# A novel mechanism of xylan binding by a lectin-like module from *Streptomyces lividans* xylanase 10A

Alisdair B. BORASTON*†‡, Peter TOMME*†‡, Emily A. AMANDORON*‡ and Douglas G. KILBURN*†‡[1]

*The Protein Engineering Network of Centres of Excellence, PENCE Inc., National Business Centre, 750 Heritage Medical Research Centre, Edmonton, Alberta, Canada T6G 2S2, †Department of Microbiology and Immunology, University of British Columbia, #300-6174 University Boulevard, Vancouver, British Columbia, Canada V6T 1Z3, and ‡The Biotechnology Laboratory, #237-6174 University Boulevard, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z3

The C-terminal module of xylanase 10A from *Streptomyces lividans* is a family 13 carbohydrate-binding module (CBM13). CBM13 binds mono- and oligo-saccharides with association constants of $\approx 1 \times 10^2$ $M^{-1}$–$1 \times 10^3$ $M^{-1}$. It appears to be specific only for pyranose sugars. CBM13 binds insoluble and soluble xylan, holocellulose, pachyman, lichenan, arabinogalactan and laminarin. The association constant for binding to soluble xylan is $(6.2 \pm 0.6) \times 10^3$/mol of xylan polymer. Site-directed mutation indicates the involvement of three functional sites on CBM13 in binding to soluble xylan. The sites are similar in sequence, and are predicted to have similar structures, to the $\alpha$, $\beta$ and $\gamma$ sites of ricin toxin B-chain, which is also in family 13. The affinity of a single binding site on CBM13 for soluble xylan is only $\approx (0.5 \pm 0.1) \times 10^3$/mol of xylan. The binding of CBM13 to soluble xylan involves additive and co-operative interactions between the three binding sites. This mechanism of binding has not previously been reported for CBMs binding polysaccharides. CBM13 is the first bacterial module from family 13 to be described in detail.

Key words: carbohydrate-binding module, cellulose-binding domain, association, protein–sugar interactions, ricin.

## INTRODUCTION

Carbohydrate-binding modules (CBMs), some of which were previously called cellulose-binding domains (CBDs), are discrete modules in many glycosyl hydrolases [1]. They are currently classified into 18 CBM families of related amino acid sequence [2–5].

Family 13 of CBMs, also referred to as the 'ricin superfamily' or 'R-type lectins' [6], is very diverse, containing members from plants, animals and microbes [2]. The only members characterized to date are the $\beta$-trefoil lectins from *Ricinus communis* (castor bean) (B-chains from ricin agglutinin and toxin, RCA and RTB respectively), *Abrus precatorius* (Jequirity bean) (*A. precatorius* agglutinin, APA), *Sambucus nigra* (elder) and *Viscum album* (mistletoe). They consist of tandem family 13 modules. Each module comprises three repeats of approx. 42 amino acids (the $\alpha$, $\beta$ and $\gamma$ domains), some of which have sugar-binding sites specific for galactose, *N*-acetylgalactosamine or sialic acid. This multivalency enhances the affinity for target glycans [7].

The endo-$\beta$-1,4 xylanase 10A (Xyn10A) from *Streptomyces lividans* includes an N-terminal catalytic module and a 130-amino-acid-long C-terminal module [8,9]. The C-terminal module binds insoluble xylan, enhancing the binding of the holoenzyme [9]. Its sequence and domain organization are similar to those of RTB [2,9], suggesting it is also a multivalent family 13 CBM, referred to as CBM13. This paper describes the binding characteristics of CBM13 and relates them to those of the lectin members of family 13.

## MATERIALS AND METHODS

### Carbohydrates and polysaccharides

Microcrystalline cellulose (Avicel™ PH101) was obtained from FMC International (Little Island, County Cork, Ireland). Bacterial microcrystalline cellulose (BMCC) was prepared from cultures of *Acetobacter xylinum* (A. T. C. C. 23769) as described previously [10]. Phosphoric acid-swollen Avicel™ (PASA) was prepared as described previously [11]. D-Glucose, *N*-acetyl-D-glucosamine, D-galactose, 6-deoxy-L-galactose (L-fucose), D-mannose, 6-deoxy-L-mannose (L-rhamnose), L-arabinose, D-ribose, D-xylose, cellobiose, lactose, gentibiose, raffinose, chitin (from shrimp shells), lichenan, yeast mannan, oat-spelt xylan, larchwood (*Larix*) xylan, soluble starch (from potato, *Solanum tuberosum*) and arabinogalactan were from Sigma. Hydroxyethylcellulose (HEC, viscosity $\approx$ 0.08–0.15 Pa [2% (w/v) solution)] was from Aldrich. Maltose was from Fisher Scientific. Barley (*Hordeum vulgare*) $\beta$-glucan, pachyman (Lot MPA80801), xylo- and arabino-oligosaccharides (> 95% pure), arabinan (from sugar beet, *Beta vulgaris*), linear L-arabinan, and pectic galactan (potato) and galactan (potato) were from MegaZyme Ltd. (North Rocks, N. S. W., Australia). Birchwood (*Betula alba*) xylan (Roth 7500; molecular mass $\approx$ 25000) was from Carl Roth RG (Karlsruhe, Germany). The xylans were separated into water-soluble and water-insoluble fractions [12,13]. Other polysaccharides (pachyman, potato galactan, linear arabinan and lichenan) used in insoluble binding

Abbreviations used: ABFB, arabinofuranosidase B from *Streptomyces lividans*; APA, *Abrus precatorius* (Jequirity bean) agglutinin; BMCC, bacterial microcrystalline cellulose; CBM, carbohydrate-binding module; CBM13, family 13 carbohydrate-binding module from *Strep. lividans* xylanase 10A; CBM2b-1, internal family 2b xylan-binding module from *Cellulomonas fimi* xylanase 11A; DP, degree of polymerization; HEC, hydroxyethylcellulose; EHEC, ethylhydroxyethylcellulose; HBAH, hydroxybenzoic acid hydrazide; IMAC, immobilized-metal affinity chromatography; MALDI-TOF MS, matrix-assisted laser-desorption time-of-flight MS; PASA, phosphoric acid-swollen Avicel™; RTB, ricin toxin B-chain from *Ricinus communis* (castor bean); RCA, *R. communis* agglutinin; Xyn10A, xylanase 10A from *Strep. lividans*; $T_m$, melting temperature; STI, soybean trypsin inhibitor.

[1] To whom correspondence should be addressed (e-mail kilburn@interchange.ubc.ca).

assays were washed with buffer prior to use to reduce the amount of associated soluble material. Holocellulose [from pine (*Pinus*)] was kindly provided by Dr. Jack Saddler of the Department of Wood Science of the University of British Columbia.

### DNA amplification and cloning

Genomic DNA from *S. lividans* was prepared as described elsewhere [14]. The gene fragment encoding the 132 residues of CBM13 was obtained and amplified by PCR. Appropriate restriction sites were introduced at the 5′ and 3′ ends of the CBM13-encoding-gene fragment for cloning in the pTug expression vectors [15,16]. Each PCR mixture (50 $\mu$l total) contained 25–50 ng of genomic DNA, 25–50 pmol of primers, 10% DMSO, 0.4 mM 2′-deoxynucleoside 5′-triphosphates and 1 unit of *Pwo* DNA polymerase in buffer (Boehringer Mannheim). A protocol of 20 successive cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 30 s, and primer extension at 72 °C for 1.5 min was followed. An *Nhe*I site (underlined) was introduced at the 5′ end of the CBM13-encoding gene fragment, using the oligonucleotide 5′-TGACTTGACGTCCGCTAGCGAGCCCCCCGCGGACGGG-3′ as primer. A *Hin*dIII (underlined) restriction site was introduced at the 3′ end of the CBM13 sequence using the oligonucleotide 5′-TGACGAGCGGCCGCAAGCTTATCAGGTGCGGGTCCAGCGTTG-3′ as primer. The resulting 0.43 kb PCR fragment was digested with *Nhe*I and *Hin*dIII and cloned in-frame with the sequence encoding the Cex leader peptide and the hexahistidine tail in pTugKH previously digested with the same restriction enzymes, to give pTugCBM13. DNA was sequenced by the AmpliTaq dye-termination-cycle sequencing protocol and an Applied Biosystems model 377 sequencer by the Nucleic Acids and Protein Services Unit in the Biotechnology Laboratory of this University.

