Identification of novel β -mannan- and β -glucan-binding modules: evidence for a superfamily of carbohydrate-binding modules

Anwar SUNNA*, Moreland D. GIBBS* and Peter L. BERGQUIST*†¹

*Department of Biological Sciences, Macquarie University, Sydney, N.S.W. 2109, Australia, and †Department of Molecular Medicine, University of Auckland Medical School, Private Bag 92019, Auckland, New Zealand

Many glycoside hydrolases, which degrade long-chain carbohydrate polymers, possess distinct catalytic modules and non-catalytic carbohydrate-binding modules (CBMs). On the basis of conserved protein secondary structure, we describe here the identification and experimental characterization of novel type of mannanase-associated mannan-binding module and also characterization of two CBM family 4 laminarinase-associated β -

glucan-binding modules. These modules are predicted to belong to a superfamily of CBMs which include families 4, 16, 17, 22 and a proposed new family, family 27.

Key words: affinity electrophoresis, modular glycoside hydrolases, non-catalytic modules.

INTRODUCTION

Polysaccharide-hydrolysing enzymes are observed frequently to be composed of distinct catalytic and non-catalytic modules that are often linked via flexible linker sequences. Separation of the individual modules has generally demonstrated that they can function independently of each other [1]. Many of the multidomain polysaccharide-hydrolysing enzymes described so far have non-catalytic domains of unknown function. The most common function identified for non-catalytic domains has been a polysaccharide-binding function. This binding function may enhance activity by increasing local enzyme concentration on the substrate surface, or by disrupting non-covalent interactions, thereby increasing substrate accessibility [2]. Cellulose-binding domains (CBDs) were the first polysaccharide-binding domains (PBDs) to be identified and studied extensively. CBDs were classified into ten families on the basis of sequence similarity [2]. However, since this classification, many PBDs with different ligand-binding specificities have been characterized. Starch-binding domains associated with amylolytic enzymes belonging to glycoside hydrolase (GH) families 13, 14 and 15 have been separated recently into bacterial and fungal subgroups on the basis of amino acid sequence similarities [3]. Two mannan-binding domains associated with a bacterial β -mannanase and a fungal β -mannosidase were reported for the first time recently [4,5]. Chitin-binding domains of several multidomain microbial chitinases have been demonstrated experimentally to bind chitin, whereas others have been deduced directly from sequence similarity [6]. Four classes of unrelated xylan-binding domains have also been reported [7–12].

All PBDs with solved structures can be divided into three groups. One group is on the basis of a three-stranded antiparallel β -sheet structure, and includes family I and family V CBDs [13]. The second group is based on β -sandwich jelly roll topology, and includes families II, III and IV CBDs [13], and starch-binding domains [3]. The third group, which includes family X CBDs, consists of two antiparallel β -sheets, one with two strands and one with three, with a short α -helix across one face of the three-stranded sheet [14].

Recently, Boraston et al. [15] proposed the use of the term ‘carbohydrate-binding modules’ (CBMs) to describe the expanded specificity of PBDs. Furthermore, these authors also describe an updated classification to include 20 families of CBMs. This classification has been expanded further to 26 families, as described in the latest update of the CAZy database (<http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html>) [16]. This new terminology will be used throughout this paper. Here, we identify and present the characterization of two new classes of CBMs; namely, novel examples of mannan-binding and β -glucan-binding modules. Furthermore, we present evidence for the definition of a superfamily of structurally related CBMs from families 4, 16, 17, 22 and a new proposed CBM family, family 27.

EXPERIMENTAL

Bacterial strains and genomic DNA

Escherichia coli strain DH5 α [17] was used as the bacterial host in all DNA cloning and expression studies. Genomic DNA from *Thermotoga neapolitana* and *Caldicellulosiruptor* strain Rt8B.4 was prepared as described previously [18]. The media and other reagents used have been described by Croft et al. [19].

Construction of pPROEX HTc recombinant plasmids

Specific primers (Table 1) were synthesized in order to PCR-amplify DNA coding for the different modules in the present study, and these were designed to include either *Bam*HI or *Eco*RI restriction enzyme sites to allow directional in-frame ligation of PCR fragments into the expression plasmid pPROEX HTc (Life Technologies, Melbourne, Australia). The expression vector pPROEX HTc encodes an N-terminal tag of six histidine residues, followed by a protease cleavage site. Both strands of recombinant plasmids were sequenced to confirm that there were no PCR-derived base changes in the DNA, except for those introduced by the addition of restriction sites to the PCR primers. The

Abbreviations used: CBD, cellulose-binding domain; CBM, carbohydrate-binding module; GH, glycoside hydrolase; GST, glutathione S-transferase; HE, hydroxyethyl; CM, carboxymethyl; LBG, locust bean gum; PBD, polysaccharide-binding domain; XBM, xylan-binding module.

