM6a-d) (Figure 1). Each subfamily consists of CBMs appended to catalytic modules that share similar characterized activities (see Table 1). This result is consistent with a recently reported study focusing on subfamily CBM6d, members of which display specificity for β- and α-agarases (*25*).

Structural Contributions to Ligand Binding Specificity. Qualitative sequence analysis is useful for categorizing individual CBMs into functional subfamilies and clusters. Within closely related CBMs, ligand binding specificity can often be predicted by comparing the characterized and predicted activities of the appended catalytic modules, as commonly, catalytic specificity will reflect CBM binding specificity. Likewise, the distribution of CBMs from the same polypeptide into homogeneous (i.e., highly similar CBMs) and heterogeneous (i.e., highly divergent CBMs) patterns provides clues about the specialization of multimodular proteins as more distantly related sequences are more likely to recognize distinct ligands or utilize divergent mechanisms of recognition. However, the potential of this approach to provide detailed information into the specific molecular determinants of ligand binding and selectivity has several limitations. First, small changes to amino acid sequence can have significant functional implications, as observed for the four-amino acid insertion within *Ct*CBM6 (*26*). Therefore, quantitative comparisons of global amino acid identity between two sequences can be misleading, and comparative analyses need to be weighted toward residues within the binding site and attributed to ligand recognition. Second, accurate correlations between parent enzyme function and potential CBM ligand specificity are dependent upon the successful characterization of the appended catalytic module(s). In some cases, this information is reliable and empirically tested; however, in many examples, it is based simply upon sequence-based predictions of enzyme activity. Third, and perhaps most importantly, is the fact that the use of sequence alignments in predicting CBM function is directly related to the relatedness that exists between the unknown CBM6 and the CBM6 for which there is biochemical and structural information available. Increasing sequence divergence will weaken the functional correlations that can be made between two targets. These limitations are even more pronounced for a CBM family such as CBM6, where significant ligand binding diversity has been well-documented.

To provide a more accurate predictive model for CBM6 function, therefore, we have performed a thorough analysis of available CBM6 complexes within the structural database with the intention of uncovering key amino acid positions for ligand binding affinity and selectivity (see Table 2). An overlay of *Cs*CBM6-1, *Cs*CBM6-3, *Bh*CBM6, *Ct*CBM6, *Cm*CBM6-2, and *Sd*CBM6-2 with a focus on the key binding residues in site I reveals, first, regions that are both spatially and functionally conserved and, second, regions of apparent diversity in structure that confer functional variation within the family (Figure 2). Three regions, A, B, and C, fall into the first category, while two regions, D and E, fall into the latter.

Regions A and B consist of structurally conserved aromatic amino acid side chains that "sandwich" a sugar monomer. Region C contains a structurally conserved asparagine whose side chain forms a direct hydrogen bond with O4 of the sugar bound by regions A and B. These three regions likely represent a universal mechanism of CBM6 carbohydrate recognition as these amino acid sequences are conserved in all but a very few CBM6 sequences (not shown). In addition to these side chains, there is a conserved water molecule that forms a hydrogen bond with the carbonyl group of asparagine in region C and to the equatorial C3 or C4 hydroxyl groups of the sugar that is sandwiched between regions A and B. The exception in this case is *Sd*CBM6-2 where, because of the cyclization of galactose at C3 and C6, the oxygen is no longer in the correct position to form a water-mediated hydrogen bond. Thus, the amino acids that we have defined as regions A, B, and C all interact with a single sugar residue of the carbohydrate ligand. However, the observation that these amino acids are conserved in all CBM6s, which, collectively, display a wide range of ligand specificities, suggests

