Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Scaffoldin-borne family 3b carbohydrate-binding module from the cellulosome of *Bacteroides cellulosolvens*: structural diversity and significance of calcium for carbohydrate binding

Oren Yaniv,[a,b] Linda J. W. Shimon,[c] Edward A. Bayer,[d] Raphael Lamed[a,b] and Felix Frolow[a,b]*

[a]Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv 69978, Israel, [b]Daniella Rich Institute for Structural Biology, Tel Aviv University, Tel Aviv 69978, Israel, [c]Department of Chemical Research Support, The Weizmann Institute of Science, Rehovot 76100, Israel, and [d]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

Correspondence e-mail: mbfrolow@post.tau.ac.il

Received 15 January 2011
Accepted 27 March 2011

PDB Reference: carbohydrate-binding module, 2xbt.

The potent cellulose-binding modules of cellulosomal scaffoldin subunits belong to the greater family of carbohydrate-binding modules (CBMs). They have generally been classified as belonging to family 3a on the basis of sequence similarity. They form nine-stranded β-sandwich structures with jelly-roll topology. The members of this family possess on their surface a planar array of aromatic amino-acid residues (known as the linear strip) that form stacking interactions with the glucose rings of cellulose chains and have a conserved Ca^{2+}-binding site. Intriguingly, the CBM3 from scaffoldin A (ScaA) of *Bacteroides cellulosolvens* exhibits alterations in sequence that make it more similar to the CBMs of free cellulolytic enzymes, which are classified into CBM family 3b. X-ray structural analysis was undertaken in order to examine the structural consequences of the sequence changes and the consequent family affiliation. The CBM3 crystallized in space group $I4_122$ with one molecule in the asymmetric unit, yielding diffraction to a resolution of 1.83 Å using X-ray synchrotron radiation. Compared with the known structures of other scaffoldin-borne CBMs, a sequence insertion and deletion appear to compensate for each other as both contained an aromatic residue that is capable of contributing to cellulose binding; hence, even though there are alterations in the composition and localization of the aromatic residues in the linear strip its binding ability was not compromised. Interestingly, no Ca^{2+} ions were detected in the conserved calcium-binding site, although the module was properly folded; this suggests that the structural role of Ca^{2+} is less important than originally supposed. These observations indicate that despite their conserved function the scaffoldin-borne CBMs are more diverse in their sequences and structures than previously assumed.

1. Introduction

Cellulosomes are multi-enzyme complexes designed for efficient degradation of plant cell-wall polysaccharides in general and cellulose in particular (Bayer *et al.*, 2004, 2008; Doi & Kosugi, 2004; Fontes & Gilbert, 2010). Cellulosomes, which were first described in the anaerobic thermophile *Clostridium thermocellum* (Bayer *et al.*, 1983; Lamed *et al.*, 1983), consist of a central scaffoldin subunit that incorporates the various enzymes into the complex, anchors the complex onto the cell surface of the bacterium and targets the complex to the substrate (Gerngross *et al.*, 1993).

The high-affinity cohesin–dockerin interaction is responsible for cellulosome architecture. For enzyme integration, the primary scaffoldin molecule carries multiple type I cohesin

modules that selectively bind to a complementary type I dockerin module located on the enzyme. An alternative type of interaction occurs in *C. thermocellum* between type II cohesin and dockerin modules located on anchoring and primary scaffoldins, respectively, and is involved in the cell-surface anchoring function in this bacterium (Leibovitz & Béguin, 1996; Salamitou *et al.*, 1992). Substrate targeting is mediated by a cellulose-specific carbohydrate-binding module (CBM) of the primary scaffoldin (Poole *et al.*, 1992) and attachment to the bacterial cell surface is mediated by an S-layer homology (SLH) module located on the anchoring scaffoldins (Lemaire *et al.*, 1995).

Bacteroides cellulosolvens is a cellulosome-producing anaerobic mesophilic cellulolytic bacterium that was originally isolated from sewage sludge. The *B. cellulosolvens* cellulosome comprises two particularly large scaffoldins: (i) a primary (enzyme-integrating) scaffoldin (ScaA) that contains 11 cohesins (Ding *et al.*, 2000), a family 3 carbohydrate-binding module (CBM3) and a C-terminal dockerin, and (ii) a cell-surface anchoring scaffoldin (ScaB) that bears ten cohesins, an X domain of unknown function and an SLH module (Xu *et al.*, 2004). The supramolecular organization of the cellulosome on the surface of the bacterium is presented schematically in Fig. 1. Surprisingly, in this particular bacterium the types of cohesins and dockerins that reside on the scaffoldins and enzymes are switched compared with those of the *C. thermocellum* standard. Thus, 11 type II (instead of type I) dockerin-borne enzymes can ostensibly be incorporated into the ScaA polypeptide by virtue of its resident type II cohesin modules. Likewise, the ten ScaB type I (instead of type II) cohesin modules can conceivably accommodate an equivalent number of type I dockerin-containing ScaA subunits, together with their complement of enzyme molecules. The cellulosome apparatus of *B. cellulosolvens* would thus comprise a total of 110 enzyme molecules. The complete cellulosome complex (~11.5 MDa in size, assuming that all cohesins are occupied by enzymes) would presumably be attached to the cell surface *via* the ScaB SLH module (Xu *et al.*, 2004) and to the cellulosic substrate *via* the ScaA CBM3b module. The *B. cellulosolvens* cellulase system shows all of the features of a powerful and dominant cellulosome assembly, including intimate association with the cell surface, a variety of different plant cell-wall-degrading enzymes (cellulases and hemicellulases), substrate targeting, enzyme amplification and enzyme-proximity effects (Xu *et al.*, 2004).

The process of cellulose degradation starts with the binding of the cellulolytic enzymes or of the entire organism to the cellulosic substrate (Bayer, Morag *et al.*, 1998). A separate module, the CBM, mediates this step. CBMs can serve as targeting agents for catalytic modules of free enzymes (Boraston *et al.*, 2004; Tomme *et al.*, 1995) or act as a separate targeting module as part of the noncatalytic scaffoldin subunit of the cellulosome (Bayer, Shimon *et al.*, 1998).

The CBM3 molecules comprise ~150 amino-acid residues. In addition to their enzyme-targeting or cell-targeting roles, CBM3 molecules have been identified in many hydrolytic enzymes that contain the family 9 catalytic module (GH9) and serve an accessory role in the enzymatic action of this family of glycoside hydrolases. Based on sequence similarity, three major subgroups (3a, 3b and 3c) have been identified in the CBM3 family (Bayer, Shimon *et al.*, 1998). Subfamilies 3a and 3b are more closely related to each other than to family 3c. In biochemical studies, members of the 3a and 3b subgroups bind strongly to the surface of microcrystalline cellulose and have also been suggested to promote the cellulolytic reaction by concentrating and properly orienting the enzyme near the cellulose surface (Gilad *et al.*, 2003; Tormo *et al.*, 1996). Recently, however, the common assumption that all family 3a and 3b CBMs function primarily in a cellulose-binding capacity was called into question by the discovery that certain CBM3b proteins (designated CBM3b′) lack some of the normally conserved residues of the cellulose-binding planar strip that had previously been identified in the 'classic' cellulose-binding CBMs (Petkun *et al.*, 2010).

