Crystal Structures of the Family 9 Carbohydrate-Binding Module from *Thermotoga maritima* Xylanase 10A in Native and Ligand-Bound Forms[†,‡]

Valerie Notenboom,[§,||] Alisdair B. Boraston,[§,⊥] Douglas G. Kilburn,*[§,⊥] and David R. Rose[§,||]

Protein Engineering Network of Centres of Excellence, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada, and Department of Microbiology and Immunology and The Biotechnology Laboratory, University of British Columbia, Vancouver, British Columbia, V6T 1Z3 Canada

Received January 26, 2001

ABSTRACT: The C-terminal module of the thermostable *Thermotoga maritima* xylanase 10A (CBM9-2) is a family 9 carbohydrate-binding module that binds to amorphous and crystalline cellulose and a range of soluble di- and monosaccharides as well as to cello and xylo oligomers of different degrees of polymerization [Boraston, A. B., Creagh, A. L., Alam, Md. M., Kormos, J. M., Tomme, P., Haynes, C. A., Warren, R. A. J., and Kilburn, D. G. (2001) *Biochemistry 40*, 6240−6247]. The crystal structure of CBM9-2 has been determined by the multiwavelength anomalous dispersion method to 1.9 Å resolution. CBM9-2 assumes a β-sandwich fold and contains three metal binding sites. The bound metal atoms, which are most likely calcium cations, are in an octahedral coordination. The crystal structures of CBM9-2 in complex with glucose and cellobiose were also determined in order to identify the sugar-binding site and provide insight into the structural basis for sugar binding by CBM9-2. The sugar-binding site is a solvent-exposed slot sufficient in depth, width, and length to accommodate a disaccharide. Two tryptophan residues are stacked together on the surface of the protein forming the sugar-binding site. From the complex structures with glucose and cellobiose, it was inferred that CBM9-2 binds exclusively to the reducing end of mono-, di-, and oligosaccharides with an intricate hydrogen-bonding network involving mainly charged residues, as well as stacking interactions by Trp175 and Trp71. The binding interactions are limited to disaccharides as was expected from calorimetric data. Comparison of the glucose and cellobiose complexes revealed surprising differences in binding of these two substrates by CBM9-2. Cellobiose was found to bind in a distinct orientation from glucose, while still maintaining optimal stacking and electrostatic interactions with the reducing end sugar.

Microbial degradation of cellulose and hemicellulose found in plant biomass is achieved by the activity of diverse polysaccharolytic enzyme systems (*2*, *3*). The majority of polysaccharidases produced by cellulolytic organisms have modular structures comprising a catalytic module and one or more ancillary modules (*4*). The most common type of ancillary module is the carbohydrate-binding module (CBM),[1] some of which were previously called cellulose-binding domains (*3*). In many cases the presence of the CBM enhances the hydrolytic activity of an enzyme. It is thought to do so by increasing the concentration of the enzyme on the substrate (*5*), though there is some evidence for nonhydrolytic disruption of cellulose by a CBM (*6*).

Structural studies of CBMs have allowed insights into the mechanisms of substrate binding. The binding sites of the CBMs for which structures have been determined are of two general types: flat surfaces comprising predominantly aromatic residues and extended shallow grooves.

The family 1 CBMs from *Trichoderma reesei* (*7*, *8*), the family 2a CBM from xylanase 10A of *Cellulomonas fimi* (*9*), the family 3 CBM from CipC of *Clostridium thermocellum* (*10*), the family 5 CBM from cellulase 5A of *Erwinia crysanthemi* (*11*), and the family 10 CBM from *Pseudomonas fluorescens* xylanase 10A (*12*) are all β-sheet polypeptides. The binding faces of CBMs from these families, identified by site-directed mutagenesis (*13*−*18*), are flat faces and composed mainly of tryptophan and tyrosine residues. The ligand of these CBMs, crystalline cellulose, is a regular array of cellulose chains with a staircase organization (*19*). The binding faces of these examples of family 1, 2a, 3, and 5 CBMs seem well suited to bind the flat surfaces of cellulose crystals. In the case of CBM2a, thermodynamic data have suggested that dehydration of the aromatic binding face provides an entropic driving force for binding (*20*). Beyond

[†]This work was supported by a grant from the Protein Engineering Network of Centres of Excellence.

[‡] PDB file codes: 1I82, 1I8A, and 1I8U.

* To whom correspondence should be addressed: The Biotechnology Laboratory, University of British Columbia, 237-6174 University Blvd., Vancouver, BC, Canada V6T 1Z3. Tel: (604) 822-4182. Fax: (604) 822-2114. E-mail: kilburn@interchange.ubc.ca.

[§] Protein Engineering Network of Centres of Excellence.

[||] Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto.

[⊥] Department of Microbiology and Immunology and the Biotechnology Laboratory, University of British Columbia.

[1] Abbreviations: CBM, carbohydrate-binding module; IPTG, isopropyl β-D-thiogalactopyranoside; PEG, poly(ethylene glycol); MAD, multiwavelength anomalous dispersion; F_o and F_c, observed and calculated X-ray structure factor amplitude; GlcNAc, *N*-acetylglucosamine; *R*-factor, X-ray residual factor; *B*-factor, X-ray temperature factor.

10.1021/bi0101704 CCC: $20.00
Published on Web 05/05/2001

this, other protein−cellulose interactions are speculative at best.

The tandem N-terminal family 4 CBMs (CBM4-1 and CBM4-2) from cellulase 9B of *C. fimi* (*21*) and the internal family 2b CBM (CBM2b-1) from xylanase 11b of *C. fimi* (*22*) bind soluble β-1,4-glucan and β-1,4-xylan chains, respectively. The binding sites of these CBMs, identified by NMR spectroscopy, are shallow grooves sufficient in width and length to accommodate individual glycan chains up to five sugar units in length (but no smaller than three). As with most carbohydrate-binding proteins, the involvement of aromatic residues in ligand binding is implicated; however, in contrast to CBM2a, thermodynamic studies of CBM4-1 and CBM2b-1 implied a predominance of polar and van der Waals interactions with the ligand (*22*, *23*). The crystal structure of the family 22 CBM from *C. thermocellum* xylanase 10B also revealed a shallow cleft, lined with tryptophan and tyrosine residues, long enough to accommodate up to four sugar moieties (*24*).

The C-terminal family 9 CBM (CBM9-2) from xylanase 10A *of Thermotoga maritima* is an extremely thermostable module (*25*) with the unique ability to bind tightly to crystalline cellulose, amorphous cellulose, and disaccharides (*1*). The thermostability and unprecedented capacity to bind all cellulose allomorphs, in addition to many types of small soluble sugars, make it an interesting candidate for structural studies. Here the crystal structure of CBM9-2 in native and ligand-bound forms to 1.9 Å resolution is described. For the first time we can describe the specific intermolecular contacts that mediate the binding of a CBM with its ligand. The accompanying paper characterizes the binding of this CBM.