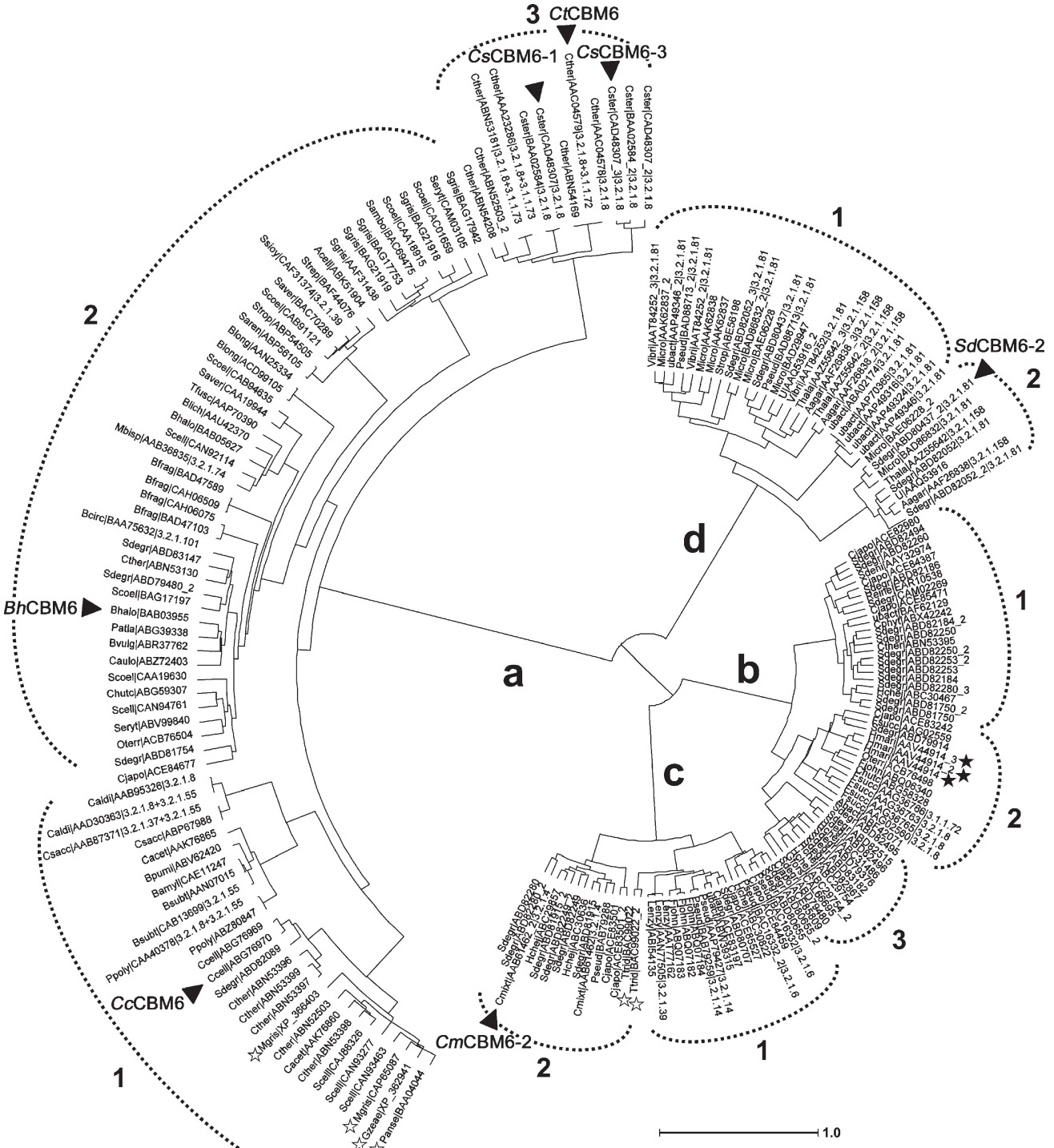

FIGURE 1: Phylogenetic tree of CBM6s displayed as a wheel diagram. Subfamily branches are labeled a−d, and the clusters within each subfamily are indicated with numbered dashed lines. The parent enzymes from each subfamily of CBM6 are associated with an activity on a general substrate: subfamily 6a, hemicellulose; subfamily 6b, xylan; subfamily 6c, β-glucans with a variety of linkages; and subfamily 6d, agarose. CBM6 sequences are labeled with corresponding species name and GenBank entries, and where appropriate, the position of the module within an enzyme containing multiple CBM6s is indicated with an underscore and number and the EC number of the parent enzyme if known. CBM6s with known structures in complex with carbohydrate ligands that were used within this study are indicated with a black triangle and labeled.

that regions A−C contribute little to this diverse specificity, although the aromatic sandwich (in which the aromatic residues are parallel to the sugar ring) does select for sugars where O2, O3, and O4 are equatorial.

Regions D and E appear to be "hot spots" of primary and tertiary structure variation that confer functional specificity at site I in CBM6s. Of the currently determined structures, region D presents amino acid side chains at one end of the binding site. Depending on the chemistry of this residue, it appears to function as a specificity determinant by closing off the core binding site through interacting with the terminus of the ligand, or opening it with hydrophobic residues that line the floor of the binding cleft (Figure 2B). If this correlation in fact represents a general principle in CBM6 function, the architecture would provide

Table 1: Subfamily Distribution of CBM6s Appended to Parent Enzymes with Assigned EC Numbers[a]

enzyme type	subfamily 6a[b] (hemicellulose)	subfamily 6b (xylan)	subfamily 6c (β-glucans)	subfamily 6d (agarose)
3.1.1.72	1	1		
acetylxylan esterase				
3.1.1.73				
feruloyl esterase	2			
3.2.1.4				
cellulase			2	
3.2.1.6				
endo-1,3(4)-β-glucanase			2	
3.2.1.8				
endo-1,4-β-xylanase	12	3		
3.2.1.14				
chitinase			2	
3.2.1.37				
xylan 1,4-β-xylosidase	1			
3.2.1.39				
glucan endo-1,3-β-D-glucosidase	1		1	
3.2.1.55				
α-N-arabinofuranosidase	4			
3.2.1.74				
glucan 1,4-β-glucosidase	1			
3.2.1.81				
β-agarase				18
3.2.1.101				
mannan endo-1,6-α-mannosidase	1			
3.2.1.158				
α-agarase				6

[a]CBM6s present in multiple copies within a multimodular enzyme are treated independently. Therefore, enzymes containing multiple CBM6s have more than one entry. [b]Generalized substrate specificity of catalytic modules appended to CBM6s within the indicated subfamily.

Table 2: List[a] of CBM6s with Known Three-Dimensional Structures

CBM6	PDB entry	parent enzyme	accession number	subfamily
*Bh*CBM6	1w9s	GH81	BAB03955	6a
xylobiose	1w9t			
laminarihexaose	1w9w			
*Bs*CBM6[b]	EC7E	XynD	CAB13699	
	EC7F			
	EC7G			
	EC7H			
	EC7O			
*Cc*CBM6	–	putative	ABG76970	6a
xylose	2V4V	galactocerebeosidase		
*Cm*CBM6-2	1uxz	CelB	AAB61462	6c
(1,3),(1,4),(1,3)-β-glc	1uy0			
cellobiose	1uyx			
cellotriose	1uyy			
xylotetraose	1UYZ			
(1,4),(1,3),(1,4)-β-glc	1UZ0			
*Cs*CBM6-1	1UY1	putative	AAL14106	6a
xylobiose	1UY2	xylanase		
xylotriose	1UY3			
xylotetraose	1UY4			
*Cs*CBM6-3	1O8P	putative	AAL14106	6a
xylotriose	1NAE	xylanase		
cellobiose	1O8S			
laminaribiose	1OD3			
*Ct*CBM6	1GMM	Xyn11A	AAC04579	6a
xylopentaose	1UXX			
*Sd*CBM6-2	–	Aga16B	ABD80437	6d
neoagarohexaose	2CDO			
neoagarohexaose	2CDP			

