

A single mutation reforms the binding activity of an adhesion-deficient family 3 carbohydrate-binding module

Oren Yaniv,^{a,b} Svetlana Petkun,^{a,b} Linda J. W. Shimon,^c Edward A. Bayer,^d Raphael Lamed^a and Felix Frolow^{a,b*}

^aDepartment of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv 69978, Israel, ^bThe Daniella Rich Institute for Structural Biology, Tel Aviv University, Tel Aviv 69978, Israel, ^cDepartment of Chemical Research Support, The Weizmann Institute of Science, Rehovot 76100, Israel, and ^dDepartment of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

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

The crystal structure of the family 3b carbohydrate-binding module (CBM3b) of the cellulosomal multimodular hydrolytic enzyme cellobiohydrolase 9A (Cbh9A) from *Clostridium thermocellum* has been determined. Cbh9A CBM3b crystallized in space group $P4_1$ with four molecules in the asymmetric unit and diffracted to a resolution of 2.20 Å using synchrotron radiation. The structure was determined by molecular replacement using *C. thermocellum* Cel9V CBM3b' (PDB entry 2wnx) as a model. The *C. thermocellum* Cbh9A CBM3b molecule forms a nine-stranded antiparallel β -sandwich similar to other family 3 carbohydrate-binding modules (CBMs). It has a short planar array of two aromatic residues that are assumed to bind cellulose, yet it lacks the ability to bind cellulose. The molecule contains a shallow groove of unknown function that characterizes other family 3 CBMs with high sequence homology. In addition, it contains a calcium-binding site formed by a group of amino-acid residues that are highly conserved in similar structures. After determination of the three-dimensional structure of Cbh9A CBM3b, the site-specific N126W mutant was produced with the intention of enhancing the cellulose-binding ability of the CBM. Cbh9A CBM3b^{N126W} crystallized in space group $P4_12_12_1$, with one molecule in the asymmetric unit. The crystals diffracted to 1.04 Å resolution using synchrotron radiation. The structure of Cbh9A CBM3b^{N126W} revealed incorporation of the mutated Trp126 into the putative cellulose-binding strip of residues. Cellulose-binding experiments demonstrated the ability of Cbh9A CBM3b^{N126W} to bind cellulose owing to the mutation. This is the first report of the engineered conversion of a non-cellulose-binding CBM3 to a binding CBM3 by site-directed mutagenesis. The three-dimensional structure of Cbh9A CBM3b^{N126W} provided a structural correlation with cellulose-binding ability, revealing a longer planar array of definitive cellulose-binding residues.

Received 15 August 2011
Accepted 26 March 2012

PDB References: Cbh9A
CBM3b^{WT}, 2ylk; Cbh9A
CBM3b^{N126W}, 3zqx.

1. Introduction

Cellulose is the major polymeric component of plant matter and is the most abundant polysaccharide on earth (Bayer, Chanzy *et al.*, 1998). The efficient degradation of plant cell-wall polysaccharides, including forms of cellulose with high crystallinity, requires the synergistic involvement of numerous enzymes. For this reason, cellulose-degrading organisms produce a remarkable variety of enzymes that exist either in the free state or in association with a multi-enzyme complex known as the cellulosome (Bayer *et al.*, 2004).

The cellulosome, which was first described for the anaerobic thermophile *Clostridium thermocellum* (Bayer *et al.*, 1983), contains over 25 different proteins, mostly hydrolytic enzymes, and is capable of efficient hydrolysis of plant cell-wall polysaccharides (Fig. 1*a*). Cellulosome components have multimodular architectures (Bayer *et al.*, 2006), being composed of one or more catalytic modules and a variety of auxiliary modules. The catalytic modules of glycoside hydrolases (GHs; EC 3.2.1.–) are a widespread group of enzymes that hydrolyze the glycosidic bond between adjacent carbohydrate residues.

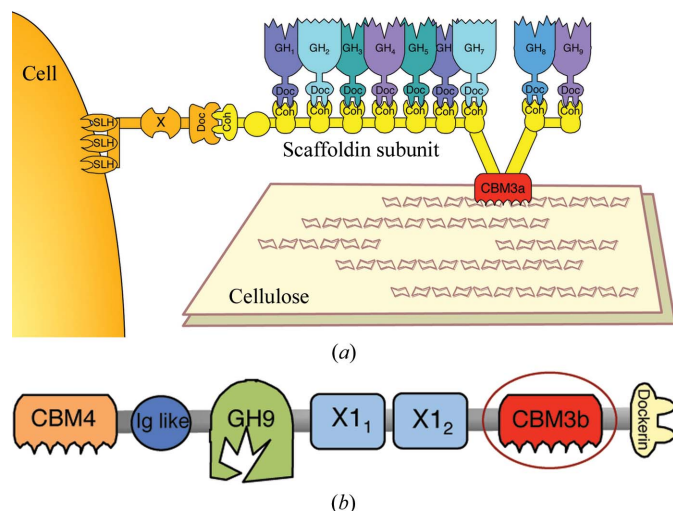


Figure 1

(*a*) Schematic representation of the proposed cell-surface disposition of the cellulosomal components of *C. thermocellum*. Dockerin (Doc)-containing enzymes (glycoside hydrolases, GHs) bind selectively to any of the nine type I cohesins (Coh) of the primary CipA scaffoldin. In turn, the terminal X-dockerin dyad of the CipA scaffoldin binds to the type II cohesins of an anchoring scaffoldin, which is connected to the cell surface via an S-layer homology (SLH) module. The cellulose-binding module (CBM) of the primary scaffoldin binds the cellulosome complex and the attached cell to the cellulosic substrate. (*b*) Schematic representation of the modular organization of *C. thermocellum* Cbh9A. The CBM3b module reported here is emphasized by a red ellipse. Ig like, immunoglobulin-like module; GH9, family 9 glycoside hydrolase; X1, X domain.

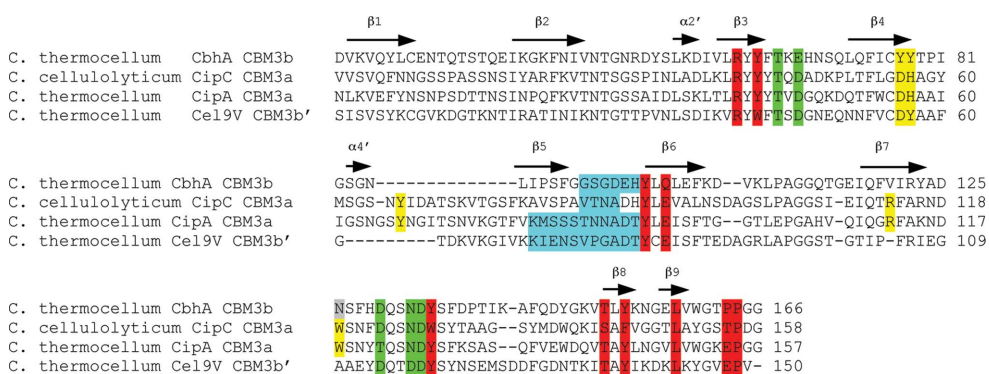


Figure 2

Sequence alignment of *C. thermocellum* Cbh9A CBM3b, *C. cellulolyticum* CipC CBM3a, *C. thermocellum* CipA CBM3a and *C. thermocellum* Cel9V CBM3b. Secondary-structural elements (β strands) are indicated and enumerated. Proposed cellulose-binding residues are shown in yellow. The mutated residue (Asn126) is shown in grey. Calcium-binding residues are shown in green. Residues that occupy the shallow groove region are shown in red. Residues forming the loops between the $\beta 5$ and $\beta 6$ strands are shown in cyan.