MATERIALS AND METHODS

Materials. All chemicals and buffer reagents used were purchased from Sigma (St. Louis, MO) unless otherwise specified. Cellotetraose was from Megazyme (Dublin, Ireland, U.K.).

Bacterial Strains and Plasmid. T. maritima CBM9-2 (*1*) was produced in *Escherichia coli* strains JM101 or the BL21-derived B834(DE3) for selenomethionine incorporation (Novagen, Madison, WI). The plasmid used was pET28a (Novagen, Milwaukee, MI).

Media and Growth Conditions. E. coli cultures were grown routinely in TB or TYP medium (*26*) unless otherwise stated at 37 °C or at 30 °C for production of protein. The medium contained 50 mg L^{-1} kanamycin of culture for strains carrying pET28 plasmids.

Production and Purification of Selenomethionine Containing CBM9-2. Unlabeled CBM9-2 was produced and purified as described previously (*1*). The pET28a plasmid containing the CBM9-2 construct was transformed into competent *E. coli* B834(DE3) methionine auxotrophs by heat shock. A 15 mL culture was grown on enriched minimal medium (2× M9 salts supplemented with all regular amino acids to 40 mg/L, vitamins, riboflavin, niacinamide, pyridoxine monohydrochloride, and thiamin to 1 mg/L, 0.4% glucose, 25 mg/L $FeSO_4$, and 2 mM $MgSO_4$) overnight at 37 °C with 50 mg/L kanamycin. Then 5 mL of this culture was added to 1 L of medium as above, with methionine substituted by seleno-L-methionine at 100 mg L^{-1}, and allowed to grow to an OD_{600} of 0.7. If the doubling time was found to be longer than 2 h, then an additional 1 mL of a 65:35 selenomethionine:methionine culture in stationary phase was added. IPTG was added to 0.3 mM to induce protein production, and cells were harvested after 5 h by centrifugation. CBM9-2 was purified as described previously (*1*) and stored at 30 mg mL^{-1} at 4 °C. The final selenomethionine protein yield from this 1 L culture was approximately 80 mg.

Crystallization of Native and Derivative CBM9-2. Tetragonal crystals were grown using the vapor diffusion technique from hanging and sitting drops in ~15% PEG 8000 and 0.2 mM sodium acetate, pH 4.5. Long thin needle crystals generally appeared immediately upon mixing the protein and reservoir solutions if either the protein concentration or the PEG concentration was too high. Optimized precipitant and protein concentration combinations did produce three-dimensional crystals that appeared after 1 week or more at room temperatures, with dimensions reaching 0.2 × 0.2 × 0.5−1.5 mm. The crystals are of space group $P4_12_12$ with cell dimensions $a = b = 56.8$ Å and $c = 123.1$ Å, allowing for 1 molecule per asymmetric unit with a solvent content of ~43%. The selenomethionine-substituted protein crystallized under conditions identical to those of the wild-type protein. Crystals of CBM9-2 complexed with glucose and cellobiose were produced by supplementing the reservoir solution with 50−100 mM substrate and adjusting precipitant concentrations.

Data Collection and Processing. Prior to data collection crystals were bathed in artificial mother liquor with 30% glycerol added as cryoprotectant and quickly frozen in a stream of nitrogen gas at 100 K in a rayon fiber loop. Data sets from a native and selenomethionine-labeled crystal were collected in-house on a Mar345 image plate using Osmic mirror focused Cu Kα X-rays, generated from a rotating anode operating at 100 mA × 50 kV. A total of 90 oscillations of 1° were collected in 5 min exposures. All data were indexed, integrated, and scaled with the Denzo/Scalepack package (*27*). Two peaks could be visualized on a difference Patterson map calculated in the program PHASES at this point (*28*), which suggested that there was significant incorporation of selenium atoms in the protein crystals (not shown). Subsequently, a three-wavelength MAD data set was collected at the Advanced Photon Source (APS) at Argonne National Laboratories at the BioCARS 14-BM-D, on a 1K CCD area detector at 105 mm crystal to detector distance. The absorption edge for selenium was detected by scanning through the theoretical X-ray absorption edge. Three wavelengths were chosen, one at the inflection point or edge of the absorption profile (f'-min, $\lambda = 0.9799$ Å), one at its peak (f''-max, $\lambda = 0.9797$ Å), and one reference set at a remote wavelength ($\lambda = 0.9537$ Å). Data sets were collected to 1.8 Å resolution, and the program package Denzo/Scalepack was used to refine experimental parameters and process data (*27*). In addition, glucose and cellobiose complexes were crystallized and data collected in-house as described above. Experimental statistics are summarized in Table 1.

Determination of Selenium Atom Positions and Phasing. Initially, the program Solve was used in an attempt to locate directly the three expected selenium atoms from the MAD data set alone (*29*). Both possible enantiomorphic space groups were tested, but none of the sites were found. Using the home data sets collected from native and SeMet crystals,

Table 1: Data Collection Statistics from Synchrotron Radiation and Rotating Anode

	(a) Synchrotron Radiation		
	peak	edge	remote
wavelength (Å)	0.9797	0.9799	0.9537
resolution (Å)	1.9	1.8	1.8
observations	318505	305530	318615
unique refl	16435	19085	19135
completeness (%)	99.80	99.00	99.50
I/σ	10.2	9.3	10.4
R_{merge}[a] (%)	6.90	7.20	5.80

	(b) Rotating Anode			
	native	SeMet	glucose	$(Glu)_2$
wavelength (Å)	1.54	1.54	1.54	1.54
resolution (Å)	1.9	1.9	1.8	1.9
observations	90745	119264	158924	272755
unique refl	15478	14908	19725	17513
completeness (%)	97.70	95.20	96.20	99.40
I/σ	8.8	16	10.7	9.5
R_{merge}[a] (%)	6.50	4.30	6.20	7.70

[a] $R_{merge} = \sum_{hkl}\sum_i|I_{hkl} - \overline{I_{hkl}}|/\sum_{hkl}\sum_i\overline{I_{hkl}}$.

