Cell. Mol. Life Sci. 63 (2006) 2954–2967
1420-682X/06/242954-14
DOI 10.1007/s00018-006-6195-3

Cellular and Molecular Life Sciences

Review

Recent structural studies of carbohydrate-binding modules

H. Hashimoto

International Graduate School of Arts and Sciences, Yokohama City University, 1-7-29 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045 (Japan), Fax: +81 45 508 7365, e-mail: hash@tsurumi.yokohama-cu.ac.jp

Received 28 April 2006; received after revision 12 July 2006; accepted 14 September 2006
Online First 27 November 2006

Abstract. Carbohydrate-binding modules (CBMs) are found in many carbohydrate-active enzymes. CBMs bind to a range of polysaccharides, their primary function being to increase the catalytic efficiency of the carbohydrate-active enzymes against soluble and/or insoluble substrates. CBMs bind to their target ligands with high specificities and affinities. Thus, CBM systems are excellent models to study the mechanism of protein-carbohydrate interaction. To date, CBMs have been classified into 45 different families and many structural and functional studies have been reported. At present, three-dimensional structures of CBMs from 31 different families have been determined. These structures demonstrate that the fold most commonly found in CBMs is the β-sandwich. In the past few years, about 10 new structures from different families have been reported. These enable detailed classification of CBM structures. This article reviews recent structural and functional studies of CBMs and discusses the sub-classification of β-sandwich CBMs.

Keywords. Carbohydrate-binding module; structure and function, classification, β-sandwich, β-jelly roll fold, immunoglobulin fold.

Introduction

Many carbohydrate-active enzymes are modular proteins composed of catalytic modules appended to one or more non-catalytic carbohydrate-binding modules (CBMs). CBMs were previously classified as cellulose-binding domains (CBDs) based on the initial discovery of several modules that bound cellulose [1, 2]. However, additional modules in carbohydrate-active enzymes are continually being found that bind carbohydrates other than cellulose. The term CBM is proposed as a more inclusive term to describe all of the non-catalytic sugar-binding modules derived from glycoside hydrolases. CAZy (Carbohydrate Active enZyme, http://www.cazy.org/CAZy) is a database of glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and CBMs. These protein groups are further subdivided into a number of families within the groups. In the CAZy database, a CBM is currently defined as a contiguous amino acid sequence within a carbohydrate-active enzyme with a discrete fold having carbohydrate-binding activity. A few exceptions are CBMs from cellulosomal scaffolding proteins and rare examples of independent putative CBMs. CBMs are currently classified into 45 primary structure-based families in the CAZy database. CBMs bind to a range of polysaccharides that include not only crystalline cellulose but also hemicelluloses such as glucan, xylan, mannan and glucomannan. The primary function of CBMs is to increase the catalytic efficiency of the enzymes against soluble and/or insoluble substrates by bringing the catalytic module into prolonged and intimate contact with substrates [3, 4].

Three-dimensional structures of CBMs from 31 different families have been determined by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. In 2004, Boraston et al. [5] presented a comprehensive review of CBM structure and function. The article is excellent and they classified 22 CBM structures into seven structural families (β-sandwich, β-trefoil, cystein knot, unique, OB fold, hevein fold and unique containing hev-

ein-like fold) [5]. Fold family 1 is β-sandwich and is the dominant fold among CBM structures. In most cases, β-sandwich CBMs have bound metal ions which mainly contribute to structural stability. However, in family 36 CBMs, a calcium ion is involved in ligand binding [6]. Fold family 2 is β-trefoil, which is commonly associated with the ricin toxin B-chain. This fold has a pseudo three-fold axis. Family 13 CBM was initially the only group to show this fold, however subsequent determination of the crystal structure of family 42 CBMs revealed that these CBMs also adopt the β-trefoil fold as described later in this review.

A classification based on substrate binding has also been proposed [5], where CBMs are grouped into three types: 'surface-binding' CBMs (type A), 'glycan chain-binding' CBMs (type B) and 'small sugar-binding' CBMs (type C). Type A CBMs are arguably distinct as their properties differ significantly from other types of carbohydrate-binding proteins. It includes members of CBM families 1, 2a, 3, 5 and 10 that bind to insoluble, highly crystalline cellulose and/or chitin. The binding sites of these CBMs, which are composed of aromatic residues, are flat or platform-like. The planar architecture of the binding site is thought to be complementary to the flat surfaces presented by cellulose or chitin crystals [7, 8]. Type A CBMs show little or no affinity for soluble carbohydrate [9].

The binding sites of type B CBMs are described as grooves, and comprise several sub-sites able to accommodate the individual sugar units of the polymeric ligand. Although ligands are recognized by aromatic side chains like type A CBMs, the side chains of tryptophan, tyrosine and less commonly phenylalanine can form planar, twisted or sandwich platforms for ligand binding. As with type A CBMs and as mentioned above, aromatic residues play a significant role in ligand binding, and the orientation of these amino acids is a key determinant of specificity [10]. This type of CBM includes examples from families 2b, 4, 6, 15, 17, 20, 22, 27, 28, 29, 34 and 36. Finally, type C CBMs have the lectin-like property of binding optimally to mono-, di- or trisaccharides, and thus lack the extended binding site grooves as observed in type B CBMs.

Since 2004, the three-dimensional structures of CBMs from families 11 [11], 25 [12], 26 [12], 30 [13], 31[14], 33 [15], 35 [16], 42 [17] and 44 [13] have been determined. These structures were determined by X-ray crystallography with one exception, namely that of family 35 CBM which was determined by NMR spectroscopy. CBMs of family 25, 26 and 42 were solved in complexes with their ligands. In this review, these structural and functional studies are considered. Also discussed is the proposal that CBM structures of the β-sandwich type be divided into two sub-families based on their fold, namely β-jelly roll and immunoglobulin folds.

Family 11 CBMs display affinity for different linked glucans. Carvalho et al. [11] reported functional and structural studies of a family 11 CBM (CtCBM11) from *Clostridium thermocellum* Lic26A-Cel5E. Lic26A-Cel5E contains family 5 GH and family 26 GH catalytic domains that display β-1,4- and β-1,3-1,4-mixed linked endoglucanase activity, respectively. Isothermal titration calorimetry (ITC) experiments determined the binding ability of CtCBM11 to both β-1,4- and β-1,3-1,4-mixed linked glucans, with K_a values of 1.9×10^5, 4.4×10^4 and 2×10^3 M^{-1} for Glcβ1,4-Glcβ1,4-Glcβ1,3-Glc, Glcβ1,4-Glcβ1,4-Glcβ1,4Glc and Glcβ1,3-Glcβ1,4-Glcβ1,3-Glc, respectively. These results demonstrate that CBMs can display a preference for mixed linked glucan.

To determine whether these ligands are accommodated in the same or diverse sites in CtCBM11, the crystal structure of the protein was determined [11]. It revealed a β-sandwich with a classical distorted β-jelly roll fold consisting of two six-stranded anti-parallel β-sheets, which form a convex and a concave side (Fig. 1). A search for proteins structurally homologous to CtCBM11 using the DALI database [18] shows that CtCBM11 is topologically similar to the CBMs of families 29, 4, 22, 15, 27, 17 and 6. These all share a β-jelly roll fold, with the main differences restricted to solvent-exposed loops. The root mean square deviations of these structures when superimposed on CtCBM11 vary in the range of 2.2–3.5 Å.

As in other CBMs, the concave side of CtCBM11 forms a cleft. The cleft appears to be the likely carbohydrate-binding site in the protein, with residues Tyr22, Tyr53, Tyr129 and Tyr152 playing a key role in ligand recognition. CtCBM11 binds two calcium ions, in a manner similar to representatives of CBM families 4, 6, 9 and 22. Usually these ions are solvent inaccessible and are suggested to stabilize the protein fold [19]. The importance of calcium ions in maintaining the structural integ-

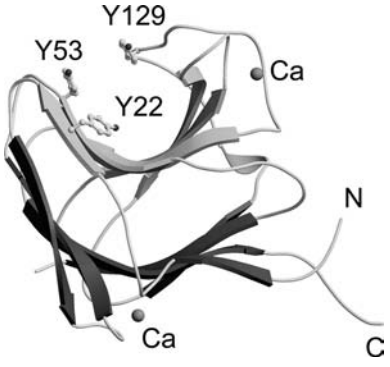

Figure 1. Overall structure of the family 11 CBM. The secondary structural elements are highlighted with the β-strands of each β-sheet colored light grey and dark grey. Tyr22, Tyr53 and Tyr129, which are involved in ligand binding, are shown as ball-and-stick models. Two calcium ions are shown as grey spheres. Figures 1, 2, 3a, 3b, 4b, 5a, 6, 7a, 7c, 7e, 7g, 8, 9a, 9c, and 9e were prepared with the programs MOLSCRIPT [61] and RASTER3D [62].

rity of family 4 CBMs was demonstrated by showing that removal of the ions reduces the melting temperature by 23 °C [20]. In CBMs from families 35 and 36, calcium was shown to play a direct role in ligand binding [6, 21]. However, because the location of the two calcium-binding sites in CtCBM11 is distant from the likely carbohydrate-binding cleft, the calcium ions are more likely to contribute to structural stability.