GHs are currently classified into 130 families based on sequence similarity (Cantarel *et al.*, 2009), and over 23 families are present in the cellulosome of *C. thermocellum* (data taken from the Carbohydrate Active Enzymes database; <http://www.cazy.org/>; Cantarel *et al.*, 2009).

The auxiliary modules serve numerous functions: carbohydrate-binding modules (CBMs) enhance the association of the catalytic modules with insoluble carbohydrates (Tomme *et al.*, 1998; Kataeva *et al.*, 1998), surface-layer homology (SLH) modules bind enzymes to cell walls (Matuschek *et al.*, 1996), other modules such as fibronectin type 3 modify substrate surfaces, probably by loosening the cellulose chains (Kataeva *et al.*, 2002), and dockerins mediate the integration of the enzymes into cellulosomes, some of which are stabilized by X modules (Adams *et al.*, 2006).

Family 3 CBMs are often found in association with GH9 enzymes and also represent the major module on the cellulosomal scaffoldin subunit for targeting the multi-enzyme complex to the crystalline cellulose substrate (Bayer, Shimon *et al.*, 1998).

Historically, the division of the >350 known CBMs in family 3 (data taken from the Carbohydrate Active Enzymes database) into subgroups (a, b and c) was based on minor sequence differences (Bayer, Morag *et al.*, 1998). Members of subgroups a and b have been shown to bind strongly to the surface of microcrystalline cellulose in biochemical studies and have been proposed to promote the cellulolytic reaction by concentrating the enzyme near the cellulose surface (Gilad *et al.*, 2003; Tormo *et al.*, 1996). Some members of subgroup b (termed subgroup b') fail to bind to crystalline cellulose and appear to deviate in sequence from the norm, particularly at selected positions considered to be important to the cellulose-binding function (Jindou *et al.*, 2006; Petkun *et al.*, 2010).

Representatives of CBMs from family 3 that belong to subgroup c fail to bind crystalline cellulosic substrates; instead, the members of this subgroup are associated with certain GH9 catalytic modules and have been shown to alter GH9 function from the standard endo-acting mode to a processive endoglucanase (Sakon *et al.*, 1997; Zverlov *et al.*, 2003).

One of the major cellulosomal multimodular hydrolytic enzymes is cellobiohydrolase 9A (Cbh9A; Fig. 1*b*) from *C. thermocellum*. This enzyme is composed, starting from the N-terminus, of a family 4 CBM, an Ig-like module, a family 9 GH module, X1₁ and X1₂ modules (formerly and erroneously called Fn3s), a family 3b CBM and a dockerin module (Zverlov *et al.*, 1998).

The family 3b CBM from Cbh9A (Cbh9A CBM3b) diverges from standard CBM3s by several changes in the residues that are highly conserved in the putative cellulose-binding surface

Table 1

Oligonucleotide primers and plasmids used in this study.

Bases that introduce the mutation in Cbh9A CBM3b^{N126W} are highlighted in bold.

| Construct | Nucleotide sequence | Comments |
|-----------------------------|---|--|
| CbhA CBM3b ^{WT} | F, 5'-CCATGGGCGATGTAAAAGTACAG-TATTTGTGC R, 5'-CTCGAGCGGCGGCGTTCCCCAAAC | General primers, used in all clones |
| CbhA CBM3b ^{N126W} | R, 5'-GATATGCAGAT TGGT CCTTCC R, 5'-GGAAGGACCAATCTGCATATC | Specific primers for overlap-extension PCR mutagenesis |

of the CBM, the so-called 'linear aromatic strip' (Jindou *et al.*, 2007), and by a unique sequence alteration that almost eliminates one of the loops in the molecule (between strands 5 and 6; highlighted in cyan in Fig. 2). Furthermore, it lacked the ability to bind crystalline cellulose in our hands (Jindou *et al.*, 2006, 2007), although others (Kataeva *et al.*, 2002) have reported that this CBM3b does bind cellulose. To clarify this ambiguity, X-ray structure analysis was initiated.

Our results revealed that Cbh9A CBM3b adopts a nine-stranded β -sandwich fold similar to the other known family 3 CBMs, but also shows a very short linear array of proposed cellulose-binding residues that is apparently not sufficient for stable substrate binding. One of the important binding residues of the linear aromatic strip, a tryptophan, is missing and has apparently been replaced in the sequence by an asparagine at position 126. Consequently, site-directed mutagenesis was performed in an attempt to reform the ability of Cbh9A CBM3b to bind cellulose. Replacement of Asn126 by Trp was indeed successful in introducing cellulose-binding function to Cbh9A CBM3b. Here, we report the determination of the crystal structures of two different constructs of Cbh9A CBM3b (wild type and N126W mutant) at atomic or near-atomic resolution. This is the first report in which cellulose-binding ability of an 'inactive' CBM3 has been introduced by site-specific mutagenesis.

2. Materials and methods

2.1. Cloning, expression and purification

Cloning of the CBM3b from Cbh9A (GenBank accession No. X80993; residues 1003–1148; Cbh9A CBM3b^{WT}) followed by protein expression and purification were carried out as described in Jindou *et al.* (2007). Site-directed mutagenesis was performed using the overlap-extension PCR mutagenesis procedure (Ho *et al.*, 1989; Meza *et al.*, 1996). The amino-acid residue of interest was changed using the oligonucleotide primers listed in Table 1. The mutated DNA fragment was cloned and inserted into pET-28a(+) as described in Jindou *et al.* (2007), yielding pET-CbhA CBM3b^{N126W}. Cbh9A CBM3b^{WT} and Cbh9A CBM3b^{N126W} were concentrated to 23 and 16.5 mg ml⁻¹, respectively, 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.

2.2. Cellulose-binding assay

A qualitative assessment of binding to insoluble microcrystalline cellulose was made using Avicel (Merck AG, Darmstadt, Germany), as reported previously (Xu *et al.*, 2004). Briefly, samples of purified CBM (20 μ g) were added to 5 mg prewashed insoluble cellulose in Tris buffer (50 mM Tris-HCl buffer pH 7.5) at a final volume of 200 μ l. The assay tubes were agitated by slow rotation at room temperature for 1 h and the samples were sedimented by centrifugation. The supernatant was set aside as the unbound fraction and a one-third volume of SDS-containing sample buffer was added. The pellets (bound fraction) were washed three times with 1 ml Tris buffer. The pellets were resuspended in 200 μ l Tris buffer and a one-third volume of SDS-containing sample buffer. Boiling the samples for 10 min eluted the bound protein. The binding of each CBM to the insoluble cellulose was evaluated by SDS-PAGE using 20 μ l of the bound and unbound fractions. Bovine serum albumin (BSA) was used in all experiments as a negative control (not shown). Each experiment was repeated at least three times.