Table 2: Structure Refinement and Model Quality Statistics[a]

	native	glucose	cellobiose
max resolution used (Å)	1.90	1.90	1.90
cell [$a = b$, c (Å), SG $P4_12_12$]	56.56, 122.57	56.76, 123.10	56.67, 122.97
R-factor, all data (%)	20.1	19.9	20.7
R-free, all data (%)	23.2	22.1	23.6
no. of protein atoms	1484	1484	1484
no. of substrate atoms	N/A	12	23
no. of solvent atoms	154	161	146
rms deviations			
bond lengths (Å)	0.005	0.005	0.006
bond angles (deg)	1.38	1.39	1.42
dihedral angles (deg)	24.1	23.2	22.7
av B-factor (Å^2)	14.01	11.23	11.04
av B-factor substr (Å^2)	N/A	14.27	14.03

[a] $R_{cryst} = \sum_{hkl}||F_o| - k|F_c||/\sum_{hkl}|F_o|$. $R_T^{free} = \sum_{hkl\subset T}||F_o| - k|F_c||/\sum_{hkl\subset T}|F_o|$.

a combined MAD/MIR routine was attempted by treating the home data set pair as an additional isomorphous heavy atom system. This approach yielded two selenium sites with occupancies of 0.63 and 0.38; the third expected site located at the N-terminus remained elusive and was presumed disordered. The phases calculated from these two positions and intensities to 1.8 Å from the home selenomethionine data set resulted in the modest average figure of merit of 0.42. Density modification using DM in the program suite CCP4 (*30*, *31*), implementing histogram matching and solvent flattening, was used to improve the map quality. The resulting map based on 43% solvent content was of sufficient quality to be completely traced, except for some loop regions. Interestingly, no density beyond C-β was seen in this experimental map for each of the selenomethionine residues, both of which are located on the surface of the protein and were later found to assume multiple conformations.

Model Building and Refinement. The initial solvent-flattened map was manually traced using the graphical interface in O (*32*). Most regions in the electron density were reasonably well defined, with distinct carbonyl oxygens and recognizable side chains in most cases, except for the two methionine residues and some disordered solvent-exposed residues. Three loop regions seemed poorly defined upon

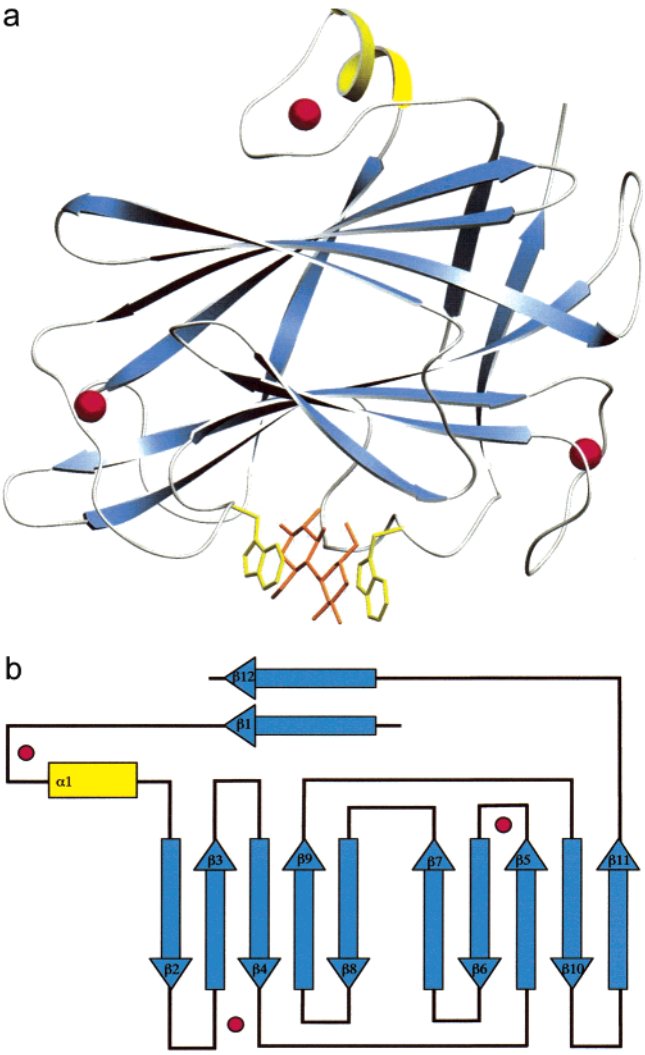
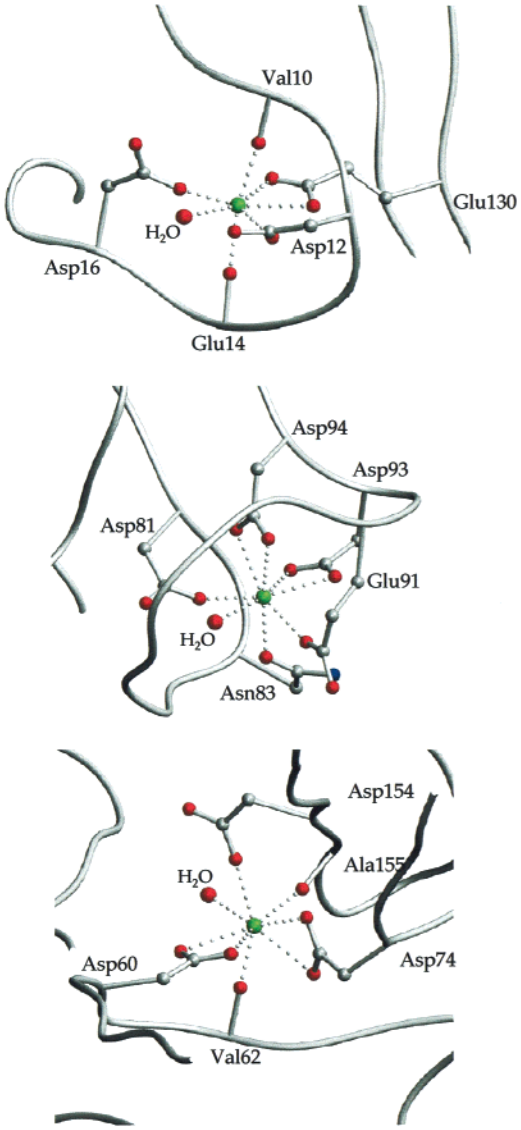


FIGURE 1: Secondary structure of CBM9-2. (a, top) Ribbon diagram of CBM9-2 showing the overall fold of the molecule. Three octahedral metal binding sites, presumed calcium, are drawn in magenta. Two Trp residues, shown in yellow, bind a cellobiose moiety in a sandwich-like fashion. The reducing end of the sugar points inward into the protein. (b, bottom) Topology map depicting the secondary structure organization of CBM9-2, with blue arrows representing β-strands and the yellow box an α-helix. Three calcium ions are shown in magenta, around the major contributing binding elements. Two antiparallel β-sheets of five strands each form the bulk of the molecule, with a parallel sheet formed by the N- and C-termini on the outside.

