Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a family IIIa scaffoldin CBD from the cellulosome of *Clostridium cellulolyticum* at 2.2 Å resolution

Linda J. W. Shimon,[a]* Sandrine Pagès,[b,c] Anne Belaich,[b] Jean-Pierre Belaich,[b,c] Edward A. Bayer,[d] Raphael Lamed,[e] Yuval Shoham[f] and Felix Frolow[e]

[a]Faculty of Chemistry, The Weizmann Institute of Science, Rehovot, Israel, [b]Bioénergétique et Ingéniérie des Protéines, Centre National de la Recherche Scientifique, IBSM-IFR1, 13402 Marseille, France, [c]University of Provence, Marseille, France, [d]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel, [e]Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Ramat Aviv, Israel, and [f]Department of Food Engineering and Biotechnology, Technion – Israel Institute of Technology, Haifa, Israel

Correspondence e-mail: linda.shimon@weizmann.ac.il

Received 9 May 2000
Accepted 19 September 2000

PDB Reference: CBD, 1g43.

The crystal structure of the family IIIa cellulose-binding domain (CBD) from the cellulosomal scaffoldin subunit (CipC) of *Clostridium cellulolyticum* has been determined. The structure reveals a nine-stranded jelly-roll topology which exhibits distinctive structural elements consistent with family III CBDs that bind crystalline cellulose. These include a well conserved calcium-binding site, a putative cellulose-binding surface and a conserved shallow groove of unknown function. The CipC CBD structure is very similar to the previously elucidated family IIIa CBD from the CipA scaffoldin of *C. thermocellum*, with some minor differences. The CipC CBD structure was also compared with other previously described CBD structures from families IIIc and IV derived from the endoglucanases of *Thermomonospora fusca* and *Cellulomonas fimi*, respectively. The possible functional consequences of structural similarities and differences in the shallow groove and cellulose-binding faces among various CBD families and subfamilies are discussed.

1. Introduction

Cellulases and related enzymes (*i.e.* glycosyl hydrolases) are multimodular proteins which typically contain a series of structurally and functionally independent modules on the same polypeptide chain (Gilkes *et al.*, 1991; Davies & Henrissat, 1995; Henrissat & Davies, 1997; Bayer, Chanzy *et al.*, 1998). The first step in the process of cellulose degradation is the binding of the cellulolytic enzyme(s) or the entire microorganism to the cellulose substrate (Tomme *et al.*, 1995; Bayer *et al.*, 1996). Binding to cellulose is mediated by a separate type of module, named the cellulose-binding domain (CBD), and a separate module, the catalytic domain, carries out the actual hydrolysis of the cellulose chain.

In many cases, a CBD is an integral component of a free enzyme. In this case, the CBD serves as a targeting agent that delivers the adjoining catalytic module to the insoluble crystalline substrate. Alternatively, an enzyme-borne CBD may recognize disengaged cellulose chains; such CBDs may play a 'helper' role whereby they assist substrate hydrolysis by selectively directing a single cellulose chain into the active site of the catalytic module. In still other cases, cellulolytic enzymes are organized into an efficient multifunctional enzyme complex called the cellulosome (Bayer, Shimon *et al.*, 1998). In this case, the CBD comprises a component part of a separate non-catalytic 'scaffoldin' subunit to which the enzyme subunits of the cellulosome are specifically but non-covalently attached. Thus, rather than targeting single enzymes to the substrate, the scaffoldin CBD alone drags along all of the cellulosomal enzymes to the insoluble cellu-

losic substrate. The importance of the cellulosomal CBD is thus multiplied tremendously, since substrate binding of all the cellulosomal enzymes as well as the entire cell is dependent on the presence of a single domain on a specialized subunit.

CBDs have been classified into about a dozen distinct families according to sequence homology (Tomme *et al.*, 1995). Some CBDs have been shown to bind selectively to insoluble cellulose or a related polysaccharide (*e.g.* xylan or chitin). In some cases, CBDs are specific for soluble cellooligosaccharides or for less crystalline forms of cellulose, which suggest that certain types of CBD may be involved in targeting an enzyme to distinct regions of the substrate. The CBDs belong to a greater class of carbohydrate-binding modules (CBMs), which emphasizes sequence and structural homology rather than substrate specificity (Coutinho & Henrissat, 1999).

Solution structures for members of CBDs from families I (Kraulis *et al.*, 1989), II (Xu *et al.*, 1995), IV (Johnson, Joshi *et al.*, 1996) and V (Brun *et al.*, 1997) have been determined. Crystal structures of only two different types of CBD are currently known (Tormo *et al.*, 1996; Sakon *et al.*, 1997), both of which are classified in family III but are of different subtypes, termed family IIIa and family IIIc. The family IIIa and IIIb CBDs are closely similar in their primary structures and both types bind strongly to crystalline cellulose. Members of the family IIIc CBDs, however, are characteristically fused to a family 9 glycosyl hydrolase. This subfamily fails to bind crystalline cellulose, but reportedly serves in a 'helper' capacity by feeding a single incoming cellulose chain into the active site of the neighbouring family 9 catalytic module pending hydrolysis (Gal *et al.*, 1997; Irwin *et al.*, 1998). The phylogenetic relationship of the family III CBDs follows the general pattern of their function (Fig. 1). All the family IIIc CBDs form a distinct cluster on one side of the tree. On the opposite side of the weighted centroid are scattered the family IIIb CBDs. The family IIIa CBDs from the clostridial scaffoldins occupy a single branch which emanates from an intermediary position among the family IIIb CBDs.