2.3. Crystallization

The protein samples were screened for crystallization conditions by the hanging-drop vapour-diffusion method (McPherson, 1982) on siliconized glass cover slips using 146 pre-formulated crystallization solutions from Hampton Research HR screens (Crystal Screen, Crystal Screen 2 and PEG/Ion). Optimization of the crystallization conditions was performed using the hanging-drop method. The initial crys-

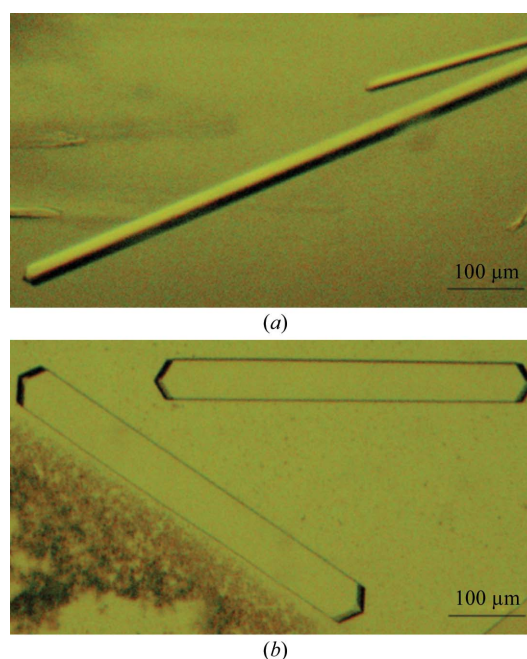


Figure 3
Crystals of (a) Cbh9A CBM3b^{WT} and (b) Cbh9A CBM3b^{N126W}.

Table 2

Crystallization condition data.

| | Cbh9A CBM3b ^{WT} | CbhA CBM3b ^{N126W} |
|---------------------|--|--|
| Original condition | PEG/Ion condition No. 48 [20% (v/v) PEG 3350, 0.2 M diammonium hydrogen citrate] | Crystal Screen 2 condition No. 17 [35% (v/v) <i>tert</i> -butanol, 0.1 M trisodium citrate dehydrate pH 5.6] |
| No. of days | 2–3 | 3 |
| Optimized condition | The same as the original condition | 30% (v/v) <i>tert</i> -butanol, 0.1 M trisodium citrate dehydrate pH 5.6 |
| Crystal picture | Fig. 3(a) | Fig. 3(b) |

tallization conditions were optimized and are presented in Table 2. The crystals are shown in Fig. 3. All crystallization experiments were performed at 293 K.

2.4. X-ray diffraction

The crystals were harvested from the crystallization drop using a MiTeGen micromount (<http://www.mitegen.com>) made of polyimide and transferred for several seconds into a cryostabilization solution comprised of equal volumes of a twofold-concentrated solution of the crystallization components and a solution consisting of 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 the MiTeGen micromount, plunged into liquid nitrogen and placed in pucks for mounting at a synchrotron. Diffraction data for Cbh9A CBM3b^{WT} and Cbh9A CBM3b^{N126W} were collected on the ID23-1 and ID14-4 beamlines, respectively, at the ESRF, Grenoble, France at a temperature of 100 K. *DENZO* and *SCALEPACK* were used for indexing, integration and scaling as implemented in *HKL-2000* (Otwinowski & Minor, 1997).

The Cbh9A CBM3b^{WT} crystals diffracted to 2.20 Å resolution and belonged to the tetragonal space group *P*₄₁, with unit-cell parameters *a* = *b* = 95.45, *c* = 82.91 Å. The Cbh9A CBM3b^{N126W} crystals diffracted to 1.04 Å resolution and belonged to the tetragonal space group *P*₄₁₂₁₂, with unit-cell parameters *a* = *b* = 50.04, *c* = 122.59 Å.

The *MATTHEWS_COEF* program as implemented in *CCP4* (Collaborative Computational Project, Number 4, 1994; Winn *et al.*, 2011) indicated four molecules in the asymmetric unit for Cbh9A CBM3b^{WT}, with a Matthews coefficient of 2.71 Å³ Da⁻¹ (Matthews, 1968), as most probable (Kantardjieff & Rupp, 2003), corresponding to a crystal solvent content of 54.7%.

For Cbh9A CBM3b^{N126W} one molecule per asymmetric unit was indicated as most probable, with a Matthews coefficient of 2.05 Å³ Da⁻¹, corresponding to a solvent content of 40.17%. The crystal diffraction and refinement statistics are presented in Table 3.

2.5. Structure determination and refinement

The structure of Cbh9A CBM3b^{WT} was determined by molecular replacement employing *MOLREP* (Vagin & Teplyakov, 2010) as implemented in *CCP4*, using the atomic coordinates of *C. thermocellum* Cel9V CBM3b' (PDB entry

Table 3

Crystal diffraction and refinement statistics.

Values in parentheses are for the highest resolution shell.

| | Cbh9A CBM3b ^{WT} | Cbh9A CBM3b ^{N126W} |
|--|---|---|
| X-ray source | ESRF beamline ID23-1 | ESRF beamline ID14-4 |
| Detector | ADSC Quantum 315 | ADSC Quantum 315 |
| Wavelength (Å) | 0.9700 | 0.93950 |
| No. of images | 130 | 260 |
| Oscillation range (°) | 1 | 0.5 |
| Space group | <i>P</i> ₄ ₁ | <i>P</i> ₄ ₁ ₂ ₁ ₂ |
| No. of crystals | 1 | 1 |
| No. of frames | 130 | 260 |
| Total rotation angle (°) | 130 | 130 |
| Unit-cell parameters (Å) | <i>a</i> = <i>b</i> = 95.45, <i>c</i> = 82.91 | <i>a</i> = <i>b</i> = 50.04, <i>c</i> = 122.59 |
| Unit-cell volume (Å ³) | 755439 | 306933 |
| No. of molecules in asymmetric unit | 4 | 1 |
| Resolution range (Å) | 50–2.20 (2.24–2.20) | 50–1.04 (1.06–1.04) |
| Total No. of reflections | 150443 | 771262 |
| Unique reflections | 37875 | 75775 |
| Mosaicity (°) | 0.26–0.53 | 0.20–0.34 |
| Multiplicity | 3.97 | 10.2 |
| Completeness (%) | 99.7 (99.9) | 97.5 (95.4) |
| Average <i>I</i> /σ(<i>I</i>) | 17.3 (1.4) | 52.9 (3.4) |
| <i>R</i> _{merge} † | 0.088 | 0.053 |
| Overall <i>B</i> factor from Wilson plot (Å ²) | 38.1 | 8.2 |
| No. of protein atoms | 4652 | 1206 |
| No. of solvent atoms | 288 | 239 |
| No. of ions | 0 | 1 |
| <i>R</i> _{cryst} | 0.173 | 0.166 |
| <i>R</i> _{free} | 0.226 | 0.170 |
| R.m.s.d. bond lengths (Å) | 0.011 | 0.005 |
| R.m.s.d. angles (°) | 1.288 | 1.251 |
| <i>MolProbity</i> validation | | |
| Ramachandran favoured (goal >98%) (%) | 99.1 | 100.0 |
| Ramachandran outliers (goal <0.2%) (%) | 0.0 | 0.0 |
| <i>C</i> ^β deviations >0.25 Å (goal 0) | 0 | 1 |
| Clash score‡ | 13.99 | 1.75 |
| Rotamer outliers (goal <1%) (%) | 0.6 | 0 |
| Residues with bad bonds (goal <1%) (%) | 0 | 0 |
| Residues with bad angles (goal <0.5%) (%) | 0 | 0 |

† *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 the sum over all equivalent reflections (Stout & Jensen, 1968). ‡ Clash score is the number of serious steric overlaps (>0.4 Å) per 1000 atoms.

2wnx; Petkun *et al.*, 2010) as a search model. The sequence identity between the model and the target protein was 31%. The *R* factor for the best solution was 56.9%, with a correlation coefficient of 47.0%. It is worth noting that other known CBM structures (PDB entry 1nbc, sequence identity 31%, Tormo *et al.*, 1996; PDB entry 1g43, sequence identity 28%, Shimon *et al.*, 2000) failed to produce a sensible molecular-replacement solution in our hands, probably owing to the structural divergence and the presence of four molecules in the asymmetric unit.