first inspection and were omitted from the polypeptide trace. Closer inspection, however, revealed that these loops included unanticipated octahedrally coordinated metal binding sites and were, in fact, well ordered. A total of 188 residues were fitted, as well as three metal ions, which were initially treated as water molecules because their identity was unknown. This model underwent rigid body refinement in the program CNS (*33*), implementing bulk solvent correction and a maximum likelihood refinement target using structure factor amplitudes, against the 1.9 Å wild-type data collected in-house with a random 10% of reflections set aside for cross-validation purposes (*33*, *34*). Visual inspection of $2|F_o| - |F_c|$ (1σ) and $|F_o| - |F_c|$ ($+3\sigma$, -3σ) electron density with calculated model phases allowed for some manual adjustments of poorly fitted regions. The three metal ions were assumed to be Ca^{2+} cations from here on, based on the



residue	atom	distance to Ca^2
Asp16	Oδ	2.39 Å
Val10	O	2.37 Å
Glu14	O	2.37 Å
Asp12	Oδ	2.43 Å
Glu130	Oε1	2.62 Å
Glu130	Oε2	2.64 Å
H_2O	O	2.40 Å
Asp 81	Oδ	2.51 Å
Asp 94	Oδ1	2.47 Å
Asp 94	Oδ2	2.80 Å
Asp 93	Oδ1	2.44 Å
Asp 93	Oδ2	2.54 Å
Glu 91	Oε	2.49 Å
Asn 83	Oδ	2.54 Å
H_2O	O	2.76 Å
Asp 74	Oδ1	2.48 Å
Asp 74	Oδ2	2.80 Å
Val 62	O	2.35 Å
Asp 60	Oδ1	2.43 Å
Asp 60	Oδ2	2.57 Å
Asp 154	Oδ	2.30 Å
Ala 155	O	2.53 Å
H_2O	O	2.63 Å

FIGURE 2: Three calcium binding sites in CBM9-2. Glu130 (a, top), Asp94 and Asp93 (b, middle), and Asp74 and Asp60 (c, bottom) bind calcium in a bifurcated fashion or through a delocalized charge on the residue.

octahedral geometry of the coordination and the size of the difference peaks (~35σ). The observation that exclusively oxygen atoms were found to coordinate the metal ions at a distance of 2.3−2.7 Å strongly supports this contention. In addition, there is precedence of Ca^{2+} ions present in thermally stable proteins in general (*35*, *36*) and CBMs in particular (*10*, *24*, *37*).

At this stage, the model was subjected to simulated annealing starting from 3000 K in steps of 25 K, followed by positional refinement and grouped thermal parameter refinement. Several cycles of manual water building were performed at this point. Upon convergence of *R*-free, individual *B*-factors were refined. Structures of the cellobiose and glucose complexed proteins were determined by molecular replacement of the unliganded model (without water molecules) against the diffraction data of the complexed crystals. The substrate was located and built into 3σ $|F_o| - |F_c|$ difference density, and these coordinates were included in the model. This model then was refined as described above, except simulated annealing was commenced at 1500 K.

RESULTS AND DISCUSSION

Structure of CBM9-2. The native protein model consists of 1484 non-hydrogen atoms, 3 calcium atoms, and 154 ordered water molecules. Structure refinement statistics are summarized in Table 2.

A ribbon diagram of the crystal structure (of the cellobiose-bound form) and a topology representation are shown in Figure 1. The three-dimensional structure consists of two twisted antiparallel β-sheets each made up of five β-strands ($\beta2-\beta3-\beta4-\beta9-\beta8$ and $\beta7-\beta6-\beta5-\beta10-\beta11$) to form a β-sandwich fold, flanked by an additional two-stranded parallel β-sheet ($\beta1-\beta12$) comprising the N- and C-termini. A short α-helix starts off a loop connecting $\beta1$ and $\beta2$, which is stabilized by a Ca^{2+} ion. Two other Ca^{2+} ions are also found at the periphery of the protein, one stabilizing a loop connecting strands $\beta5$ and $\beta6$ and one sandwiched between $\beta3$ and $\beta4$. This last site also forms contacts with $\beta10$ and $\beta11$, thus connecting the two β-sheets. The coordination of this site and the two other Ca^{2+} sites, depicted in Figure 2, exhibits close to perfect octahedral geometry, and all involve oxygens originating from a combination of main-chain

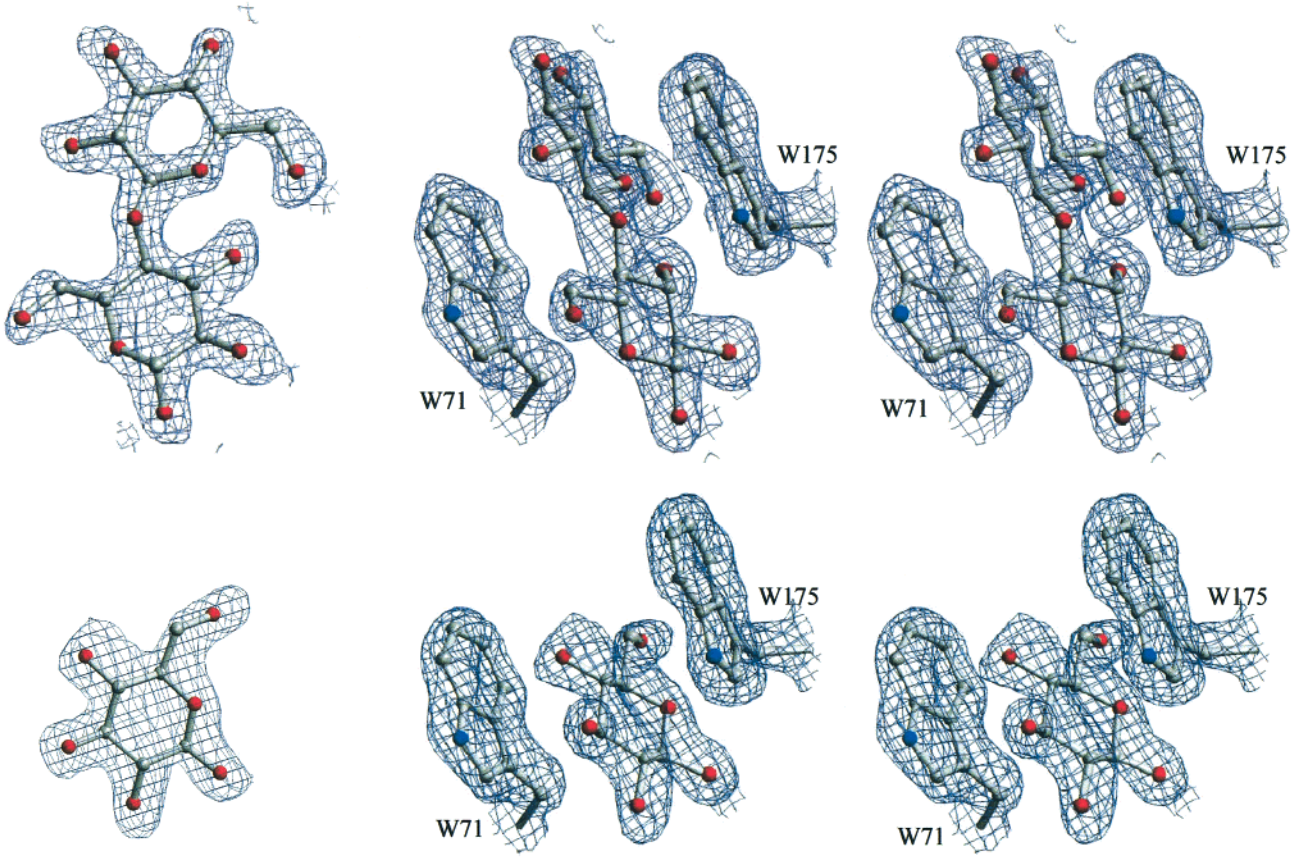


FIGURE 3: $2F_o - F_c$ electron density contoured at 1σ from the cellobiose (a, top) and glucose (b, bottom) complexes, using phases from the protein model only. The ligands could be unambiguously modeled into the density with one clear orientation. In both complexes two Trp residues (right, stereo) stack the sugar in a sandwich.

