

# Mutational Insights into the Roles of Amino Acid Residues in Ligand Binding for Two Closely Related Family 16 Carbohydrate Binding Modules\*<sup>§</sup>

Received for publication, July 25, 2010, and in revised form, August 13, 2010. Published, JBC Papers in Press, August 25, 2010, DOI 10.1074/jbc.M110.168302

Xiaoyun Su<sup>‡§</sup>, Vinayak Agarwal<sup>¶</sup>, Dylan Dodd<sup>‡§||</sup>, Brian Bae<sup>\*\*</sup>, Roderick I. Mackie<sup>‡§‡‡</sup>, Satish K. Nair<sup>‡§¶\*\*1</sup>, and Isaac K. O. Cann<sup>‡§||‡‡2</sup>

From the <sup>‡</sup>Energy Biosciences Institute, <sup>§</sup>Institute for Genomic Biology, Departments of <sup>||</sup>Microbiology and <sup>\*\*</sup>Biochemistry, <sup>¶</sup>Center for Biophysics and Computational Biology, and <sup>‡‡</sup>Department of Animal Sciences, University of Illinois, Urbana, Illinois 61801

Carbohydrate binding modules (CBMs) are specialized proteins that bind to polysaccharides and oligosaccharides. *Caldanaerobius polysaccharolyticus* Man5ACBM16-1/CBM16-2 bind to glucose-, mannose-, and glucose/mannose-configured substrates. The crystal structures of the two proteins represent the only examples in CBM family 16, and studies that evaluate the roles of amino acid residues in ligand binding in this family are lacking. In this study, we probed the roles of amino acids (selected based on CBM16-1/ligand co-crystal structures) on substrate binding. Two tryptophan (Trp-20 and Trp-125) and two glutamine (Gln-81 and Gln-93) residues are shown to be critical in ligand binding. Additionally, several polar residues that flank the critical residues also contribute to ligand binding. The CBM16-1 Q121E mutation increased affinity for all substrates tested, whereas the Q21G and N97R mutants exhibited decreased substrate affinity. We solved CBM/substrate co-crystal structures to elucidate the molecular basis of the increased substrate binding by CBM16-1 Q121E. The Gln-121, Gln-21, and Asn-97 residues can be manipulated to fine-tune ligand binding by the Man5A CBMs. Surprisingly, none of the eight residues investigated was absolutely conserved in CBM family 16. Thus, the critical residues in the Man5A CBMs are either not essential for substrate binding in the other members of this family or the two CBMs are evolutionarily distinct from the members available in the current protein database. Man5A is dependent on its CBMs for robust activity, and insights from this study should serve to enhance our understanding of the interdependence of its catalytic and substrate binding modules.

Plant cell walls, composed mainly of cellulose, hemicellulose, and lignin, represent one of the most abundant sources of biomass on Earth (1), and microbial glycoside hydrolases play a

critical role in their degradation into oligosaccharides and monosaccharides (2). These enzymes possess highly diversified biochemical activities, a reflection of the complexity of the chemical bonds by which the sugars are linked in plant cell walls.

The glycoside hydrolases commonly occur as modular polypeptides that contain catalytic domains and other modules, including carbohydrate binding modules (CBMs).<sup>3</sup> CBMs exhibit three auxiliary effects that serve to improve the enzymatic functions of the catalytic domain (3). First, CBMs are able to increase the proximity of the catalytic domain to the substrates, thereby increasing the local concentration of the catalytic domain on the substrate (4). Second, CBMs recognize and bind to specific substrates, and this targets the catalytic domain to specific parts or regions of the complex plant cell wall (5, 6). Third, some types of CBMs function to physically disrupt the intermolecular interactions within certain polysaccharide substrates, thus facilitating subsequent hydrolysis by the associated catalytic domain (7, 8).

The carbohydrate binding modules are divided into three types based on their substrate preference as well as their binding mechanisms (3). Type A CBMs have a flat binding platform, composed mainly of aromatic amino acid residues, which enables them to bind to the surface of crystalline cellulose (9). In contrast, the binding cleft of type B CBMs is concave, forming a cleft that allows accommodation of up to six sugar residues in a single polysaccharide chain (10–12). Type C CBMs lack the extended binding cleft that type B CBMs possess and only bind to small sugars such as mono-, di-, or trisaccharides (13).

The large variation in substrate specificities of CBMs suggests that the structures of CBMs have plasticity that can be evolutionarily selected to accommodate different ligands. A single CBM may be able to bind to differently configured substrates, and this has been demonstrated for the *Piromyces equi* NCP1 CBM29-1 and CBM29-2, *Caldanaerobius polysaccharolyticus* Man5A CBM16-1 and CBM16-2, *Clostridium thermocellum* Lic26A-Cel5E CBM11 and CtCel9D-Cel44A CBM44 (5, 12, 14–16). The *P. equi* NCP1 CBM29-2 can bind amorphous cellulose and mannan and also glucomannan.

<sup>3</sup> The abbreviations used are: CBM, carbohydrate binding module; LBG, locust bean gum; CMC-Na, sodium carboxymethyl cellulose; ITC, isothermal titration calorimetry; AGE, affinity gel electrophoresis; KGM, Konjac glucomannan.

\* This work was supported by a grant from the Energy Biosciences Institute and in part by National Institutes of Health Fellowship 1F30DK084726 from NIDDK (to D. D.).

<sup>§</sup> The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Fig. 1.

The atomic coordinates and structure factors (codes 30EA and 30EB) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

<sup>1</sup> To whom correspondence may be addressed. Tel.: 217-333-0641; Fax: 217-244-5858; E-mail: snair@uiuc.edu.

<sup>2</sup> To whom correspondence may be addressed: 1105 Institute for Genomic Biology, 1206 West Gregory Dr., University of Illinois, Urbana, IL 61801. Tel.: 217-333-2090; Fax: 217-333-8286; E-mail: icann@illinois.edu.

## Role of Amino Acids in Family 16 CBM

Similarly, the two family 16 CBMs of *C. polysaccharolyticus* (abbreviated as CBM16-1 and CBM16-2 below) are also able to bind to mannose- and glucose-configured substrates (16). The CBM16 polypeptides constitute a large family of 59 members that are linked to diverse catalytic domains (17–21), suggesting that these CBMs possess distinct binding specificities. Until recently, the residues that are key to ligand binding in this family were unknown. Insights were therefore gained when the crystal structures of the *C. polysaccharolyticus* Man5A CBM16-1 and CBM16-2 were solved by Bae *et al.* (16). Furthermore, the high resolution co-crystal structure of CBM16-1 in complex with mannopentaose and cellopentaose provided a sound basis for a detailed mutational analysis of these two CBMs. The Man5A CBM16-1 and CBM16-2 share high amino acid sequence homology (62% identity and 74% similarity), which forms the basis of their similar spatial structures (16). However, not all of the polar residues predicted to make direct contact with the ligand in CBM16-1 are conserved in CBM16-2. The high amino acid sequence conservation, accompanied by the minor differences in polar residues, as well as the 2-fold difference in binding affinity for these CBMs, provides an excellent model for probing the importance of key residues in binding to differently configured substrates (16).

In this study, each of the key residues of both CBM16-1 and CBM16-2, proposed to interact with ligands by either hydrophobic stacking interaction or forming direct hydrogen bonds, were mutated to alanine. Furthermore, nonconserved polar residues, predicted to be critical to binding, were mutated to their equivalent residues in the counterpart CBM. Subsequently, the relative binding affinities of the wild-type and mutant derivatives were compared by affinity gel electrophoresis (AGE) using sodium carboxymethyl cellulose (CMC-Na), locust bean gum (LBG), and konjac glucomannan, which represent glucose-configured, mannose-configured, and glucose/mannose mixed polysaccharide substrates, respectively. Furthermore, the binding parameters of the different derivatives of CBM16-1 and CBM16-2 were determined by isothermal titration calorimetry using  $\beta$ -1,4-cellopentaose and  $\beta$ -1,4-mannopentaose as substrates. Results from these studies revealed that Trp-20, Trp-125, Gln-81, and Gln-93 in the two CBMs play critical roles, although other polar residues targeted for mutation play moderately important roles in ligand binding. Interestingly, CBM16-1 Q121E increased the binding affinities for all substrates tested, suggesting that this noncritical polar residue can be used to manipulate the binding affinities of the *C. polysaccharolyticus* CBMs.

### EXPERIMENTAL PROCEDURES

**Materials**—The construction of the plasmids expressing the *C. polysaccharolyticus* (formerly *Thermoanaerobacterium polysaccharolyticum* and reclassified by Lee *et al.* (22)) Man5A carbohydrate binding modules CBM16-1 and CBM16-2 were described by Bae *et al.* (16). The *Escherichia coli* JM-109 strain from Promega (Madison, WI) was used for maintaining and propagating all the plasmids in this study. The oligosaccharides, including  $\beta$ -1,4-D-cellopentaose,  $\beta$ -1,4-D-mannopentaose, and the polysaccharide konjac glucomannan, were purchased from Megazyme (Wicklow, Ireland). The other two polysaccharides,

sodium carboxymethyl cellulose and locust bean gum, were from Sigma.

**Site-directed Mutagenesis**—The generation of point mutations for CBM16-1 or CBM16-2 was conducted using the QuikChange site-directed mutagenesis kit according to the manufacturer's instructions (Stratagene, La Jolla, CA). Briefly, 25 ng of plasmid DNA encoding either CBM16-1 or CBM16-2 wild-type was used as template in a PCR amplification. In addition, the reaction mixture contained 250 nM each of the primers, 0.2 mM dNTPs and 1  $\mu$ l of PfuTurbo DNA polymerase. The PCR amplification steps were as follows: 95 °C for 30 s for 1 cycle, then 95 °C for 30 s, 55 °C for 1 min, and 68 °C for 6 min for 18 cycles. The PCR products were visualized by running 5- $\mu$ l aliquots on a 1% agarose gel in TAE buffer. If a band appeared on the gel with the correct size, the rest of the PCR product was digested with DpnI at 37 °C overnight. A 1- $\mu$ l aliquot of the digested PCR product was then used to electroporate *E. coli* JM-109 electrocompetent cells. To select for recombinant *E. coli* cells, the products of the electroporation were spread on LB plates containing 100  $\mu$ g/ml ampicillin and incubated at 37 °C overnight. A single colony from each plate was inoculated in 10 ml of LB medium supplemented with ampicillin and cultured at 37 °C with shaking overnight. The plasmids in the transformants were purified, and the DNA inserts were sequenced (W. M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champaign) to confirm the presence of the introduced mutation. To generate the plasmids encoding CBM16-1 Q21G/N97R/Q121E and CBM16-2 G21Q/R97N/E121Q triple mutants, the QuikChange multisite-directed mutagenesis kit was used with a mixture of primers containing the desired mutations.