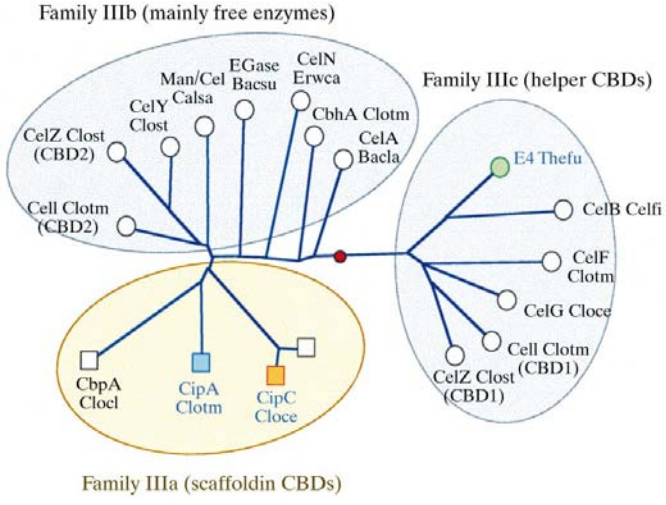

Figure 1
Phylogenetic analysis of the family III CBDs. Scaffoldin CBDs are shown as squares and enzyme-borne CBDs as circles. Sequences were obtained from the respective GenBank accession codes (included parenthetically) as follows: CipC Cloce, scaffoldin from *C. cellulolyticum* (U40345); CipA Clotm, scaffoldin from *C. thermocellum* (L08665); CbpA Clocl, scaffoldin from *C. cellulovorans* (M73817); CipA Clojo, scaffoldin from *C. josui* (AB004845); CelF Clotm and CelI Clotm, endoglucanases F and I from *C. thermocellum* (X60545 and L04735, respectively); CelG Cloce, cellulase G from *C. cellulolyticum* (M87018); CelY Closr and CelZ, endoglucanses Y and Z from *C. stercorarium* (Z69359 and X55299, respectively); Man/Cel Calsa, mannanase/cellulase from *Caldocellum saccharolyticum* (L01257); EGase Bacsu, endoglucanase from *Bacillus subtilis* (M16185); CelN Erwca, cellulase N from *Erwinia carotovora atroseptica* (L39788); CelA Bacla, cellulase A from *Bacillus lautus* (M76588); E4 Thefu, cellulase E4 from *Thermobifida fusca* (M73322).

The first crystal structure of a CBD (Tormo *et al.*, 1996) was determined for the family IIIa scaffoldin CBD from the cellulosome of *C. thermocellum* (Poole *et al.*, 1992; Gerngross *et al.*, 1993). By comparing its structure with those of other families, a general molecular mechanism was proposed for the binding of CBDs to cellulose (Tormo *et al.*, 1996). Sequences of three other scaffoldin CBDs are currently known, all of which are very similar to each other (Shoseyov *et al.*, 1992; Kakiuchi *et al.*, 1998; Pagès *et al.*, 1999). Nevertheless, minor differences are evident from sequence alignment (Fig. 2). It was thus of interest to compare the possible structural consequences that such 'natural' substitutions would impose. For this purpose, the homologous CBD from a related but distinct cellulosome-producing bacterium, *C. cellulolyticum*, was crystallized in the present work. Its structure was nearly identical to that of *C. thermocellum*, indicating that nature appears to maintain strict conservation in this type of protein.

2. Materials and methods

2.1. Crystallization and data collection

The 18 kDa recombinant protein consisting of the 160-residue CBD from the scaffoldin subunit (CipC) of *C. cellulolyticum* with portions of the adjoining N-terminal linker regions was overexpressed in *Escherichia coli* and purified as described previously (Pagès *et al.*, 1997). The protein was crystallized using the hanging-drop vapour-diffusion method. Crystals were grown from drops where 2 µl of protein solution at an initial concentration of 7.3 mg ml^{-1} was mixed with 2 µl of precipitant solution of 17% PEG 8000, 100 m*M* sodium cacodylate buffer pH 6.5 and 15 m*M* zinc acetate. Typical prismatic crystals grew in about two weeks with maximum dimensions of 0.3 × 0.2 × 0.2 mm. The space group is $P6_522$, with unit-cell parameters $a = 46.86$, $b = 46.86$, $c = 244.12$ Å and one molecule in the asymmetric unit.

The crystal was mounted in a sealed glass capillary and X-ray diffraction data were collected at room temperature using an R-AXIS II imaging-plate detector system (equipped with blue image plates) mounted on a Rigaku FR-C rotating-anode generator (Cu $K\alpha$, focal cup 0.1 × 0.1 mm, 50 kV, 55 mA) equipped with focusing mirrors. A rotation angle of 1.0° was used per exposure. All data were integrated and scaled using the *HKL* program suite (Otwinowski, 1993). The data-collection statistics are summarized in Table 1.

2.2. Structure determination

The structure of the CipC CBD was obtained by molecular replacement using *AMoRe* (Navaza, 1994) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). Since the CBD from *C. cellulolyticum* displays 43% identity with that from *C. thermocellum*, the atomic coordinates of the latter CBD (PDB code 1nbc) were used as a search model.

A radius of 20 Å was chosen for the rotation function, with data in the resolution range 20–3 Å. No clearly distinguishable peak was seen, but the first 100 solutions were placed in a translation search. Data in the resolution range 8.0–3.0 Å were used for the translation function. One clear solution (number 41) was obtained with a correlation coefficient of 0.35. After 20 cycles of rigid-body fitting using data in the resolution range 20–2.5 Å, the correlation coefficient was 0.54 and the *R* factor was 0.48.