carbonyls, waters and side-chain carbonyls. Several side-chain carbonyls [Glu130 in (a), Asp93 and 94 in (b), and Asp74 and Asp60 in (c)] appear to form interactions with the calcium ion through their deprotonated resonance form, with the negative charge distributed over both oxygens. None of the calcium sites is in direct contact with the binding site, suggesting that the observed calcium binding sites most likely serve a structural role, stabilizing the three-dimensional organization of the polypeptide thereby contributing to the thermostability of the protein.

CBM9-2 has been shown to bind amorphous and crystalline cellulose, xylan, and a range of small soluble sugars including mono- and disaccharides and cello- and xylo-oligosaccharides. Cellobiose bound with the highest affinity of the substrates tested, and cellooligosaccharides with a degree of polymerization greater than 2 did not bind with higher affinity, suggesting that a disaccharide might fully occupy the binding region (*1*). Two tryptophan residues, Trp175 and Trp71, were found in parallel orientation at the base of the second twisted β-sheet such that both residues could interact with a carbohydrate placed between these residues. The stacking of aromatic amino acid side chains against the apolar surfaces of carbohydrates is common in protein−carbohydrate interactions. For this reason, and the prevalence of planar polar amino acids at the base of the cleft formed by these tryptophans, this was predicted to constitute the binding region of CBM9-2.

Structure of CBM9-2 in Complex with Saccharides. The native protein model was used in rigid body refinement against data of the cellobiose−CBM9-2 cocrystallized complex to optimize the orientation of the model in the unit cell. Difference electron density from $|F_o| - |F_c|$ and $2|F_o| - |F_c|$ (contoured at 3σ and 1σ, respectively) clearly revealed a single cellobiose molecule sandwiched between Trp71 and Trp175 as anticipated (Figure 3a). A cellobiose molecule was easily fitted into the density in a single orientation, with the reducing end of the sugar pointed into the protein. An extensive network of hydrogen-bonding interactions involving all hydroxyl groups and the cyclic oxygen is formed around the reducing end (proximal) sugar. Of particular interest are interactions formed with the OH1 and OH2 groups of the reducing sugar. Binding experiments have shown that substrates with modified functional groups at these positions (such as 1-*O*-methylglucopyranosides, GlcNAc, or mannose) (*1*) are unable to compete with cellulose for binding to CBM9-2, suggesting that interactions at these positions are extremely important. Indeed, the anomeric hydroxyl group (OH1) forms three hydrogen bonds with Asn172, Arg161, and Gln151; OH2 also forms two interactions, with Glu77 and an ordered water molecule, which is traced back to the cyclic nitrogen of Trp175. A schematic representation of interactions between cellobiose and CBM9-2 is shown in Figure 4a. The tryptophan residues provide planar hydrophobic stacking interactions spanning the whole disaccharide, centering on the glycosidic bond and seemingly involving both sugar rings equally. The nonreducing or distal saccharide receives fewer hydrogen bonds from the protein than the proximal sugar; its OH6 hydroxyl forms a hydrogen bond with Arg98.

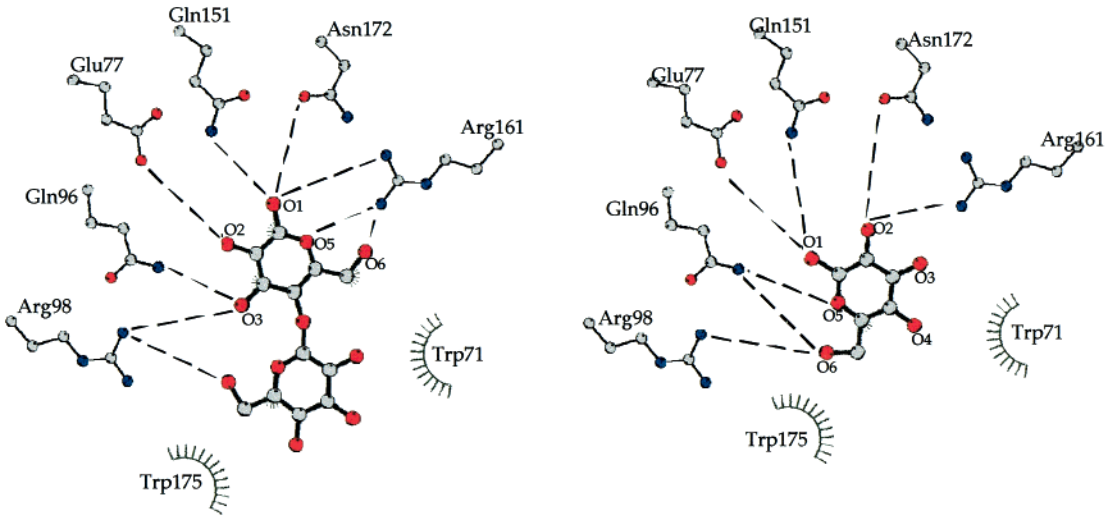


FIGURE 4: Schematic representations of cellobiose (a, left) and glucose (b, right) binding.

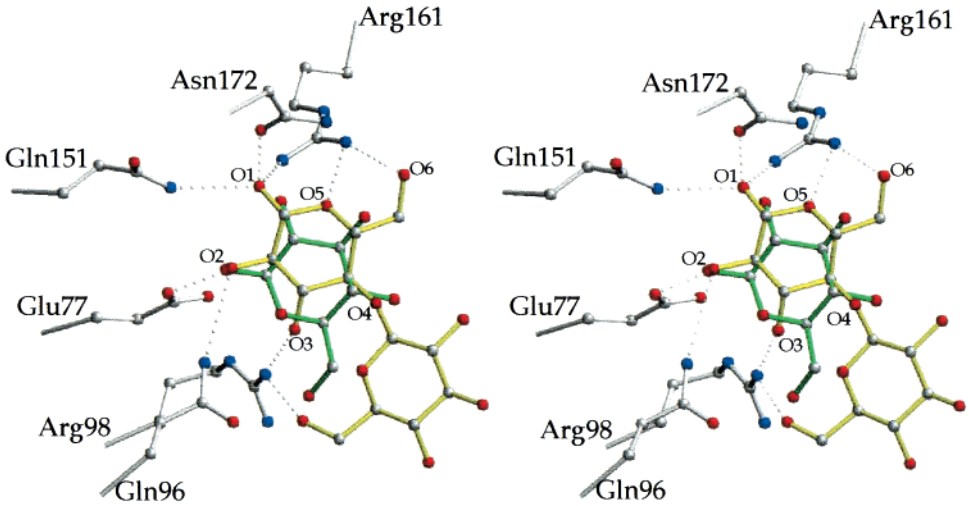


FIGURE 5: Stereo representation of the binding site interactions in CBM9-2 with cellobiose (yellow) and glucose (green). Only the hydrogen-bonding interactions with cellobiose are shown, for clarity (see Figure 4 for glucose interactions).