Site-directed mutations were introduced by 'mega-primer' PCR. The first PCR step to introduce the mutation was performed using the mutagenic primer and the appropriate 5′ or 3′ flanking primer that was used to amplify the fragment containing the mutation. The conditions outlined above were used for amplification. Amplified products were purified using Qiaex II (Qiagen, Chatsworth, CA, U.S.A.) after electrophoresis through a 1%-agarose gel. This amplified DNA was then employed as a 'mega-primer' in a second PCR reaction utilizing the appropriate 5′ or 3′ flanking primer to amplify the entire desired gene fragment. The gene fragments encoding the mutated *CBM13* genes were then inserted into pTugKH as described above. Restriction digests targeting silent restriction sites inserted into the mutagenic primer sequences were used to screen for positive clones containing the mutations. Constructs were sequenced as described above.

### Protein purification

Overnight cultures of *Escherichia coli* strain JM101/pTugCBM13 were diluted 500-fold in tryptone/yeast extract/phosphate (TYP) [17] medium supplemented with 100 $\mu$g of kanamycin/ml, and grown at 30 °C to a cell density ($D_{600}$) of 0.3. Isopropyl 1-thio-$\beta$-D-galactopyranoside was added to a final concentration of 0.2 mM and incubation was continued for a further 36 h at 30 °C. The cells were harvested by centrifugation (8500 ***g***) for 10 min at 4 °C and resuspended to about 1/50 of the original culture volume in 20 mM Tris/HCl buffer, pH 7.9, containing 10 mM imidazole and 0.5 M NaCl. Cells were ruptured by two passages through a French pressure cell (144.9 MPa; 21000 lbf/in$^2$) and cell debris was removed by centrifugation for 20 min at 27000 ***g*** and 4 °C. CBM13 was purified from the clarified cell extract by immobilized-metal affinity chromatography (IMAC) using a column containing 10 ml of His-Bind resin (50% slurry) (Novagen, Milwaukee, WI, U.S.A.) to give a final bed volume of $\approx$ 5 ml. Standard column preparation and purification were employed as recommended by the resin supplier. Protein fractions were analysed for purity on 20% Phast gels (Pharmacia). Pure CBM13 fractions were pooled and exchanged into the appropriate buffer and concentrated in a stirred ultrafiltration unit (Amicon, Beverly, MA, U.S.A.) with a 1 kDa cut-off filter (Filtron, Northborough, MA, U.S.A.). The purity was estimated by SDS/PAGE and matrix-assisted laser-desorption time-of-flight MS (MALDI-TOF MS).

### Protein determination

The concentration of purified protein was determined by $A_{280}$ using a calculated molar absorption coefficient [18] of 32342 $M^{-1} \cdot cm^{-1}$.

### Thermal denaturation

All thermal denaturations were performed on a Cary 100e UV–visible spectrophotometer (Varian, Melbourne, Australia). UV difference spectra were taken by collecting a baseline from 230 nm to 330 nm at 25 °C using 800 $\mu$l of polypeptide at 15 $\mu$M in 25 mM Tris/HCl, pH 7.4. The temperature of the cuvette block was ramped to 80 °C and the sample allowed to equilibrate for 20 min. The sample was rescanned over the same wavelength range.

Denaturation experiments were performed using 800 $\mu$l samples of polypeptide at 15 $\mu$M in 25 mM Tris/HCl, pH 7.4. The temperature of the cuvette block was ramped at 2 °C/min from 25 to 75 °C. Data were collected at a wavelength of 272 nm at 0.2 °C intervals, with the baseline value taken at 25 °C. Linear regression was used to fit trend lines to the linear pre- and post-transition portions of the raw denaturation curve. The results of the linear regression were used to predict absorbance values for fully native and fully unfolded polypeptide over the complete temperature interval. The fraction of native protein was determined by using the equation:

$$f_n = (r - r_u)/(r_n - r_u)$$

where $r$ is the measured absorbance value. $r_n$ and $r_u$ correspond to the predicted absorbance values of fully native and fully unfolded polypeptide respectively, at a given temperature. Melting-temperature ($T_m$) values were determined graphically by determining the temperature at which 50% of the polypeptide was unfolded.

### Fluorescence analysis of protein–carbohydrate binding

All fluorescence experiments were performed on a Perkin–Elmer LS-50 luminescence spectrometer. Fluorescence emission scans were performed using 5 $\mu$M CBM13 with or without 25 mM ligand in 25 mM Tris/HCl buffer, pH 7.5. The excitation wavelength was 280 nm. Emission intensities were collected over the wavelength range of 300–400 nm. The excitation and emission slit widths were 5 nm. Five scans were averaged.

Solute quenching experiments were performed using KI and 5 $\mu$M CBM13. Ionic strength was kept constant by the addition of NaCl to samples such that the final concentration of salt (total KI and NaCl concentration) was 1 M. Samples were buffered with 25 mM Tris/HCl buffer, pH 7.5. The excitation wavelength was 295 nm. Emission intensities were collected at 350 nm with integration times of 2 s. The excitation and emission slit widths

were 5 nm. Four readings were averaged. Data were plotted as relative fluorescence intensity ($F_0/F$, where $F_0$ is the fluorescence intensity in the absence of KI; $F$ is the fluorescence intensity with KI) against the KI concentration [19].

Quantitative binding experiments were performed by adding amounts of the test carbohydrate in 25 mM Tris/HCl buffer, pH 7.5, to 800 $\mu$l of CBM13 (5 $\mu$M in the same buffer) with continuous stirring. Fluorescence intensity was measured 2 min after each addition of ligand using excitation wavelengths of 275, 285, and 295 nm and measuring emission at 350 or 330 nm wavelengths and a slit width of 5 or 10 nm. Three to five 30 s integration periods were averaged for each data point. The fluorescence yield was constant over this time. The emission spectra of all solutions were corrected for background fluorescence caused by buffer and carbohydrates and for dilution and inner-filter effects [20] as required. Relative fluorescence ($F/F_0$, where $F$ is the initial fluorescence; $F_0$ is the fluorescence in the presence of carbohydrate) was plotted against carbohydrate concentration. Association binding constants $K_a$ ($M^{-1}$) and the maximal fluorescence change ($F_{max}$) in the protein upon full complexation with sugar ligand were derived by a non-linear least-squares fit of the corrected data to a one-site binding model using Origin v.5.0 (Microcal, Northampton, MA, U.S.A.) software. To evaluate the requirement for metal ions for binding, $CaCl_2$ or $MgCl_2$ (0–1 M) was added to both carbohydrate (xylose) and protein solutions. The pH-dependence for binding of xylose by CBM13 was determined in 25 mM citrate/phosphate/borate buffer, pH 3.0–10.5, as described above.