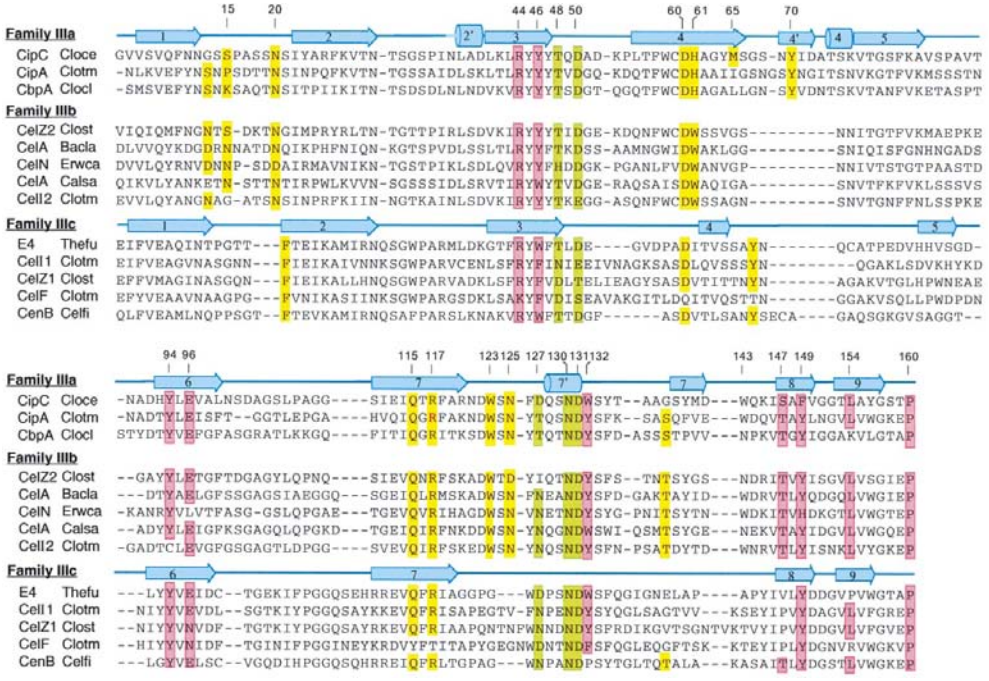

Figure 2
Structure-based sequence alignment of family III CBDs. Secondary-structural elements are indicated and enumerated. Proposed cellulose-binding residues are shown in yellow, as are the homologous residues in the other CBDs from this family. Calcium-binding residues are shown in green. Red indicates highly conserved residues, which occupy the shallow groove of unknown function on the top face of the molecule. The CBDs indicated in the figure are from the parent proteins among the list in the legend to Fig. 1.

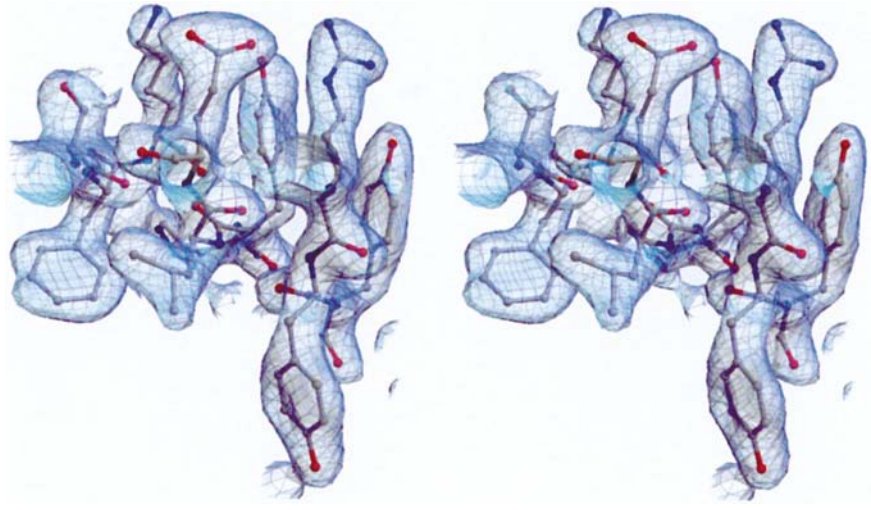

Figure 3
Stereoview of the $2|F_o| - |F_c|$ electron-density map contoured at 1.2σ in the region of the shallow-groove region, which includes residues in the region of β-strands 3, 5 and 6.

2.3. Refinement

A random subset of the data (10%) was set aside for cross-validation and assessment of the refinement (Brünger, 1992). The search model was then subjected to refinement against data in the resolution range 20.0–2.2 Å using *CNS* (Brunger *et al.*, 1998). This converged to an *R* factor of 0.52 and an $R_{\rm free}$ of 0.53. The total omit-map procedure was applied in *CNS* and then examined in *O* in order to assess the quality of the solution. The $2F_o - F_c$ map revealed the positions of the non-conserved side chains of the CipC sequence, thereby allowing them to be manually built into the map using the program *O* (Jones *et al.*, 1991).

Additional density for the small insertion (residues 101–102) and the large shift of the extended loop between strands 4 and 4′ were clear. At the N-terminus, four additional residues could be clearly built.

Several rounds of refinement of the model in the *CNS* torsional dynamic method with the bulk-solvent option activated were performed, alternating with refinement of group *B* factors. One calcium ion and two zinc ions were located in the electron density and were built into the model. The *R* factor at this stage was 0.261 and $R_{\rm free}$ was 0.319.

Next, individual *B* factors were introduced. During the final stages of refinement water molecules were inserted into the model contingent upon the following three criteria: (i) if the

Table 1
Data collection.

Values for the outermost shell are given in parentheses.

Crystal	
Space group	$P6_522$
Temperature	RT
Unit-cell parameters (Å)	a = 46.86, b = 46.86, c = 244.12
Internal scaling	
Resolution (Å)	20–2.20
No. of crystals	1
Reflections measured	41312
Unique relections	8466
Completeness (%)	94.3 (76.6)
$I/\sigma(I)$ (average)	25.9
R_{sym}† (%)	7.3 (33.7)

† $R_{sym} = \sum|I_{obs} - I_{avg}|/\sum I_{avg}$.

Table 2
Model refinement.

Resolution range (Å)	20.0–2.20
No. non-H atoms	1212
No. of solvent atoms	96
No. of ion atoms	3
R_{cryst}†	0.190
R_{free}	0.264
Geometry	
RMS bonds (Å)	0.012
RMS bond angles (°)	1.70
RMS dihedral angles (°)	27.90
RMS impropers (°)	0.99

† $R_{cryst} = \sum ||F_{obs}|_{(hkl)} - |F_{calc}|_{(hkl)}|/\sum|F_{obs}|_{(hkl)}$.

peaks in the $F_o - F_c$ electron-density maps exhibited heights greater than 3σ, (ii) if the peaks were within hydrogen-bonding distance of appropriate atoms and (iii) if their placement caused an improvement in R_{free}. Using these criteria, 96 well ordered water molecules were placed. The model underwent several more cycles of manual rebuilding and refinement, ending with individual temperature-factor refinement of the final model. The refinement converged at an R factor of 0.190 and an R_{free} of 0.264 (see Table 2 for refinement statistics).