Calorimetric analysis of binding indicated that glucose may undergo a different mechanism of binding compared to the disaccharides tested (*1*). Thus it was of interest to investigate the crystal structure of CBM9-2 in complex with glucose. The structure was acquired as described above. Again, difference electron density using the unbound structure as a starting model clearly revealed a glucose moiety bound in the binding pocket of the protein, allowing for straightforward building of the sugar into the structure (Figure 3b). Surprisingly, glucose is bound in the pocket in a markedly different orientation from cellobiose. Although the reducing end of the sugar remains oriented inward, none of the specific hydrogen-bonding interactions is maintained, but they are replaced by a set of new interactions (Figure 4b). Roughly, the glucose molecule, compared to the proximal ring of cellobiose, has rotated 180° around the axis defined by C1−C4 and an additional 60° in the plane of the ring. This not only maintains proper stacking interactions with the Trp "sandwich" but allows for a crude overlap of functional groups in space, and therefore an almost equal number of hydrogen bonds are formed with the protein as seen with cellobiose (Figure 5). Interestingly, OH1 and OH2 in glucose

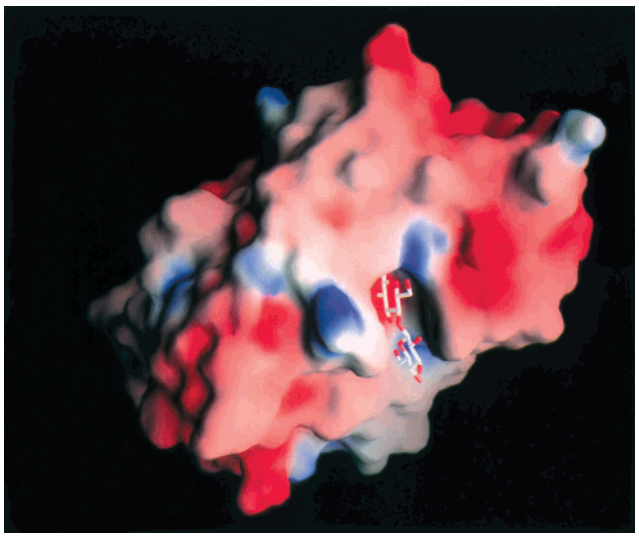


FIGURE 6: Surface representation of CBM9-2 in complex with cellobiose, where red areas represent negative surface potential and blue regions depict positive surface potential. The binding region resembles a blind canyon which can accommodate two units from one end of a polysaccharide chain. Internal units from a polysaccharide chain are not able to bind in such a pocket.