**Gene Expression and Protein Purification**—The plasmids encoding either the wild-type or a mutant of CBM16-1 or CBM16-2 were transformed individually into BL21(DE3) CodonPlus RIL competent cells (Stratagene, La Jolla, CA) that were then selected on LB plates supplemented with 100  $\mu$ g/ml ampicillin and 50  $\mu$ g/ml chloramphenicol and cultured at 37 °C. A single colony containing the desired plasmid was then cultured in LB broth supplemented with the same antibiotics at 37 °C until the absorbance at 600 nm ( $A_{600}$ ) reached 0.3. The temperature for culturing the cells was decreased to 16 °C, and isopropyl thio-galactopyranoside was added (0.1 mM, final concentration) to induce the expression of the particular CBM16 derivative. The culturing was continued for 16 h, and the cells were harvested by centrifugation. The cell pellets were resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 20 mM imidazole), and the cell contents were released by passage through an EmulsiFlex C-3 cell homogenizer (Avestin, Ottawa, Canada). The supernatants were collected for subsequent protein purification.

All recombinant proteins were purified by immobilized metal ion affinity chromatography according to the manufacturer's instructions (Talon resin, Clontech). The eluted proteins were further purified by gel filtration using an AKTExpress FPLC equipped with a HiLoad 16/60 Superdex 200 column (GE Healthcare) and a buffer composed of 50 mM Tris-HCl, 150 mM NaCl, pH 7.5.

AGE—AGE was used to compare the relative affinities of the various mutant proteins with those of the wild-type CBM16-1 or CBM16-2 in binding to CMC-Na, KGM, and LBG. The methods employed were essentially the same as that described by Tomme *et al.* (23) with minor modification. Thus, all gels were composed of two layers. The upper layer was the stacking gel, which contained 3% acrylamide in 1.5 M Tris-HCl buffer, pH 8.3. The lower layer was the separating gel, which contained 12% acrylamide in the same buffer as the stacking gel. Varying concentrations of polysaccharides were added in the separating gels (CMC-Na, 0.25, 0.5, 1, 2 mg/ml; KGM, 3.1, 6.2, 12, 25  $\mu$ g/ml; LBG, 7.8, 16, 32, 62  $\mu$ g/ml). As a control, the proteins were also resolved in a gel containing no polysaccharide. One microgram each of CBM16-1 or CBM16-2 wild-type and their mutants were subjected to electrophoresis at 100 V for 7 h at a constant temperature of 4 °C. After electrophoresis, the gels were stained with Coomassie Brilliant Blue G-250 for visualization.

To calculate the relative affinities for the polysaccharides, the migration distances were plotted against the substrate concentrations, using  $1/(R - r)$  as orthogonal and  $1/[S]$  as horizontal coordinates, respectively. The  $R$  represents the relative migrating distance of a specific carbohydrate binding module compared with BSA in the absence of polysaccharides, and the  $r$  represents the relative migrating distance of this protein compared with BSA in the presence of a polysaccharide at a given concentration. For the horizontal coordinates,  $S$  represents the concentration of the polysaccharide, which was expressed as milligrams/ml. The methods used in calculating the binding constants and relative affinities were essentially the same as described by Takeo *et al.* (24).

**Circular Dichroism Spectroscopy**—Far-UV CD spectroscopy was used to analyze the secondary structures of the wild type and mutants of CBM16-1 or CBM16-2. The spectroscopy was performed using a J-815 CD spectropolarimeter (Jasco, Japan) equipped with a constant temperature cell holder. All protein samples (300  $\mu$ l, 0.2 mg/ml) in a 1-mm quartz cuvette were kept at 25 °C for 5 min before scanning. The measurements started with an initial wavelength of 260 nm and a final wavelength of 190 nm with a wavelength step of 0.1 nm. The analysis of the secondary structure of the proteins was conducted on the Dichroweb website using an algorithm of CDSSTR with a reference Set 4 optimized for 190–240 nm (25, 26).

**Isothermal Titration Calorimetry (ITC)**—Isothermal titration calorimetry measurements were carried out at 25 °C using a VP-ITC microcalorimeter from MicroCal Inc. (Northampton, UK). All proteins were extensively dialyzed against a buffer composed of 50 mM sodium phosphate, 150 mM NaCl, pH 7.0; and both oligosaccharide ligands (cellopentaose and mannopentaose) were dissolved in the same buffer. Typically, 100  $\mu$ M protein was injected with 28 successive 10- $\mu$ l volumes of ligand (1–2.6 mM) at 300-s intervals. The data collected were fitted to a nonlinear regression using a single site binding model (MicroCal Origin software), and the thermodynamic parameters were calculated with both the Gibbs free energy equation  $\Delta G = \Delta H - T\Delta S$  and the relationship  $\Delta G = -RT \ln K_a$ .

**Crystallization and X-ray Data Collection**—Initial co-crystallization efforts of CBM16-1 variants in complex with either

mannopentaose or cellopentaose focused around the crystallization conditions originally reported for wild-type CBM16-1, with the respective oligosaccharide (16). Although the co-crystallization conditions were similar to those previously reported, crystals of the site-specific variants demonstrated markedly different morphology and occupy different space groups than for the corresponding wild-type complexes. Briefly, for the variant CBM16-1 mannopentaose complex, 2  $\mu$ l of complex (20 mg/ml in 100 mM KCl, 10 mM HEPES, pH 7.5, plus 5 mM mannopentaose) was added to 2  $\mu$ l of precipitant (1.6 M ammonium sulfate, 100 mM MES, pH 6.5, 10% dioxane) and equilibrated over the precipitant solution at 20 °C. For the variant CBM16-1 cellopentaose complex, 2  $\mu$ l of complex (20 mg/ml in 100 mM KCl, 10 mM HEPES, pH 7.5, plus 5 mM cellopentaose) was added to 2  $\mu$ l of precipitant (30% polyethylene glycol 3350, 100 mM Tris-HCl, pH 8.3, 200 mM  $MgCl_2$ ) and equilibrated over the precipitant solution at 20 °C. Prior to data collection, crystals of each complex were supplemented with a cryoprotectant and vitrified by direct immersion into liquid nitrogen.

Crystals of the CBM16-1 Q121E-mannopentaose complex occupy space group  $C22_1$  with unit cell parameters  $a = 56.4$  Å,  $b = 96.5$  Å,  $c = 72.9$  Å, and contain one molecule in the crystallographic asymmetric unit. An 8-fold redundant data set was collected to a limiting resolution of 1.55 Å at LS-CAT (Advanced Photon Source, Argonne, IL) utilizing a Mar CCD detector (overall  $R_{\text{merge}} = 7.2\%$  and  $I/\sigma(I) = 4$  in the highest resolution shell). Crystals of the CBM16-1 Q121E-cellopentaose complex occupy space group  $P1$  with cell parameters  $a = 32.4$  Å,  $b = 48.4$  Å,  $c = 48.9$  Å,  $\alpha = 62.5$ ,  $\beta = 85.2$ ,  $\gamma = 86.2$ , with two molecules in the asymmetric unit. A 5-fold redundant data set was collected to a limiting resolution of 1.35 Å at LS-CAT (Advanced Photon Source, Argonne, IL) utilizing a Mar CCD detector (overall  $R_{\text{merge}} = 4.7\%$  and  $I/\sigma(I) = 9$  in the highest resolution shell). All data were indexed, merged, and scaled using the HKL2000 package (27).

**Structure Determination and Refinement**—Crystallographic phases for the CBM16-1 Q121E complexes were determined by molecular replacement (28, 29), utilizing the coordinates for the corresponding wild-type complex structures (minus all ligand and solvent molecules) as the search probe. Clear density for the appropriate ligand could be observed prior to any crystallographic refinement. Following rigid body refinement of the initial molecular replacement solution, each structure was subjected to automatic rebuilding using ARP/warp (30), followed by manual rebuilding using XtalView (31). The respective ligands were manually built into difference Fourier maps after the free  $R$  factors dropped below 25%. Solvent molecules and alternate conformers were added to yield final models with good free  $R$  factors, no Ramachandran outliers, and excellent stereochemistry.

Cross-validation, using 5–7% of the data for the calculation of the free  $R$  factor, was utilized throughout the model building process to monitor building bias (32). The stereochemistry of the models was routinely monitored throughout the course of refinement using PROCHECK (33) and MolProbity.