¹ To whom correspondence should be addressed (e-mail peter.bergquist@mq.edu.au).

Table 1 Oligonucleotides used in this study

Engineered restriction sites are underlined.

Primer	Sequence
TNEALAMF	5'-TTTCCTGTTTGGATCCTGGCTCAGAATATTTTAC-3'
TNEALAMR	5'-CTGACTTCCAGAAATTCACACATCTTCCATCACTAC-3'
TNEALAMF3	5'-ACAAACAGGGATCCAAAGTGACCTATGAACAGA-3'
TNEALAMR3	5'-TTTCCTGGGAGAATTCATTGAGGGCTACCGA-3'
RTMANF	5'-ATTCCTGAGGATCCCTGGTGTGTTGAGTCG-3'
RTMANF2	5'-AATGTAAGAAGGATCCCTGAGGATGCTTCAAATC-3'
RTMANR1	5'-TCTTTTGACAGAATTCAGCATCCTCAGGTCTTA-3'
RTMANR2	5'-ACTTTTGGAGGAATTCACCTTCTGCCACAAGTT-3'

recombinant peptides produced by each plasmid are shown in Figure 2.

Production and purification of fusion proteins

Recombinant fusion proteins were produced as described previously by Sunna et al. [12], and purification was performed with Ni²⁺-nitrilotriacetate Magnetic Agarose Beads (Qiagen, Victoria, Australia) following the manufacturer's instructions. Purity of the final sample was determined by SDS/PAGE using 12% pre-cast gels (Gradipore, Sydney, Australia) as described by the method of Laemmli [20]. Proteins were stained with Coomassie Brilliant Blue R-250 (Sigma Chemical Co., St. Louis, MO, U.S.A.). Pure samples were finally desalted using Microcon ultrafiltration spin columns (cut-off 3 kDa; Millipore, Sydney, Australia), and these were then stored in 50 mM sodium phosphate buffer, pH 6.0, at 4 °C. Protein concentrations were determined using the Micro BCA protein quantification kit (Pierce, Rockford, IL, U.S.A.).

Affinity electrophoresis

Qualitative and quantitative binding of CBMs to soluble polysaccharides were evaluated by affinity electrophoresis. Non-denaturing continuous PAGE was performed using gels containing 7.5% (w/v) acrylamide in 1.5 M Tris/HCl buffer, pH 8.2. For soluble ligand-containing gels, birchwood xylan (Sigma), hydroxyethyl (HE)-cellulose (Merck, Darmstadt, Germany), locust bean gum (LBG)-galactomannan (Sigma), lichenan (Sigma), barley β -glucan (Sigma), carboxymethyl (CM)-pachyman (Megazyme International, Bray, Ireland), *Saccharomyces cerevisiae* mannan (Sigma) or laminarin (Sigma) were added to the gel mixtures at a final concentration of 0.1% (w/v) before polymerization. Native gels without ligand were run simultaneously under the same conditions. Proteins were electrophoresed at room temperature and 100 V. Chicken egg albumin, BSA and urease (all purchased from Sigma) were used as non-binding protein controls. The dissociation constant (K_d) values for polysaccharides were determined by performing affinity electrophoresis with various concentrations of ligand [21].

RESULTS AND DISCUSSION

Identification of conserved motif in some CBM families

Recently, we showed that an example of a class of xylanase-associated module, previously ascribed a thermostabilizing function, was also a xylan-binding module (XBM). We proposed that xylan binding was likely to be the primary function of this class of module, and that the thermostabilizing function was merely a

consequence of close association with the adjacent catalytic module [12]. This XBM and modules of similar amino acid sequence have since been classified as a member of the CBM family 22 (<http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html>). Sequence alignments of members of this class of XBM allowed us to identify two conserved motifs near the N- and C-terminal regions of this module. The N-terminal motif, in PROSITE pattern notation [22], consisted of the amino acids phenylalanine and glutamate (F-E), whereas the C-terminal motif (using the single-letter amino acid notation) consisted of [FY]-Y-[IV]-D, where square brackets indicate acceptable amino acids for a given position. A survey of the translated GenBank® database (release no. 115.0) revealed that representatives from CBM families 4, 16, 17 and 22 also possessed these motifs, separated by approx. 120–140 amino acids (Figure 1). A potentially unclassified new family of CBMs was also identified on the basis of possession of these motifs. These potential CBMs were associated with both GH family 5 and 26 mannanases (Figure 1).