### Affinity electrophoresis

Qualitative and quantitative binding of CBM13 to soluble polysaccharides was evaluated by affinity electrophoresis [21,22] in 10 % (w/v) polyacrylamide gels polymerized in the absence or presence of various amounts of polysaccharide (usually 0–1 %, w/v). Electrophoresis was for 1.5 h at 4 °C, pH 8.8 and 150 V in a Mini PROTEAN II system (Bio-Rad). To prevent the soybean trypsin inhibitor (STI, 10 $\mu$g) used as an internal standard from running off the gel, it was loaded after 45 min of electrophoresis. After electrophoresis the gels were stained with Coomassie Blue. The migration distances of CBM13 and the reference protein were measured directly on the gels and used to determine the association constants ($K_a$) as described below.

Dissociation constants ($K_d$) were obtained from plots of $1/(R_0-r)$ versus $1/C$ according to the affinity equation [21]:

$$1/(R_0-r) = 1/(R_o-R_c)(1+K_d/C)$$

where $r$ is the relative migration distance of the CBM in the presence of ligand in the gel, $R_o$ is the relative migration distance of the free CBM in the absence of ligand, $R_c$ is the relative migration distance of the complex at high excess of ligand, where all CBM molecules are fully complexed, $C$ is the molar concentration of the ligand in the gel and $K_d$ is the dissociation constant of CBM for the macromolecular ligand. $K_d$ values were determined as the inverse of the absolute value of the intercept on the abscissa of data plotted according to the affinity equation. $K_a$ values were calculated as the reciprocals of the $K_d$ values. All migration distances of CBM were measured relative to the migration of the reference protein (STI).

Interaction of CBM13 with low-molecular-mass ligands was assessed by competition affinity electrophoresis [21] in gels containing both soluble birchwood xylan (0.43 mM or 2.5 g/l) and a competing mono- or di-saccharide (0.1 M).

### MALDI-TOF MS

Purified CBM13, drop-dialysed overnight into water using a 0.025 $\mu$m vs-membrane (Millipore) was diluted in water to 40 pmol/$\mu$l and mixed 1:1 with a saturated matrix solution of sinapinic acid (Sigma) in 70 % acetonitrile and 0.1 % trifluoroacetic acid. A 1 $\mu$l portion was spotted on to the MALDI target and allowed to dry. The mass was obtained by positive-ion MALDI-TOF MS on a SELDI-MassPhoresis system (Ciphergen, Palo Alto, CA, U.S.A.). Bovine superoxide dismutase (12230.6 Da) was used as a calibrant.

### Adsorption assays on insoluble polysaccharides

For qualitative adsorption experiments, 75 $\mu$g of purified CBM13 were mixed end-over-end at 4 °C with BMCC (2 mg), Avicel (10 mg), PASA (2 mg), pine holocellulose (20 mg), oat-spelt or birchwood xylan (10 mg) or linear arabinan (5 mg) in a final volume of 1 ml of potassium phosphate buffer (50 mM, pH 7.0). After 1 h, polysaccharides were collected by centrifugation (13000 ***g***, 4 °C) and washed with 1 ml of potassium phosphate buffer (50 mM, pH 7.0). The washing step was repeated three times. After the last washing step, insoluble polysaccharides were collected by centrifugation and boiled for 5 min after addition of 40 $\mu$l of SDS loading buffer. Fractions (20 $\mu$l each) were then analysed for protein by SDS/PAGE through 13 % (w/v) gels.

For semi-quantitative binding analyses (percentage of CBM13 bound for a fixed amount of each polymer), 75 $\mu$g of CBM13 were incubated with 1 mg of each polysaccharide as described above. After 1 h incubation, polysaccharides were removed by centrifugation (15 min at 13000 ***g***, 4 °C) and unbound protein left in the supernatants was measured ($A_{280}$) and used to calculate the amount of CBM bound to each polysaccharide as described previously [22,23].

### Sugar analysis

The total sugar concentration of mono- and di-saccharide solutions and of the soluble polysaccharides was determined by the phenol/$H_2SO_4$ method [24]. Reducing sugar was measured with hydroxybenzoic acid hydrazide (HBAH) [25]. Appropriate monosaccharide standards were used for both assays.

## RESULTS AND DISCUSSION

### Production and purification of CBM13

CBM13 (purity > 95 %) was obtained in yields of 60–80 mg/l. Its calculated molecular mass was 15250.5 Da; its experimentally determined molecular mass was 15240.9 Da, a value within the 0.1 % error of the mass spectrometer.

### Amino acid similarity to ricin

Each module of RTB has a $\beta$-trefoil fold – a reflection of repeated domain architecture. Of the six domains, only the 1$\alpha$, 1$\beta$ and 2$\gamma$ domains of RTB have functional sugar-binding sites [26–30]. The remaining domains contain regions with similar conformations to the functional binding sites, but lack key amino acid residues required for binding.

CBM13 contains three repeated sequences (domains) of 42 amino acids that share 55 % sequence similarity with one another,

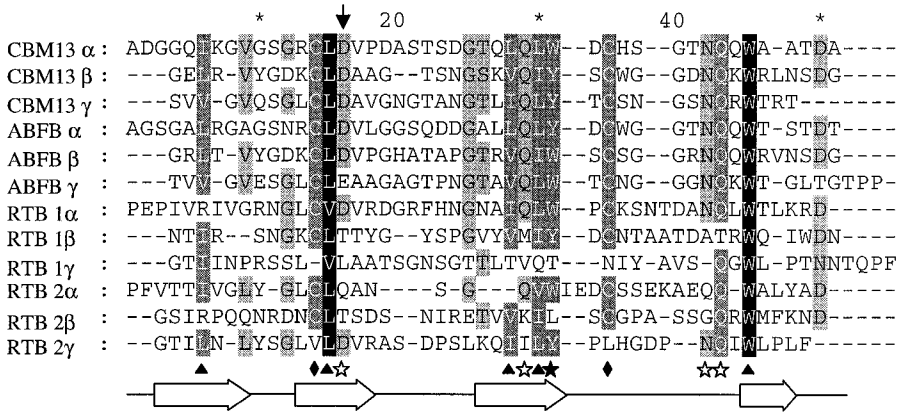


**Figure 1 Amino acid alignment of the $\alpha$, $\beta$ and $\gamma$ repeats of selected family 13 carbohydrate-binding modules**

CBM13 is the binding module from xylanase A from *Strep. lividans*; ABFB, the binding module from arabinofuranosidase B from *Strep. lividans*; RTB 1, module 1 of ricin toxin B chain; RTB 2, module 2 of ricin toxin B chain. $\alpha$, $\beta$ and $\gamma$ refer to the repeated sequences as in ricin toxin B chain. ☆ indicates residues involved in substrate binding in RTB; ★ Denotes the aromatic residues involved in substrate binding. ▲ Denotes residues forming the hydrophobic core of RTB. ◆ Indicates cysteine residues forming disulphide bonds (no disulphide bonds are present in the $\gamma$ domains of RTB). The arrow above the sequences indicates the residues targeted for mutation in CBM13. Regions of $\beta$-sheet are indicated beneath the sequences by arrows. Connecting lines indicate loops. This alignment was prepared using ClustalW [52].