[a]List taken from the CAZy website [www.cazy.org (5)]. [b]This structure was not included in this analysis because it is an apo form.

key selectivity in ligand binding by determining whether the CBM binds to internal regions within a polysaccharide or to their termini. In the known cases, *Cs*CBM6-1, *Cs*CBM6-3, and *Ct*CBM6 all bind to internal stretches of polysaccharide within xylan, having at least four binding subsites that interact with the ligand. Region D for these proteins consists of an isoleucine in *Cs*CBM6-1 and *Ct*CBM6, while *Cs*CBM6-3 contains a phenylalanine (Figure 2). The isoleucine side chain is positioned in a trajectory parallel to the binding pocket, which allows xylan to extend out over the aliphatic group into the bulk solvent. The phenylalanine in *Cs*CBM6-3 is disposed in an analogous orientation creating a similar effect. In contrast, for reported CBM6s that bind terminal sugars, in particular, *Bh*CBM6, *Sd*CBM6-2, and *Cm*CBM6, all have amino acid side chains that transform the binding site within region D from an open cleft into a closed pocket (Figure 2B). A glutamate residue that is conserved between *Bh*CBM6 and *Cm*CBM6-2 performs this function and forms a hydrogen bond with the nonreducing terminus of the ligand. In *Sd*CBM6-2, a similar structural effect is created by a tyrosine residue that is conserved in three-dimensional space with the glutamates of *Bh*CBM6 and *Cm*CBM6. Currently, using region D as a predictive tool is complicated by the sequence heterogeneity present within the stretch of amino acids containing this residue (Figure 2A). Further structural investigation into end-binding and chain-binding CBM6s, therefore, will be required to clarify the position of the region D signature in the absence of an available structure and determine if the correlation between polarity and function is consistent throughout the entire CBM6 family.

Region E is perhaps the most poorly defined in terms of structure but appears to have the most profound impact on ligand selectivity. This region is composed of a loop (E_{loop}) linking β-strands 10 and 11, which is variable in length and amino acid composition (Figure 2A). In *Ct*CBM6, *Cs*CBM6-1, *Cs*CBM6-3, and *Cm*CBM6, the E_{loop} is short, leaving the binding site open and allowing the ligand to extend out into solvent. This architecture facilitates the internal binding mode displayed by these proteins. *Bh*CBM6 and *Sd*CBM6-2 have larger E_{loop}s containing signature aromatic amino acid side chains. In *Bh*CBM6, this feature enlarges the binding site (Figure 2B). These contours accommodate the curved shape of its β-1,3-glucan ligand, allowing it to wrap around the CBM surface. Likewise, *Sd*CBM6-2 has a similar extended loop, but it contains a tryptophan residue in a structurally distinct position. This signature creates a second subsite for the galactose sugar in neoagarobiose and allows the linear sugar to extend outward. Interestingly, the presence of a lengthened E_{loop} appears to correlate with a glycine at the beginning of β-strand 7 (Figure 2A). The loops from *Ct*CBM6, *Cs*CBM6-1, and *Cs*CBM6-3 on the