3. Results and discussion

3.1. Overview

The CBD of the cellulosomal scaffoldin protein CipC from *C. cellulolyticum* was cloned and expressed. The N-terminus of the original clone included six residues (MFAAGT) upstream of the sequence shown in Fig. 2. The protein was crystallized by the hanging-drop vapour-diffusion method. The crystals belonged to the hexagonal space group $P6_522$ and contained one molecule per asymmetric unit. The crystal was determined by molecular replacement using the *C. thermocellum* CBD as a search model. The final atomic model was refined against the data to 2.2 Å to a crystallographic R factor of 0.190 (Table 2). The final electron-density map is of high quality (Fig. 3). The first three amino acids in the original clone, however, were not seen in the electron-density maps; numbering of the CipC CBD commences with the first ordered residue, *i.e.* alanine, followed by glycine and threonine. The stereochemical quality of the structure is very good, with over 88% of the amino-acid residues in the most favourable regions of the Ramachandran plot and no outliers (data not shown). The structure also contains one calcium ion, two zinc ions and 96 water molecules.

3.2. Overall description

The 160 amino-acid CipC CBD is a nine-stranded β-sandwich of overall dimensions 48 × 28 × 23 Å (Fig. 4) and is very similar to the structure reported by Tormo *et al.* (1996). The top face of the β-sandwich comprises relatively short antiparallel strands (5, 6, 3, 8 and 9) that form a shallow groove. Both the N- and C-termini are found in this region; proline 160, the last residue in the protein, packs tightly into this groove. The bottom face is flat, comprising the relatively long strands 1, 2, 7 and 4 (Fig. 5).

3.3. Calcium binding

One Ca atom, which probably serves a structural role, was found bound in a cavity between loops 3/4 and the region between strands 7 and 8, *i.e.* the short 7″ helical stretch. The coordinating ligands are Thr48 main-chain O (2.41 Å), OG1 (2.48 Å), Asp50 OD2 (2.54 Å), Asp127 main-chain O (2.41 Å), Asn130 OD1 (2.41 Å), Asp131 OD1 (2.43 Å) and OD2 (2.47 Å). A buried water molecule forms the eighth ligand (2.68 Å).

This pentagonal bipyramidal arrangement of Ca^{2+} was also seen in *C. thermocellum* CBD CipA (Tormo *et al.*, 1996) and in the *T. fusca* E4 CBD (Sakon *et al.*, 1997). Residues 48, 50, 130

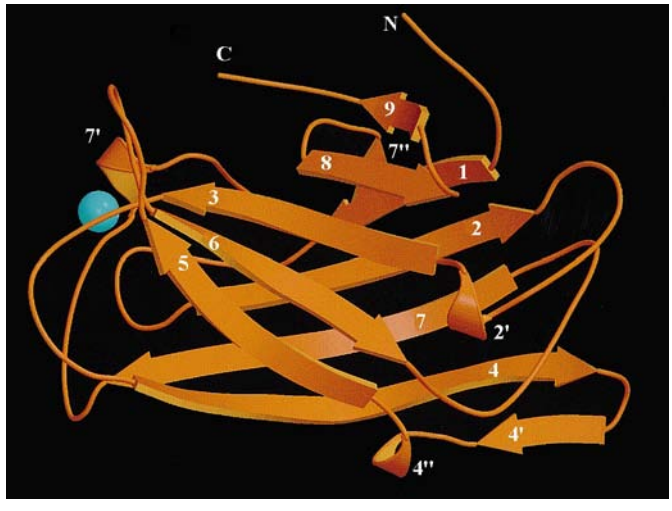

Figure 4
The structure of CipC CBD from *C. cellulolyticum*. Schematic ribbon diagram of the overall three-dimensional structure. The figure was generated using *Molscript* and *Raster3D* (Kraulis, 1991; Merrit & Murphy, 1994). The calcium ion is shown as a cyan-coloured ball and the N- and C-termini are labelled. The major secondary structures are numbered according to Fig. 2.

and 131 are well conserved among many of the family III CBDs. Residue 127, which interacts *via* its main chain, varies; it is Asp in CipC and E4 and is Thr in CipA. This implies that Ca^{2+}-binding sites will also be found in other family III CBDs.

CipC also contains a Zn^{2+}-binding site that was not seen in either the CipA or the E4 CBD. During refinement and examination of the difference Fourier maps, two peaks of 14σ were found at the protein–protein interface. One of these peaks occupies a special position on the crystallographic twofold axis, thereby forming a three-metal-atom cluster coordinated *via* water molecules. The coordination is *via* His93, Gln49 and the symmetry-related residues across the crystallographic dyad. These observations are consistent with the zinc acetate used in crystallization and are likely to represent a crystallization artifact, although Zn^{2+} binding has been reported for the N-terminal family IV CBD from *Cellulomonas fimi* CenC (Johnson *et al.*, 1998).

3.4. Cellulose binding

As was hypothesized from the previously reported structure of the family IIIa cellulosomal CBD from CipA of *C. thermocellum* (Tormo *et al.*, 1996), the smooth β-sheet surface comprising strands 1, 2, 4 and 7 appears to interact with three adjacent chains of the cellulose (Fig. 6). Projecting from that surface is a linear array of aromatic residues which are conserved across family IIIa and IIIb CBDs (Fig. 2) and which are thought to interact with the rings of the cellulose substrate primarily *via* aromatic stacking. The residues involved in the stacking interactions are Trp123, His61 and Tyr70, all of which form a contact with the glucose rings of the substrate by aromatic stacking. Arg117 and Asp60 form a salt bridge which presents both the guanidium group of the arginine and a closed hydrogen-bonded ring, which can interact with the cellulose sugar. In addition, the CipC CBD includes an unconserved residue (Met65), which is also aligned in this planar strip and is positioned in such a way as to allow its S atom to interact with the cellulose.

In addition to the aromatic stacking interactions *via* the conserved planar strip, the CBD presents from the flat β-sheet a large number of uncharged polar residues, which can also be considered to anchor the CBD to its cellulose substrate *via* interactions with the sugar rings and sugar hydroxyl O atoms. These residues are arranged in two strips running parallel to the planar aromatic strip. In the strip adjacent to the planar strip are the totally conserved uncharged polar residues Asn125, Asn20, Ser21 and Gln115. Also aligned in this strip are the non-conserved residues Tyr23, Arg25, Glu113 and Ser111. These eight residues together form a strip that runs the entire length of the molecule. The third strip comprises residues that are partially conserved: Ser18, Ser19, Asn11, Tyr140, Lys27 and Thr29.