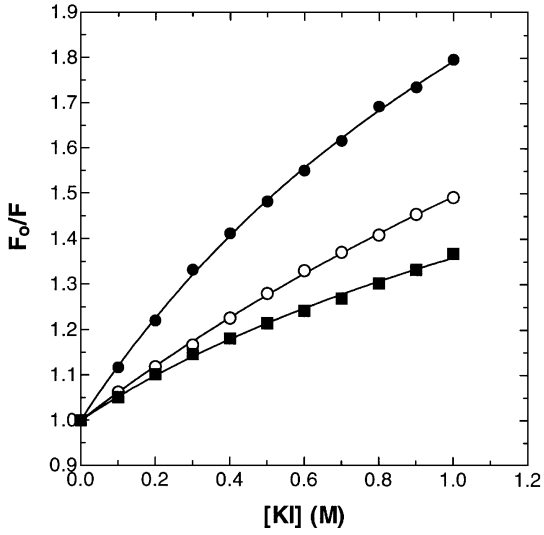


**Figure 2 Stern–Volmer analysis of CBM13 in the absence of ligand (●), the presence of 25 mM xylose (○) or the presence of 25 mM galactose (■)**

$F_0$ is the initial fluorescence of the sample in the absence of KI. $F$ is the fluorescence in the presence of quencher. All samples were in 25 mM Tris/HCl, pH 7.5.

greater than 50% sequence similarity with the repeated domains of the family 13 CBM from *S. lividans* arabinofuranosidase B (ABFB), and up to 70% sequence similarity with the domains of RTB (Figure 1). Comparing CBM13 and RTB, the repeats of CBM13 share the greatest similarity with the 1$\alpha$ domain of RTB.

The amino acids in all three domains of RTB that form the hydrophobic core of the $\beta$-trefoil fold are well conserved in CBM13 (Figure 1). The disulphide-bonded cysteine pairs in the $\alpha$ and $\beta$ domains of both RTB modules are also conserved in CBM13. The $\gamma$ domain of CBM13 has an additional pair of cysteine residues. There are no free thiol groups in CBM13 [9], suggesting that CBM13 has three disulphide bonds, one in each domain. The repeated three-domain motif and conservation of important structural residues imply that CBM13 shares a similar $\beta$-trefoil fold with RTB.

The binding residues in the 1$\alpha$ site of RTB are highly conserved in all three domains of CBM13. In particular, a key aspartic acid residue (D22 of RTB) that forms strong hydrogen bonds with the $C^3$ and $C^4$ hydroxyl groups of galactose, and a tryptophan residue that interacts with the B-face of the galactose, are conserved [26] (Figure 1).

On the basis of the implied structural similarities with RTB and the conservation of binding residues, it appeared likely that CBM13 could bind monosaccharides, such as galactose, and has three sites involved in substrate binding.

## Mono-, di- and oligo-saccharide binding by CBM13

The binding of most sugars tested to CBM13 shifted the emission maximum from 335 to 330 nm and decreased the emission intensity. Galactose and lactose caused the same shift, but the emission intensity increased (results not shown). Stern–Volmer plots of the relative fluorescence of CBM13 at different concentrations of KI were concave, possibly indicating classes of tryptophan residues with differing solvent exposure (Figure 2) [19]. Xylose and galactose reduced the quenching of tryptophan fluorescence by KI (Figure 2), as did arabinose and ribose (results not shown).

The emission spectrum of CBM13 was dominated by tryptophan, despite the presence of three tyrosine residues in the polypeptide. The blue-shifted wavelength of maximum fluorescence emission upon sugar binding suggested that the environment of a tryptophan residue(s) became more hydrophobic. Similarly, the reduced quenching of tryptophan fluorescence by KI in the presence of sugar indicated the shielding of an exposed tryptophan(s). These changes could result from the stacking of a sugar against a tryptophan residue – an interaction common to many protein↔carbohydrate interactions [31]. Only the putative binding site in the $\alpha$ domain of CBM13 contains a tryptophan

**Table 1 Specificity and affinity of CBM13 for soluble sugars**

Abbreviations used: GlcNAc, *N*-acetylglucosamine; GlcD, D-glucose; XylD, D-xylose; GalD, D-galactose; AraL, L-arabinose; FruD, D-fructose.

| Sugar | Composition* | $10^2 \times K_a$ ($M^{-1}$)† | $-\Delta G$ (kJ/mol)‡ |
|---|---|---|---|
| D-Mannose | – | 0.6 ± 0.0 | 9.9 |
| L-Fucose | – | 1.3 ± 0.1 | 11.9 |
| D-Xylose | – | 1.4 ± 0.0 | 12.0 |
| L-Arabinose | – | 1.5 ± 0.1 | 12.3 |
| D-Glucose | – | 1.7 ± 0.0 | 12.4 |
| D-GlcNAc | – | 3.7 ± 0.1 | 14.4 |
| D-Galactose | – | 6.1 ± 0.8 | 15.6 |
| D-Ribose | – | 6.4 ± 0.2 | 15.8 |
| L-Rhamnose | – | 10.4 ± 0.1 | 16.9 |
| Cellobiose | GlcD($\beta$1-4)GlcD | 0.3 ± 0.0 | 8.4 |
| Maltose | GlcD($\alpha$1-4)GlcD | 1.5 ± 0.0 | 12.2 |
| Xylobiose | XylD($\beta$1-4)XylD | 1.8 ± 0.1 | 12.7 |
| Gentibiose | GlcD($\beta$1-6)GlcD | 4.6 ± 0.1 | 14.9 |
| Lactose | GalD($\alpha$1-4)GlcD | 9.7 ± 0.4 | 16.8 |
| Arabinotriose | $[(\alpha 1\text{-}5)AraL]_3$ | 1.4 ± 0.1 | 12.0 |
| Arabinohexaose | $[(\alpha 1\text{-}5)AraL]_6$ | 3.6 ± 0.2 | 14.3 |
| Xylotetraose | $[(\beta 1\text{-}4)XylD]_4$ | 6.2 ± 0.4 | 15.7 |
| Birchwood xylan | $[(\beta 1\text{-}4)XylD]_{44}$ | 62.0 ± 6.0§ | 21.3 |

* Sugar linkage and composition; subscript letters denote DP.

† Errors represent standard errors determined from the non-linear least-squares fit to a one-site binding model.

‡ 20 °C (293 °K) was taken as the reference temperature for the calculation of binding energies.

§ Determined by affinity electrophoresis.

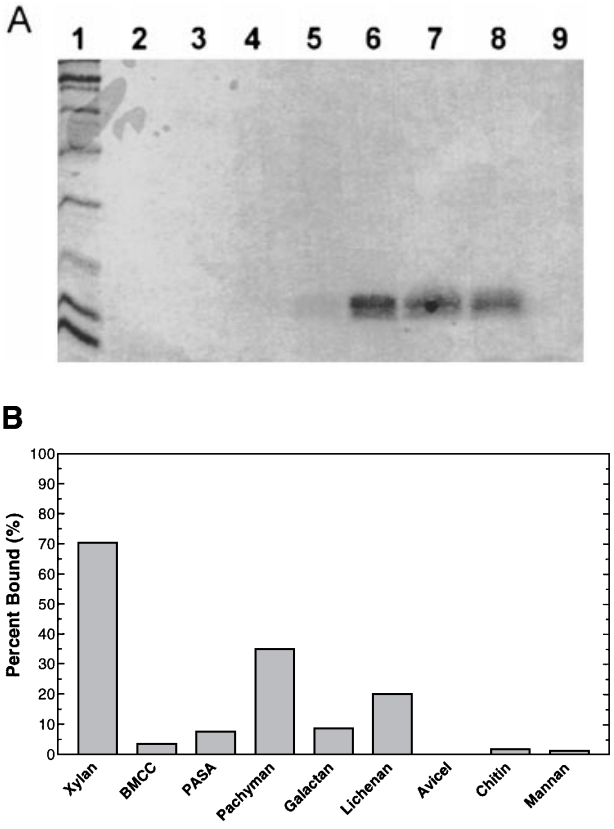


**Figure 3 Insoluble polysaccharide-binding characteristics of CBM13**

(**A**) SDS/PAGE analysis of bound samples desorbed by boiling with SDS/PAGE loading buffer (see the Materials and methods section for sample preparation). Lane 1, molecular-mass standards; lane 2, BMCC; lane 3, Avicel(TM); lane 4, PASA; lane 5, 0.5 $\mu$g of purified CBM13; lane 6, holocellulose; lane 7, birchwood xylan; lane 8: oat-spelt xylan; lane 9, linear arabinan. (**B**) Depletion analysis of CBM13 adsorption to insoluble polysaccharides [the xylan used was birchwood xylan; the mannan was ivory-nut (*Phytelephas macrocarpa*) mannan; the galactan was potato galactan].

residue; the other putative binding sites contain tyrosine residues. Thus the changes in fluorescence emission upon the addition of sugar to CBM13 likely indicate binding at the site in the $\alpha$ domain.