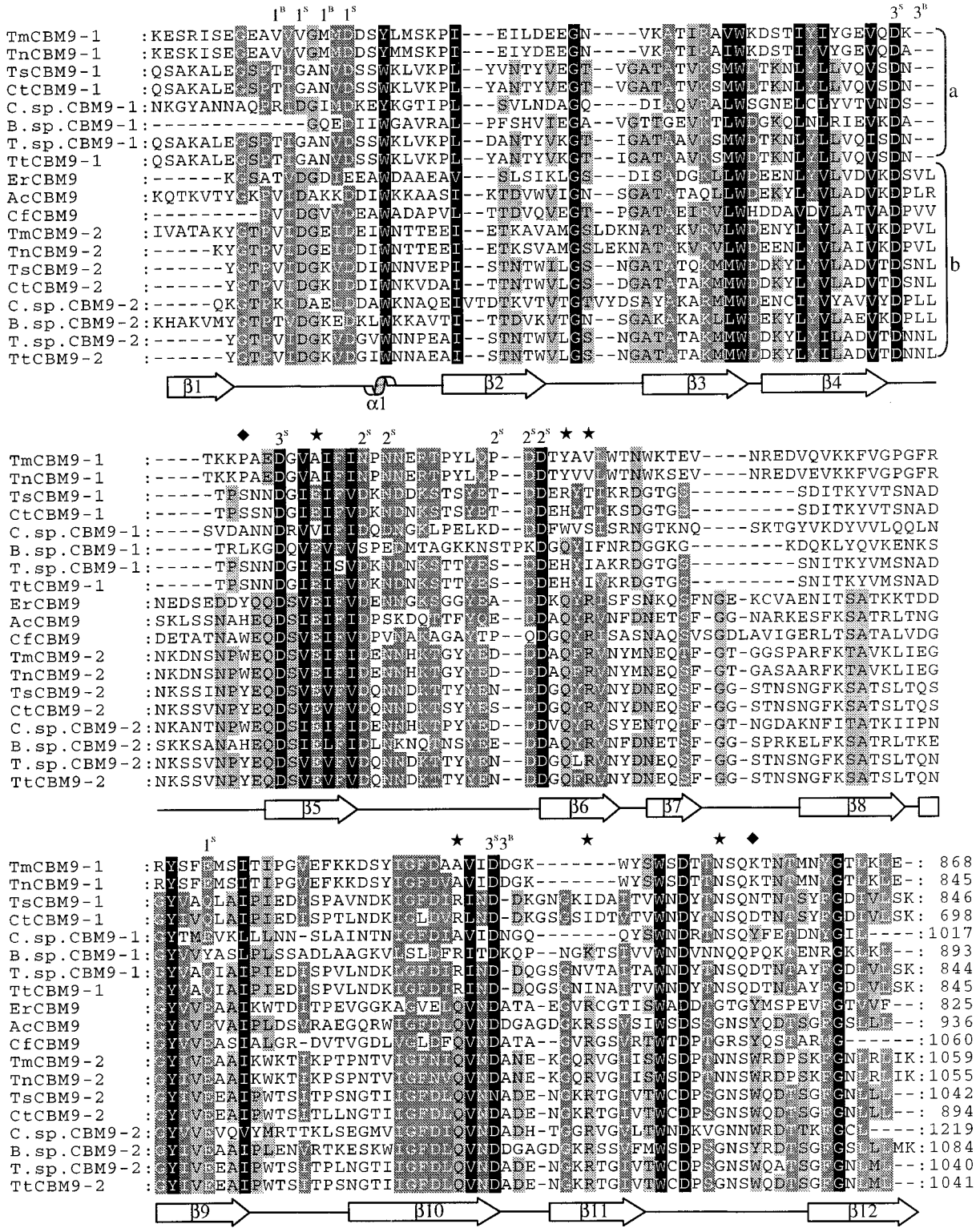


FIGURE 7: Alignment of the protein sequences of the members of family 9 CBMs. Sequences labeled −1 or −2 indicate the N-terminal and C-terminal repeat of a tandem, respectively. Prefixes refer to organisms as follows: Tm sequences from *Thermotoga maritima* MSB8 (accession no. S61311); Tn, *Thermotoga neopolitana* (Z46945); C.sp., *Caldicellulosiruptor* sp. (AF036925); Tt, *Thermoanaerobacterium thermosulfurigenes* (U50592); Ts, *Thermoanaerobacterium saccharolyticum* (M97882); Ct, *Clostridium thermocellum* (S41797); T.sp., *Thermoanaerobacterium* sp. (Z46264); Ac, *Aeromonas caviae* (AB019244); B.sp., *Bacillus* sp. (AJ006645); Er, *Eubacterium ruminatum* (D63938); Cf, *Cellulomonas fimi* (Z50866). The secondary structure of CBM9-2 is shown beneath with arrows indicating sheet, curls indicating helix, and lines indicating coil. ◆ denotes aromatic residues in CBM9-2 involved in ligand binding; ★ denotes residues involved in H-bonding with the ligand. Numbered columns show residues involved in binding metal ions; superscript letters indicate whether the backbone carbonyl (B) or side chain (S) interacts with the metal. Bracketed sequences labeled a or b indicate subfamilies a and b. Alignment was prepared using ClustalW (*45*).

overlap almost perfectly with OH2 and OH1 of cellobiose, respectively, supporting the importance of these two hydroxyl groups as discussed above, but not in the conserved fashion that was anticipated.

On the assumption that only one of the above modes of binding corresponds to that of "natural" substrates, we investigated two larger saccharides, cellotetraose and xylopentose, in complex with CBM9-2 in the crystal. In both of these structures, only the first two sugars had visible electron density, in agreement with the thermodynamic data indicating that all the essential interactions were made by the two reducing end sugars (*1*). The mode of binding of the proximal sugar was exactly that seen in the glucose structure described above, with all of the interactions shown in Figure 4b maintained (except where the C6 substituent is absent in xylose), suggesting that the glucose, rather than the cellobiose, orientation was the normal binding mode. The second, distal sugar in both of the polysaccharide structures forms the same Trp stacking interaction seen in the cellobiose structure. It is not clear why cellobiose forms an anomalous interaction with the binding site at the proximal sugar position. Interestingly, the second sugar unit of xylopentaose was rotated 180° about the glycosidic bond relative to the first, rather than the expected 120° found in xylan chains (*38*). The $\Delta\Delta G$ of xylobiose relative to xylose was ~2.5 kJ/mol less favorable than the $\Delta\Delta G$ of cellobiose relative to glucose (*1*). The reduced energetic contribution of the second sugar in xylobiose may result from the additional energetic cost of imposing conformation on the glycosidic bond of the xylooligosaccharide.

CBM9-2 Binds to the Reducing Ends of Carbohydrates. The results presented here demonstrate that CBM9-2 recognizes the reducing ends of small carbohydrates. It has also been shown that altering the reducing ends of polysaccharides by $NaBH_4$ reduction destroys the binding of CBM9-2 to these polysaccharides (*1*). The shape of the CBM9-2 binding site reveals the structural basis of this specificity. The carbohydrate-binding site of CBM9-2 can be described as a "blind canyon", a groove in the protein surface that is blocked at one end (Figure 6). The reducing ends of small sugars point into the blocked end of this canyon and make several hydrogen bonds with the protein. Such a conformation cannot accommodate binding internally on a polysaccharide chain or to side chains where the reducing ends of the sugars are involved in the glycosidic linkages.

Tandem Family 9 Carbohydrate-Binding Modules. Family 9 of carbohydrate-binding modules contains approximately 19 putative CBM sequences from 11 organisms (Figure 7) (*39*). These sequences frequently occur as tandems in polysaccharide-degrading enzymes, particularly in the thermophilic and hyperthermophilic organisms. *T. maritima* xylanase 10A contains such a tandem repeat of family 9 CBMs. To further elucidate the function of this doublet, we have investigated the binding properties of the first family 9 CBM in xylanase 10A from *T. maritima*, CBM9-1, and the tandem of both family 9 CBMs, CBM9-1.2.

CBM9-1 did not bind in an affinity electrophoresis assay to any of the polysaccharides tested (barley β-glucan, birchwood xylan, pectic galactan, xyloglucan, or arabinoxylan) nor did it bind to phosphoric acid swollen cellulose, bacterial microcrystalline cellulose, or insoluble oat spelt xylan. The affinity of CBM9-1.2 for barley β-glucan, PASA, and cellooligosaccharides was unchanged relative to CBM9-2 (results not shown). This indicates that CBM9-1 does not have any obvious binding function either alone or in the context of the second binding module. The lack of binding activity of CBM9-1 may be explained by the lack of conservation of most residues that interact with ligand in CBM9-2 (Figure 7). Particularly noticeable is the absence of the two aromatic residues involved in "sandwiching" the sugar. Generation of a structural model (*40*) of CBM9-1 using CBM9-2 as a template implies that there is no binding cleft (not shown).

We propose that carbohydrate-binding module family 9 contains two subfamilies: 9a and 9b. Family 9a, which includes CBM9-1, comprises the N-terminal modules of tandem family 9 modules. Members of family 9a are easily identified by deletions in loop regions between β-sheets 4 and 5, a loop making up part of the CBM9-2 binding site, and between β-sheets 7 and 8 of CBM9-2 (Figure 7). The similarity of CBM9-1 to the other family 9a modules suggests a comparable lack of carbohydrate-binding function throughout this subfamily.

Some members of CBM family 22 found in thermophilic xylanases have been shown to protect the catalytic module from thermal inactivation (*41*−*43*). This family has recently been shown to comprise xylan-binding modules but, similar to family 9, also contains modules with no apparent binding function (*24*, *44*, *45*). It is possible that CBM9-1 has a similar thermostabilizing effect. However, no thermostabilizing role has been unequivocally demonstrated as the sole function for any ancillary module. Thus, one cannot ignore alternatives such as the potential importance of the spacing between the catalytic module and CBM9-2 imposed by CBM9-1; CBM9-1 may perform the role of a linker.