As was discussed by Tormo *et al.* (1996), the periodicity of the residues in the planar strip and the adjacent anchoring strips appears to complement the consecutive positions of the glucose rings in the cellulose fibres; however, in crystalline cellulose this type of fully exposed sugar ring would be seen only in surface defects or at the obtuse corners of the cellulose crystals at the slightly exposed (010) surface. It is anticipated, however, that as the crystalline surface of cellulose is degraded, these obtuse corners would be most accessible to attack, thereby increasing the available surface area.

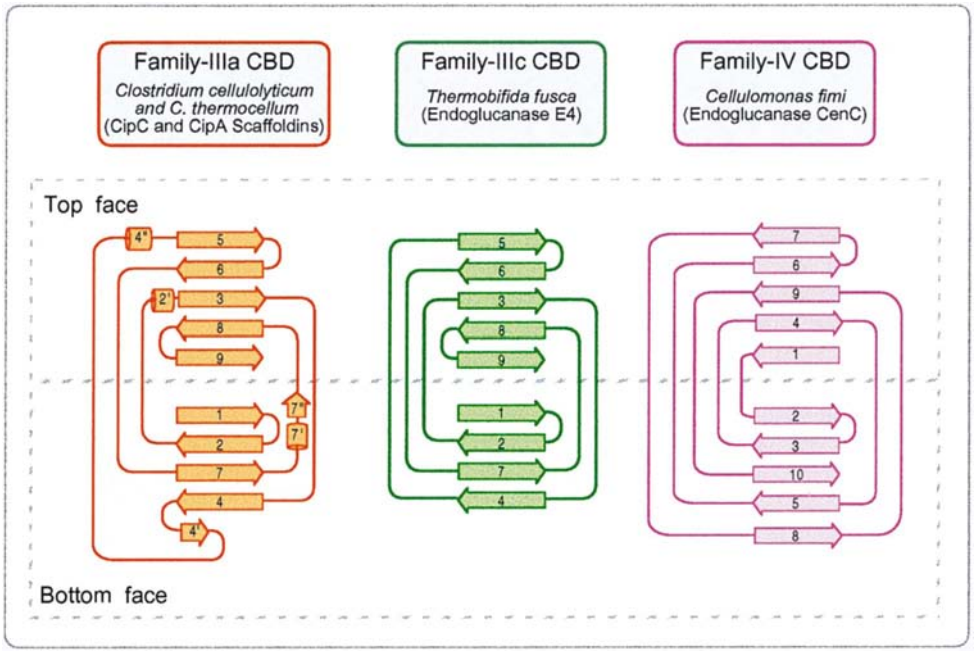

Figure 5
Topology diagram of the family IIIa CipC CBD (orange) compared with those of other bacterial CBD structures. The structural elements are labelled according to the respective published report: family IIIc (Sakon *et al.*, 1997) and family IV CBD (Johnson, Joshi *et al.*, 1996).

3.5. Comparison of the two scaffoldin family IIIa CBD structures

The CipC CBD from *C. cellulolyticum* has a 43% identity with the CipA CBD from *C. thermocellum*. The structure of the CipC CBD is very similar to that of CipA (Tormo *et al.*, 1996), with an RMS deviation of 1.38 Å for the fit of 152 corresponding equivalent C^{α} atoms (Fig. 7*a*). Comparison of the crystal structures revealed a number of minor differences in the C^{α} backbone, some of which were produced by the small insertion and deletion (Figs. 7*a* and 8*b*). The two-residue insertion with respect to the family IIIa CipA, which contains Ser101 and Asp102, is located in the loop between strands 6 and 7. Interestingly, residues at these positions are also seen in other

family IIIa and IIIb CBDs, but are not found in the family IIIc CBDs. The largest difference between the two known family IIIa structures is a 7.93 Å shift of the extended loop between strands 4 and 4′, centred on residue Asn69 of the CipC structure (Fig. 7*a*, arrow). This shift also brings Met65 into alignment with the proposed cellulose-binding residues of the planar strip.

The cellulose-binding face of the CipC CBD is more highly charged than that of CipA (compare Figs. 7*c* and 7*e*) with the inclusion of Arg25 and Glu113 (Gln and Gln in CipA). On the opposite face of the molecule, there are differences in the residues which line the shallow groove; differences which also lead to a significantly different pattern of charge distribution in this groove (Figs. 7*d* and 7*f*). These include the substitution of Lys42, Lys76 and Lys82 for Asn, Thr and Val in the respective positions in CipA. There are a number of residues in the CipC CBD shallow-groove region that are totally conserved in all of the family III CBDs. These include Arg44, Tyr46, Tyr94, Glu96, Leu154 and Pro160. Nearly conserved in this groove are Trp132 and Ser147 (Tyr and Thr, respectively, in CipA) and Trp143. As in CipA, Arg44 is sandwiched between Tyr94 and Phe149; Pro160, the last residue of all known family III CBDs, is packed tightly against Trp143.

3.6. Comparison of the family IIIa *versus* family IIIc CBD structures

The CipC scaffoldin CBD from *C. cellulolyticum* has a 21% identity with the family IIIc CBD from cellulase E4 of *T. fusca*. The structures can be superimposed with an RMS deviation of 2.34 Å for the 131 corresponding C^{α} atoms of the E4 CBD (Fig. 7*b*). Both molecules share a similar jelly-roll fold of a nine-stranded antiparallel β-sandwich as well as the Ca^{2+}-binding site. The CipC CBD, however, lacks the disulfide bond seen in the E4 CBD although, on the basis of sequence alignment, the disulfide bond is not strictly conserved in the family IIIc CBDs. There are several other differences in the superposition of these two subtypes of CBD. One striking difference is the marked truncation of the loop between strands 4 and 4′ (residues 69–76) which in the CipC CBD is adjacent to the cellulose-binding face.

Figure 6
View of the flat cellulose-binding face of CipC CBD. Residues proposed to interact with the three adjacent cellulose chains are shown in blue. The planar-strip stacking residues at top (dark blue) from left to right: Trp123, the Arg117–Asp60 salt bridge, His61, Tyr70 as well as the non-conserved Met65. The adjacent anchoring-strip residues (medium blue) include (from left to right) Asn125, Asn20, Ser21, Tyr23, Arg25, Gln115, Glu113 and Ser111. The second anchoring strip (light blue) at bottom includes (from left to right) residues Ser18, Ser19, Asn11, Tyr140, Lys27 and Thr29.