Crystal parameters, data collection parameters, and refinement statistics for each of the structures are summarized in Table 9. The refined coordinates have been deposited in the

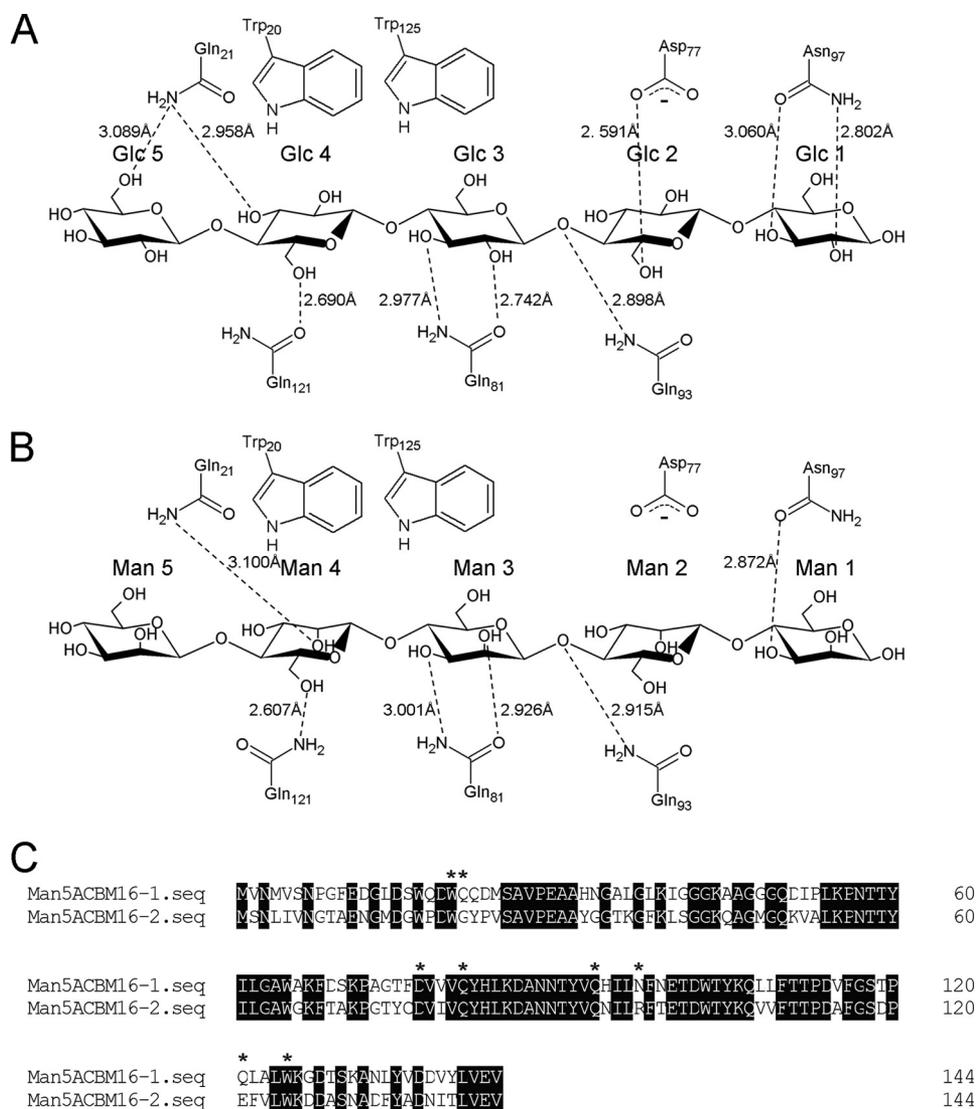


FIGURE 1. Schematic diagram of the interaction of CBM16-1 with cellopentaose (A) and mannopentaose (B). The direct H-bonds are expressed as dashed lines. C, sequence alignment of CBM16-1 and CBM16-2. The identical residues are in a black background, and the mutated residues are indicated with asterisks.

PDB with accession 30EA (CBM16-1 Q121E-cellopentaose) and 30EB (CBM16-1 Q121E-mannopentaose).

**Amino Acid Sequence Alignment**—The amino acid sequences of members of CBM family 16 were retrieved from the Carbohydrate Active enZYme database and the GenBank™ database and aligned with the sequences of Man5A CBM16-1 and CBM16-2 by using ClustalX.

## RESULTS

**Identification of Residues Involved in Ligand Binding and Construction of Mutants**—In a previous study, we reported the crystal structure of CBM16-1 in complex with  $\beta$ -1,4-cellopentaose and  $\beta$ -1,4-mannopentaose (16). These structures revealed a binding groove that accommodated the mannose- and glucose-configured oligosaccharides. Within the binding cleft, two tryptophans (Trp-20 and Trp-125) and six polar residues (Gln-21, Asp-77, Gln-81, Gln-93, Asn-97, and Gln-121) were predicted to contribute to ligand binding by either hydrophobic stacking interactions or hydrogen bond formation with the

ligands (Fig. 1, A and B). To explore the roles of these residues in ligand binding, each residue was individually mutated to alanine. These mutations allowed subsequent comparison of the binding affinities of the wild-type and the mutant forms of each protein by AGE and isothermal titration calorimetry.

The primary sequences of CBM16-1 and CBM16-2 each include 143 amino acid residues, because the first methionine was included to facilitate gene expression. The two CBMs share 62 and 74% identity and similarity, respectively, at the amino acid sequence level (Fig. 1C). For CBM16-2, many of the residues predicted to be involved in binding for CBM16-1 are conserved. Thus, the two tryptophans (Trp-20 and Trp-125) and five polar residues (Asp-77, Gln-81, Gln-93, Arg-97, and Glu-121) of CBM16-2 were mutated to alanine. The Gly-21 in CBM16-2 (corresponding to Gln-21 in CBM16-1) was also mutated to alanine for a direct comparison with the CBM16-1 Q21A mutant.

Of the aromatic and polar residues, three differed in the two CBMs (Gln-21, Asn-97, and Gln-121 in CBM16-1 corresponded to Gly-21, Arg-97, and Glu-121 in CBM16-2). Each of these residues was mutated to the respective residue in the other CBM. The mutations therefore led to the mutants CBM16-1 Q21G, N97R, and Q121E and mutants CBM16-2 G21Q, R97N, and E121Q. Furthermore, two mutants carrying triple mutations were constructed, and these were CBM16-1 Q21G/N97R/Q121E and CBM16-2 G21Q/R97N/E121Q. The triple mutations made the two CBMs even more similar to each other at the amino acid sequence level.

All wild-type and mutant genes were well expressed heterologously in *E. coli*, and the recombinant proteins were purified by affinity chromatography followed by size exclusion chromatography. On SDS-polyacrylamide gels, the wild-type and the mutant proteins appeared as single bands (data not shown).

**Structural Integrity of CBM16-1 and CBM16-2 Mutants**—To evaluate whether the mutations affected the structural integrity of the CBMs, the secondary structural elements of all proteins were compared using circular dichroism scans, and the spectra were analyzed using the Dichroweb on-line server (Tables 1 and 2). For all 12 mutants of CBM16-1, there were no significant changes in the secondary structure compared with the wild-type. For CBM16-2, most of the mutations did not cause significant changes to the secondary structural elements with the

**TABLE 1****CD spectroscopy analysis of CBM16-1 and its mutants**

The values in the table are given as percentage. All values are expressed as mean  $\pm$  S.D. of three independent experiments.

Protein	$\alpha$ -Helix	$\beta$ -Sheet	Turn	Unordered
WT	4.7 $\pm$ 0.6	44.7 $\pm$ 1.5	20.3 $\pm$ 0.6	29.0 $\pm$ 1.7
W20A	4.3 $\pm$ 0.6	42.7 $\pm$ 0.6	21.0 $\pm$ 1.0	30.7 $\pm$ 0.6
W125A	5.3 $\pm$ 0.6	42.3 $\pm$ 1.5	21.0 $\pm$ 0.0	30.3 $\pm$ 0.6
Q21A	4.8 $\pm$ 1.0	45.3 $\pm$ 1.5	20.0 $\pm$ 0.8	28.8 $\pm$ 1.5
D77A	4.3 $\pm$ 1.2	44.0 $\pm$ 1.0	19.7 $\pm$ 0.6	30.0 $\pm$ 1.0
Q81A	3.7 $\pm$ 0.6	44.7 $\pm$ 0.6	20.7 $\pm$ 0.6	30.0 $\pm$ 1.0
Q93A	4.5 $\pm$ 0.7	44.0 $\pm$ 1.4	20.5 $\pm$ 0.7	29.0 $\pm$ 0.0
N97A	4.7 $\pm$ 0.6	45.3 $\pm$ 1.2	19.7 $\pm$ 1.2	28.3 $\pm$ 0.6
Q121A	4.3 $\pm$ 0.6	44.3 $\pm$ 0.6	20.0 $\pm$ 1.0	29.7 $\pm$ 1.2
Q21G	5.3 $\pm$ 0.6	45.3 $\pm$ 1.5	18.7 $\pm$ 1.2	29.0 $\pm$ 1.7
N97R	4.0 $\pm$ 1.0	44.7 $\pm$ 1.2	19.3 $\pm$ 1.5	30.3 $\pm$ 1.2
Q121E	4.3 $\pm$ 0.6	44.0 $\pm$ 2.0	20.7 $\pm$ 1.5	29.0 $\pm$ 1.7
Q21G/N97R/Q121E	4.3 $\pm$ 0.6	44.7 $\pm$ 0.6	20.3 $\pm$ 0.6	29.7 $\pm$ 0.6

**TABLE 2****CD spectroscopy analysis of CBM16-2 and its mutants**

The values in the table are given as percentage. All values are expressed as mean  $\pm$  S.D. of three independent experiments.

Protein	$\alpha$ -Helix	$\beta$ -Sheet	Turn	Unordered
WT	5.3 $\pm$ 0.6	42.7 $\pm$ 0.6	21.3 $\pm$ 0.6	30.0 $\pm$ 0.0
W20A	4.5 $\pm$ 0.6	42.0 $\pm$ 1.4	21.5 $\pm$ 0.6	31.0 $\pm$ 0.8
W125A	4.3 $\pm$ 0.6	42.7 $\pm$ 2.1	21.3 $\pm$ 1.2	30.0 $\pm$ 1.0
G21A	4.5 $\pm$ 0.6	43.5 $\pm$ 1.0	20.8 $\pm$ 1.0	29.8 $\pm$ 1.3
D77A	4.5 $\pm$ 1.0	43.0 $\pm$ 0.8	21.0 $\pm$ 0.0	30.5 $\pm$ 1.0
Q81A	4.7 $\pm$ 0.6	43.7 $\pm$ 1.1	20.7 $\pm$ 0.6	29.7 $\pm$ 0.6
Q93A	4.3 $\pm$ 0.6	45.7 $\pm$ 0.6 <sup>a</sup>	20.7 $\pm$ 0.6	29.0 $\pm$ 1.7
R97A	5.3 $\pm$ 0.6	42.0 $\pm$ 0.0	21.3 $\pm$ 0.6	30.3 $\pm$ 0.6
E121A	5.0 $\pm$ 1.0	43.3 $\pm$ 0.6	21.3 $\pm$ 0.6	29.3 $\pm$ 0.6
G21Q	4.7 $\pm$ 0.6	43.0 $\pm$ 0.0	21.0 $\pm$ 0.0	30.3 $\pm$ 0.6
R97N	4.3 $\pm$ 0.5	43.3 $\pm$ 1.0	21.0 $\pm$ 0.8	30.8 $\pm$ 1.0
E121Q	4.7 $\pm$ 0.6	42.7 $\pm$ 1.2	21.7 $\pm$ 0.6	30.3 $\pm$ 0.6
G21Q/R97N/E121Q	4.3 $\pm$ 0.5	41.3 $\pm$ 1.0	21.5 $\pm$ 0.6	32.0 $\pm$ 0.1 <sup>b</sup>

<sup>a</sup>  $p = 0.024$ .