We wanted to determine whether the identified putative mannanase-associated modules were also CBMs, and, if so, whether it would be possible to predict their polysaccharide-binding specificity on the basis of the catalytic function of their associated catalytic module. Accordingly, specific primers were designed to amplify the putative mannanase-associated modules of the extremely thermophilic bacterium *Caldicellulosiruptor* strain Rt8B.4 [23]. The putative N- and C-terminal family 4 CBMs of a modular laminarinase from *T. neapolitana* were also examined. The modular structure of each protein and the relative binding position of all primers used are presented in Figure 2. The sequences of the primers are given in Table 1. The selected modules were expressed as fusion proteins containing an N-terminal tag of six histidine residues for easy purification. Recombinant fusion proteins were purified to at least 90% electrophoretic homogeneity (results not shown), and then were tested for polysaccharide binding using affinity electrophoresis.

Affinity binding experiments

Affinity electrophoresis was used to determine binding specificity of the different purified fusion proteins against soluble polysaccharides (Figure 3). The migration of all the proteins tested was not affected in control gels that did not contain polysaccharides. In the case of *Caldicellulosiruptor* strain Rt8B.4 mannanase, ManA, the two N-terminus-associated modules in tandem (ManAm12) and the N-terminal module alone (ManAm1) were retarded most severely by LBG-galactomannan, and also to a lesser extent by barley β -glucan, CM-pachyman and HE-cellulose. However, the second module alone (ManAm2) showed no affinity for any of the polysaccharides tested. Furthermore, none of the ManA modules were retarded in the presence of *S. cerevisiae* mannan. LBG-galactomannan is a β -1,4-linked D-mannose polysaccharide with α -1,6-linked galactose (mannose:galactose ratio 4:1), whereas *S. cerevisiae* mannan is a polysaccharide of α -1,6-, α -1,2- and α -1,3-mannosidic linkages. This indicates that ManAm1 and ManAm12 bind specifically to the β -1,4-linked D-mannose rather than to the α -1,6-linked galactose of LBG-galactomannan.

Interestingly, the ManAm2 module alone did not bind to any of the ligands tested. Similar results have been reported for the XBMs of *T. maritima* XynA [24] and *Clostridium thermocellum* Xyn10B (formerly XynY) [25]. The properties of the two non-catalytic modules flanking Xyn10B of *Clo. thermocellum* were investigated recently [25]. Only the C-terminal module X6b was found to have a carbohydrate-binding property, whereas the

© 2001 Biochemical Society

Figure 2 Modular structures of *T. neapolitana* LamA and *Caldicellulosiruptor* strain Rt8B.4 ManA

Relative binding positions of oligonucleotide primers (arrows) used to PCR-amplify, clone and overexpress the putative CBMs encoded by each gene. The nucleotide sequence of each primer is shown in Table 1. Diagonal-line-shaded box, leader sequence; black box, putative CBMs; M, modules; hatched-shaded box, not defined; GH modules (GH *n*) are as indicated on the Figure.

N-terminal module X6a did not bind to any of the carbohydrates tested. In *T. maritima* XynA, glutathione S-transferase (GST) hybrid proteins containing both N-terminal-associated modules in tandem (GST–A1A2) or the second module alone (GST–A2) did interact with various soluble xylans and mixed β -1,3-/ β -1,4-glucans, whereas the first module alone (GST–A1) did not bind to any of the ligands tested [24]. Differences in polysaccharide-binding specificity of similar CBMs have been attributed to changes in the orientation of a single residue on their binding surfaces [7]. Thus, in the case of X6a, GST–A1 and ManAm2, amino acid residue differences in the ligand-binding site of these modules may indicate a binding specificity for an as-yet-undefined or untested polysaccharide.