As described above, inspection of the surface of the cleft of CtCBM11 reveals the presence of several aromatic residues which are conserved in the other CBM11 family members. To evaluate the importance of residues Tyr22, Tyr53, Tyr129 and Tyr152 in ligand binding, Carvalho et al. [11] produced the mutant proteins Y22A, Y53A, Y129A and Y152A and compared their biochemical properties with wild-type CtCBM11. The data show that Y22A displayed no detectable affinity for the ligand used in the assay, while Y53A and Y129A exhibited significantly reduced binding for these polysaccharides. In contrast, Y52A showed similar affinity as the wild-type. This suggests that the three residues, Tyr22, Tyr53 and Tyr129, play an important role in the recognition of glucomannan, xylan, and β-1,4- and β-1,4-β-1,3-linked glucans. Qualitative binding assays were also performed to investigate the capacity of CtCBM11 to interact with the insoluble polysaccharides avicell and acid-swollen cellulose. The data showed that the mutants Y22A, Y53A and Y129A displayed dramatically reduced binding to the insoluble ligands, suggesting that cellulose binds to CtCBM11 within the extended cleft. Y152A had no apparent effect on binding to the insoluble ligands. These results suggest that the groove located in CtCBM11 is the carbohydrate-binding cleft and that the residues Tyr22, Tyr53 and Tyr129 play a major role in the interaction of CtCBM11 with its ligands.

Structural studies of CBM family 6, which is closely related to the CBM11 family, demonstrate that these proteins contain two topologically and spatially different clefts that display different ligand specificities [22]. One of the two binding clefts (cleft B) in CmCBM6 from *Cellvibrio mixtus* endoglucanase 5A was shown to interact with β-1,3-1,4-mixed linked and β-1,4-linked glucans with similar affinities and binds only weakly to β-1,3-glucan [23]. The crystal structure of CmCBM6 demonstrates that in cleft B, only a β-1,4 glucose can be accommodated at two of the sugar-binding sub-sites whereas the third sub-site can interact with either a β-1,4- or β-1,3-linked glucose molecule and the fourth sub-site displays exclusive specificity for β-1,3-linked glucose [22]. Analysis of the CtCBM11 structure, together with the mutagenesis data, indicates that unlike CBM6, the family 11 CBM contains only a single ligand-binding site, in common with the majority of CBMs. Furthermore, the preference shown by CtCBM11 may reflect a general feature of CBMs that recognize β-1,3-1,4-mixed linked glucans.

CBM30 and CBM44 bind to decorated glucans, such as xyloglucan

Although xyloglucan, which includes a backbone of β-1,4-glucan decorated primarily with xylose resides, is a key component of the plant cell wall, CBMs that bind to this polymer had not been identified. To investigate whether CBMs can display high affinity for heavily branched β-1,4-glucans such as xyloglucan and to investigate the molecular determinants that govern this binding specificity, Najmudin et al. [13] have studied the structure and function of the N- and C-terminal domains of CtCel9D-Cel44A from *C. thermocellum*. This protein is a component of the bacterial cellulosome [24]. This enzyme contains internal GH family 9 and 44 catalytic domains in addition to an N-terminal CBM30, a type I dockerin and a C-terminal module of unknown function. CBM30, which displays affinity for β-1,4-glucopolymers [25], plays a significant role in the function of GH9, a typical processive endoglucanase, whereas GH44 was assigned as displaying endo-xylanase activity [24]. The first 90 amino acids of the 250-residue C-terminal region of CtCel9D-Cel44A comprises a PKD module, while the remaining 160 amino acids display no extensive sequence identity with other modules present in GHs. The data presented by Najmudin et al. [13] show that this C-terminal region of CtCel9D-Cel44A binds to plant structural polysaccharides and is thus classified as a novel CBM, representing the founder member of CBM family 44.

Biochemical properties of CBM44 and PKD-CBM44 were evaluated. Affinity gel electrophoresis for various ligands revealed that both CBMs target β-glucose polymers, including xyloglucan. The affinities of CBM44 and PKD-CBM4 for the various ligands were similar, suggesting that the PKD domain does not contribute to carbohydrate recognition. The observation that CBM30 and CBM44 bind to xyloglucan provides the first evidence that CBMs are able to accommodate the side chains of this decorated glucan.

The structure of PKD-CBM44 reveals the presence of two separated domains, PKD and CBM44, both adopting a β-sandwich fold (Fig. 2a). Two calcium ions are found in the structure, one bound to the PKD module and the other bound to CBM44. Most β-sandwich CBMs have at least one calcium ion, which is often solvent inaccessible and is suggested to stabilize the protein fold [5]. It has been shown recently, however, that in a family 35 and a family 36 CBM, a calcium ion plays a direct role in ligand recognition [16, 26]. The calcium-binding sites in the PKD domain and CBM44 are buried inside the protein structures supporting, in both cases, the idea of a structural role for the metal. Alignment of CBM44 with CBM families 4, 6, 9, 16, 17, 22, 27, 29, 33, 35 and 37 reveals the presence of a very short conserved motif which plays a key role in the coordination of structural

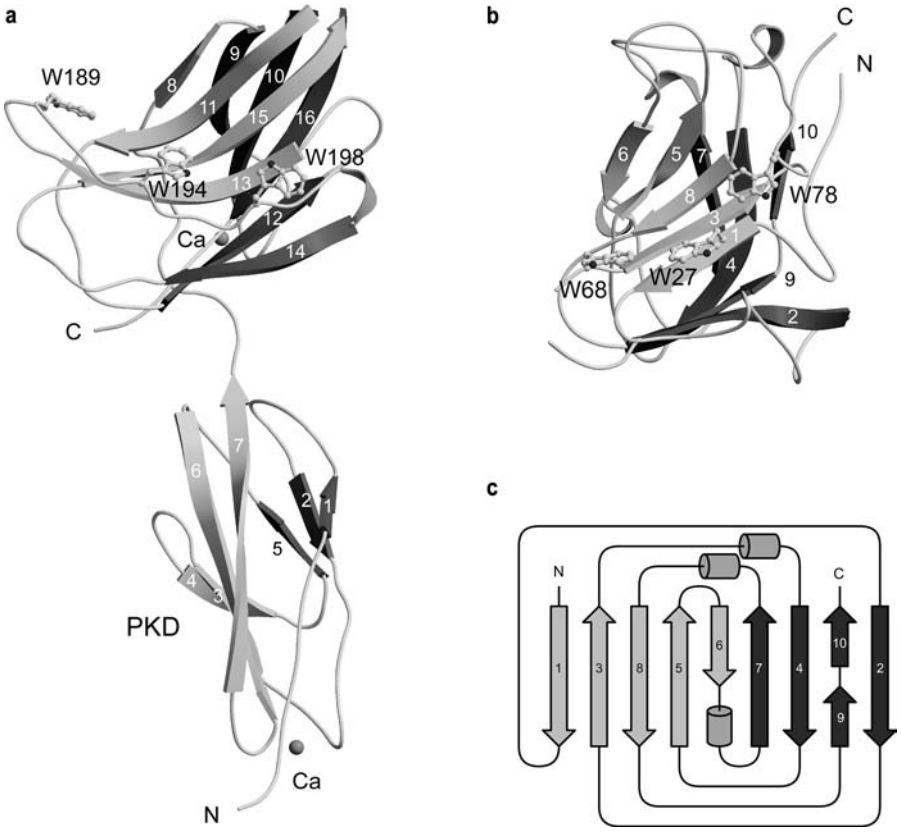

Figure 2. (*a*) Ribbon drawing of PKD-CBM44. The secondary structural elements are highlighted with the β-strands of each β-sheet colored light grey and dark grey. Each domain binds one calcium ion, shown as grey spheres. Aromatic residues, Trp189, Trp194 and Trp198, which are involved in ligand binding, are shown as ball-and-stick models. (*b*) Ribbon drawing of CBM30. The secondary structural elements are highlighted, with the β-strands of each β-sheet colored light grey and dark grey. Aromatic residues, Trp27, Trp68 and Trp78, which are involved in ligand binding, are shown as ball-and-stick models. (*c*) Topology diagram of CBM30. Colors of secondary structural elements correspond with those of the ribbon drawing. The diagram clearly shows β-jelly roll topology.

calcium. The PKD domain is built from two β-sheets with three and four β-strands packed face to face as is typical of a β-sandwich (Fig. 2a). The number and arrangement of these strands are identical to several members of the immunoglobulin and fibronectin type III superfamilies, and thus the PKD domain can be defined as displaying an immunoglobulin β-sandwich fold [27]. Biochemical data suggested that the PKD domain in CtCel9D-Cel44Al did not modulate the function of CBM44 when binding to soluble and insoluble polysaccharides [13]. However, the PKD module may be involved in specific protein-protein interactions within the cellulosome.

The CBM44 structure reveals a classic distorted β-jelly roll fold consisting of two anti-parallel β-sheets, which form a concave (β-strands 8, 11, 13 and 15) and a convex surface (β-strands 9, 10, 12, 14 and 16). The concave surface is the putative ligand-binding site, because aromatic residues are aligned on the surface (Fig. 2a). DALI [18] structural similarity searches reveal that CBM44 is most similar to CBMs of families 22, 17, 27, 11, 4, 29, 28 and 17, all of which adopt a β-jelly roll fold.