Family 9b, which includes CBM9-2, comprises the C-terminal modules of tandem family 9 modules and the family 9 modules that appear alone. The conservation of structural components (residues found in the secondary structure and metal binding sites) and functional components (residues involved in substrate binding) of this subfamily suggests that structure and carbohydrate-binding function is also relatively well conserved among these modules.

REFERENCES

1. Boraston, A. B., Creagh, A. L., Alam, M., Kormos, J. M., Tomme, P., Haynes, C. A., Warren, R. A. J., and Kilburn, D. G. (2001) *Biochemistry 40*, 6240−6247.
2. Beguin, P., and Aubert, J. P. (1994) *FEMS Microbiol. Rev. 13*, 25−58.
3. Tomme, P., Warren, R. A., and Gilkes, N. R. (1995) *Adv. Microb. Physiol. 37*, 1−81.
4. Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. J., and Warren, R. A. (1991) *Microbiol. Rev. 55*, 303−315.
5. 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) *Biochem. J. 331*, 775−781.
6. Din, N., Gilkes, N. R., Tekant, B., Miller, R. C., Warren, R. A., and Kilburn, D. G. (1991) *Bio/Technology 9*, 1096−1099.
7. Kraulis, J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989) *Biochemistry 28*, 7241−7257.
8. Mattinen, M. L., Linder, M., Drakenberg, T., and Annila, A. (1998) *Eur. J. Biochem. 256*, 279−286.
9. Xu, G. Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E., and Harvey, T. S. (1995) *Biochemistry 34*, 6993−7009.
10. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., and Steitz, T. A. (1996) *EMBO J. 15*, 5739−5751.

11. Brun, E., Johnson, P. E., Creagh, A. L., Tomme, P., Webster, P., Haynes, C. A., and McIntosh, L. P. (2000) *Biochemistry 39*, 2445−2458.
12. Raghothama, S., Simpson, P. J., Szabo, L., Nagy, T., Gilbert, H. J., and Williamson, M. P. (2000) *Biochemistry 39*, 978−984.
13. McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A., and Kilburn, D. G. (2000) *Protein Eng. 13*, 801−809.
14. Goldstein, M. A., and Doi, R. H. (1994) *J. Bacteriol. 176*, 7328−7334.
15. Linder, M., Mattinen, M. L., Kontteli, M., Lindeberg, G., Stahlberg, J., Drakenberg, T., Reinikainen, T., Pettersson, G., and Annila, A. (1995) *Protein Sci. 4*, 1056−1064.
16. Nagy, T., Simpson, P., Williamson, M. P., Hazlewood, G. P., Gilbert, H. J., and Orosz, L. (1998) *FEBS Lett. 429*, 312−316.
17. Simpson, H. D., and Barras, F. (1999) *J. Bacteriol. 181*, 4611−4616.
18. Ponyi, T., Szabo, L., Nagy, T., Orosz, L., Simpson, P. J., Williamson, M. P., and Gilbert, H. J. (2000) *Biochemistry 39*, 985−991.
19. Blackwell, J. (1982) in *Cellulose and Other Natural Polymer Systems. Biogenesis, Structure, and Degradation* (Brown, R. M., Jr., Ed.) pp 403−428, Plenum Press, New York.
20. Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G., and Haynes, C. A. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 12229−12234.
21. Johnson, P. E., Joshi, M. D., Tomme, P., Kilburn, D. G., and McIntosh, L. P. (1996) *Biochemistry 35*, 14381−14394.
22. Simpson, P. J., Bolam, D. N., Cooper, A., Ciruela, A., Hazlewood, G. P., Gilbert, H. J., and Williamson, M. P. (1999) *Struct. Fold. Des. 7*, 853−864.
23. Tomme, P., Creagh, A. L., Kilburn, D. G., and Haynes, C. A. (1996) *Biochemistry 35*, 13885−13894.
24. Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J., and Fontes, C. M. (2000) *Biochemistry 39*, 5013−5021.
25. Wassenberg, D., Schurig, H., Liebl, W., and Jaenicke, R. (1997) *Protein Sci. 6*, 1718−1726.
26. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) in *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
27. Carter, C., Jr., and Sweet, R. M. (1997) *Methods Enzymol. 276*, 307−326.
28. Furey, W., and Swaminathan, S. (1990) Proceedings of the American Crystallographic Society Meeting, New Orleans, LA.
29. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 55*, 849−861.
30. Collaborative Computational Project, No. 4 (1994) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 50*, 760−763.
31. Cowtan, K. D., and Zhang, K. Y. (1999) *Prog. Biophys. Mol. Biol. 72*, 245−270.
32. Jones, T. A., Bergdoll, M., and Kjoller, A. (1990) in *Crytallographic and modeling methods in molecular design* (Bugg, C. E., and Ealick, S. E., Eds.) pp 63−72, Springer-Verlag, New York.
33. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 905−921.
34. Kleywegt, G. J., and Brunger, A. T. (1996) *Structure 4*, 897−904.
35. Hwang, K. Y., Song, H. K., Chang, C., Lee, J., Lee, S. Y., Kim, K. K., Choe, S., Sweet, R. M., and Suh, S. W. (1997) *Mol. Cells 7*, 251−258.
36. Daniel, R. M., Toogood, H. S., and Bergquist, P. L. (1996) *Biotechnol. Genet. Eng. Rev. 13*, 51−100.
37. Johnson, P. E., Creagh, A. L., Brun, E., Joe, K., Tomme, P., Haynes, C. A., and McIntosh, L. P. (1998) *Biochemistry 37*, 12772−12781.
38. Atkins, E. D. T. (1992) Three-dimensional structure, interactions and properties of xylan, in *Xylan and Xylanases: Progress in Biotechnology* (Visser, J., Beldman, G., van Kusters, S., and Voragen, A. G. L., Eds.) Vol. 7, pp 39−50, Elsevier, Amsterdam.
39. Coutinho, P. M., and Henrissat, B. (1999) in *Recent advances in carbohydrate bioengineering* (Gilbert, H. J., Davies, G. J., Henrissat, B., and Svensson, B., Eds.) pp 3−12, Royal Society of Chemistry, Cambridge.
40. Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995) *Proteins 23*, 318−326.
41. Fontes, C. M., Hazlewood, G. P., Morag, E., Hall, J., Hirst, B. H., and Gilbert, H. J. (1995) *Biochem. J. 307*, 151−158.
42. Kim, H., Jung, K. H., and Pack, M. Y. (2000) *Appl. Microbiol. Biotechnol. 54*, 521−527.
43. Winterhalter, C., Heinrich, P., Candussio, A., Wich, G., and Liebl, W. (1995) *Mol. Microbiol. 15*, 431−444.
44. Sunna, A., Gibbs, M. D., and Bergquist, P. L. (2000) *Biochem. J. 346*, 583−586.
45. Meissner, K., Wassenberg, D., and Liebl, W. (2000) *Mol. Microbiol. 36*, 898−912.
46. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res. 22*, 4673−4680.

BI0101704