The CBMs of families 3a and 3b were proposed in a previous study to be distinguished by the nature of the parent protein, with the family 3a CBMs (CBM3a) being components of cellulo-

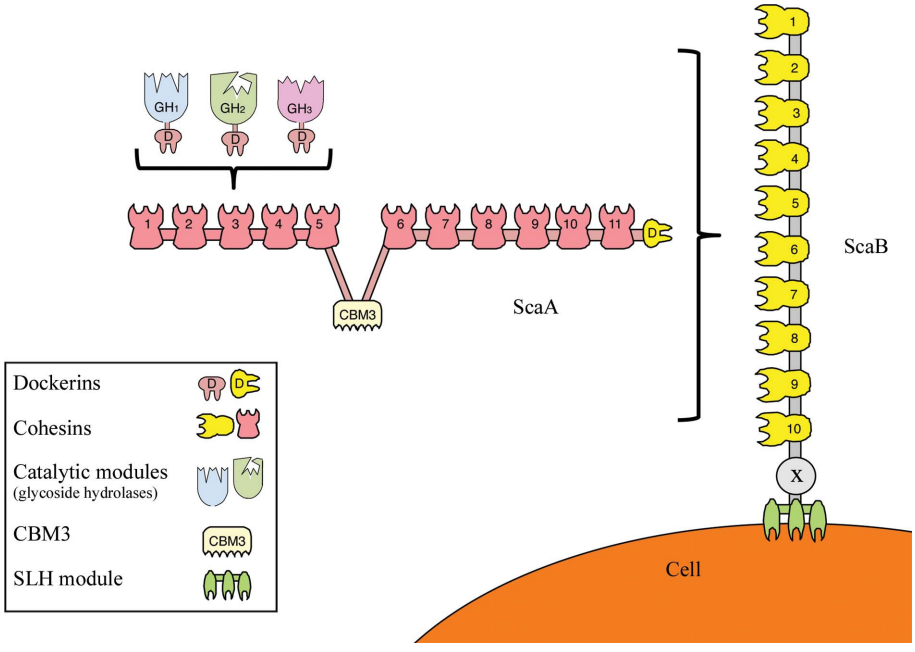


Figure 1
Schematic representation of the proposed cell-surface disposition of the cellulosomal components of *B. cellulosolvens*. Dockerin-containing enzymes (glycoside hydrolases; GHs) are incorporated into the ScaA scaffoldin owing to the interaction of their resident type II dockerin modules with type II ScaA cohesins. ScaA in turn, together with its complement of enzymes, is attached in multiple copies to type I ScaB cohesins. The cellulosome complex is attached to its cellulosic substrates *via* the ScaA CBM3 and to the cell surface *via* the ScaB SLH module.

somal scaffoldin subunits and the family 3b CBMs (CBM3b) being the targeting agents for free noncellulosomal enzymes (Bayer, Morag *et al.*, 1998). CBM3a and CBM3b possess similar primary sequences, except for a few known examples of CBM3a that contain an extra loop that includes an additional cellulose-binding aromatic residue (Jindou *et al.*, 2006). All CBM crystal structures that belong to family 3 (Tormo *et al.*, 1996; Shimon *et al.*, 2000; Sakon *et al.*, 1997; Petkun *et al.*, 2010) harbour a Ca^{2+}-binding site usually composed of five conserved residues. The five residues are located in loops connecting the 'top face' and the 'bottom face' of the module, so the effect of Ca^{2+} binding is probably structural, with the two loops being secured and consequently tethering the upper and lower β-sheets together (Petkun *et al.*, 2010).

The *B. cellulosolvens* ScaA CBM3b is the second scaffoldin-borne CBM after that of CipV from *Acetivibrio cellulolyticus* (Ding *et al.*, 1999) to be classified into family 3b instead of family 3a, supporting the notion that the scaffoldin CBMs are more diverse than was originally thought. To gain further insight into the nature of this diversity, we undertook an X-ray structural analysis of the *B. cellulosolvens* ScaA CBM3b. The procedure that we followed and the molecular structure that was obtained are described in this report.

2. Materials and methods

2.1. Cloning of the CBM3b from ScaA of *B. cellulosolvens*

A DNA fragment encoding ScaA CBM3b (GenBank accession No. AAG01230.2) was amplified by PCR from *B. cellulosolvens* ATCC 35603 genomic DNA isolated as described by Murray & Thompson (1980) using two specific primers: 5′-CCATGGGCCCTGTACAAGTTAACAGCGAC-3′ and 5′-GAATTCTTATGGTGGCGTACCATATACCAA-AG-3′. The PCR products were purified, cleaved with the restriction enzymes *Nco*I and *Eco*RI and inserted into the pET-28a(+) expression vector (Novagen, Madison, Wisconsin, USA) without any additional tag, yielding pET-ScaA-CBM3b.

2.2. Protein expression and purification

Escherichia coli BL21 (DE3) cells harbouring pET-ScaA-CBM3b were aerated at 310 K in 4 l Terrific Broth supplemented with 25 µg ml^{-1} kanamycin. After 2.5 h, when the culture had reached an A_{600} of 0.6, 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce gene expression and cultivation continued at 310 K for an additional 12 h. The cells were then harvested by centrifugation (5000*g* for 10 min) at 277 K and subsequently resuspended in 50 m*M* phosphate buffer pH 8.0 containing 300 m*M* NaCl at a ratio of 1 g wet pellet per 4 ml buffer solution. A few micrograms of DNase powder were added prior to the sonication procedure. The suspension was kept on ice during sonication, after which it was centrifuged (20 000*g* at 277 K for 30 min) and the supernatant fluids were collected. The recombinant *B. cellulosolvens* CBM3b was purified by cellulose-affinity purification using crystalline cellulose (microcrystalline cellulose, type 50; Sigma, St Louis, Missouri, USA). The supernatant fluid was incubated with cellulose for 1 h with gentle stirring at 277 K. The cellulose pellet was recovered by centrifugation and washed three times with 1 *M* sodium bicarbonate and three times with 200 m*M* sodium bicarbonate. The resulting cellulose pellet was resuspended in 1%(*w*/*v*) aqueous diethylamine solution to elute *B. cellulosolvens* CBM3b protein and the cellulose powder was removed by centrifugation. The eluate was neutralized to pH 7.5 using 1 *M* Tris–NaCl buffer pH 7. Protein purity was evaluated by SDS–PAGE (15%) stained with Coomassie Brilliant Blue. The protein was dialyzed for 12 h with gentle stirring at 277 K against a buffer consisting of 150 m*M* Tris–HCl pH 7.5, 300 m*M* NaCl and 0.05% sodium azide. The protein was concentrated to 20 mg ml^{-1} using Centriprep YM-3 centrifugal filter devices (Amicon Bioseparation, Millipore, Billerica, Massachusetts, USA). The protein concentration was determined by measuring the UV absorbance at 280 nm.

2.3. Crystallization and X-ray diffraction

The protein sample was screened for crystallization conditions by the microbatch method (Chayen *et al.*, 1990) using 288 pre-formulated crystallization solutions from Hampton Research HT screens (SaltRx, Index HT and Crystal Screen HT; http://www.hamptonresearch.com/) and 96 conditions of the Wizard I and II sparse-matrix crystallization screens from Emerald BioSystems (http://www.emeraldbiosystems.com/). Samples were dispensed using an Oryx-6 crystallization robot from Douglas Instruments (http://www.douglas.co.uk/). A 1 µl sample of the protein solution together with a 1 µl aliquot of the crystallization condition was dispensed into each well. A mixture of silicone and paraffin oils combined in a 1:1 volume ratio was used to cover the crystallization wells. Crystallization was performed at 293 K in a temperature-controlled room.