CBM30 also adopts a β-sandwich fold in which the two anti-parallel β-sheets each contain five β-strands (Fig. 2b). β-Strands 1, 3, 5, 6 and 8 form a concave surface (cleft) with aromatic residues aligned in the cleft. This cleft is the putative ligand-binding site. The CBM30 structure clearly displays a β-jelly roll fold (Fig. 2c) but with insertion of helical elements, although it was not mentioned that CBM30 adopts a β-jelly fold in the publication [13]. The DALI database [18] shows that CBM30 is topologically similar to families 11, 29, 4, 27, 15, 22, 6 and 36 CBMs. All of those CBMs adopt a β-jelly roll fold.

Inspection of the surface clefts of CBM30 and CBM44 reveals the presence of three aromatic residues on the edge of both putative binding sites (Fig. 2). To prove the importance of the solvent-exposed residues Trp27, Trp68 and Trp78 in CBM30 and Trp27, Trp68 and Trp78 in CBM44, mutant proteins in which the aromatic residues were changed to alanines were produced and their biochemical properties investigated by affinity gel electrophoresis (AGE) or ITC. In CBM30, W27A and W68A displayed no significant affinity for decorated and un-

decorated β-glucans or glucomannan. In contrast, W78A displayed reduced, but still significant affinity for xyloglucan and barley β-glucan and no detectable binding to glucomannan. These results confirmed that the concave surface does form the ligand-binding site and that the three tryptophans are involved in ligand recognition. In CBM44, W194A displayed no significant affinity for ligand while W189A and W198 showed a relatively modest decrease in affinity. Trp194 is the central aromatic residue of the binding site. Thus, Trp194 possibly makes a stronger hydrophobic interaction with the glucan than the flanking tryptophans (Trp189 and Trp198), although the dominance of this amino acid may also reflect an additional hydrogen bond between the indole nitrogen of Trp194 and the polysaccharide. Indeed, the central tryptophan residue in the ligand-binding site of the CBM2a of *Cellulomonas fimi* Cel6A also plays a more important role in cellulose binding than the flanking tryptophans [8]. The results described above clearly indicate that the hydrophobic platform provided by the three tryptophan residues in both CBM30 and CBM44 plays a significant role in ligand binding.

Family 35 CBMs undergo significant conformational change upon ligand binding

The three-dimensional structure of family 35 CBM (Man5C-CBM35), derived from *Cellvibrio japonicus* β-1,4-mannanase Man5C, was determined by NMR spectroscopy [16]. The structure is a β-jelly roll fold, similar to that observed with many CBMs, consisting of two five-stranded anti-parallel β-sheets and a helical turn (Fig. 3a). To investigate interaction with ligands, NMR titration experiments were performed with β-1,4-linked mannopentaose [16]. The largest changes in chemical shift observed were for residues at the end of the protein furthest from both N and C termini, in a region containing five adjacent loops containing three aromatic residues exposed to the solvent (Tyr60, Trp109 and Tyr111). These three aromatic residues, along with Gly110 and Lys63, showed the largest chemical shift changes upon titration, implicating them strongly in ligand binding. From fitting of the chemical shift curves of Tyr60, Trp109 and Tyr111, an approximate K_d of 100 µM was derived. As Man5C-CBM35 had been shown to bind to galactomannan and undecorated mannopentaose with similar affinity [21], a further set of titrations was performed with a typical 'decorated' mannan derivative, namely 6^3,6^4-α-D-galactosyl-mannopentaose. Again, the largest chemical shift changes were observed for residues Tyr60, Lys63, Trp109, Gly110 and Trp111, although the magnitude of these and other changes differed slightly from those previously observed for mannopentaose. A K_d of ~300 µM was determined from the fitting of chemical shift curves.

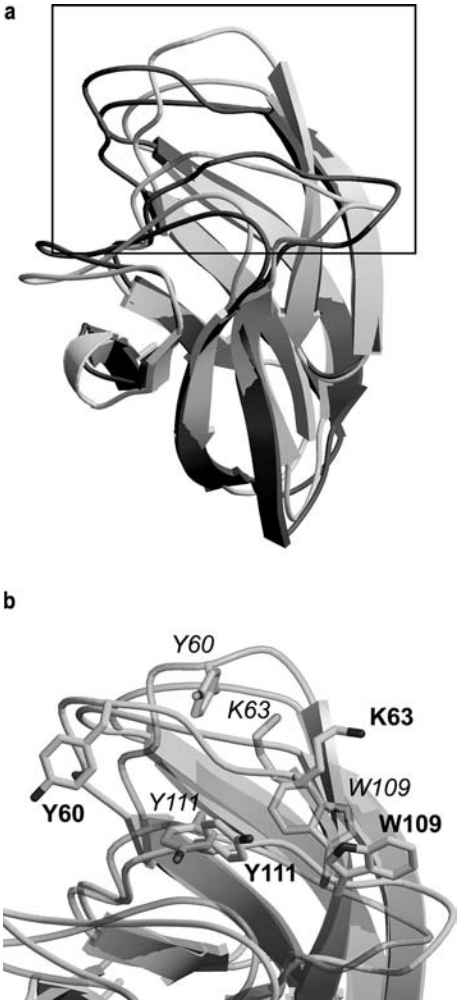

Figure 3. (*a*) Structural overlay of family 35 CBM in ligand-free (light grey) and ligand-bound (dark grey) states. (*b*) Close-up view of the region boxed in *a.* Residues showing the largest chemical shift changes are shown as stick models (Tyr60, Lys63, Trp109 and Tyr111).

The changes in chemical shift observed upon titration with mannopentaose suggest significant structural changes upon ligand binding. The protein sample was therefore saturated with 20-fold excess mannopentaose and the structure of Man5C-CBM35 bound to ligand determined (Fig. 3). It should be noted that in this work, the signals from sugars could not be assigned due to excessive overlap of bound and free sugar resonances. Upon ligand binding, Man5C-CBM35 maintains its β-sandwich fold, while towards the binding site it undergoes a minor but significant conformational change, allowing Man5C-CBM35 to accommodate the carbohydrate. The largest structural change is in the aromatic side chains implicated in mannan binding from the titration data. The most obvious change is the reorientation of the aromatic ring of Trp109, from lying flat on the protein surface in the free structure to protruding out from the surface in an appar-

ently more mobile conformation in the bound structure. This shift in Trp109 is accompanied by the movement of the adjacent loops of residues 57–60 and 80–84 inward towards the binding site, which results in an extension of the region that is able to contact the carbohydrate chain. Thus, these results demonstrate that the protein changes its conformation to bind its substrate in a shallow groove lined primarily on one side by Trp109 and Tyr111 and on the other by Tyr60 and Lys63.

A structural homology search using DALI showed strongest similarity to the CBM6 of *C. mixtus* endoglucanase 5A (CmCBM6-2) [22, 23] and other CBM6s. CBM6 is an interesting family, as it displays considerable promiscuity in ligand binding, with different modules showing affinity for amorphous cellulose, xylans and β-glucans. Even more unusual is the fact that different members of CBM6 had been found to bind substrate in two different places, known as cleft A and cleft B [22]. The ligand-binding site in Man5C-CBM35 is in approximately the same location as cleft A in CBM6, but is narrowed by approximately 2 Å. CBM6 from *C. thermocellum* (CtCBM6) utilizes residues Trp92 and Tyr34 for binding xylan [28]. The equivalent residues are not present in Man5C-CBM35, and Tyr60, Trp109 and Tyr111 appear to form a much narrower binding site. Thus, the CBM35 groove is too restrictive for interaction with xylan.

In most β-jelly roll CBMs, the ligand-binding site extends across and is perpendicular to one of the β-sheets that form the jelly roll. In contrast, the primary binding site in three of the four CBM6 modules studied to date and in Man5C-CBM35 is not located in the central groove, but is formed by the loops that connect the two β-sheets on the edge of the jelly roll. Therefore, it is proposed that CBM6 and CBM35 should be viewed as a sub-family of the large β-jelly roll CBM superfamily [16].

Family 42 CBM structure reveals a β-trefoil fold related to family 13 CBM

Aspergillus kawachii IFO4308 α-L-arabinofuranosidase (AkAbfB) comprises two domains: a catalytic domain and an arabinose-binding domain (ABD, which is classified as CBM42). α-L-arabinofuranosidase catalyzes the hydrolysis of α-1,2-, α-1,2-, α-1,3- and α-1,5-L-arabinofuranosidic bonds in L-arabinose-containing hemicelluloses such as arabinoxylan and L-arabinan [29, 30]. *A. kawachii* is an industrially important fungus in traditional Japanese liquor brewing and produces two different α-L-arabinofuranosidases, namely arabinofuranosidase A (AkAbfA) and arabinofuranosidase B (AkAbfB) [31]. AkAbfA and AkAbfB are assigned to families GH51 and GH54, respectively. Members of the GH51 family only hydrolyze small substrates, including short-chain arabino-oligosaccharides, whereas members of the GH54 family are able to hydrolyze polymeric substrates such as arabinoxylans in addition to small substrates [32]. The crystal structure and catalytic residues of GH51 *Geobacillus stearothermophilus* T-6 α-L-arabinofuranosidase A have been determined [33–35]. In contrast, the catalytic residues and three-dimensional structure of GH54 enzymes were completely unknown until Miyanaga et al. [17] reported the crystal structures of AkAbfB in complex with arabinose. These structures reveal that the ABD of AkAbfB is a novel CBM that recognizes the arabinose side chains of arabinoxylans.