The affinities for monosaccharides varied from less than $1 \times 10^2$ $M^{-1}$ to $1 \times 10^3$ $M^{-1}$ and of the monosaccharides investigated only fructose did not bind (Table 1). The association constants for binding to xylose were unchanged in the presence of 0, 1.0, 10.0, and 100.0 mM $CaCl_2$ or $MgCl_2$ (results not shown). The pH optimum for binding was 7.5 (results not shown).

Although the binding of CBM13 to monosaccharides appears to be relatively weak, it is of the same order as for other sugar ↔ lectin interactions. The association constants of lactose and galactose for CBM13 were approximately one-third of those for the related lectins RTB [28], RCA [32], and APA [33]. The association constants were also on the same order of magnitude as those found for the unrelated *Mordica charantia* (bitter gourd) lectin [34] and many other unmodified carbohydrates binding to lectins [35]. CBM13's broad specificity for monosaccharides is unusual for carbohydrate-binding proteins. The highly directional nature of hydrogen bonding between amino acid residues on the protein and multiple hydroxy groups on the sugar are proposed to determined the specificity of protein ↔ carbohydrate interactions [36,37]. Though this appears to be relaxed with CBM13, the lack of binding specificity is not unprecedented, as demonstrated by the bitter-gourd lectin [34]. Although the literature appears to indicate that RTB is specific for galactose and *N*-acetylgalactosamine, it is unclear whether its binding to the monosaccharides that are substituents of hemicellulose has been investigated. Should RTB represent a truly galactose-specific class of family 13 CBMs, then the specificity of CBM13 is clearly lower, though it does have a slight preference for galactose. This may reflect adaptations in CBM13 that allow it to bind to polysaccharides.

## Polysaccharide binding by CBM13

CBM13 bound to $\beta$1-4-linked xylose polymers (birchwood xylan and oat-spelt xylan), but not polymers of other $\beta$1-4-linked sugars (Figure 3). Xylan is a structural polysaccharide associated with other polysaccharides, mainly cellulose, in plant cell walls [38]. Pine holocellulose, a delignified wood pulp, was used as a substrate more closely resembling the natural substrate of xylanase 10A. CBM13 bound to holocellulose, indicating that a mixed preparation of cellulose and xylan does not alter the presentation of the xylan sufficiently to prevent binding (Figure 3a). There was less, but still significant, binding to pachyman, a (1,3)-$\beta$-D-glucan, and lichenan, a (1,3)(1,4)-$\beta$-D-glucan (Figure 3b). The heterogeneous nature and partial solubility of many of these polysaccharides precluded quantitative assessment of binding. CBM13 bound relatively tightly to soluble birchwood xylan (Figure 4) and arabinogalactan ($\beta$1-3-linked D-galactose with $\alpha$1-3,5-linked L-arabinose substitutions linked $\alpha$1-3 to the galactan backbone) and weakly to arabinan ($\alpha$1-5-linked L-arabinose) and laminarin ($\beta$1-3-linked D-glucose). It did not bind to ethylhydroxyethylcellulose (EHEC), HEC, yeast mannan, agarose,

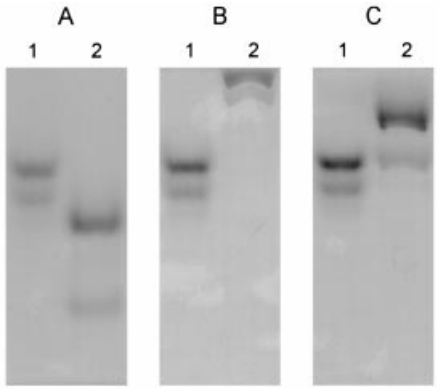


**Figure 4 Affinity electrophoresis of CBM13 on birchwood xylan in the absence and presence of xylose**

(**A**) Native gel without xylan or xylose. (**B**) Electrophoresis in the presence of 0.25% (w/v) (0.43 mM) birchwood xylan. (**C**) Electrophoresis in the presence of 0.25% w/v (0.43 mM) birchwood xylan and 0.1 M xylose as competing ligand. Lane 1, STI (10 $\mu$g); lane 2, CBM13 (10 $\mu$g).

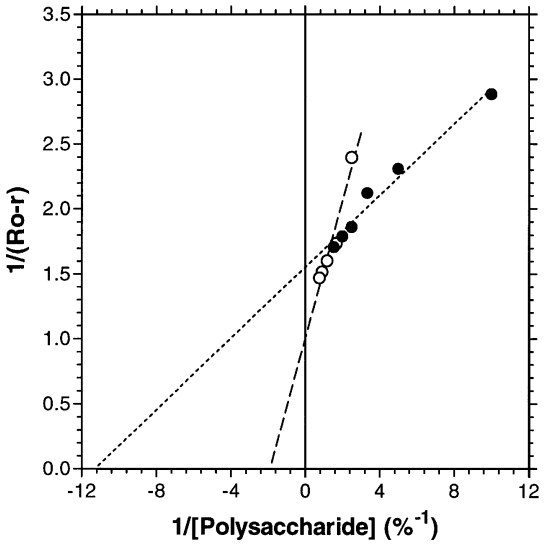


**Figure 5 Quantitative affinity electrophoresis of CBM13 binding to xylan and arabinogalactan**

Samples (10 $\mu$g) of CBM13 were run on affinity gels with increasing concentrations of either xylan (●) or arabinogalactan (○). Relative mobility values ($r$) were calculated for CBM13 relative to a non-binding standard (10 $\mu$g of soybean trypsin inhibitor). $R_0$ corresponds to the relative mobility of CBM13 in the absence of polysaccharide. $1/(R_0 - r)$ values were plotted against the reciprocal of the polysaccharide concentration and the association constants determined as the absolute value of the intercept on the abscissa.

dextran, barley $\beta$-glucan, oat $\beta$-glucan, starch or pectic galactan. The partitioning values for the binding of CBM13 to xylan and arabinogalactan were 1.13 l/g and 0.19 l/g respectively (Figure 5). In order to determine the molar association constants, the average molecular masses of the polysaccharides must be known. This was calculated for xylan by determining the average degree of polymerization (DP) from a comparison of total and reducing sugars. The average DP of the xylan preparation was 44, giving

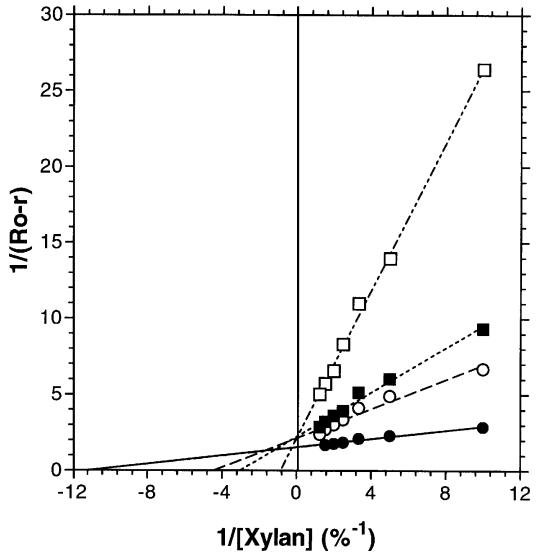


**Figure 6 Quantitative affinity electrophoresis of CBM13 mutants binding to xylan**

The values calculated for the association constants were $(6.2 \pm 0.6) \times 10^3$, $(1.8 \pm 0.2) \times 10^3$, $(2.6 \pm 0.3) \times 10^3$, and $(0.5 \pm 0.1) \times 10^3$/mol of xylan polymer for wild-type (●), D58A (○), D99A (■) and D16A/D99A (□) respectively. See the Materials and methods section for experimental details.