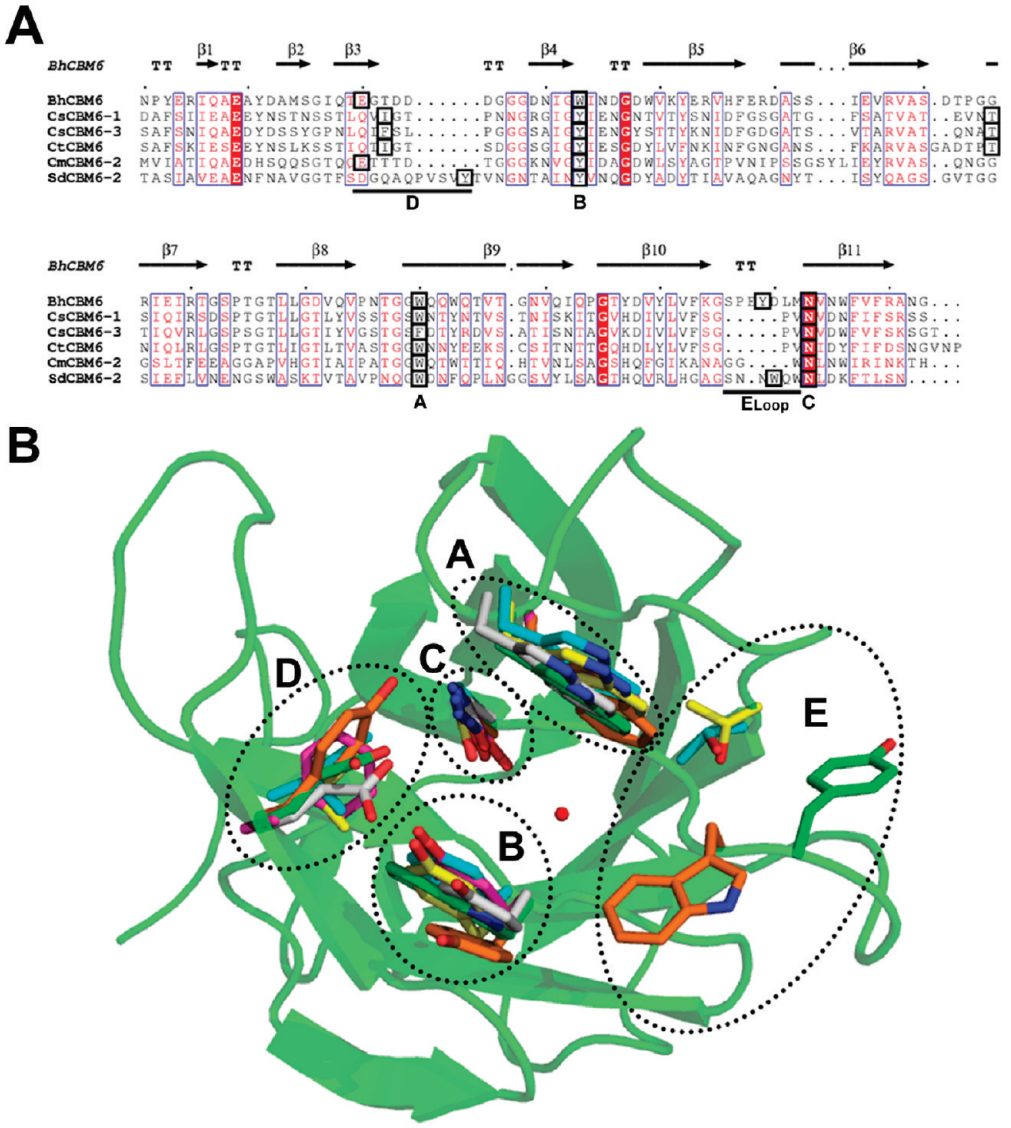

FIGURE 2: Universal mechanisms and ligand selectivity within CBM6 binding sites. (A) Structural alignments of known CBM6 structures. Secondary structure elements extracted from the *Bh*CBM6 coordinates are displayed above the sequences. Regions important for ligand binding within site I (A−E_{loop}) are indicated by black boxes and labeled below the alignment. (B) Structural superimposition of amino acids from regions A−E contributing to formation of the complex in CBM6s. The transparent ribbon diagram displayed as a backbone colored green is from *Bh*CBM6. The amino acid side chains are displayed in different colors for the sake of clarity: green for *Bh*CBM6, yellow for *Cs*CBM6-1, magenta for *Cs*CBM6-3, cyan for *Ct*CBM6, gray for *Cm*CBM-2, and orange for *Sd*CBM6-2. The conserved water molecule found within the core of all CBM6 binding sites is shown as a red sphere.

other hand have a threonine in this position. *Cm*CBM6 appears to be an exception to this observation as it does contain a conserved glycine and a shortened E_{loop}. The functional significance of this observation is currently unknown. To determine if this correlation will emerge as a useful feature for identifying putative CBM6s with specialized binding properties, further structural and functional studies, including mutagenesis to display loss or gain of function phenotypes, will be required.

At the most detailed level, our bioinformatic model enables us (i) to predict from the nature of the residue at position D whether a protein is likely to bind internally or to the termini of their complex carbohydrate ligands and (ii) to predict from the sequence in region E whether the protein is likely to display novel specificity. We should emphasize that at this stage our model does not allow the identification of the target ligand. Thus, these analyses reveal their greatest power in predicting novel binding properties in family 6 CBMs. It should also be made clear that our analyses lack the resolution to predict single-amino acid changes that would allow the identification of the predicted novel CBM specificity. For example, a glutamate in region D is associated with "end-binding" CBMs while an aliphatic, such as isoleucine or phenylalanine, is associated with internal binding on chains. However, a simple glutamate to isoleucine mutation, or vice versa, is unlikely to be a successful approach to manipulating the specificity of the CBM. Though the glutamate is a key determinant, its role is dictated by surrounding structural features that enable it to adopt a conformation appropriate to its function.

With this predictive model in hand, we sought to explore its utility in predicting functionality by characterizing the CBM6 from the *C. cellulolyticum* predicted GH59 enzyme Gal59A (*Cc*CBM6). Sequence alignment of *Cc*CBM6 with *Cs*CBM6-1 and *Ct*CBM6 (all from subfamily 6a), proteins that bind to xylose polymers and have known three-dimensional structures, reveals conserved residues in regions A (Trp110), B (Phe52), and C (Asn142) (Figure 3A). As described above, regions A and B contribute to