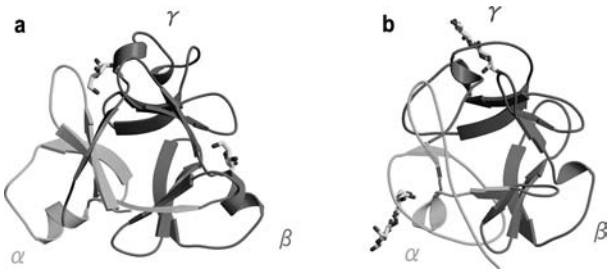

Figure 4. Structures of families 42 and 13 CBMs. The α, β and γ sub-domains are shown as light-grey, grey and dark-grey drawings, respectively. (*a*) Ribbon drawings of the ABD of AkAbfB (CBM44). Two arabinofuranoses, which are bound by the β and γ sub-domains, are shown as stick models. (*b*) Ribbon diagram of CBM13 from *Streptomyces lividans*. The xylobiose molecules are shown as stick models.

The ABD of AkAbfB possesses a β-trefoil fold (Fig. 4a). β-trefoil proteins with a G-X-X-X-Q-X-(W/Y) motif are classified into the ricin B-like family [36]. This family is divided into two types, namely galactose-binding lectins and xylan-binding domains. The xylan-binding domain belongs to family 13 CBMs (Fig. 4b). The ABD contains three segments (α, β and γ sub-domains) of about 50 amino acid residues each. Each sub-domain consists of four-stranded β-hairpin turns and one additional α-helix, which is inserted between the β-3 and β-4 strands in each sub-domain. There is structural similarity between each pair of sub-domains and they assemble with pseudo-threefold symmetry. However, amino acid sequence homology between the sub-domains is almost undetectable, with only three residues common (serine in the β-1 strand and tyrosine and histidine in the β-2 strand).

Although sharing a common structure, at least three clear differences are observed between CBM13 and the ABD. First, the ABD of AkAbfB does not contain the G-X-X-X-Q-X-(W/Y) motif of CBM13 [36]. Experimental data support the fact that the glutamine residue plays an important role in substrate recognition and that the tryptophan residue is involved in formation of the hydrophobic core. Second, in the ABD of AkAbfB, hydrophobic residues such as tyrosine are located in the position of a relatively conserved disulfide bond in each sub-domain

of CBM13. The ADB contains only one disulfide bond on the surface of the β sub-domain. Third, the sugar-binding sites of the ABD of AkAbfB are completely different from those of CBM13. Thus, Miyanaga et al. [17] proposed that the ABD, the non-catalytic β-trefoil domain of GH54, be classified as a new CBM. Subsequently, the ABD has been classified into family 42 CBM in the CAZy database.

The ABD binds two arabinofuranose molecules in the structure of the complex with arabinose. These are located in pockets in the β and γ sub-domains. A number of structures of β-trefoil lectins and CBM13 in complex with ligands have been reported. The lactose-bound structure of the ricin toxin B-chain (RTB) from *Ricinus communis* [37] demonstrates the presence of three galactose-binding sites. The structures of CBM13 with various mono- or oligosaccharides have been reported [38–40]. RTB and CBM13 bind xylose or xylo-oligosaccharides at the same sugar-binding sites which are formed by two β-strands (β-2 and β-3) and a loop between β-3 and β-4. However, the location of the arabinose-binding sites of ABD of AkAbfB are clearly different and are shifted 8 Å in the direction of the next sub-domain. The structures around the binding sites are also very different. Furthermore, an additional helix is involved in the sugar recognition by ABD of AkAbfB.

Family 25 and 26 CBMs are structurally related to family 20 CBMs

There are currently only six classified families of starch-binding CBMs: families 20, 21, 25, 26, 34 and 41 [41]. Of these groups, the starch-binding properties have only been thoroughly characterized for two CBMs, namely those from families 20 and 41. NMR and thermodynamic studies by ITC indicate that CBM20 has two binding sites with approximately similar affinities for malto-oligosaccharides (ranging from ~6×10^2 M^{-1} for triose to ~3×10^3 M^{-1} for a pentaose) [42–43]. Remarkably, studies have shown that the concerted action of these two binding sites results in the disruption of amylase structure [44, 45]. In contrast, CBM41 from *Thermotoga maritima* Pul3 has only a single carbohydrate-binding site in the protein and the affinity of this module for malto-oligosaccharide is substantially higher, in the range of 10^6 M^{-1} [46]. Of the starch-binding CBMs, three-dimensional structures from members of families 20 and 34 are the only ones that have been determined. The NMR structure of the CBM20 from *A. niger* was the first to be determined [47]. Subsequently, numerous CBM20s have been solved by X-ray crystallography. The first CBM34 to have its structure solved by X-ray crystallography was the N-terminal CBM from *Thermoactinomyces vulgaris* TVAII [48].

Bacillus halodurans C-125 maltohexaose-forming amylase is a five-module protein comprised of an N-terminal family 13 catalytic module followed in order by two modules of unknown function, a family 26 CBM (BhCBM26) and a family 25 CBM (BhCBM25). Recently, Boraston et al. [12] presented an excellent structure-function analysis of starch and α-gluco-oligosaccharide recognition by BhCBM25 and BhCBM26. The two CBMs were both able to bind granular starch as assessed by depletion binding isotherms. BhCBM25 binds with an association constant (K_a) of 3.3×10^4 M^{-1}. The corresponding value for BhCBM26 was 3.7×10^4 M^{-1}. The BhCBM26/CBM25 tandem bound with a K_a of 1.6×10^6 M^{-1}. The roughly 50-fold affinity enhancement for the tandem CBMs relative to the single CBMs is common with tandem CBMs [49–51].

The crystal structures of BhCBM25 and BhCBM26 revealed that both CBMs have extremely similar β-sandwich folds (Fig. 5). Boraston et al. [12] state that BhCBM25 and BhCBM26 adopt an 'immunoglobulin-like' topology. The structures of BhCBM25 and BhCBM26 overlap with a root mean square deviation of 1.12 Å. Structural comparisons with DALI [18] show that the closest CBM structures are family 20 and 34 CBMs, which also adopt an immunoglobulin-like fold, as described below. BhCBM25 was initially crystallized in the space group $P2_1$ when using a large molar excess of maltotetraose. Clear electron density indicated that two molecules of maltotetraose were bound to BhCBM25. This phenomenon is somewhat at odds with solution studies, which indicate a single binding site on the protein. To clarify ligand binding of BhCBM25, Boraston et al. [12] cocrystallized BhCBM25 with a ~1:1 molar ratio of maltotetraose to protein to improve the probability of obtaining a crystal form with the predominant binding site occupied. Crystals obtained under these conditions belong to the space group $P4_32_12$. The electron density clearly showed only one ligand bound to one site, despite full accessibility of the other site. This result was consistent with solution studies. Thus, BhCBM25 has only one predominant malto-oligosaccharide-binding site with the pyranose rings of two adjacent sugars stacked against Tyr34 and Tyr74 (Fig. 5a).

BhCBM26 crystallized in the presence of a large excess of maltose. In contrast to BhCBM25, BHCBM26 was found to bind only a single molecule of maltose under these conditions (Fig. 5b). The architecture of the BhCBM26 binding site is similar to that of BhCBM25, as described above. The base of the binding site is formed by Trp36, Tyr23 and Tyr25, with the pyranose rings of the two sugar residues stacked against Trp36 and Tyr25. The binding of maltose to BhCBM26 appears to cause a small structural change in the protein. In contrast, no substantial conformational changes upon ligand binding to BhCBM25 were observed. Notable conformational

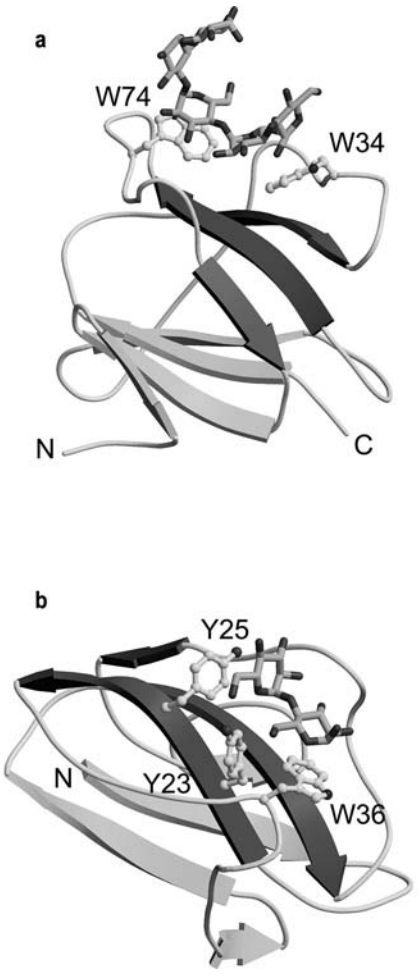

Figure 5. (*a*) Ribbon drawing of BhCBM25 complexed with maltotetraose. The secondary structural elements are highlighted, with the β-strands of each β-sheet colored light grey and dark grey. Maltotetraose is shown as a stick model, with oxygen atoms in dark grey. Trp34 and Trp74 are shown as ball-and-stick models. (*b*) Ribbon drawing of BhCBM26 complexed with maltose. The secondary structural elements are highlighted, with the β-strands of each β-sheet colored light grey and dark grey. Maltose is shown as a stick model, with oxygen atoms in dark grey. Tyr25 and Trp36 are shown as ball-and-stick models.

changes upon ligand binding have only been observed in CBM36 [26] and CBM35 [16]. The NMR structures of CBM35 are reviewed below.