<sup>b</sup>  $p < 0.0001$ .

exception of the Q93A mutant, which showed a small but significant increase in  $\beta$ -sheet composition, and the G21Q/R97N/E121Q mutant, which showed a small increase in unordered structure.

*Two Conserved Tryptophan Residues Are Critical for Binding to Substrate in Both CBM16-1 and CBM16-2*—Sample plots used to determine the relative affinities of the wild-type and mutants of CBM16-1 and CBM16-2 from AGE are given in Fig. 2. In addition, representative ITC data are presented in Fig. 3. Mutation of Trp-20 and Trp-125 in CBM16-1 to alanine attenuated the capacity of CBM16-1 to bind to polysaccharides (Table 3), cellopentaose (Table 5), and mannopentaose (Table 7) to levels that were undetectable by the assays employed. One exception to this observation was that CBM16-1 W125A retained 13% binding affinity to CMC-Na relative to the wild-type protein (Table 3). CMC-Na is an artificial soluble cellulose analog with carboxymethyl groups linked to some hydroxyl groups of the glucopyranose chain. Therefore, the residual binding of W125A to this artificial substrate may not be biologically relevant. These results demonstrated that CBM16-1 Trp-20 and Trp-125 play critical roles in binding to polysaccharide and oligosaccharide substrates. Furthermore, the observation that binding to both cellopentaose and mannopentaose was attenuated indicates that the roles of the two amino acid residues in binding are irrespective of the orientation of the C2 hydroxyl.

Mutation of CBM16-2 Trp-20 and Trp-125 to alanine also abolished binding to all three polysaccharides (Table 4) as well as cellopentaose (Table 6) and mannopentaose (Table 8). These results further support the critical role of these two conserved tryptophan residues in ligand binding.

*Two Conserved Polar Residues, Gln-81 and Gln-93, Are Essential for Ligand Binding in Both CBM16-1 and CBM16-2*—Mutation of Gln-81 to alanine completely abolished the binding of CBM16-1 to each oligosaccharide (Table 5 and Table 7) and also the polysaccharides (Table 3), indicating that, similar to Trp-20 and Trp-125, this residue plays a critical role in ligand binding. Similarly, CBM16-2 Q81A did not show binding to any polysaccharides (Table 4) and oligosaccharides (Table 6 and Table 8), thus confirming the important role of Gln-81 in ligand binding in each of the two family 16 CBMs.

The binding of CBM16-1 Q93A was severely impaired with cellopentaose (1% residual binding, Table 5) and CMC-Na (13% residual binding, Table 3) as ligands, and for mannopentaose (Table 7), LBG, and KGM (Table 3), our methods did not detect any binding. As expected, the CBM16-2 Q93A mutant also had no detectable binding affinity for mannopentaose (Table 8), KGM, and LBG (Table 4); however, this mutant retained some affinity for cellopentaose (6% residual binding, Table 6) and CMC-Na (28% residual binding, Table 4). These results therefore suggested that Gln-93 is a critical residue for binding to mannose-configured substrates for both CBM16-1 and CBM16-2. Furthermore, the residual binding to cellopentaose indicates that this residue may be important for discriminating between glucose- and mannose-configured substrates.

*Other Polar Residues, Gln-21, Asp-77, Asn-97/Arg-97, and Gln-121/Glu-121, Play Moderately Important Roles in Ligand Binding by CBM16-1 and CBM16-2*—The CBM16-1 Q21A mutant decreased the binding affinity by about 8- and 14-fold for cellopentaose (Table 5) and mannopentaose (Table 7), respectively. The mutation also reduced the affinity by about 3-fold for CMC-Na, 8-fold for LBG, and 7-fold for KGM, respectively (Table 3). Not surprisingly, the binding affinities of CBM16-2 G21A remained almost the same for polysaccharides in AGE analysis (Table 4). However, the binding affinities of G21A for both oligosaccharides were decreased 2-fold in the ITC measurements (Tables 6 and 8).

The CBM16-1 D77A mutation decreased the binding affinities for all polysaccharides and oligosaccharides, with a range from 1.3 to 4.3-fold (Tables 3, 5, and 7). The CBM16-2 D77A mutation also decreased the binding affinities for both the oligosaccharides (Tables 6 and 8) and polysaccharides (Table 4).

The CBM16-1 N97A mutation resulted in decreases in affinity of about 2-fold for both cellopentaose (Table 5) and mannopentaose (Table 7). The folds of decrease in the affinities with CMC-Na (2-fold), LBG (3-fold), and KGM (2-fold) were similar to those for the two oligosaccharides (Table 3). The CBM16-2 R97A decreased the binding affinities for cellopentaose by 2-fold (Table 6) and for mannopentaose by 3-fold (Table 8). On the other hand, the decreases in affinity for the polysaccharides were mostly higher as follows: CMC-Na (5-fold), LBG (2-fold), and KGM (7-fold) (Table 4).

The Q121A mutation also moderately decreased the binding affinities of CBM16-1 to all substrates, with the residual affini-

## Role of Amino Acids in Family 16 CBM

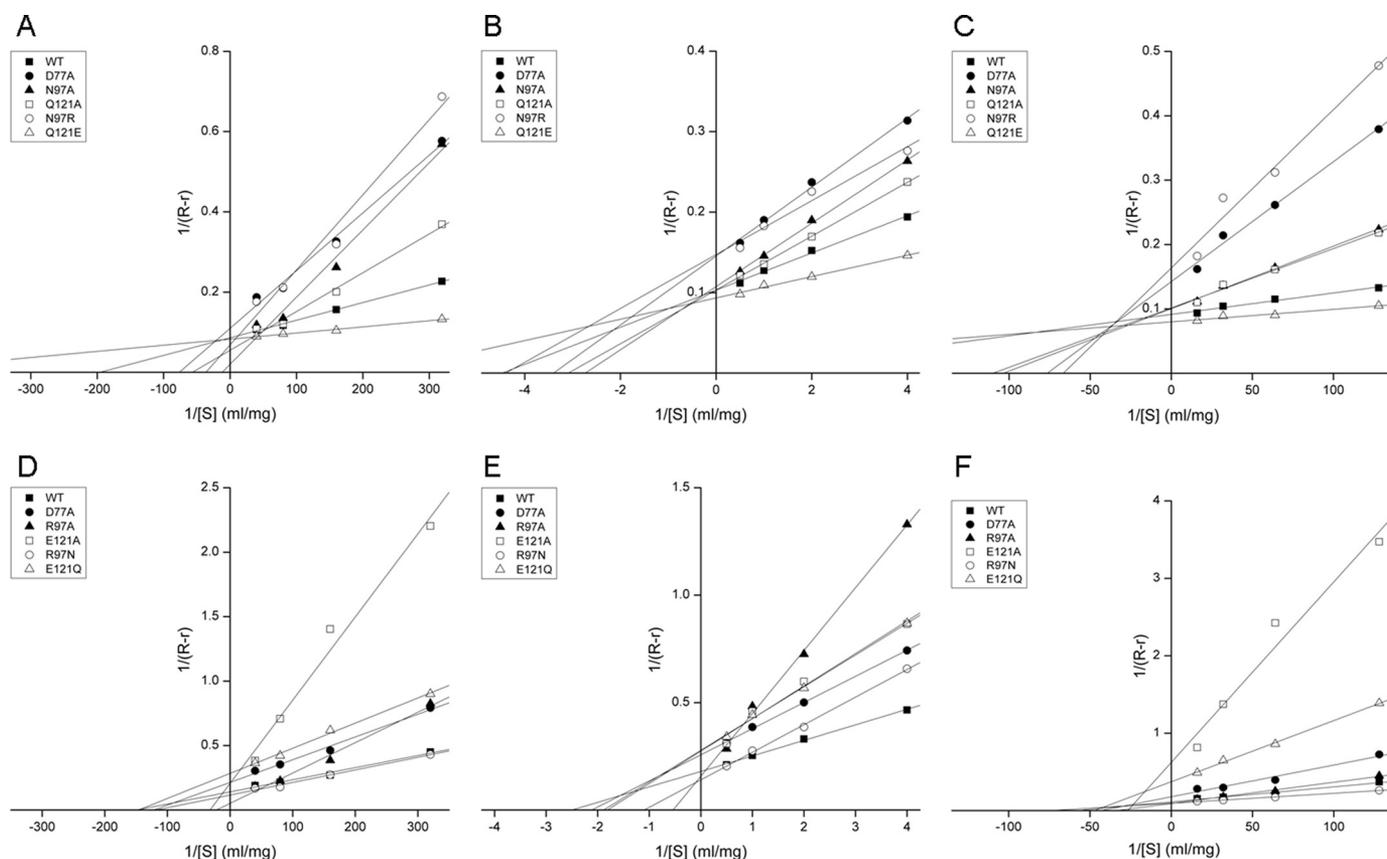


FIGURE 2. Sample plots used to determine the relative affinities of CBM16-1 with KGM (A), CMC-Na (B), and LBG (C) or CBM16-2 with KGM (D), CMC-Na (E), and LBG (F) in AGE.

ties for mannopentaose and CMC-Na determined to be 21 and 69%, respectively (Tables 3, 5, and 7). The CBM16-1 Gln-121 equivalent in CBM16-2 was Glu-121, and the CBM16-2 E121A mutation decreased the affinities for all substrates, with the residual affinities ranging from 5% for mannopentaose to 72% for CMC-Na (Tables 4, 6, and 8). In summary, these results demonstrated that Gln-21, Asp-77, Asn-97/Arg-97, and Gln-121/Glu-121 all contribute to ligand binding, but their roles are less critical than Trp-20, Trp-125, Gln-81, and Gln-93.