The first crystals appeared after 11 d in condition No. 35 of the Index kit [1.0 *M* ammonium sulfate, 0.1 *M* HEPES pH 7.0, 0.5%(*w*/*v*) polyethylene glycol 8000]. Crystals were harvested from the crystallization drop using a MiTeGen MicroMount (http://www.mitegen.com) made of polyimide and transferred for a short time into a cryostabilization solution mimicking the mother liquor supplemented with 18%(*w*/*v*) sucrose, 16%(*w*/*v*) glycerol, 16%(*w*/*v*) ethylene glycol and 4%(*w*/*v*) glucose. For data collection, crystals were mounted on MiTeGen MicroMounts and flash-cooled in a nitrogen stream at a temperature of 100 K produced by an Oxford Cryostream low-temperature generator (Cosier & Glazer, 1986).

Diffraction data were measured on the ID29 beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. An ADSC Q315 detector and X-ray radiation of 0.9763 Å wavelength were used. Diffraction data consisting of 100 images with 1° oscillation per frame were collected. The data were indexed and integrated with *DENZO* and scaled with *SCALEPACK* as implemented in *HKL*-2000 (Otwinowski & Minor, 1997). During the diffraction data-scaling procedure in *SCALEPACK*, Friedel pairs were kept separated in order to preserve the anomalous signal for further use. The crystals diffracted to 1.83 Å resolution and

Table 1
Crystal and diffraction data.

Values in parentheses are for the highest resolution shell.

	Parent crystal	Ca^{2+}-regrown crystal
X-ray source	ESRF, beamline ID29	ESRF, beamline ID14-4
Space group	$I4_122$	$I4_122$
No. of crystals	1	1
No. of frames	100	140
Total rotation angle (°)	100	140
Unit-cell parameters		
$a = b$ (Å)	83.19	83.50
c (Å)	96.14	96.21
V (Å^3)	665386	670914
No. of molecules in asymmetric unit	1	1
Resolution range (Å)	50–1.83 (1.86–1.83)	50–1.98 (2.01–1.98)
Total No. of reflections	131442	129913
Unique reflections	15209	12241
Mosaicity (°)	0.24–0.5	0.42–0.68
Multiplicity	8.64	10.61
Completeness	98.3 (97.0)	100 (99.8)
Average $I/\sigma(I)$	29.2 (1.9)	28.3 (2.5)
R_{merge}†	0.084	0.108
Overall B factor from Wilson plot (Å^2)	25.6	27.7

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\sum_{hkl}$ denotes the sum over all reflections and $\sum_i$ denotes the sum over all equivalent and symmetry-related reflections (Stout & Jensen, 1968).

Table 2
Summary of refinement statistics.

	Parent crystal	Ca^{2+}-regrown crystal
X-ray source	ESRF, beamline ID29	ESRF, beamline ID14-4
Space group	$I4_122$	$I4_122$
Resolution range (Å)	50–1.83	50–1.98
No. of protein atoms	1211	1211
No. of solvent atoms	135	63
No. of ions	0	0
R_{cryst}	0.191	0.182
R_{free}	0.242	0.212
Geometry		
R.m.s.d. bonds (Å)	0.007	0.006
R.m.s.d. angles (°)	1.091	1.039
MolProbity validation		
Ramachandran favoured (%) (goal > 98%)	98.1	99.4
Ramachandran outliers (%) (goal < 0.2%)	0.6	0.0
C^β deviations > 0.25 Å (goal 0)	0	0
Clash score†	12.64	11.81
Rotamer outliers (%) (goal < 1%)	0.8	0.8
Residues with bad bonds (%) (goal < 1%)	0	0
Residues with bad angles (%) (goal < 0.5%)	0	0

† Clash score is the number of serious steric overlaps (>0.4 Å) per 1000 atoms

belonged to the tetragonal space group $I4_122$, with unit-cell parameters $a = b = 83.19$, $c = 96.14$ Å. The calculated Matthews coefficient (Matthews, 1968) of 2.41 Å^3 Da^{-1} gave a solvent content of 49%, corresponding to the presence of one monomer in the asymmetric unit.

Parent *B. cellulosolvens* CBM3b crystals were immersed in a mother-liquor solution supplemented with 1 m*M* $CaCl_2$ at 293 K. Soaking of Ca^{2+} into the crystals resulted in their complete dissolution. After two months, several crystals reappeared. Diffraction data were collected on the ID14-4 beamline at the ESRF. An ADSC Q315 detector and X-ray radiation of 0.9393 Å wavelength were used. The data were processed as described above. The crystals diffracted to 1.98 Å resolution and belonged to the tetragonal space group $I4_122$, with unit-cell parameters $a = b = 83.50$, $c = 96.21$ Å. The calculated Matthews coefficient (Matthews, 1968) of 2.43 Å^3 Da^{-1} gave a solvent content of 49.33%, which corresponded to the presence of one monomer in the asymmetric unit. The statistics of the X-ray data analysis are presented in Table 1. Coordinates and structure-factor amplitudes for the parent protein structure have been deposited in the PDB with code 2xbt.

2.4. X-ray fluorescence measurements

Online X-ray fluorescence (XRF) spectra (Leonard *et al.*, 2009) of *B. cellulosolvens* CBM3b protein solution and crystals were measured on the ID23-1 and ID14-4 beamlines at the ESRF in order to analyze the metal content of the samples (Garcia *et al.*, 2006). Crystalline samples were mounted as described for X-ray diffraction analysis on MiTeGen MicroMounts or on Hampton cryoloops. A drop of protein solution was placed on a circular siliconized glass cover slide (Hampton Research), collected using a MiTeGen MicroMount and then frozen using liquid nitrogen. Crystals were harvested from the crystallization drop using the same procedure as used for the crystals used in the X-ray diffraction experiments.

A silicon drift diode detector coupled with X-flash MMAX signal processing units (Xflash 5010, Bruker AXS Microanalysis, Germany) was used. The incident beam energies were 12.7 keV (ID23-1) and 13.2 keV (ID14-4). Each sample was exposed for 20–60 s, with maximal beam transmission of up to 2.5%. The collected spectrum was analyzed using the *PyMCA* program (Sole *et al.*, 2007).

2.5. Structure determination and refinement

The structure of the *B. cellulosolvens* CBM3b crystal was determined by molecular replacement employing *Phaser* v.2.1.4 (Storoni *et al.*, 2004; McCoy *et al.*, 2005) as implemented in *CCP*4 (Winn *et al.*, 2011), using the atomic coordinates of *C. thermocellum* CBM3a (PDB code 1nbc; Tormo *et al.*, 1996) as a search model. The sequence identity between the model and the target protein was 41%. A clear solution with a Z score of 10.6 and a log-likelihood gain of 97.94 was obtained. The R_{cryst} and R_{free} factors were 0.423 and 0.464, respectively, after the first round of ten cycles of restrained refinement in *REFMAC*5 (Murshudov *et al.*, 1997). The model was subjected to several rounds of restrained refinement of positional and thermal parameters using *REFMAC*5, followed by manual building of the nonconserved side chains of *B. cellulosolvens* CBM3b using *Coot* (Emsley & Cowtan, 2004). The R_{cryst} and R_{free} factors converged to 0.280 and 0.281, respectively.