The N-terminal module of *T. neapolitana* LamA (LamAm1) was retarded to some extent in the presence of lichenan, laminarin and barley β -glucan, whereas the migration of the C-terminal module (LamAm3) was severely retarded by barley β -glucan, lichenan and CM-pachyman. The retardation of LamAm3 was less significant in the presence of HE-cellulose, whereas laminarin had no effect on the mobility of the module. In the case of LamAm1, the degree of binding to laminarin was not improved in gels containing higher concentrations (up to 1.0%) of the polysaccharide (results not shown), whereas LamAm3 did not bind to this ligand at the higher concentration. The low affinity of LamAm1 to the ligands tested may suggest a binding specificity for an as-yet-untested ligand.

LamAm3 bound strongly to mixed β -1,3-/ β -1,4-glucans (barley β -glucan and lichenan) and β -1,3-glucans (CM-pachyman). Barley β -glucan is mainly a non-substituted unbranched β -1,3-/ β -1,4-linked D-glucose polysaccharide (β -1,3: β -1,4 ratio

$\approx 1:3$), whereas lichenan has a higher frequency of β -1,3-linkages (β -1,3: β -1,4 ratio $\approx 2:1$). Interestingly, LamAm3 did bind to CM-pachyman, but not to laminarin. CM-pachyman is a carboxymethylated, highly purified polymer of β -1,3-linked D-glucose, whereas laminarin has a backbone of β -1,3-linked D-glucose with some β -1,6-linkages and 50% terminally located mannitol. Since LamAm3 does not bind to laminarin, it is unclear at this moment whether the binding to CM-pachyman is specific or due to the CM groups substitution in the ligand. The laminarinase-associated CBMs studied here are related to the family 4 CBMs. The few CBMs of family 4 that have been studied experimentally have been shown to bind xylan and amorphous cellulose. These are the first examples of β -glucan-specific binding modules characterized so far.

In both cases examined here, the predicted carbohydrate-binding function of putative binding modules has been shown to correlate directly with the substrate specificity of covalently associated GH modules.

Quantitative binding of CBMs to soluble polysaccharides was evaluated by affinity electrophoresis in the presence of various amounts of polysaccharide (Figures 4 and 5). The reciprocal relative migration distance of the CBMs ($1/r$) was plotted against the polysaccharide concentration ($\mu\text{g/ml}$). The apparent dissociation constant, K_a , for the binding of the CBMs to the ligands was determined as the inverse of the absolute value of the intercept on the abscissa of data plotted (Figures 4B and 4C, and Figure 5B). As previously observed by other investigators [5,24], the binding of the CBMs to the polysaccharides often results in strong smears of the protein band observed. Nevertheless, we obtained an apparent dissociation constant by defin-

Figure 3 Qualitative affinity electrophoresis of purified modules in the absence and presence of various soluble polysaccharides (0.1%, w/v)

Lane 1, BSA; lane 2, chicken egg albumin; lane 3, *T. neapolitana* LamAm1; lane 4, *T. neapolitana* LamAm3; lane 5, *Caldicellulosiruptor* strain Rt8B.4 ManAm12; lane 6, *Caldicellulosiruptor* strain Rt8B.4 ManAm1; lane 7, *Caldicellulosiruptor* strain Rt8B.4 ManAm2. No ligand, no soluble polysaccharide; LBG-mannan, LBG-galactomannan; SC-mannan, *Sac. cerevisiae* mannan; BW-xylan, birchwood xylan. Gels were scanned and image-processed with Adobe Photoshop 6.0.

Figure 4 Quantitative affinity electrophoresis of *Caldicellulosiruptor* strain Rt8B.4 ManA purified modules

(A) Affinity electrophoresis at various concentrations (0–62.5 $\mu\text{g/ml}$) of LBG-galactomannan. Lane 1, urease; lane 2, ManAm12; lane 3, ManAm1; lane 4, ManAm2; lane 5, BSA. (B) Reciprocal relative migration distance ($1/r$) of ManAm12 and (C) ManAm1 against various LBG-galactomannan concentrations. Gels were scanned and image-processed with Adobe Photoshop 6.0.

ing the centre of the leading edge of each protein band as the migration distance for smeared bands. Using this method, the dissociation constant for the binding of ManAm12 and ManAm1 modules to LBG-galactomannan was determined to be about 6.2 and 2.0 $\mu\text{g/ml}$ respectively (Figures 4B and 4C). These values indicate that both modules have a remarkably strong binding affinity for soluble galactomannan. The only other characterized mannan-binding module from Man26A of *Cellulomonas*

fimi has a dissociation constant for galactomannan of about 4.6 $\mu\text{g/ml}$ [5].