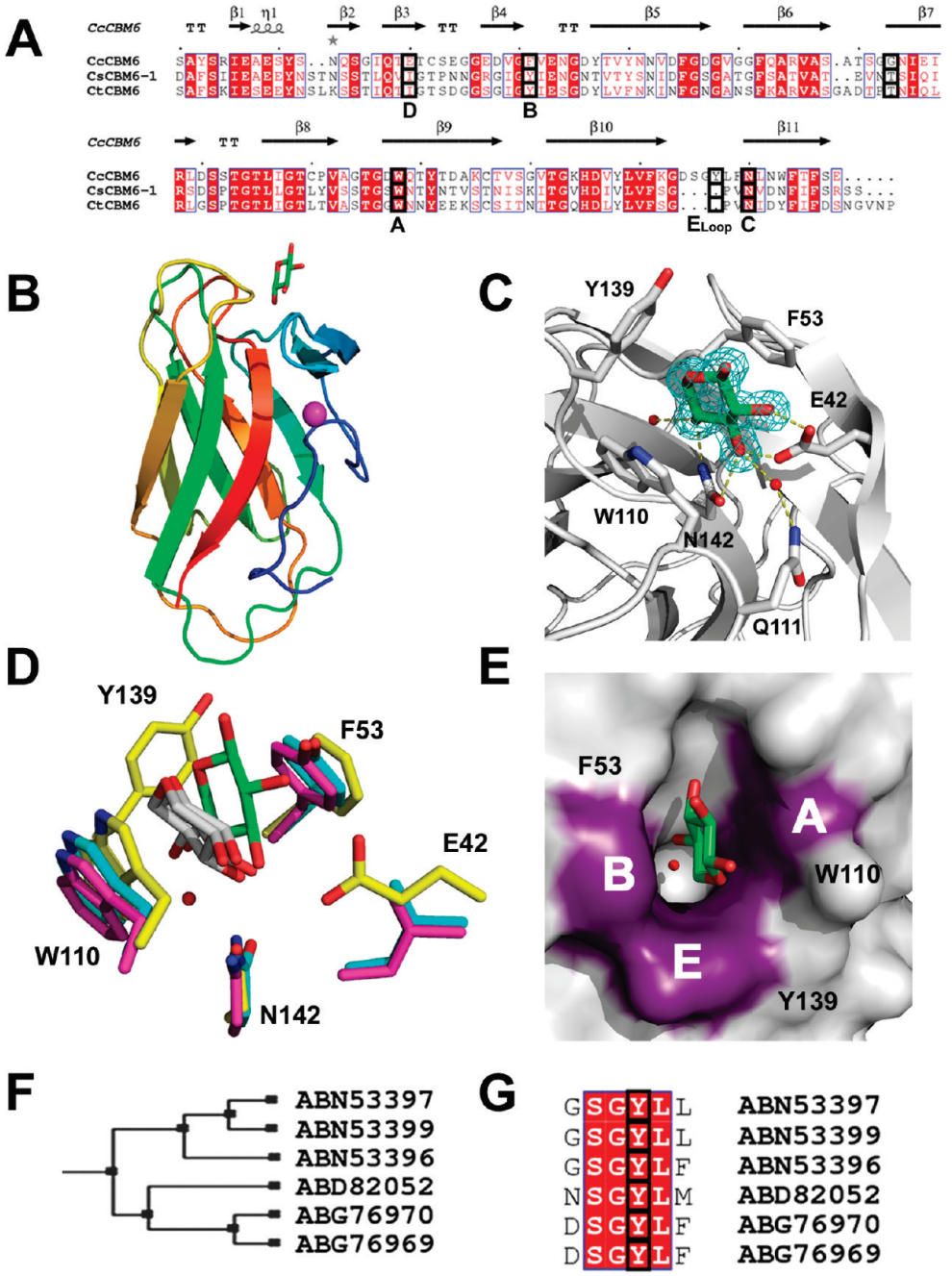

FIGURE 3: Structural analysis of the phylogeny and function of *Cc*CBM6. (A) Structural alignment of xylose binding CBM6s. The secondary structure elements extracted from the *Cc*CBM6 coordinates are indicated above the alignment. The important amino acids for binding are indicated with black boxes and labeled below the sequences. (B) Secondary structure representation of the overall fold of *Cc*CBM6 at 1.5 Å. The model is color-ramped from blue (N-terminus) to red (C-terminus). The structural sodium ion is represented as a magenta sphere, and the bound xylose molecule is shown as green sticks (PDB entry 2V4V). The maximum-likelihood (*29*)/σ_A (*30*) $2F_o - F_c$ map is contoured at 1.0σ (0.49 electron/Å^3). (C) Binding site architecture of *Cc*CBM6. Amino acids and xylose are displayed as sticks with H-bonds depicted as yellow dashes. The two ordered water molecules in contact with xylose are shown as red spheres. (D) Structural superimposition of *Cs*CBM6-1 (cyan), *Ct*CBM6 (magenta), and *Cc*CBM6 (yellow) in complex with xylose. The xylooligosaccharides from the internal binding *Cs*CBM6-1 and *Ct*CBM6 (both colored gray) have been trimmed down to a monosaccharide for the sake of clarity. The xylose ligand from *Cc*CBM6 is colored green. The amino acids are also labeled with the appropriate region. (E) Solvent accessible surface model of the complex binding site depicting the "coin-slot"-shaped binding site of *Cc*CBM6. The ordered water molecule in contact with O4 of xylose is shown as a red sphere. The aromatic residues involved in binding are colored purple and labeled as a region. (F) Alignment of CBM6s that cluster with *Cc*CBM6 (subfamily 6a, cluster 2). (G) Amino acid sequence of E_{Loop} regions from CBM6s shown in panel E highlighting the strict conservation of tyrosine residues.

the overall binding energy but are not necessarily important for ligand selectivity, and region C interacts with hydroxyl groups in the equatorial configuration. In contrast, when regions D and E are compared among the three proteins, there are distinct differences in *Cc*CBM6, which hint at unique molecular determinants for selective carbohydrate recognition. Within region D, Glu42 aligns with the glutamates in *Bh*CBM6 (subfamily 6a) and *Cm*CBM6 (subfamily 6c), residues that confer a pocket topology to the binding site corresponding to their end binding functions. In addition, there is a signature glycine at the beginning of β-strand 7, consistent with a lengthened E_{loop}, and an aromatic residue (Tyr139) within the predicted E_{loop} of *Cc*CBM6.