The model was further refined with *PHENIX* v.1.6-328 (Adams *et al.*, 2002) and assessed using *Coot*. After several rounds of refinement, manual correction of the model and addition of solvent atoms, the final *B. cellulosolvens* CBM3b model was obtained. The R_{cryst} and R_{free} factors for this model were 0.191 and 0.242, respectively. The structure was validated using the *MolProbity* suite (Chen *et al.*, 2010) as implemented in *PHENIX*. Refinement statistics and validation results are summarized in Table 2. The structure of the crystal regrown in the presence of Ca^{2+} was refined using the atomic coordinates of *B. cellulosolvens* CBM3b as an initial model. The refinement and validation processes were conducted similarly to those for the parent crystal.

2.6. Cellulose-binding assay

Proteins (with and without the addition of 20 m*M* $CaCl_2$) were mixed with microcrystalline cellulose (5 g l^{-1}; Merck AG, Darmstadt, Germany) in 50 m*M* sodium citrate buffer pH 6.0 to a final volume of 0.5 ml in 1.5 ml tubes at room temperature. The tube contents were continuously mixed by rotation. After an equilibration time of 2 h, the cellulose and bound proteins were removed by centrifugation (10 000*g* for 10 min). Centrifugation was repeated twice to ensure removal of the cellulose. The amount of unbound protein was determined by measurement of the UV absorbance of the supernatant fluids. The amount of bound protein was calculated as the difference between the initial and the unbound amounts. Each experiment was repeated four times.

Binding to insoluble microcrystalline cellulose was qualitatively assessed as reported by Xu *et al.* (2004). The unbound fraction (supernatant fluids) was set aside and one-third of its volume of SDS-containing sample buffer was added. The pellets (bound fraction) were washed three times with 1 ml citrate buffer and then resuspended in 500 µl citrate buffer with the addition of one-third of its volume of SDS-containing sample buffer. Boiling of the samples for 10 min eluted the bound protein. The binding of the protein to cellulose was evaluated by SDS–PAGE using 10 µl aliquots of the bound and unbound fractions. Each experiment was repeated at least three times.

3. Results and discussion

3.1. Overview

The CBM3b module of the cellulosomal scaffoldin ScaA from *B. cellulosolvens* (GenBank accession No. AAG01230.2), consisting of 158 amino-acid residues, was cloned and expressed. Using the microbatch method to crystallize the protein, we obtained crystals belonging to the tetragonal space group $I4_122$ with one molecule in the asymmetric unit.

We determined the structure of the crystals by molecular replacement using the coordinates of *C. thermocellum* CBM3a as a search model (PDB code 1nbc; Tormo *et al.*, 1996). The final atomic model was refined to R_{cryst} and R_{free} factors of 0.191 and 0.242, respectively, at 1.83 Å resolution. The numbering of the amino-acid residues in the structure corresponds to that of the cloned sequence. The stereochemical quality of the structure was good, with over 98% of the residues in the most favourable regions of the Ramachandran plot (Table 2). The first two residues (Fig. 2) were not observed in the electron-density maps. The structure of the crystals regrown in the presence of Ca^{2+} strongly resembled the

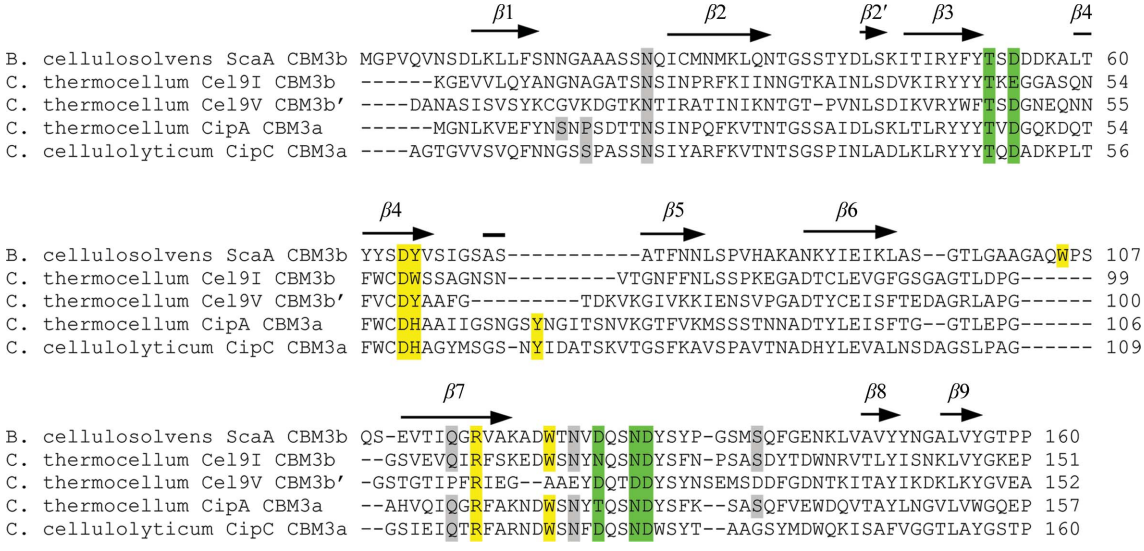
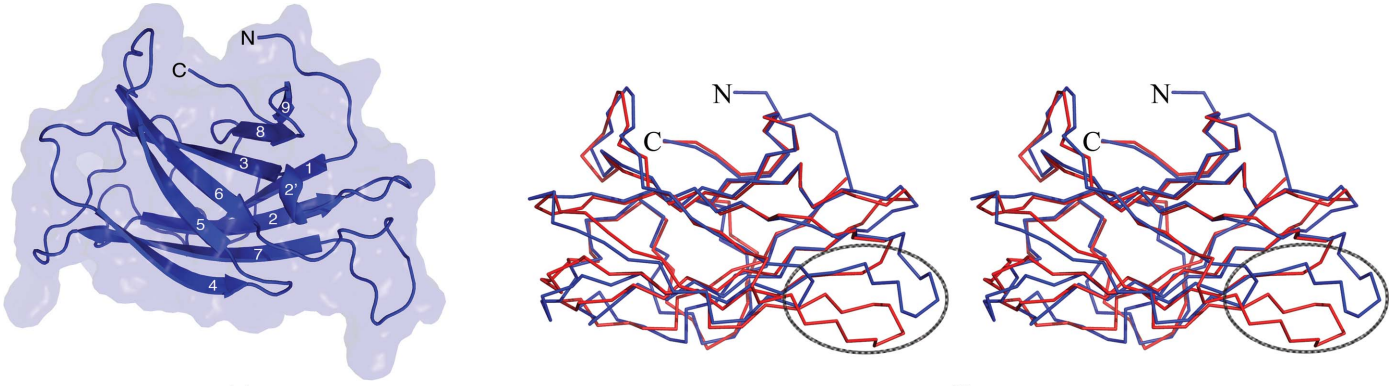


Figure 2
Sequence alignment of the *B. cellulosolvens* ScaA CBM3b with the sequences of other representative family 3 CBMs: *C. thermocellum* Cel9I CBM3b, *C. thermocellum* Cel9V CBM3b′ (PDB code 2wnx; Petkun *et al.*, 2010), *C. thermocellum* CipA CBM3a (PDB code 1nbc; Tormo *et al.*, 1996) and *C. cellulolyticum* CipC CBM3a (PDB code 1g43; Shimon *et al.*, 2000). The alignment was performed using *ClustalW* v.2.0.12 (Larkin *et al.*, 2007). The sequence identities between the *B. cellulosolvens* ScaA CBM3b and the other three molecules vary between 31 and 42%. Secondary-structural elements (β-strands) corresponding to the *B. cellulosolvens* ScaA CBM3b structure are indicated as arrows and labelled. Proposed cellulose-binding and anchoring residues are shown in yellow and grey, respectively. Putative calcium-binding residues are shown in green. The *B. cellulosolvens* CBM3b exhibits an 11-residue gap in the sequence (after residue 72) relative to the CBM3as and a unique six-residue insertion (after residue 101).