The family 31 CBM structure reveals a typical immunoglobulin fold

Hashimoto et al. [14] reported the crystal structure of the family 31 CBM from *Alcaligenes* sp. strain XY-234 β-1,3-xylanase. The enzyme consists of two modular domains, a xylanase catalytic domain (GH26) and a family 31 CBM (AlcCBM31). The catalytic domain demonstrates hydrolysis activity toward β-1,3-xylan. The binding abilities of AlcCBM31 to a number of polysaccharides including β-1,3-glucan, β-1,4-glucan, β-1,4-mannan, β-1,3-xylan and β-1,4-xylan have previously been investigated [52]. These results demonstrated that β-1,3-xylan was the only one of the polysaccharides investigated for which there was binding activity [52].

The crystal structure shows that AlcCBM31 is assembled with 322 symmetry in the asymmetric unit and forms a spherical hexamer (Fig. 6a). The hexamer seems to be an artifact of crystallization or truncation, because gel filtration analysis of the full-length β-1,3-xylanase shows that the full-length protein exists as a monomer in solution. The monomeric structure of AlcCBM31 reveals a β-sandwich with a classical immunoglobulin fold, where β1, β2, β6 and β5, which are in one anti-parallel β-sheet, are in face contact with β4, β3, β7 and β8 in the other anti-parallel β-sheet (Fig. 6a, 7a, 7b). AlcCBM31 contains two intra-molecular disulfide bonds in the N- and C-terminal loop regions which bridge the N and C termini and stabilize the loop regions. A search for proteins structurally homologous to AlcCBM31 using the DALI database [18] shows that AlcCBM31 is topologically similar to family 34 (PDB ID; 1BVZ) and 9 (PDB ID; 1I82) CBMs. CBM34 forms the N-terminal domain of α-amylase II from *T. vulgaris* R-47 [53] and adopts an immunoglobulin fold (Fig. 7c, d). Structural studies reveal that the N-terminal domain of β-amylase I, which is also classified as a family 34 CBM, acts as a starch-binding domain [48]. CBM9 is the CBM of endo-1,4-β-xylanase [54]. The folding topology of CBM9 displays a structure based on an immunoglobulin fold (Fig. 7e, f), although additional secondary structures are inserted. Thus, a new sub-family is proposed to classify the CBMs with immunoglobulin folds. Recently, the crystal structure of a family 33 CBM was determined and also revealed an immunoglobulin fold [15]. The structure of CBM33 is reviewed in the next section (Fig. 7g, h). Further detailed classification of β-sandwich CBMs is provided in the last section.

The molecular surface of AlcCBM31 does not show any noticeable cleft likely to form a carbohydrate-binding site. Several aromatic residues (Tyr390, Trp409, Tyr411, Tyr417 and Tyr450), however, are solvent exposed (Fig. 6c). Tyr390 and Tyr411 are invariant in all five CBM31 family member sequences. Trp409 and Tyr417 are conserved in at least four of the family. Thus, these aromatic residues are candidates for involvement in binding to the β-1,3-xylan, although to be certain which residues form the ligand-binding site, further studies are required.

The family 33 CBM utilizes polar side chains for binding to chitin

Chitin is a linear insoluble polymer of β-1,4-linked *N*-acetylglucosamine and is a common constituent of fun-

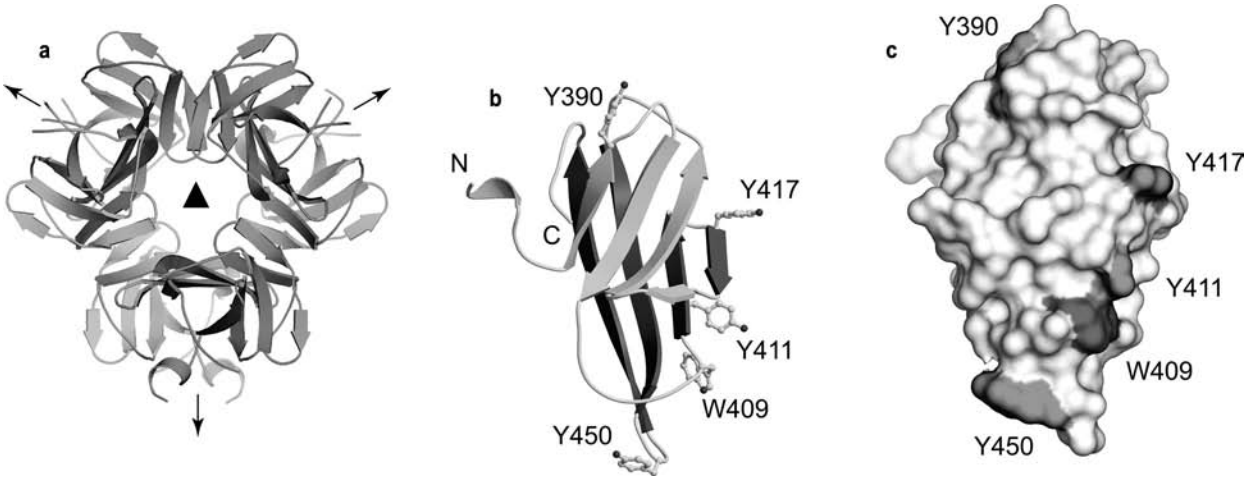

Figure 6. (*a*) Hexameric assembly of AlcCBM31 in an asymmetric unit. Non-crystallographic symmetries are shown; arrows and triangles show pseudo-twofold and pseudo-threefold axes. Each monomer is shown in grey. This figure was prepared using the program PyMOL (http://pymol.sourceforge.net/). (*b*) Ribbon drawing of a representative monomer. The secondary structural elements are highlighted, with the β-strands of each β-sheet colored light grey and dark grey. Solvent-exposed aromatic residues, which are candidate key residues for ligand binding, are shown as ball-and-stick models. (*c*) Surface representation of AlcCBM31. Candidate residues for ligand binding are highlighted.

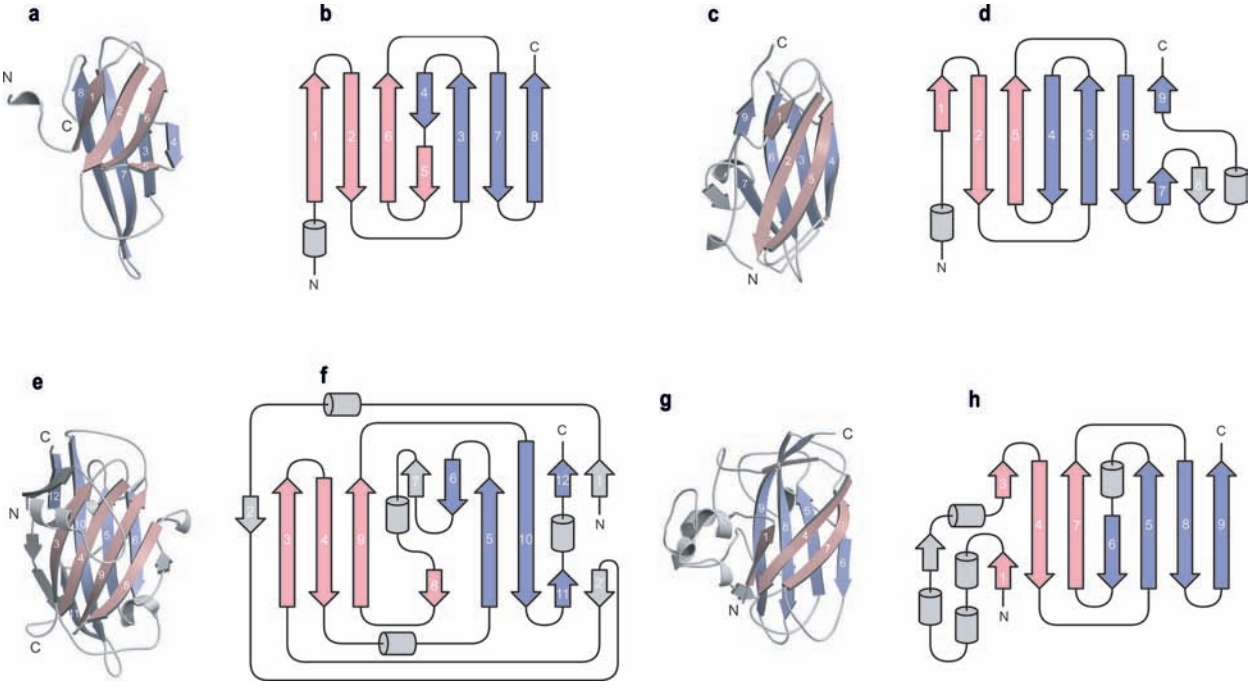

Figure 7. Ribbon representation and topology diagram of CBMs with immunoglobulin fold Ribbon representation of family 31 (*a*), 34 (*c*), 9 (*e*) and 33 (*g*) CBMs are shown. β-Strands colored pink and light purple are formed into the immunoglobulin fold. Topology diagrams of family 31 (*b*), 34 (*d*), 9 (*f*) and 33 (*h*) CBMs are also shown. β-Strands forming the immunoglobulin fold are highlighted in pink and purple. Pink and purple strands form anti-parallel β-sheets.