Significantly, this aromatic is not positionally conserved with the tyrosine from the *Bh*CBM6 sequence, which contributes to its unique binding site architecture. Although the structural significance of this observation remains to be determined, the flexibility of a glycine residue at this position may facilitate remodeling of the E_{loop} of the binding site. Taken together, these predictions, which are based upon our combined phylogenetic and structural analyses of CBM6s, suggest that *Cc*CBM6 may bind terminal ends of its ligand in a pocket or blind canyonlike binding site (similar to both *Cm*CBM6 and *Bh*CBM6); however, the Tyr139 present within the E_{loop} may contribute to unique binding site selectivity.

The binding specificity of *Cc*CBM6 was screened, and binding affinities were quantified using isothermal titration calorimetry (see Figure 1 of the Supporting Information). The module interacts with xylose and xylooligosaccharides with similar affinities [xylose $K_A = (1.7 \pm 0.2) \times 10^3\ M^{-1}$; xylohexaose $K_A = (2.1 \pm 0.2) \times 10^3\ M^{-1}$; xylobiose $K_A = (6.3 \pm 1.6) \times 10^3\ M^{-1}$]. *Cc*CBM6 also bound xyloglucan, which displays α-linked xylopyranose decorations, with an affinity of $\sim 2.0 \times 10^3\ M^{-1}$. This observation indicates that *Cc*CBM6 can bind to both α- and β-linked xylose residues, suggesting that there is inherent flexibility within the binding site to accommodate the linked sugars that could be positioned in either configuration. The protein did not bind to gluco-configured ligands (glucose, β-glucans, or cellooligosaccharides) and did not recognize the stereochemically distinct glycoses (galactose, arabinose, or mannose). The relatively similar affinities for xylose-containing carbohydrates suggest that the CBM is binding to the xylose decorations in xyloglucan and the nonreducing termini in the xylooligosaccharides, which is consistent with the functional predictions garnered from the phylogenetic analysis. Significantly, this specificity has not previously been observed in a CBM.

To define the molecular determinants of the *Cc*CBM6−xylose interaction, we determined the three-dimensional structure of *Cc*CBM6 in complex with xylose (see Table 3 for crystallization statistics). Similar to other CBM6s, *Cc*CBM6 displays a β-sandwich fold consisting of seven β-strands that form two antiparallel β-sheets (Figure 3B). In addition, there are two antiparallel strands at its N-terminus that border the loop regions of the module. Between these N-terminal strands and the convex β-sheet, there is a coordinated sodium ion that plays a structural role in stabilizing the fold. As predicted, site I in *Cc*CBM6 comprises the ligand binding site as structural analysis reveals electron density for a single xylose residue (Figure 3C). Because of the shape symmetry displayed by xylose, the orientation of the sugar was determined by *B*-factor analysis of the sugar refined in both possible orientations. One orientation gave a large discrepancy in *B* values between O5 and C5, whereas the chosen orientation gave good agreement between these values. Indeed, if the reducing end of the sugar was pointing into the pocket, the protein would not be able to bind to xyloglucan. Analysis of the binding site reveals that there is a constellation of direct and water-mediated hydrogen bonds between the hydroxyl groups of xylose and *Cc*CBM6 (Figure 3C). Key among these is Asn142, which hydrogen bonds with O3 and O4 of the xylose molecule and forms the floor of the binding site; Phe53 and Trp110 form the sides of a hydrophobic pocket that sandwich the planar pyranose ring. As described above, these features are conserved among members of CBM6, composing the A, B, and C regions, respectively, which confer specificity for sugars in which O2 and O3 are equatorial.

Table 3: Crystallization and Refinement Statistics of *Cc*CBM6

Data Collection	
space group	$P2_12_12_1$
cell dimensions	
a (Å)	31.03
b (Å)	43.60
c (Å)	93.32
α (deg)	90.00
β (deg)	90.00
γ (deg)	90.00
resolution (Å)	93.25−1.50 (1.54−1.50)
R_{merge}	0.06 (0.33)
$I/\sigma I$	13.1 (3.6)
completeness (%)	99.6 (99.9)
redundancy	4.37 (3.67)
Refinement	
resolution (Å)	1.50
no. of reflections	91777
R_{work}/R_{free}	0.19/0.22
no. of atoms	1147
protein	974
ligand/ion	10
water	162
B-factor (Å^2)	
protein	10.3
ligand/ion	10.0
water	21.9
root-mean-square deviation	
bond lengths (Å)	0.02
bond angles (deg)	1.58
Ramachandran statistics (%)	
allowed	99.1
generously allowed	0.9
disallowed	0.0

Within the more variable regions of the binding site, regions D and E, there are noticeable distinctions within *Cc*CBM6 that contribute to ligand selectivity. Region D consists of Glu42, the residue that aligns with glutamates in both *Bh*CBM6 and *Cm*CBM6, which forms one wall of the binding pocket and forms hydrogen bonds with O2 and O3 (Figure 3C). These interactions are critical for stabilizing the ligand in a "flipped" orientation, when compared to the other characterized CBM6−xylooligosaccharide complexes. This orientation results in the presentation of the O1 atom toward the bulk solvent, which is in agreement with the potential role of *Cc*CBM6 in binding the β-1,4-linked xylooligosaccharides and the α-1,6-linked decorations of xyloglucan. In this conformation, Asn142 in region C still interacts with O3 and O4 of xylose, but the hydroxyl groups are in the opposite orientation compared to those observed in the xylose chain binding *Cs*CBM6-1 and *Ct*CBM6 (Figure 3D). Tyr139, which is present in the E_{loop} (comprising residues Lys135−Phe141) and was predicted to confer unique properties to the binding site, contributes to the formation of a wall on the opposing side of the binding site with respect to region C and is in hydrophobic contact with C5 of xylose. This is a key determinant in the selectivity of xylose binding for two reasons: (1) it excludes the positioning of a C1 hydroxyl, which would be present if the xylose was in the same orientation as in the *Ct*CBM6 and *Cs*CBM6-1 complexes, and (2) it presents a steric block preventing the decoration of C5, explaining why the CBM is not able to bind to hexose sugars such as glucose. In this regard, Tyr139 is perhaps the most important determinant in directing