(*a*) (*b*)

Figure 3
Overall structure of *B. cellulosolvens* ScaA CBM3b. (*a*) Cartoon representation of the major secondary-structural elements (β-strands), numbered according to Fig. 2. The N- and C-termini are indicated. (*b*) Stereo diagram of the superimposed C^{α} traces of *B. cellulosolvens* ScaA CBM3b (blue) and *C. thermocellum* CipA CBM3a (red; PDB code 1nbc; Tormo *et al.*, 1996). Note the deviation (designated by the ellipse) in the respective conformation of the loop between strands 6 and 7 in the *B. cellulosolvens* CBM3b compared with that of *C. thermocellum* CBM3a.

structure of the native crystals, with a root-mean-square deviation of 0.143 Å on C^{α} atoms.

3.2. Overall structural analysis of *B. cellulosolvens* ScaA CBM3b and comparison with *C. thermocellum* CipA CBM3a

The 158 residues of CBM3b from ScaA of *B. cellulosolvens* formed a nine-stranded antiparallel β-sandwich comprising 63 of the 158 residues (about 40%) of the protein. The remaining residues were in extended loops that probably contribute to the overall stability of the molecule by forming a network of interactions (Fig. 3*a*). The 'bottom' β-sheet of the β-sandwich was flat and was formed by strands 1, 2, 7 and 4, while the 'top' β-sheet comprises antiparallel strands 5, 6, 3, 8 and 9. Both the N-terminus and the C-terminus are located in the 'top' sheet region.

Typical of β-sandwich proteins such as *B. cellulosolvens* ScaA CBM3b is the formation of numerous intramolecular core interactions, mainly hydrophobic, that serve to assist in folding and help to stabilize the protein (Alber *et al.*, 2009). These interactions were formed at the interface between the two β-sheets of the sandwich. To identify the intramolecular interactions that build the hydrophobic core of the protein, we used the *Protein Interfaces, Surfaces and Assemblies* server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.htm; Krissinel & Henrick, 2007) and the *Protein Interactions Calculator* server (http://crick.mbu.iisc.ernet.in/~PIC; Tina *et al.*, 2007). The hydrophobic core area was 1292 $Å^2$ and was formed by 18 hydrophobic residues, 12 aromatic residues and one hydrogen bond. The content of the hydrophobic core of *B. cellulosolvens* ScaA CBM3b is similar to those of *C. thermocellum* CBM3a (Tormo *et al.*, 1996; 19 hydrophobic residues and 11 aromatic residues) and *C. cellulolyticum* CipC CBM3a (Shimon *et al.*, 2000; 16 hydrophobic residues and 12 aromatic residues).

Likewise, the overall structure is similar to the related structures of *C. thermocellum* CipA CBM3a (Tormo *et al.*, 1996) and *C. cellulolyticum* CipC CBM3a (Shimon *et al.*, 2000; Fig. 3*b*), with pairwise root-mean-square deviations (r.m.s.d.s) on C^{α} atoms of about 0.9 Å (using 118 atoms) and 0.8 Å (using 109 atoms) calculated using *PyMOL* (DeLano, 2002).

Sequence comparison between *B. cellulosolvens* CBM3b and *C. thermocellum* CBM3a revealed two major differences. The first was a six-residue insertion after residue 101 in the CBM3b from *B. cellulosolvens* (Fig. 2). This insertion, which is evidently unique to *B. cellulosolvens* and is not found in any other known CBM3, extended the loop between β-strands 6 and 7. The second difference was an 11-residue deletion after residue 72 in *B. cellulosolvens* CBM3b (Fig. 2). This eliminated β-strand 4′ (observed in the CBM3a modules) and shortened the loop between β-strands 4 and 5; it also shortened β-strand 4 by five residues (Fig. 4). In this context, closer inspection of the sequence of CBM3b from *B. cellulosolvens* revealed an interesting substitution at one of the postulated cellulose-binding residues (Fig. 2), in which a tyrosine (Tyr65) replaced

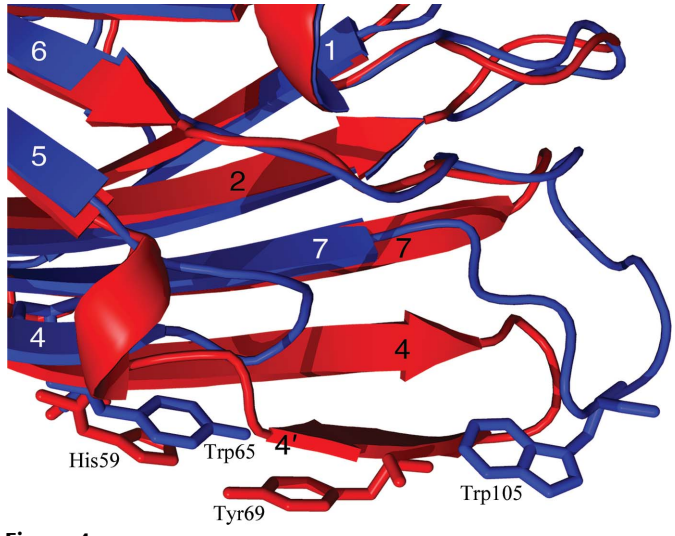


Figure 4
Major structural differences between *B. cellulosolvens* CBM3b and *C. thermocellum* CBM3a. Secondary structures are coloured and numbered according to their origin (*B. cellulosolvens* CBM3b, blue; *C. thermocellum* CBM3a, red). Proposed cellulose-binding residues are shown in stick representation and are coloured according to their origin.

the histidine found in the family 3a CBMs. Nevertheless, the appearance of tyrosine at this position is consistent with its function in cellulose binding as part of a planar aromatic strip, as described by Tormo *et al.* (1996).

3.3. Cellulose binding

According to the accepted hypothesis (Tormo *et al.*, 1996; Lehtiö *et al.*, 2003; Ding *et al.*, 2006), the cellulose-binding CBM3 module interacts with the glucose rings of cellulose *via* a linear array of aromatic residues, which are conserved across family 3a and 3b (but not 3c) CBMs (Shimon *et al.*, 2000). The aromatic residues (Fig. 5) are located on the planar 'bottom' β-sheet comprising β-strands 1, 2, 4 and 7. In the *B. cellulosolvens* CBM3b (found here experimentally to bind cellulose), the aromatic residues thought to be involved in the stacking interactions with the glucose rings are Tyr65 (replacing His59 of the *C. thermocellum* CipA CBM3a) and Trp122. In addition, like the other scaffoldin-borne CBM3s from *C. thermocellum* and *C. cellulolyticum*, Asp64 and Arg116 of *B. cellulosolvens* CBM3b form a salt bridge, creating a closed hydrogen-bonded ring that is capable of interacting with glucose rings in a manner similar to aromatic residues (Fig. 6).