After the first round of ten cycles of restrained refinement in *REFMAC5* (Murshudov *et al.*, 2011), the *R* and *R*_{free} values were 0.421 and 0.487, respectively. Iterative cycles of model

building and refinement were carried out using the programs *Coot* (Emsley *et al.*, 2010) and *REFMAC5* from the *CCP4* suite. NCS (noncrystallographic symmetry) refinement was used during structure refinement and the *R* and *R*_{free} factors converged to 0.246 and 0.305, respectively. Further refinement was performed using *PHENIX* (Adams *et al.*, 2010) and was assessed using *Coot*. NCS restraints were released and after several rounds of refinement, manual model corrections and addition of solvent atoms convergence of refinement of the Cbh9A CBM3b^{WT} model at 2.20 Å resolution was reached, with corresponding *R* and *R*_{free} values of 0.173 and 0.226, respectively. The structure was validated using *MolProbity* (Chen *et al.*, 2010) as implemented in *PHENIX*.

The structure of Cbh9A CBM3b^{N126W} was determined by molecular replacement employing *MOLREP* as implemented in *CCP4*, using the atomic coordinates of chain *A* of Cbh9A CBM3b^{WT} as a search model. The *R* factor for the best solution was 54.2%, with a correlation coefficient of 63%. The model was subjected to several rounds of restrained refinement of positional and thermal parameters using *REFMAC5*, followed by manual building of the mutated side chain of Cbh9A CBM3b^{N126W} using *Coot*. The *R* and *R*_{free} values converged to 0.218 and 0.224, respectively. Further refinement was performed using *PHENIX* and was assessed using *Coot*. After several rounds of refinement with anisotropic

temperature parameters, solvent atoms, H atoms as the riding model and manual model corrections, a final Cbh9A CBM3b^{N126W} model at 1.04 Å resolution was obtained. The corresponding *R* and *R*_{free} values for this model were 0.166 and 0.170, respectively. The structure was validated using *MolProbity* as implemented in *PHENIX*. The stereochemical quality of the structures was good, with over 99% of the residues in the most favourable regions of the Ramachandran plot (Table 3).

The refinement statistics and validation results are summarized in Table 3. The final coordinates and structure factors for Cbh9A CBM3b^{WT} and Cbh9A CBM3b^{N126W} were submitted to the PDB (entries 2ylk and 3zqx, respectively). The first 21 residues (comprising the N-terminal histidine tag) were not observed in the electron-density maps. The protein structure is numbered according to the cloned sequence.

3. Results and discussion

3.1. Overall structure analysis of Cbh9A CBM3b from *C. thermocellum* and comparison with CipA CBM3a and Cel9V CBM3b'

The crystals of Cbh9A CBM3b^{WT} grew as very long tetragonal needles that were square in cross-section, with the *c* axis in the long crystal axis direction. They exhibited exceptionally large elasticity perpendicular to the needle length (Fig. 3*a*). The crystals of Cbh9A CBM3b^{N126W} formed elongated prisms with tetragonal pyramidal facets on the distal ends (Fig. 3*b*).

The Cbh9A CBM3b^{WT} module, with 166 amino-acid residues, forms a nine-stranded antiparallel β -sandwich of approximate overall dimensions 40 × 25 × 22 Å (measured on the C α skeleton; Fig. 4*a*). The first 21 amino-acid residues of the cloned protein were not visible in the $2F_{\text{obs}} - F_{\text{calc}}$ and $F_{\text{obs}} - F_{\text{calc}}$ electron-density maps. Similar to the other published structures of family 3 CBMs, the 'bottom' β -sheet of the β -sandwich is flat and is formed by strands 1, 2, 7 and 4, while the 'top' β -sheet is comprised of antiparallel strands 5, 6, 3, 8 and 9. The N- and C-termini are both located in the 'top' sheet region. The top and bottom β -sheets are formed by 25 and 26 residues, respectively, out of a total of 145 residues (about 35%) in the protein. The remainder of the residues are in extended loops that apparently contribute to the overall stability of the molecule.

In order to identify the interactions at the interface between the top and bottom β -sheets, the *Protein Interactions Calculator* server (<http://crick.mbu.iisc.ernet.in/~PIC>; Tina *et al.*, 2007) was used. The interface