an average molecular mass of approx. 5826 Da and a molar association constant of $(6.2 \pm 0.6) \times 10^3$ $M^{-1} \cdot mol^{-1}$ of xylan polymer. Owing to the nature of affinity electrophoresis, this value does not reflect the polydispersity of the xylan preparation, but rather the average DP, and it does not account for potential multiple binding sites on the polymer – an exclusion that may result in an apparently higher association constant. Its substituents precluded the same calculation for arabinogalactan: the many branches would have led to an incorrect DP, hence an erroneous association constant. Birchwood xylan is relatively unsubstituted [39,40], so the estimate of its DP was valid. The association constant for CBM13 binding to xylan compared well with the internal family 2b CBM, CBM2b-1 (or XBD1), from xylanase 11A of *C. fimi*, which binds xylan with an association constant of $2.5 \times 10^3$ $M^{-1}$ [41].

D-Xylose, L-arabinose, D-galactose, D-mannose, D-glucose, D-ribose, and cellobiose, but not fructose, prevented the binding of CBM13 to xylan (Figure 4). Thus it appeared that CBM13 bound xylan and small sugars at either the same binding site(s) or overlapping binding sites.

It is noteworthy that CBM13 has a higher specificity for polysaccharides than it does for monosaccharides. CBM13 appears to be unable to bind polymers of a sugar that it would otherwise bind as a monosaccharide (e.g. CBM13 binds galactose, but not galactan). The reasons for this are unknown, but masking of the recognition determinants by either a particular glycosidic linkage or the conformation adopted by the polysaccharide seem likely factors.

### Role of multiple binding sites in xylan binding by CBM13

Lectins often contain contiguous modules with multiple binding sites or form multimeric complexes consisting of two or more non-covalently linked modules. The association constants of

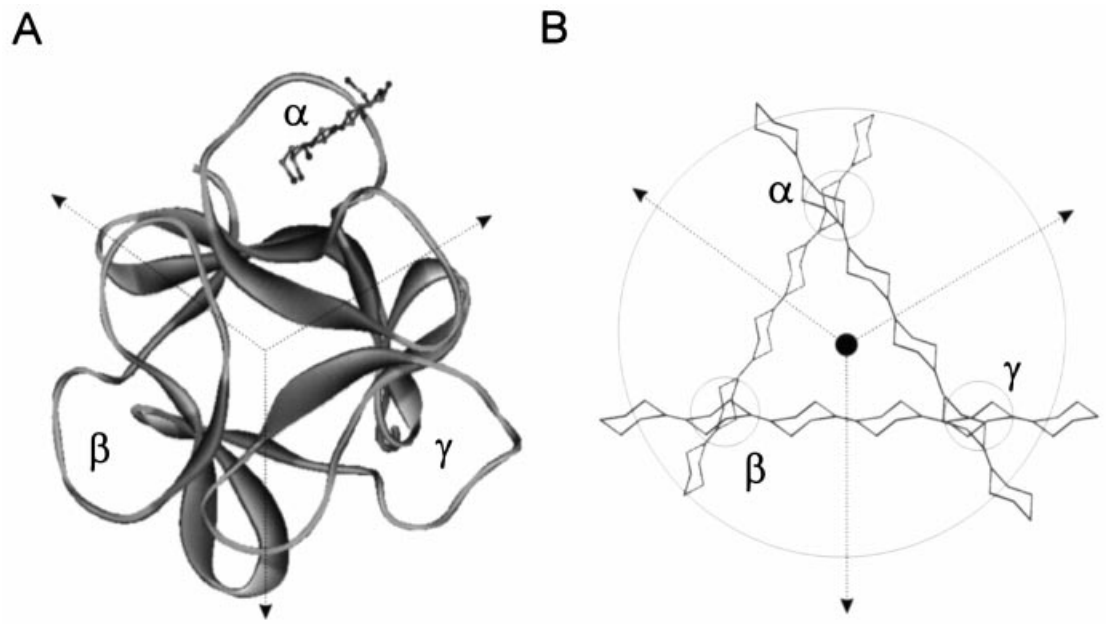


**Figure 7 Schematic domain architecture of CBM13 in comparison with ricin toxin B-chain**

(**A**) Structure of module 1 of RTB shown with lactose bound in the $\alpha$ site. (**B**) Schematic structure of CBM13 shown with possible orientations of bound xylan. The small circles indicate the binding sites in CBM13. Dotted lines with arrowheads show the three-fold axis of symmetry. The domains are labelled $\alpha$, $\beta$ and $\gamma$ according to the convention used throughout the present paper.

individual binding sites are frequently low ($\approx 10^3$–$10^5$ $M^{-1}$). However, extremely tight binding ($K_a$ values $\approx 10^8$–$10^9$ $M^{-1}$) can result from multiple binding sites interacting with ligands [35,37]. RTB is an example of this. The 1$\alpha$ and 2$\gamma$ binding sites separately bind sugars with affinities on the order of $10^3$ and $10^4$ $M^{-1}$ respectively; however, RTB interacts with larger, branched ligands (e.g. cell surface glycans) with affinities three orders of magnitude higher [7]. CBM13 appeared to have a similar potential with three putative binding sites. Substitution of the aspartic acid residues at the base of the 1$\alpha$ and 2$\gamma$ binding sites in RTB resulted in loss of binding to those sites [29,30]. The homologous aspartic acid residues D16, D59 and D99 (numbering conforms to that in Figure 1), in the $\alpha$, $\beta$, and $\gamma$ domains respectively, were mutated in CBM13. The single mutations D58A ($Asp^{58} \rightarrow Ala$) and D99A decreased the binding affinity approx. 3-fold (Figure 6). The double mutation D16A/D99A decreased the affinity 12-fold (Figure 6). This supported three binding sites in CBM13.

The melting temperatures were 56.5, 54.5, 57.5, and 56.5 °C for wild-type, D58A, D99A and D16A/D99A mutants respectively. The observation that the stabilities of the mutants were relatively unchanged compared with the wild-type CBM13 suggested that the changes in the xylan binding affinity of the mutants were due to the importance of the substituted amino acid to binding rather than gross structural perturbations.

The structural similarity of CBM13 and RTB suggests the presence of three binding sites in CBM13, evenly spaced radially around the axis of symmetry present in the $\beta$-trefoil fold (Figure 7). Alanine substitution of the conserved residues D16, D59 and D99 in the $\alpha$, $\beta$ and $\gamma$ domains of CBM13 decreased the affinity for xylan, indicating that these sites do participate in xylan binding. Furthermore, at least one of these sites (in the $\alpha$ domain) is able to bind mono- and di-saccharides that are too small to involve the co-operation of second and third binding sites. These results are most consistent with the presence of functional binding sites in each of the three domains of CBM13. If the three binding sites are independent and equal with respect to xylan binding, additive affinities would be expected: abolition of binding at one site would only result in a two-thirds decrease in binding affinity and abolition of binding at two sites would result in a 3-fold decrease in binding affinity. The 3–4-fold and 12-fold decreases in xylan binding affinity for single and double mutations respectively in CBM13 indicate a co-operative component between the sites to increase the affinity for xylan. Although it is possible that the binding sites in the three domains combine to make a single xylan-binding site, this is not consistent with the postulated $\beta$-trefoil structure, which could only allow at most two of the binding sites to interact simultaneously with a single xylan molecule (Figure 7). In this case, mutation of a single binding site would result in a reduction from three possible binding modes to only one. If the probabilities of binding via one mode of the three were additive, such a mutation should result in a 3-fold decrease in binding affinity. This is consistent with the observed effects of the single mutations in CBM13 and implies an additive component to the binding affinity. Furthermore, this also suggests that the 3–4-fold decrease in affinity from a single to a double mutant (i.e. D99A compared with D16A/D99A) gives an approximation of the degree of co-operativity between two binding sites.