Trp105 is a constituent of the six-residue polypeptide insertion that is unique to the *B. cellulosolvens* CBM3b. This insertion extends the loop between β-strands 6 and 7 and was aligned with the other aromatic residues of the linear array (Figs. 5 and 6). The additional tryptophan apparently serves as a replacement for the tyrosine (Tyr69 in *C. thermocellum* CipA or Tyr70 in *C. cellulolyticum* CipC) located in the 11-residue peptide stretch that is missing from the *B. cellulosolvens* sequence, providing evidence that CBM3b modules display a conserved function that is modulated by structural alterations. Comparison of the structures determined by X-ray crystallography appeared to support this assumption. Significantly, in *B. cellulosolvens* CBM3b there is an 11-residue deletion in which one of the planar-strip residues shown to be important for cellulose binding is removed (corresponding to Tyr69 in *C. thermocellum* CBM3a; Benhar *et al.*, 2001). In apparent compensation, the scaffoldin-borne *B. cellulosolvens* CBM3b contains a tryptophan residue in its unique six-residue loop. The new loop serves to place Trp105 in close planar alignment with the other cellulose-binding residues. The additional tryptophan could thus compensate for the deleted tyrosine (Figs. 4 and 5) by forming stacking interactions with the glucose rings.

A closer view of the proposed cellulose-binding residues of *B. cellulosolvens* CBM3b (Fig. 6) revealed another interesting difference from the CBM3a modules of *C. thermocellum* and *C. cellulolyticum*. The length of the linear array in the *C. thermocellum* and *C. cellulolyticum* CBM3a modules is approximately 32 Å. In contrast, the equivalent length in *B. cellulosolvens* CBM3b was approximately 40 Å when Trp105 was included and 28.5 Å without Trp105, thus indicating that the pattern of aromatic residues that participate in stacking interactions with glucose rings is more diverse than previously anticipated (Tormo *et al.*, 1996).

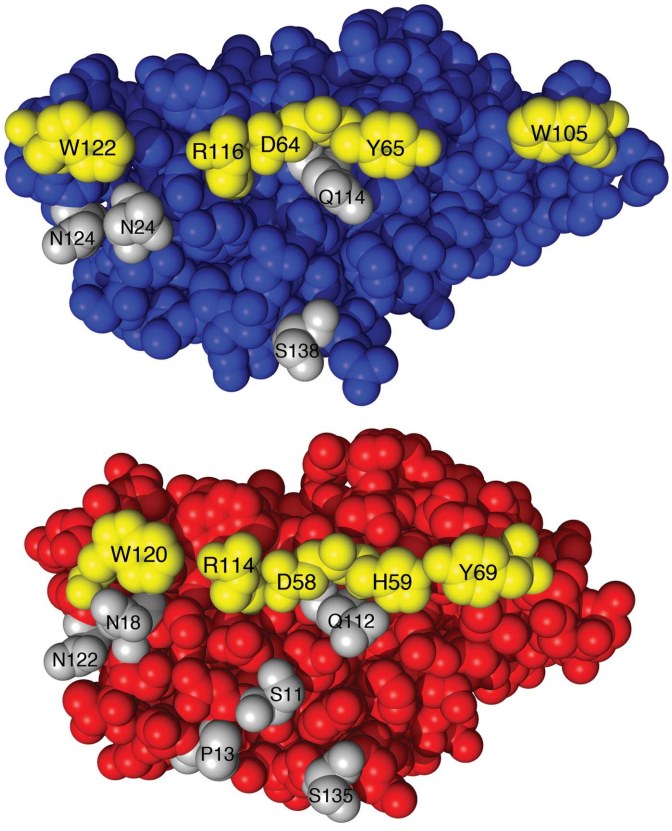


Figure 5
Space-filling representations of the proposed cellulose-binding regions of *B. cellulosolvens* CBM3b (blue) and *C. thermocellum* CipA CBM3a (red). The residues forming the cellulose-binding strip are coloured yellow. Putative cellulose-anchoring residues are coloured grey.

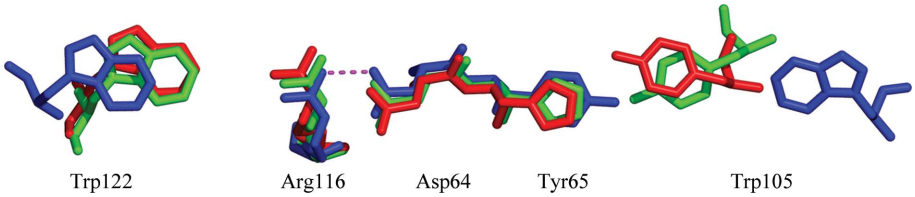


Figure 6
Superposition of the amino-acid residues forming the linear strip that purportedly binds crystalline cellulose in the known structures of CBM3s shown as a stick diagram, *B. cellulosolvens* ScaA CBM3b (blue), *C. thermocellum* CipA CBM3a (red; PDB entry 1nbc; Tormo *et al.*, 1996) and *C. cellulolyticum* CipC CBM3a (green; PDB entry 1g43; Shimon *et al.*, 2000). Residues from *B. cellulosolvens* CBM3b are numbered. The salt bridge between Arg16 and Asp64 is marked as a magenta dashed line.

3.4. Ca^{2+}-binding site

In contrast to the other known CBM3 structures (Tormo *et al.*, 1996; Sakon *et al.*, 1997; Shimon *et al.*, 2000; Petkun *et al.*, 2010), all of which contain a single Ca^{2+} ion (in the region between strands 3 and 4), in the *B. cellulosolvens* CBM3b structure reported here Ca^{2+} ions were absent from the electron-

density Fourier maps calculated with $2F_o - F_c$, $F_o - F_c$ and anomalous difference coefficients with the phases of final refinement (using *PHENIX*). In addition, the absence of the Ca^{2+} ion in our *B. cellulosolvens* CBM3b protein sample and crystals was verified by X-ray fluorescence (XRF) spectrum measurements, a technique that can identify the presence of elements in the crystals and solutions (Thompson *et al.*, 2001; Leonard *et al.*, 2009) and was manifested by the absence of emission lines that are characteristic for calcium ($K_{\alpha1}$ of 3.69 keV, $K_{\alpha2}$ of 4.01 keV).

The absence of Ca^{2+} may reflect crystallization using high concentrations of ammonium sulfate, despite several other studies in which Ca^{2+} was retained in the relevant sites of calcium-binding proteins grown under similar conditions (Szebenyi & Moffat, 1986; Swain *et al.*, 1989; Cook *et al.*, 1993; Andersson *et al.*, 1997; Buchanan *et al.*, 2005).

Despite the absence of Ca^{2+} in the structure, the five residues that form the calcium-binding site were conserved in the sequence (Figs. 2 and 7*b*). The lack of Ca^{2+} in *B. cellulosolvens* CBM3b was accompanied by changes in the molecular structure. The loop between β-strands 3 and 4 was thus found to have diverged by about 2–3 Å from the large loop between β-strands 7 and 8 (Fig. 7*b*). This shift broadens the entire molecule by about 3–5 Å relative to *C. thermocellum* CBM3a and *C. cellulolyticum* CipC CBM3a. In particular, Asp54 Cα of *B. cellulosolvens* CBM3b was shifted by about 1.6 Å from the corresponding Asp residues in *C. thermocellum* CBM3a (Asp48) and *C. cellulolyticum* CipC CBM3a (Asp46). This shift apparently disrupted the precise spatial organization that is essential for octahedral coordination of the calcium ion.