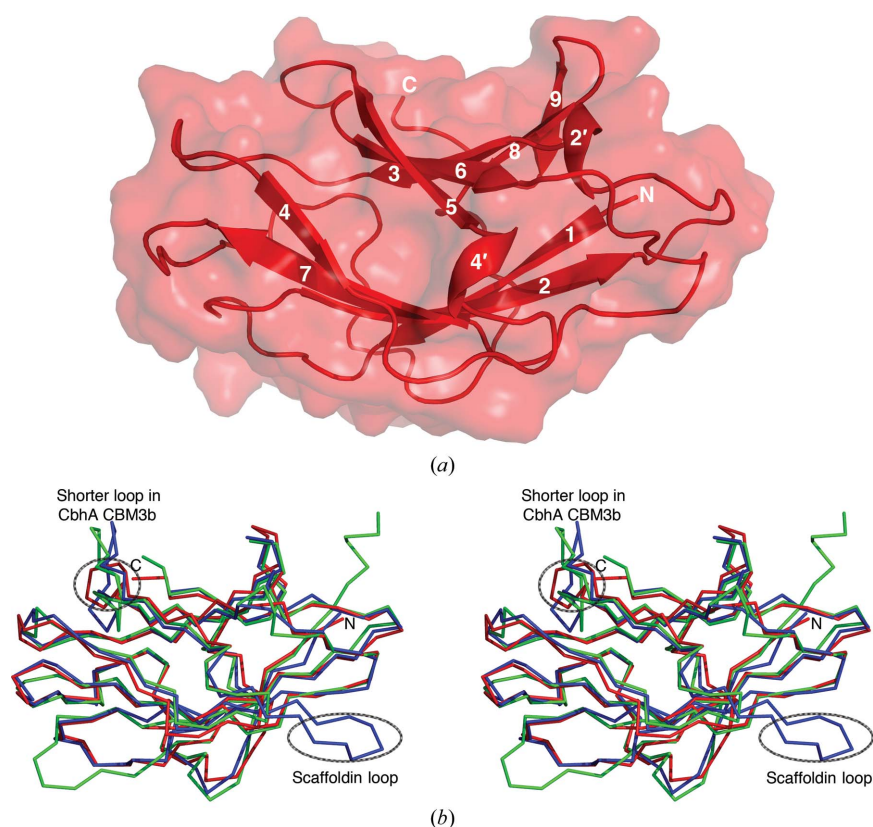


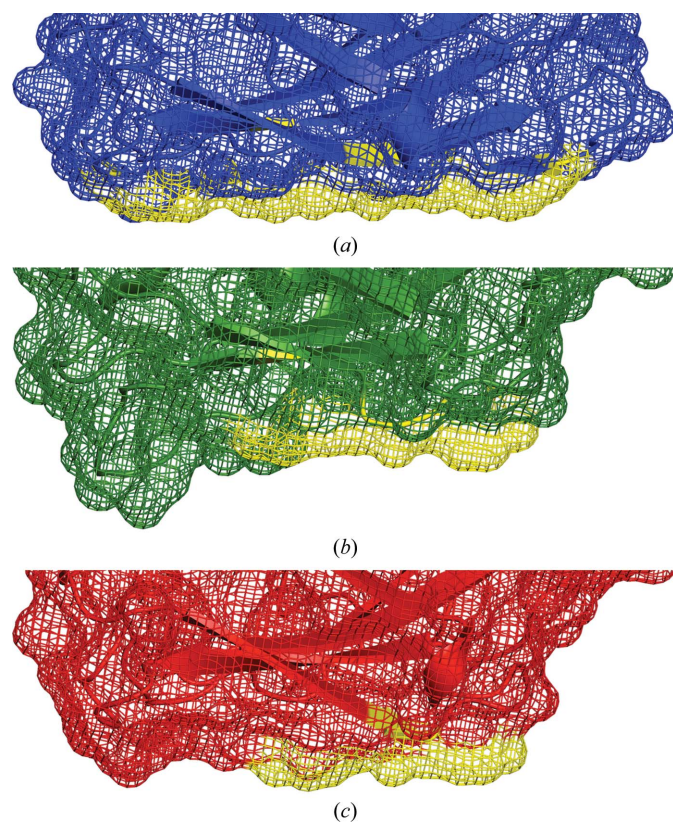
Figure 4
(*a*) Cartoon representation of the overall structure of Cbh9A CBM3b from *C. thermocellum*. The major secondary-structural elements are numbered according to Fig. 2. (*b*) Stereo diagram of the superimposed C α traces of *C. thermocellum* Cbh9A CBM3b (red), *C. thermocellum* Cel9V CBM3b (green) and *C. thermocellum* CipA CBM3a (blue). (*a*) and (*b*) are rendered in the same molecular orientation.

Table 4

R.m.s.d. values of C α atoms for superposed structures (annotated with their PDB codes) of several CBM3s.

Calculated with *PyMOL*. The number of C α atoms used for alignment is shown in parentheses. Values are in Å.

| Protein | CipA CBM3a, 1nbc | Cel9V CBM3b, 2wnx | Cbh9A CBM3b ^{WT} , 2ylk | | | |
|-------------------------------------|---------------------|----------------------|----------------------------------|-------------|-------------|-------------|
| | | | Chain A | Chain B | Chain C | Chain D |
| Cel9V CBM3b, 2wnx | 0.804 (109) | | | | | |
| Cbh9A CBM3b ^{WT} , 2ylk | | | | | | |
| Chain A | 0.968 (105) | 0.875 (102) | | | | |
| Chain B | 0.941 (105) | 0.868 (103) | 0.164 (135) | | | |
| Chain C | 0.920 (104) | 0.819 (100) | 0.149 (125) | 0.161 (129) | | |
| Chain D | 0.882 (103) | 0.794 (100) | 0.203 (134) | 0.151 (129) | 0.185 (129) | |
| Cbh9A CBM3b ^{N126W} , 3zqx | 0.949 (104) | 0.874 (99) | 0.449 (127) | 0.428 (124) | 0.420 (125) | 0.384 (121) |

**Figure 5**

Comparison of the proposed cellulose-binding faces in *C. thermocellum* CipA CBM3a, Cel9V CBM3b and Cbh9A CBM3b^{WT}. The surfaces of the molecules are represented by a mesh. Putative cellulose-binding residues are shown in yellow. (a) *C. thermocellum* CipA CBM3a (blue). (b) *C. thermocellum* Cel9V CBM3b (green). (c) *C. thermocellum* Cbh9A CBM3b^{WT} (red). In Cbh9A CBM3b (as in CipA CBM3a), the proposed cellulose-binding face is relatively flat with no major topographic obstacles. In Cel9V CBM3b there are structural changes in the backbone, creating a surface 'hump'.

analysis revealed a hydrophobic core with a total contact area of 1100 Å² which is formed by 21 hydrophobic and four aromatic interactions. The formation of such intramolecular interactions, which are mostly hydrophobic, is typical for cellulosomal β -sandwich proteins; these interactions serve to

assist in folding and help to stabilize the molecule (Alber *et al.*, 2009).

The overall structure of Cbh9A CBM3b is similar to the related structures of CBM3a of the CipA scaffoldin of *C. thermocellum* (PDB entry 1nbc; Tormo *et al.*, 1996) and of CBM3b' located in endoglucanase Cel9V of *C. thermocellum* (PDB entry 2wnx; Petkun *et al.*, 2010) (Figs. 2 and 4b), with pairwise r.m.s.d.s calculated on C α atoms of 0.968 and 0.804 Å, respectively (Table 4), as calculated using *PyMOL*. Despite the similarity in the overall structure and topology of the CBM3s, several distinctive differences that might influence the cellulose-binding

ability can be noted. Cbh9A CBM3b (like all family 3b CBMs) lacks the scaffoldin loop (Fig. 4b; Ding *et al.*, 1999) observed in all CBM3as; on the other hand, the putative cellulose-binding face of Cbh9A CBM3b is flat, without any topographic obstacles, as opposed to Cel9V CBM3b (Fig. 5). Another distinctive difference is the variation in the length of the loop between β -strands 5 and 6. In Cbh9A CBM3b the length of this loop is only six residues, as opposed to about 11 residues in CipA CBM3a and Cel9V CBM3b' (Fig. 2, highlighted in cyan).

3.2. Cellulose binding

The cellulose-binding ability of CBM3s is attributed to a planar linear array (also known as the 'linear strip') of aromatic residues that form hydrophobic stacking interactions with glucose rings from the cellulose chains (Tormo *et al.*, 1996; Shimon *et al.*, 2000). The residues are widely distributed in the primary sequence, but in the folded protein they align to form a linear strip embedded in a smooth planar surface. In CBM3a from *C. thermocellum* this strip is composed of five residues and is 31.8 Å in length; it is accompanied by two additional strips of polar anchoring side chains (Fig. 6a). In Cbh9A CBM3b the linear array is composed of only two residues, Tyr77 and Tyr78, and is 15.8 Å in length (Figs. 2 and 6b). This observation is in agreement with its inability to bind crystalline cellulose (Jindou *et al.*, 2006).

In previous work (Petkun *et al.*, 2010), the structure of *C. thermocellum* CBM3b' from Cel9V (Cel9V CBM3b') was reported. In Cel9V CBM3b', in addition to the loss of several cellulose-binding residues (similar to Cbh9A CBM3b), there are structural changes in the backbone that create a surface 'hump' that apparently interferes with the formation of cellulose–protein surface interactions. Site-directed mutation (A121W) of the nonconserved linear-strip residue in Cel9V CBM3b' did not reform the cellulose-binding ability of Cel9V CBM3b', confirming that the loss of the conserved cellulose-binding residues alone fails to explain the loss of cellulose-binding ability of this molecule (Petkun *et al.*, 2010).

In Cbh9A CBM3b, however, the putative cellulose-binding face is relatively flat, with no major topographic obstacles that can prevent the formation of stacking interactions with cellulose (Fig. 5c). Therefore, an attempt to introduce the cellulose-binding function into 'inactive' Cbh9A CBM3b using mutagenesis was initiated.

Selection of the point mutation was based on putative knowledge of which residues participate in and contribute to cellulose binding and on a structural alignment of the crystal structures of *C. thermocellum* CipA CBM3a (PDB entry 1nbc; Tormo *et al.*, 1996), *C. cellulolyticum* CipC CBM3a (PDB entry 1g43; Shimon *et al.*, 2000), *C. thermocellum* Cel9V CBM3b' (PDB entry 2wnx; Petkun *et al.*, 2010) and *C. thermocellum* Cbh9A CBM3b. Site-specific mutagenesis performed on *C. thermocellum* CipA CBM3a (Jindou *et al.*, 2006) demonstrated that Trp118 plays an important role in cellulose binding and that its mutation to Ala has a deleterious effect on the binding of the mutant to cellulosic polymers. This Trp118 corresponds to Asn126 in Cbh9A CBM3b (Fig. 7a), so a logical

selection for mutation was the replacement of Asn126 by Trp. Site-specific mutagenesis was performed for this purpose, yielding Cbh9A CBM3b^{N126W}.