the interaction with the nonreducing end of xylose. We confirmed the importance of each of the residues that are predicted to interact with xylose using a site-directed mutagenesis approach, as conversion of Glu42, Phe53, Asn142, Trp110, and Tyr139 to alanine resulted in the complete ablation of xylose binding (as determined by the loss of quantifiable binding by ITC). It is possible that the Y139A mutation could make C5 available to decoration, and thus, potentially, the CBM variant will bind to glucose. While the Y139A mutant did exhibit trace affinity for glucose, it was far too weak to quantify by ITC or any other method, demonstrating that the hydrophobic interaction at C5 was indispensible. Overall, the important contributions of Glu42, Phe53, Trp110, and Tyr139 in contouring the binding site are evident when viewing the solvent-accessible surface model, which reveals that the binding site of *Cc*CBM6 has a coin-slot shape that would snugly accommodate a single terminal sugar pointed directly into the protein (Figure 3E).

Closer analysis of the sequence cluster for *Cc*CBM6 in the family 6 tree (Figure 1) reveals the presence of five other closely related CBM6 sequences from various organisms (Figure 3F). All six of these CBM6s are components of uncharacterized multimodular proteins belonging to *C. thermocellum* (Cthe_2195, ABN53397; Cthe_2197, ABN53399; Cthe_2194, ABN53396), *Saccharophagus degradans* (Cbm6H-CBM32F, ABD82052), and *C. cellulolyticum* (Gal59A, ABG76970; Gal27A, ABG76969). Predicted activities for the parent catalytic modules include a β-glycosidase GH2 (ABN53399), α-galactosidase GH27 (ABG-76969), a putative galactocerebrosidase GH59 (ABG76970), and an esterase (ABN53396). Despite the noticeable differences in parent enzyme structure, the CBM6s share between 50 and 91% identity, with the highest degree of homology belonging to the two sequences from *C. cellulolyticum*. These sequences display absolute conservation of the E_{loop} tyrosine (Figure 3G) and an acidic residue in region D (not shown). To determine if the relationship between regions D and E in these cases translates into a conserved xylose biding function, we produced a recombinant form of CBM6 from *C. cellulolyticum* GH27 (AG76969). ITC experiments confirmed that it also binds to xylose with an affinity similar to that of *Cc*CBM6 from the putative GH59 enzyme described above (not shown). This result suggests that two enzymes with notably different enzymatic activities utilize a targeting mechanism based upon a CBM with overlapping specificity, an observation that was recently reported for family 35 CBMs (*27*). To rule out the possibility that this observation was not simply a case of two enzymes from different families displaying redundant activities on common xylo-containing substrates, we produced and purified the catalytic modules of *Cc*GH59 and *Cc*GH27 and performed biochemical analyses. No activity for the GH59 catalytic module could be detected by thin layer chromatography, even against galactocerebrosides, the glycolipid previously demonstrated to be hydrolyzed by the human enzyme from this family (*28*) (see the Supporting Information for a list of tested substrates). Biochemical studies, however, confirmed that the GH27 (ABG76969) enzyme displays α-galactosidase activity as it hydrolyzes 4-nitrophenyl α-galactosylpyranoside, with a k_{cat} of 11400 ± 1700 min^{-1} and a K_M of 495 ± 3.2 μM, but failed to attack any other α- or β-arylglycosides, including aryl-xylosides (see Figure 2 of the Supporting Information). We have also explored the activity of GH27 against α-D-galactose-containing oligosaccharides, which revealed that the enzyme was more active against galactosyl-mannotrisose (k_{cat} = 51.2 ± 8.6 min^{-1}, and K_M = 3.90 ± 0.6 mM) than stachyose (k_{cat} = 22.8 ± 1.0 min^{-1}, and K_M = 19.8 ± 1.0 mM) or raffinose (k_{cat} = 29.9 ± 1.4 min^{-1}, and K_M = 101 ± 12.1 mM).

Genomic analysis of the locus containing the CBM6s appended to GH27 and GH59 catalytic modules within *C. cellulolyticium* revealed the presence of a significantly expanded gene cluster that contains 14 genes (Ccel1229−Ccel1242) and one outlier (Ccel1656) with annotated CBM6s (*5*) (see the Supporting Information). The predicted activities of the parent enzymes within this locus display a wide variety of specificities, including potential activities against xylan, arabinose side chains, and esterase linkages, and belong to families unrelated by sequence, including GH2, GH10, and GH43. All but one of these genes belong to the Ccel_1229−Cecl_1242 gene cluster, suggesting that they are coregulated. Interestingly, within every *C. cellulolyticum* CBM6, there is absolute conservation of the tyrosine residue that provided terminal xylose binding specificity. In contrast, within region D, there is notable sequence divergence at this position, suggesting that within the entire CBM6 cluster there is the potential for both overlapping and novel binding specificities. Interestingly, in four of the CBM6 sequences from C. *cellulolyticum*, the region D residue is present as a valine and in one other sequence as a glycine. These residues may confer a chain end binding function to these modules.