These two CBDs appear to play a different role in cellulolysis. As described above, the family IIIa CBD (as well as the family IIIb CBD) exhibits strong binding to crystalline cellulose, whereas the family IIIc CBD plays a helper role in catalysis by feeding single chains of the substrate into the catalytic site.

It is interesting to note the sequence (and presumed structural) similarity of the shallow groove exhibited by all the family III CBDs. In this area of the molecule, the surface residues that occupy the groove are quite well conserved and only relatively minor differences can be noted (Fig. 2). It is especially interesting that the groove region is retained in the family IIIc 'helper' CBDs. This implies that the function of the groove would be similar despite the dissimilarity in their presumed primary function, *i.e.* cellulose binding. In other words, the role played by this conserved groove is independent of the character of the substrate-binding role (*i.e.* single chain *versus* surface binding) which characterizes the given CBD subfamily. Since the groove region together with its conserved residues is essentially identical in both the scaffoldin CBDs and the free cellulase CBDs from families IIIb and IIIc, we can deduce that its function would be similar irrespective of the nature of its neighbouring modules.

3.7. Comparison of the family IIIa *versus* family IV CBD structures

The NMR structure of the family IV CBD, the N-terminal CBD from *Cellulomonas fimi* CenC (Johnson, Joshi *et al.*, 1996), is quite different overall from that observed in the family III CBDs, *e.g.* from CipC and CipA. Automated structure-based similarity search yielded no similarity between the family III and family IV CBDs. Manual inspection, however, revealed an interesting topological resemblance in the groove region of the two structures. Although both family III and family IV CBDs exhibit an all-β jelly-roll topology with a groove or cleft on their 'top' faces, the family IV CBD lacks the flat 'bottom' cellulose-binding surface and planar strip residues. In the case of the family IV CBDs, the cleft on the top of the molecule is thought to be the cellulose-binding site. The family IV CBD fails to bind crystalline cellulose but binds to amorphous cellulose or soluble cellodextrins (Johnson, Tomme *et al.*, 1996; Tomme *et al.*, 1996). Since the family III CBDs also recognize amorphous cellulose, it was of interest to compare the nature of its groove with the cleft of the family IV CBD in order to see if any similarity could be discerned.