**Binding Affinities of CBM16-1 and CBM16-2 Can Be Tuned by Replacing a Nonconserved Polar Residue with Its Equivalent in the Other CBM**—When CBM16-1 Gln-121 was mutated to glutamic acid, the binding affinities were increased for all of the substrates tested, ranging from 1.3-fold for LBG to 2.8-fold for KGM (Tables 3, 5, and 7). Meanwhile, the enthalpies were more favorable for oligosaccharide binding:  $\Delta\Delta H = -1.02$  kJ for Q121E-cellopentaose (Table 5) and  $\Delta\Delta H = -0.83$  kJ for Q121E-mannopentaose reaction (Table 7). This result suggests that the amino acid residue at position 121 may be a critical determinant in manipulating the binding affinities for these two CBMs.

Interestingly, the CBM16-2 G21Q mutant also showed increased binding affinities for polysaccharides from 1.6-fold (for LBG) to 2-fold (for KGM, Table 4). However, decreases were rather observed for cellopentaose (Table 6) and mannopentaose (Table 8), as observed for the CBM16-2 G21A mutant.

For CBM16-1, N97R showed almost unchanged binding affinity for CMC-Na (Table 3), whereas the affinities for other

substrates decreased from 1.9 to 6.3-fold (Tables 3, 5, and 7). The binding affinities of CBM16-1 Q21G decreased from 1.5 (mannopentaose) to 16.7 (KGM)-fold for all substrates (Tables 3, 5, and 7). For the oligosaccharide substrates, CBM16-2 E121Q exhibited considerable decreases in affinities (Tables 6 and 8), although for the polysaccharides the decreases in affinity were only moderate (Table 4). CBM16-2 R97N also decreased affinities for the substrates except for LBG, where the affinity rather increased (Tables 4, 6, and 8). Taken together, although CBM16-1 Q121E and CBM16-2 G21Q (for polysaccharides) increased the binding affinities, the other reciprocal mutants (CBM16-1 Q21G, CBM16-1 N97R, CBM16-2 R97N, and CBM16-2 E121Q) exhibited decreased binding affinities. These findings therefore suggest that the nonconserved polar residues in the binding cleft can be modified to fine-tune the affinities of these two CBMs.

The triple mutant CBM16-1 Q21G/N97R/Q121E completely lost its binding affinity for mannopentaose (Table 7) and LBG (Table 3). In the case of CBM16-2 G21Q/R97N/E121Q, it also resulted in the loss of binding affinity for mannopentaose (Table 8) and 10% residual binding activity for cellopentaose (Table 6). The binding affinities of this triple mutant for CMC-Na, KGM, and LBG were 28, 19, and 83%, respectively.

**Structural Basis for Increased Oligosaccharide Affinity of CBM16-1 Q121E**—Given that our calorimetric analysis suggests that Gln-121 in CBM16-1 is a critical determinant for modulating both oligosaccharide and polysaccharide binding affinity, we determined the co-crystal structures of CBM16-1

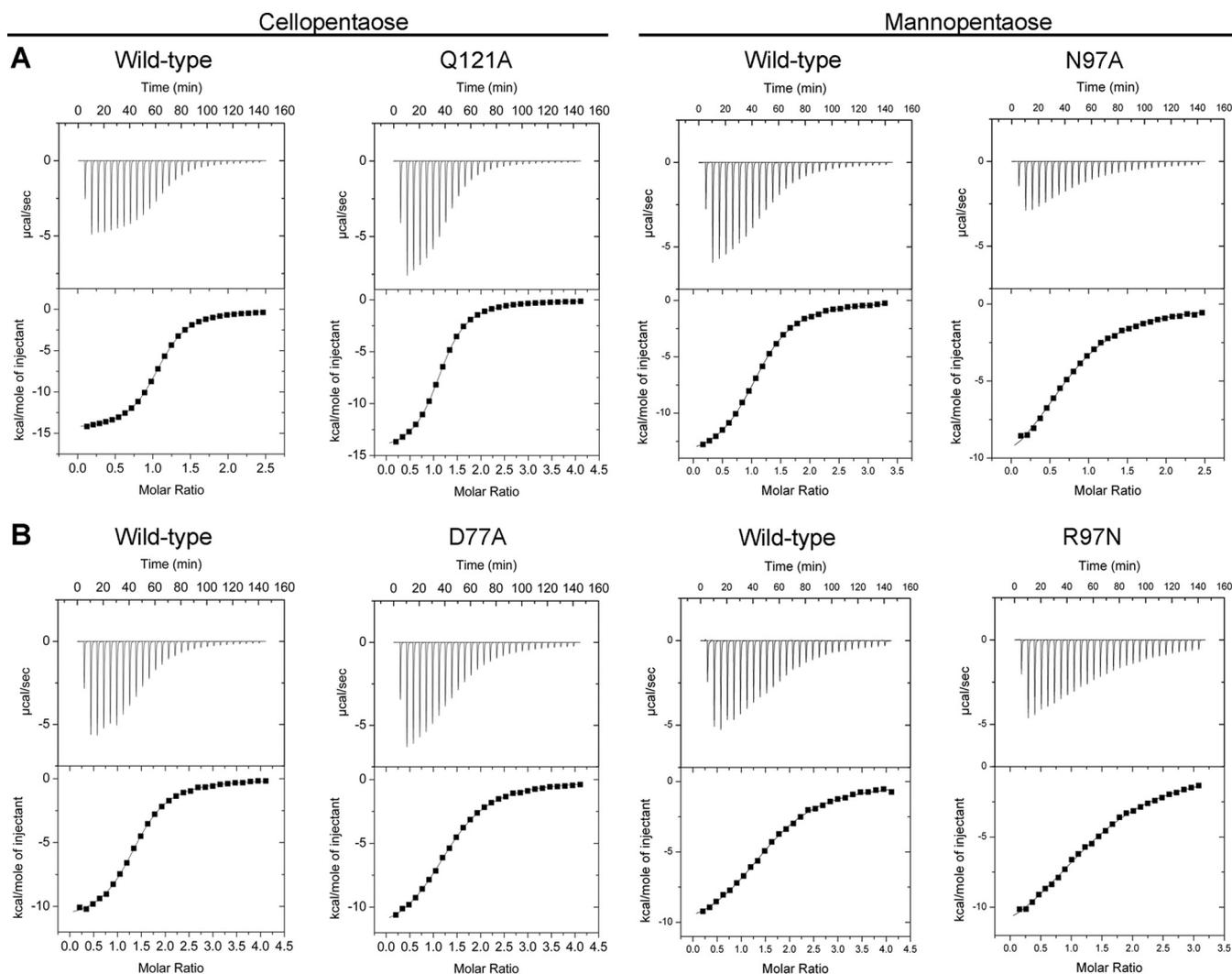


FIGURE 3. Representative ITC data for the binding of the wild-type or mutants of CBM16-1 (A) or CBM16-2 (B) with cellopentaose or mannopentaose. All experiments were carried out in 50 mM sodium phosphate buffer, pH 7.0, at 25 °C.

TABLE 3

Relative affinities of CBM16-1 wild-type and mutants for CMC-Na, KGM, and LBG as determined by AGE

Protein	Interaction <sup>a</sup>	Ligand					
		CMC-Na		KGM		LBG	
		$K_a$	Relative affinity	$K_a$	Relative affinity	$K_a$	Relative affinity
		<i>ml/mg</i>	%	<i>ml/mg</i>	%	<i>ml/mg</i>	%
Wild type		4.5	100	195.7	100	305	100
W20A	HS	ND <sup>b</sup>		ND		ND	
W125A	HS	0.6	13	ND		ND	
Q21A	HB	1.8	40	28.1	14	40.9	13
D77A	HB	3.4	76	76.2	39	74.9	25
Q81A	HB	ND		ND		ND	
Q93A	HB	0.6	13	0.8	0	ND	
N97A	HB	2.7	60	85.5	44	100.4	33
Q121A	HB	3.1	69	56.6	29	112.7	37
Q21G	HB	2.7	60	12	6	59.4	19
N97R	HB	4.4	98	36.4	19	47.5	16
Q121E	HB	7	156	540.1	276	399.5	131
Q21G/N97R/Q121E	HB	2.2	49	10.4	5	ND	

<sup>a</sup> The interactions are expressed as HS, residues predicted to be involved in hydrophobic stacking with ligand; HB, residues predicted to be involved in hydrogen bonding with ligand.

<sup>b</sup> ND means no binding was detected.

Q121E in complex with mannopentaose (1.55 Å resolution) and cellopentaose (1.35 Å resolution) to further delineate the basis for increased binding affinity observed for this variant.

Each co-crystal structure was determined by molecular replacement, and clear and continuous electron density, corresponding to each of the five  $\beta$ -1,4-linked sugars, could be

**TABLE 4**

Relative Affinities of CBM16-2 wild-type and mutants for CMC-Na, KGM and LBG as determined by AGE

Protein	Interaction <sup>a</sup>	Ligand					
		CMC-Na		KGM		LBG	
		$K_a$	Relative affinity	$K_a$	Relative affinity	$K_a$	Relative affinity
		ml/mg	%	ml/mg	%	ml/mg	%
Wild-type		2.5	100	149.3	100	59.1	100
W20A	HS	ND <sup>b</sup>		ND		ND	
W125A	HS	ND		ND		ND	
G21A		2.2	88	178.1	119	47.6	81
D77A	HB	2.1	84	121.8	82	43.5	74
Q81A	HB	ND		ND		ND	
Q93A	HB	0.7	28	ND		ND	
R97A	HB	0.5	20	21.9	15	33.9	57
E121A	HB	1.8	72	32.1	22	26.7	45
G21Q	HB	3.7	168	301.8	202	96.8	164
R97N	HB	1.1	44	118.4	79	72.9	123
E121Q	HB	1.9	76	145.9	98	47.2	80
G21Q/R97N/E121Q	HB	0.7	28	27.7	19	49.2	83

<sup>a</sup> The interactions are expressed as HS, residues predicted to be involved in hydrophobic stacking with ligand; HB, residues predicted to be involved in hydrogen bonding with ligand.

<sup>b</sup> ND means no binding was detected.