The structure revealed that the newly introduced Trp126 is aligned with the other two aromatic residues to form an improved and longer linear array of 25.4 Å in length (Fig. 7b). Moreover, owing to different packing arrangements, the relative orientation of Tyr77 and Tyr78 in the mutant structure diverges from that in the wild-type structure by the rotation of Tyr77 by 122° clockwise around the Tyr77 C^β–C^γ bond. This rotation places the aromatic ring of Tyr77 of Cbh9A CBM3b^{N126W} in plane with the aromatic rings of Tyr78 and Trp126, thereby forming a planar strip. The cellulose-binding

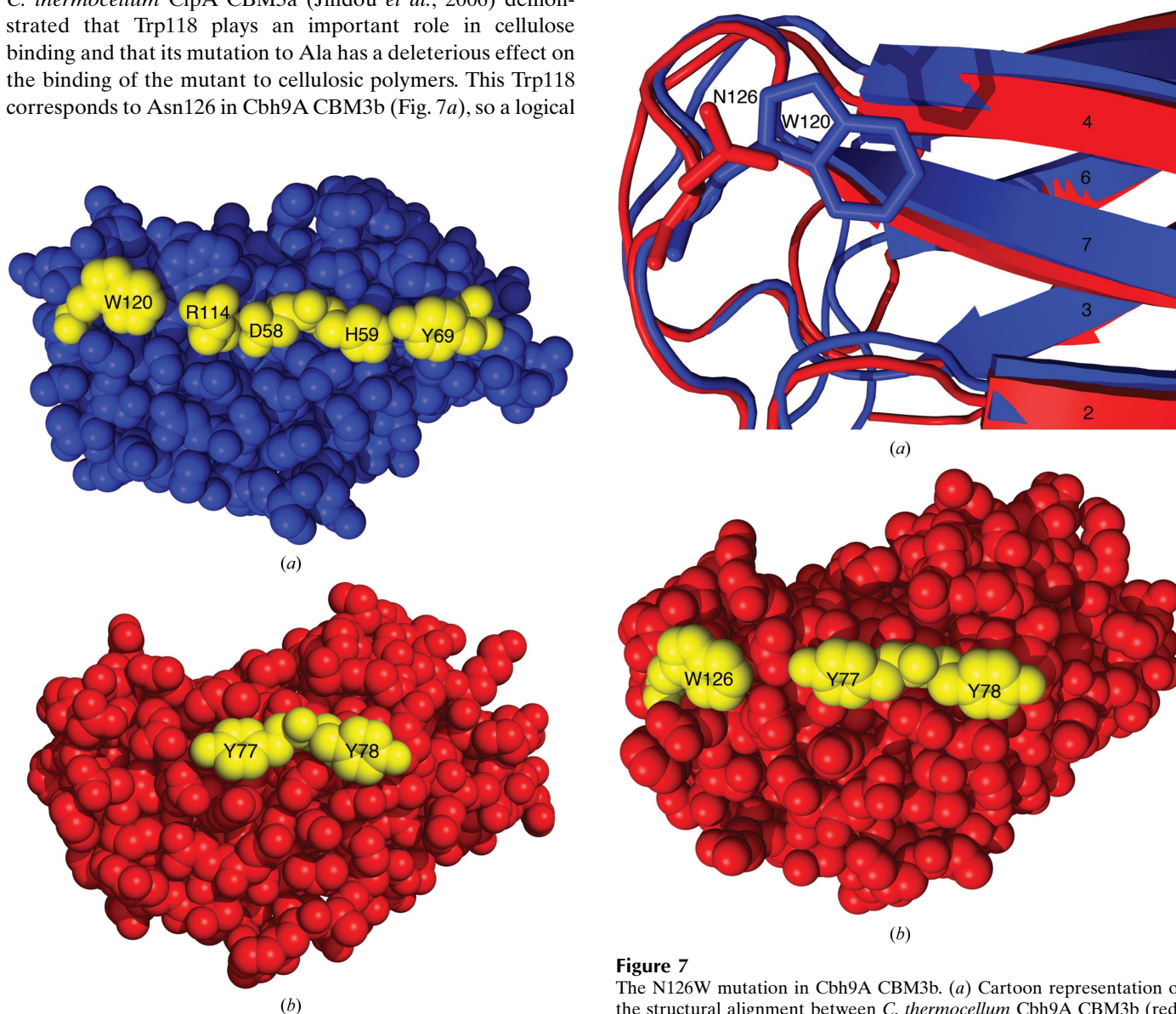


Figure 6
Proposed cellulose-binding region of Cbh9A CBM3b^{WT} compared with that of CipA CBM3a. Molecules are represented as space-filling atomic models. Residues forming the cellulose-binding strip are highlighted in yellow and numbered according to Fig. 2. (a) *C. thermocellum* CipA CBM3a (blue). (b) *C. thermocellum* Cbh9A CBM3b^{WT} (red).

Figure 7
The N126W mutation in Cbh9A CBM3b. (a) Cartoon representation of the structural alignment between *C. thermocellum* Cbh9A CBM3b (red) and *C. thermocellum* CipA CBM3a (blue) (r.m.s.d. 0.92 Å). The main-chain atoms and C^β of amino-acid residue Asn126 of Cbh9A CBM3b^{WT} are in close proximity to those of Trp120 of CipA CBM3a. (b) Proposed cellulose-binding region represented by a space-filling atomic model in *C. thermocellum* Cbh9A CBM3b^{N126W}. Residues forming the cellulose-binding strip are highlighted in yellow.

function of the new Cbh9A CBM3b^{N126W} was tested using binding assays (Fig. 8), which indicated that the protein is 'active' in the sense that it can bind microcrystalline cellulose. It therefore seems that three aromatic residues are sufficient to establish a stable substrate–protein interaction.

3.3. Ca²⁺-binding site

Cbh9A CBM3b contains a solvent-accessible calcium-binding site that is conserved in CBMs that display a β -sandwich fold, in which it apparently plays a structural role (Hashimoto, 2006; Tormo *et al.*, 1996; Figs. 2 and 9). Interestingly, even though the two constructs (Cbh9A CBM3b^{WT} and Cbh9A CBM3b^{N126W}) are identical with respect to the calcium-binding amino-acid residues and in the purification steps of the expressed proteins, a calcium ion was found in the crystal structure of the Cbh9A CBM3b^{N126W} mutant but not in that of the wild-type protein. The anomalous difference Fourier map for the Cbh9A CBM3b^{N126W} structure was calculated and the largest peak at 34.7 σ was located in a position with coordination geometry typical for Ca²⁺, with the next peak at 12.9 σ located at the position of the S atom of Cys76 (here, σ is the r.m.s. of the anomalous Fourier map). A comparable peak was absent from the anomalous Fourier map of the Cbh9A CBM3b^{WT} structure (there are no peaks above 5 σ in this map and a lot of noise below 4.9 σ). The typical octahedral coordination of the bound calcium is formed by interactions with seven O atoms: the main-chain carbonyl O atom (2.42 Å) and OG1 (2.58 Å) of Thr65, OE1 (2.54 Å) and

OE2 (2.52 Å) of Glu67, the main-chain carbonyl O atom (2.40 Å) of Asp130, OD1 (2.42 Å) of Asn133 and OD1 (2.41 Å) of Asp134. The eighth interaction is formed with a buried water molecule (2.53 Å; Fig. 10). This arrangement is similar to that found in other CBM3 structures, *i.e.* PDB entries 1nbc (Tormo *et al.*, 1996), 2wnx (Petkun *et al.*, 2010) and 1g43 (Shimon *et al.*, 2000), with a maximum r.m.s.d. of 0.366 Å for the atoms of the calcium-binding residues.