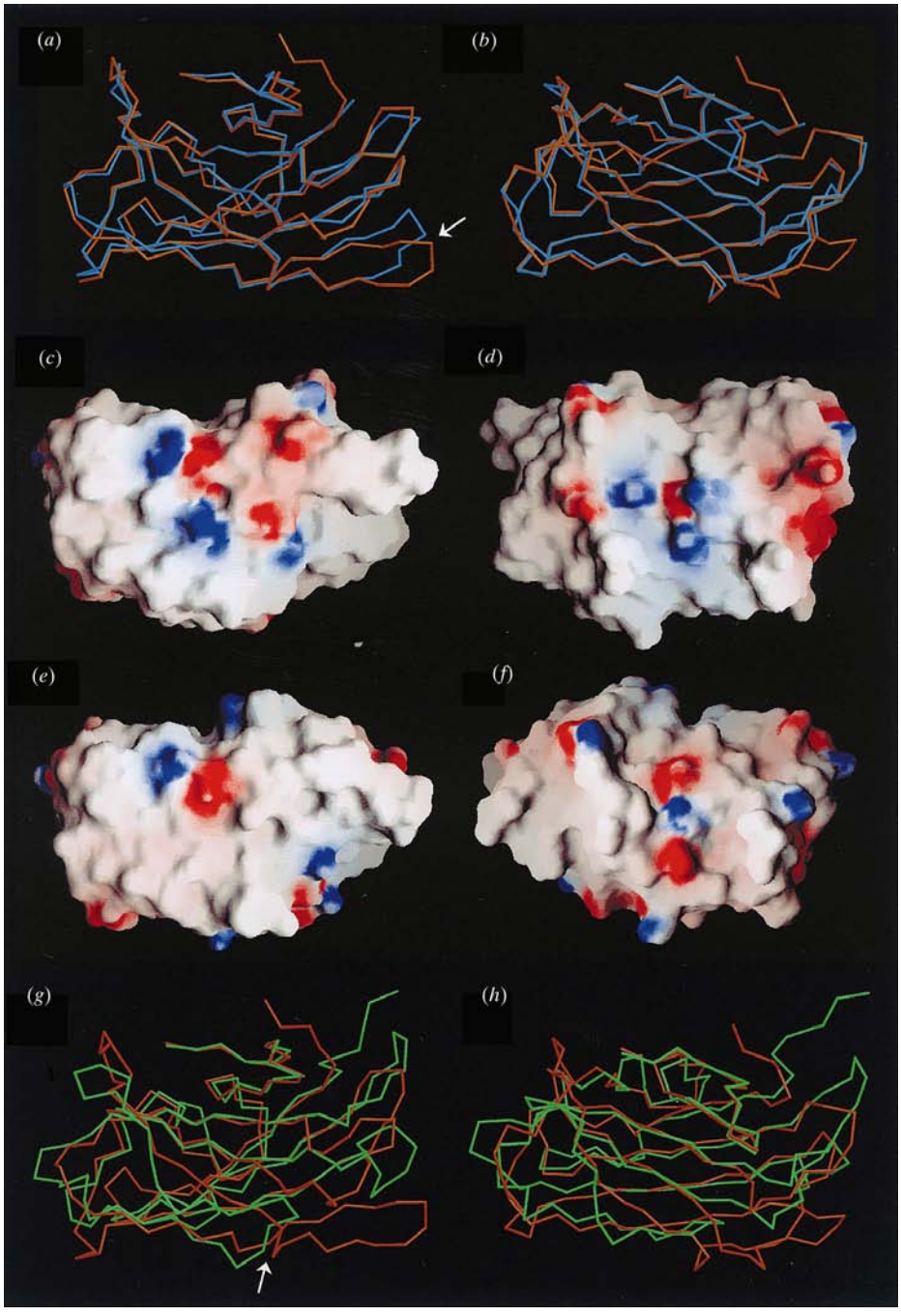

Figure 7
Structural comparisons of the CipC CBD with other family III CBDs. (*a*) Superposition (C^{α} trace) of the two known scaffoldin family IIIa CBD structures: CipC CBD from *C. cellulolyticum* (orange) *versus* the *C. thermocellum* CipA CBD (blue). The arrow highlights the position of the pronounced displacement of Asn69. (*b*) 90° rotation of the superimposed molecules in (*a*) about the horizontal axis, showing a bird's-eye view of the shallow groove. (*c*) Molecular surface of CipC CBD, colour-coded by electrostatic potential: red, negative; blue, positive. Shown in this view is the cellulose-binding face. (*d*) 180° rotation of the CipC CBD (*c*), viewing the shallow-groove face. Analogous views of electrostatic potential of CipA CBD are shown in (*e*) and (*f*), respectively. (*c*), (*d*), (*e*) and (*f*) were rendered with the program *GRASP* (Nicholls *et al.*, 1993). Note the relatively charged nature of the cellulose-binding face of the CipC CBD *versus* that of CipA. (*g*) Superposition of the scaffoldin CipC CBD from *C. cellulolyticum* (orange) *versus* the free enzyme family IIIc CBD from the *T. fusca* E4 cellulase (green). The arrow highlights the position of the truncated part of the E4 molecule relative to the CipC CBD. (*h*) 90° rotation of the superimposed molecules in (*g*) about the horizontal axis. The respective orientation of the CipC CBD molecule in views (*a*) and (*g*) and in views (*b*) and (*h*) are identical.

By superimposing the C^{α} positions in the cleft of CenC and the shallow groove on the upper face of CipC CBD, a fit with an RMS deviation of 2.5 Å was obtained for a limited number of 22 residues. The best fit of the CenC cleft with the CipC shallow grooves is obtained with the directionality of the β-strands reversed (see Figs. 8*a* and 8*b*). Most remarkable about this superposition was that it revealed that the seemingly similar topologies of these two molecules are actually inverted (see Fig. 5). Moreover, while the C^{α} positions are not very different, the character of the residues that line these grooves is markedly different. In this context, the groove of the CipC CBD (and all the other family III CBDs) contains long, polar or charged groups such as Arg, Lys and large aromatic residues such as Tyr (see Figs. 8*c* and 8*e*). In contrast, the cleft of the CenC CBD displays short hydrophobic residues such as Val and Leu (see Figs. 8*d* and 8*f*).

There can be no doubt that any proposed function of the CipC CBD shallow groove or its mode of interaction with cellulose must be quite different from that of the family IV CBDs. One intriguing possibility is that the shallow groove may interact with peptide segments such as the Pro/Thr-containing linker segments of the scaffoldin and other cellulosomal components. In this regard, the terminal proline residue of the CipA CBD, which characteristically stacks against Trp143, was deleted by genetic means and the resultant protein was expressed, purified and crystallized (Jose Tormo *et al.*, unpublished results). The ensuing structure showed that the deleted proline residue was replaced by a different proline from the N-terminal linker sequence of a neighboring molecule, which assumed its place in the crystal. The propensity of the shallow groove to bind to proline residues might reflect its involvement in processing conformational changes in the quaternary structure, which may accompany the binding of the cellulosome (and

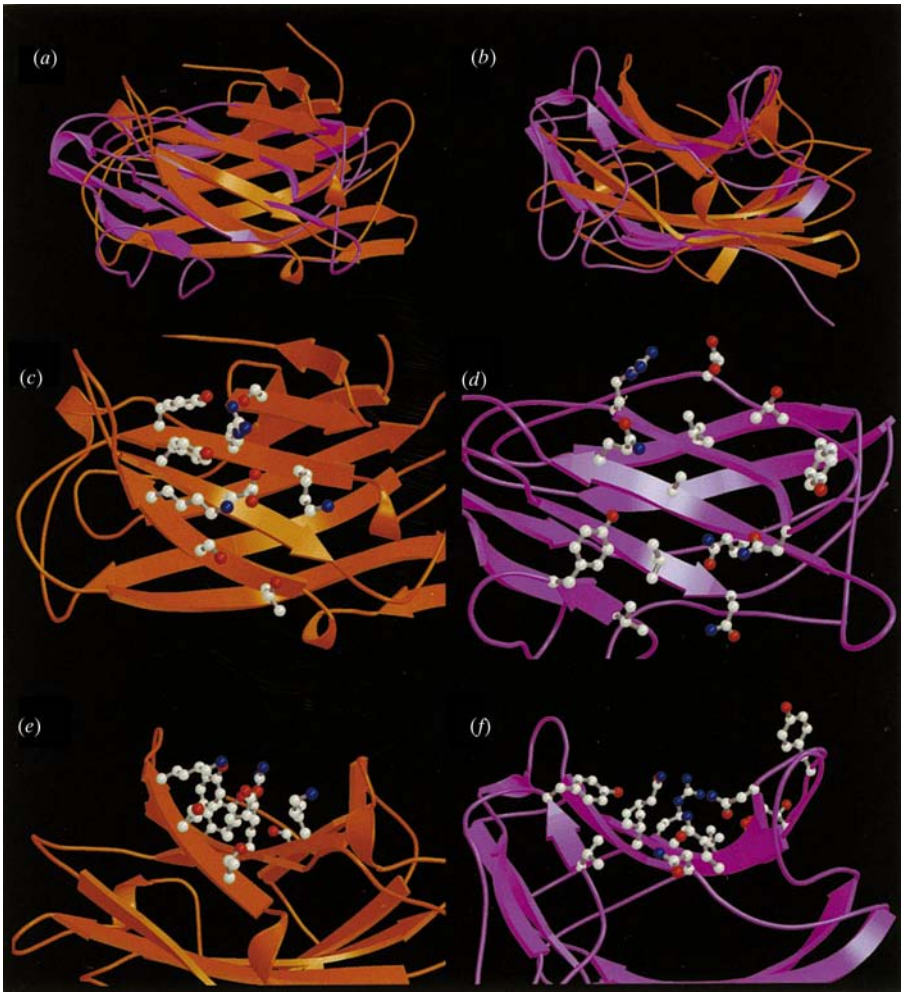

Figure 8
Structural comparison of the groove region of CipC CBD (orange) and the cleft of the family IV CBD (violet) from *Cellulomonas fimi* β-1,4-glucanase CenC. (*a*) Bird's-eye view of the superimposed ribbon diagrams of the groove regions of the two molecules. Note that the direction of the respective β-strands in the two molecules is reversed. (*b*) 90° rotation of the superimposed molecules in (*a*) about the horizontal axis. (*c*) Detailed view of the CipC CBD shallow groove as shown in (*a*) with ball-and-stick display of critical residues. (*d*) Corresponding view of the critical residues of the CenC CBD cleft. (*e*) 'Side view' of the CipC CBD groove packed mainly with large charged and aromatic residues. (*f*) 90° rotation of the CenC CBD (*d*) showing a side view of the cleft filled with relatively short hydrophobic residues.