**TABLE 5**

Affinity of CBM16-1 wild-type and mutants binding to cellopentaose determined by ITC

Protein	$n^a$	$K_a \times 10^{3a}$	$\Delta G^a$	$\Delta H^a$	$T\Delta S^a$	Relative affinity <sup>b</sup>
		$M^{-1}$	$kJ$	$kJ$	$kJ$	
Wild-type	1.09 ± 0.03	261.3 ± 3.8	-7.38 ± 0.01	-14.32 ± 0.34	-6.94 ± 0.35	100
W20A		ND <sup>c</sup>				
W125A		ND				
Q21A	1.10 ± 0.01	30.3 ± 2.5	-6.11 ± 0.04	-13.40 ± 0.15	-7.28 ± 0.25	12
D77A	1.15 ± 0.02	84.8 ± 0.4	-6.75 ± 0.05	-14.09 ± 0.07	-7.34 ± 0.03	32
Q81A		ND				
Q93A	0.94 ± 0.06	3.7 ± 0.1	-4.88 ± 0.01	-14.83 ± 0.01	-9.96 ± 1.35	1
N97A	0.98 ± 0.03	134.3 ± 9.0	-6.99 ± 0.04	-13.44 ± 0.62	-6.45 ± 0.61	51
Q121A	1.13 ± 0.02	131.0 ± 2.5	-6.98 ± 0.03	-14.83 ± 0.45	-7.85 ± 0.42	50
Q21G	1.08 ± 0.04	38.0 ± 0.7	-6.24 ± 0.01	-13.96 ± 0.36	-7.71 ± 0.36	15
N97R	1.14 ± 0.01	146 ± 9.5	-7.04 ± 0.02	-13.71 ± 0.10	-6.67 ± 0.12	54
Q121E	0.92 ± 0.03	416.3 ± 27.0	-7.66 ± 0.03	-15.34 ± 0.03	-7.67 ± 0.55	159
Q21G/N97R/Q121E	1.15 ± 0.04	25.1 ± 3.0	-6.01 ± 0.07	-12.60 ± 0.14	-6.59 ± 0.21	10

<sup>a</sup> Values are reported as means ± S.E. from the mean of three independent experiments.

<sup>b</sup> Relative affinity was calculated as (wild-type or mutant  $K_a$ /wild-type  $K_a$ ) × 100.

<sup>c</sup> ND means no binding was detected.

**TABLE 6**

Affinity of CBM16-2 wild-type and mutants binding to cellopentaose determined by ITC

Protein	$n^a$	$K_a \times 10^{3a}$	$\Delta G^a$	$\Delta H^a$	$T\Delta S^a$	Relative affinity <sup>b</sup>
		$M^{-1}$	$kJ$	$kJ$	$kJ$	
Wild-type	1.33 ± 0.04	96.4 ± 1.3	-6.80 ± 0.00	-11.36 ± 0.13	-4.56 ± 0.13	100
W20A		ND <sup>c</sup>				
W125A		ND				
G21A	1.13 ± 0.05	41.5 ± 1.4	-6.31 ± 0.02	-15.68 ± 0.92	-9.37 ± 0.91	43
D77A	1.26 ± 0.07	58.0 ± 6.9	-6.37 ± 0.19	-12.68 ± 0.16	-6.18 ± 0.10	60
Q81A		ND				
Q93A	1.41 ± 0.72	5.4 ± 0.5	-5.09 ± 0.04	-10.99 ± 6.28	-5.90 ± 6.25	6
R97A	1.33 ± 0.02	40.3 ± 2.2	-6.28 ± 0.03	-11.91 ± 0.22	-5.63 ± 0.24	42
E121A	1.37 ± 0.01	15.5 ± 0.9	-5.74 ± 0.03	-11.50 ± 0.29	-5.74 ± 0.32	16
G21Q	1.52 ± 0.30	56.8 ± 4.7	-6.45 ± 0.00	-12.56 ± 2.22	-5.47 ± 2.72	59
R97N	1.20 ± 0.05	46.8 ± 5.3	-6.37 ± 0.06	-12.74 ± 0.59	-6.37 ± 0.65	49
E121Q	1.31 ± 0.09	28.8 ± 2.8	-6.09 ± 0.05	-11.73 ± 0.56	-5.65 ± 0.55	30
G21Q/R97N/E121Q	1.19 ± 0.06	9.5 ± 0.7	-5.43 ± 0.05	-13.73 ± 0.76	-8.30 ± 0.80	10

<sup>a</sup> Values are reported as means ± S.E. from the mean of three independent experiments.

<sup>b</sup> Relative affinity was calculated as (wild-type or mutant  $K_a$ /wild-type  $K_a$ ) × 100.

<sup>c</sup> ND means no binding was detected.

observed in initial electron density maps, prior to any crystallographic refinement. Each structure was refined without restraints to free *R* factors below 22%. Detailed information on the collected data, phasing, and refinement is provided in Table 9.

A comparison of each variant co-crystal structure against the appropriate wild-type complex suggests only modest changes with root mean square deviations of 0.2 Å for all *C* $\alpha$  atoms for

the cellopentaose complex structures and 0.3 Å for the manno-pentaose complex structures. In each of these variant co-crystal structures, the oligosaccharide is buried in an extended conformation within a 25-Å cleft orthogonal to the  $\beta$ -sheets of the polypeptide. As observed in each of the wild-type co-crystal structures, interactions with the oligosaccharide are mediated by hydrophobic stacking (with Trp-20 and Trp-125) and extensive hydrogen bonding interactions.

TABLE 7

Affinity of CBM16-1 wild-type and mutants binding to mannopentaose determined by ITC

Protein	$n^a$	$K_a \times 10^{3a}$ $M^{-1}$	$\Delta G^a$ $kJ$	$\Delta H^a$ $kJ$	$T\Delta S^a$ $kJ$	Relative affinity <sup>b</sup>
Wild-type	1.13 ± 0.01	91.8 ± 1.56	-6.79 ± 0.01	-14.12 ± 0.24	-7.33 ± 0.25	100
W20A		ND <sup>c</sup>				
W125A		ND				
Q21A	0.95 ± 0.00	6.8 ± 0.4	-5.24 ± 0.03	-14.30 ± 0.18	-9.06 ± 0.21	7
D77A	0.94 ± 0.05	21.6 ± 1.9	-5.91 ± 0.07	-14.01 ± 0.21	-8.10 ± 0.27	23
Q81A		ND				
Q93A		ND				
N97A	0.78 ± 0.00	39.45 ± 1.06	-6.25 ± 0.02	-12.25 ± 0.02	-5.99 ± 0.00	43
Q121A	0.60 ± 0.03	19.4 ± 0.44	-5.85 ± 0.01	-12.70 ± 0.62	-6.85 ± 0.63	21
Q21G	0.99 ± 0.09	62.7 ± 5.9	-5.18 ± 0.05	-13.05 ± 1.56	-7.87 ± 1.60	68
N97R	1.05 ± 0.03	22.8 ± 0.4	-5.95 ± 0.03	-12.41 ± 0.53	-6.46 ± 0.54	25
Q121E	0.96 ± 0.01	163.0 ± 8.7	-7.12 ± 0.02	-14.95 ± 0.18	-7.83 ± 0.18	178
Q21G/N97R/Q121E		ND				

<sup>a</sup> Values are reported as means ± S.E. from the mean of three independent experiments.<sup>b</sup> Relative affinity was calculated as (wild-type or mutant  $K_a$ /wild-type  $K_a$ ) × 100.<sup>c</sup> ND means no binding was detected.

TABLE 8

Affinity of CBM16-2 wild-type and mutants binding to mannopentaose determined by ITC

Protein	$n^a$	$K_a \times 10^{3a}$ $M^{-1}$	$\Delta G^a$ $kJ$	$\Delta H^a$ $kJ$	$T\Delta S^a$ $kJ$	Relative affinity <sup>b</sup>
Wild-type	1.22 ± 0.03	48.8 ± 3.6	-6.03 ± 0.60	-14.18 ± 0.52	-8.15 ± 0.67	100
W20A		ND <sup>c</sup>				
W125A		ND				
G21A	1.39 ± 0.07	20.0 ± 1.0	-5.87 ± 0.03	-14.63 ± 0.42	-8.75 ± 0.44	41
D77A	1.40 ± 0.02	19.9 ± 2.2	-5.86 ± 0.07	-14.52 ± 0.49	-8.67 ± 0.55	41
Q81A		ND				
Q93A		ND				
R97A	1.33 ± 0.01	16.6 ± 0.6	-5.75 ± 0.01	-15.04 ± 0.39	-9.29 ± 0.40	34
E121A	1.68 ± 0.80	2.5 ± 0.6	-4.64 ± 0.14	-15.34 ± 7.23	-10.70 ± 7.08	5
G21Q	1.44 ± 0.02	14.1 ± 0.6	-5.54 ± 0.25	-17.09 ± 0.16	-11.56 ± 0.30	29
R97N	1.37 ± 0.02	20.9 ± 0.2	-5.89 ± 0.01	-14.67 ± 0.14	-8.79 ± 0.14	43
E121Q	1.65 ± 0.05	6.5 ± 0.8	-5.20 ± 0.48	-14.70 ± 0.48	-9.50 ± 0.46	13
G21Q/R97N/E121Q		ND				

<sup>a</sup> Values are reported as means ± S.E. of the mean from three independent experiments.<sup>b</sup> Relative affinity was calculated as (wild-type or mutant  $K_a$ /wild-type  $K_a$ ) × 100.<sup>c</sup> ND means no binding was detected.

TABLE 9

Data collection, phasing, and refinement statistics

	CBM16-1-cellopentaose	CBM16-1-mannopentaose
<b>Data collection</b>		
Space group	P <sub>1</sub>	C222 <sub>1</sub>
Cell dimensions		
<i>a, b, c</i>	32.4, 48.4, 48.9 Å	56.4, 96.5, 72.9 Å
$\alpha, \beta, \gamma$	62.5, 85.2, 86.2°	90.0, 90.0, 90.0°
Resolution <sup>a</sup>	50-1.35 Å (1.37-1.35 Å)	50-1.55 Å (1.58-1.55 Å)
$R_{\text{sym}}$	4.7% (14.5%)	7.2% (30.0%)
$I/\sigma(I)$	45.9 (9.4)	41.4 (3.7)
Completeness	94.9% (91.5%)	99.8% (96.6%)
Redundancy	4.9 (4.7)	7.7 (5.9)
<b>Refinement</b>		
Resolution	25.0-1.35 Å	25.0-1.55 Å
No. of reflections	51,900	27,723
$R_{\text{work}}/R_{\text{free}}^b$	17.4/20.3	19.9/21.8
No. of atoms		
Protein	2316	1138
Sugar	112	56
Water	497	142
<b>B-factors</b>		
Protein	11.89	17.28
Sugar	14.16	23.01
Water	25.06	33.56
Root mean square deviations		
Bond lengths	0.001 Å	0.006 Å
Bond angles	1.03°	1.00°

<sup>a</sup> Highest resolution shell is shown in parentheses.<sup>b</sup>  $R$ -factor =  $\Sigma(|F_{\text{obs}}| - k|F_{\text{calc}}|)/\Sigma|F_{\text{obs}}|$  and  $R_{\text{free}}$  is the  $R$  value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement.