gal cell walls, shells of crustaceans and exoskeletons of insects. Chitin binding is an important event in mechanisms varying from plant responses to chitin-containing pathogens to chitin degradation by micro-organisms. Genes putatively encoding chitin-binding proteins or proteins containing chitin-binding domains have been found in many organisms. In the CAZy database they are classified into several CBM families (families 1, 2, 12, 14, 18, 19 and 33). The majority of these modules are domains of larger enzymes, where they are thought

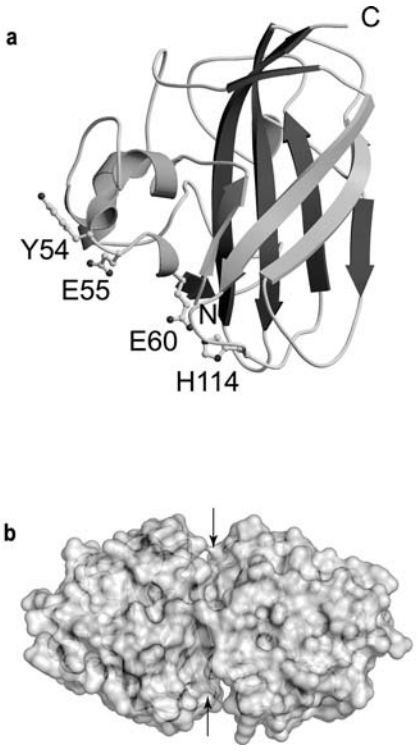

Figure 8. (*a*) CBP21, CBM33 member, is shown as a ribbon model. The secondary structural elements are highlighted with the β-strands of each β-sheet colored light grey and dark grey. Residues involved in ligand binding are shown as ball-and-stick models. (*b*) The molecular surface of the dimer is shown as a semi-transparent representation. Two clefts produced by dimer formation are indicated by arrows. This figure was prepared using the program PyMOL (http://pymol.sourceforge.net/).

to increase substrate affinity and enhance catalytic efficiency [55]. Chitin-binding modules also exist as independent, non-catalytic chitin-binding proteins (CBPs). Non-catalytic CBPs are found in family 14, 18 and 33 CBMs. Family 14 and 18 CBMs are small anti-fungal proteins that share a structurally similar chitin-binding motif [56]. Until recently, family 33 CBMs had not been structurally characterized, although low-resolution structural information was provided by X-ray solution scattering [57]. CBP21 from *Serratia marcescens* is classified as a family 33 CBM. Suzuki et al. [58] revealed that CBP21 has a specific preference for binding β-chitin and that the affinity is strongly pH dependent. Vaaje-Kolstad et al. [15] determined the crystal structure of CBP21. The structure consists of a three-stranded and a four-stranded β-sheet that form a β-sandwich (Fig. 8a). There is a bud-like 65-residue pseudo-domain between β-strands one and two, which consists of loops, one short α-helix and two short 3_{10} helices (Figures 8a). The β-sandwich core structure adopts a compact fibronectin III (FnIII)-type fold. Thus, the structure of CBP21 is constructed via an immunoglobulin fold (Fig. 7g, h). A search in the DALI database revealed that the closest structural match is to the FnIII-like chitin-binding domain of chitinase A (ChiA) from *S. marcescens* with a root mean square deviation of 2.5 Å. The ChiA-FnIII also adopts an immunoglobulin fold.

Three crystal forms, $P2_12_12_1$, $P3_22_1$ and $I2_12_12_1$, were obtained. In all crystals, two CBP21 molecules were packed together through pseudo-twofold symmetry (Fig. 8b). Through this dimerization, two identical clefts on both sides of the dimer are formed (Fig. 7b). To investigate the potential relevance of the dimer observed in the crystal, four CBP21 mutants (Y147A, Q161A, A152R and N163R) were constructed. Tyr147 and Gln161 are responsible for dimer interactions and Ala152 and Asn163 are on the cleft surface. All four mutants bound chitin with equal affinity, suggesting that neither the individual residues nor the cleft formed by dimerization play a role in the interaction with chitin. Sequence alignment revealed the presence of a cluster of conserved, mainly hydrophilic residues, and these are found at the surface of CBP21. In contrast to most other CBMs, CBP21 does not have a cluster of aromatic acids on the surface used for carbohydrate binding. To investigate the function of these conserved surface residues, a subset was mutated to alanine (Tyr54, Glu55, Glu66, His114, Asp182 and Asp185). It should be noted that Tyr54 is the only conserved surface-located aromatic residue in CBP21. A mutational study of a homologous CBP had previously indicated that an aromatic residue at this position is important for chitin binding [59]. The equilibrium dissociation constants (K_d) for each CBP21 mutant were measured against β-chitin. The largest increases in K_d were observed for Y54A (7.9-fold) and E60A (5.9-fold). E55A and H114A decreased binding capacity 2.8- and 2.0-fold, respectively. These results suggest that binding of CBM21 to chitin could be primarily governed by polar interactions.

CBM structures with a β-sandwich fold are divided into twofold sub-families

In 2004, Boraston et al. [5] classified 22 families of CBMs into seven fold families. Currently, CBMs are classified into 45 different families and structures from 31 of these families have been determined, as described above. This enables detailed structural classification of CBMs. Of these structures, 23 families have a β-sandwich motif. It is therefore clear that the β-sandwich is the major fold in CBM structures. To date, CBMs with a β-sandwich fold are found in families 2, 3, 4, 6, 9, 11, 15, 17, 20, 22, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 36 and 44. At the beginning of 2004, the majority of CBM structures with a β-sandwich motif had a β-jelly roll fold. However, in the last few years, a couple of CBMs with β-sandwich

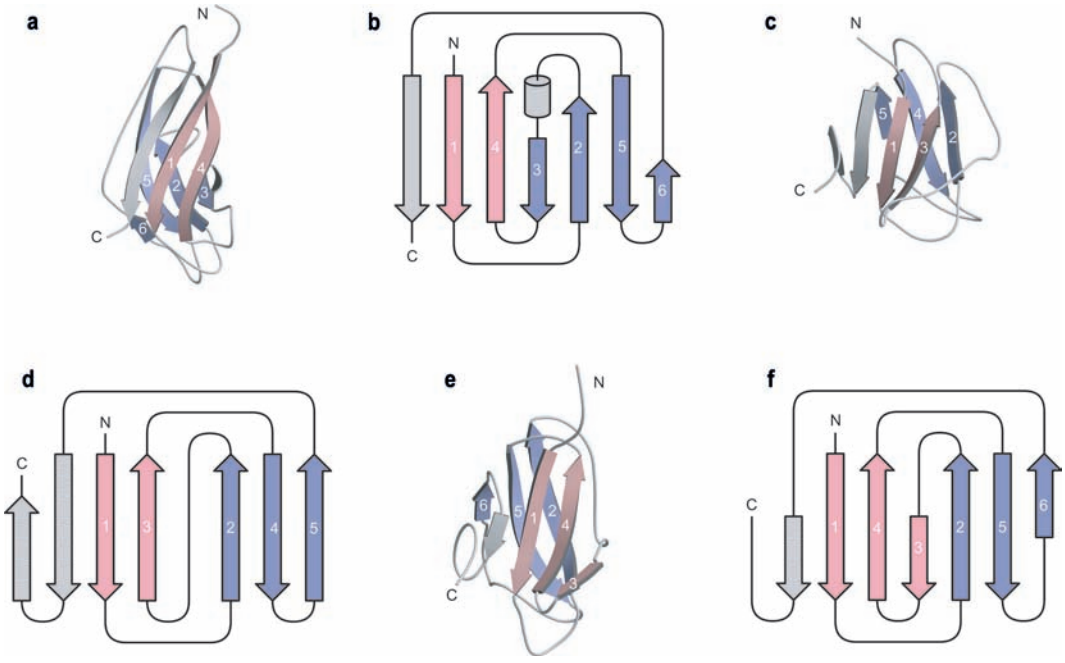

Figure 9. Ribbon representation and topology diagram of CBMs with an immunoglobulin-like fold. Ribbon representation of families 20 (*a*), 25 (*c*) and 26 (*e*) are shown. β-strands colored pink and light purple are formed into the immunoglobulin fold. Topology diagrams of families 20 (*b*), 25 (*d*) and 26 (*e*) are also shown. β-Strands forming the immunoglobulin-like fold are highlighted in pink and purple. Pink and purple strands form anti-parallel β-sheets.