perhaps free cellulases) to insoluble cellulosic substrates [Morag (Morgenstern) *et al.*, 1991; Morag *et al.*, 1992]. In this context, we are currently exploring the possible co-crystallization of the CBD together with soluble cellulosome-derived linker segments.

The authors appreciate the assistance of Dr Haim Rozenberg in the preparation of the figures. This work was supported by a contract from the European Commission (Fourth Framework, Biotechnology Programme, BIO4-97-2303) and grants from the Israel Science Foundation (administered by the Israel Academy of Sciences and Humanities, Jerusalem) are gratefully acknowledged. Additional support was provided by the Otto Meyerhof Center for Biotechnology, established by the Minerva Foundation (Munich, Germany), and funds from the Technion-Niedersachsen Cooperation (Hannover, Germany).

References

Bayer, E. A., Chanzy, H., Lamed, R. & Shoham, Y. (1998). *Curr. Opin. Struct. Biol.* **8**, 548–557.
Bayer, E. A., Morag, E., Shoham, Y., Tormo, J. & Lamed, R. (1996). *Bacterial Adhesion: Molecular and Ecological Diversity*, edited by M. Fletcher, pp. 155–182. New York: Wiley-Liss, Inc.
Bayer, E. A., Shimon, L. J. W., Lamed, R. & Shoham, Y. (1998). *J. Struct. Biol.* **124**, 221–234.
Brun, E., Moriaud, F., Gans, P., Blackledge, M., Barras, F. & Marion, D. (1997). *Biochemistry*, **36**, 16074–16086.
Brunger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Coutinho, P. M. & Henrissat, B. (1999). *Genetics, Biochemistry and Ecology of Cellulose Degradation*, edited by K. Ohmiya, K. Hayashi, K. Sakka, Y. Kobayashi, S. Karita & T. Kimura, pp. 15–23. Tokyo: Uni Publishers Co.
Davies, G. & Henrissat, B. (1995). *Structure*, **3**, 853–859.
Gal, L., Gaudin, C., Belaich, A., Pagès, S., Tardif, C. & Belaich, J.-P. (1997). *J. Bacteriol.* **179**, 6595–6601.
Gerngross, U. T., Romaniec, M. P. M., Kobayashi, T., Huskisson, N. S. & Demain, A. L. (1993). *Mol. Microbiol.* **8**, 325–334.
Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. J. & Warren, R. A. J. (1991). *Microbiol. Rev.* **55**, 303–315.
Henrissat, B. & Davies, G. (1997). *Curr. Opin. Struct. Biol.* **7**, 637–644.
Irwin, D., Shin, D.-H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A. & Wilson, D. B. (1998). *J. Bacteriol.* **180**, 1709–1714.
Johnson, P., Joshi, M., Tomme, P., Kilburn, D. & McIntosh, L. (1996). *Biochemistry*, **35**, 14381–14394.
Johnson, P. E., Creagh, A. L., Brun, E., Joe, K., Tomme, P., Haynes, C. A. & McIntosh, L. (1998). *Biochemistry*, **37**, 12772–12781.
Johnson, P. E., Tomme, P., Joshi, M. D. & McIntosh, L. P. (1996). *Biochemistry*, **35**, 13895–13906.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kakiuchi, M., Isui, A., Suzuki, K., Fujino, T., Fujino, E., Kimura, T., Karita, S., Sakka, K. & Ohmiya, K. (1998). *J. Bacteriol.* **180**, 4303–4308.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J. & Gronenborn, A. M. (1989). *Biochemistry*, **28**, 7241–7257.
Merrit, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Morag, E., Bayer, E. A. & Lamed, R. (1992). *Appl. Biochem. Biotechnol.* **33**, 205–217.
Morag (Morgenstern), E., Bayer, E. A. & Lamed, R. (1991). *Appl. Biochem. Biotechnol.* **30**, 129–136.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nicholls, A., Bharadwaj, R. & Honig, B. (1993). *Biophys. J.* **64**, 166–170.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Pagès, S., Belaich, A., Fierobe, H.-P., Tardif, C., Gaudin, C. & Belaich, J.-P. (1999). *J. Bacteriol.* **181**, 1801–1810.
Pagès, S., Gal, L., Belaich, A., Gaudin, C., Tardif, C. & Belaich, J.-P. (1997). *J. Bacteriol.* **179**, 2810–2816.
Poole, D. M., Morag, E., Lamed, R., Bayer, E. A., Hazlewood, G. P. & Gilbert, H. J. (1992). *FEMS Microbiol. Lett.* **99**, 181–186.
Sakon, J., Irwin, D., Wilson, D. B. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 810–818.
Shoseyov, O., Takagi, M., Goldstein, M. A. & Doi, R. H. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 3483–3487.
Tomme, P., Creagh, A. L., Kilburn, D. G. & Haynes, C. A. (1996). *Biochemistry*, **35**, 13885–13894.
Tomme, P., Warren, R. A. J., Miller, R. C., Kilburn, D. G. & Gilkes, N. R. (1995). *Enzymatic Degradation of Insoluble Polysaccharides*, edited by J. M. Saddler & M. H. Penner, pp. 142–161. Washington, DC: American Chemical Society.
Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y. & Steitz, T. A. (1996). *EMBO J.* **15**, 5739–5751.
Xu, G.-Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E. & Harvey, T. S. (1995). *Biochemistry*, **34**, 6993–7009.