Among the carbohydrate binding modules found in glycosyl hydrolases, CBM13 binds its target polysaccharide, xylan, by a previously undescribed mechanism. Other CBMs have single binding sites on each module. Regardless of whether these modules bind insoluble or soluble polysaccharides, the binding sites are generally extended, spanning five or more sugar residues [42–46]. In particular, the xylan-binding module CBM2b-1 has a single elongated and twisted binding site that accommodates the helical turn of five xylopyranoside residues in xylan [41]. In contrast, CBM13 has three small binding sites, each of which appears to be a general, low-affinity sugar-binding site. Its

relatively increased affinity for xylan comes by virtue of a macroscopic association constant that is a function of additive and co-operative effects of the interaction of three binding sites. This is much like the avidity effects seen in the association of RTB with cell-surface glycans [7], but with much less profound effects.

### Biological implications

The galactose-binding $\beta$-trefoil lectins may have evolved from a small galactose-binding polypeptide, similar to the 1$\alpha$ domain of RTB, that forms a trimer with similar architecture to a $\beta$-trefoil fold [26,47]. Subsequent gene triplication and fusion would lead to a contiguous polypeptide with a $\beta$-trefoil fold. Further gene duplication and fusion would lead to the binding-site configuration of RTB and related eukaryotic lectins. CBM13, and all of the prokaryotic members of family 13, would be derived from the first triplication step resulting in the three repeated domains with very similar amino acid sequences. The similarity between repeated fibronectin type III (Fn3) repeats in glycosyl hydrolases has been used to estimate the length of time since the organism acquired the module [48]. Fewer repeats and dissimilarity between the repeats indicate that the modules have been in place for a long time. The domains of the RTB-like lectins show a relatively high degree of dissimilarity, and the modules have been duplicated, suggesting that they have been in plants for a long time. In contrast, prokaryotic enzymes contain single family 13 CBMs that have very similar domains. This evidence implies an evolutionary history analogous to that of the Fn3 repeats [48]. Plants were likely the first to acquire the family 13 modules. A prokaryotic ancestor then received DNA encoding a family 13 CBM, probably from decaying plant matter. Subsequent horizontal transfer between microbes has given the distribution of the family 13 CBMs that we now see.

Proposed functions of CBMs include targeting of the enzymes to specific regions of the substrate [49], assisting in enzyme processivity [50], and non-hydrolytic disruption of the substrate [51]. Enhanced association of an enzyme with its substrate is commonly observed. CBMs specific for cellulose have association constants that range from $10^5$ $M^{-1}$ to $10^6$ $M^{-1}$ [3]. With affinities of this magnitude a large proportion of the enzyme is bound, even in the absence of abundant amounts of substrate. CBM13 and CBM2b-1 have much lower affinities for their substrate (i.e. $10^3$ $M^{-1}$). Thus a large excess of xylan relative to the parent enzymes, *S. lividans* Xyn10A and *C. fimi* Xyn11A, is required for efficient binding of the enzymes. This implies that, in the biological system involving these enzymes: (1) large amounts of substrate are present; (2) the amount of enzyme produced is low; and/or (3) the local concentration of the micro-organism is low. Thus, if one assumes an evolutionary selection for efficient use of enzymes, the low affinity of CBM13, and low affinity of other binding modules, has important implications for the availability of enzyme and substrate in the micro-organism's natural environment.

We thank R. A. J. Warren for his valuable suggestions in preparing this manuscript, Linda Sandercock for assistance in preparing *Strep. lividans* genomic DNA, and Bradley McLean for helpful insights and discussion. This research was supported by the Protein Engineering Network of Centres of Excellence, the Natural Sciences and Research Council of Canada, and CBD Technologies Inc.

## REFERENCES

1 Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. J. and Warren, R. A. (1991) Domains in microbial beta-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol. Rev. **55**, 303–315

2 Tomme, P., Warren, R. A., Miller, Jr., R. C., Kilburn, D. G. and Gilkes, N. R. (1995) in Enzymatic Degradation of Insoluble Polysaccharides (Saddler, J. N. and Penner, M., eds.), Cellulose-binding domains: classification and properties, pp. 142–163, American Chemical Society, Washington

3 Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A. L., Sturch, K., Gilkes, N. R., Haynes, C. A., Warren, R. A. J. and Kilburn, D. G. (1998) Characterization and affinity applications of cellulose-binding domains. J. Chromatogr. B **715**, 283–296

4 Coutinho, P. M. and Henrissat, B. (1999) Carbohydrate-active enzymes: an integrated database approach. 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

5 Boraston, A. B., McLean, B. W., Kormos, J. M., Alam, M., Gilkes, N. R., Haynes, C. A., Tomme, P., Kilburn, D. G. and Warren, R. A. J. (1999) Carbohydrate-binding modules: diversity of structure and function. In Recent Advances in Carbohydrate Bioengineering (Gilbert, H. J., Davies, G. J., Henrissat, B. and Svensson, B., eds.), pp. 202–211, Royal Society of Chemistry, Cambridge

6 Hirabayashi, J., Dutta, S. K. and Kasai, K. (1998) Novel galactose-binding proteins in Annelida. Characterization of 29-kDa tandem repeat-type lectins from the earthworm *Lumbricus terrestris*. J. Biol. Chem **273**, 14450–14460

7 Baenziger, J. U. and Fiete, D. (1979) Structural determinants of *Ricinus communis* agglutinin and toxin specificity for oligosaccharides. J. Biol. Chem. **254**, 9795–9799

8 Vincent, P., Shareck, F., Dupont, C., Morosoli, R. and Kluepfel, D. (1997) New $\alpha$-L-arabinofuranosidase produced by *Streptomyces lividans*: cloning and DNA sequence of the *abfB* gene and characterization of the enzyme. Biochem. J. **322**, 845–852

9 Dupont, C., Roberge, M., Shareck, F., Morosoli, R. and Kluepfel, D. (1998) Substrate-binding domains of glycanases from *Streptomyces lividans*: characterization of a new family of xylan-binding domains. Biochem. J. **330**, 41–45

10 Gilkes, N. R., Jervis, E., Henrissat, B., Tekant, B., Miller, R. C. J., Warren, R. A. and Kilburn, D. G. (1992) The adsorption of a bacterial cellulase and its two isolated domains to crystalline cellulose. J. Biol. Chem. **267**, 6743–6749

11 Coutinho, J. B., Gilkes, N. R., Warren, R. A., Kilburn, D. G. and Miller, R. C. J. (1992) The binding of *Cellulomonas fimi* endoglucanase C (CenC) to cellulose and Sephadex is mediated by the N-terminal repeats. Mol. Microbiol. **6**, 1243–1252

12 Blake, J. D. and Richards, G. N. (1971) An examination of some methods for fractionation of plant hemicelluloses. Carbohydr. Res. **17**, 253–268