structures have been shown to have an immunoglobulin fold. Currently, 7 families of CBMs are known to adopt the immunoglobulin fold (family 9, 20, 25, 26, 31, 33 and 34 CBMs). Of these, family 9, 31, 33 and 34 CBMs have a typical immunoglobulin fold, although some secondary structural elements are inserted (Fig. 7). The folding topologies of families 20, 25 and 26 CBMs are very similar to each other (Fig. 9), although some components of the immunoglobulin fold are deficient. In any case, these CBMs can be classified and grouped into the immunoglobulin fold sub-family. Thus, recent structural studies enable a detailed classification of β-sandwich structure. The structural classification based on Boraston's classification [5] is updated in Table 1, with new structures included. The β-sandwich CBM family is divided into two fold sub-families: β-jelly roll and immunoglobulin. To date, 16 families can be classified into the β-jelly roll sub-family and 7 into the immunoglobulin sub-family. It is unclear whether CBMs in the immunoglobulin sub-family have a functional correlation with each other. Because families 20, 25, 26 and 34 possess an ability to bind starch, this binding property may indicate some structure-function relationship in these CBMs. These families may be grouped as a clan, because members of a clan are not only structurally but also functionally related. In fact, family 20, 21, 25, 26, 34 and 41 CBMs are functionally grouped into starch-binding CBMs [41], as described above. Furthermore, a recent bioinfomatics study proposed that family 20 and 21 CBMs be classified into a new starch-binding domain-containing clan [60]. In contrast to these families, CBM family 9 and 31 are not functionally related, because CBM9 is a binding module of endo-1,4-β-xylanase while CBM31 is that of β-1,3-xylanase.

Table 1. CBM fold families based on the review by Boraston et al. [5].

Fold family	Fold		CBM families
1	β-sandwich		
		β-jelly roll	2, 3, 4, 6, **11**, 15, 17, 22, 27, 28, 29, **30**, 32, **35**, 36, **44**
		immunoglobulin	9, 20, **25**, **26**, **31**, **33**, 34
2	β-trefoil		13, **42**
3	cystein knot		1
4	unique		5, 12
5	OB fold		10
6	hevein fold		18
7	unique		14

The nine novel structures are highlighted in bold.

In future, novel CBM families are likely to be found, and progress in structural and functional studies of novel CBM families should bring about a more detailed classification into sub-families or clans. This will lead to further understanding of the structure-function and evolutionary relationships of these molecules.

1 Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T. and Claeyssens, M. (1988) Studies of the cellulolytic system of *Trichoderma reesei* QM 9414: analysis of domain function in two cellobiohydrolases by limited proteolysis. Eur. J. Biochem. 170, 575–581.

2 Gilkes, N. R., Warren, R. A., Miller, R. C. Jr and Kilburn, D. G. (1988) Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. J. Biol. Chem. 263, 10401–10407.

3 Tomme, P., Warren, R. A. and Gilkes, N. R. (1995) Cellulose hydrolysis by bacteria and fungi. Adv. Microb. Physiol. 37, 1–81.

4 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.

5 Boraston, A. B., Bolam, D. N., Gilbert, H. J. and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. 382, 769–781.

6 Jamal, S., Nurizzo, D., Boraston, A. B. and Davies, G. J. (2004) X-ray crystal structure of a non-crystalline cellulose-specific carbohydrate-binding module: CBM28. J. Mol. Biol. 339, 253–258.

7 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.

8 McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A. and Kilburn, D. G. (2000) Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. Protein Eng. 13, 801–809.

9 Nagy, T., Simpson, P., Williamson, M. P., Hazlewood, G. P., Gilbert, H. J. and Orosz, L. (1998) All three surface tryptophans in Type IIa cellulose binding domains play a pivotal role in binding both soluble and insoluble ligands. FEBS Lett. 429, 312–316.

10 Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J. and Williamson, M. P. (2000) The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. J. Biol. Chem. 275, 41137–41142.

11 Carvalho, A. L., Goyal, A., Prates, J. A., Bolam, D. N., Gilbert, H. J., Pires, V. M., Ferreira, L. M., Planas, A., Romao, M. J. and Fontes, C. M. (2004) The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates beta-1,4- and beta-1,3-1,4-mixed linked glucans at a single binding site. J. Biol. Chem. 279, 34785–34793.

12 Boraston, A. B., Healey, M., Klassen, J., Ficko-Blean, E., Lammerts van Bueren, A. and Law, V. (2006) A structural and functional analysis of alpha-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. J. Biol. Chem. 281, 587–598.

13 Najmudin, S., Guerreiro, C. I., Carvalho, A. L., Prates, J. A., Correia, M. A., Alves, V. D., Ferreira, L. M., Romao, M. J., Gilbert, H. J., Bolam, D. N. and Fontes, C. M. (2006) Xyloglucan is recognized by carbohydrate-binding modules that interact with beta-glucan chains. J. Biol. Chem. 281, 8815–8828.

14 Hashimoto, H., Tamai, Y., Okazaki, F., Tamaru, Y., Shimizu, T., Araki, T. and Sato, M. (2005) The first crystal structure of a family 31 carbohydrate-binding module with affinity to beta-1,3-xylan. FEBS Lett. 579, 4324–4328.

15 Vaaje-Kolstad, G., Houston, D. R., Riemen, A. H., Eijsink, V. G. and van Aalten, D. M. (2005) Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21. J. Biol. Chem. 280, 11313–11319.

16 Tunnicliffe, R. B., Bolam, D. N., Pell, G., Gilbert, H. J. and Williamson, M. P. (2005) Structure of a mannan-specific family 35 carbohydrate-binding module: evidence for significant conformational changes upon ligand binding. J. Mol. Biol. 347, 287–296.

17 Miyanaga, A., Koseki, T., Matsuzawa, H., Wakagi, T., Shoun, H. and Fushinobu, S. (2004) Crystal structure of a family 54 alpha-L-arabinofuranosidase reveals a novel carbohydrate-binding module that can bind arabinose. J. Biol. Chem. 279, 44907–44914.

18 Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. J. Mol. Biol. 233, 123–138.

19 Notenboom, V., Boraston, A. B., Chiu, P., Freelove, A. C., Kilburn, D. G. and Rose, D. R. (2001) Recognition of cello-oligosaccharides by a family 17 carbohydrate-binding module: an X-ray crystallographic, thermodynamic and mutagenic study. J. Mol. Biol. 314, 797–806.

20 Abou-Hachem, M., Karlsson, E. N., Simpson, P. J., Linse, S., Sellers, P., Williamson, M. P., Jamieson, S. J., Gilbert, H. J., Bolam, D. N. and Holst, O. (2002) Calcium binding and thermostability of carbohydrate binding module CBM4–2 of Xyn10A from *Rhodothermus marinus*. Biochemistry 41, 5720–5729.

21 Bolam, D. N., Xie, H., Pell, G., Hogg, D., Galbraith, G., Henrissat, B. and Gilbert, H. J. (2004) X4 modules represent a new family of carbohydrate-binding modules that display novel properties. J. Biol. Chem. 279, 22953–22963.

22 Pires, V. M., Henshaw, J. L., Prates, J. A., Bolam, D. N., Ferreira, L. M., Fontes, C. M., Henrissat, B., Planas, A., Gilbert, H. J. and Czjzek, M. (2004) The crystal structure of the family 6 carbohydrate binding module from *Cellvibrio mixtus* endoglucanase 5a in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. J. Biol. Chem. 279, 21560–21568.

23 Henshaw, J. L., Bolam, D. N., Pires, V. M., Czjzek, M., Henrissat, B., Ferreira, L. M., Fontes, C. M. and Gilbert, H. J. (2004) The family 6 carbohydrate binding module CmCBM6–2 contains two ligand-binding sites with distinct specificities. J. Biol. Chem. 279, 21552–21559.

24 Ahsan, M. M., Kimura, T., Karita, S., Sakka, K. and Ohmiya, K. (1996) Cloning, DNA sequencing, and expression of the gene encoding *Clostridium thermocellum* cellulase CelJ, the largest catalytic component of the cellulosome. J. Bacteriol. 178, 5732–5740.

25 Arai, T., Araki, R., Tanaka, A., Karita, S., Kimura, T., Sakka, K. and Ohmiya, K. (2003) Characterization of a cellulase containing a family 30 carbohydrate-binding module (CBM) derived from *Clostridium thermocellum* CelJ: importance of the CBM to cellulose hydrolysis. J. Bacteriol. 185, 504–512.

26 Jamal-Talabani, S., Boraston, A. B., Turkenburg, J. P., Tarbouriech, N., Ducros, V. M. and Davies, G. J. (2004) Ab initio structure determination and functional characterization of CBM36; a new family of calcium-dependent carbohydrate binding modules. Structure 12, 1177–1187.