The presence of closely related CBMs (≥60% amino acid sequence identity) on enzymes with diverse functionalities is similar to what was recently reported for CBM35s (*27*). This observation, that CBM6s may confer similar binding properties to their parent enzymes with diverse activities, is not a common feature within CBM-associated carbohydrate active enzymes, as conventionally the CBM specificity mirrors the catalytic module specificity. Such overlapping CBM functions suggest a conserved mechanism of enzyme substrate targeting to regions of the cell wall undergoing active depolymerization that contain heterogeneous polysaccharides or concentrating the enzymes on a common xylose-containing scaffold. This observation will be fundamental for streamlining the characterization of these functionally related targets in the future.

CONCLUSIONS

The increasing awareness and prevalence of CBM6 modules in known plant cell wall hydrolases underpins the importance of these protein modules in the degradation of complex, recalcitrant polysaccharides. The deployment of bioinformatic analyses to predict CBM6 specificity, or at least identify modules that are likely to recognize novel ligands, has the potential to make a significant contribution to our understanding of the plant cell wall degradation process. In this study, we have identified a CBM6 which displays a novel specificity in that it recognizes a terminal xylose that can be either α- or β-linked to other sugars. It is intriguing, therefore, that a CBM6 with xylose specificity is linked to a GH27 α-galactosidase as polysaccharides containing α-galactosyl residues, such as galactomannan, are not known to contain xylose residues. This observation is not restricted to this example as phylogenetic analysis indicates that other CBM6s contain the "xylose-binding" motif. This unusual arrangement of functional modules within a glycoside hydrolase with specificities that appear uncoupled suggests two possibilities: (1) components of the plant cell wall contain terminal Xyl and αGal residues, and (2) polysaccharides containing large numbers of terminal xylose residues (i.e., as that which occurs in xyloglucan) are in the proximity of polymers, containing α-1,6-Gal side chains

(i.e., galactomannan and galactoglucomannan) and other diverse sugars and linkages. Further investigation into the biological function of the enzymes within this CBM6-enriched genomic locus, and the identification of new CBM6s using the predictive model presented here, will help to crystallize our understanding of the growing repertoire of expanding CBM functions.

SUPPORTING INFORMATION AVAILABLE

Primer sequences, ITC data, enzyme kinetics, a list of enzyme substrates, and *C. cellulolyticum* genome data. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: Fine tuning polysaccharide recognition. *Biochem. J. 382*, 769–782.
2. Gregg, K. J., Finn, R., Abbott, D. W., and Boraston, A. B. (2008) Divergent modes of glycan recognition by a new family of carbohydrate-binding modules. *J. Biol. Chem. 283*, 12604–12613.
3. van Bueren, A. L., Higgins, M., Wang, D., Burke, R. D., and Boraston, A. B. (2007) Identification and structural basis of binding to host lung glycogen by streptococcal virulence factors. *Nat. Struct. Mol. Biol. 14*, 76–84.
4. Boraston, A. B., Ficko-Blean, E., and Healey, M. (2007) Carbohydrate recognition by a large sialidase toxin from *Clostridium perfringens*. *Biochemistry 46*, 11352–11360.
5. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2008) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Res. 37*, D233–D238.
6. 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.
7. 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.
8. Pires, V. M., Henshaw, J., Prates, J. A., Bolam, D., Ferreira, L. M., Fontes, C. M., Henrissat, B., Planas, A., Gilbert, H. J., and Czjzek, M. (2004) The crystal structure of the family 6 carbohydrate-binding module from *Cellvibrio mixtus* lichenase 5A in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. *J. Biol. Chem. 279*, 21560–21568.
9. Henshaw, J., 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.
10. Liu, Q. P., Yuan, H., Bennett, E. P., Levery, S. B., Nudelman, E., Spence, J., Pietz, G., Saunders, K., White, T., Olsson, M. L., Henrissat, B., Sulzenbacher, G., and Clausen, H. (2008) Identification of a GH110 subfamily of α-1,3-galactosidases: Novel enzymes for removal of the α-3Gal xenotransplantation antigen. *J. Biol. Chem. 283*, 8545–8554.
11. Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J., and Williamson, M. P. (2000) The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. *J. Biol. Chem. 275*, 41137–41142.
12. Edgar, R. C. (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf. 5*, 113.
13. 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.
14. 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.
15. Boraston, A. B., Warren, R. A., and Kilburn, D. G. (2001) Glycosylation by *Pichia pastoris* decreases the affinity of a family 2a carbohydrate-binding module from *Cellulomonas fimi*: A functional and mutational analysis. *Biochem. J. 358*, 423–430.
16. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem. 179*, 131–137.
17. 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.
18. Read, R. J. (2001) Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr. D57*, 1373–1382.
19. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.
20. 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.
21. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol. 6*, 458–463.
22. Brunger, A. T. (1992) Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature 355*, 472–475.
23. Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) SFCHECK: A unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D55*, 191–205.
24. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr. 26*, 283–291.
25. Michel, G. B. T., Kloareg, B., and Czjzek, M. (2009) The family 6 carbohydrate-binding modules have coevolved with their appended catalytic modules toward similar substrate specificity. *Glycobiology 19*, 615–623.
26. 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.
27. 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.
28. Chen, Y. Q., and Wenger, D. A. (1993) Galactocerebrosidase from human urine: Purification and partial characterization. *Biochim. Biophys. Acta 1179*, 53–61.
29. 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.
30. Read, R. J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A42*, 140–149.