The Asp shift can possibly be explained by the replacement of Tyr121 in *C. thermocellum* CBM3a by Val125 in *B. cellulosolvens* CBM3b (Fig. 7*b*). Tyr121 participates in a network of hydrophobic interactions that eventually secure and almost lock the Asp residue in its required position for calcium binding. The absence of the aromatic residue in *B. cellulosolvens* CBM3b created a virtual void in the hydrophobic interaction network and therefore allowed the Asp residue greater flexibility.

Cellulose-binding assays revealed that despite the apparent distortion of the module the ability of *B. cellulosolvens* CBM3b to bind cellulose was not affected significantly by the

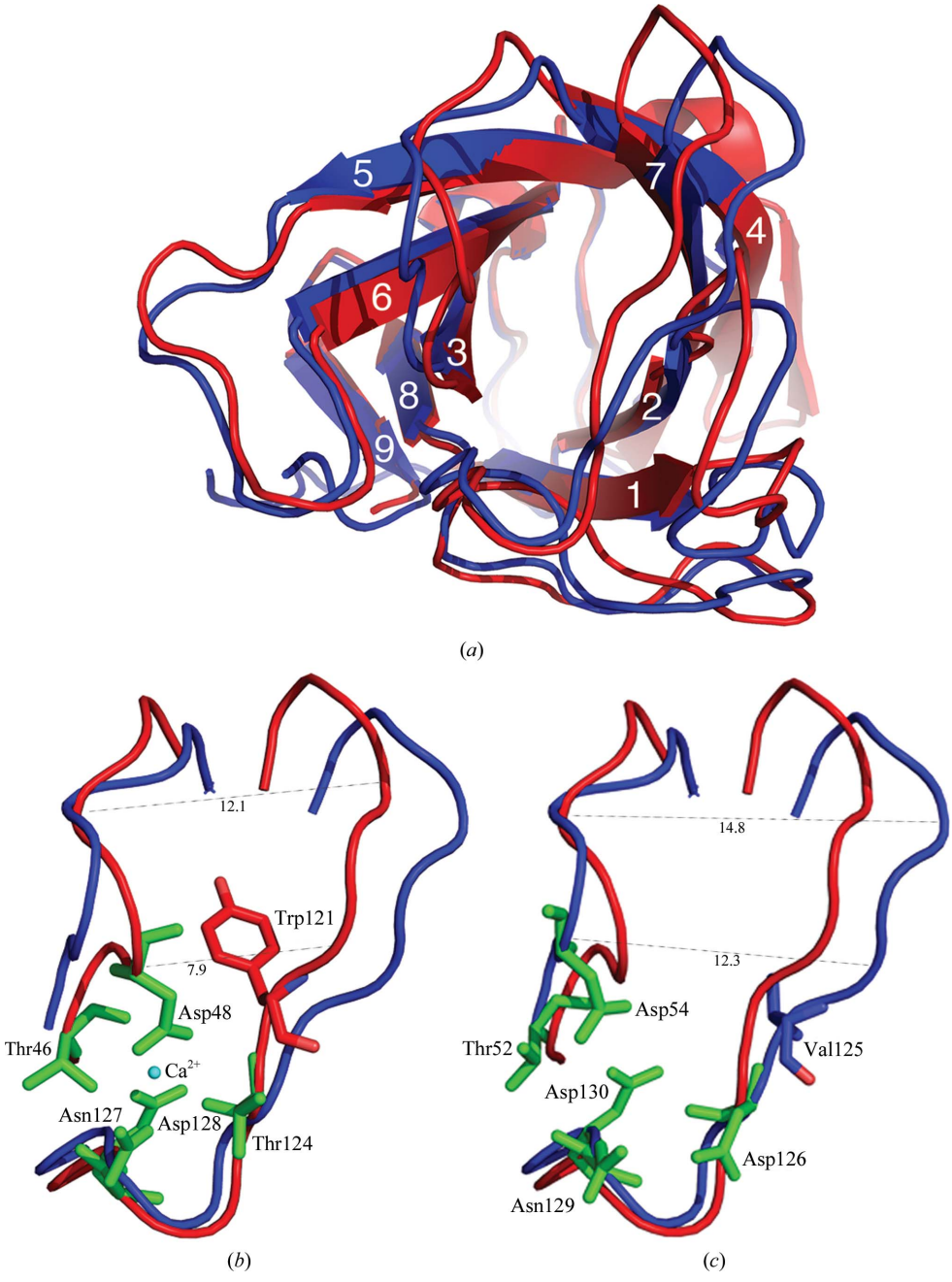


Figure 7
Loop shift and calcium-binding site. (*a*) Structural alignment of *B. cellulosolvens* CBM3b (blue) and *C. thermocellum* CBM3a (red). Major secondary structures are numbered according to Fig. 2. (*b*, *c*) Calcium-binding residues (green) and proposed determinant residues (red and blue) of *C. thermocellum* CBM3a (*b*) and *B. cellulosolvens* CBM3b (*c*). Loops that are shifted between *B. cellulosolvens* CBM3b (blue) and *C. thermocellum* CBM3a (red) are shown in cartoon representation. Distances between the loops are indicated (in Å).

absence of Ca^{2+} ions (data not shown). Comparison of the partitioning of the protein between the bound and the unbound fractions reveals almost identical results with or without addition of $CaCl_2$. In apparent agreement with this finding, mutation analysis of the CpbA CBM3a from the scaffoldin subunit of the mesophilic *C. cellulovorans* showed that substitution of the aspartate residues occupying the Ca^{2+}-binding site does not appear to have any effect on its binding to cellulose (Goldstein & Doi, 1994).

4. Summary

The three-dimensional structure of the ScaA CBM3b from *B. cellulosolvens* was determined by X-ray analysis. Despite some differences between the CBM3b from *B. cellulosolvens* and known scaffoldin-borne CBM3a structures (Tormo *et al.*, 1996; Shimon *et al.*, 2000), they all share certain common structural and functional properties that allow the targeting of the multienzyme cellulosome complex to its cellulosic substrate. The unique features of CBM3b from *B. cellulosolvens* provide molecular evidence for the versatility of this complicated system, indicating that the relationship between the family 3a and 3b CBMs is not necessarily a precise function of the parent protein. Consequently, the scaffoldin setting for CBMs is not restricted to the CBM3as, and CBM3b can also accommodate scaffoldin subunits and function as the major substrate-binding entity of the cellulosome. It would therefore be intriguing to determine the three-dimensional structure of the *A. cellulolyticus* CipV CBM3b, which is another known scaffoldin-borne CBM3b.

Historically, the division of the family 3 CBMs into subgroups (3a, 3b and 3c) was based on minor sequence differences, combined with the fact that the known scaffoldin-borne CBMs from four different clostridia could be differentiated by the existence of a distinctive Trp-containing loop (Bayer, Morag *et al.*, 1998). It seems that this distinction no longer holds for the newly discovered scaffoldin- and enzyme-containing CBMs and therefore subgroups 3a and 3b can be combined. The major features that appear to govern strong binding to crystalline cellulosic substrates are the presence of a flat surface on the CBM molecule and a planar linear strip of aromatic residues (Tormo *et al.*, 1996; Shimon *et al.*, 2000; Sakon *et al.*, 1997; Petkun *et al.*, 2010).

We thank the ESRF, Grenoble, France for use of the macromolecular crystallographic data-collection facilities and the ID29, ID23-1 and ID14-4 staff for their assistance. This research was supported by the Israel Science Foundation (ISF; Grant No. 293/08). EAB holds The Maynard I. and Elaine Wishner Chair of Bio-Organic Chemistry at the Weizmann Institute of Science.