In comparison with the wild-type CBM16-1-mannopentaose structure, the position of the oligosaccharide in the Q121E-mannopentaose co-crystal structure deviates significantly at sites 1 and 2. This is not unexpected as these sites are liganded by the polypeptide only along one face of the sugar-binding cleft. As a consequence, the C6 hydroxyl at site 2 now orients toward the top part of the cleft, resulting in a hydrogen bond interaction with the Oδ2 of Asp-77 (O–O distance of 2.7 Å) that is absent in the wild-type co-crystal structure.

In the variant structure, the side chain of Glu-121 is situated closer to the pyranoside in site 4 than in the wild-type structure, resulting in a favorable electrostatic interaction between Oε1 and the C6 hydroxyl (Fig. 4A). The introduction of an additional negative charge in this variant does not disrupt the hydrogen-bonding interaction with neighboring ligand residue Gln-81, which interacts with the C3 hydroxyl in site 3. However, the Q121E mutation results in the replacement of a nitrogen atom with oxygen as the hydrogen-bonding partner. The Glu-121 side chain is positioned directly in line with the side chain of Gln-81, resulting in a stereochemistry that may orient the lone pair electrons on Nε2 of Gln-81 for more favorable interactions with the C3 hydroxyl of the ligand at site 3.

The CBM16-1 Q121E-cellopentaose co-crystal structure reveals a similar set of compensatory movements, in comparison with the wild-type co-crystal structure (Fig. 4B). The side chain of Glu-121 is shifted even more toward the pyranoside at site 4,

## Role of Amino Acids in Family 16 CBM

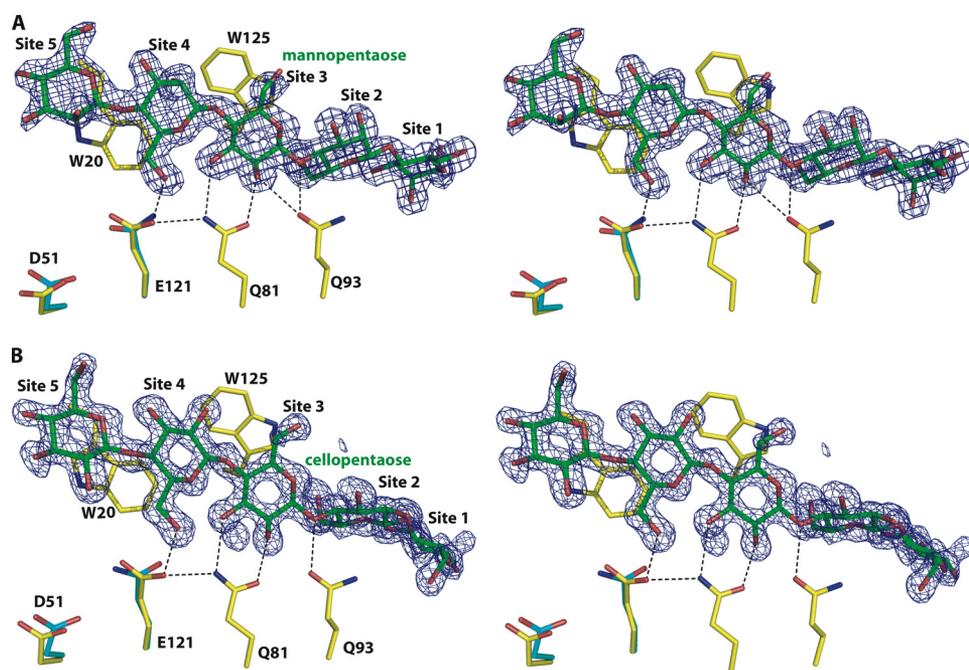


FIGURE 4. Difference Fourier electron density maps (contoured at  $3\sigma$  over background) calculated with coefficients  $F_{\text{obs}} - F_{\text{calc}}$  calculated from phases from the final refined model with the coordinators of the ligand deleted prior to one round of crystallographic refinement. Shown are the corresponding maps derived from the 1.55-Å resolution co-crystal structure of the CBM16-1 Q121E-mannopentaose complex (A), and the 1.35 Å resolution co-crystal structure of the CBM16-1 Q121E-cellopentaose complex (B). The final refined coordinates are superimposed with protein residues colored in yellow and the ligand residues colored in green. For comparison, the coordinates of residues Asp-51 and Gln-121 from the corresponding wild-type structures are also superimposed and colored in cyan. Hydrogen bonds are denoted with black dashed lines.

relative to the mannopentaose co-crystal structure, resulting in a long hydrogen bond (3.5 Å) between the C6 hydroxyl and O $\epsilon$ 2 of Glu-121. This may account for the modestly more favorable enthalpic difference of  $-0.21$  kJ (relative to the wild-type) for the interaction of this variant with cellopentaose over mannopentaose.

**Amino Acid Sequence Analysis**—When we used CBM16-1 to search the GenBank™ protein database, we found that the protein with the top hit score was the CenC domain protein from *Paenibacillus* sp. JDR-2 (GenBank™ accession number YP\_003011283; supplemental Fig. 1). The result from our search was similar for CBM16-2. Despite the important roles in substrate binding of the two tryptophan residues and six polar residues of CBM16-1/-2 along the binding cleft, the amino acid sequence alignment revealed that only Trp-20 is conserved in the CenC domain protein and an endo- $\beta$ -galactosidase (GenBank™ accession number AAD04036), and Trp-125 is replaced by a conservative residue, phenylalanine, in the CenC domain protein. Interestingly, the polar residues are not conserved. Furthermore, we only detected limited conservation of the residues studied in this study in other members of CBM family 16 (supplemental Fig. 1).

## DISCUSSION

The *C. polysaccharolyticus* CBM16-1 and CBM16-2 are the only two proteins within the carbohydrate binding module family 16 with known three-dimensional structures (16). From the co-crystal structure of CBM16-1 with substrate, two tryptophan residues and six polar residues were predicted

to contribute to ligand binding by providing hydrophobic stacking interaction or direct hydrogen bonding interactions. From the current mutational studies, we have demonstrated that the roles of these residues in CBM16-1 as well as the equivalent residues in CBM16-2 in ligand binding are similar. The two tryptophans (Trp-20 and Trp-125) in both CBMs were found to be critical for ligand binding, which is consistent with many studies emphasizing the importance of aromatic residues in the binding cleft of CBMs (5, 34, 35). Two polar residues (Gln-81 and Gln-93) are also critical for ligand binding in both CBM16-1 and CBM16-2, although Gln-81 seems indispensable. The two hydrogen bonds formed between the  $\gamma$ -amino of Gln-81 and the C3 hydroxyl at site 3 and between the  $\gamma$ -carbonyl group of Gln-81 and the C2 hydroxyl at the same site not only ensure that the two CBMs bind to glucose-configured or mannose-configured substrates rather than other substrates

but also provide an essential force to form a stable CBM-substrate complex. Similar to Gln-81, the CBM16-1 Gln-93 is relatively buried at the bottom of the binding cleft, possibly excluding solvent from ligand binding. Similar critical roles for buried polar residues have been observed for *C. thermocellum* Xyn10B CBM22-2 and *C. thermocellum* Xyn11A CBM6 (11, 36). Thus, Trp-20, Trp-125, Gln-81, and Gln-93 collectively form a core binding platform, which interacts with sugar units at four subsites (site 5, 4, 3, and 2) of the pentasaccharides.

In contrast, the other polar residues (Gln-21, Asp-77, Asn-97/Arg-97, and Gln-121/Glu-121) play moderately important roles in ligand binding. These residues are scattered around the binding core, providing auxiliary interactions with the ligand. The four positions contain the nonconserved polar residues (Gln-21/Gly-21, Asn-97/Arg-97, and Gln-121/Glu-121) among the eight positions investigated in this study, suggesting that the residues at these positions may permit manipulation of the binding affinities of the CBMs. Indeed, changing the nonconserved residue Gln-121 to its CBM16-2 equivalent glutamate led to a CBM16-1 mutant (CBM16-1 Q121E) with an increased binding affinity for each of the substrates tested. Although there have been many mutational studies, in regard to amino acid residues predicted to be responsible for substrate binding by type B CBMs, to our knowledge, this is the second example of a single mutation reported to increase the binding affinity for the original substrates of a CBM (5, 34–39). The *P. equi* CBM29-2 K85A was the first example; Lys-85 makes a steric clash with O $_2$  of glucoside (but not man-

noside) 4, explaining why its mutation to alanine increases the affinity for glucose-containing substrates.

To investigate the structural basis for its increased affinity, we have determined high resolution crystal structure of the CBM16-1 Q121E variant in complex with the oligosaccharides mannopentaose and cellopentaose. Although the overall structures of the variant complexes are nearly identical to those of the corresponding wild-type complex, modest movements at and around the side chain of residue 121 results as a consequence of the Q121E mutation. These changes, along with a more favorable electrostatic environment, result in the improved affinity of this variant toward oligosaccharides.

Although residues Trp-20, Trp-125, Gln-81, and Gln-93 coordinately hold the ligand tightly from its upper and lower faces, Gln-21/Gly-21, Asp-77, Asn-97/Arg-97, and Gln-121/Glu-121 are residues that can be fine-tuned for important auxiliary function in binding the ligand. This viewpoint is further supported by the data from the CBM16-1 triple mutant. Although the Q121E mutation increased the binding affinity of CBM16-1 for all substrates, the combination of this mutation with N97R and Q21G mutations, which have negative effects on binding activity by CBM16-1, counteracted the positive effect of Q121E. The fine-tuning process of the key residues has likely taken place in these tandemly linked family 16 CBMs to optimize their function in Man5A. The tandem CBMs are similar to each other in both binding specificity and strength. These similarities might be functionally important for their roles in the intact Man5A.