The calcium-binding residues are located in two loops: Thr65 and Glu67 in the loop connecting β -strands 3 (from the upper sheet) and 4 (from the bottom sheet) and residues Asp130, Asn133 and Asp134 in the loop connecting β -strands 7 (from the bottom sheet) and 8 (from the upper sheet). Therefore, it seems that the role of the calcium ion is structural to facilitate holding the two loops together and as a result to clamp the top and bottom β -sheets.

As mentioned previously, a calcium ion was found in Cbh9A CBM3b^{N126W} but not in Cbh9A CBM3b^{WT}. Despite the absence of the metal ion in the wild-type protein the two structures are almost identical, as can be seen in Table 4. Thus,

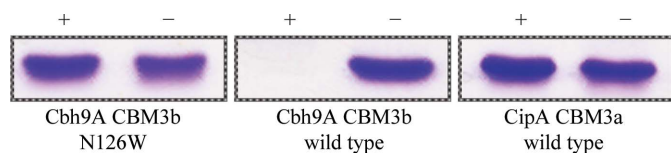


Figure 8

Cellulose-binding assay. The partition of CBM3b bands between the bound (+) and unbound (−) states is shown. See text for details.

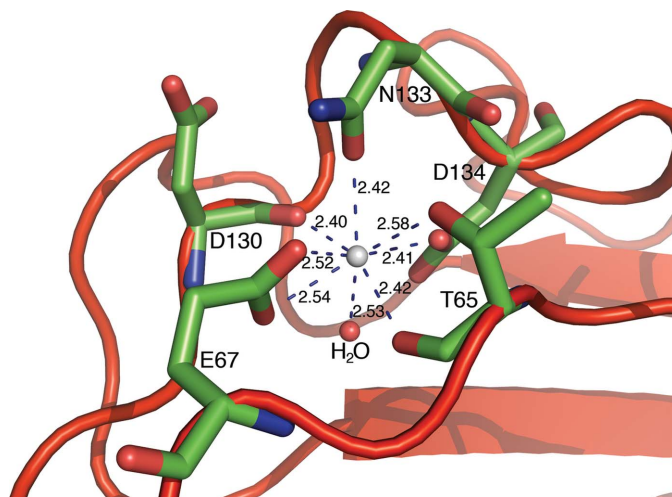


Figure 9

The Ca²⁺-binding site in Cbh9A CBM3b, with secondary-structural elements represented as a cartoon and residues that coordinate Ca²⁺ (shown as spheres) represented as sticks. Distances are in Å.

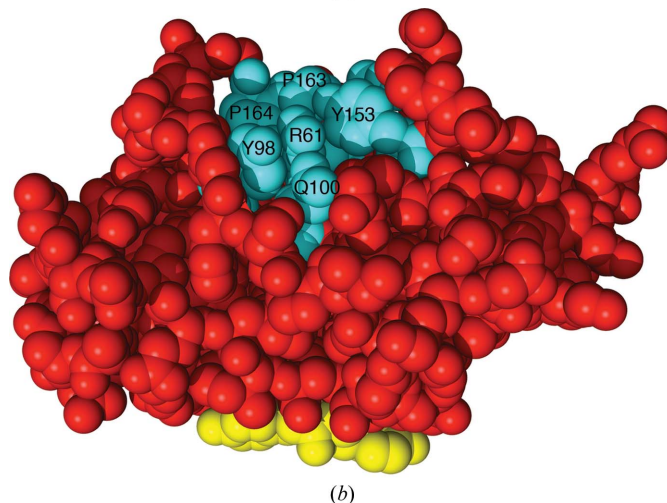
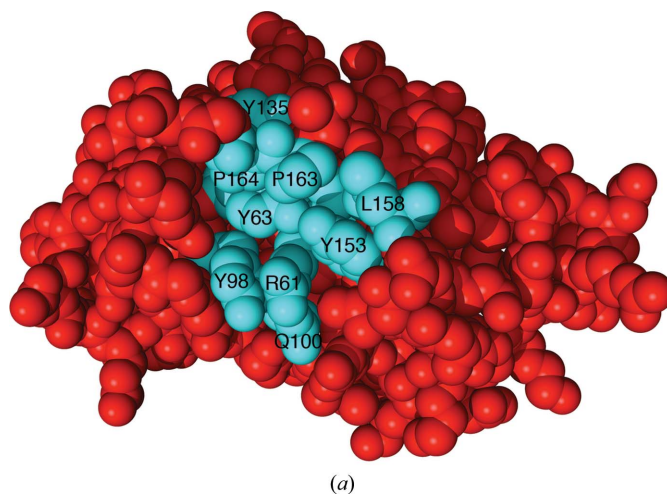


Figure 10

Shallow-groove region of *C. thermocellum* Cbh9A CBM3b represented by a space-filling atomic model. (a) View from the top. (b) Lateral view. Residues forming the shallow groove are highlighted in green; residues forming the cellulose-binding strip are highlighted in yellow.

the calcium ion may play a relatively minor structural role and may not be essential for proper protein folding and conformation. The lack of a calcium ion in the recently published *Bacteroides cellulosolvens* ScaA CBM3b structure (PDB entry 2xbt; Yaniv *et al.*, 2011) provides further supporting evidence for this premise.

3.4. The shallow groove

The Cbh9A CBM3b structure includes a surface topological feature called the 'shallow groove' which is well conserved in all family 3 CBMs (Tormo *et al.*, 1996; Shimon *et al.*, 2000; Petkun *et al.*, 2010).

The shallow-groove region of Cbh9A CBM3b is located on strands 3, 5, 6, 8 and 9 and its surface is delineated by four aromatic rings (Tyr63, Tyr98, Tyr135 and Tyr153), three aliphatic residues (Leu158, Pro163 and Pro164) and three polar residues (Arg61, Gln90 and Thr151) (Fig. 10). The slopes of the shallow groove are formed by the loops between β -strands 5 and 6 and between β -strands 8 and 9. In Cbh9A CBM3b the length of the loop between β -strands 5 and 6 is relatively short, comprising only six residues, as opposed to about 11 residues in *C. thermocellum* CipA CBM3a and *C. thermocellum* Cel9V CBM3b'. The structural consequence of the shortening of this loop is apparently a shallow groove with a smaller estimated depth of 2.8 Å compared with an average of 5 Å in similar molecular structures [*C. thermocellum* CipA CBM3a (PDB entry 1nbc; Tormo *et al.*, 1996), *C. cellulolyticum* CipC CBM3a (PDB entry 1g43; Shimon *et al.*, 2000), *C. thermocellum* Cel9V CBM3b' (PDB entry 2wnx; Petkun *et al.*, 2010) and *B. cellulosolvens* ScaA CBM3b (PDB entry 2xbt; Yaniv *et al.*, 2011)]. The biological significance of this variation and the function of the shallow-groove region are unknown but may be related to interactions with the long cellulosomal intermodular linkers to establish and stabilize the superstructure of cellulosomal elements.