27 Bycroft, M., Bateman, A., Clarke, J., Hamill, S. J., Sandford, R., Thomas, R. L. and Chothia, C. (1999) The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease. EMBO J. 18, 297–305.

28 Czjzek, M., Bolam, D. N., Mosbah, A., Allouch, J., Fontes, C. M., Ferreira, L. M., Bornet, O., Zamboni, V., Darbon, H., Smith, N. L., Black, G. W., Henrissat, B. and Gilbert, H. J. (2001) The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved. J. Biol. Chem. 276, 48580–48587.
29 Kaneko, S., Arimoto, M., Ohba, M., Kobayashi, H., Ishii, T. and Kusakabe, I. (1998) Purification and substrate specificities of two alpha-L-arabinofuranosidases from *Aspergillus awamori* IFO 4033. Appl. Environ. Microbiol. 64, 4021–4027.
30 Saha, B. C. (2000) Alpha-L-arabinofuranosidases: biochemistry, molecular biology and application in biotechnology. Biotechnol. Adv. 18, 403–423.
31 Koseki, T., Okuda, M., Sudoh, S., Kizaki, Y., Iwano, K., Aramaki, I. and Matsuzawa, H. (2003) Role of two alpha-L-arabinofuranosidases in arabinoxylan degradation and characteristics of the encoding genes from shochu koji molds, *Aspergillus kawachii* and *Aspergillus awamori.* J. Biosci. Bioeng. 96, 232–241.
32 De Ioannes, P., Peirano, A., Steiner, J. and Eyzaguirre, J. (2000) An alpha-L-arabinofuranosidase from *Penicillium purpurogenum*: production, purification and properties. J. Biotechnol. 76, 253–258.
33 Hovel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y. and Schomburg, D. (2003) Crystal structure and snapshots along the reaction pathway of a family 51 alpha-L-arabinofuranosidase. EMBO J. 22, 4922–4932.
34 Shallom, D., Belakhov, V., Solomon, D., Gilead-Gropper, S., Baasov, T., Shoham, G. and Shoham, Y. (2002) The identification of the acid-base catalyst of alpha-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. FEBS Lett. 514, 163–167.
35 Shallom, D., Belakhov, V., Solomon, D., Shoham, G., Baasov, T. and Shoham, Y. (2002) Detailed kinetic analysis and identification of the nucleophile in alpha-L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. J. Biol. Chem. 277, 43667–43673.
36 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.
37 Montfort, W., Villafranca, J. E., Monzingo, A. F., Ernst, S. R., Katzin, B., Rutenber, E., Xuong, N. H., Hamlin, R. and Robertus, J. D. (1987) The three-dimensional structure of ricin at 2.8 Å. J. Biol. Chem. 262, 5398–5403.
38 Fujimoto, Z., Kuno, A., Kaneko, S., Kobayashi, H., Kusakabe, I. and Mizuno, H. (2002) Crystal structures of the sugar complexes of *Streptomyces olivaceoviridis* E-86 xylanase: sugar binding structure of the family 13 carbohydrate binding module. J. Mol. Biol. 316, 65–78.
39 Fujimoto, Z., Kaneko, S., Kuno, A., Kobayashi, H., Kusakabe, I. and Mizuno, H. (2004) Crystal structures of decorated xylooligosaccharides bound to a family 10 xylanase from *Streptomyces olivaceoviridis* E-86. J. Biol. Chem. 279, 9606–9614.
40 Notenboom, V., Boraston, A. B., Williams, S. J., Kilburn, D. G. and Rose, D. R. (2002) High-resolution crystal structures of the lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A with bound substrates reveal a novel mode of xylan binding. Biochemistry 41, 4246–4254.
41 Rodriguez-Sanoja, R., Oviedo, N. and Sanchez, S. (2005) Microbial starch-binding domain. Curr. Opin. Microbiol. 8, 260–267.
42 Sigurskjold, B. W., Svensson, B., Williamson, G. and Driguez, H. (1994) Thermodynamics of ligand binding to the starch-binding domain of glucoamylase from *Aspergillus niger.* Eur. J. Biochem. 225, 133–141.
43 Sorimachi, K., Le Gal-Coeffet, M. F., Williamson, G., Archer, D. B. and Williamson, M. P. (1997) Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to beta-cyclodextrin. Structure 5, 647–661.
44 Giardina, T., Gunning, A. P., Juge, N., Faulds, C. B., Furniss, C. S., Svensson, B., Morris, V. J. and Williamson, G. (2001) Both binding sites of the starch-binding domain of *Aspergillus niger* glucoamylase are essential for inducing a conformational change in amylose. J. Mol. Biol. 313, 1149–1159.
45 Southall, S. M., Simpson, P. J., Gilbert, H. J., Williamson, G. and Williamson, M. P. (1999) The starch-binding domain from glucoamylase disrupts the structure of starch. FEBS Lett. 447, 58–60.
46 Lammerts van Bueren, A., Finn, R., Ausio, J. and Boraston, A. B. (2004) Alpha-glucan recognition by a new family of carbohydrate-binding modules found primarily in bacterial pathogens. Biochemistry 43, 15633–15642.
47 Sorimachi, K., Jacks, A. J., Le Gal-Coeffet, M. F., Williamson, G., Archer, D. B. and Williamson, M. P. (1996) Solution structure of the granular starch binding domain of glucoamylase from *Aspergillus niger* by nuclear magnetic resonance spectroscopy. J. Mol. Biol. 259, 970–987.
48 Abe, A., Tonozuka, T., Sakano, Y. and Kamitori, S. (2004) Complex structures of *Thermoactinomyces vulgaris* R-47 alpha-amylase 1 with malto-oligosaccharides demonstrate the role of domain N acting as a starch-binding domain. J. Mol. Biol. 335, 811–822.
49 Boraston, A. B., Kwan, E., Chiu, P., Warren, R. A. and Kilburn, D. G. (2003) Recognition and hydrolysis of noncrystalline cellulose. J. Biol. Chem. 278, 6120–6127.
50 Boraston, A. B., McLean, B. W., Chen, G., Li, A., Warren, R. A. and Kilburn, D. G. (2002) Co-operative binding of triplicate carbohydrate-binding modules from a thermophilic xylanase. Mol. Microbiol. 43, 187–194.
51 Bolam, D. N., Xie, H., White, P., Simpson, P. J., Hancock, S. M., Williamson, M. P. and Gilbert, H. J. (2001) Evidence for synergy between family 2b carbohydrate binding modules in *Cellulomonas fimi* xylanase 11A. Biochemistry 40, 2468–2477.
52 Okazaki, F., Tamaru, Y., Hashikawa, S., Li, Y. T. and Araki, T. (2002) Novel carbohydrate-binding module of beta-1,3-xylanase from a marine bacterium, *Alcaligenes* sp. strain XY-234. J. Bacteriol. 184, 2399–2403.
53 Kamitori, S., Kondo, S., Okuyama, K., Yokota, T., Shimura, Y., Tonozuka, T. and Sakano, Y. (1999) Crystal structure of *Thermoactinomyces vulgaris* R-47 alpha-amylase II (TVAII) hydrolyzing cyclodextrins and pullulan at 2.6 Å resolution. J. Mol. Biol. 287, 907–921.
54 Notenboom, V., Boraston, A. B., Kilburn, D. G. and Rose, D. R. (2001) Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. Biochemistry 40, 6248–6256.
55 Hashimoto, M., Ikegami, T., Seino, S., Ohuchi, N., Fukada, H., Sugiyama, J., Shirakawa, M. and Watanabe, T. (2000) Expression and characterization of the chitin-binding domain of chitinase A1 from *Bacillus circulans* WL-12. J. Bacteriol. 182, 3045–3054.
56 Suetake, T., Tsuda, S., Kawabata, S., Miura, K., Iwanaga, S., Hikichi, K., Nitta, K. and Kawano, K. (2000) Chitin-binding proteins in invertebrates and plants comprise a common chitin-binding structural motif. J. Biol. Chem. 275, 17929–17932.
57 Svergun, D. I., Becirevic, A., Schrempf, H., Koch, M. H. and Gruber, G. (2000) Solution structure and conformational changes of the *Streptomyces* chitin-binding protein (CHB1). Biochemistry 39, 10677–10683.
58 Suzuki, K., Suzuki, M., Taiyoji, M., Nikaidou, N. and Watanabe, T. (1998) Chitin binding protein (CBP21) in the culture supernatant of *Serratia marcescens* 2170. Biosci. Biotechnol. Biochem. 62, 128–135.
59 Zeltins, A. and Schrempf, H. (1997) Specific interaction of the *Streptomyces* chitin-binding protein CHB1 with alpha-chitin

– the role of individual tryptophan residues. Eur. J. Biochem. 246, 557–564.
60 Machovic, M., Svensson, B., MacGregor, E. A. and Janecek, S. (2005) A new clan of CBM families based on bioinformatics of starch-binding domains from families CBM20 and CBM21. FEBS J. 272, 5497–5513.
61 Kraulis, P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24, 946–950.
62 Merritt, E. A. and Murphy, M. E. (1994) Raster3D Version 2.0: a program for photorealistic molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 50, 869–873.