Distinctive responses for differently configured substrates were observed for some mutants. Specifically, two critical residues, *i.e.* Trp-125 of CBM16-1 and Gln-93 of CBM16-1/-2, have higher residual binding affinities for glucose-configured than mannose-configured substrates after they were mutated to alanine. For the alanine mutants of the polar residues with moderate importance, there were still subtle differences, with a trend of slightly higher affinities remaining for glucose-configured oligosaccharide or polysaccharide for Gln-21, Asp-77, Asn-97, and Gln-121 of CBM16-1, and Asp-77 and Glu-121 of CBM16-2.

Interestingly, the mutants of CBM16-2 seemed to be more susceptible to new H-bond formation, because more mutants (CBM16-2 D77A, R97A, E121A, R97N, and E121Q) either had their binding enthalpy nearly unchanged or increased for both cellopentaose and mannopentaose (Tables 6 and 8). In contrast, CBM16-1 had less number of mutants (Q93A and Q121A) showing this characteristic (Tables 5 and 7). This implies that, although CBM16-1 and CBM16-2 are similar to each other in both primary sequence and spatial structure, CBM16-2 may have a more flexible binding cleft than that of CBM16-1.

Man5A is a robust enzyme that releases oligosaccharides from polysaccharides such as locust bean gum, guar gum,  $\beta$ -mannan, glucomannan, lichenan, and carboxymethyl cellulose. In the hydrolysis of the first four substrates, the enzyme is highly dependent on its two CBMs for release of end products, because deletion of the two CBMs almost abolishes product detection (40, 41). Our preliminary studies suggest that the two CBMs may be important in destabilization of the polysaccharide substrates, thus increasing substrate availability to the cat-

alytic module. We look forward to investigating the effects of the mutations reported in this study on the enzymatic activity of Man5A, especially on glucose-, mannose-, and glucose/mannose-configured polysaccharides. Of significance to our findings in this study is also the observation that the four critical residues (Trp-20, Trp-125, Gln-81, and Gln-93) in the two family 16 CBMs are not highly conserved in the other assigned members of this family. We postulate that most of the members in this family either bind to different substrates or bind to ligands similar to those of the Man5A CBMs but with different strategies.

*Acknowledgments*—We thank Alexandra Sexton for technical assistance. We thank Charles M. Schroeder, Shosuke Yoshida, Yejun Han, Michael Iakiviak, and Young-Hwan Moon of the Energy Biosciences Institute and M. Ashley Spies for scientific discussions. We also thank Keith Brister and Joseph Brunzelle at LS-CAT (23-ID) at Argonne National Laboratories, Advanced Photon Source) for facilitating and technical assistance during crystallographic data collection.

## REFERENCES

- Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6578–6583
- Coughlan, M. P. (1985) *Biotechnol. Genet. Eng. Rev.* **3**, 39–109
- Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) *Biochem. J.* **382**, 769–781
- 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
- Najmudin, S., Guerreiro, C. I., Carvalho, A. L., Prates, J. A., Correia, M. A., Alves, V. D., Ferreira, L. M., Romão, M. J., Gilbert, H. J., Bolam, D. N., and Fontes, C. M. (2006) *J. Biol. Chem.* **281**, 8815–8828
- Henshaw, J., Horne-Bitsch, A., van Bueren, A. L., Money, V. A., Bolam, D. N., Czjzek, M., Ekborg, N. A., Weiner, R. M., Hutcheson, S. W., Davies, G. J., Boraston, A. B., and Gilbert, H. J. (2006) *J. Biol. Chem.* **281**, 17099–17107
- Din, N., Damude, H. G., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A., and Kilburn, D. G. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11383–11387
- Gao, P. J., Chen, G. J., Wang, T. H., Zhang, Y. S., and Liu, J. (2001) *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao* **33**, 13–18
- Mattinen, M. L., Kontteli, M., Kerovuo, J., Linder, M., Annala, A., Lindberg, G., Reinikainen, T., and Drakenberg, T. (1997) *Protein Sci.* **6**, 294–303
- Boraston, A. B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D. R., Kilburn, D. G., and Davies, G. J. (2002) *J. Mol. Biol.* **319**, 1143–1156
- 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) *J. Biol. Chem.* **276**, 48580–48587
- Carvalho, A. L., Goyal, A., Prates, J. A., Bolam, D. N., Gilbert, H. J., Pires, V. M., Ferreira, L. M., Planas, A., Romão, M. J., and Fontes, C. M. (2004) *J. Biol. Chem.* **279**, 34785–34793
- Boraston, A. B., Notenboom, V., Warren, R. A., Kilburn, D. G., Rose, D. R., and Davies, G. (2003) *J. Mol. Biol.* **327**, 659–669
- Flint, J., Nurizzo, D., Harding, S. E., Longman, E., Davies, G. J., Gilbert, H. J., and Bolam, D. N. (2004) *J. Mol. Biol.* **337**, 417–426
- Charnock, S. J., Bolam, D. N., Nurizzo, D., Szabó, L., McKie, V. A., Gilbert, H. J., and Davies, G. J. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14077–14082
- Bae, B., Ohene-Adjei, S., Kocherginskaya, S., Mackie, R. I., Spies, M. A., Cann, I. K., and Nair, S. K. (2008) *J. Biol. Chem.* **283**, 12415–12425
- Blaak, H., Schnellmann, J., Walter, S., Henrissat, B., and Schrempf, H. (1993) *Eur. J. Biochem.* **214**, 659–669
- Leng, L., Zhu, A., Zhang, Z., Hurst, R., and Goldstein, J. (1998) *Gene* **222**, 187–194
- Barbeyron, T., Gerard, A., Potin, P., Henrissat, B., and Kloreg, B. (1998)

## Role of Amino Acids in Family 16 CBM

- Mol. Biol. Evol.* **15**, 528–537
20. Yamagishi, K., Suzuki, K., Imai, K., Mochizuki, H., Morikawa, K., Kyo-gashima, M., Kimata, K., and Watanabe, H. (2003) *J. Biol. Chem.* **278**, 25766–25772
  21. Matsushima, R., Danno, H., Uchida, M., Ishihara, K., Suzuki, T., Kaneniwa, M., Ohtsubo, Y., Nagata, Y., and Tsuda, M. (2010) *Appl. Microbiol. Biotechnol.* **86**, 567–576
  22. Lee, Y. J., Mackie, R. I., Cann, I. K., and Wiegel, J. (2008) *Int. J. Syst. Evol. Microbiol.* **58**, 666–670
  23. Tomme, P., Creagh, A. L., Kilburn, D. G., and Haynes, C. A. (1996) *Biochemistry* **35**, 13885–13894
  24. Takeo, K. (1984) *Electrophoresis* **5**, 187–195
  25. Whitmore, L., and Wallace, B. A. (2008) *Biopolymers* **89**, 392–400
  26. Johnson, W. C. (1999) *Proteins* **35**, 307–312
  27. Otwinowski, Z., Borek, D., Majewski, W., and Minor, W. (2003) *Acta Crystallogr. A* **59**, 228–234
  28. McCoy, A. J. (2007) *Acta Crystallogr. D Biol. Crystallogr.* **63**, 32–41
  29. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J. Appl. Crystallogr.* **40**, 658–674
  30. Perrakis, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53**, 448–455
  31. McRee, D. E. (1999) *J. Struct. Biol.* **125**, 156–165
  32. Brünger, A. T. (1992) *Nature* **355**, 472–475
  33. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR* **8**, 477–486
  34. Pell, G., Williamson, M. P., Walters, C., Du, H., Gilbert, H. J., and Bolam, D. N. (2003) *Biochemistry* **42**, 9316–9323
  35. Flint, J., Bolam, D. N., Nurizzo, D., Taylor, E. J., Williamson, M. P., Walters, C., Davies, G. J., and Gilbert, H. J. (2005) *J. Biol. Chem.* **280**, 23718–23726
  36. Xie, H., Gilbert, H. J., Charnock, S. J., Davies, G. J., Williamson, M. P., Simpson, P. J., Raghothama, S., Fontes, C. M., Dias, F. M., Ferreira, L. M., and Bolam, D. N. (2001) *Biochemistry* **40**, 9167–9176
  37. Notenboom, V., Boraston, A. B., Chiu, P., Frelove, A. C., Kilburn, D. G., and Rose, D. R. (2001) *J. Mol. Biol.* **314**, 797–806
  38. Xie, H., Bolam, D. N., Nagy, T., Szabó, L., Cooper, A., Simpson, P. J., Lakey, J. H., Williamson, M. P., and Gilbert, H. J. (2001) *Biochemistry* **40**, 5700–5707
  39. Henshaw, J. L., Bolam, D. N., Pires, V. M., Czjzek, M., Henrissat, B., Ferreira, L. M., Fontes, C. M., and Gilbert, H. J. (2004) *J. Biol. Chem.* **279**, 21552–21559
  40. Cann, I. K., Kocherginskaya, S., King, M. R., White, B. A., and Mackie, R. I. (1999) *J. Bacteriol.* **181**, 1643–1651
  41. Han, Y., Dodd, D., Hespen, C. W., Ohene-Adjei, S., Schroeder, C. M., Mackie, R. I., and Cann, I. K. (2010) *J. Bacteriol.* **192**, 4111–4121

**Mutational Insights into the Roles of Amino Acid Residues in Ligand Binding for Two Closely Related Family 16 Carbohydrate Binding Modules**

Xiaoyun Su, Vinayak Agarwal, Dylan Dodd, Brian Bae, Roderick I. Mackie, Satish K. Nair and Isaac K. O. Cann

*J. Biol. Chem.* 2010, 285:34665-34676.

doi: 10.1074/jbc.M110.168302 originally published online August 25, 2010

---

Access the most updated version of this article at doi: [10.1074/jbc.M110.168302](https://doi.org/10.1074/jbc.M110.168302)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2010/09/02/M110.168302.DC1.html>

This article cites 41 references, 15 of which can be accessed free at <http://www.jbc.org/content/285/45/34665.full.html#ref-list-1>