4. Summary

Contradictory results have been obtained in studies of the cellulose-binding properties of *C. thermocellum* Cbh9A CBM3b. Recombinant Cbh9A CBM3b failed to bind crystalline cellulose in our hands (Jindou *et al.*, 2006, 2007), although others (Kataeva *et al.*, 2004) have reported that this CBM3b does bind cellulose. The three-dimensional structure solution reported in this communication revealed a very short putative cellulose-binding region (composed of Tyr77 and Tyr78) which is probably not sufficient for stable cellulose binding. The insertion of a third aromatic residue into the cellulose-binding linear strip by site-specific mutagenesis was effective in introducing cellulose-binding ability into the molecule. The three-dimensional structure of this mutant CBM confirmed that the inserted residue, Trp126, may assume an orientation that brings it into conjunction with Tyr77 and Tyr78, thus increasing the length of the cellulose-binding strip in an appropriate manner. This is the first report to describe the reformation of a native adhesion-deficient CBM to acquire

cellulose-binding features similar to those of other related CBMs.

The biochemical and structural evidence for the inability of Cbh9A CBM3b to bind cellulose and the high energy cost required for the production of additional modules possibly indicate that this CBM3b does not function as a cellulose-binding module *per se*, but possesses other as yet unknown functions.

In another known case of a CBM3b that does not bind cellulose (Petkun *et al.*, 2010), it was suggested that the *C. thermocellum* Cel9V CBM3b serves to further modify the function of the parent GH9 enzyme owing to its close proximity to the adjacent CBM3c and to the enzyme catalytic module. In the multimodular enzyme Cbh9A the CBM3b is much more distal from the catalytic module, so it can be assumed that the CBM3b might possess other functions.

The existence of deviant forms of the CBM3b family (Jindou *et al.*, 2006; Petkun *et al.*, 2010), together with the non-cellulose-binding *C. thermocellum* Cbh9A CBM3b reported here, raises questions about the historical division of the family 3 CBMs, which was based on minor sequence differences. It seems that this distinction may no longer hold for the newly discovered scaffoldin- and enzyme-containing CBMs and therefore subgroups a and b can be combined (Yaniv *et al.*, 2011). A new and more accurate method for the classification of CBM3 subfamilies (based on sequence homology and functional and structural studies) may be adopted for more apparent differentiation between subgroups a and b. Such a classification system should include characterization according to cellulose-binding abilities.

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

References

- Adams, J. J., Pal, G., Jia, Z. & Smith, S. P. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 305–310.
- Adams, P. D. *et al.* (2010). *Acta Cryst. D* **66**, 213–221.
- Alber, O., Noach, I., Rincon, M. T., Flint, H. J., Shimon, L. J. W., Lamed, R., Frolow, F. & Bayer, E. A. (2009). *Proteins*, **77**, 699–709.
- Bayer, E. A., Belaich, J. P., Shoham, Y. & Lamed, R. (2004). *Annu. Rev. Microbiol.* **58**, 521–554.
- Bayer, E. A., Chanzy, H., Lamed, R. & Shoham, Y. (1998). *Curr. Opin. Struct. Biol.* **8**, 548–557.
- Bayer, E. A., Kenig, R. & Lamed, R. (1983). *J. Bacteriol.* **156**, 818–827.
- Bayer, E. A., Morag, E., Lamed, R., Yaron, S. & Shoham, Y. (1998). *Carbohydrases from Trichoderma reesei and Other Microorganisms*, edited by M. Claeysens, W. Nerinckx & K. Piens, pp. 39–65. London: The Royal Society of Chemistry.
- Bayer, E. A., Shimon, L. J. W., Shoham, Y. & Lamed, R. (1998). *J. Struct. Biol.* **124**, 221–234.
- Bayer, E. A., Shoham, Y. & Lamed, R. (2006). *The Prokaryotes*, Vol. 2, 3rd ed., edited by M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt, pp. 578–617. New York: Springer.

- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
- 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.* **D66**, 12–21.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Ding, S. Y., Bayer, E. A., Steiner, D., Shoham, Y. & Lamed, R. (1999). *J. Bacteriol.* **181**, 6720–6729.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Gilad, R., Rabinovich, L., Yaron, S., Bayer, E. A., Lamed, R., Gilbert, H. J. & Shoham, Y. (2003). *J. Bacteriol.* **185**, 391–398.
- Hashimoto, H. (2006). *Cell. Mol. Life Sci.* **63**, 2954–2967.
- Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989). *Gene*, **77**, 51–59.
- Jindou, S., Petkun, S., Shimon, L., Bayer, E. A., Lamed, R. & Frolov, F. (2007). *Acta Cryst.* **F63**, 1044–1047.
- Jindou, S., Xu, Q., Kenig, R., Shulman, M., Shoham, Y., Bayer, E. A. & Lamed, R. (2006). *FEMS Microbiol. Lett.* **254**, 308–316.
- Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- Kataeva, I. A., Li, X.-L., Chen, H. & Ljungdahl, L. G. (1998). *Genetics, Biochemistry and Ecology of Cellulose Degradation*, edited by K. Ohmiya, K. Sakka, S. Karita, K. Hayashi, Y. Kobayashi & T. Kimura, pp. 456–460. Tokyo: Uni Publishers.
- Kataeva, I. A., Seidel, R. D. III, Shah, A., West, L. T., Li, X.-L. & Ljungdahl, L. G. (2002). *Appl. Environ. Microbiol.* **68**, 4292–4300.
- Kataeva, I. A., Uversky, V. N., Brewer, J. M., Schubot, F., Rose, J. P., Wang, B.-C. & Ljungdahl, L. G. (2004). *Protein Eng. Des. Sel.* **17**, 759–769.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Matuschek, M., Sahm, K., Zibat, A. & Bahl, H. (1996). *Mol. Gen. Genet.* **252**, 493–496.
- McPherson, A. (1982). *Preparation and Analysis of Protein Crystals*. New York: Wiley.
- Meza, R., Nuñez-Valdez, M. E., Sanchez, J. & Bravo, A. (1996). *FEMS Microbiol. Lett.* **145**, 333–339.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- 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. & Frolov, F. (2010). *Acta Cryst.* **D66**, 33–43.
- Sakon, J., Irwin, D., Wilson, D. B. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 810–818.
- Shimon, L. J. W., Pagès, S., Belaich, A., Belaich, J.-P., Bayer, E. A., Lamed, R., Shoham, Y. & Frolov, F. (2000). *Acta Cryst.* **D56**, 1560–1568.
- Stout, G. H. & Jensen, L. H. (1968). *X-ray Structure Determination: A Practical Guide*. New York: MacMillan.
- Tina, K. G., Bhadra, R. & Srinivasan, N. (2007). *Nucleic Acids Res.* **35**, W473–W476.
- Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A. L., Sturch, K., Gilkes, N. R., Haynes, C. A., Warren, R. A. & Kilburn, D. G. (1998). *J. Chromatogr. B Biomed. Sci. Appl.* **715**, 283–296.
- Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y. & Steitz, T. A. (1996). *EMBO J.* **15**, 5739–5751.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Xu, Q., Morrison, M., Nelson, K. E., Bayer, E. A., Atamna, N. & Lamed, R. (2004). *FEBS Lett.* **566**, 11–16.
- Yaniv, O., Shimon, L. J. W., Bayer, E. A., Lamed, R. & Frolov, F. (2011). *Acta Cryst.* **D67**, 506–515.
- Zverlov, V. V., Velikodvorskaya, G. A. & Schwarz, W. H. (2003). *Microbiology*, **149**, 515–524.
- Zverlov, V. V., Velikodvorskaya, G. V., Schwarz, W. H., Bronnenmeier, K., Kellermann, J. & Staudenbauer, W. L. (1998). *J. Bacteriol.* **180**, 3091–3099.