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.

Alber, O., Noach, I., Rincon, M. T., Flint, H. J., Shimon, L. J., Lamed, R., Frolow, F. & Bayer, E. A. (2009). *Proteins*, **77**, 699–709.

Andersson, M., Malmendal, A., Linse, S., Ivarsson, I., Forsén, S. & Svensson, L. A. (1997). *Protein Sci.* **6**, 1139–1147.

Bayer, E. A., Belaich, J. P., Shoham, Y. & Lamed, R. (2004). *Annu. Rev. Microbiol.* **58**, 521–554.

Bayer, E. A., Kenig, R. & Lamed, R. (1983). *J. Bacteriol.* **156**, 818–827.

Bayer, E. A., Lamed, R., White, B. A. & Flint, H. J. (2008). *Chem. Rec.* **8**, 364–377.

Bayer, E. A., Morag, E., Lamed, R., Yaron, S. & Shoham, Y. (1998). *Carbohydrases from Trichoderma reesei and Other Microorganisms*, edited by M. Claeyssens, W. Nerinckx & K. Piens, pp. 39–65. London: The Royal Society of Chemistry.

Bayer, E. A., Shimon, L. J., Shoham, Y. & Lamed, R. (1998). *J. Struct. Biol.* **124**, 221–234.

Benhar, I., Tamarkin, A., Marash, L., Berdichevsky, Y., Yaron, S., Shoham, Y., Lamed, R. & Bayer, E. A. (2001). *Glycosyl Hydrolases in Biomass Conversion*, edited by M. F. Himmel, J. O. Baker & J. N. Saddler, pp. 168–189. Washington DC: American Chemical Society.

Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). *Biochem. J.* **382**, 769–781.

Buchanan, K. T., Ames, J. B., Asfaw, S. H., Wingard, J. N., Olson, C. L., Campana, P. T., Araújo, A. P. & Engman, D. M. (2005). *J. Biol. Chem.* **280**, 40104–40111.

Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Cook, W. J., Jeffrey, L. C., Cox, J. A. & Vijay-Kumar, S. (1993). *J. Mol. Biol.* **229**, 461–471.

Cosier, J. & Glazer, A. M. (1986). *J. Appl. Cryst.* **19**, 105–107.

DeLano, W. L. (2002). *PyMOL.* http://www.pymol.org.

Ding, S.-Y., Bayer, E. A., Steiner, D., Shoham, Y. & Lamed, R. (1999). *J. Bacteriol.* **181**, 6720–6729.

Ding, S.-Y., Bayer, E. A., Steiner, D., Shoham, Y. & Lamed, R. (2000). *J. Bacteriol.* **182**, 4915–4925.

Ding, S.-Y., Xu, Q., Ali, M. K., Baker, J. O., Bayer, E. A., Barak, Y., Lamed, R., Sugiyama, J., Rumbles, G. & Himmel, M. E. (2006). *Biotechniques*, **41**, 435–443.

Doi, R. H. & Kosugi, A. (2004). *Nature Rev. Microbiol.* **2**, 541–551.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Fontes, C. M. & Gilbert, H. J. (2010). *Annu. Rev. Biochem.* **79**, 655–681.

Garcia, J. S., Magalhães, C. S. & Arruda, M. A. (2006). *Talanta*, **69**, 1–15.

Gerngross, U. T., Romaniec, M. P., Kobayashi, T., Huskisson, N. S. & Demain, A. L. (1993). *Mol. Microbiol.* **8**, 325–334.

Gilad, R., Rabinovich, L., Yaron, S., Bayer, E. A., Lamed, R., Gilbert, H. J. & Shoham, Y. (2003). *J. Bacteriol.* **185**, 391–398.

Goldstein, M. A. & Doi, R. H. (1994). *J. Bacteriol.* **176**, 7328–7334.

Jindou, S., Xu, Q., Kenig, R., Shulman, M., Shoham, Y., Bayer, E. A. & Lamed, R. (2006). *FEMS Microbiol. Lett.* **254**, 308–316.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Lamed, R., Setter, E. & Bayer, E. A. (1983). *J. Bacteriol.* **156**, 828–836.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.

Lehtiö, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M. & Teeri, T. T. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 484–489.

Leibovitz, E. & Béguin, P. (1996). *J. Bacteriol.* **178**, 3077–3084.

Lemaire, M., Ohayon, H., Gounon, P., Fujino, T. & Béguin, P. (1995). *J. Bacteriol.* **177**, 2451–2459.

Leonard, G., Solé, V. A., Beteva, A., Gabadinho, J., Guijarro, M., McCarthy, J., Marrocchelli, D., Nurizzo, D., McSweeney, S. & Mueller-Dieckmann, C. (2009). *J. Appl. Cryst.* **42**, 333–335.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
Murray, M. G. & Thompson, W. F. (1980). *Nucleic Acids Res.* **8**, 4321–4325.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Petkun, S., Jindou, S., Shimon, L. J. W., Rosenheck, S., Bayer, E. A., Lamed, R. & Frolow, F. (2010). *Acta Cryst.* D**66**, 33–43.
Poole, D. M., Morag, E., Lamed, R., Bayer, E. A., Hazlewood, G. P. & Gilbert, H. J. (1992). *FEMS Microbiol. Lett.* **78**, 181–186.
Sakon, J., Irwin, D., Wilson, D. B. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 810–818.
Salamitou, S., Tokatlidis, K., Béguin, P. & Aubert, J. P. (1992). *FEBS Lett.* **304**, 89–92.
Shimon, L. J. W., Pagès, S., Belaich, A., Belaich, J.-P., Bayer, E. A., Lamed, R., Shoham, Y. & Frolow, F. (2000). *Acta Cryst.* D**56**, 1560–1568.
Sole, V. A., Papillon, E., Cotte, M., Walter, P. & Susini, J. (2007). *Spectrochim. Acta B*, **62**, 63–68.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Stout, G. H. & Jensen, L. H. (1968). *X-ray Structure Determination. A Practical Guide.* London: Macmillan.
Swain, A. L., Kretsinger, R. H. & Amma, E. L. (1989). *J. Biol. Chem.* **264**, 16620–16628.
Szebenyi, D. M. & Moffat, K. (1986). *J. Biol. Chem.* **261**, 8761–8777.
Thompson, A., Attwood, D., Gullikson, E., Howells, M., Kim, K.-J., Kirz, J., Kortright, J., Lindau, I., Pianetta, P., Robinson, A., Scofield, J., Underwood, J., Vaughan, D., Williams, G. & Winik, H. (2001). *X-ray Data Handbook*, edited by A. Thompson & D. Vaughan. Lawrence Berkeley National Laboratory, Berkeley, California, USA.
Tina, K. G., Bhadra, R. & Srinivasan, N. (2007). *Nucleic Acids Res.* **35**, W473–W476.
Tomme, P., Driver, D. P., Amandoron, E. A., Miller, R. C., Antony, R., Warren, J. & Kilburn, D. G. (1995). *J. Bacteriol.* **177**, 4356–4363.
Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y. & Steitz, T. A. (1996). *EMBO J.* **15**, 5739–5751.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Xu, Q., Bayer, E. A., Goldman, M., Kenig, R., Shoham, Y. & Lamed, R. (2004). *J. Bacteriol.* **186**, 968–977.