The dissociation constant for the binding of LamAm3 to barley β -glucan was determined to be about 65 $\mu\text{g/ml}$ (Figure 5B). The GST-A2 xylan-binding module of *T. maritima* XynA bound to barley β -glucan with a dissociation constant of approx. 55 $\mu\text{g/ml}$ [24]. However, GST-A2 also bound to different xyans, whereas LamAm3 bound specifically to β -glucans.

Figure 5 Quantitative affinity electrophoresis of *T. neapolitana* LamAm3 purified module

(A) Affinity electrophoresis at various concentrations (0–500 $\mu\text{g/ml}$) of barley β -glucan. Lane 1, BSA; lane 2, chicken egg albumin; lane 3, LamAm3. (B) Reciprocal relative migration distance ($1/r$) of LamAm3 against various barley β -glucan concentrations. Gels were scanned and image-processed with Adobe Photoshop 6.0.

Definition of a new CBM family

We propose that the characterized CBMs from *Caldicellulosiruptor* strain Rt8.B4 presented here should be defined as the first examples of a new family of CBMs, namely family 27. The three known examples of this family are listed in Table 2. All are associated with 1,4- β -mannanase catalytic modules. Both GH family 5 mannanases have single C-terminal CBMs, whereas the GH family 26 mannanase has two N-terminal CBMs in tandem. The duplicated N-terminal CBMs from *Caldicellulosiruptor* strain Rt8.B4 ManA presented in this work bind with high specificity to the soluble polysaccharide LBG-galactomannan, but not to α -mannan from *S. cerevisiae*. Family 27 represents a new class of CBMs that are likely to function primarily as β -mannan-binding modules ('MBMs').

Definition of a superfamily of CBMs

We propose the consolidation of four previously identified CBM families (4, 16, 17 and 22) and a novel family (tentatively classified here as family 27) into a superfamily of CBMs. These relationships are defined on the basis of shared amino acid similarity and the retention of two conserved motifs (Figure 1).

A broader consensus motif describing this superfamily can be defined when the conserved motifs from all above families are examined. The respective motifs from 51 representatives of families 4, 16, 17, 22 and 27 CBMs were compared, and the PROSITE motif defined as F-[DEN]-x(110,145)-[YFIVM]-[YWCFH]-[IVLFA]-[DES]-[DNYQ]-[VIFYA] (where 'x' indicates a position at which any amino acid is accepted). Each residue at each position in the motif was present in at least 5% of the 51 aligned sequences (results not shown). Residues that were present in less than 5% of examples were not included in the motif. To test the usefulness of this motif, the SwissProt and TrEMBL databases were scanned using the ScanProsite tool (<http://expasy.proteome.org.au/tools/scnpsit2.html>).

Representatives of CBM families 4, 16, 17, 22 and 27 were detected using the PROSITE motif. Owing to the high degeneracy of the motif, 1787 matches were obtained. However, the search was refined to 89 potential CBMs by considering only those matches with a described polysaccharide-hydrolysing function. Out of the 89, a total of 57 (64%) were known examples of families 4, 16, 17, 22 or 27 CBMs. The remaining 32 matches were not examined further, although they could represent potential novel classes of CBMs.

Shared sequence identity derived from pairwise alignment between representatives of CBM families 4, 16, 17, 22 and 27 is generally lower than 25% (results not shown), and is therefore too low to allow meaningful alignment of the sequences outside of the two small conserved motifs presented here. Abou Hachem et al. [11] have predicted that the CBMs of *Rhodothermus marinus* XynA were structurally related to family 4 CBMs on the basis of a 25% shared sequence identity. These CBMs, previously classified as family 16 CBM [15], have been recently reclassified as members of family 4 CBM (<http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html>). An examination of the three-dimensional structure for CenCm1 (family 4 CBM), previously referred to as CenC CBD_{N1} [26], shows that the N- and C-terminal conserved motifs identified here lie adjacent to, and interact directly with, each other (Figure 6). These two conserved motifs correspond to areas that are not involved in polysaccharide binding and, since the folding of proteins is better conserved than their sequences, the identified motifs most probably represent a common folding signature among related CBM families. Charnock et al. [25] have reported the structure of the X6b CBM from *Clo. thermocellum* Xyn10B, classified as a family 22 CBM. Xyn10B X6b shares only 23% identity with CenCm1; however, they have a very similar overall structure, with the ligand-binding site forming a shallow cleft [26]. Moreover, the two conserved motifs lie adjacent to one another in X6b in the same position as that seen for *Cel. fimi* CenCm1 (Figure 6). This retention of position of the conserved motifs with respect to overall structure supports further our grouping of these identified CBMs as members of a superfamily of CBMs.