13 Selvendran, R. R. and O'Neill, M. A. (1987) Isolation and analysis of cell walls from plant material. Methods Biochem. Anal. **32**, 25–153

14 Betzler, M., Dyson, P. and Schrempf, H. (1987) Relationship of an unstable *argG* gene to a 5.7-kilobase amplifiable DNA sequence in *Streptomyces lividans* 66. J. Bacteriol. **169**, 4804–4810

15 Graham, R. W., Greenwood, J. M., Warren, R. A., Kilburn, D. G. and Trimbur, D. E. (1995) The pTugA and pTugAS vectors for high-level expression of cloned genes in *Escherichia coli*. Gene **158**, 51–54

16 Tomme, P., Creagh, A. L., Kilburn, D. G. and Haynes, C. A. (1996) Interaction of polysaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC. 1. Binding specificity and calorimetric analysis. Biochemistry **35**, 13885–13894

17 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

18 Mach, H., Middaugh, C. R. and Lewis, R. V. (1992) Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins. Anal. Biochem. **200**, 74–80

19 Eftink, M. R. and Ghiron, C. A. (1981) Fluorescence quenching studies with proteins. Anal. Biochem **114**, 199–227

20 Eftink, M. R. (1997) Fluorescence methods for studying equilibrium macromolecule–ligand interactions. Methods Enzymol. **278**, 221–257

21 Takeo, K. (1985) Electrophoresis **5**, 187–195

22 Johnson, P. E., Tomme, P., Joshi, M. D. and McIntosh, L. P. (1996) Interaction of soluble cellooligosaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC 2. NMR and ultraviolet absorption spectroscopy. Biochemistry **35**, 13895–13906

23 Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P. and Gilbert, H. J. (1998) *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. Biochem. J. **331**, 775–781

24 Chaplin, M. F. (1986) Monosaccharides. In Carbohydrate Analysis: a Practical Approach (Chaplin, M. F. and Kennedy, J. F., eds.), pp. 1–36, IRL Press, Oxford

25 Lever, M. (1973) Colorimetric and fluorometric carbohydrate determination with *p*-hydroxybenzoic acid hydrazide. Biochem. Med. **7**, 274–281

26 Rutenber, E. and Robertus, J. D. (1991) Structure of ricin B-chain at 2.5 A resolution. Proteins **10**, 260–269

27 Houston, L. L. and Dooley, T. P. (1982) Binding of two molecules of 4-methylumbelliferyl galactose or 4-methylumbelliferyl *N*-acetylgalactosamine to the B chains of ricin and *Ricinus communis* agglutinin and to purified ricin B chain. J. Biol. Chem. **257**, 4147–4151

28 Zentz, C., Frenoy, J. P. and Bourrillon, R. (1978) Binding of galactose and lactose to ricin: equilibrium studies. Biochim. Biophys. Acta **536**, 18–26

29 Frankel, A., Tagge, E., Chandler, J., Burbage, C. and Willingham, M. (1996) Double-site ricin B chain mutants retain galactose binding. Protein Eng. **9**, 371–379

30 Frankel, A. E., Burbage, C., Fu, T., Tagge, E., Chandler, J. and Willingham, M. C. (1996) Ricin toxin contains at least three galactose-binding sites located in B chain subdomains 1$\alpha$, 1$\beta$, and 2$\gamma$. Biochemistry **35**, 14749–14756

31 Quiocho, F. A. (1988) Molecular features and basic understanding of protein–carbohydrate interactions: the arabinose-binding protein-sugar complex. Curr. Top. Microbiol. Immunol. **139**, 135–148

32 Sharma, S., Bharadwaj, S., Surolia, A. and Podder, S. K. (1998) Evaluation of the stoichiometry and energetics of carbohydrate binding to *Ricinus communis* agglutinin: a calorimetric study. Biochem. J. **333**, 539–542

33 Yamasaki, N., Absar, N. and Funatsu, G. (1989) The interaction of *Abrus precatorius* agglutinin with saccharides as analyzed by fluorescence spectroscopy. J. Biochem. (Tokyo) **105**, 708–711

34 Das, M. K., Khan, M. I. and Surolia, A. (1981) Fluorimetric studies of the binding of *Momordica charantia* (bitter gourd) lectin with ligands. Biochem. J. **195**, 341–343

35 Lis, H. and Sharon, N. (1998) Lectins: carbohydrate-specific proteins that mediate cellular recognition. Chem. Rev. **98**, 637–674

36 Quiocho, F. A. (1986) Carbohydrate-binding proteins: tertiary structures and protein–sugar interactions. Annu. Rev. Biochem. **55**, 287–315

37 Weis, W. I. and Drickamer, K. (1996) Structural basis of lectin–carbohydrate recognition. Annu. Rev. Biochem. **65**, 441–473

38 Heredia, A., Jimenez, A. and Guillen, R. (1995) Composition of plant cell walls. Z. Lebensm. Unters. Forsch. **200**, 24–31

39 Bouveng, H. O., Garegg, P. G. and Lindberg, B. (1958) Position of the O-acetyl groups in Birch xylan. Chem. Ind. **52**, 1727

40 Glaudemans, C. P. J. and Timell, T. E. (1958) The polysaccharides of White Birch (*Betula papyrifera*). III. Determination of composition and identification of 2-*O*-(4-*O*-methyl-D-glucopyranosyluronic acid)-D-xylopyranose. J. Am. Chem. Soc. **80**, 941–943

41 Simpson, P. J., Bolam, D. N., Cooper, A., Ciruela, A., Hazlewood, G. P., Gilbert, H. J. and Williamson, M. P. (1999) A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity. Struct. Fold. Des. **7**, 853–864

42 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) Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. Biochemistry **34**, 6993–7009

43 Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y. and Steitz, T. A. (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. EMBO J. **15**, 5739–5751

44 Brun, E., Moriaud, F., Gans, P., Blackledge, M. J., Barras, F. and Marion, D. (1997) Solution structure of the cellulose-binding domain of the endoglucanase Z secreted by *Erwinia chrysanthemi*. Biochemistry **36**, 16074–16086

45 Johnson, P. E., Joshi, M. D., Tomme, P., Kilburn, D. G. and McIntosh, L. P. (1996) Structure of the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC determined by nuclear magnetic resonance spectroscopy. Biochemistry **35**, 14381–14394

46 Linder, M., Mattinen, M. L., Kontteli, M., Lindberg, G., Stahlberg, J., Drakenberg, T., Reinikainen, T., Pettersson, G. and Annila, A. (1995) Identification of functionally important amino acids in the cellulose-binding domain of *Trichoderma reesei* cellobiohydrolase I. Protein Sci. **4**, 1056–1064

47 Rutenber, E., Ready, M. and Robertus, J. D. (1987) Structure and evolution of ricin B chain. Nature (London) **326**, 624–626

48 Little, E., Bork, P. and Doolittle, R. F. (1994) Tracing the spread of fibronectin type III domains in bacterial glycohydrolases. J. Mol. Evol. **39**, 631–643

49 Coutinho, J. B., Gilkes, N. R., Kilburn, D. G., Warren, R. A. and Miller, Jr., R.C (1993) The nature of the cellulose-binding domain affects the activities of a bacterial endoglucanase on different forms of cellulose. FEMS Microbiol. Lett. **113**, 211–218

50 Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A. and Wilson, D. B. (1998) Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. J. Bacteriol. **180**, 1709–1714

51 Din, N., Gilkes, N. R., Tekant, B., Miller, R. C., Warren, R. A. and Kilburn, D. G. (1991) Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. Bio/Technology **9**, 1096–1099

52 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**, 4673–4680


Received 16 February 2000/22 May 2000; accepted 20 June 2000