The results presented in this paper and previous studies clearly show that modules of similar amino acid sequence of previously classified CBDs have specificities for polysaccharides other than

Table 2 Representatives of new proposed family 27 CBM

With respect to the adjacent catalytic module, 'N' represents the N-terminus; 'C' represents the C-terminus. For *T. maritima*, the catalytic function has not been proven experimentally.

Organism	Gene product name	GenBank® accession no.	No. of CBMs	Position	Function of associated catalytic module	GH family
<i>T. maritima</i>	TM1227	AE001779	1	C	Mannanase	5
<i>B. stearothermophilus</i>	ManF	AF038547	1	C	Mannanase	5
<i>Caldicellulosiruptor</i> strain Rt8.B.4	ManA	U39812	2	N-N	Mannanase	26

Figure 6 Three-dimensional ribbon structures of *Cel. fimi* CenCm1 and *Clo. thermocellum* Xyn10B X6b CBMs

Comparison showing the overall three-dimensional similarity between CenCm1 (CBM family 4) and Xyn10B X6b (CBM family 22), and the positions of their ligand-binding clefts, as described previously [25,26]. The backbone atoms of residues in the conserved N-terminal motif are shown (dark-grey spheres), along with the backbone atoms of the residues in the conserved C-terminal motif (light-grey spheres). Images were generated using SwissPDB viewer version 3.6b2 [28].

cellulose. CBM family 2 contains members that bind to cellulose (CBM2a) and xylan (CBM2b). The differing polysaccharide-binding specificity between these modules has been attributed to changes in the orientation of one tryptophan residue on the binding surface of the proteins [7]. Recently, Simpson et al. [27] reported that, by changing a single amino acid residue, a CBM2b XBM lost its affinity to bind xylan and became a CBM2a cellulose-binding module. Fernandes et al. [9] have shown that a *Clo. thermocellum* xylanase, XylV, has a binding module with significant sequence similarity to family 6 CBMs. However, this type 6 module had highest affinity for xylan, and bound only weakly to Avicel and acid-swollen cellulose. Dupont et al. [10] have shown that the *Streptomyces lividans* xylanases AxeA and XlnB possess family 2 CBMs, which specifically bind insoluble xylan and do not bind cellulose. The family 5 CBMs exhibit sequence similarity to family 12 CBMs, which are mainly chitin-binding modules [13].

The examples presented above strongly suggest that structurally related members of CBM families have evolved to have altered binding capabilities. Presumably, they evolved an altered specificity either in concert with their respective associated catalytic modules, or subsequent to recombination and association with a catalytic module with a different substrate specificity.

This work was supported by grants from Macquarie University Research Grant Fund and an Australian Research Council Small Grant.

REFERENCES

- Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. J. and Warren, R. A. J. (1991) Domains in microbial β -1,4-glycanases: sequence conservation, function and enzyme families. *Microbiol. Rev.* **55**, 303–315.
- Tomme, P., Warren, R. A. J., Miller, R. C. J., Kilburn, D. G. and Gilkes, N. R. (1995) Cellulose-binding domains: classification and properties. In *Enzymatic Degradation of Insoluble Carbohydrates* (Saddler, J. N. and Penner, M. H., eds.), pp. 142–163. American Chemical Society, Washington, DC.
- Janecek, S. and Sevcik, J. (1999) The evolution of starch-binding domain. *FEBS Lett.* **456**, 119–125.
- Kulminkaya, A. A., Eneiskaya, E. V., Isaeva-Ivanova, L., Savel'ev, A. N., Sidorenko, I. A., Shabalin, K. A., Golubev, A. M. and Nuestrrov, K. N. (1999) Enzymatic activity and β -galactomannan binding property of β -mannosidase from *Trichoderma reesei*. *Enzyme Microb. Technol.* **25**, 372–377.
- Stoll, D., Boraston, A., Ståhlbrand, H., McLean, B. W., Kilburn, D. G. and Warren, R. A. J. (2000) Mannanase Man26A from *Cellulomonas fimi* has a mannan-binding module. *FEMS Microbiol. Lett.* **183**, 265–269.
- Svitil, A. L. and Kirchman, D. L. (1998) A chitin-binding domain in a marine bacterial chitinase and other microbial chitinases: implications for the ecology and evolution of 1,4- β -glycanases. *Microbiology* **144**, 1299–1308.
- Simpson, P. J., Bolam, D. N., Cooper, A., Ciruela, A., Hazlewood, G. P., Gilbert, H. J. and Williamson, M. P. (1999) A family 11b xylan-binding domain has a similar secondary structure to a homologous family 11a cellulose-binding domain but different ligand specificity. *Structure* **7**, 853–864.
- Black, G. W., Hazlewood, G. P., Millward-Sadler, S. J., Laurie, J. I. and Gilbert, H. J. (1995) A modular xylanase containing a novel non-catalytic xylan-specific binding domain. *Biochem. J.* **307**, 191–195.
- Fernandes, A. C., Fontes, C. M. G. A., Gilbert, H. J., Hazlewood, G. P. and Fernandes, T. H. (1999) Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. *Biochem. J.* **342**, 105–110.
- 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.
- Abou Hachem, M., Nordberg Karlsson, E., Bartonek-Roxå, E., Raghothama, S., Simpson, P. J., Gilbert, H. J., Williamson, M. P. and Holst, O. (2000) Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: cloning, expression and binding studies. *Biochem. J.* **345**, 53–60.
- Sunna, A., Gibbs, M. D. and Bergquist, P. L. (2000) The thermostabilizing domain, XynA, of *Caldibacillus cellulovorans* xylanase is a xylan binding domain. *Biochem. J.* **346**, 583–586.
- Bayer, E. A., Chanzy, H., Lamed, R. and Shoham, Y. (1998) Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol.* **8**, 548–557.
- Raghothama, S., Simpson, P. J., Szabó, L., Nagy, T., Gilbert, H. J. and Williamson, M. P. (2000) Solution structure of the CBM10 cellulose binding module from *Pseudomonas* xylanase A. *Biochemistry* **39**, 978–984.
- 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.

- 16 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
- 17 Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–580
- 18 Morris, D. D., Reeves, R. A., Gibbs, M. D., Saul, D. J. and Bergquist, P. L. (1995) Correction of the β -mannanase domain of the *celC* pseudogene from *Caldicellulosiruptor saccharolyticus* and activity of the gene product on kraft pulp. *Appl. Environ. Microbiol.* **61**, 2262–2269
- 19 Croft, J. E., Love, D. R. and Bergquist, P. L. (1987) Expression of leucine genes from an extremely thermophilic bacterium in *Escherichia coli*. *Mol. Gen. Genet.* **210**, 490–497
- 20 Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**, 680–685
- 21 Mimura, Y., Nakamura, K. and Takeo, K. (1992) Analysis of the interaction between an alpha (1 $\rightarrow$ 6) dextran-specific mouse hybridoma antibody and dextran B512 by affinity electrophoresis. *J. Chromatogr.* **597**, 345–350
- 22 Hofmann, K., Bucher, P., Falquet, L. and Bairoch, A. (1999) The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27**, 215–219
- 23 Gibbs, M. D., Elinder, A. U., Reeves, R. A. and Bergquist, P. L. (1996) Sequencing, cloning and expression of a β -1,4-mannanase gene, *manA*, from the extremely thermophilic anaerobic bacterium, *Caldicellulosiruptor* Rt8B.4. *FEMS Microbiol. Lett.* **141**, 37–43
- 24 Meissner, K., Wassenberg, D. and Liebl, W. (2000) The thermostabilizing domain of the modular xylanase XynA of *Thermotoga maritima* represents a novel type of binding domain with affinity for soluble xylan and mixed-linkage β -1,3/ β -1,4-glucan. *Mol. Microbiol.* **36**, 898–912
- 25 Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M. A., Davies, G. J. and Fontes, C. M. G. A. (2000) The X6 “thermostabilizing” domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the *Clostridium thermocellum* X6b domain. *Biochemistry* **39**, 5013–5021
- 26 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**, 13895–13906
- 27 Simpson, P., Hefang, X., Bolam, D., Gilbert, H. and Williamson, M. (2000) The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. *J. Biol. Chem.* **275**, 41137–41142
- 28 Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modelling. *Electrophoresis* **18**, 2714–2723

Received 8 February 2001/13 March 2001; accepted 30 March 2001